

Development of Self-Archiving Tools to Support Archiving, Analysis and Re-Use of Qualitative Data

Agustina Martínez-García

A thesis submitted in partial fulfilment of the requirements
of Liverpool John Moores University for the degree of Doctor
of Philosophy (PhD)

August 2013

To Rob y mi familia: Toni, Agustín y Miguel

ABSTRACT

The potential to share and re-use qualitative archived data has garnered much interest in recent years. This increased attention can be attributed mainly to advances in both data documentation standards and digital archiving technologies, which provide users with the ability to archive, share and disseminate qualitative research materials. However, there remain theoretical and epistemological barriers to and implications for the sharing and re-use of qualitative study data. One way to address these issues is by studying research practices (with practitioners' active involvement), in combination with developing software tools that support digital archiving of qualitative studies.

Semantic technologies, combined with metadata standards and documentation schemas have the potential to enhance qualitative data documentation, archiving and analysis. In fact, it has been established that data documentation is one of the key elements that enables data archiving. The use of appropriate standard documentation frameworks is crucial to data archives' exposure and has a direct impact on the discoverability, search and retrieval of archived data. The technological aspect of this study has been the development of a self-archiving toolkit that makes use of such technologies. The purpose of this work was to allow users, with varying levels of research experience (e.g. from undergraduate student researchers up to more experienced senior researchers) to avail of the benefits offered by qualitative digital archiving.

To complement the technological developments undertaken, the present study also explored the practices of different researchers: undergraduate student researchers, researchers involved in teaching research-oriented modules, as well as senior researchers. This exploration focused on the collection, organisation, analysis and presentation of qualitative data and how these relate to and can be supported by digital archiving to enable researchers to organise, disseminate, and visualise research collections.

ACKNOWLEDGEMENTS

I would like to thank a number of people who helped and supported me throughout my research, the data collection and writing up stages, without whom I would not have succeeded in completing this study. I would especially like to thank the following:

First of all I would like to thank my supervisors Professor Patrick Carmichael and Professor Dave Huddart, who have encouraged me to embark on this research project and have provided me with great support throughout my time as a PhD student. Particularly, I would like to thank Patrick for supporting me during the qualitative practical research stages of the study, encouraging me to use a combined research and technology development approach and to direct my focus towards the potential for using these tools within undergraduate research and teaching/learning research activities. Furthermore, I would like to thank Liverpool John Moores University (LJMU), Faculty of Education, Community and Leisure (ECL) for providing me with the opportunity, resources and range of useful research seminars which I enjoyed attending.

I would also like to acknowledge the work of Arofan Gregory (Metadata Technology) in the design of the model underlying the implemented set of tools, and the advice and support from Louise Corti (UK Data Archive (UKDA)) on key aspects of archived qualitative studies and metadata standards for qualitative data.

Special thanks go to my respondents who gave their valuable time to be interviewed, and who engaged in very useful discussions around their research projects and practices. Great thanks must also go to my colleagues, especially Kate Litherland, who spent many hours reading drafts of my thesis and giving me their valuable comments.

Finally, I would like to say 'a big thank you' to my family and friends, particularly to my parents, Rob and Julia, who have provided great support throughout all of the stages of my work.

SUPPLEMENTARY MATERIALS AND ADVANCE PUBLICATIONS

The self-archiving toolkit implemented in this study has been developed as open source software licensed under the Apache License 2.0¹. The toolkit is freely available for anyone to download, install and use. It is hosted in a Google Projects site², *QuDEx Repository Tools*, which includes detailed documentation and access to the toolkit's source code. Additionally, the toolkit is hosted in an access-controlled LJMU server, so that researchers, and students, within the faculty can make internal use of the tools for testing and research purposes.

A journal article has been published (in advance of the thesis) in a special issue of *Technology, Pedagogy and Education* on the Semantic Web in Education³:

Martínez-García, A. and Corti, L. (2012). Supporting student research with semantic technologies and digital archives. *Technology, Pedagogy and Education*, 21(2), 273-288. doi:10.1080/1475939X.2012.704320

The author is one of the authors of a paper published in the Data Documentation Initiative (DDI) Working Papers Series⁴, which presents a model describing qualitative data. This publication resulted from the work of the DDI Qualitative Working Group, of which the author is a member.

Hoyle, L., Corti, L., Gregory, A., Martínez-García, A., Wackerow, J., Alvar, E., Betancort-Cabrera, N., Gallagher, D., Gebel, T., Haumataki, J., Kuula, A., McEachern, S., and Zuell, C. (2013). A Qualitative Model for DDI. *DDI Working Papers Series*. doi:10.3886/DDIWorkingPaper05

Notes

¹<http://www.apache.org/licenses/LICENSE-2.0.html>

²<https://code.google.com/p/qudex-repository-tools/>

³Accessible online at: <http://www.tandfonline.com/toc/rtpe20/21/2#.UdrVCxaS1vY>

⁴<http://www.ddialliance.org/resources/publications/working/workingpapers>

CONTENTS

Abstract	I
Acknowledgements	II
Supplementary Materials and Advance Publications	III
Contents	IV
List of Figures	VII
List of Tables	X
List of Acronyms	XI
1 Introduction	1
2 Literature Review: On The Potential Of Qualitative Digital Archiving	7
2.1 ICTs, the Knowledge Society and Digital Environments	8
2.2 From Digital Libraries to Digital Archives - Areas of Development	12
2.2.1 Digital Repositories at the Core of Research Archiving	15
2.2.2 Digital Repositories, Open Access and the 'Web of Data'	16
2.3 Archiving and Re-Using Qualitative Studies	18
2.4 Data Management Strategies and Qualitative Archives	20
2.4.1 Data Management Methodologies: Generalising Practices	22
2.5 Technological Barriers to and Enablers of Qualitative Research Archiving	24
2.6 CAQDAS Tools. Towards Research Data Exchange and Archiving	26
2.6.1 CAQDAS Evolution: Supporting Collaborative Research Approaches	28
2.6.2 CAQDAS and Data Archiving	32
2.7 The Implications of Research Archiving on Qualitative Research	36
2.7.1 The Nature of Qualitative Inquiry, Epistemological Challenges of Archiving	37
2.8 Qualitative Research and Data Sharing	44
3 Research Design and Methodology	48
3.1 Context of the Research	49
3.2 Research Objectives	50

3.3	Research Design	51
3.4	Locating the Research Participants	54
3.5	Research Instruments: Data-Gathering Methods	56
3.5.1	Interview Design	56
3.6	Data Storage Methods	58
3.7	Data Analysis	59
3.8	Conclusion	60
4	QuDEX Repository: a Self-Archiving Toolkit for Qualitative Studies	62
4.1	The Toolkit: Technical Overview	63
4.2	Documenting Qualitative Data Collections	65
4.3	Qualitative Collections - QuDEX-based Data Model	67
4.4	Archiving Qualitative Collections	71
4.4.1	Toolkit Visualisation Interfaces	76
4.5	Advanced Qualitative Archiving Models	80
5	Exploring Research Practices and Qualitative Studies Archiving	86
5.1	Exploring Student Research Projects	87
5.2	Researchers' Practices and Perspectives on Qualitative Archiving and Sharing	93
5.2.1	Exploring Research Projects and Digital Archiving	94
5.2.2	Research Practices, Archiving and Data Re-use	98
5.2.3	Teaching Practices: Supporting Undergraduate Research	104
5.3	Exemplary Case Studies Digital Archives	108
5.3.1	Multimedia Case Study of a School	108
5.3.2	BrainAble Multimedia Case Studies	112
5.3.3	Implications for the Self-Archiving Toolkit	117
5.4	The Self-Archiving Toolkit's Support of Research Practices	120
6	Conclusions	131
6.1	Suggestions for Further Work	136
	Appendices	141
A	QuDEX Repository Tools v0.1 – Technical Documentation	141
A.1	General Overview	141
A.2	Technology Background	143
A.3	Technical Documentation	144
A.3.1	Java Libraries	145
A.4	Toolkit Web Applications	153
A.4.1	QuDEX Template Builder	153
A.4.2	QuDEX Collection Manager	180
A.5	Toolkit's QuDEX Support	208
A.5.1	Sample QuDEX Instance	209

B CADQAS Design Approaches	214
C Research Design	224
C.1 Study Topics	224
C.2 Participant Information Sheet	230
C.3 Consent Form	232
C.4 Correspondence with Participants	232
C.5 Interview Schedules	233
C.5.1 Interview Schedule - Students	233
C.5.2 Interview Schedule - Researchers/Lecturers	235
C.5.3 Interview Schedule - Senior Researchers	238
C.6 Interview Transcripts	240
C.6.1 Interview Transcript Student 4	241
C.6.2 Interview Transcript Researcher 3	248
C.6.3 Exemplary Archives: Sample Transcript Interview with Researcher 5	261
C.6.4 Exemplary Archives: Sample Transcript Interview with Researcher 6	281
Bibliography	290

LIST OF FIGURES

2.1	CAQDAS evolution, collaborative research and data sharing	29
2.2	Summary of QuDEx model elements	35
4.1	Data model: object types and their relationships	69
4.2	Example of a ‘work-in-progress’ archive organisation	73
4.3	Framework for data organisation, collection generation and publishing	74
4.4	Spreadsheet template builder: generates spreadsheet templates with appropriate meta-data fields	75
4.5	Resource bundles upload: creates the collection structure in the repository and stores the associated data resources	76
4.6	Data Model used to represent data collections in Fedora Repository based on the QuDEx Model	76
4.7	Faceted collection exploration: the user can explore the contents of a given collection and how the different resources relate to each other	77
4.8	Resource Description Framework (RDF) graph example	79
4.9	Graph visualisation: exploring in detail individual sub-collections or files, along with their related content and analytical information	80
4.10	QuDEx schema and Toolkit support for a standard research data lifecycle	82
4.11	Qualitative Data Exchange Model	83
5.1	Data collection methods employed	114
5.2	Self-Archiving toolkit test archive: sample resources	119
A.1	General overview of tools’ architecture	143
A.2	Components diagram	145
A.3	Fedora Connection classes diagram	146
A.4	FedoraOperationsAPIA classes diagram	147
A.5	FedoraOperationsAPIM classes diagram	147
A.6	Mulgara library main classes	149
A.7	MulgaraOperations class diagram	151
A.8	HttpManagement class diagram	152
A.9	Components Diagram	153
A.10	General use case for the web application	154

A.11 Create Collection Template use case	154
A.12 Generate Collection Template use case	155
A.13 Create Files Template use case	156
A.14 Generate Files Template use case	156
A.15 Metadata Flow activity diagram	158
A.16 Generate Flow activity diagram	159
A.17 Upload Flow activity diagram	159
A.18 Flows and components interaction	160
A.19 Metadata sequencing diagram (scenario of selecting the included metadata elements and then proceed to spreadsheet generation)	161
A.20 Sequence Diagram of Collection Spreadsheet generation	163
A.21 Files Spreadsheet Template sequence diagram	165
A.22 CheckSpreadsheetAction class diagram	166
A.23 GenerateSpreadsheetAction class diagram	166
A.24 ParseDdiAction class diagram	166
A.25 Element class diagram	167
A.26 Elements class diagram	168
A.27 VocabularyElement class diagram	169
A.28 File class diagram	170
A.29 MdSchema class	171
A.30 MdSchemaInstance class	172
A.31 GenerateFlowHandlerClass class diagram	173
A.32 DefaultController class diagram	174
A.33 MetadataService classes diagram	174
A.34 MetadataServiceImpl class diagram	177
A.35 Constants class diagram	178
A.36 RdfFunctions class diagram	178
A.37 ExcelManagement class diagram	180
A.38 QuDEX Repository deployment diagram	181
A.39 General use case for the repository application	182
A.40 Collection Operation Selection use case diagram	183
A.41 New Collection use case	183
A.42 Existing Collection use case	184
A.43 Upload Files Spreadsheet use case	184
A.44 Process Flow activity diagram	186
A.45 Upload Flow activity diagram	187
A.46 Add Relations Flow activity diagram	188
A.47 Flows collaboration diagram	189
A.48 Operation Selection sequence diagram	190
A.49 Collection Update/Select sequence diagram	191
A.50 Adding Collection Relationships sequence diagram	192

A.51	Select Type of Collection Creation sequence diagram	192
A.52	Existing Collection Modification sequence diagram	194
A.53	New Collection Creation sequence diagram	195
A.54	CollectionActions class diagram	196
A.55	EmailActions class diagram	196
A.56	FilesActions class diagram	197
A.57	FilesModel class diagram	199
A.58	ProcessFlowController class diagram	200
A.59	ApplicationContextProvider class diagram	201
A.60	ApplicationServletContextListener class diagram	201
A.61	CustomServletContextProvider class diagram	202
A.62	FileServlet class diagram	203
A.63	Service classes diagram	203
A.64	SpreadsheetServiceImpl class diagram	205
A.65	Constants class diagram	206
A.66	ExcelManagement class diagram	207
C.1	Digital archiving, key areas	225
C.2	Digital archiving, sub-area QDA	226
C.3	Digital archiving, sub-area Technological Themes	227
C.4	Digital archiving, sub-area Data Management Strategies	228
C.5	Digital archiving, sub-area Research Practices	229
C.6	Participant recruitment email	233
C.7	Research process flow	235

LIST OF TABLES

4.1	Documentation of “Collection” and “File” objects	71
4.2	Comparison of Models of Archiving	72
5.1	Description of selected student projects	88
5.2	Research projects overview	95
5.3	Barriers to and benefits of qualitative archiving and data sharing	102
5.4	Kinds of materials used in teaching activities	105
5.5	Student data organisation practices and materials	121
5.6	Researcher data organisation practices and materials	125
5.7	Summary of the contribution of the empirical cases to the toolkit design and development	129
B.1	CAQDAS comparison	216

LIST OF ACRONYMS

- AAC** Advanced Audio Compression
- API** Application Programming Interface
- ASSDA** Australian Social Sciences Data Archive
- CAQDAS** Computer-Assisted Qualitative Data Analysis Software
- CENDARI** Collaborative European Digital Archive Infrastructure
- CESSDA** Council of European Social Science Data Archives
- CMS** Content Management System
- CRUD** Create, Read, Update and Delete
- CSS** Cascade Style Sheet
- CSV** Comma Separated Values
- BNB** British National Bibliography
- BCI** Brain-Computer Interface
- BNCI** Brain/Neural Computer Interface
- DB** Database
- DEXT** Data Exchange Tools
- DDI** Data Documentation Initiative
- DC** Dublin Core
- DEXT** Data Exchange Tools
- DL** Digital Library
- DPE** Digital Preservation Europe
- DRS** Digital Replay System
- ECL** Education, Community and Leisure

- EJME** Enhanced Journals...Made Easy!
- EPSRC** Engineering and Physical Sciences Research Council
- ESDS** Economic and Social Data Service
- ESRC** Economic and Social Research Council
- Fedora** Flexible Extensible Digital Repository Architecture
- FSD** Finnish Data Service
- FOXML** Fedora Object XML
- FQS** Forum Qualitative Social Research
- GESIS** Leibniz Institute for the Social Sciences
- GIS** Geographic Information Systems
- GPS** Global Positioning System
- GT** Grounded Theory
- H.264/MPEG-4 AVC** Audio Video Compression
- HCI** Human Computer Interface
- HRAF** Human Relations Area Files
- HTML** Hypertext Transfer Mark-up Language
- HTTP** Hypertext Transfer Protocol
- ICT** Information and Communication Technology
- IEEE** Institute of Electrical and Electronics Engineers
- iTQL** Tucana Query Language
- JISC** Joint Information Systems Committee
- JSON** JavaScript Object Notation
- KE** Knowledge Ecosystem
- KMS** Knowledge Management System
- LJMU** Liverpool John Moores University
- LLP** Learning Landscape Project
- LOM** Learning Object Metadata

MARC MACHine-Readable Cataloging

MIT Massachusetts Institute of Technology

MM multimedia

MMR Mixed Methods Research

MODS Metadata Object Description Schemas

MVC Model View Controller

N3 Notation3

OA Open Access

OAI Open Archives Initiative

OAIS Open Archival Information System

OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting

OASIS Organization for the Advancement of Structured Information Standards

OECD Organisation for Economic Co-operation and Development

OWL Web Ontology Language

PDF Portable Document Format

PID Permanent Identifier

PREMIS Preservation Metadata Implementation Strategies

QDA Qualitative Data Analysis

OER Open Educational Resource

QG Qualitative Geography

QSS Qualitative Social Sciences

QuDEX Qualitative Data Exchange

RELU Rural Economy and Land Use Programme

RISK Rapid Integration of Skills and Knowledge

RDF Resource Description Framework

RDFS RDF Schema

REST REpresentational State Transfer

SA secondary analysis

SAS Statistical Analysis System

SIMILE Semantic Interoperability of Metadata In like and Unlike Environments

SKOS Simple Knowledge Object System

SPARQL SPARQL Protocol and RDF Query Language

SPSS Statistical Package for the Social Sciences

SS Social Sciences

TAG Technical Architecture Group

TEL Technology-Enhanced Learning

TEI Text Encoding Initiative

UKDA UK Data Archive

UML Unified Modelling Language

URI Uniform Resource Identifier

URL Uniform Resource Locator

VLE Virtual Learning Environment

VRE Virtual Research Environment

W3C World Wide Web Consortium

WADL Web Application Description Language

WG Working Group

WWW World Wide Web

XACML eXtensible Access Control Markup Language

XML eXtensible Mark-up Language

INTRODUCTION

The potential to share and re-use qualitative archived data has garnered much interest in recent years. The increased attention can be attributed mainly to advances in both data documentation standards and digital archiving technologies (Corti, 2005). However, digital archiving of qualitative studies raises important scientific, ethical and legal issues, including, but not limited to considerations surrounding data anonymisation, data access-control policies and restrictions, and confidentiality (Bishop, 2009). To address the challenges of archived qualitative data, the need to develop new methodologies and models for archiving, based on a combination of social science methodology and archival descriptions, then becomes apparent (Corti, 2007).

The technological aspect of this study has been the development of a self-archiving toolkit that makes use of semantic technologies, digital repositories and data documentation standards. The purpose of this work was to allow users, with varying levels of research experience (from undergraduate student researchers up to more experienced senior researchers) to avail of the benefits offered by qualitative digital archiving. The design and development of this toolkit has been informed by analysis of a range of research practices that are used in primary analysis, archiving and re-use of qualitative data. This has been supported by empirical data gathered from interviews conducted with a range of participants, and by active involvement in an international Working Group (WG), the Qualitative Data Model WG¹, as part of the DDI Alliance² to develop international standards based on eXtensible Mark-up Language (XML) for metadata describing social science data. Three different groups of participants, referred to hereafter as cohorts³, were involved in this study: undergraduate students conducting small-scale research projects; experienced researchers conducting qualitative and mixed-methods studies; and senior researchers involved in large-scale research projects which led to the development of archives. The active participation in the above-mentioned DDI working group allowed for an exploration of a set of use cases from various research archives, which were presented as part of the work of the group. These use cases helped to inform the initial design of the toolkit's data model and the archiv-

ing processes that users are required to follow when using the toolkit. Qualitative interviews conducted with researchers with varying levels of research experience informed subsequent refinements and further development of the initial, generic design of the toolkit. The design of the self-archiving toolkit was thus informed by both generic archiving practices as well as those more specific to actual researchers, thereby ensuring a toolkit that could support researchers' existing data documentation, organisation and archiving practices and that would also be applicable to a wider community.

One of the most common uses of existing research archives is that of secondary analysis (SA). An in-depth exploration of how digital archives can facilitate secondary analysis as well as the epistemological and ethical considerations surrounding secondary analysis was undertaken. The use of existing qualitative archives for the purposes of research data re-use and SA has been widely discussed and indeed is still an ongoing debate (Hammersley, 1997; Heaton, 2004a; Hammersley, 2010; Mauthner et al., 1998; Parry and Mauthner, 2004); and it has an impact on the processes of designing and implementing qualitative digital archives. Therefore, a substantial part of the literature review that was carried out in the course of the present study has facilitated a better understanding of researchers' perceptions about primary research data re-use and whether the latter could constrain or limit qualitative data archiving. Alongside the technological advances in the areas of digital archiving and data documentation, their challenges and limitations, we have to bear in mind the theoretical, epistemological barriers and implications of re-using, sharing and conducting secondary analysis of qualitative studies (Heaton, 2004a). Moreover, the need for qualitative data archiving is still being questioned by researchers, who are resistant to archive their research materials for a variety of reasons. They highlight the nature of research projects; researchers' epistemological positions; preservation of confidentiality and informed consent; scepticism about usefulness to secondary researchers; criticism of the research methods (i.e., researchers' exposure); copyright and ownership of the research materials, among others (Corti et al., 1995; Cheshire, 2009; Moore, 2007). Moreover, the initial literature review for this study has revealed more specific issues related to secondary analysis and qualitative data re-use and archive. Another important issue is the availability of datasets from primary studies, which is quite limited but has increased considerably in the last 10 years (Corti and Thompson, 1998; Thompson, 2004; Kuula and Borg, 2008). Although a large number of datasets, primarily historical, are already archived, the lack of standards when documenting and archiving makes it difficult to discover, access and re-use them. In this respect, new technologies, especially the ones related to online access to digital resources, data description/documentation and digital archiving can potentially provide a solution to those issues. This research is located where there is a demand or desire to archive, analyse, and re-use data, but current technological solutions - particularly across networks - are inadequate. A number of studies with a focus on the use of electronic environments, more concretely Virtual Research Environments (VREs), have highlighted both barriers and enablers to their use for research purposes (Carmichael and Procter, 2006; Laterza et al., 2007; Wilson et al., 2007; Carusi and Reimer, 2010). Such online tools provide users with shared storage, collaboration and communication tools, that are combined with

secure and access-controlled environments. However, challenges such as data access, difficulty with learning software and technologies, lack of appropriate skills, extensive training needs and ready access to technical support remain present (Carusi and Reimer, 2010). Additionally, whilst VREs represent an intermediate solution between online learning environments (with a focus on supporting teaching and learning activities rather than research tasks) and specialised research data archives, they provide limited support for key functionalities to facilitate data sharing and re-use. That is, a combination of data storage capabilities with appropriate documentation frameworks. This is particularly important for qualitative research data archiving since the nature of these data requires the use of open descriptive standards (Kuula, 2000) that include suitable terms for describing both research data, and perhaps more importantly, the context in which the research was produced.

While the implementation of qualitative digital archives has, to date, been primarily concerned with long-term data preservation, and more recently with sharing and re-use, it has been widely recognised that qualitative digital archives have many potential applications in teaching and learning environments (Corti and Bishop, 2005; Bishop, 2012). Students can make use of archived classic studies to gain a better understanding of the rationale for a particular research design or data collection approach, can learn to appreciate the complexity of data analysis and can critically evaluate particular research strategies, approaches to ethical issues or fieldwork methods. A good example of this type of strategy is the one followed in the UKDA⁴ where archives of qualitative classic studies are used to support students studying Masters' courses on research methods. This still, however, positions students not as producers of research, but as users of existing resources who learn through engagement with expert discourses. However, if they are encouraged to undertake projects, build portfolios of their own work and contribute to knowledge-building communities, they also need to be supported in contributing to archives and engaging with the same challenges that other researchers face. In this respect, an online self-archiving toolkit has been developed as part of the present study. The toolkit has been designed to help users with a requirement for archiving, searching and disseminating collections of qualitative data, regardless of the type of media. While useful for any kind of research community, these tools have been specifically designed to be sufficiently accessible that they could be used by non-experts such as individual students conducting small-scale research projects. They provide easy ways of organising and archiving research data so that the created collections can be visualised and exposed in ways that enable integration with the increasing 'web of linked data', as well as being presented in support of dissertations. Interviews with undergraduate students conducting research projects have enabled the exploration of the kinds of materials collected by the students, their practices and the processes of organising and archiving of their data. Interviews with researchers involved in the development of students' projects, either teaching or supervising students, have also enabled the exploration of the kinds of materials the participants use to support teaching research modules to undergraduates, and teachers' perceptions about research archiving and its potential to be used with teaching purposes. The exploration of both the materials gathered by the students in the course of their projects and the materials used

by teachers to support teaching research methods allowed for a refinement of the toolkit's data model to ensure that the documentation of such materials was fully supported. More generally, this empirical work has allowed an exploration of how student projects might be supported with the self-archiving toolkit developed in this study.

In relation to researchers' practices of data archiving, a substantial part of the study has investigated actual data management practices and archival procedures carried out by researchers during qualitative research projects. In this respect, two exemplary archives of qualitative studies were explored to gain a better understanding of the research approaches followed during the archiving of the materials, to identify how the archive design and implementation was performed and to discover the role of technologies, and their impact, within those processes. Additionally, ongoing work carried out in the aforementioned Qualitative Data Exchange WG has highlighted important unresolved issues within certain areas of qualitative archiving, which could be resolved with the use of new digital repositories and semantic technologies. Those issues focused primarily on: a) the implementation of appropriate data documentation frameworks, b) analysis of the role of documentation standards in Social Sciences, c) integration of those documentation standards with archiving systems and d) the basis for the development of tools facilitating data sharing, re-use and dissemination. There has been agreement that data documentation is one of the key elements of data archiving (Kuula and Borg, 2008; Vardigan et al., 2008). In Open Archival Information System (OAIS) (ISO OAIS Reference Model for OAIS) terms, data documentation refers to the 'representation information', that is, "The information that maps a Data Object into more meaningful concepts" (Consultative Committee for Space Data Systems [CCSDS], 2012, p.1-14). Examples of data documentation elements for quantitative data include the 'codebook' (technical document to understand and interpret the numeric codes in a survey data file) for a numeric survey file or descriptions of the sampling procedures used. For qualitative data, information about the methodology of a study or contextual information describing data collection methods, research instruments, or analytical elements like 'codebooks' for theme-based analysis, are examples of data documentation. The use of appropriate and standard documentation frameworks is crucial in relation to the exposure of data archives and it has an impact on the discovery, search and retrieval (data access) of the available datasets. One of the most popular and consolidated standards for the compilation, presentation, and exchange of documentation for datasets in the social sciences is the DDI, which provides a framework to document primarily quantitative datasets. The most important output from the work carried out in the working group has been the development of a robust XML-based schema for qualitative data exchange (Hoyle et al., 2013), which can be integrated with the DDI standard. It is expected that this schema will serve as the basis for the development of tools for visualising, analysing and disseminating qualitative research data. These kinds of models, or schemas, and their associated tools are necessary to support complex and large-scale research projects. For such studies, the archiving processes, along with the design and implementation of data management plans, requires researchers to perform an extra set of tasks in addition to those directly related to their research. In this respect, the self-archiving toolkit implemented in this study

seeks to provide users with a simpler means of building a digital archive than would be the case for users involved in larger, more complex research projects. At the same time, the toolkit seeks to facilitate the documentation of qualitative data in standard and consistent ways, similarly to those more complex schemas, so that the archived qualitative data can be shared and re-used.

The dissertation is structured as follows. The initial literature review summarises the investigative research around four different areas. The first section explores key Information and Communication Technologies (ICTs), and then it points up how they have changed the digital landscape more generally, and education, libraries and archives accordingly. The second section explores, firstly, a range of technologies (digital repositories and semantic technologies) to support digital archiving, and secondly, it explores digital archiving practices. Although linked to research practices, it focuses on research data management approaches. The third section reviews the evolution of Computer-Assisted Qualitative Data Analysis Software (CAQDAS)⁵ packages focusing on the functionalities of a set of packages. As part of this evolution of Qualitative Data Analysis (QDA) software, there has been a shift from individual use and tools design targeting specific research approaches to networked, collaborative tools that have the potential to be integrated with archiving processes, facilitating data sharing and re-use. The latter highlights a number of challenges that existing software packages are facing and it also opens up numerous possibilities for the design of qualitative archiving models that can be integrated with CAQDAS packages. These aspects are also discussed in Chapter 2. The last section provides critical background commentary on qualitative inquiry, methodologies of qualitative research, and the relationships and implications they have on secondary analysis and the re-use of qualitative archived data.

Chapter 3, 'Research design and methodology', describes in detail the research methodology used for this study, introducing the main objectives, the design approach of the self-archiving toolkit developed in this study, followed by sections describing the research instruments designed, data collection stages and the analysis approach for the data gathered. Chapter 4 firstly provides a technical overview of the implemented self-archiving toolkit and its underlying technologies. Secondly, it introduces the context for data documentation and the design decisions made with respect to the implemented toolkit's underpinning data model along with an overview of the key issues around data documentation of qualitative materials. Thirdly, the toolkit's data model is described in detail and lastly, an enhanced model for qualitative data archiving and exchange is presented and compared with both the Qualitative Data Exchange (QuDEX) schema and the self-archiving toolkit's data model. This enhanced model is part of the modelling and implementation work carried out within the DDI Qualitative Data Model WG, of which the author is a member.

Chapter 5 discusses the findings from the interviews conducted as part of the empirical work and is structured into three different sections. The first section discusses the interview data from undergraduate students who conducted small-scale research projects. It focuses on as-

pects such as students' approaches to research; the issues encountered at the various stages of their projects; how they structured, organised and subsequently analysed their data; and the potential for using archiving tools to support the development of students' projects. The second section provides a descriptive account from interviews carried out with active researchers who were also involved in research-oriented undergraduate modules. A number of aspects were covered highlighting: prior experiences from research projects; their perceptions of qualitative archiving and secondary analysis; and their experiences of teaching (supervising) undergraduate students' research, with a focus on the kinds of materials used to support teaching and the issues students faced. The third section of the chapter provides a detailed analysis of two selected case studies of exemplary qualitative, mixed methods digital archives. Researchers' practices and the design/development approaches to archiving; issues faced during those processes, as well as the role of technologies in the archive production process as a whole are covered. Lastly, Chapter 6 provides a reflective account of the work that has been carried out. It reviews areas where the self-archiving toolkit implemented in this study could potentially be used and highlights those where work remains to be done, or where potential tensions have been identified. The chapter also summarises those aspects of the toolkit that could be enhanced or extended in the future.

Notes

¹<http://www.ddialliance.org/alliance/working-groups#qdewg>

²<http://www.ddialliance.org/alliance>

³The term cohort is used here to distinguish between three groups of participants: undergraduate students conducting small-scale research projects, more experienced researchers who are also involved in teaching undergraduate-level research modules, and senior researchers involved in qualitative archiving projects. In the context of Social Sciences research, the term cohort usually implies a longitudinal aspect of a study. In the case of this study, the term has been used to distinguish between groups who "share a common characteristic" (Cohort [Def. 1], nd) and does not imply that they were involved in a longitudinal study.

⁴The UK Data Archive is the curator of the largest collection of digital data in the social sciences and humanities in the United Kingdom. Qualidata, established in 1994, emerged as part of a project to create a national archiving centre for qualitative fieldwork data. At present, Qualidata is a specialist service of the Economic and Social Data Service (ESDS), which provides access and support for a range of social science qualitative datasets, promoting and facilitating increased and more effective use of data in research, learning and teaching. UK Data Archive: <http://www.data-archive.ac.uk/>

⁵CAQDAS is a term created by the directors of the CAQDAS networking project at the University of Surrey, Guilford, UK. Project site: <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas>

LITERATURE REVIEW: ON THE POTENTIAL OF QUALITATIVE DIGITAL ARCHIVING

There is widespread research interest in ICTs and the interactions between these and society. The importance of innovation on ICTs has been agreed, since they are key defining technologies. Their significance is not only technology related but it also spreads across social and economic factors (Dutton, 2004). We live in a technology-driven, information society in which getting high quality information quickly is very important. The early advances in computing and the different information technology areas, of which one of the most visible manifestations has been the Internet (International Institute for Educational Planning, 2007), have made it possible to find and distribute information, moving not only towards an ‘information society’ but to a ‘network society’. Castells (2000) argues that

“the information society is the new mode of human existence, in which the production, recording, processing, and retrieving of information in organised networks plays the central role”. (Castells, 2000)

The dominant functions and processes in the information age are increasingly organised around networks, in which structures and activities are organised around networks of information that is electronically-processed (Castells, 2005). More recent technologies, such as ubiquitous networks (cloud computing), Web 2.0 or Social Web, and linked data approaches, now offer many more possibilities for sharing, archiving, retrieving, combining and generating new knowledge. The increased use of new technologies in work, leisure and daily life has been one of the factors that has raised the importance of ICT for teaching and learning. Such importance on ICT for learning has been reinforced, both in terms of digital competences as an essential life and career competence, as well as the enabling role of ICT for creativity and innovation (E&T 2010 Programme, 2010). For instance, the number of international policy developments is increasing. Examples include the policies developed by the European Commission, more concretely

the work of the *Cluster on ICT for learning* which is part of the Education & Training (E&T) 2010 work programme¹, and the i2010 initiative². This initiative groups all the policies around information society, focusing on the development of three broad areas: development of a borderless information space and promoting national level markets for e-communications and digital services; innovation and investment on research and ICTs; and promotion of an open, transparent and accessible information society, transitioning to a 'knowledge society' (United Nations Educational, Scientific and Cultural Organization, 2005). However, Castells (2005, p. 16) argues that most of these policies have placed much of the emphasis on technological upgrading and enhancement of research capabilities, while indeed the improvements in productivity, learning or creativity have been very limited: "introducing technology *per se* does not ensure productivity, innovation or greater human (social) development". The latter also applies to Education, i.e. the introduction of the technologies is not sufficient without developing new strategies of teaching and learning, which aim to prepare learners for participation in a networked/information society in which knowledge will be the most critical resource for social and economic development (Lehtinen et al., 1999). The different implementations of ICTs along with new ways of social interaction play a very important role in Education, as they are contributing to recent theoretical approaches to teaching and learning, such as 'innovative knowledge communities' or 'knowledge-creation' learning approaches, in which knowledge creation is not originated by creative individuals but takes part in certain kinds of social practices of working for advancing knowledge (Paavola et al., 2002, p. 7). The development of social structures and practices that support desired interaction between participants is central to facilitate educational change through ICT tools.

2.1 ICTs, the Knowledge Society and Digital Environments

The wider societal impacts of technology have been explored in descriptions of "knowledge economy" (Peters, 2007; Vallima and Hoffman, 2008); "knowledge society" (United Nations Educational, Scientific and Cultural Organization, 2005; Weert, 2006); "learning society" (National Committee of Inquiry into Higher Education, 1997; Laurillard, 2002) or even "network society" (Castells, 2000, 2005). The role of educational systems and processes as elements within these new social formations are described by Weert (2006); Vallima and Hoffman (2008). In a knowledge society supported by ICT, the ability to locate, classify and sort information is essential (United Nations Educational, Scientific and Cultural Organization, 2005). One marker of a knowledge society is continuous innovation that demands lifelong learning, knowledge development, and knowledge sharing (Weert, 2006). ICTs allow learners to seek information and develop knowledge at any time and any place where access is available and unrestricted. The evolution of the web has created a shift from a web centred on information access and retrieval where users are *passive consumers* of information (Web 1.0) to a web aiming at the concept of information being socially and collaboratively constructed by users (Web 2.0 (O'Reilly, 2006)). Additionally, innovation in digital technologies and mobile devices offers individuals a means to connect anywhere, anytime where digital technologies are accessible.

Innovation in ICT also represents a challenge for the internal development of Higher Education, particularly concerning three elements of 'learning society' (teaching, research and service), owing to such a rapid change on information technologies: "institutions are not only producing and supporting technological innovations but are at the same time intensive users and subject to the limitations of ICT" (Vallima and Hoffman, 2008, p. 278). For example, evolving ICTs in education have made available Virtual Learning Environments (VLEs) and new sources of information, which have an impact on students' learning processes. In the early stages, most of the e-learning platforms were mostly content-management oriented (with a focus on content delivery) and learning activities were poorly designed (Britain and Liber, 2006, p. 28), with a clear separation between teaching/learning theories; pedagogies and the functionalities of the different platforms. That is, the online learning platform was seen as a place-holder for teaching/learning materials that were provided by the teacher (a knowledge-transmission model rather than knowledge-construction) and the role of the student was as a knowledge consumer. While some of the early learning environments maintained a model of 'students as consumers' where students were seen as the customers of technology-facilitated knowledge authored and delivered by teachers, more recent Technology-Enhanced Learning (TEL) environments support approaches in which the teacher is not simply a transmitter of knowledge, but is more like a mediator of teaching and learning activities that are co-constructed. Such environments present the characteristic of involving interactions with other people, i.e. learning environments supported by technology, like other learning environments, are places where learning takes place, in part, through social interaction (Winn, 2002). In this respect, although Web 2.0 applications are not developed specifically for educational purposes, they have a number of affordances that can make them useful in teaching and learning environments. Web 2.0 applications, such as social networks, wikis, and blogs, focus on social connectivity, providing the space for collaboration and sharing of information to support the networks for social learning (Shaohua and Peilin, 2008). VLEs, as well as research-oriented environments like VREs, incorporate such kinds of collaboration/social applications, and support more recent ideas in relation to electronic networks, such as 'communities of inquiry' (Wilson et al., 2007), or 'networked communities' (Hakkarainen et al., 2004). These ideas expand previous learning models such as 'communities of practice', developed by Lave and Wenger (1991) - see also (Wenger, 1998, 2009). In contrast to communities of practice, which are based on social learning theories about cognition as situated and learning as participation, communities of inquiry are more focused on knowledge-creation as opposed to knowledge-transfer through socialisation (Hakkarainen et al., 2004). VREs and VLEs provide users with the tools and technologies to support them doing their tasks (whether teaching and learning, or research oriented), to interact with other users (collaboration), and to make use of/generate resources (Voss and Procter, 2009; Peters et al., 2012).

So far, the growth of ICT has significantly increased the global capacity for creation of raw data and the speed at which it is produced. Web 1.0, and more concretely the Internet, has facilitated access to a wide variety of information. However, this capacity for producing, us-

ing and sharing data does not necessarily result in knowledge creation. For knowledge creation to take place, reflection is required to create awareness, meaning and understanding (Paavola and Hakkarainen, 2005). In this respect, a more recent idea discussed in relation to electronic networks, is that of 'innovative knowledge communities' (Hakkarainen et al., 2004). Knowledge advancement, discovery, and innovation is a process of creating and developing new materials (through collective inquiry), in which information is disseminated alongside practices, to construct a *collective memory*. That is, "knowledge that is distributed among members of a community although it might have consisted, originally, in individual observations, experiences or insights... Common experience becomes encapsulated in shared narratives, practices of work and formal databases that no longer depend on individuals" (Hakkarainen et al., 2004, p.73). The introduction of inquiry-based and collaborative teaching and learning processes, supported by learning environments, in which students collaborate actively with teachers and tutors in the construction of knowledge, helps to rethink and reconstruct the relationship between teaching and research within the university more generally (Neary and Winn, 2009, p.193). In this respect, Neary and Winn (2009) propose a model of 'students as producers' as an alternative to a model of consumers in which there is an emphasis on academic delivery to students (e.g. acquisition of competencies for the labour market). In the latter model, the teaching and learning processes are very directed rather than aiming to set the basis for students production of new knowledge artefacts, or promoting research activities and contribution to academic communities. In this context, there are numerous attempts (Paavola et al., 2002; Healey and Jenkins, 2009; Lakkala et al., 2008) to explore pedagogical models, for 'students as producers' supported by the use of collaborative technologies. These models aim at developing more innovative communities of inquirers and knowledge-building communities within educational systems, in which knowledge creation occurs in community-based environments rather than from creative individuals. A very good example of knowledge-building projects carried out by undergraduates is the Finnish project "Citizen Memory Project" (Hakkarainen et al., 2004) where students constructed an online database storing local history information that could be consulted and used by other students, teachers/researchers, or the general public. Students were engaged with local communities in an inquiry project to collect information about local history. Other examples include Slotta's 'Knowledge Community and Inquiry' model, where students in a classroom work together to create 'persistent knowledge resources' which then serve as sources of materials and inspiration to subsequent inquiry projects (Slotta, 2010). A further example is Rapid Integration of Skills and Knowledge (RISK) which is concerned with moving from a top-down, teacher-delivered curriculum, to one of distributed teaching and learning in which everyone in the classroom is both a teacher and a learner (Kinney, 2012). Approaches like these facilitate a change in the roles of both teachers and students: while the teacher maintains his role in designing the focus of the teaching-learning processes, he also becomes a co-learner and the students become co-researchers, constructing resources in collaborative ways, with the participation of the teacher, by using available technologies, web-based resources and their own discoveries.

Web 2.0 applications provide the space for collaboratively constructed information, as well

as open access to this information. While universal access to information is key within the knowledge society (United Nations Educational, Scientific and Cultural Organization, 2005), for both research and learning it is not just access to content but access to trusted resources via discovery tools that is important (Harris, 2012). Students already use collaboratively constructed content, such as Wikipedia, as a starting-off point. Users are already aware that such content is not always accurate, however they know that there may be links to peer-reviewed articles at the end. In this respect, involvement from educational institutions (schools, universities) to build up a culture of knowledge sharing may contribute to not only provide open resources but also to facilitate access to quality-controlled, relevant information. This is the objective of the Open Educational Resources (OERs) movement, which seeks to help equalise access to knowledge and educational opportunities, by offering digitised materials freely and openly to educators, students and independent learners to use and re-use for teaching, learning and research (International Institute for Educational Planning, 2007). In addition to OERs, libraries and educational/research organisations are moving towards Open Access (OA). OA is defined as “free, immediate, permanent online access to the full text of research articles for anyone, web-wide” (Harnad, 2012). It has an impact on future research, teaching and learning (Harris, 2012) and it also encompasses changes in research patterns, e.g. simplifying information access. The discoverability of OA content is key to its usefulness, and in meeting student requirements, e.g. OA may help students exploring what resources other students are using, or if supported by appropriate systems, networks of OA related content could be established on a global basis rather than locally or per institution.

The Semantic Web, also referred to as *Web of Data* or Web 3.0, is conceptualised as “an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation... data on the web [is] defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications” (Berners-Lee et al., 2001). The Semantic Web is contributing towards moving both information and knowledge creation forward in their capacities to use intelligence to digitally create meaning independent of user-driven ICT (Shadbolt et al., 2006; Magnan et al., 2007; Harvey, 2010). Independently from providing open access to resources, another key issue is the current access to different types of resources, i.e. data lives in different systems (digital repositories, information management systems, or databases) and formats. This has important implications for data discoverability and information processing, so that integrated access to data coming from multiple sources, such as user-generated content, institutional repositories, digital archives, or electronic libraries, can be provided. In this respect, there are a number of key developments that are contributing to overcoming such issues:

- Data representation and documentation standards, such as RDF³ or ontology languages like Web Ontology Language (OWL)⁴, in combination with tools to convert information between different formats and visualisation tools, facilitate integrating user-generated content with those coming from larger data providers such as publishers, libraries, or institutional repositories.

- Digital repositories and semantic databases such as triplestores⁵ enable storing heterogeneous information in an integrated manner, regardless of original sources or formats. As such, they offer opportunities for 'legacy' systems, databases, catalogues and other collections to be integrated into semantic web and linked data applications, such as query portals, or visualisation tools. Examples of such portals include the DBPedia Query Editor⁶, or the British Library's British National Bibliography (BNB) SPARQL Protocol and RDF Query Language (SPARQL) portal⁷.

There are interesting applications that make use of standard data formats and documentation vocabularies, facilitating a shift from collaborative applications (underpinned by Web 2.0 technologies) that are often domain-specific and operate in closed environments (Magnan et al., 2007). For Magnan et al. (2007), semantic technologies can enhance Knowledge Ecosystems (KEs)⁸ by transforming unstructured data into structured data which has meaning and relationships attached; and by facilitating interoperability to store and share knowledge in any form. For digital libraries and archives, the use of appropriate, and open, standard documentation frameworks is crucial in relation to information exposure and has direct impact on the discoverability, search and retrieval of the available datasets (Gartner, 2008), i.e. the adoption of an open set of standards, linked in a clearly-established manner will ultimately make the process of creating digital libraries much easier for all concerned. Libraries, such as the British Library, are publishing open metadata about their records, including selective coverage of both printed and electronic publications (e.g. the above mentioned BNB⁹). Although such efforts include mostly making available bibliographic information, by using metadata records, rather than providing access to the data themselves, libraries are starting to explore ways in which they could make their data available to wider, and potentially non-library, audiences. A very good example of this is the British Library 'Research Explorer', which aims at migrating library formats to cross-domain standards such as RDF and linked data formats, and adopting licensing for enabling a wider re-use of data (British Library, 2012). While libraries are changing (Harris, 2012; British Library, 2012) they continue to play important roles within emergent networked information societies, acting as what Castells (2005) describes as 'hubs' rather than simply repositories. Web 3.0, semantic or linked data technologies, in combination with archiving tools such as digital repositories or digital libraries have the potential for the realisation of the former. Key linked data developments, along with digital archiving tools and data documentation standards are described more in-depth throughout the rest of this chapter.

2.2 From Digital Libraries to Digital Archives - Areas of Development

Libraries play a very important role in modern *information society*. They are knowledge organisers - cataloguing, classifying and describing knowledge - and enable the preservation of heritage and history. ICTs, the Internet and web technologies provide a great opportunity to circulate this heritage and knowledge by making them available on a global scale. There are multiple views of what a digital library is: e.g. an electronic source of (digitised) materials; a specific digitised collection or publisher's collection; a scientific repository - using a wide variety of

software technologies and data models. There are a large number of early definitions of the term Digital Library (DL), mostly coming from research and information science communities owing to the fact that research and practice are conducted concurrently at each stage of the continuum from basic research on digital libraries to their implementation and usage (Borgman, 1999). One of the definitions that is particularly of interest is from Borgman et al. (1996) who define a digital library as:

a set of electronic resources and associated technical capabilities for creating, searching and using information. In this sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds; static or dynamic images) and exist in distributed networks. The content of digital libraries includes data, metadata that describe various aspects of the data (e.g. representation, creator, owner, reproduction rights) and metadata that consist of links or relationships to other data or metadata, whether internal or external to the digital library.

Whilst the aforementioned definition is not a recent one, it still includes current key areas of development within digital libraries. The full cycle of 'creating, searching and using information' is covered rather than simply collecting and storing information. Additionally, the contents of a DL include not only data but metadata which is related to the key area of data documentation. Lastly, the concepts of networked and distributed access to information are also important functionalities of a DL, along with the ability to link information, both externally and internally. The latter relates to the area of information accessibility, dissemination and exchange. Most of the present technology developments within digital libraries, and more generally within information management systems, are concerned with:

- *Digitisation and standardisation of data formats.* This area covers developments concerned with the definition and evaluation of open or standard machine-readable data representation formats, that most software is capable of interpreting, thus facilitating data interchange and transformation. Along with the standardisation of data formats and digitisation, which have a clear impact on data preservation, developments on the area of data documentation, such as open metadata and descriptive data documentation standards, are key to data discoverability within networked environments.
- *Preservation and archiving of digital resources.* With the advances in the implementation of e-libraries or digital libraries, numerous projects have appeared, which cover different areas within the development, evaluation and implementation of digital preservation strategies and the modelling of digital archiving (and the development of software systems to support archiving). For example, the CEDARS project¹⁰ in the UK addressed the main methodological and practical issues of digital preservation; and Digital Preservation Europe (DPE)¹¹ focused on the need to improve coordination, cooperation and consistency in current activities to guarantee effective preservation of digital materials.
- *Dissemination of information resources and semantic technologies.* This area is related to the implementation of technological services that enable the storage of heterogeneous

information in an integrated manner (semantic databases or triplesores), to convert information between different formats, and to access information in networked environments (visualisation tools).

The toolkit developed in this study incorporates elements from the three areas described above. In relation to the first two areas, the data model that underpins the toolkit combines elements from standard documentation vocabularies and relational ontologies that support the documentation, organisation and storage of information resources. Collections of materials that are created with the toolkit are preserved in a digital repository. Additionally, the digital repository is enhanced with additional technology developments that make use of and contribute to implementations within the third area. Those are data visualisation and dissemination tools which can facilitate data sharing and exchange. Whilst the use of digital repositories clearly contributes to the preservation aspect of digital archives, the aim of engaging researchers, students and the general public with existing archived data so that resources can be shared and re-used, is equally important. Thus, visualisation tools underpinned by data expressed and documented in standard and open formats, have also been implemented as part of the developed toolkit.

While preservation and access control are still two key areas within digital archiving of research collections, linked data technologies (W3C, 2013) and visualisation tools open up new ways of disseminating digitally archived materials and promote open access, that is, linked data facilitates data exchange between information systems and connects these materials with other related sources. There has been a growing realisation by educational and governmental organisations, of the power of linked data for exposing, sharing, and connecting data and information available online, using uniform and standard mechanisms (Berners-Lee, 2009; Bizer et al., 2009) like Uniform Resource Identifiers (URIs)¹². Linked data approaches aim to use the Web to connect related data that was not previously linked, or using the Web to lower the barriers to linking data currently linked using other methods, so that data is published in ways that allow people to use them easily and reliably (HM Government, 2012). There are a number of examples of existing uses of linked data approaches, such as direct.gov.uk (as a provider of government policy datasets, national statistics) or UK Government Linked Data¹³, to improve access to and engagement with official and trusted data, by multiple and even new audiences, such as the general public. In the case of educational institutions, these initiatives aim to develop networks of reliable, trusted shared educational materials. For example, the UK Government Linked Data is not only making available a large and increasing number of datasets, but it is also implementing technological infrastructures and tools such as programming Application Programming Interfaces (APIs), so that third parties and software development communities can access and integrate these datasets with their own, contributing this way to the increasing “Web of Data”¹⁴. Other examples related to education include the LinkedEducation platform¹⁵, or the LinkedUp project¹⁶ (Linking Web Data for Education), which aim to promote the use of linked data for educational purposes.

2.2.1 Digital Repositories at the Core of Research Archiving

At present, digital repositories are developing rapidly and the role of digital libraries in those implementations is very important. Digital libraries are making diverse contributions in the digital repositories development area (ARL Digital Repository Issues Task Force, 2009) such as housing local digital contents, and facilitating the digitisation of classic datasets. They represent a key element of electronic research infrastructures by providing users - authors, contributors - with a set of services that facilitate long-term archiving and migration of content; dissemination and access management; metadata and format management; search and discovery tools, among others.

In the simplest terms, a digital repository is “where digital content, assets, are stored and can be searched and retrieved for later use... A repository supports mechanisms to import, export, identify, store and retrieve digital assets”(Hayes, 2005). There are numerous ways of classifying digital repositories, e.g. by type of content (e-prints repository, data repository, learning object repository), by how it is managed (institutional, subject, research), or by purpose or services provided (preservation of digitalised contents, unified access to heterogeneous data or data management), among others. But, what differentiates digital repositories from other technological platforms like content management systems, data catalogues or databases? One key difference between a digital repository and other systems is that a repository is a more flexible system in terms of the type of contents and the uses of the data that it holds. Content Management Systems (CMSs) are normally tied to one particular use of the data, for example a CMS used in a university stores learning materials for specific courses or modules. In contrast, a digital repository might hold the core intellectual assets of the whole university, thus facilitating their use to support a variety of information processes. Repositories can serve content that can be used in departmental VLEs, can underpin educational websites and can disseminate research outputs across particular subject areas or time periods (Hayes, 2005). While the uses of digital repositories vary considerably - institutional, research organisations, personal archives - this study focuses on those that serve as digital archiving systems for research data, more specifically in the discipline of social sciences, and those that manage qualitative studies, although they still provide basic support for quantitative data such as survey data, and questionnaires. Both raw, and derived, research data are considered, as well as other supplementary materials that enrich research collections.

Currently, developments in the area of digital repositories, especially within research and education, are primarily concerned with facilitating access to different types of resources, not only enabling access from a particular system or repository but also connecting and enabling integrated access to a whole network of digital repositories holding datasets from across a wide range of disciplines (Heery and Anderson, 2005; Hedges and Blanke, 2008; Jacob, 2009). A very important example of this type of implementation is the European project DRIVER¹⁷ (Digital Repository Infrastructure Vision for European Research), which is a multi-phase initiative whose primary objective is to establish an infrastructure of digital repositories at European level, offer-

ing sophisticated repository services to both researchers and more general users. The implementation and provision of mechanisms to deliver content resources in a wide variety of formats, e.g. any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data, is central for such kinds of projects. Additionally, the current research and developments in the areas of systems interoperability, data documentation, and metadata¹⁸ standards, play a crucial role to successfully facilitate content discovery and dissemination. Examples of implementations that aim to achieve those include the development of open data description standards, which range from generic description vocabularies like Dublin Core (DC) to describe bibliographical records, or Learning Object Metadata (LOM) to describe learning objects or electronic educational resources, to more specialised ones such as Data Documentation Initiative (DDI¹⁹). DDI aims to provide a common framework to describe mostly statistical (quantitative data), although begins to support the description of qualitative data. Data documentation standards combined with the implementation of mechanisms to expose digital scientific resources like the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)²⁰, allow for interoperability within networks of repositories by providing a common repository output. Open Archives Initiative (OAI) has made it possible to make all institutional repositories seamlessly interoperable, i.e., OAI allows the contents of digital archives to be exposed and metadata collected and aggregated at archive, collection or individual item level.

2.2.2 Digital Repositories, Open Access and the 'Web of Data'

Across a wide range of disciplines, the provision of flexible and configurable digital repositories that offer interfaces such as OAI-PMH, provide support to document or describe datasets using multiple documentation standards or enable combinations of internal metadata (or data) with external data coming from multiple sources by using integrated search endpoints like semantic triplestores. Semantic triplestores are special databases that allow the storage of information in ways that preserve its semantic meaning so that it can be processed in machine-readable ways. In contrast to more traditional databases such as relational databases, semantic triplestores can provide integrated access to a range of different sources rather than access to a single source. For example, one could perform 'aggregated searches' in which authoritative content on a specific subject, and sourced from a digital library, could be combined with newspaper archives, or with crowd-sourced content related to the topic of interest. Such technological developments represent some of the key enablers of the adoption of the next generation of the web, the Semantic Web (Berners-Lee, 2000; Shadbolt et al., 2006) more generally.

There are a number of open-source repository implementations that are community driven, such as Duraspace²¹ (DSpace, Fedora) and ePrints. What most of them have in common is their flexibility in terms of the types of content that they could manage, their extensibility (how easy is to extend and enhance them) and their support for mechanisms to import, export, identification, storage and retrieval of digital assets. Flexible Extensible Digital Repository Architecture (Fedora) is of particular interest in relation to semantic web developments and it has been

selected as the basis for the digital archiving tools implemented in this study. This system offers many useful features such as an open digital object model (Lagoze et al., 2006) that allows users to define rich content assets and it is optimised for handling a wide variety of physical types of data. Another important advantage is its architectural separation between the repository itself and user and programmable interfaces. This means that there are very few limitations to the features that can be added to it, whilst still maintaining its interoperability with other software applications and systems. This flexibility is also an important factor in educational contexts where it allows the development of the different interfaces that might be required to address the needs of teachers and learners, rather than archivists and expert researchers. Fedora offers several features that facilitate the integration of existing datasets with those available online in semantic-ready formats. The digital assets stored in the repository can be described using metadata annotations under different schemas, so bibliographical records could use Dublin Core, MACHine-Readable Cataloging (MARC)-Metadata Object Description Schemas (MODS)²² and other vocabularies. Moreover, Fedora is closely coupled with Mulgara²³ semantic triplestore, meaning that the contents of the digital repository, regardless of original sources or formats, can be exposed as 'semantic web ready' data owing to its support for formats like RDF (introduced earlier) or Notation3 (N3)²⁴. As such, it offers opportunities for 'legacy' systems, databases, catalogues and teaching collections to be integrated into semantic web and linked data applications. Data and metadata from the digital repository can then be combined with data from other sources and the results of online searches.

Supporting the integration of different datasets within a common storage system requires expressing data in a common, processable format, hence a very important development area concerned with linked data is the implementation of exchangeable and standard data formats and conversion tools. One of the most popular formats that enables both describing information in machine-readable ways and integrating together information coming from multiple sources is RDF. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning (W3C, 2004). Since the description of information in RDF is essential to the development of linked data applications, there are numerous development projects implementing conversion tools highlighting the World Wide Web Consortium (W3C) "ConverterToRDF"²⁵, which compile the most popular available converters (OAI-PMH, MARC, Excel spreadsheets, or BibText) or the Massachusetts Institute of Technology (MIT) directory of "RDFizers"²⁶.

Once all the data is expressed and integrated in suitable formats, it is very important to provide interactive and easy-to-use applications to visualise and navigate through the data. There has been significant work in the development of web application frameworks designed to include semantically rich data, and to interact directly with the aggregated datasets. These include lightweight data visualisation frameworks like 'Exhibit' from the Semantic Interoperability of Metadata In like and Unlike Environments (SIMILE) Project (Huynh et al., 2007; Mazzochi et al., 2005) or data discovery interfaces like Blacklight²⁷. Alongside these developments, another im-

portant area concerned with dissemination of information, mostly academic and research oriented, is the development of OA, which was introduced in Section 2.1. There are two ways to provide OA. The first of these makes use of open access journal-publishing (i.e. “the golden road”), in which journals provide OA to their articles. The second approach involves using open access self-archiving (i.e. “the green road”), in which authors provide OA to their own published articles by making their own eprints²⁸ free for all. Additionally, OA can refer to any form of scientific information - publications, primary data, etc; but it is considered as essential to, at least, include access to research articles published in peer-reviewed journals (or conferences) in all scholarly and scientific disciplines worldwide. Optionally, one can self-archive other types of materials under OA (books, media content such as audio or video, software...) or supplementary resources like primary data or ‘pre-prints’.

Linked Data on the web initiatives, in combination with open access policies, are clearly contributing to the construction of “enhanced publications”, which combine interrelated information objects into a logical whole, e.g., publications coupled with relevant presentations and associated datasets (Vanderfeesten, 2011). For example, a journal article could include not only the final paper but also the underlying research data, illustrative images and other publication data such as comments and ratings. These approaches present the benefits of maximising the uptake, usage, applications and impact of the research output of universities and also enable the collection, exhibition and management of the research output and impact of a given institution (Harnad, 2007). There a number of projects aiming at implementing systems to support the presentation of publications in combination with related research data. For example, the Enhanced Journals...Made Easy! (EJME) project²⁹ aims to to design a practical work process for publishing scientific/scholarly journals with enhanced publications, that is, supporting the researcher in all the steps involved in publishing online a research article, from the online submission of the text, to the online presentation of both the text and related research data. Another example is the ‘inContext Visualiser’³⁰, which is an open-source visualisation tool that allows the display of research material and the connections between the various different related items, such as related multimedia materials, information about the authors, or related articles.

From here on, the rest of the chapter focuses on digital archiving of qualitative data studies, highlighting technology developments and current challenges/issues around qualitative data sharing and exchange. It then provides an account of the key applications of qualitative research archives (secondary analysis, teaching and learning) and the implications of archiving for qualitative inquiry, with a focus on the epistemological, methodological and ethical challenges.

2.3 Archiving and Re-Using Qualitative Studies

Before the establishment of research data archives in the social sciences, the main sources of secondary data were public records, organisations’ in-house records, publications and researchers’ personal collections. The advances in computing technologies and digital archiving

have contributed substantially to the availability of secondary data (Heaton, 2004b). While these advances applied initially more to quantitative datasets, e.g. to the establishment of first major social sciences archives like the Human Relations Area Files (HRAF)³¹ or the Roper Center for Public Opinion Research (both in the United States), major interest has grown in the possibility of archiving and re-using existing qualitative research studies in recent years (Corti and Thompson, 1998; Moore, 2007). Archived qualitative data are a rich source of research material that can be re-analysed, re-worked and compared with contemporary data. Qualitative studies present an enormous potential for their sharing, re-use and secondary analysis. There are a number of reasons for developing qualitative archives: as a mechanism for safe storage facilities, thus preventing data loss; to provide enhanced opportunities for data sharing and comparative research - new digital systems provide richer applications to visualise, navigate through and search for materials; or using archives as teaching resources for qualitative methods courses, among others (Sobal, 1981; Corti, 2005; Bishop, 2012). However, whilst some of the applications of the use of existing research studies in social sciences are contested (this is developed further in the next sections), it has been widely recognised that teaching applications are useful. Research students use classic studies as secondary sources, or as primary sources with which they can explore new research questions. Such uses present opportunities for them to gain a better understanding of the rationale for selecting particular data collection instruments, or the complexity of data analysis. They can also evaluate particular research strategies or fieldwork methods.

Traditionally, there was a general preference among international data archives to deal predominantly with machine-readable statistical or quantitative data. This information can be processed more easily than qualitative data. This, combined with researchers' familiarity with existing quantitative databases, has made secondary analysis to quantitative data a very common and popular practice among researchers in social sciences (Kiecolt and Nathan, 1985). Furthermore, researchers' familiarity with these practices has had a clear impact in technology developments in the area of quantitative data management and analysis. There is a wide range of computer software packages and statistical techniques for quantitative data analysis, which have increased the utility of large-scale quantitative datasets (e.g. survey data) for purposes other than those originally intended, thus making secondary analysis and data re-use more important. The latter, in combination with community-driven efforts in the areas of digital archiving and re-use, have contributed to the development and adoption of quantitative data documentation standards. Such standards aim to ensure datasets' compatibility and portability, highlighting the work carried out by the DDI Alliance. While there is a well established tradition of archiving and reusing quantitative data in social sciences (Kuula, 2000) and quantitative social scientists take the lead in the development of particular archives, there is no similar tradition with the re-analysis of qualitative primary data from other researchers and therefore, the level of commitment to data sharing and re-use is not as strong among the qualitative research community.

As for archiving and data sharing, the latter had implications on the number of demands from qualitative researchers for digital archiving facilities to be established, which remained

much lower. In the UK, for example, the Economic and Social Research Council (ESRC) Data Archive was established in 1967 to preserve quantitative datasets but it was not until 1994, with the development of Qualidata³², that plans for including qualitative data were made. In this respect, one of the main issues was the availability of datasets from primary studies, which was quite limited, but has increased considerably. There have been some important initiatives to archive qualitative data, both data from classic studies and contemporary ones, like the above mentioned 'Qualidata' (UK) or the Murray Research Centre³³ (US). Although large amount of datasets, primarily historical, are already archived, the lack of standards when documenting and archiving presents challenges for data discovery, access and therefore, re-use. Other examples of archived materials include universities' repositories holding researchers' data collections, including interview data (transcripts, audio or video files) or reports from past studies but again, they remain isolated and public access is mostly not possible. New and current technology developments, especially those related to information management and online access such as digital repositories, federated search and retrieve services, data representation formats, provide a solution to these issues. The availability of better technological infrastructures facilitates the discoverability of and provides access to a greater number of qualitative datasets, promoting secondary uses, like the recent initiative promoted by the ESRC that makes freely available over 200 datasets covering every major project funded by this council (ESRC, 2012). However, there is a particularly important technological area that needs further development alongside the implementation directions that the other highlighted areas are taking. This area is concerned with the development of open descriptive documentation standards. The availability of consistent frameworks for documenting and describing qualitative datasets is key to archiving and data sharing. It not only provides mechanisms for structuring the datasets but it also enhances discoverability, search mechanisms, so that existing sources can be connected or aggregated to secondary ones.

2.4 Data Management Strategies and Qualitative Archives

The main motivations for establishing qualitative archives are enabling and facilitating secondary analysis, data re-use and teaching. Nowadays with the advances in technologies, it is much easier and less time-consuming to store, disseminate and make data accessible online, for example, many institutions are keen to share research data to increase the impact and visibility of their research. Among the research community, there is a growing recognition of the importance of data sharing in principle and in practice. Many data collections have value beyond the original research purposes. Data sharing promotes innovation and potential new data usages; enables verification of research findings; encourages the improvement and validation of research methods; reduces the cost of duplicating data collection; promotes the research that created the data and its outcomes, and provides important resources for education or training (Van den Eynden et al., 2011).

There are a number of ways for sharing research data. Some of the most common include:

- depositing research data with a specialist data centre or archive,
- including research data as part of a journal submission to support a publication like enhanced publications which were introduced previously,
- making data available online through a project or institutional website and,
- depositing data in an institutional repository or sharing informally with colleagues (peer-to-peer).

The criteria for selecting among the different approaches varies owing to the nature of the data generated, their characteristics, and it also depends on the research environments and disciplines. With respect to the latter, researchers' attitudes and practices in relation to the creation, sharing, re-use and preservation of research data are closely linked to the discipline in which they work. Within the social sciences, the degree of researchers' preparation for data sharing depends greatly on the traditional characteristics of individual sub-disciplines. Additionally, most of them share the fact that their focus on collecting and using data is bounded by agreements relating to confidentiality, together with legal and ethical considerations, which are significant barriers to data sharing and re-use. However, as it will be described in Chapter 4, these barriers could be overcome with some of the existing technologies. For example, new digital repositories provide robust access control mechanisms that could be used in combination with anonymised data rendering techniques. Notwithstanding, there are different kinds of research data, such as audio-visual materials, that are challenging for the existing technologies. It is very common in many studies in social sciences to use video and analyse participants' interaction.

Differences in disciplines or sub-disciplines, e.g. research around sensitive issues like violence, trauma, post-conflict societies in contrast to theoretical research, or research around methodologies, combined with the issues derived from the nature of the qualitative data being collected reveal that there is a need for developing new methodologies and models for data archiving when dealing with qualitative data. While there is often a tendency of trying to produce generic models, evaluation projects researching sets of case studies within multiple disciplines have revealed that it is preferred to define domain-specific strategies (Key Perspectives Ltd., 2010). These strategies should not only take into account data preservation or curation but they should also capture and document, as well as possible, the research process and they have to be based on a combination of social sciences methodologies and archival descriptions (Corti et al., 2007).

A number of initiatives, carried out by higher education institutions and research centres, for the development of data sharing infrastructures have appeared (JISC, 2009; The Europeana Foundation, 2010; CENDARI, 2012). Such initiatives have a number of aims: 1) supporting researchers to manage and share data through tools - virtual research environments, research social media, 2) providing practical guidance and training - ethical procedures, data protection

and management training courses and 3) enabling data citation and linking data with publications - to increase visibility and accessibility of data and the research itself. On the other hand, data management is not only crucial for quality research data but also for facilitating data sharing, long-term preservation and therefore, re-use. Furthermore, the potential value that can be gained from encouraging researchers to make their research data available for others to find, review and re-use, possibly in ways not imagined when the data was collected, has been recognised (Key Perspectives Ltd., 2010).

2.4.1 Data Management Methodologies: Generalising Practices

The current tendency across research funding bodies of requiring that research grants application specify a data management plan in the proposals was introduced previously. In this respect, there are a few projects that have implemented data management approaches to train and support researchers for data preparation and storage. One example is the Rural Economy and Land Use Programme (RELU)³⁴ programme which has assisted researchers with processes such as planning custom project-specific data management procedures, or with the design of their own data management practices through programme data policies - for data archiving - and by providing a dedicated support service, funded by the research councils (Van den Eynden et al., 2010). Another project of interest is Timescapes³⁵, which has developed a multimedia archive of longitudinal studies that incorporates textual, audio and visual data. Such an archive offers great opportunities to analyse existing research data through time. Moreover, its spin-off research project explores the possibilities of conducting and evaluating secondary analysis approaches which make use of the archive (Irwin et al., 2012).

With respect to data management, the key areas in the social sciences are: data management planning - primarily at the research design stages; ethics and intellectual property rights (IPR) policies; consent and confidentiality; and data contextualisation, description and documentation (Van den Eynden et al., 2010). The data management practices may vary depending on the nature of the research programme, i.e. 'what emphasis is placed on data management and data sharing', or on practices applied by individual researchers. The first area, data management planning, is concerned with helping researchers considering, at the time of designing and planning research, how data will be managed during the research process, and shared afterwards. The most common areas of work identified in the process of planning a project's data management strategy (Van den Eynden et al., 2011) are:

- to describe the need for access to existing data sources and existing access limitations,
- to identify what data will be generated during the research,
- to establish plans for management and archiving of the collected data including plans for sharing data and the description of the expected difficulties when making data available for re-use and the provision of alternatives to solve those difficulties, and

- to manage the ethical or legal issues on data sharing and definition of data management roles and responsibilities within the research team.

One of the most difficult tasks during the design of data management strategies is planning how to manage research data archiving, especially identifying and describing expected difficulties as well as the ethical issues that could arise when preparing data for their re-use. Furthermore, evaluations carried out by the Qualidata centre (Van den Eynden et al., 2011) highlight that the case of not identifying any difficulties for secondary use of research data is one of the most common cases that challenge data sharing and re-use. For example, researchers often do not outline in the consent forms or instruments for data collection agreement sections for data archiving and sharing. In such situations, research data could not be archived owing to potential ethical and confidentiality breaches. Most of the times these limitations can be overcome by anonymising data, aggregating data, obtaining consent to share data, or even discussing data archiving with the owners of licensed data. Lack of awareness of the importance of discussing data sharing with research participants during consent processes may lead to the failure of getting consent for data sharing and therefore truncate the plans for subsequent archiving and data re-use. Additionally, data management plans need to provide clear and detailed information for researchers, pointing out the issues relevant during the process and should be complemented with open discussions about data management with experts (support services, senior archivists), which will help ensuring that researchers are well informed about the data sharing limitations and issues.

On the other hand, as mentioned previously, there are two major concerns about data archiving and sharing, and the implications for researchers. The first one is that archiving is often associated with an institutional management culture and therefore, researchers perceive it as a process that requires them to make significant efforts in addition to them carrying out research tasks. Researchers have to participate in the archive design, data management plan specification, alongside conducting primary research and analysis, preparing data for archiving, and at the same time, anticipating the implications of data re-use. The second concern is derived from the personal implications that data archiving and deposit preparation for open access have on researchers, i.e., researchers' reluctance to sharing their data owing to the possible effects on research participants, or their awareness of their research being exposed. In the initial stages of the data management process, issues like the protection of participants have been at the forefront, whereas when the data has to be sent to the archive, the different concerns about exposure arise. The most common ones are fear of revealing personal ways of working, consequences falling more heavily on early career researchers, who are often the ones doing primary data collection and preparation for archiving, and possible (self-) consequences for the quality of the data, among others (Cheshire, 2009; Van den Eynden et al., 2010).

The design and implementation of appropriate data management practices plays an important part in the success of the archiving of qualitative research studies, therefore an important

part of this study has explored data management strategies, the archival procedures followed and the issues that were raised during the implementation of exemplary research projects.

2.5 Technological Barriers to and Enablers of Qualitative Research Archiving

There has been agreement that data documentation, along with the advances in the area of data management systems like digital repositories, is one of the key elements of data archiving from a technological perspective. The use of appropriate standard documentation frameworks is crucial in relation to data archives' exposure: it has an impact on accessing, discovering, searching across and retrieving available datasets. The volumes of data that are archived continue to increase and 'open data' initiatives are making many of these more evident. However, for archived data to be useful, data has to be accurately, richly and contextually described (Corti, 2005). The representation of original data, methods and analytic interpretation and their relationships requires agreed and open standards and procedures if they are to be archived. In this respect, the number of widely accepted standards for selecting, structuring and documenting archive content is increasing considerably. The first of these is the DDI, which was introduced in Chapter 1 and seeks to provide a standardised means of describing all the elements of a research study. This covers not only data but research instruments, designs and protocols, along with analytical schemes and the results of data analysis. However, DDI started off with a focus on quantitative and statistical data, rather than qualitative data. Another significant development is the OAI-PMH (introduced earlier), which allows the contents of digital archives to be exposed and metadata collected and aggregated at archive, collection or individual item level. A third is the Data Exchange Tools (DExT) Project, which aims to provide researchers, and data archive staff working with primary research data, with tools that enable not only long-term preservation but to develop, improve and test models for data exchange for both survey data and qualitative research data.

With regards to qualitative data, there are different initiatives bringing together the archiving community to help in the definition of standards that cover qualitative data since, for most of the data, the ways in which they can be described do not differ much from the description of survey or aggregate data. Both types of data, that is qualitative and quantitative, are structured to some extent and present common features or description elements, such as data provenance (for both types the provenance is some kind of social science investigation), methodological information, or elements to describe the data collection protocol (sampling and collection methods). The similarities between both types of research data, in terms of how to describe them, has been one of the motivating factors which led the DDI to establish in 2010 an international working group: the Qualitative Data Exchange DDI Working Group³⁶. The work of this group is particularly of interest to this study since it involved the implementation of a test model for qualitative data that could be integrated within the existing DDI standard. A first draft of the schema has been implemented, and it is described in a paper published in the DDI working papers series (Hoyle

et al., 2013). The development of such a kind of schema is highly important for both the research and archiving communities, and represents an attempt to overcome the lack of standards to document and archive qualitative data.

Open descriptive standards are not only useful for qualitative data description and documentation but also provide useful mechanisms to structure collections of data and to define archival models. Traditional “data archival models” normally look at a collection of research data and assemble the data together in a logical way, with assistance from the investigators involved in the research project being archived. However, it is very important to provide mechanisms that allow not only the definition of terms describing research data but also the inclusion and preservation of the different interrelationships between the data being described. These models require common descriptive standards to document, store and provide access to the data. Examples of descriptive elements include study description - research design, collection methods, data provenance; data types and files included in the collection - along with their associated relationships and access conditions. Currently, standards like [DDI](#) (in its most basic form) are being used by data archives as agreed standards to describe their qualitative data collections. However, the extent of the documentation only covers high-level methodological description (research design, data collection methods), that is supplemented by ‘user guides’, prepared by the original researcher, and a file list of the resources included in a given archive. In this respect, archiving models such as the one developed in this study which underpins the implemented self-archiving toolkit (described in detail in Chapter 4), or the more advanced schema model proposed by the [DDI](#) working group attempt to provide enhanced solutions for the archiving of qualitative data. Firstly, these models support more detailed documentation, both at study (methodology, research design) and data levels. Secondly, they provide mechanisms that allow the expression of the relationships between the different research elements included in a research study in machine-processable ways. For example, transcriptions can be associated with audio/video materials, or annotations linked to specific segments or sections of transcriptions. All the relationships can be expressed in standard ways facilitating data exchange between different systems.

In addition to open data documentation standards, the standards for the representation of texts in digital form are key to data documentation, sharing and exchange, thus they are of interest to most of the research data archives. The Text Encoding Initiative ([TEI](#)) is one of the most widely adopted standards of this type. It specifies encoding methods for machine-readable texts within the humanities, social sciences and linguistics, among others. [TEI](#) guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation ([TEI, 1994](#)). Such standards enable the addition of documentation at research data level, in the form of electronic mark-up (normally XML-based) that adds more value to the qualitative data. Useful contextual data can be added to qualitative data like descriptive headers in transcripts or field notes, [XML](#) mark-up of data such as speaker tags, corrections made in a transcript, or researchers’ annotations. Such [XML](#) tags enable web-based

searching, browsing and retrieval of research data. Whilst there are still open issues of re-using qualitative data in sensible ways without all the background knowledge and tacit understanding acquired by the original researchers (Corti and Gregory, 2011a), the provision of this enriched mark-up offers the possibility of generating more detailed documentation of the data. Examples of the latter include describing audio recordings, attaching field notes, fieldwork letters and memos to the original data, which can help aiding the original fieldwork experience. Additionally, making available researchers' coding, the classification frameworks used during analysis or the annotation of the relationships and dependencies within the collected data, provides re-users of the data with more contextual information, thereby working across collections can be facilitated. However, the current ways of capturing this type of information, which make use of software tools or packages, creates problems in sharing, or re-using, research data.

2.6 CAQDAS Tools. Towards Research Data Exchange and Archiving

Unlike to statistical packages commonly used in social sciences research such as Statistical Package for the Social Sciences (SPSS), CAQDAS packages or QDA tools are designed to facilitate 'qualitative approaches' to qualitative research (Silver, 2012). The purpose of this type of software is not to provide researchers with a generic methodological or analytical framework: the different available packages provide support for certain analysis tasks differently, and there are still debates around whether a particular tool or package might drive the way researchers perform the analysis (Lewins and Silver, 2009). It was quite common for these packages to be designed based on a particular research methodology, or to provide solutions to specific types of research projects. The different methodological approaches which underpinned the early development of some of the main CAQDAS packages is described more in detail in Appendix B.

Since qualitative research studies can create huge amounts of raw data (Miles and Huberman, 1994a), the tasks of organising, tracking, encoding, and managing the data are very important to researchers, and there is an awareness of the efforts needed to accomplish them. Early developments of this type of tools came primarily from the work of academics that were involved in qualitative data analysis in the late 1980s. The first software program implemented to support the process of qualitative data analysis was the Ethnograph (Seidel, 1998), developed by John Seidel in 1985. In the early 1990s, a number of other software packages were launched, for example QSR NUD*IST (Richards and Richards, 1991) and ATLAS.ti (Muhr, 1991). The first versions of these packages were designed and implemented by researchers, as software-based solutions for specific projects, that is, the research data itself was central to the tools' requirements. Aspects such as research topics and the chosen methodological approach had an impact on the software development and design. For example, Ethnograph was designed to support a qualitative approach, where the data collection method was interviews, by outputting interview segments sorted by selected topics or code words. In contrast, ATLAS.ti design was driven by project needs as well as by a combination of various methods, i.e., phenomenology, hermeneutics and grounded theory. Despite this, they have evolved into more general tools which incorpo-

rate a set of functionalities that support multiple research approaches. The most common (core) functionalities of the different CAQDAS packages are the following (Lewins and Silver, 2009):

- *Project and data structure management.* A key feature is to facilitate access to, and management of, project data in ways that facilitate access to parts or whole project data. Most of the packages handle a single 'project file' that enables instant access to the research project's descriptive information and supports maintaining the links between all the data included in a particular project. The ability to make associations between data elements is central and facilitates organising ideas and analysis by creating paths through the data based on both relationships and concepts: e.g. when working with interview data, one could navigate through segments based on associations or relate segments to other sections of the data.
- *Code and retrieve functionalities.* When performing analysis tasks, it is very important for researchers to be able to manage self-defined keywords and/or conceptual categories (codes) and apply them to selections of text, still images, or of segments of audio/video. Most of the packages provide support for those in addition to features for code generation in easy and flexible ways, supporting different combinations of strategies.
- *Project management and data organisation.* In addition to the capability of creating and maintaining project's descriptive information along with research data, CAQDAS tools offer mechanisms to manage and store additional types of supplementary information, such as literature lists and abstracts. These can be cross-referenced and coded within the project.
- *Searching and interrogating the datasets.* Researchers are always testing ideas and relationships between themes and issues, by asking questions or querying a project's database. All the packages offer means by which one can interrogate the dataset. Searches might produce an additional level of analytic coding and allow combining coding (interpretive or conceptual) with the organisational (descriptive) dimensions of the work. The results of the searches can be saved in report files or, with some of the packages, they can be coded and integrated into existing coding schemas.
- *Closeness to the data.* CAQDAS include different mechanisms for maintaining and providing instant access to source data files. It is very important for researchers to have instant access to the primary source data. These packages allow the extraction of specific segments or parts of the data that have been categorised or annotated using different coding schemes in ways that can be more accurate than when working manually. However, these capabilities often require researchers to keep the data in flat and static ways that can lead to narrower explorations of the data. There are some critiques of these approaches (Richards, 1998; Seidel, 1998) which, while recognising the importance of being close to the data, maintain that some distance from the primary data is critical to good analysis. The focus of the analysis should be on the interrelated aspects of a setting or subjects under investigation rather than breaking the whole into separate parts.

- *Annotation tools.* This feature is particularly important to qualitative analysis, both in terms of the source data (analytical memo-ing) and the analysis process itself (procedural memo-ing). Researchers often create interpretive materials during data collection, in the form of notes or observations, that supplement the data collected. During analysis, descriptive and contextual information can be added to the source data. Memos or comments can be linked to documents, either embedded in the actual document or separate in textual form. Notes or comments can be added to explain codes and concepts, how they are defined and how they change, in order to keep track of the analysis process and the rationale for certain tasks being performed. Most of the CAQDAS packages provide support in one way or another to include these as ‘annotations’ of source data along with descriptive and contextual information at research project level.

Over the last decade, mixed methods approaches are becoming more popular and have been adopted in different fields like sociology, anthropology, education and geography. Geography, more concretely the sub-disciplines of geo-referenced Qualitative Social Sciences (QSS) or Qualitative Geography (QG), provide an interesting example of fields that have found pathways through mixed-methods approaches, enabled by technologies such as CAQDAS packages and Geographic Information Systems (GIS) applications (Fielding and Cisneros-Puebla, 2009). Examples of such packages include Dedoose³⁷, MaxQDA³⁸ or Digital Replay System (DRS)³⁹. Some of the functionalities of those packages include support for combining quantitative and qualitative data (survey data and interview data from the same participants); inclusion of geo-referencing information; identification of types (categories, codes, themes) in the qualitative material and comparison with selected quantitative information, or cross-tabulation to compare coded segments from interviews according to selected groups (Kuckartz, 2010).

2.6.1 CAQDAS Evolution: Supporting Collaborative Research Approaches

Initial generations of CAQDAS included existing word processing software and database systems which were re-purposed to aid researchers with data organisation and search tasks but did not provide or integrate encoding functionalities to support data analysis. Second generation tools, commonly referred to as “code and retrieve” tools, started adding functionalities for creating code schemes or networks to support theory building. Lastly, third generation includes the latest versions of the most commonly used packages such as NVivo, Atlas.ti, HyperResearch. Most of them are often classified as “code-based theory building software” and provide support for many of the activities involved in text analysis and interpretation, such as selecting, coding, annotating, and comparing relevant segments of the data. They also support other approaches by incorporating more advanced multimedia capabilities that allow users to work not only with texts but with a wide variety of data such as video, graphics and audio. Although there has been a continuous evolution and new developments are taking place in the area of QDA software programs, most CAQDAS packages are still single-user oriented tools in which the user creates a ‘project’ file (the terminology differs among packages) and assigns or imports all the data files relevant to the research into that ‘project’. In this case, the research data is locked into one par-

ticular software solution. There is no support for sharing or making research data accessible from outside the software program, and the analytical approaches to the data are constrained to the software built-in methods (first scenario in Figure 2.1).

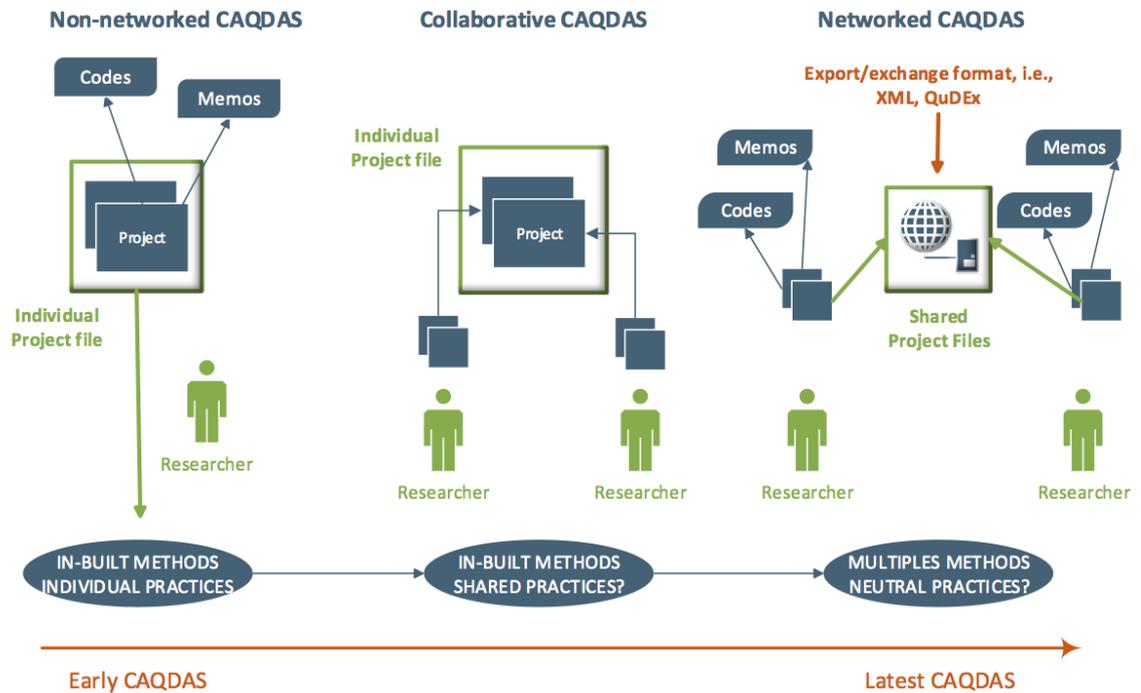


Figure 2.1: CAQDAS evolution, collaborative research and data sharing

Research approaches within the social sciences, such as ethnographic research, often involve gathering a combination of qualitative data from fieldwork (observation notes, research diaries) with quantitative data (surveys, questionnaires). For ethnography in particular, the number of research projects in which research is undertaken by multidisciplinary teams working collaboratively is increasing substantially (Lieber et al., 2003). The central instrument for this type of research are field notes and related data: researchers write notes both on topics that emerge from fieldwork and on topics based on prior theory or literature. Unlike other kinds of qualitative research, ethnographers, especially within collaborative ethnography, are often less reluctant to share their research notes for use within the research team, or with research participants, and they often work geographically dispersed. The technological requirements for this type of work differ notably from those that are based on individual research projects, in which shared access to research data, or collaborative analysis, are not needed and the outputs from the project are normally publications or reports incorporating elements from the analysis in the form of vignettes, quotes, or interview excerpts. In situations where geographically distributed research teams contribute to and work with the same research data, the affordances of QDA technologies and web-based technologies to support this type of environment are apparent. A combination of CAQDAS functionalities with online systems for storing and retrieving information provides a solution for the previous situation. Researchers can store and modify their notes,

which can then be edited or annotated (with codes and categories) by other researchers within the team. Research notes can be linked to relevant literature or underpinning theory, or to other forms of research data such as visual data or statistical data, and researchers can construct and manage collaboratively codebooks or coding summaries so that the data can be interrogated and searched across looking for patterns (Lieber et al., 2003). Whilst this category of CAQDAS software enables collaborative work and provides a means to share research data (within the research team, not with the wider community), the source and analytical data are still locked into a particular database solution and research projects in which a variety of analytical approaches are involved are not supported (collaborative CAQDAS in Figure 2.1).

Although a few of the early QDA software packages started to provide support for team-based collaborative research (Dedoose, formerly EthnoNotes), there is still a lack of support for complex multidisciplinary and collaborative projects that require multiple analytical approaches and data sharing. The dynamics of research teams can vary quite significantly, as can the role of software within them. The Learning Landscape Project (LLP)⁴⁰, that aimed to recognise and describe the special nature of teaching and learning in a research-intensive, collegiate university, provides a good example of Mixed Methods Research (MMR). While most of the researchers of the project team came from a social sciences background, the nature of the research required different analytical approaches to the same data - questionnaires producing both quantitative and qualitative information, focus groups and students' diaries. The analysis was performed in collaborative ways and both data and analytic outputs were shared. Additionally, different researchers took part at different stages of the project, that is, in situations in which data was already collected, either re-analysis or new analysis took place. In this particular example, the analysis tasks can be seen as collaborative in a sense. Some data was already analysed and included in a previous report, and then another researcher came in and drew upon those to perform new analyses or completion; or research instruments which required mixed-method approaches - questionnaires with qualitative and quantitative questions - were analysed by different researchers. The use of an online VRE was crucial to support these kinds of activities, collaborative work and data sharing. The VRE was used as a sort of database/archive in which the entry point was a web page containing a graphical representation of the collected primary data as a way of navigating through the data, via hyperlinks that redirect to the relevant data folders in the VRE. The model for organising, storing and presenting the data was grounded on the concept of 'case record', introduced by Stenhouse (1978), as it helped in structuring the data and designing useful ways of presenting them. This model is a top-to-bottom (tier-based) representation in which the top layer contained all the elaborated (edited) outputs such as reports; the middle tier included links to all the intermediate 'case records', which are descriptive and lightly edited accounts of a piece of source data; and the bottom layer contained links to the original data collected.

Approaches like the one described above show that in reality researchers use a combination of different kinds of tools that provide support for the activities in the course of a research

project: archiving systems, such as digital repositories or VREs, are used to manage, share and visualise research data, and QDA tools enable to explore and analyse different kinds of data. The tendency is for users to take advantage of specific functionalities of a range of tools rather than buying-in a particular software solution, e.g. collaborative networking tools or online archives to manage and share data, and a variety of CAQDAS packages for analytic purposes (Carmichael, 2002). However, whilst current information technologies such as Access GRID⁴¹ or semantic technologies enable distributed data archiving and collaborative ways of working with data, CAQDAS packages' lack of data exchange functionalities, and the fact that the tools manage data in different ways, represent a barrier and add substantial difficulties to data sharing and research collaborative tasks that are technology-supported. In this respect, XML-based technologies can enable the inclusion of capabilities for web use in CAQDAS packages (networked CAQDAS in Figure 2.1). Likewise, exchange formats that are XML-based have the potential to provide solutions for incorporating qualitative data analysis functions into networked applications, such as digital archives, or groupware systems, as well as interchanging data between the different CAQDAS tools. This represents a shift from research project's data locked into one particular tool (non-networked CAQDAS in Figure 2.1) to shared project's data, which is managed by different tools, and that can also be integrated with data from other systems, such as research archives, or e-publishing systems. There has been a growing interest in collaborative research archival (Fielding, 2004). However, archived qualitative data remains under-exploited in comparison to quantitative data within the social sciences. The availability of datasets from primary studies in electronic format for online access is quite limited, although it has increased considerably in the last ten years. As mentioned in section 2.3, initiatives such as Qualidata (UK), the Council of European Social Science Data Archives (CESSDA) in Europe or the Australian Social Sciences Data Archive (ASSDA) are promoting and contributing to this increase. More recently, international and multidisciplinary research projects like the Collaborative European Digital Archive Infrastructure (CENDARI)⁴² are also providing and facilitating access to existing research archives and resources, to promote the development of inquiry environments using secondary data. These initiatives and research projects, in combination with e-social science technologies (online data collection tools, GRID technologies for fieldwork) applied to digital archives enable researchers access to a wide variety of online research materials, that can be annotated, shared and used to produce multiple and interrelated narratives. These types of scenarios, in combination with the use of CAQDAS packages with analytic purposes using a wide variety of data, highlights the fact that the distinction between research archiving (and their associated tools) and CAQDAS tools is blurring. Moreover, the latest developments in the area of qualitative research archiving are exploring and implementing new models of archiving that include support for documenting, storing and disseminating qualitative data resources, and they also facilitate the exchange of these kinds of data within other systems like research online environments, or QDA software, so that new analyses or re-analyses can be performed over aggregated data (Corti, 2008; Corti and Gregory, 2011a,b).

2.6.2 CAQDAS and Data Archiving

Current trends in data archiving developments, along with the adoption of innovative research methods within qualitative social research involving collaborative work and mixed methods approaches, are factors that contribute to newer CAQDAS implementations and are particularly important to data sharing and re-use. As described in the previous section, some CAQDAS packages are evolving into online tools which begin to support collaboration. However, there is still little support for working with research data, either primary or secondary, which is coming from different systems. This represents an important barrier for the use of CAQDAS in research environments where researchers' collaboration is a requirement and there is a need to share and re-use data. While some of the most popular packages support some form of import/export, the formats used are proprietary, therefore data that has been annotated and described using one package cannot be used in other packages. However, some packages offer partial export functionalities compatible with other vendors. Basic coding and annotations can be shared but the most important analysis information such as links to the source data, on which the annotations are based, is not preserved. Examples of this kind of information include 'segment to document' association, relationships between memos, and codes and their associated document(s). Additionally, while the core data elements of their underlying model for handling both data and analysis elements (categories, codes, memos, annotations) are very similar, if not the same, there are substantial differences around terminology. In this respect, the development of open descriptive standards, aiming at the exchangeability and portability of research projects between the different software applications, attempts to provide a technological solution to the previous issues. Moreover, open descriptive standards are highly important for data interpretation and long-term preservation to help future-proof data, since they offer an underlying open model which could serve as the basis for the development of open source (non-proprietary) and platform independent software.

Alongside open documentation standards, non-proprietary data formats are central to the development of interoperable QDA and archiving systems. For example, for statistical data there is a widely adopted common portable exchange format, SPSS⁴³, which is used by a range of proprietary packages. Moreover, many data types have already open descriptive standards: documents, spreadsheets and databases can be exported into XML as a common format; or there are open standards to describe audio such as Advanced Audio Compression (AAC) or Audio Video Compression (H.264/MPEG-4 AVC). This is not the case for CAQDAS packages since the most popular packages use proprietary formats for the data they handle. This has implications for both research data and researchers in that data will be locked into one technical solution owing to data being loaded, classified and annotated in one particular system. All the descriptive value added to the data - e.g. codes attached to particular segments of text, and variable classifications attached to documents - cannot be exported in ways that preserve the different relationships between data and annotations nor could they be archived in useful ways since this would imply the implementation of a system that supports the multiple vendors' formats. Although some of the current packages are starting to develop export/import capabilities (e.g. the import of anno-

tations into XML format so they could be sharable) from, and to, the other packages, none of them use any standard documentation schema. Furthermore, all the solutions from the different vendors vary and therefore, researchers are currently not able to take their own collections of annotations and prepare them for archiving purposes, or to visualise them using open source tools, or to publish them in the web. In this respect, Corti and Gregory (2011a, p. 3) suggest that “unless there is a move towards more robust interchangeability – export features and import from a common format, enriched data could be locked into a software system that may be inaccessible and redundant in the future”.

The next two sections describe more in detail how the DEX project and the advanced archiving model implemented in the DDI Qualitative Exchange Working Group (both introduced in Section 2.5), attempt to provide a solution to the latter, and other issues.

2.6.2.1 DEX Project: QuDEX Schema

The technologies for conversion, and entry, of data that is in proprietary formats are particularly important and problematic aspects of data management and curation (Corti, 2008). In this area, the work carried out by the above-mentioned DEX project has become particularly important for the archiving and dissemination of qualitative research data from technological and methodological (research practices) perspectives. With respect to the technology, it has addressed some of the ever present issues around exchangeability and portability among the different QDA tools, and information systems more generally. The work of the project produced two separate and distinct deliverables. The first of these was the implementation and testing of a formal XML schema (namely QuDEX) to support data exchange of qualitative interview-based multimedia data. The second deliverable from the project included work with survey data, mostly in SPSS format, and the creation of a Java-based conversion tool (named Survey-DEX), which exported multiple versions of SPSS to an ‘open’ format plus associated formal metadata, and to other major data analysis formats such as STATA⁴⁴ software proprietary format or Statistical Analysis System (SAS). It is the QuDEX schema that is of interest to this study since it has been incorporated, along with other schemas, into the archiving model that underpins the toolkit (for documenting, archiving and visualising small-scale research data collections) that has been developed as part of this study.

In the early stages of the design of the QuDEX schema, the researchers involved in the project analysed a set of the most popular CAQDAS packages. This provided a good insight into the most relevant analytical features, since CAQDAS packages embody a variety of methodological and analytical approaches⁴⁵. This review also enabled the identification of the commonalities among them; the terminology used (while working with similar data types and analytical elements, the terminology sometimes differs considerably between software solutions) and the kinds of data from primary research that those packages support. Both exploring the analytical elements and the kinds of research data that are managed by those applications, especially focusing in social research methods and techniques based on fieldwork, provided the basis for the

design of a ‘software neutral’ model as an attempt to solve two of the major issues faced by data archives in relation to acquisition and dissemination of data from qualitative research studies. The first of these is that most of the research data managed and transformed within analytical software are stored in proprietary formats, therefore it is challenging to archive and re-use them. Examples of these analytical data include annotations, coding and linkage of those with primary data. The second issue is related to the ability to capture the links and relationships between outputs from primary research and the source data upon which they are based. In this respect, CAQDAS packages make these relationships explicit and manage them internally, but they cannot be exported across systems. Additionally, most of the research online environments and archives do not provide support for managing these data relationships in computer-automated ways. They allow the inclusion of hyperlinks to source data (interviews, field notes) within research diaries, or outputs such as project reports, or online publications. In contrast, recent developments of digital repositories (Fedora, ePrints, DSpace), along with the maturity of semantic technologies and data documentation standards are adding important contributions to this area. This is due to their capabilities for linking heterogeneous data in computer-readable formats by using technologies such as RDF, or relationship ontologies expressed in languages such as OWL or RDF Schema (RDFS).

The QuDEx schema developed in the DExT project attempts to solve the above mentioned issues by preserving annotations of, and relationships between, data and other related materials included in a qualitative research study while, at the same time, it incorporates elements from metadata standards such as DC, OAI, DDI, TEI and Preservation Metadata Implementation Strategies (PREMIS). To achieve this, the schema is based on a number of key concepts (see Figure 2.2) and elements that can represent coding, classifying, annotating, and linking functionalities. A qualitative research study is represented in a single QuDEx XML file, in which description and location information about all the resources included the study is provided. In QuDEx terms, a research collection is composed of ‘documents’ which are logical units representing single data sources, e.g. research design, interview transcript, interview audio file. A ‘document’ presents an associated ‘physical’ unit which provides information about the physical data source file to which the document refers to. This information includes file attributes such as location (e.g. full path in a file system, or Uniform Resource Locator (URL) for files that are available online), file format, or creation date. Additionally, a document can be annotated by using QuDEx ‘categories’, which are similar to keyword classifications used to describe data; ‘codes’, which are single words, defined by the researcher, or coming from authoritative controlled vocabularies like a thesaurus, and ‘memos’, which are additional notes provided by the researcher. QuDEx categories and codes support ‘code and retrieve’ capabilities within CAQDAS packages. Finally, relating documents, codes, categories and memos is possible by using QuDEx ‘relation’ elements, i.e. “transcription A isTextFormatOf Audio Interview A” or “Consent Form isInstrumentFor Focus Group C”.

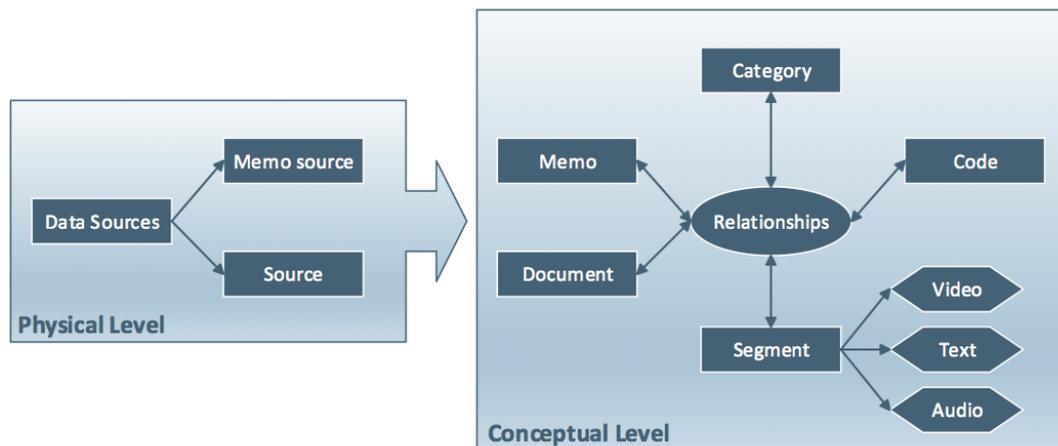


Figure 2.2: Summary of QuDEx model elements

2.6.2.2 DDI ‘Qualitative Data Exchange’ Model

The increasing interest of the CAQDAS community around re-using research data, especially in those projects where researchers work collaboratively and a variety of research methods are employed along with the maturity and recognition of mixed methods approaches (where qualitative and quantitative data are gathered) to qualitative research, are factors that have contributed to recent developments to customise existing well-known and recognised standards for the exchange and re-use of quantitative data so that qualitative data and analysis documentation could be supported and integrated within those.

The design of the qualitative exchange schema developed by the DDI WG addressed three distinct concerns. The first of these was the storage and exchange format of qualitative analysis data, meaning ‘analytical data’ in a CAQDAS sense. This refers to the elements used to describe and store data that comes from the analysis process, such as codes and categories for thematic analysis or memos or supplementary annotations of field notes. When addressing the design of this aspect, QuDEx schema (described in the previous section) proved useful and was used as the basis for modelling analytical information since it provides a mechanism for exchanging analytical data between different CAQDAS systems. The second was study documentation, that is, exploring what aspects of the research should be documented and what metadata standards are appropriate to provide richer description of information. This includes research design, methodological approaches, data collection methods and relation to external study-level documentation, such as literature and underpinning theories for a given study. Thirdly, qualitative data documentation, focusing on data archiving procedures and mechanisms for processing and storing documentation that is embedded in the data. Examples of this kind of information include headers in an interview transcript, data lists describing cases, individuals, or items being studied.

The identification of requirements and model design processes were informed by the differ-

ent use cases provided by the participants of the working group. These range from specific qualitative archive implementations from various institutions, such as Finnish Data Service (FSD) in Finland, ASSDA in Australia, or Leibniz Institute for the Social Sciences (GESIS) in Germany, to examples of research projects, that are implementing archiving tools and models, like this study, or research projects that use specific research approaches and provide valuable input about the research processes and the aspects that need to be modelled and documented. These use cases provided insightful information about two aspects of the work of this study. Firstly, data documentation at a conceptual understanding level by addressing what aspects of the research process and what elements should be documented. Secondly, technical aspects by considering what technologies are currently being used and the issues that specific systems are facing.

The rest of the chapter focuses on the challenges, issues and implications of qualitative inquiry for qualitative archived data, exploring the key difficulties in re-using data. Qualitative data re-use for the purposes of secondary analysis is a complex research activity that encompasses substantial epistemological, methodological and ethical problems that the user has to consider, compared to working with numerical data, therefore these aspects have to be taken into account, technical challenges and issues aside, when designing archiving models that underpin digital archiving tools.

2.7 The Implications of Research Archiving on Qualitative Research

While archiving technology developments and information technologies more generally are fairly unproblematic, the implementation of research archives holding qualitative social sciences data, aiming particularly at data sharing and re-use often present numerous, sometimes insurmountable, issues. There has been a debate for more than a decade about whether or not social scientists should engage in the re-use or secondary analysis of qualitative data. In the debate, the opposing positions are on the one hand one that states that qualitative data cannot be re-used on epistemological or ethical grounds, and on the other a pragmatic position that says that data should be open for their use by others (Mason, 2007). From the qualitative inquiry community perspective, many researchers are still questioning archiving qualitative research data. There are many arguments supporting the reluctance to archive qualitative research materials for a wide variety of reasons, highlighting a number of areas that will be developed more in detail in the following sections (Corti et al., 1995):

- Qualitative Inquiry and the associated epistemological barriers to data archiving, highlighting the nature of qualitative data; the methodological implications of data re-use; the loss of contextual information when archiving qualitative data; and the role of researchers in data archiving alongside the impact of data archiving on researchers,
- Ethical considerations, including preservation of informed consent and participants' confidentiality; research participants' protection; and its impact on the research itself and,

- Openness of the research data and the issues related to data access-control, copyright, confidentiality, data ownership.

2.7.1 The Nature of Qualitative Inquiry, Epistemological Challenges of Archiving

As introduced previously in section 2.4.1, the establishment of qualitative archives (such as [ESDS Qualidata](#)), policies from funding bodies such as the [ESRC](#) in the UK, which require depositing research data and associated materials for archiving at the end of research projects, as well as a growing interest on preserving and re-using qualitative data (Organisation for Economic Co-operation and Development (OECD)'s Declaration on Access to Research Data from Public funding) are factors that have clearly contributed to a significant increase of archived qualitative data ([Bishop, 2009](#)). However, these policies provoked mixed reactions from qualitative social scientists and started an ongoing debate, mostly in the UK, about the potential benefits and drawbacks of the archiving of qualitative data. Advocates of qualitative archiving saw the value in preserving significant social sciences studies; in adding transparency to qualitative research by revealing its processes of knowledge construction; and in using archived studies for historical and methodological research, as well as for teaching purposes ([Mauthner and Parry, 2009](#)). However, they also expressed ethical and legal issues regarding the disclosure of personal/confidential information; the difficulties of providing ethical assurances to respondents about how their data might be reused in the future; and, the epistemological challenges of using data taken out of their contexts of production ([Mauthner and Parry, 2009](#); [Slavnic, 2013](#)).

This polarised debate is not so much concerned with disregarding engagement in data preservation and sharing activities, but rather how this could be done without compromising the ethical and philosophical integrity of qualitative research. While few researchers would be comfortable with a qualitative archive simply for data preservation purposes, the suggestion that data can be picked up and re-used by different researchers not involved in the original data collection process is rather more contentious. Such concerns are primarily reflexivity in qualitative inquiry; the provision of context; the researcher-participant relationship; and, the risks of data misrepresentation. They stem from the more philosophical underpinnings of qualitative research. Some authors suggest that social sciences research data, generated from interpretive approaches, typically involve subjectivities and epistemologies that do not lend themselves to data archiving ([Hammersley, 1997](#); [Parry and Mauthner, 2004](#)). However, the archiving of qualitative studies, data preservation purposes aside, is central to secondary analysis and primary research data re-use. With respect to secondary analysis, one of the main reasons for undertaking it is that it provides access to rich data, particularly those, which are collected in unusual research settings, or which give voice to rarely heard participants ([Carmichael, 2008](#)). Archived qualitative materials can be used to check the findings of existing studies through re-analysis of existing data, and they can also provide researchers with banks of data that can be used for secondary analysis, either to supplement their own primary data or to carry out free-standing, comparative research which draws on a range of studies ([Hammersley, 1997](#)). The diversity of qualitative research and the epistemological underpinnings of different perspectives are reflected in attitudes towards,

and approaches to, secondary analysis.

The constructed nature of social reality for some areas of qualitative research such as Social Constructivism and the types of data generated through the research process represent an important issue while approaching secondary analysis. Some researchers argue that the fact of storing data in an archive, making it available and ready for others to use, can deny the constructed nature of research data since it eliminates prior meaning(s) or interpretation(s). The significance of this construction process is that subsequent analyses of the archived data may differ from those undertaken by the original researcher. The use of different theoretical or conceptual frameworks when re-analysing archived data, or the fact that new researchers have only partial access to the research project data (they have not participated in the original studies) could be some of the reasons why the construction processes differ. Such arguments raise the need to distinguish between analyses of qualitative data carried out by the original researcher from those undertaken by other researchers, and whether “secondary analysis is tenable, given that it is often thought to involve an inter-subjective relationship between the researcher and the researched” (Heaton, 1998, p. 3). In response to this, it may be argued that even where primary data is gathered via interviews or observation in qualitative studies, there may be more than one researcher involved. Hence within the research team the data still has to be contextualised and interpreted by those who were not present. For example, in numerous research studies, detailed interview guidelines are produced and often the researchers who were involved in data collection are not the ones subsequently interpreting, analysing the data. Another response to the issue of the constructed nature of qualitative data is to argue that the design, conduct and analysis of both qualitative and quantitative research are always contingent upon the contextualisation and interpretation of participants’ situation and responses. Thus, secondary analysis is no more problematic than other forms of empirical inquiry, all of which, at some stage, depend on the researcher’s ability to form critical insights based on subjective understanding (Heaton, 1998; Hammersley, 2010).

Some authors, such as Mauthner et al. (1998), place themselves in a very distinctive position with respect to secondary analysis. They claim that one of the barriers to secondary analysis is that the fact of new researchers not having been there to share the epistemological perspectives of the original researchers is a burdensome barrier to secondary analysis and qualitative data sharing and therefore, the primary role of SA should be methodological exploration rather than substantive engagement with research data. However, this raises the question of ‘what is secondary analysis?’. For example, for Heaton (1998), “secondary analysis involves the utilisation of existing data, collected for the purposes of a prior study, in order to pursue a research interest which is distinct from that of the original work”. Moreover, the issue of the secondary analyst not having been there could be solved by, to some extent, involving the original researchers in the follow-up study. The development of some research archives, where its implementation “was assisted by the original researchers, who provided contextual information and guidance as to the roles and interrelationships of specific elements of the primary data”, represents a pragmatic

response to the issue of subsequent researchers not having been present in the original research project (Carmichael, 2008, p. 390).

Hammersley (1997) questions the usability of qualitative data archives, when the purpose of the approach is solely to check the findings of existing studies through re-analysis, as new data generated from re-analysis can be seen as not valid. This is based on the assumption that only the original analysis is considered authentic and the re-analysis of it, via an archive, is partial and not entirely true. However, in those situations where the use of archived materials, drawing from a range of studies, aims to supplement researchers' own primary data, re-analysis is no longer an issue. An alternative view (Heaton, 2004a) argues that SA could be seen as a 'bricolage process' that often draws on multiple datasets, sources and methodological approaches and therefore, it de-emphasises the importance of personal involvement in the original research. Moreover, it is important that re-use of qualitative data is not perceived as a replication of qualitative research (Kuula, 2010), i.e. re-use is partial and, more importantly, it usually asks quite different questions from the original research. Additionally, the fact that in some disciplines re-analysis is a norm, but also a distinctive aspect of a disciplinary practice, challenges the distinction between primary and secondary analysis (Heaton, 1998; Fielding, 2004), which leads to the question of 'what are the boundaries between primary analysis and secondary analysis'? For example, with respect to researchers re-using their own data it may be difficult to determine whether the research is part of the original inquiry or sufficiently new and distinct from it to qualify as secondary analysis. In the case of secondary analysts re-using other researchers' data, issues about the degree of overlap between their respective work might also be present. Moreover, some authors refer to re-studies which start as secondary analysis but that could be termed as 're-use of data' since it constitutes new data (Moore, 2007), i.e., when re-using primary data these are continuously reconstructed and placed into new contexts, which might suggest that secondary analysis could be considered as primary analysis of a different set of data.

Mauthner et al. (1998) argue that the effect is that the initial context of the primary data, from the perspective of the original researcher, is no longer considered. The main arguments here are the concerns of taking part in data archiving and sharing practices based on foundational terms (Mauthner and Parry, 2009; Slavnic, 2013). From a foundational perspective, "data (knowledge) are conceptualised as separate from the subjectivities that generate them, and independent of the relational and inter-subjective contexts that give rise to them" (Mauthner and Parry, 2009, p. 294). In short, the main critique is that qualitative data preservation and sharing within the social sciences have been developed on implicit foundational terms, that is, the archiving of qualitative data has been addressed following the existing approaches to the archiving of quantitative data. In contrast, more scientific rationales for data archiving and sharing have also been taken up (Corti and Thompson, 2004). They argue that qualitative data preservation and sharing lead to 'better' science through scientific transparency, innovation and accuracy. Innovation is possible through archiving and re-use because new questions can be asked, and data can be approached in ways that were not originally approached. New themes, find-

ings and perspectives can be generated allowing for new understandings of the data (Corti and Thompson, 2004; Bishop, 2007). Additionally, and perhaps more importantly, Fielding (2000a) argues that secondary analysis can be less subjective since the analytic interests of the contemporary researchers will not have shaped, or influenced, the data collected towards particular analytic purposes.

2.7.1.1 The Importance of Context

The preservation of contextual information is one of the epistemological challenges of qualitative research archiving. Before defining 'contextual information', it is important to take into account and explore the different types of context that are relevant to archived data and (as well as) how to best document them. The first type of context is the context in which the study data collection took place, which can often be confused with the interpretive processes of the primary researchers owing to the effects of reflexivity in primary research data (Parry and Mauthner, 2005). Moreover, context and reflexivity are seen as intrinsic to the process of qualitative research, and therefore the degree of access to these has a key impact on the levels of re-usability of the research data (Moore, 2006). While for some researchers the issue of context is highly grounded in the epistemology of qualitative research, for those that are supporters of qualitative research archiving, the solution to the epistemological problem is a 'practical' one: the provision of context (Fielding and Fielding, 2000). These researchers see qualitative research data as the product of the reflexive relationship between the researcher and the researched. They agree that researchers are used to the multiple effects of reflexivity in primary research data, thus the provision of contextual features is more a practical issue rather than an epistemological one. The documentation of data, and its associated context at different levels, could potentially solve the issue of providing sufficient contextual information to archived data. The first type of context (data collection context) could be documented by fully reporting the issues related with research design and methods used. Examples of this type of documentation include elements such as an outline of the original study, data collection procedures, description of the processes of data categorisation and summarisation, and so on.

The second type of context, which is related to the interpretive processes that took place during the research, requires more complex description elements that differ considerably, depending on the meaning of the specific context taken into consideration. Different contexts such as political or policy contexts in which data were collected, methodological context and interpretive frameworks, or even research approaches used, need to be described differently. When looking at the multiple contexts in which the primary research takes place, different authors/researchers propose different approaches to context framing. Bishop (2006) compiles multiple framings by different authors and two main framings are highlighted. The first of these classifies context along a continuum of scale or distance and the other, considers context at different points in time, which implies looking in two dimensions: multiple levels of data granularity and across time. In the first framing, two main elements are considered when capturing context: the interaction itself (researchers and participants) and the overarching cultural fac-

tors, which are still related to the interaction since they are part of its context. Examples of this framing include the approach followed by [Holstein and Gubrium \(2004\)](#), which uses a framing of 'proximal' and 'distal' and highlights three key aspects: conversation or interaction; situation; and institutional or cultural elements. These three levels are equally important and mutually constitutional - each element depends on the other two to make full sense of the data.

Moving into the second proposed framing, context based on multiple levels of data granularity implies looking at the levels of interaction, between researchers and participants; the situation itself - settings researched and project characteristics; and culture/institutional characteristics. On the other hand, context analysed across time highlights the temporal nature of the contexts themselves. Context can be considered in at least two different periods: one can look at the context at the time of the original research (when the project was done), or when the re-analysis takes place (time of re-use). With respect to the context surrounding the interactions or conversations between researchers and participants, what the researcher doing the re-analysis is going to discover depends on the initial preservation when archiving, since it is clear that the secondary researcher will have less materials available. Additionally, the secondary researcher was not present at the time of the original research and therefore, some authors state that this implies losing an irreplaceable bit of context that rules out most of the data re-use ([Parry and Mauthner, 2005](#)). However, there are ways of recovering this context, like involving the original researcher(s) in some way, for example, inviting them to participate in the new project, by asking them questions of the study or by inviting them to comment on the findings from the re-analysis. Moreover, in order to maximise the archived contextual information and materials, it is very important to consider what elements to look at when analysing context and when the data is being produced: when exploring the background literature of the study, when creating samples, during interviews or while reading transcripts or composing analytical memos. Another context level - when framing data granularity - is the situational context. The concept of situation here, particularly for [Holstein and Gubrium \(2004\)](#), refers to the setting and there are multiple elements to consider when defining context at this level. In some cases, the project itself is included as a special part of the situation. The setting is normally the context of the 'qualitative work' and it captures a wide range of factors such as who is present, how the participants are related, the physical setting itself, and interactions like body language. Several of these factors are often not recorded or not archived due to rigorous research ethics, or fears of violating research ethics. This fact clearly affects and constrains what is archived or even recorded. Furthermore, most of this information is taken tacitly, therefore is very difficult to recall or record, and even if recorded, the information provided is minimal. In this respect, the specification of a minimal set of information that has to be completed will serve as a guide for researchers and could minimise the loss of this tacit and very useful information. For example, several archives like [UKDA](#) or [FSD](#) provide researchers with a set of guidelines for documenting these kinds of information when preparing data for their archive⁴⁶.

In relation to the primary research and the interpretive frameworks used, the analysis of con-

text is affected by the fact that Interpretivism in social sciences emphasises human subjectivities and the meanings people attach to the world when attempting to make sense of it (Cheshire, 2009). These meanings are socially constructed through interactions with others in specific historical and cultural contexts. To understand this process of meaning construction, Cheshire recognises that we need to understand the specific contexts in which the construction process takes place. She argues that archiving data separates them from the broader context of the study and from the important contextual clues that allow for correct interpretation. This could lead to misinterpretation of the primary data when re-using or re-analysing it, and raises the question of how best we could re-contextualise qualitative data in practice. Some authors recognise that this issue could be addressed practically by defining frameworks for expressing context and providing researchers with guidelines to provide context for archived qualitative research in order to facilitate its re-use and/or secondary analysis (Bishop, 2006). Moore (2007) suggests that secondary analysis is, in fact, re-contextualisation, and this highlights how important contextual information is to the process of re-using data. In this respect, both primary and secondary researchers have the responsibility to be reflexive in a manner suited to their specific projects. In the case of secondary analysis, reflexivity requires consideration of both the contemporary context and that of the original project (Fielding, 2004). Additionally, another response to the issue of misinterpretation of the primary data with the purposes of data re-use, or to conduct secondary analysis, is that the approach (SA) does not necessarily preclude the possibility of collecting primary data. This may be required to obtain additional data or to pursue in a more controlled way the findings emerging from the original analysis (Heaton, 1998). Consultation with the primary researchers, if they are available, in order to investigate the circumstances of the original data generation and processing can also help to solve the issue of limited contextual information as well as partial access to the original primary data.

Putting the debates about the usefulness or the feasibility of qualitative data archiving (with re-analysis and data re-use purposes) aside, most researchers agree that the value of archived qualitative data for re-use is highly enhanced when more detailed context is provided. The purposes of secondary analysis and the epistemological positions of the researchers provide different perspectives when looking at context. For some researchers, the provision of context aims at re-creating the context of the original study, and for others, like Moore (2006, p. 26), context is redefined or is looking at the re-contextualisation of 'the process of the production of data' rather than attempting to recreate the original context.

2.7.1.2 The Role of the Researcher and Ethical Issues for Archiving

Another epistemological challenge of archiving qualitative data is the role that researchers play during the qualitative research process. It is agreed that qualitative data is jointly constructed through the interactions of researchers and research participants. Additionally, opinions and subjectivities of the researcher are often bounded to the research data and therefore, it is important that they are preserved since they affect data collection and analysis processes. This co-constructed process is also influenced by the trust relationships that are built up with re-

search participants in the research project. While these relations can contribute to the richness of data, we have to bear in mind that, particularly in projects where the nature of the research is highly sensitive, researchers often perceive that the archiving of data might affect the interaction with participants (researchers feel the need to protect the participants) on the one hand, and on the other they assume that research participants would not accept the idea of archiving. Additionally, it is perceived that archiving could affect research participants' consent on the grounds that research data are planned to be used by others. Furthermore, even if consent was granted, archiving and the possibility of further analyses of the archived data could result in participants being portrayed in negative ways. However, those concerns could be partially solved by imposing very specific access conditions to the archive and if this was the case, both researchers and participants would have to make joint decisions when defining the access conditions (Fink, 2000). In relation to those issues, a number of studies have been conducted exploring participants' perceptions about archiving and research interaction, in which participants were also contacted for permission for archiving research data from studies in which they participated. (Kuula, 2005; Graham et al., 2007). Researchers' concerns are grounded in the relationships established with participants through the research and data collection, which are perceived as unpredictable and private, and participants are in need of protection. However, the results of these studies showed that in most of the cases research participants did not share some of the researchers' preconceived arguments against archiving and about the nature of the relationship (between researcher and participant). Moreover, the participants of these studies believed that they had control over the interview and they did not interpret qualitative interviews as "secret engagements that would hinder the archiving of the data for further use" (Kuula, 2010, p. 12). Instead, they perceived open access to research data for further use as self-evident and a way for them to engage in the advancement of science (Kuula, 2010).

Researchers involved with qualitative data archiving have been responsive to the issues of contextual information and how best to archive it, and have assisted data archives with defining best practices to improve the documentation of the research process and the generated data (Corti and Blackhouse, 2000). Most social science data are subject to confidentiality clauses, therefore issues such as informed consent, participants' anonymity and confidentiality must be addressed in the context of archiving qualitative material. Parry and Mauthner (2004, p. 97) state that:

"... the key issues [for archiving] involve respecting copyright and ownership, ensuring confidentiality and anonymity of respondents and researchers, and securing informed consent."

Although numerous debates around sharing qualitative data indicate that researchers are not very keen to archive their data, mainly because of ethics concerns, feedback from participants (Corti and Blackhouse, 2000; Kuula, 2010) suggests that, in some situations, researchers may be unnecessarily worried about participants' consent and the ethical issues derived from archiving. Indeed, participants, when asked, state that they have spared time - especially with qualitative

research - and expect use to be made of it. The latter refers back to the issue of 'overprotection' of participants, and researchers' assumptions of participants' protection and the possible implications of archiving for the participants and the research process. Additionally, other studies such as Thompson's reflections on The Edwardians project (Thompson, 2004), comment on the issues of anonymity and *participants' voices* and highlight the importance of distinguishing practices in different disciplines, i.e. oral history versus other sociological research. Participants were asked whether they would prefer to be cited anonymously rather than with their real names, and an overwhelming majority was very proud to be quoted by name. By not asking, researchers could be neglecting attention to the participants' context and their expectations from the research process. From the archiving perspective, it is crucial that the assurances made to the participants involved in the research projects are maintained. Whilst the archived data should only be made available for access when informed consent has been obtained, and anonymity and confidentiality are ensured, researchers involvement in discussions with participants, where possible, around the usage of data could help overcome any ethical issues, while at the same time, ensuring that participants get appropriate attention regarding their perceptions on archiving and data re-use.

2.8 Qualitative Research and Data Sharing

Along with the concerns mentioned above about research ethics, epistemologies and the issue of context preservation, the reluctance of qualitative researchers to archive their data is partially derived from the absence of any culture of data sharing. One important concern regarding data sharing is that archiving seems to be associated with an institutional management culture, which tends to add additional administrative tasks to researchers unless complementary resources are provided to assist them with the archiving processes. For example, there have been complaints in some countries, like the UK, where researchers applying for funding had to agree to provide data management plans in the early stages of the project and to archive their research studies once they have finished. While researchers have to accept this requirement when applying for money and for public accountability of research, they have also to bear in mind that the archiving process has to be driven by intellectual and methodological frameworks. Another important concern that qualitative researchers have about data sharing is that it exposes their research practices to others, which they may not be comfortable doing. This issue is underpinned by some researchers' view of qualitative research as a personal attempt, which is characterised for presenting a significant degree of personal involvement from the researcher in the process of constructing research data. Therefore, this can generate the fear that others may scrutinise very personal interpretations and declare them inferior (Broom et al., 2009).

In relation to the technologies, data sharing presents two major challenges. The first of these is producing exhaustive documentation about the collected/generated data and the research processes involved in conceptualising, collecting, managing, processing and analysing those data. Full documentation enables effective resource discovery - within data catalogues -

of distributed data sources and enables informed re-use. The second challenge for sharing data is facilitating data dissemination in the most flexible way so it enables using multiple access methods along with innovative uses of the archived data. Corti (2005) identifies that the previous challenges require high standard data collection, full documentation of research methods and practices (including the consent process), capture and preserve the context of the data collection. Additionally, it is very important to make available both the richness and structure of data and to implement mechanisms to capture the interrelationships between data and analyses alongside to represent and disseminate data in appealing ways, like for example presenting academic findings combined with evidence from the raw data (Corti, 2005).

The latter represents one attempt to solve the problem of capturing sufficient context by storing additional or supplementary contextual data alongside the primary data associated with the study. This way the risk of subsequent users of the archived materials misinterpreting the original research can be minimised. While this approach has been used to encourage researchers in some qualitative archives like Qualidata in UK, other researchers (Mauthner, et al., 1998) critique this approach since they argue that the inclusion of supplementary contextual materials does not really overcome the epistemological issues of reusing data. The provision of contextual information, supplementary information along with the original researcher's involvement in the archive construction process might provide a partial solution, to the issue of new researchers or subsequent users of data, who have not been involved in the primary research and therefore, have not shared the epistemological standpoints of the original researchers (Carmichael, 2008).

An in-depth exploration of a number of areas related to digital archiving of qualitative research data has been made. A certain degree of cross-over between existing technologies for digital archiving and qualitative data analysis; qualitative research epistemologies and research practices in relation to data archiving and re-use; and, research data management more generally was identified. The literature exploration enabled the identification of key technologies for digital archiving and semantic technologies suitable for the toolkit (and its underlying data model) developed in this study. The exploration of the perceptions of researchers and archiving communities in the social sciences to the archival processes for qualitative data and the use of technologies to support those processes has highlighted a number of barriers to as well as the benefits of qualitative digital archiving. These perceptions then formed the basis for a further exploration of the impact of researchers' epistemologies, and other factors, on the research process by analysing current archiving practices of researchers from the three different participant groups introduced in Chapter 1. Lastly, and perhaps more importantly, the literature review has highlighted the limited overlap that exists between research practices and archiving practices, in relation to qualitative digital archiving. Research practices tend to be separate from archiving practices, i.e. the development of research archives is normally performed at the end of a research project, and researchers are required to perform extra tasks in addition to those directly related to their research. In developing a self-archiving toolkit, this divide between research and archiving practices has been the main motivation for undertaking a design approach that has

been informed by existing practices of both archivists and researchers alike, but that also takes into account the best ways to document qualitative data so as to address the more technical issues described throughout this chapter.

Notes

- ¹http://ec.europa.eu/education/lifelong-learning-policy/framework_en.htm
- ²http://europa.eu/legislation_summaries/information_society/strategies/c11328_en.htm
- ³RDF is part of the W3C specification, considered one of the Semantic Web standards, and is a language for describing information resources in standardised ways.
- ⁴<http://www.w3.org/TR/owl-ref/>
- ⁵A triplestore is a purpose-built database for the storage and retrieval of triples, a triple being a data entity composed of subject-predicate-object, like “Bob has written an article” or “Bob attended to a conference” (*Triplestore*, 2013).
- ⁶<http://dbpedia.org/sparql>
- ⁷<http://bnb.data.bl.uk>
- ⁸The term is traditionally used to describe communities of practice using collaborative applications to build knowledge in a *bottom-up* fashion. See ‘Pyramid of Knowledge’ (Cleveland, 1982).
- ⁹<http://www.bl.uk/bibliographic/natbib.html>
- ¹⁰The project aims to investigate strategies which will ensure that the digital information resources typically included in library collections may, with other non-digital objects, be preserved over the longer term. Project information: <http://www.ukoln.ac.uk/services/elib/projects/cedars/>
- ¹¹<http://www.digitalpreservationeurope.eu/about/>
- ¹²<http://tools.ietf.org/pdf/rfc3986.pdf>
- ¹³<http://data.gov.uk/>
- ¹⁴<http://www.w3.org/standards/semanticweb/data>
- ¹⁵linkededucation.wordpress.com/
- ¹⁶<http://linkedup-project.eu/>
- ¹⁷The vision of the DRIVER project is to establish the successful interoperation of both data network and knowledge repositories as integral parts of the E-infrastructure for research and education in Europe. Project’s Official Site: <http://www.driver-community.eu/>
- ¹⁸Metadata is often defined as ‘descriptive data about data’
- ¹⁹<http://www.ddi-alliance.org>
- ²⁰<http://www.openarchives.org/pmh>
- ²¹<http://www.duraspace.org/>
- ²²<http://www.loc.gov/standards/mods/mods-mapping.html>
- ²³<http://mulgara.org/>
- ²⁴<http://www.w3.org/TeamSubmission/n3/>
- ²⁵<http://www.w3.org/wiki/ConverterToRdf>
- ²⁶<http://simile.mit.edu/wiki/RDFizers>
- ²⁷<http://projectblacklight.org/>
- ²⁸Eprints are the digital texts of peer-reviewed research articles, before and after refereeing. <http://www.eprints.org/openaccess/self-faq/#What-is-Eprint>
- ²⁹<http://www.surf.nl/en/projecten/Pages/EJME.aspx>
- ³⁰<http://www.surf.nl/en/projecten/Pages/EscapeVisualisationcomponent.aspx>
- ³¹HRAF- Human Relations Area Files, aiming at cataloguing the results of anthropological research and to make the datasets available for secondary historical and cross-cultural research.
- ³²ESDS Qualidata is a specialist service of the ESDS (Economic and Social Data Service) led by the UK Data Archive (<http://www.data-archive.ac.uk>) at the University of Essex. The service provides access and support for a range of social science qualitative datasets, promoting and facilitating increased and more effective use of data in research, learning and teaching.

³³<http://www.murray.harvard.edu/>

³⁴<http://relu.data-archive.ac.uk/>

³⁵Timescapes project is the first major qualitative longitudinal study funded in the UK. It is funded by ESRC and ran during 5 years, starting in 2007. Project's website: <http://www.timescapes.leeds.ac.uk/>

³⁶<http://www.ddialliance.org/alliance/working-groups#qdewg>

³⁷<http://www.dedoose.com/>

³⁸<http://www.maxqda.com/>

³⁹<http://thedrs.sourceforge.net/>

⁴⁰<http://www.caret.cam.ac.uk/blogs/llp/>

⁴¹<http://www.accessgrid.org>

⁴²<http://www.cendari.eu/>

⁴³<http://www.spss.com.hk/>

⁴⁴<http://www.stata.com>

⁴⁵For a more complete overview of some of the most popular CAQDAS packages see Table B.1 in appendix B.

⁴⁶UKDA, interview transcription guidelines:

<http://www.data-archive.ac.uk/create-manage/document/data-level?index=2>; or FSD, Anonymisation of qualitative data guidelines: <http://www.fsd.uta.fi/en/anonymisation/index.html>

RESEARCH DESIGN AND METHODOLOGY

This research and development study is concerned with the areas of qualitative research, digital archiving and semantic technologies. It has involved the development of a set of tools to support digital archiving, analysis and re-use of qualitative data, as well as the collection, analysis and interpretation of qualitative data gathered from semi-structured interviews. A qualitative exploratory case study approach was adopted, so as to gain a greater understanding of the topics (current research and archiving practices of qualitative studies) from the point of view of specific respondents from three different groups, and to inform the design and development of the self-archiving toolkit by obtaining detailed descriptions of the participants' research and archiving practices. The aims of this case study were firstly, to obtain example data models that could be used to test and refine the toolkit's data model and the archiving processes that would be required of users of the toolkit to create research collections and secondly, to identify potential scenarios for its use in these, and similar, research settings. The first group consisted of final year undergraduate students conducting small-scale research projects. For this first cohort, the topics of interest were the development of their research projects, the practices they employed and the kinds of materials they produced. The second group was made up of researchers-lecturers who are involved in both research and research-informed teaching and learning activities. In this case three main topics were covered: a) the description of a research project of their choice, with a focus on archiving practices and the kinds of materials produced in this project, b) the participants' research practices and their perceptions of qualitative archiving and data sharing, and c) their teaching practices in research-oriented undergraduate level modules. The third group consisted of two senior researchers who were involved in the development of qualitative research archives for different purposes. With regards to this third cohort, the topic explored was the design and development of digitally archived case studies, with a focus on the design decisions that were taken, how the development stage was approached, and the purposes and uses of the archives that were developed.

An exploratory method was selected to gain new insights, discover new ideas and increase knowledge of the practices of participants during the development of research projects (all cohorts), research and teaching activities (second cohort only), and the design and development processes of qualitative digital archives (third cohort only). The research approach used by the author was primarily concerned with obtaining detailed descriptions of the processes that were followed by participants in the course of their research. This aimed to develop a better understanding of how the participants approached the different stages of their projects, the decisions taken by the participants and the rationale for those decisions (Yin, 2009). Additionally, this approach was taken so as to explore the potential benefits that self-archiving tools, such as the toolkit implemented in this study, might have for researchers undertaking the types of processes and activities encountered in the research environments considered here.

3.1 Context of the Research

The technological aspect of this study was the development of a self-archiving toolkit to document qualitative data in standard ways, and to enable data sharing and re-use. The motivations for developing such software tools have been the unresolved issues that arise during the development of archives for qualitative research. The explored literature combined with the the author's own experience of developing archiving tools, as well as the work of the DDI Qualitative WG, have all helped in identifying these issues. The literature review highlighted that, while a number of widely adopted standards and documentation vocabularies exist for the description of quantitative data, these are lacking when it comes to describing qualitative data. The work of the DExT project and the DDI Qualitative WG, which have been described in Chapter 2, have attempted to provide partial solutions to this issue. The DExT project developed an XML-based schema for the documentation of qualitative materials and analytical information. This schema was designed to enable the exchange of those kinds of information within QDA software or CAQDAS packages. The complex XML model developed by the DDI Qualitative WG attempts to provide a standard schema for the archiving and exchange of qualitative data in digital form, which integrates the documentation of research processes and activities with the data generated during those processes. These kinds of models, or schemas, and their associated tools are necessary to support complex and large-scale research projects as introduced in Chapter 1. For such studies, the archiving processes, along with the design and implementation of data management plans, requires the researchers to perform an extra set of tasks in addition to those directly related to their research. These extra tasks can be both unfamiliar to the researcher and time-consuming, and since a researchers' expertise lies in their particular field of research there are arguments in favour of delegating archiving and data curation duties to archivists and data centres. However, there remains a concern for those small-scale projects, or individual researchers, which do not have the funds or the infrastructure to preserve and look after datasets after the completion of the research. This study therefore seeks to address this particular problem by implementing a self-archiving toolkit and an underlying data model that facilitates the documentation of qualitative data in standard and consistent ways, in order to facilitate data

sharing and re-use. At the same time the toolkit seeks to provide users with a simpler means of building a digital archive than would be the case for users involved in larger, more complex research projects. Additionally, the introduction of archiving practices within the research process could facilitate, firstly, an early identification of ethical and methodological design issues, such as those described in Chapter 2, which are often associated with the lack of early data preparation for subsequent archiving. Secondly, it could facilitate awareness of data documentation standards and their impact on the dissemination of research.

3.2 Research Objectives

The literature exploration of this study involved investigative research to:

- identify key technologies in the areas of digital archiving and semantic technologies that are suitable for the set of tools - and the underlying data model - that are developed in the present study to support researchers with research archiving processes, and
- explore the perceptions held by researchers and archiving communities in the social sciences of the archival process for qualitative materials and the use of technologies to support those processes.

A range of journals have provided critical background commentary on the methodologies of qualitative research by secondary analysis and in particular on its relation to and impact on the re-use of qualitative archived data. The initial exploration of the available literature found, in the Forum Qualitative Social Research (FQS) Journal ¹, a rich source of materials relevant to qualitative archives, secondary analysis and re-use of qualitative research data. There has been ongoing debate, largely but not exclusively in FQS, into the barriers to and potential benefits of qualitative archiving that concerns secondary analysis of existing research data (Corti et al., 2005; Corti, 2006; Moore, 2006; Parry and Mauthner, 2005; Bishop, 2005; Hammersley, 2010). The initial literature review was also used to determine the questions that were most relevant to the topics explored. Additionally, online journals and databases in the field of Computer Science, such as the Institute of Electrical and Electronics Engineers (IEEE) Computer Society, were consulted during the design and implementation of the software tools developed in this study. Since the field of Computer Science is one that is continuously changing with new developments being carried out all the time, an important part of the reviewed literature consists of working papers, technical documentation and online resources.

The specific practical objectives of this study were to:

- develop a set of online tools to document, archive, explore and re-use qualitative data in research settings in which researchers (and students) are engaged in the development of research projects,
- analyse practices and key elements within the processes of design and development of qualitative digital archives by studying existing complex and challenging digital archives,

- gain a rich understanding of the impact of researchers' epistemologies, and other factors, on the research process, exploring key issues researchers face in the course of qualitative studies. This objective focused especially on the design, data collection, analysis and use of the data collected, as well as on identifying potential benefits of and barriers to digital archiving.
- analyse existing practices within research-based teaching and learning environments, with a focus on the sort of materials used by lecturers to support teaching activities; and to investigate their perceptions of the benefits of using existing archive materials to enhance students' understanding of research projects and the research process.

3.3 Research Design

An extensive literature review of existing technologies in the areas of digital archives and data documentation, and technological approaches undertaken in the development of digital archives was conducted in relation to the technological work. This review informed the initial design of the toolkit developed in this study. The author's participation as an active member of the [DDI Qualitative Data Exchange Working Group](#) (in the role of a model architect) and the exploration of a set of use cases from various research archives (some of them presented in the [DDI WG](#)) served to obtain key input on the design of the toolkit's data model. The [DDI](#) group provided use cases from existing archives, as well as the space to discuss the challenges and limitations of current technologies, used in the development of archives, and existing practices of qualitative data archiving. These aimed to explore the kinds of data held in existing archives that researchers produce and use, across different disciplines within the social sciences, to identify key priorities and basic requirements for the developed self-archiving toolkit in order to:

- support the documentation, organisation, storage and visualisation of qualitative materials,
- support the documentation of contextual information for qualitative materials,
- support data exchange/sharing with other systems (e.g. with [CAQDAS](#) packages, or research data published online), and
- incorporate mechanisms for the documentation and management of a basic set of analytical data.

A more detailed description of the technological aspects of the design of the toolkit, the implementation processes and the technologies involved is provided in Chapter 4. The exploration of use cases from existing qualitative research archives contributed to the design and implementation of a generic data model which supports the documentation and storage of qualitative research materials, basic analytical information and the different kinds of relationships between qualitative materials, in ways that facilitate data sharing and re-use. However, as described in Chapter 2, those activities required for the archiving of research materials, when performed by

archivists or data curators, often differ from those performed by researchers in the course of their research projects. This consideration is particularly important in the design of the archiving processes supported by the developed toolkit, and thus required further analysis of a range of research practices used in primary analysis, archiving and re-use of qualitative data. In this respect, rich descriptions of the data organisation and archiving processes followed by the participants were needed so as to obtain detailed input for the design of a flexible model of archiving processes required of users of the toolkit so as to support a range of archiving practices of researchers with varying levels of expertise. At the same time, such a design and implementation still meets the more general requirements and practices of well-established qualitative research archives. This was achieved by exploring the different approaches taken at each stage of the research process by each participant. Special attention was paid to data organisation and archiving activities and the role these played in their research, as well as any issues encountered during their research. The information that needed to be gathered from each group of participants to inform the design of the archiving model was organised according to the different stages of the research process: research and data collection instruments' design, data collection, analysis and interpretation, and presentation or publication of results. Aspects that needed to be covered included: detailed description of the process; the kinds of materials produced and how these are described and related; data organisation and archiving practices and their role/impact on the research process; and noteworthy issues, difficulties encountered during the development of the projects. The research design of the empirical case studies is described in more detail in the rest of this section.

According to Yin (2009), five components of a case study research design are especially important:

- a study's questions;
- its propositions, if any;
- its unit(s) of analysis;
- the logic that links the data to the propositions; and
- the criteria for interpreting the findings.

In relation to the first component, a case study approach is appropriate when 'how' and 'why' questions are being posed. For this study, interviews with students participating in small-scale research projects were used to gain a better understanding of the research process and the issues they faced while developing their projects. For example questions such as 'how students organised and structured the data gathered', or 'why they selected particular data collection methods' were of particular interest. Additionally, an in-depth exploration of the kinds of materials they produced, and their data organisation practices, was needed to identify additional requirements for the archiving processes and to obtain examples against which to test the toolkit

developed in this study.

As for the second component, the purpose of identifying propositions is to direct attention to those aspects that should be examined within the scope of the study (Miles and Huberman, 1994b), as well as to limit the scope of the cases (Baxter and Jack, 2008). In this respect, the interviews with researchers involved in teaching undergraduate level research modules were conducted to gain a rich understanding of their practices when conducting qualitative studies, the issues they faced and their perceptions of qualitative archiving and data sharing. The topic of qualitative archiving with a focus on data sharing and re-use was central to this study. The literature review highlighted a number of issues in relation to this topic, which needed to be addressed. Examples of the kinds of propositions that were developed included 'the importance of documenting the context of the research when archiving qualitative data', or 'highly interpreted data represents a barrier to qualitative data archiving'. For the researchers involved in teaching the aim of those interviews was to develop an understanding of their practices, and the kinds of materials they used to support students conducting small-scale research projects so as to explore how the toolkit might support such activities. These aspects were more exploratory, and therefore the identification of a set of propositions was not necessary (Yin, 2009; Baxter and Jack, 2008) since the purpose of the exploration was clearly established.

The units of analysis (third component) of this study were introduced previously as: undergraduate students conducting small-scale research projects (first cohort), researchers involved in teaching research-oriented modules (second cohort), and senior researchers who had been involved in the design and development of digitally archived case studies (third cohort). The interviews from the first cohort were designed to reveal students' practices and the processes they followed when approaching the development of their projects. The interviews with researchers involved in teaching were similar to the first cohort in that the participants selected a research project of their choice to discuss. Additionally, it was important to explore the research practices of these participants and their perceptions of qualitative digital archiving, especially when the aims are data sharing and re-use. Lastly, interviews with senior researchers who had conducted research projects in which multimedia case studies were developed, were carried out to develop an in-depth understanding of the processes followed during the design and implementation of the archives developed during these studies. The first archive was a multimedia case study exploring a school-based research setting, which was developed to support Master's level students undertaking action research. The second digital archive was a set of multimedia case studies that was developed in the course of a large-scale research and development project. This project was concerned with assistive technologies and Human Computer Interface (HCI) and Brain/Neural Computer Interfaces (BNCIs). The rationale for the selection of these digital archives was to explore: a) different models of archiving, b) the role of technologies in the archiving processes, c) key issues faced during the selected archives' development, and d) the uses and purposes of those archives. As with the student interviews, the purpose of interviews with the senior researchers was to identify additional requirements and to obtain examples against which

the toolkit could be tested and optimised.

Lastly, the analytic strategy which corresponds to the last two components of the design recommendations for case study research (Yin, 2009) that was followed was cross-case synthesis. Cross-case synthesis is an analytic technique that applies specifically to multiple-case studies (Yin, 2009, p. 156). Cross-case synthesis is also likely to produce more robust results as opposed to having only single cases. For this study, cross-case synthesis was used to facilitate a comparison of aspects, such as participants' perceptions of the benefits of and barriers to qualitative data archiving and sharing, or data organisation practices, across the student and researcher participants (first and second cohorts). As for the third cohort, the purpose of the interviews with these participants was to explore the design and development of two particular multimedia case studies in which the nature of the data and the purposes of the archives differed substantially. Here the analytic approach was concerned with obtaining rich descriptions of the design and implementation approaches followed, rather than with comparing the two cases.

The rest of this chapter describes in detail the aspects of the research methodology such as data collection methods and data analysis activities, which constituted the empirical work of this study.

3.4 Locating the Research Participants

According to Hycner (1999, p. 156) "the phenomenon dictates the method (not vice-versa) including even the type of participants". Purposive sampling, considered by Welman et al. (2006) as the most relevant kind of non-probability sampling, was chosen to identify the primary participants. However, the subjectivity and 'non-probability' based nature of the unit of selection in purposive sampling often means that it can be difficult to defend the representativeness, or appropriateness of the sample. In this respect, the number of participants, that is 10 participants from the three cohorts introduced previously, seemed sufficient and appropriate for the following reasons. Firstly, the participants were selected to provide a purposive sample that would represent a range of practices. The selection was based on conversations with researchers and staff involved in research and research-oriented teaching and learning environments, the kinds of projects in which the participants had been involved in the past, and 'the purpose of the research' (Schwandt, 2007, p. 271). The students who participated in this study had conducted small-scale research projects, in which they typically used mixed-methods approaches to explore different topics, which included aspects of educational practice, the implementation of national or local policy, or the experience of learners in different educational settings. The researchers (second cohort) who were selected were also engaged in a wide variety of research projects. Additionally, their research approaches ranged from phenomenological, ethnographic research to participatory approaches, or action research. This could allow for comparison of their perceptions of qualitative archiving, data sharing and re-use. Lastly, the two existing qualitative archives that were selected differed considerably in terms of their content, the purposes of

their development, and their intended uses. The purpose of collecting data from three different groups of participants was to achieve ‘data triangulation’ to corroborate evidence from different sources so as to shed light on a theme or perspective thereby providing ‘validity’ to the findings (Creswell, 2013, p. 251).

Researcher participants included LJMU students and staff, and senior researchers based at other institutions. The inclusion criteria for the students and staff was that students were participating in a research-oriented module (the Independent Project² from the Education Studies degree at LJMU) and that researchers were involved in that module (or equivalent) either as lecturers or project supervisors. Initially, informal conversations were held with these potential participants to explore the nature and scope of their involvement. More detailed information such as ‘participant information’ sheets were provided electronically via email. Student participants were recruited as part of the established discussion with staff who run the modules and supervise the student projects which accompany Independent Study Modules (another component of the Education Studies degree).

With regards to the senior researchers, participants were identified and recruited through existing networks of researchers who already worked with the ESRC/Engineering and Physical Sciences Research Council (EPSRC) funded Ensemble project³ and with the ESDS⁴ Qualidata Archive⁵. The inclusion criteria for this group was for principal investigators of research projects that led to the development of qualitative digital archives. Potential participants who were involved in highly contentious or confidential research studies were excluded. This was due to the nature of the research conducted in such projects, which did not lend itself to the development of publicly available archives; as well as the approach of this study which aims at obtaining in-depth descriptive data.

In accordance with LJMU ethics guidelines, a full application was submitted to and approved by the Research Ethics Committee of the University (Application for Ethical Approval of Undergraduate, Postgraduate or Staff Research involving Human Participants or the Use of Personal Data). Informed consent was used, including participant information sheets and consent forms. The participant consent form outlined:

- that they were participating in research,
- the purpose and procedures of the research,
- the risk and benefits of the research,
- the voluntary nature of research participation and the participant’s right to stop the research at any time and,
- the procedures to protect confidentiality: use of access-controlled digital environments; personal data were kept confidential and not reflected in subsequent publications or reports.

The consent form was explained to participants at the beginning of each interview. Most potential subjects signed the agreement, and those who did not, were not pressured to participate in the study. All who became participants were in agreement with its content and signed.

3.5 Research Instruments: Data-Gathering Methods

Semi-structured, in-depth interviews were conducted with the above mentioned groups of participants. Kvale (2007, p. 10) defines qualitative semi-structured interviews as “attempts to understand themes from the subject’s own perspectives”. Interviews should therefore seek to obtain descriptions of the interviewees’ experiences with respect to the topics explored. For this study, the interviews were used to obtain rich, descriptive information about researchers’ practices, the processes involved and the issues faced during the development of their research projects. While an ‘interview schedule’ including topics and questions that needed to be covered was developed, the interviews allowed for an open discussion to develop so that new and emergent themes could be explored. Prior to the formal interviews, informal conversations were held with the participants. For the student participants, these conversations were held with their module leader or project supervisor(s). The informal discussions in combination with an extensive literature review by the author provided a keen understanding of the topic of interest, and facilitated the development of relevant questions for the formal interviews. Semi-structured interviews were considered the most appropriate data collection method since the intent of the approach was to understand the topics of interest (development of a research project, research archive, or teaching practices with the support of archived research materials) in the participant’s terms, and to obtain a ‘description of the experience as it is experienced by the person herself’ (Bentz and Shapiro, 1998, p. 96). Additionally, interviewing also facilitates probing, i.e. interesting themes that emerge during the interviews can be followed up to understand as much as possible about the subject and the participants’ perspective (Welman et al., 2006). Note taking during interviews can make it difficult to focus on the interview itself and could result in poor note taking and also hinder the development of a good rapport between interviewer and interviewee. To avoid this, interviews were audio-recorded, without additional notes being taken besides reminders or specific instructions such as “go back to question X”.

3.5.1 Interview Design

Three different interview schedules were developed, one for each cohort of participants. For the student researchers (first cohort), the questions focused on the participants’ own experiences, and were structured according to the different identified stages of their research projects: research proposal and research design development; literature review; data collection and organisation; analysis and interpretation; and lastly, dissertation writing. Specific attention was given to the participants’ perceptions of research in the context of their project, and the aim therefore was to have them to describe in depth all the processes involved during the development of their project as well as the issues they faced along the way. The interviews were reciprocal in that both the researcher and the participant engaged in the dialogue, in which views and

perspectives around a theme of mutual interest were shared. Additionally, each participant was asked to bring along a set of materials of their choice (produced in the course of their project, i.e. analysis documents, research instruments, anonymised data) to facilitate discussion during the interview, but also to provide additional data. For example, participants who brought some of their analysis documents, or summaries of articles that were consulted, provided additional information about how they described and classified their data. This allowed for comparison of their practices with those of potential users of the toolkit implemented in this study, so as to test and identify new requirements.

For the researchers/lecturers (second cohort), the questions in the interview schedule fell into three broad categories:

- Description of a research project of their choice,
- Research practices, epistemologies and researchers' perceptions of data archiving (further research or secondary analysis) and,
- The kinds of materials that were used to support practices in teaching undergraduate level research-oriented modules.

The first set of questions sought to explore the kinds of projects and areas of research in which the participants were involved, by focusing on a specific medium-to-large scale project of their choice. The interview was designed to focus on the kinds of data produced in the course of the project in relation to data archiving. For those projects that ultimately produced a research archive, the focus was on data management and the barriers to and benefits of data sharing and dissemination. For those projects that did not lend themselves to data archiving, the focus was on research practices and the participant's perception of archiving, potential or identified issues with or barriers to data sharing in the context of their project. Additionally, the exploration of their experiences with collaborative work was also of interest owing to its inherent relationship with and impact on research archiving and data sharing.

The second set of questions dealt with the participants' research practices more generally, focusing on aspects such as the kinds of materials produced in the course of research, their own data organisation and archiving practices, and their perceptions of archiving of qualitative research data, with an emphasis on secondary analysis and data re-use. The final set of questions dealt with the materials and approaches that participants use to support teaching research methods to undergraduate students: the participants were asked to describe in detail those modules related to research; the kinds of materials used to support these teaching activities; and their perceptions of using existing archived materials as a way of supporting students' understanding of research projects. Additionally, the last set of questions was concerned with the issues faced by students during the development of their own projects and how they were helped to overcome those. This linked to some of the topics covered by the interviews with the

undergraduate students.

Finally, the interview schedule developed for the interviews conducted with researchers who developed archives was less structured than those developed for students and researchers. This, more flexible approach, made it possible to ask further questions beyond the ones originally planned, and to follow up emergent themes of interest, and to clarify the meaning of participants' responses. The two multimedia case studies were selected for detailed analysis of the approaches undertaken in the design and development of case study archives. These archives compiled very rich data: the first archive included a range of qualitative, survey and technical data, and the second included mostly qualitative data (audio and video interviews with a wide range of participants) as well as statistical data. Consequently, there was a focus on discussing aspects such as materials selection and structure of the archive, the intended uses and purposes of the developed archives, and the issues and advantages of sharing results and data. Other aspects, such as a description of the issues faced during data collection and organisation processes as well as the technologies used for the development of the archives were also discussed.

3.6 Data Storage Methods

All of the interviews were audio-recorded with the permission of the participants. Each interview was assigned a code based on the data and the type of participant, for example "Student1-interview-date" for an interview audio file and "Student1-interview-date-transcript" for the associated transcript. Whilst transcription is often part of the analysis process, it can also enhance the potential for sharing and re-using qualitative data. To achieve this it is recommended to produce full transcriptions and to supplement them with contextual information in the form of supplementary information attached to the transcription file (UKDA, 2013). This usually includes information such as a unique identifier for the interview; event details like date, place or interviewee details; and anonymised references to participants. It is recommended that the level of transcription complements the level of the analysis in those cases for which analysis focuses on providing in-depth description of the knowledge, attitudes and experiences of an individual (McLellan et al., 2003). Consequently, full verbatim transcripts were produced, word by word, retaining frequent repetitions, and including pauses or emotional expressions which provide preliminary analytical information, such as participants' reflection on or deep thinking about particular topics. The interview transcription files also include a summary section at the beginning which contains information about the participant, interview date and brief details of the interview.

The data collected from each interview include a variety of hard copy documentation and electronic files generated during the interview itself and provided by the participants. For each interview, the following kinds of data were produced and stored:

- The informed consent agreement (both a hard copy and a scanned electronic copy),

- Any additional information and documentation that the participant offered during the interview, for example research materials such as sample interviews, questionnaires or analysis documents,
- Any notes made during the author's analysis, e.g. emergent themes, relevant quotes grouping or interview summaries with analytical notes and,
- Raw transcripts and 'analysis' of the interviews.

The electronic data was stored in a secure and password protected online storage space, on a laptop and on an external hard drive. The electronic data stored on a laptop and external hard drive was secured by using authentication and data encryption mechanisms.

3.7 Data Analysis

In the data analysis process, the transcript(s)⁶ from each participant were reviewed a number of times to allow for a holistic picture of each participant's experience to emerge, while at the same time facilitating initial comparison between participants. Following this review, an initial summary for each participant was produced. For example, the summaries included brief descriptions of their projects, the kinds of data that were produced/generated, and a summary of the participants' practices during the different stages of their projects. This facilitated the creation of summary tables such as the ones presented in Chapter 5. This process was followed for the transcripts from the three cohorts of participants.

Following production of the initial summaries, the transcripts were analysed and a list of significant statements from each participant was developed: different emerging ideas were highlighted (and colour-coded by theme) and separated whenever one idea ended and a new one began. These elements were also accompanied by annotations in the form of descriptive notes or analytical comments. The elements were then labelled as part of a theme, one such theme being 'perceived issues of qualitative data sharing'. For example, the quoted text below is taken from the response of one participant with the set of relevant ideas highlighted.

“research should be shared as widely as possible. I think you would need to be careful about the level of analysis that you present because things could be interpreted [pause] in a negative way. If you just present the raw data, or if you just presented it half-formed and it doesn't have the right context... so I would probably would have quite a lot of labelling and contextual information related to the data...”

This process was repeated a number of times for each transcript to ensure consistency of analysis between interviews from the same cohort. Additionally, these revisions also served to identify elements that were out of scope or that were repeated and could therefore be omitted. For example, as mentioned previously, there was one student who was interviewed twice. In the second interview with this student some elements related to data collection that had been covered in the first interview. Those statements were carefully analysed and compared with those

from the first interview to ensure that they were not revealing new issues or elements of interest, and those that were repeated were therefore not taken into account.

With regards to the first two cohorts of participants (students and researchers), once the relevant elements of each transcript had been revealed, common elements were highlighted. For example, all comments relating to the participants' data organisation practices once they collected their data were noted, as were comments referring to issues encountered during that process. These common elements were then organised into meaningful sets. The use of semi-structured interviews was particularly useful for this organisation process since the scope was quite focused. This facilitated, for example, the organisation of the different statements from the participants by 'stage of the process', so that the participants' practices and issues identified could then be associated with a particular stage of the research process. In contrast, the transcripts from the two senior researchers were mainly analysed individually. As mentioned earlier, these interviews were more open. In this case, the analysis focused on providing rich descriptions, including verbatim examples, of the researchers' practices and processes followed in the design and development of those archives. However, the analysis of the first multimedia case study revealed ideas and issues that were related to some of the themes that emerged from the interviews with researchers from the second cohort. Those common aspects are described in Chapter 5.

Lastly, cross-case synthesis was used to determine what, if any, common themes were present within the first two cohorts. This allowed for comparison to be made between participants from within these two cohorts. For example, this proved particularly useful for comparing the data organisation processes that the students followed in order to draw cross-case conclusions concerning some of the aspects explored. For the participants from the second cohort, this also helped in making comparisons between participants' perceptions of the benefits of and barriers to qualitative data archiving and sharing. The results of such comparisons are presented in summary tables in Chapter 5.

3.8 Conclusion

The development of a self-archiving toolkit to document, archive, explore and re-use qualitative data, so that it provides researchers with a sufficiently accessible framework for doing so, has involved design and development activities informed by present technological issues faced in the development of qualitative digital archives as well as current practices (research and archiving) of researchers with varying levels of expertise.

More technical input from use cases from existing research archives helped to define base requirements, which take account of key issues that are faced when preparing, documenting and archiving qualitative research materials, in order to guarantee that the toolkit incorporates mechanisms that can help to address or overcome such issues. The key issues are summarised

as follows: a) lack of common documentation standards for qualitative data, b) limited support of existing archiving systems for accurately describing qualitative data, and c) limited support of existing archiving models for describing the relationships between the generated/archived qualitative materials. Moreover, another issue to the archiving of qualitative data is that the design and development of qualitative archives is often driven by archivists or specialised data centres, without taking into account researchers' practices. That is, archiving and research practices are separated: the production of a research archive is performed as the final step, and once the investigation is complete.

The complexity of the processes of data documentation, research description and archiving of qualitative materials can be quite challenging for researchers, especially in scenarios such as the ones explored in the research settings of this study: undergraduate students conducting research, or researchers involved in small-scale or individual research projects. Firstly, the tasks associated with digital archiving introduce a further level of complexity and requires additional efforts and time-consuming tasks on the part of researchers. Secondly, the introduction of archiving elements within researchers' practices is essential for data sharing and re-use although it can also introduce procedures that may require researchers to perform tasks in a different manner. In this respect, the empirical case studies explored in the course of this study helped to minimise the impact of the introduction of archiving practices in addition to those directly related with research. The empirical case studies informed the design of the self-archiving toolkit developed in this study so that it provides researchers with a basic framework to create and store qualitative data collections in simple ways and by following processes with which students and researchers are already familiar.

Notes

¹<http://www.qualitative-research.net/index.php/fqs>

²The 'Independent Project' is an optional module included in the Education Studies degree, and its specialisation routes - Early Years Education, Physical Education and Special and Inclusive Needs. In this module students are introduced to research methods and theories and conduct small-scale research projects involving empirical work and the production of a final dissertation.

³<http://www.ensemble.ac.uk/wp/>

⁴<http://www.esds.ac.uk/>

⁵<http://www.esds.ac.uk/qualidata/about/>

⁶During data collection, there was one participant from the student cohort who was interviewed twice. The first interview was conducted when the student had finished data collection and had just begun analysis and writing. The second interview was conducted once the project was finished and the student had submitted their dissertation.

QUDEX REPOSITORY: A SELF-ARCHIVING TOOLKIT FOR QUALITATIVE STUDIES

The system that has been implemented in the course of this study, ‘QuDEX Repository Tools’, is a set of online applications designed to help users, familiar with online information resources such as data archives, digital libraries, with a requirement for archiving, searching and disseminating collections of qualitative data, regardless of the type of media. While useful for any kind of research community, these tools have been specifically designed to be sufficiently accessible that they could be used by non-experts, such individual students conducting small-scale research projects; or researchers involved in teaching undergraduate level research-oriented modules. They provide users with easy ways of organising, documenting, and archiving their research data so that the created collections can be visualised and exposed in ways that enable their integration with the increasing ‘web of linked data’ (Bizer et al., 2009), as well as being presented in support of dissertations, as part of electronic portfolios, presentations or enhanced publications.

The design and implementation of this kind of tool has to take into account a wide spectrum of aspects from three different areas. The first of these areas is of a technological nature. A number of open-source technologies, for digital archiving, semantic databases, and visualisation frameworks, have been explored and evaluated, focusing on aspects such as:

- the degree of adoption within different communities (software communities, digital archives, research communities) and active development and engagement,
- compatibility between the different technologies that underpin the system, and
- the ease of integration and deployment.

The second area, which was described in detail in Chapter 2 (section 2.6.2), covers data documentation and data exchange standards. Data documentation standards are highly important within information systems ranging from digital libraries, archives, or publishing systems to less specialised systems, such as learning management systems or virtual research environments. These standards provide a means to accurately describe data in digital forms, facilitating their dissemination and discovery across the web and enabling data sharing and exchange between different information systems. While there are a number of widely adopted standards for describing different kinds of data, like for example, Dublin Core for publications, DDI for quantitative and statistical data, or OAIS for archival systems, there is no agreed documentation standard for describing qualitative data. Additionally, the increasing adoption of CAQDAS packages for managing and analysing qualitative data in digital form, along with the issues explored in Chapter 2 (section 2.6), are key motivations for the emergent work on the implementation of qualitative data exchange schemas and archiving models. Open qualitative data exchange models could provide a solution to problems arising from:

- the use of proprietary formats between CAQDAS packages,
- the differences in the terminology used to describe data and their relationships,
- the lack of support for online and collaborative modes of work, and
- the lack of interoperability between different software solutions.

In this respect, this study and the proposed tools seek to contribute to this emerging area, by implementing an underlying data model that facilitates the documentation of qualitative data in standard and consistent ways. This is because it makes use of a number of existing data documentation standards, and it provides mechanisms for integrating/exchanging data collections with other systems. The latter could potentially allow to re-use qualitative collections stored with the support of the toolkit with QDA tools for further data analysis.

Lastly, the third area is related to qualitative data archives, focusing specifically on aspects such as data management practices, the kinds of data included in qualitative data archives, and the ways in which these are described and presented via online data catalogues. In this respect, being an active member of the DDI Qualitative Data Exchange Working Group has provided highly important insights, related to existing practices of qualitative data archiving carried out by well-established international data archives, such as UKDA, FSD in Finland, or ASSDA in Australia, that have informed the design of the toolkit developed.

4.1 The Toolkit: Technical Overview

This section provides an overview of the technologies underpinning the toolkit developed in this study: Qudex Repository Tools. Appendix A includes the technical documentation of all the components of the self-archiving toolkit, in which aspects such as the toolkit's architecture, the

code libraries that compose the toolkit, and the toolkit's functionalities are described in detail. [QuDEX Repository Tools](#) are deployed in an application server, e.g. [Apache Tomcat](#)¹, and rely on a number of existing open-source tools that are in common use in data archives and institutional repositories. These include digital repositories and [RDF](#) databases; metadata standards and [RDF](#) vocabularies; or the [QuDEX](#) schema, published by the UK Data Archive for describing [CAQDAS](#) metadata and qualitative data collections. These tools are entirely developed in [Java](#)² and were released under the [Apache 2.0](#) license³. These tools rely on and communicate with two different external systems that need to be installed and configured before installing these tools. The systems are:

- *Fedora Digital Repository*. Across a wide range of disciplines, a key enabler of semantic web adoption more generally is the provision of flexible, configurable digital repositories which are 'semantic web ready' as they have interfaces, such as [OAI-PMH](#)⁴ for disseminating metadata about the collections they handle, or they provide mechanisms for storing metadata in semantic databases such as triplestores. Following analysis of the functionalities of several repositories, Fedora was selected as the digital repository underlying the self-archiving toolkit developed in this project. Fedora offers many useful features, such as a digital object model that allows storage, within an object, of metadata annotations using different schemas, or storage of data in differing formats and semantic-ready data in the form of internal [RDF/XML](#) data. Additionally, it is closely coupled with the [Mulgara](#) [RDF](#) database. This approach enhances the management of semantic-ready metadata and data stored in Fedora by aggregating them in the triplestore instance.
- *Mulgara RDF Database*. This database implements many of the [W3C](#) Semantic Web concepts⁵. [Mulgara](#) is designed to hold metadata in the form of subject-object-predicate statements. Its integration with a Fedora repository facilitates access to data and metadata from the digital repository, which can then be combined with data from other sources. Query interfaces can be used to populate web interfaces, or lightweight visualisation frameworks like 'Exhibit' from the [MIT SIMILE](#) Project. It is possible to pass complex queries to [Mulgara](#) using a number of query languages such as [SPARQL](#), a query language similar to the ones used to query relational databases, or [Tucana Query Language \(iTQL\)](#) which is [Mulgara's](#) proprietary query language. The results of these queries can then be used to drive web applications, be presented using the Exhibit framework, or be exposed as web services since it provides a [REpresentational State Transfer \(REST\)](#)⁶ web interface.

The toolkit developed in this project incorporates three different applications, which communicate both with Fedora repository and [Mulgara](#) triplestore, and allow the organisation, documentation, storage and visualisation of qualitative collections. The first of these, 'Spreadsheet builder', is a web application underpinned by a set of [Java](#) libraries to assist users with the organisation and documentation of their data. It presents users with a stage-by-stage process, beginning with the selection of metadata terms that are necessary to describe their data. This process generates a set of spreadsheets that can then be completed to express both resources and collection structures while avoiding the tendency for individuals to improvise metadata fields, or

to use terms from established standards in inconsistent ways. The second application, ‘QuDex Collection Manager’, is also underpinned by a set of Java libraries and its web interface makes use of the Spring Model View Controller (MVC) framework⁷, and its Web Flow extension, to implement the ‘flows’ of the web application, which are a sequence of steps that guide the user to perform a series of operations in the digital repository. This application is concerned with the creation and storage of the documented datasets in the digital repository. Again, it presents users with a stage-by-stage process that facilitates, firstly, creating the collection structure in the repository, and secondly, integrating the documented files and their relationships within the collection structure. Lastly, the third application, ‘QuDex collection explorer’, is a web application to visualise the qualitative collections once they have been stored in the digital repository, and in the Mulgara RDF database. The collections, their associated data and the relationships between them are visualised using the SIMILE’s Exhibit visualisation framework, by performing a set of queries in the RDF database. As opposed to the first two applications, the collection explorer accesses the collections’ information from the RDF database rather than from the digital repository. The RDF database includes an RDF graph that mirrors the semantic information about the collections stored in the digital repository, so that it can be accessed via SPARQL queries. The inclusion of the contents of the collections in the RDF database allows this information to be made available, not only to the visualisation applications developed in this toolkit, but also to a wide variety of applications such as other archives, linked data portals, and personal websites.

4.2 Documenting Qualitative Data Collections

While archiving technology developments, and information technologies more generally, are fairly unproblematic, the implementation of research archives holding qualitative data with data sharing and re-use purposes presents numerous, sometimes insurmountable, issues from ethical and research methodological perspectives. As introduced in Chapter 2, many researchers are still concerned with these issues and are reluctant to undertake qualitative research data archiving. The nature of the data, the role of the researcher in the research process, the context in which the research took place, along with the epistemologies of qualitative research more generally, are the most common arguments against archiving. However, current technologies provide solutions to some of these issues. In this chapter the author focuses on those for which the proposed set of tools attempts to provide a solution.

In relation to confidentiality and data ownership, much research data about people, even sensitive data, can be shared ethically and legally if researchers employ strategies of informed consent, anonymisation and data access control. Sensitive and confidential data can be safeguarded effectively by controlling access to data, or use of them. As well as securely storing data, such safeguards enable data to be shared with others for research and educational purposes. In this respect, the implemented toolkit, more specifically the collection manager, makes use of the mechanisms provided by Fedora security architecture. A major feature of this architecture is the integration of the eXtensible Access Control Markup Language (XACML) and an XACML-based

policy enforcement module. Developed by the Organization for the Advancement of Structured Information Standards (OASIS) Consortium ⁸, XACML is an XML-based mark-up language to encode access-control policies. It enables the specification of fine-grained, machine-readable policies that can be applied to control access to the repository in different levels. These levels range from the whole repository contents, or specific collections, to a single object, or even specific data within a digital object. For example, if we had a digital object in the repository that holds the audio representation of an interview accompanied by descriptive metadata, the repository could provide external access to the metadata only, while the raw audio data could remain private.

The preservation of contextual information was discussed in Chapter 2. It is one of the epistemological challenges of qualitative data archiving. One of the different types of context that is relevant to data archiving is the context of the original study, e.g. political or policy context of the study, methodological context, or research approaches taken. Another type of context that is equally important is the context of the data collection itself. That is, describing aspects related to the conversation or interaction, between researcher and participant, the situation (the research setting), or external contextual elements that are relevant to the particular event taking place, like for example institutional or cultural elements. From a data description perspective, the documentation of these different aspects could be addressed by defining different levels of data, and their associated contextual information. The general context of the study can be documented by including information such as the research proposal, details about the funding body for a particular research study, and a description of the research design and the methods used. The role of descriptive standards is very important in this respect, since they provide an agreed and standardised set of metadata terms that supports the documentation of this type of information. For example, the DDI codebook supports study-level documentation by including terms to describe aspects such as the methodological approach of the study, geographical and temporal coverage, sampling, and data collection methods used. Additionally, other external sources such as related publications could also be included and linked to the study resources to which they are relevant. With respect to the documentation of data collection, contextual information related to a particular event, e.g. interview, focus group, observation, can be provided by including descriptive metadata elements, such as the method(s) for data collection, data collection procedures, researcher's notes or memos, or descriptive information about the participants. More specific information concerning the interactions and events that occurred during a specific data collection situation could even be added in the form of electronic mark-up, adding more value to the qualitative data. One of the documentation formats to enrich digital textual data, by adding electronic mark-up, is TEI. An interview transcript in XML format could be marked-up in TEI to include tags:

- for interview description (e.g. location of the interview, information about the participants),
- for the documentation of events that occurred during the interview (pauses, interrup-

tions), or

- for including the original *verbatim* in those cases where corrections of typos have been made in the transcript.

The provision of this enriched mark-up offers the possibility of generating more detailed documentation of the data and, when combined with other technologies such as search engines and visualisation frameworks, enhances not only the way data is displayed (by enabling multiple representations of the same data) but it also provides more accurate systematic searching, retrieval and browsing of textual data. Additionally, more contextual information for the data collection materials could be provided if analytical documentation such as the codebooks, or the classification frameworks that researchers used during data analysis accompanied the archived data. The inclusion of such information, in combination with annotation of the relationships and dependencies within the collected data, can help re-users of the qualitative materials and facilitate working across different collections. However, the current ways of capturing this type of information, which make use of software tools or packages, are very heterogeneous, thereby it is difficult to share, or re-use, research data.

The provision of very rich documentation data, and all the processes involved, lead to a complex archiving process that often requires significant efforts on the part of the researchers. Researchers and users depositing their data have to perform time-consuming tasks, and learn new procedures that involve using new technologies. The desire for a less heavyweight solution has been one of the motivations for the development of this self-archiving toolkit which, while presenting robust access control mechanisms and advanced services to manage data collections, also provides researchers with a basic framework to create and store small-scale research data collections. The main goal is to provide mechanisms to create data collections in simple ways and by following processes with which students and researchers are familiar. Users can archive small-scale qualitative data collections by simply organising their data and creating data spreadsheets that contain resource descriptions annotated using appropriate metadata terms. The data model underpinning the implemented toolkit is described in detail in the next section.

4.3 Qualitative Collections - QuDEX-based Data Model

Participation in the DDI Qualitative Exchange working group has allowed an in-depth exploration of a wide variety of use case scenarios for non-numeric, multimedia data in the social sciences. It has provided insights about the kinds of data that researchers produce and use, the methods of their choice, and the types of software and tools that they use to work with their data, across different disciplines within the social sciences. One example is sociology, in which the most common collected data is audio-visual data from interviews, observations or focus groups that are transcribed into text. CAQDAS packages are often used to perform analytical tasks such as “code and retrieve”, which can also integrate text with other forms of data such as images and video. CAQDAS packages also allow for discourse analysis, in which researchers

collect interviews or record spoken data and make use of text mining techniques, such as text categorisation, or concept/entity extraction. To support those analytical tasks *CAQDAS* packages also work with text encoded or marked-up using a variety of *XML* schemas for analysis. While different disciplines collect and work with different kinds of data, what they have in common is that most of them make use of data represented in textual form. Therefore, the core data that is handled by most of the current qualitative data archives is textual, although support for audio-visual and multimedia materials is also provided. The contents of a typical archived qualitative collection are mostly interviews, diaries, writings and transcriptions of audio, or audio-visual materials that are accompanied by a variety of documentary data, including user guides prepared by the original researchers to assist re-users exploring the collection, the research proposal and methodological summaries, and topic guides. With regards to the presentation of these collections via the web archive interfaces, the user has access to the study documentation for which the source is a *DDI XML* file that is processed and presented visually, and typically includes information such as:

- *Study description*. A record containing information about the depositors, principal investigator(s), data access conditions.
- *Abstract*. Short description of the study and keywords or related topics.
- *Documentation data-lists*. Summary of all the data files included in the collection.
- *Citation information and related materials*.

The documentation information is usually available for download in multiple formats (e.g. pdf, excel spreadsheets) and, depending on the access conditions for the collection, a bundle containing the datasets is usually also available for download. These data catalogues offer the user advanced search interfaces based on the documentation of the different collections, and easy ways of exploring the data available. Moreover, the catalogue contents are highly discoverable owing to the fact that the collections are described using data documentation standards, such as *DDI*. The data models underpinning these representations of qualitative data collections provide very detailed study-level information and to some extent, they also provide other kinds of metadata information at data level, such as title, description, file format, number of pages or duration. However, these models do not include mechanisms for expressing the different relationships between the qualitative materials, nor do they handle the inclusion of data collection contextual information and the direct association of that contextual information with the data files themselves. From the user perspective, the visualisation of these relationships, in addition to basic analytical information such as the themes associated with a particular interview transcript, provides richer contextual information and enhances the exploration of qualitative collections.

In this respect, the data model of the self-archiving toolkit developed in this study includes support, firstly, for managing basic analytical units such as categories and codes, which can be

attached to any data file included in the collection. Secondly, it uses a set of relationships for linking the materials in the collection, while at the same time, maintaining descriptive documentation at the study level, in a similar fashion to data archives. The implemented data model is based on a combination of the Fedora object model to represent the types of objects within a collection, and the *QuDEx* schema (introduced in Chapter 2) to represent the above mentioned relationships. The model includes three basic types of objects, which are generic enough to represent both the structure of the collection and the types of qualitative materials that can be included (see Figure 4.1).

The digital repository object model is designed based on a “compound digital object” that allows the aggregation of one or more content items into the same digital object. These items can be of any format and can either be stored locally in the repository, or stored externally and then referenced by the digital object. This design principle has facilitated the design of the data model, which required incorporation of a combination of data documentation content items, using a variety of metadata standards, along with their associated qualitative materials. The data model is composed of three different types of digital objects, which are linked to each other by using semantic relationships that are defined in an ontology expressed in *RDF* format: the *Fedora* relationships ontology⁹. While the relationships expressed in this ontology are used internally as a mechanism for linking the digital objects, they are generic and conform to linked data standards. This has the advantage of enabling re-usability and dissemination of the digital objects, and their relationships.

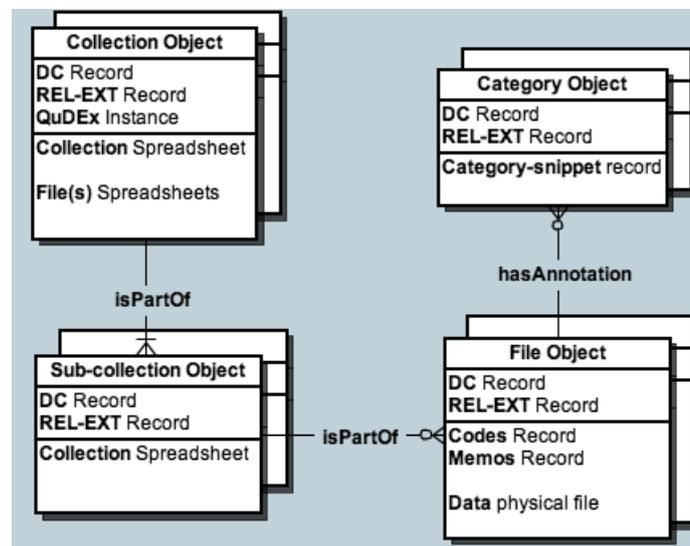


Figure 4.1: Data model: object types and their relationships

The first of these objects, ‘collection object’, is the core object of the model and it represents the qualitative collection as a whole (the study). Following the principles of Fedora’s compound digital object model, it incorporates the following content items:

- *Dublin Core record*. This is a documentation content item, which contains a metadata

record describing a qualitative item. This metadata record contains those terms from the Dublin Core standard that are relevant to a description of the study represented in a particular collection. The contents of this record are indexed automatically, in the form of triples, into the [RDF](#) database (Mulgara) so that they are accessible to the collection visualisation tools.

- *REL-EXT record*. This is a special documentation record that contains the relationship information, that is, it expresses all the objects with which this collection is related, and the type of relationship. It also holds the metadata terms associated with the collection, which are included in metadata standards other than Dublin Core. The contents of this record are also indexed into the [RDF](#) database.
- *QuDEx instance record*. This content item is particularly important since it contains a representation of the whole collection, including all the files, analytical elements and relationships, expressed in a [QuDEx XML](#) instance. This is highly important in terms of data sharing and re-usability, since the description of the files included in the collection, along with their relationships and attached analytical elements, could be imported into other applications that support the [QuDEx](#) schema. A more detailed description of the toolkit's support of [QuDEx](#) is included in [Appendix A.5](#).
- *"Collection spreadsheet" content item*. This is a content item that contains the original spreadsheet used to generate a collection in the digital repository. This is useful in that it provides a *backup* mechanism.
- *A number of "File spreadsheet" content items*. Similarly to the collection spreadsheet content item, the collection object contains a list including all the file spreadsheets used to aggregate the qualitative materials to the collection.

The second of the objects included in the model is the 'sub-collection object'. This object is quite similar to the collection object in that it holds documentation and relationship records and a collection spreadsheet content item. However, it is used with collection organisation purposes so that the qualitative materials can be grouped by different criteria and then attached to the sub-collection object. The third type of object is the file object, which is used to represent and store any type of qualitative material. It includes similar documentation records to the ones for a collection or sub-collection object, although the metadata terms used to describe this type of object differ (for more detailed information about the metadata vocabularies used to describe collections and files, see [Table 4.1](#) overpage). Additionally, it can have analytical documentation attached to it. This information is included in special content item records, codes and memos, which are represented as [XML](#) snippets expressing what codes and memos are associated with a particular file object, and more importantly, listing all the files that are also associated with the given code/memo. These codes and memos are analytical annotations of a given file, similar to the ones used within [CAQDAS](#) packages. Lastly, the fourth type of object is the 'category object', which is an analytical element that can be attached to file objects. The important records associated with this type of object are the RELS-EXT documentation record, and the 'category-snippet'

content item. The first of these includes important relational information which describes all the files within the collection to which a given category object is attached and it also describes the source(s) for a given category. Analytical categories, like the ones used in CAQDAS packages, are normally hierarchical, therefore relationships defining the provenance are needed. The second is a content item record that holds the information about a given category object as a QuDEX XML snippet.

Table 4.1: Documentation of “Collection” and “File” objects

Type	Description	Schema	Metadata Terms
Collection	Two different types of documentation are provided for collection and sub-collection objects, which use a combination of metadata terms from multiple schemas. The first type of documentation is related to the methodology of the study, and uses a reduced set from DDI ^a schema (version 2.1); and geographical information, from WGS84 lat/long ^b vocabulary. The second type of documentation is general description of the collection using mostly Dublin Core terms.	DC	description; language; coverage; rights; contributor; type; creator; identifier; subject; date; publisher and title
		QuDEX	updatePid and deletePid (used for deleting and updating collections)
		DDI	samplingProcedure; universal; timeMethod; dataCollector and collection-Mode
		RDF	type (collection or file)
File	Two different types of documentation are provided for file objects. The first one is a more general description of the qualitative material physical file, including information such as file format, date of creation, author and title whereas the second type is analytical information attached to the file such as memos or annotations, codes within the file or categories.	GEO	latLong and location
		DC	description; language; subject; format; title; coverage; type; creator and source
		QuDEX	labelCode; labelCategory; category; labelMemo; memo; isOriginal and code
		RDF	type (collection or file)
		GEO	latLong and location

^a <http://www.ddialliance.org/specification/ddi2.1/lite/index.html>

^b <http://www.w3.org/2003/01/geo/>

4.4 Archiving Qualitative Collections

The developed toolkit provides mechanisms to create research collections in simple ways and by following processes with which students and other researchers who are not archivists or database designers are familiar. The empirical data gathered from interviews conducted with researchers with varying levels of experience allowed for an exploration of a range of research practices from primary analysis, archiving and re-use of qualitative data. This facilitated the identi-

fication of key requirements for the archiving model that underpins the self-archiving toolkit so that the initial design, which was mostly informed by existing practices of research data archives, could be adapted and refined to support archiving practices of researchers. Such a combined design approach is key to developing an accessible set of archiving tools that lowers the bar to its adoption by less experienced users who are less familiar with archiving activities, but that aligns well with more general archiving practices such as those of archivists or data curators. Table 4.2 compares the different features of two models of archiving: the more traditional model associated with research archives/archivists, and the toolkit's archiving model which combines aspects of the previous model with the requirements gathered from analysis of research practices.

Table 4.2: Comparison of Models of Archiving

Archivists/Data Archives	Self-archiving Toolkit
The creation of an archive is a one-time operation, in which final output(s) from a research project and accompanying primary data collected and/or generated are stored in a digital repository or archiving system.	The creation of a 'final archive', that is, an archive that is highly unlikely to be modified is supported by the toolkit's archiving model. This type of archive will often include final outputs from a research project and carefully selected primary supporting materials.
Common to both models	
<ul style="list-style-type: none"> a Research design, data management plan and consent plan for archiving and sharing are defined at the beginning of a project. b Data collection, documentation and organisation can happen throughout the research process. 	
Specific to each model	
<ul style="list-style-type: none"> c Data preservation/archive creation (digitisation, metadata creation and data storage) is performed once the investigation is complete. 	<ul style="list-style-type: none"> c The toolkit model of archiving supports the creation of 'pilot collections' at any stage of the research. These collections can be visualised through the toolkit's visualisation tools to support researchers' analysis and interpretation activities during their research. d Creation of 'work-in-progress' collections that can be referenced from research instruments such as research diaries, or project reports. Archiving activities to extend/modify these collections can be performed at any stage of the research project.

In short, the model of archiving underpinning the self-archiving toolkit supports both practices of archiving: those of archivists or research data centres and those of researchers. The exploration of two empirical cases that represent complex digital archives (the third cohort of participants as introduced in Chapter 3) provided examples of archives for which the accurate documentation of a research study, and the provision of sufficient contextual information are

central, especially when the purposes of the archived materials is to support data re-use and/or further exploration. Additionally, the design and creation of the archives in these two cases was performed incrementally, in that new, interpreted data was added to the archives over multiple iterations. In this respect, the toolkit's model of archiving was extended so as to provide support for incremental processes of archiving (e.g. pilot collections described in Table 4.2). The ability to create small research collections alongside data analysis activities can be useful, firstly to explore the kinds of data that can be displayed together and how the different resources relate to each other, and secondly, to help researchers identify appropriate ways of documenting their data. Figure 4.2 shows another example from one of the student empirical cases. To support the development of their research project the student created a 'work-in-progress' archive, which was organised according to the different sections of the dissertation (the output of the project) and the types of materials associated with each section.

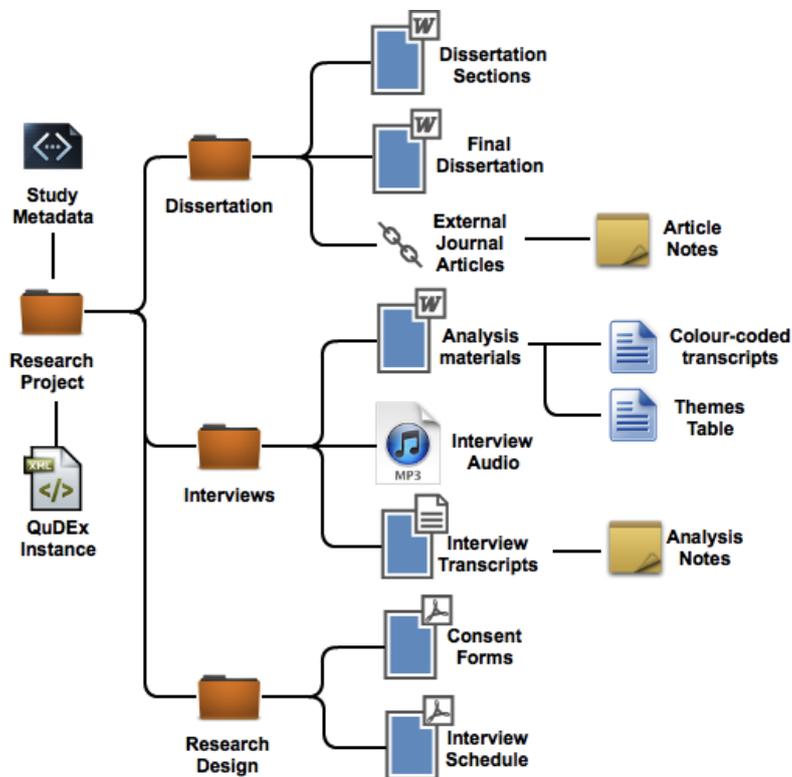


Figure 4.2: Example of a 'work-in-progress' archive organisation

A more detailed account of how the exploration of the different empirical cases informed, or aligned, with the design and implementation of the self-archiving toolkit is provided in Chapter 5. Figure 4.3 shows the process of organising, documenting and creating an archive with the support of the toolkit. A 'collection' can be seen as a set of resources associated with a project, dissertation or article, data are primarily collected materials (survey data, interview audio files, transcripts or notes) and analytical outputs may be numerical summaries, generated charts or tables, and completed pieces of writing. These are collated using spreadsheets either in Excel or CSV formats, and once stored into the archive, they are organised in a manner that is very similar to a 'files and folders' structure on a desktop computer.

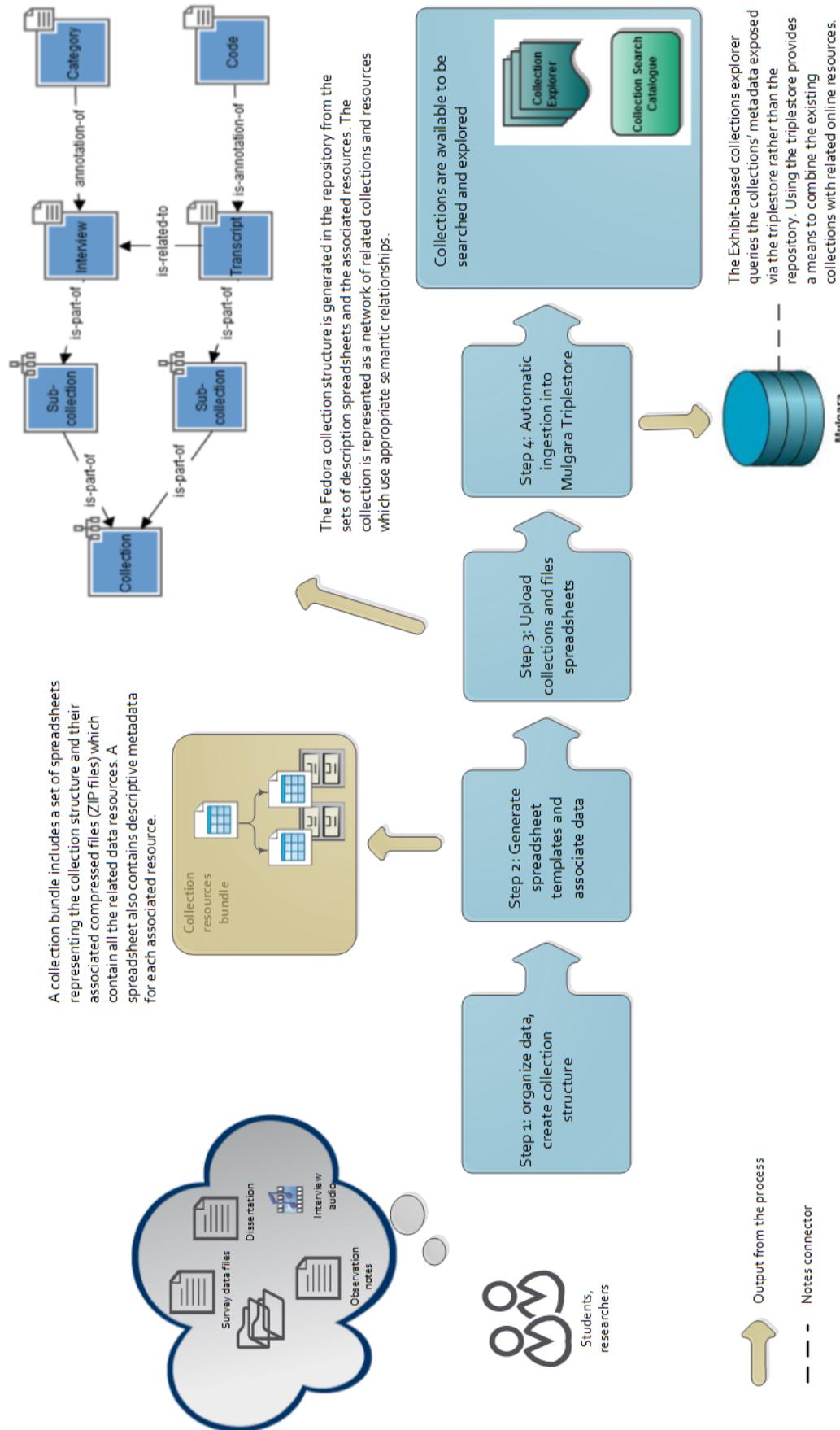


Figure 4.3: Framework for data organisation, collection generation and publishing

The toolkit guides the users (students, researchers) through the process of structuring and describing their data using a series of web applications (steps 1, 2 and 3 in Figure 4.3) without their having to deal directly with the internal working of the underlying systems (Fedora or Mulgara), or to understand too much about the metadata schemas that are listed in Table 4.1. The first steps in the collection creation process are concerned with describing and organising the data (step 1 in Figure 4.3) in appropriate ways, to ensure that the data can be shared and integrated with existing online content. As it has been previously highlighted, standardised data documentation is key to digital archiving and has an impact on dissemination, data discovery and effective re-use. To ensure that a research study and all its associated materials are described in consistent ways, the first stage of the archiving process, data organisation and documentation, is supported by the template builder application (see Figure 4.4 below). This application is a spreadsheet-based tool for capturing metadata about qualitative and mixed-methods studies, providing an easy-to-use mechanism for describing all the materials included in a particular project or collection (step 2 in Figure 4.3).

QuDex Repository Tools - Template Builder

Home | Tutorial | Further Information

Generate Spreadsheet

The spreadsheet will contain the following header types:

- Selected metadata elements from the schemas

Review the included metadata elements

Review the selected metadata elements that will be added to the Collection-level spreadsheet to be generated.

Review Metadata Schemas

Re-start the process

Refresh the spreadsheet headers and start the process again

Start new spreadsheet

Spreadsheet Headers Overview

Metadata Elements selected

You have selected elements from the following schemas. If you want to go back and change those just click the button below.

Schema Prefix	Schema Name	Optional Elements	Compulsory Elements
dc	Dublin Core Vocabulary	description, language, identifier, subject, coverage, rights, contributor, type, creator, date, publisher	title
rdf	RDF Vocabulary	---	type
geo	Basic Geo (WGS84 lat/long) Voc.	lat_long, location	---
qudex	QuDex Vocabulary	deletePid, updatePid	---
rel	Fedora relationships ontology	isPartOf	---
ddi2	DDI2 Vocabulary	samplingProcedure, universe, timeMethod, dataCollector, collectionMode	---
skos	Skos Vocabulary	note	---

Auto fill-in DDI2 Study level metadata (optional)

If the DDI metadata are included in the spreadsheet, you can upload a DDI2 instance so the spreadsheet can be auto filled-in. The application will parse the DDI2 file looking for the DDI fields and will include their values in the generated template.

Upload DDI2 XML Instance (optional)

No file chosen * Upload one XML file at a time

Add DDI Metadata

When finished adding DDI Codebook files, click "Generate Spreadsheet" button.

Generate Spreadsheet

Figure 4.4: Spreadsheet template builder: generates spreadsheet templates with appropriate metadata fields

Once the descriptive spreadsheets have been created and populated with descriptions of the data, research instruments, analytical tools and outcomes of interpretation of analysis, the user can then generate and archive the collection in the repository (step 3 in Figure 4.3). This involves using another web application, 'QuDex collection manager' (see Figure 4.5), which allow users to, firstly, generate the collection structure and secondly, upload all the data resources with which the collection is associated. The process of uploading the spreadsheets in the repository transforms each described element into the data structures included in the model. These structures are shown in Figure 4.6. The 'collection structure' spreadsheets are used to generate collection and sub-collection elements representing the organisation of the research collection

QuDex Repository Tools - Collection Manager

Collections/Sub-collections Spreadsheet upload

Description of Collection-level Spreadsheet

Notes

* If a collection is selected, the relationship information in the spreadsheet (field `rel:isPartOf`) will be ignored. In other words, all the collections described in the spreadsheet will be subcollections of selected Collection.

You can skip the process of adding collection/subcollections to the selected collection and go directly to "Upload Files Spreadsheet"

Upload Files Spreadsheet

Collection-level Spreadsheet

Selected collection navigation

The selected collection presents the following structure in the repository:

- WPPS Case Study (Repository Identifier: qudex:1)
 - About (Repository Identifier: qudex:2)
 - Other resources (Repository Identifier: qudex:3)
 - PL (Repository Identifier: qudex:4)
 - PP (Repository Identifier: qudex:5)
 - Assessment (Repository Identifier: qudex:6)
 - OS (Repository Identifier: qudex:7)
 - PE (Repository Identifier: qudex:8)
 - ET (Repository Identifier: qudex:9)
 - Community (Repository Identifier: qudex:10)
 - CR (Repository Identifier: qudex:11)

Upload a Collection/Sub-collection Spreadsheet. This will create the QuDEX collection/sub-collection objects in Fedora Repository.

Options:

1. Upload a Collection-level spreadsheet

Use this option is uploading a top-level collections spreadsheet which does not contain relationship information.

No file chosen

Figure 4.5: Resource bundles upload: creates the collection structure in the repository and stores the associated data resources

and the 'resources' spreadsheets are used to generate files objects representing the different instruments produced during the research process.

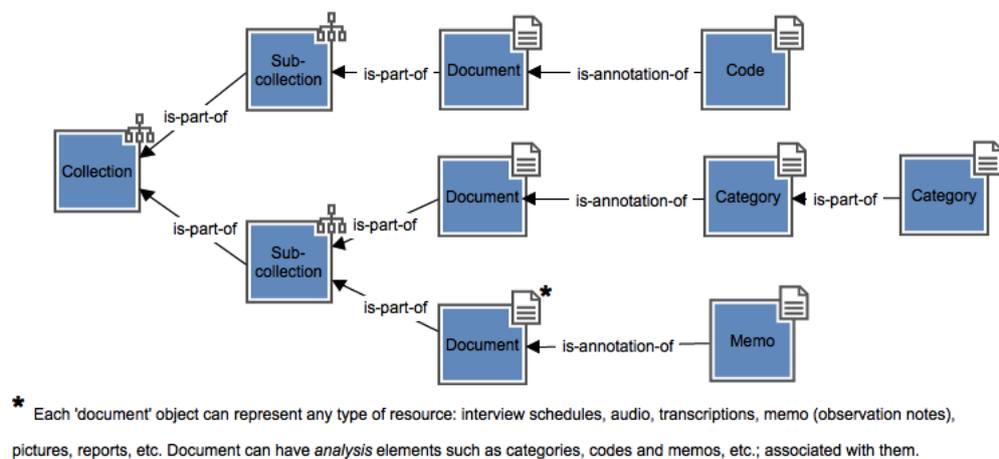


Figure 4.6: Data Model used to represent data collections in Fedora Repository based on the QuDex Model

4.4.1 Toolkit Visualisation Interfaces

The visualisation interface included in the implemented toolkit is a front-end to the digital repository and the associated triplestore. Once the collections are archived, they can be visualised through one interface: 'QuDex Collection Explorer'. It relies on the Exhibit visualisation framework to implement a dedicated, faceted browser that allows an exploration of an individual digital collection (see Figure 4.7). This application implements a set of Java servlets that access the repository contents from the triplestore instance. The triplestore mirrors the contents

of the repository (collection metadata and relationships information) and exposes them as [RDF](#). The interface performs a set of [SPARQL](#) queries over the triplestore and the results are then processed by the Exhibit interface to generate the collection visualisation. More importantly, one of the advantages of mirroring the contents of the repository in the triplestore is that the data associated with a particular collection could also integrate relevant external information that is not included in the repository. Examples of such additional resources include citation information drawn from publishers' websites, or 'open data' from elsewhere on the web, which can be related to themes explored in the archived collection.

QuDEX Collection Explorer

Search Facet

Filters

Collection Structure
(others) (2)

Keywords (Subject)

- 2 collaborative design
- 1 collaborative learning
- 2 collaborative work
- 1 Common Wealth of Australia
- 1 communication
- 1 communication technologies

Category Hierarchy

- project (2)
 - educational (1)
 - physical (1)

Type of Element

- 2 File

File Format

- 2 application/x-shockwave-flash
- 2 text/rtf

Collection elements

- 1 Transcripts (ET)
- 1 Transcripts (PE)
- 2 WPPS Case Study

Save Track

Searching the Archive

This search tool allows you to search across the elements of a particular collection: individual documents by using a free facets. The linking of the facets means that as you make selections, the available options in other facets will change, a you get. 'External Links' open in a new browser window or tab; other documents will download or open according to you

Search Results

Data has been loaded with the following filters set: collaborative design

2 File filtered from 385 originally ([Reset All Filters](#))

sorted by: [date](#); [then by...](#) • grouped as sorted

Interview with Ray Trotter about 'what works' (transcript) (2010)
(related to 'Transcripts (PE) and WPPS Case Study' Collection) [View Graph Demo](#)
No Description Available
Total related categories = 2
[Collapse](#)

This item is format of 'Interview with Ray Trotter about 'what works' (video)'.

Keywords: architectural design, school environment, project development, co-operative learning
Coverage: North Dandenong, Melbourne, Australia
DC Type: Document
Date: 2010
Latitude/longitude: -37.95974,145.21705
Location: North Dandenong, Melbourne, Australia

Interview with Mary Featherston about curiosity (transcript) (2010)
(related to 'Transcripts (ET) and WPPS Case Study' Collection) [View Graph Demo](#)
No Description Available
Total related categories = 2
[Collapse](#)

This item is format of 'Interview with Mary Featherston about curiosity (video)'.

Keywords: collaborative design, industrial revolution, school spaces, and school classroom
Coverage: North Dandenong, Melbourne, Australia
DC Type: Document
Date: 2010
Latitude/longitude: -37.95974,145.21705
Location: North Dandenong, Melbourne, Australia

Figure 4.7: Faceted collection exploration: the user can explore the contents of a given collection and how the different resources relate to each other

The collection explorer presents the users with a 'faceted' (property filters) navigation of the collection, by applying a set of filters. Faceted search enables users to explore a multi-dimensional information space, the collection structure, associated resources and analytical information such as codes, categories or notes/memos, by combining text search with a progressive narrowing of selections in each dimension. Each facet corresponds to a property or set of properties of the information elements included in the collection, which allows a filtering of the information by applying either an individual filter, or a combination of them. This is similar to 'AND' searches over a catalogue or database. Given the nature of the information included in a collection, the most relevant facets (filters) that have been selected are the following:

- *Collection Structure*. This filter allows for an exploration of the collection based on its or-

ganisation.

- *Keywords (Subject)*. This filter lists all the topics that have been used to describe the materials included in the collection.
- *Category Hierarchy*. This filter allows users to search across the collection based on the category framework used to annotate the resources of the collection.
- *Resource Type*. This filter allows users to visualise the resources of specific types (objects of the data model), e.g. File, Collection, or Category.
- *File Format*. This filter allows users to visualise the resources of specific formats, e.g. audio files, video files, Portable Document Format (PDF).
- *'Collection elements'*. This filters allows users to explore the files included in the archive by collection/sub-collection provenance.

The way in which the user queries the data, by selecting the different filters, produces a reduced set of results that can be very useful when one is working with large datasets. For example, one could visualise only the transcriptions associated with interviews that have been coded according to a selected set of topics.

Another functionality of the collection explorer is its ability to keep track of the user navigation, that is, the application can generate 'snapshots' of the search results after the user has applied some filters to the dataset. This is performed by simply generating a URL which contains a set of parameters to identify what filters have been applied. The latter allows the user to save the results of a set of queries against the dataset, which are then permanently accessible by simply using this URL in the web browser. This functionality could be quite useful to support the presentation of results, for example in 'enhanced publications' (as discussed in Chapter 2) in which a publication is accompanied by research data, extra materials or post-publication data. Student dissertations could be enriched if accompanied by the underlying data (evidence of the research), or online research diaries. Alternatively, research portfolios could reference specific sub-sets of the archived collection by simply including a reference (URL) to a particular pre-filtered search. As the web pages generated using 'collection explorer' are entirely Hypertext Transfer Mark-up Language (HTML) based responses to a set of predefined queries, all that users need to do is construct that query to present their whole uploaded collection, a subset of that collection, or individual items in web pages, virtual learning environments, electronic portfolios or online social environments. The aim is to make this process as simple and streamlined as the integration 'by URL' of a resource such as a Youtube video into a personal web page, but with the provenance and referential integrity of what is presented also being 'explorable' by other users.

Lastly, the collection explorer incorporates a supplementary visualisation interface: 'graph explorer'. This application presents a different approach to visualising the data from a particular collection. With the collection explorer, the user performs searches across an entire collection,

by following a filtering and data reduction approach. In contrast, the graph explorer allows a fine-grained, in-depth exploration of an individual sub-collection. The sub-collection, all the related files, and how they relate to that particular sub-collection are visualised in a directed graph (more concretely an **RDF** graph). An **RDF** graph is composed of a set of triples. Each triple is represented in the graph by two nodes (subject and object) that are linked (predicate) (W3C, 2004). Figure 4.8 shows an example of an **RDF** graph.

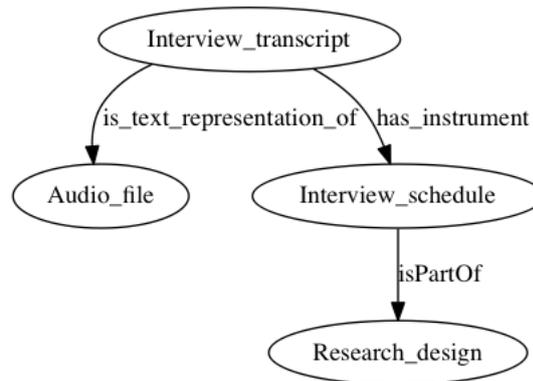


Figure 4.8: **RDF** graph example

This visualisation also shows any analytical annotations that have been attached to the visualised materials. Alternatively, when the user selects an individual file from the collection explorer visualisation, he is presented with all the files, either belonging to the same sub-collection, or across the entire collection, that are related to the previously selected file (see Figure 4.9). When selecting any element included in the graph, the user has then access to the displayed elements' metadata information and specific details, in a similar fashion to how the items' information was presented in the collection explorer. Such a visualisation approach is possible owing to the definition and modelling of a set of relationships, both from Fedora's relationship ontology (to represent the structure of a study) and from the QuDEx schema (to represent all the relationships between files and their associated analytical elements). A person exploring a particular instrument from the collection, for example a transcription of an interview, is presented with methodological information, such as the interview schedule used for that particular interview, a sub-set of the coded themes or notes associated with it, along with supplementary information such as policy documents or publications related to those identified themes.

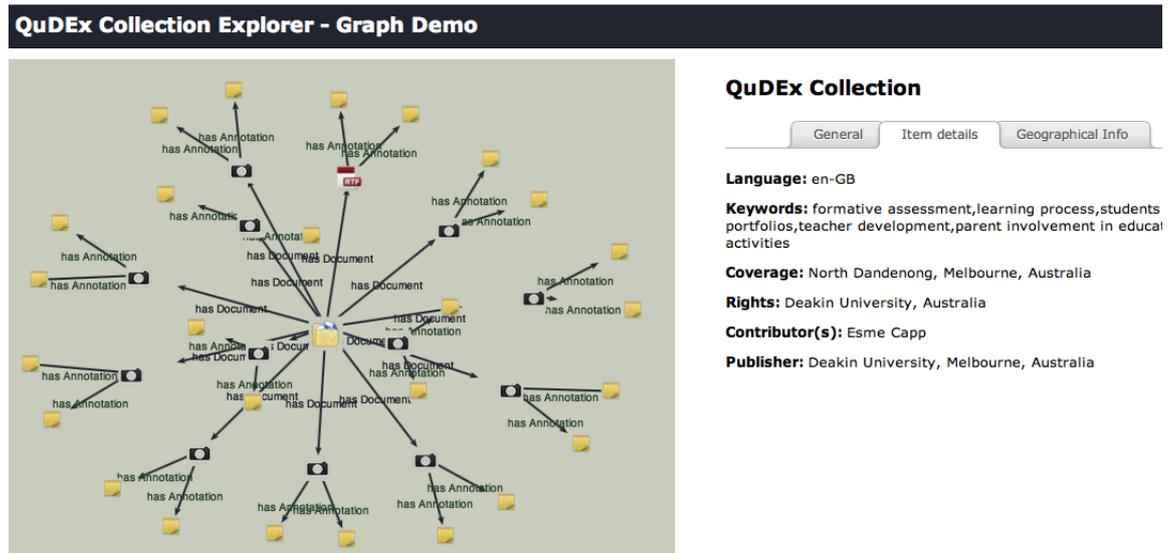


Figure 4.9: Graph visualisation: exploring in detail individual sub-collections or files, along with their related content and analytical information

4.5 Advanced Qualitative Archiving Models

The implementation of appropriate frameworks and tools for documenting and archiving larger research projects, where collaborative work is a requirement and multi-method research approaches are involved, becomes quite challenging owing to the technological requirements, and the underpinning data models, being much more complex. Such models and their associated tools have to reflect and incorporate current and standard archiving procedures, from the perspectives of a variety of stakeholders, e.g. researchers, archivists, information systems architects, and throughout the full research data/process lifecycle. There is an existing need to both define and implement research archiving models that provide a robust and complete support for the full research data management lifecycle with a focus on qualitative studies. In this respect, substantial work has been carried out by the [DDI Qualitative Exchange](#) working group, which has led to the implementation of an [XML](#)-based schema for the archiving and exchange of qualitative data, that integrates the documentation of research processes and activities, along with the data generated during those processes. This model is currently being reviewed by a number of stakeholders from the research and archiving communities.

As introduced in Chapter 2, the [QuDEX](#) schema is underpinned by a model that provides mechanisms to support the analysis stage of a research project, i.e. it provides rich mechanisms to store both the generated analytical data and the relationships between them, therefore facilitating the interchange of analytical data between systems or tools. The self-archiving toolkit developed in the present study provides a simple but flexible archiving model for qualitative data that facilitates the documentation and storage of research data, covering research design

and methodological aspects, such as study description or data collection methods used, and the documentation of qualitative materials at individual file level (e.g. for an interview file both analytical metadata and attributes of the audio file can be provided). Additionally, analytical data is also supported by modelling QuDEx schema analytical elements that can be exported and therefore potentially re-used within CAQDAS, in combination with data relationship information which enables richer visualisation and integration with external information sources by using linked data technologies. While both the QuDEx schema and the toolkit include capabilities for most of the processes and tasks involved in a typical research data lifecycle¹⁰ (see Figure 4.10 overpage), the design of the application logic and activities involved in those processes is one that focuses on the researchers' users' practices rather on supporting the documentation of the processes that take place when depositing data into a digital archive. Data processing activities, such as transcription templating and its processing, data anonymisation and formatting, or data processing events and the description of the tools used to process the data, are neither included in the toolkit's data model nor are they embedded in the process of self-archiving a qualitative collection. Despite this, some of those processes could be documented by using generic metadata terms and added to the collection in the form of textual documentation materials. Moreover, these limitations do not represent a critical issue since the purpose of the tools is to provide researchers and students with a basic framework for documenting and archiving their data without introducing a high level of complexity in those processes, while at the same time, allowing them to share and re-use their data in consistent ways so that data could still be integrated in larger-scale systems.

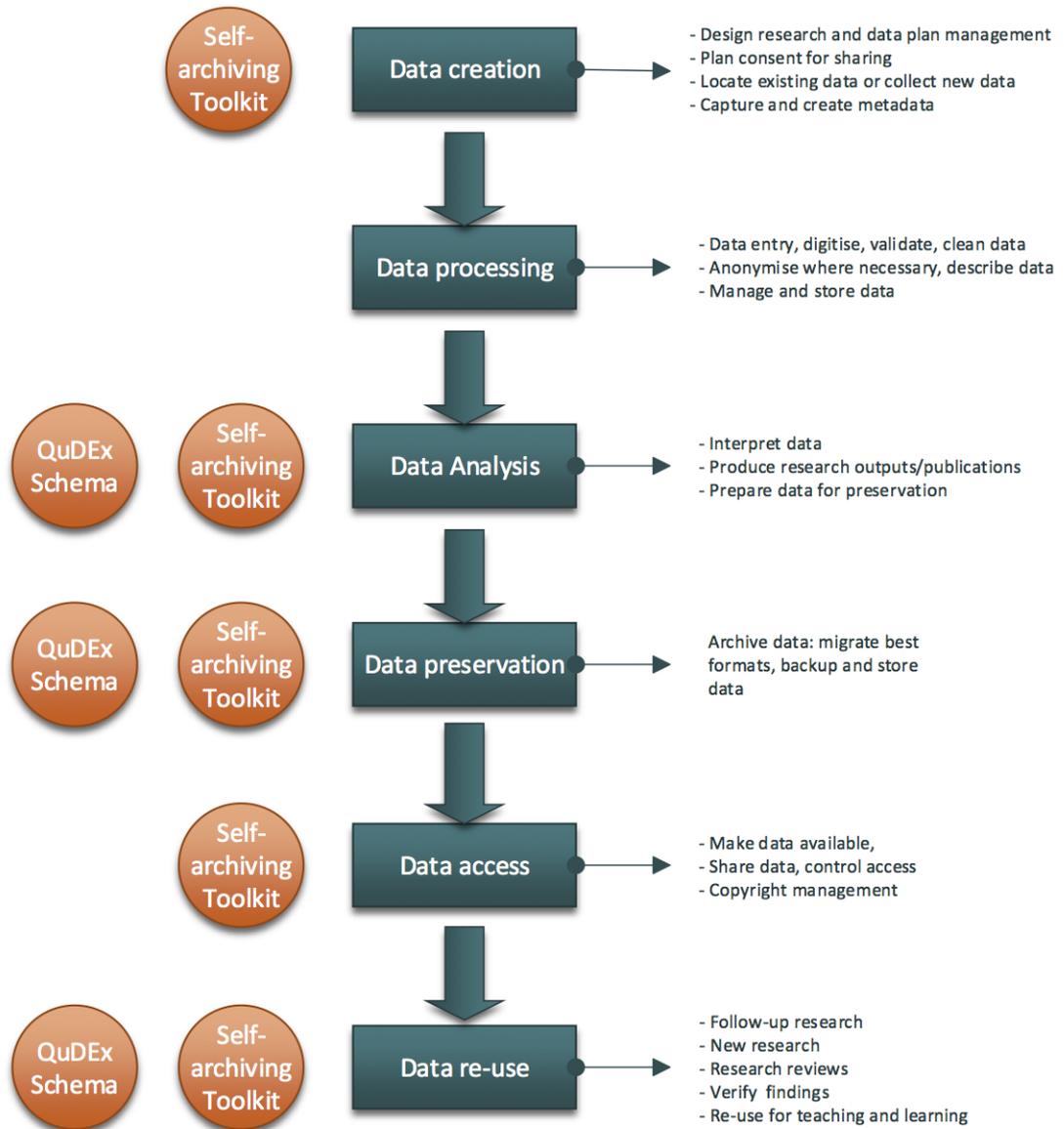


Figure 4.10: QuDEX schema and Toolkit support for a standard research data lifecycle

In contrast, the qualitative schema model proposed by the working group represents an enhanced combination of the previous two approaches, QuDEX schema and the toolkit's model designed in this study, while at the same time it provides full support for documenting data processing procedures in machine-readable ways. Firstly, it attempts to document the practices and processes involved in the course of qualitative studies by providing mechanisms to document in great detail individual events like 'an interview'. Examples of this type of information include description of the event (interview, focus group, ...), information about the participants which can be linked to related contextual information, and the data generated in those events. The latter could be a recorded audio file, metadata of the devices used to produce the audio file, associated transcription and how the transcription was produced. Figure 4.11 shows an abstraction of how a generic qualitative research project could be documented and stored using the DDI model.

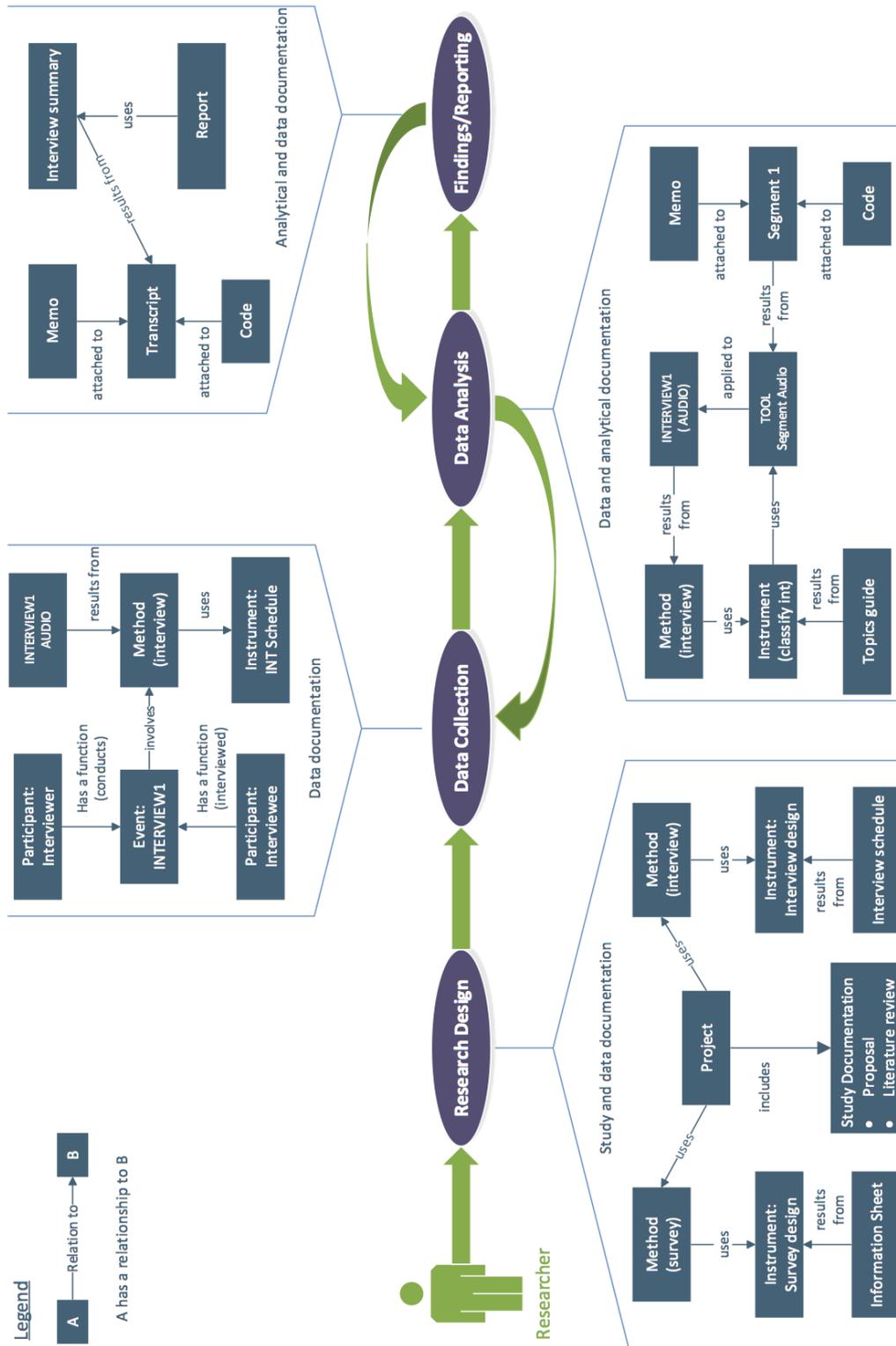


Figure 4.11: Qualitative Data Exchange Model

Each stage of the project presents an example of the kinds of materials and processes that could be documented. In the research design stage of a mixed methods research project 'A', the kinds of data that one could document (study documentation in Figure 4.11) and store include: study documentation data, such as the research proposal; information about the project's funding bodies or a link to the literature (publications, other research projects) underpinning the documentation about the research approach (e.g. Ethnography), and data collection methods that are employed (surveys, interviews). The data that is produced during data collection could be richly described by including detailed information about the process, e.g., one interview could be described in terms of an event (the interview) which involves entities participating - interviewer and interviewee. The event involves using one particular data collection method (audio-recorded interview) that is supported by using a particular research instrument (interview schedule) based on a data collection mode (semi-structured interview). The data collection method could even include more specific information about the data that is produced, for example, an audio file which results from using a tool (audio-recorder). Finally, the analysis process and the data that it generates could also be described in great detail (data and analytical documentation in Figure 4.11). An audio file 'interview 1' results from a data collection method (interview), which is an analysis instrument described in a 'topics guide'. The audio file is segmented by using a specific audio tool that produces different segments from the audio file that are then annotated by using codes (more specific information such as the coding scheme used could be provided), memos or notes.

As a result of this emphasis on modelling the research process, quite important discussions emerged around specific research methodologies and analytical approaches and what information should be included in the model in order to describe those. One of the approaches explored was Grounded Theory, in which the central analytical elements are codes and categories. The analysis process is iterative and needs to be described richly. However, the schema should be as generic as possible so that both research and analytical data can be re-used by as many information systems (CAQDAS tools, digital archives, VREs) as possible, and the model can serve as the basis for the development of open tools, underpinned by software-neutral and 'practices-neutral' approaches, which aim to support as many research approaches as possible.

What has been described in this chapter is a set of accessible self-archiving tools for qualitative research studies, which enable the documentation, storage and visualisation of collections in a variety of ways. The toolkit developed in this study provides users with a simple and structured approach for creating web archives of qualitative collections, which works well for small-scale research projects. In the course of research projects, especially those carried out by less experienced researchers or research students, there are important factors such as time-constraints, or lack of knowledge about research methodologies and the processes involved, which have an impact on the development of the projects. In order to support such situations, it has been a requirement to find a balance between providing users with a framework to support the documentation of research data, and research processes, in great levels of detail, and the degree

of complexity that such level of documentation could potentially add to the processes involved during the creation of a web archive. This complexity increases, when introducing the element of research archiving supported by new technologies with which users might not be familiar. On the other hand, more advanced research archiving models, and their associated tools, are necessary to support complex and large-scale research studies. In such cases, the archiving processes involved, along with the design and implementation of data management plans, requires researchers performing a set of tasks on top of those that are directly related with research. These extra tasks can be both new to researchers and time-consuming. Researchers' expertise lies in research, and there are arguments in favour of delegating archiving and data curation activities to archivists and data centres. Despite the latter, there is a concern for those small-scale projects, or individual researchers, which do not have the funds or the infrastructure to preserve and look after datasets beyond the lifespan of their research projects. In this respect, the introduction of lighter archiving processes, supported by digital repositories and a set of accessible online tools, can serve as the basis for future more complex archiving processes. This, at the same time, can serve to introduce researchers, and research students, to those more complex procedures that are required to prepare research data for sharing and re-use, by providing them with tools which support simple data organisation and management processes, and that can be embedded within their own practices.

Notes

¹<http://tomcat.apache.org/>

²<http://www.java.com/en/>

³<http://www.apache.org/licenses/LICENSE-2.0.html>

⁴<http://www.openarchives.org/pmh/>

⁵<http://www.w3.org/RDF/>, <http://www.w3c.org/2001/sw>

⁶The REST architectural style was developed by W3C Technical Architecture Group (TAG) in parallel with Hypertext Transfer Protocol (HTTP) version 1.1, based on the existing design of HTTP version 1.0. The World Wide Web (WWW) represents the largest implementation of a system conforming to the REST architectural style. REST-style architectures conventionally consist of clients and servers. Clients initiate requests to servers; servers process requests and return appropriate responses. Requests and responses are built around the transfer of representations of resources. For a complete description see Fielding (2000b).

⁷<http://www.springsource.org/>

⁸OASIS Consortium, <https://www.oasis-open.org/>

⁹<http://www.fedora.info/definitions/1/0/fedora-relsext-ontology.rdfs>

¹⁰Diagram adapted from UKDA's research data lifecycle. Source:

<http://data-archive.ac.uk/create-manage/life-cycle>

EXPLORING RESEARCH PRACTICES AND QUALITATIVE STUDIES ARCHIVING

This chapter describes the empirical work carried out in the course of this research and development study. The case study approach has provided, firstly detailed descriptions of research and archiving practices from researchers with varying levels of research experience to inform subsequent refinements, and further development, of the initial design of the self-archiving toolkit. The different empirical cases have contributed to the design of a toolkit which incorporates simpler archiving processes to create web archives of qualitative collections and that can be embedded within researchers' existing practices. Secondly, the exploration of researchers' epistemologies, their perceptions of data archiving and secondary analysis has helped to identify potential barriers and benefits of digital archiving, as well as those areas where the self-archiving toolkit has the potential to support research-oriented, teaching and learning activities. Interview data from each cohort of participants (as introduced in Chapter 3) is discussed in three different sections as follows. Section 5.1 discusses the students' perceptions of research, their practices and the processes followed during the development of their projects, with an emphasis on aspects such as their research interests, the importance of their background within the approaches chosen, and epistemology-related issues. The interviews conducted with students (first cohort, as introduced in Chapter 3) have provided important insights into their practices and perceptions of research while conducting their projects as well as a suitable setting to explore how the affordances of the implemented self-archiving toolkit might be put to use in undergraduate student research projects. Section 5.2 discusses the insights and findings from semi-structured interviews conducted with more experienced researchers that were also involved in teaching and learning activities with undergraduate students conducting research projects (second cohort of participants, as described in Chapter 3). This has facilitated the analysis of existing practices within research-based teaching and learning environments, with a focus on the sort of materials used by lecturers to support teaching activities; and to investigate their perceptions of the ben-

efits of using existing archive materials to enhance students' understanding of research projects and the research process in general. Additionally, the analysis of the participants' own research practices, with a focus on data organisation and archiving, has helped to identify additional requirements for toolkit's archiving model and to obtain examples against which the toolkit could be tested and optimised. Section 5.3 discusses the design and development processes followed by the researchers who were interviewed to produce two multimedia case study archives (third cohort of participants, as described in Chapter 3) and the implications of the analysis of those processes for the self-archiving toolkit developed in this study. Aspects, such as the selection of the materials to be archived, the archives' organisation approaches, the technologies used, and the key issues or difficulties faced by the two researchers during the production of their archives were of particular interest. These two empirical cases allowed for an exploration of different models of archiving, the key issues faced during the development of the research archives, as well as the uses and purposes of those archives. Lastly, section 5.4 provides a descriptive account of how the exploration of the different empirical cases informed/aligned with the design and implementation of the self-archiving toolkit developed in this study. The section summarises the data organisation practices of the students and researchers who were interviewed, the implications for the archiving processes implemented in the toolkit and discusses the self-archiving toolkit's support for the participants' practices.

5.1 Exploring Student Research Projects

As introduced in Chapter 3, the projects developed by the students who participated in this study were located in the Independent Project module. This module seeks to address knowledge-creation and knowledge-building, that is, showing students the relevance of understanding how knowledge is created, tested and questioned. At the same time, the module provides an environment in which students can begin to familiarise themselves with real research rather than just being taught about it. Four student projects, investigating a topic of their choice in school-based settings, were selected from over fifty examples (see summary in Table 5.1) on the basis that the projects the students proposed involved the collection and analysis of multiple types of data. Such projects have the potential to lend themselves to presentation as small archive collections for re-use or further exploration - either by the students themselves, or by a wider audience.

Table 5.1: Description of selected student projects

| Student | Project Summary | Research Approach | Research Methods | Data collected |
|------------------|--|-------------------|--|---|
| <i>Student 1</i> | 'Dyslexia: Approaches to supporting a child within a mainstream school at Key Stage 2'. This project investigated teachers' own understanding of what dyslexia is and the approaches they had taken to support children with dyslexia in a mainstream school. | Mixed-methods | Questionnaires and semi-structured interviews. | <ul style="list-style-type: none"> • Consent forms • Questionnaires (hard copy) • Interview files (audio-recorded) |
| <i>Student 2</i> | 'A study to examine the extent to which the National Curriculum and an individual's personal appearance have an effect on female participation in PE'. This project explored year 9 students' perceptions about physical education and the reasons behind the decrease of enrolments in this subject in a mainstream school. | Single case study | Focus groups conducted with year 9 students, with a sample of 12 participants. | <ul style="list-style-type: none"> • Consent forms • Focus group files (audio-recorded) |
| <i>Student 3</i> | 'An investigation comparing the perceived importance of physical education between two specialist sports colleges'. Exploring the various perceived roles of physical education and its purpose, around three main themes: obesity prevention; facilitation of social personal development; and encouragement of positive behaviour. | Comparative study | Questionnaires and semi-structured interviews. | <ul style="list-style-type: none"> • Consent forms • Questionnaires (hard copy) • Interview files (audio-recorded) |
| <i>Student 4</i> | 'A research study exploring the connection between parents' and childrens' attitudes towards Physical activity'. Investigating parental attitudes in regards to physical activity, focusing on gender stereotypes and the impact of parents' attitudes on children. | Survey | Quantitative questionnaires (year 9 students and their parents). | <ul style="list-style-type: none"> • Consent forms • Questionnaires (hard copy) |

The development of their research projects followed structured and linear approaches and the students were assisted by the teachers and tutors involved in the module. That is, the module was structured in a set of stages (e.g. topic selection, initial literature review and research proposal) where students had to produce a preliminary research proposal prior to the development of their projects. By the end of their projects students produced a bundle of resources associated with a single case study, a survey or a comparative research study, which included all the materials produced and used during the research. Students were then offered the opportunity to reflect on the research process during semi-structured interviews. These interviews were based on these materials and the inquiry practices and the issues that arose around them, especially focusing on research design, data collection, analysis and interpretation. Analysis of student interviews and the research designs they developed highlighted a number of features and issues:

- students' research and learning epistemologies,
- students' practices and the processes followed while developing their projects, and
- other more general issues, some of them related to the curriculum structure, while others highlighted students' perceptions of research more generally.

Firstly, students had epistemologically naive positions when approaching research. This revealed a limited understanding of the nature of inquiry. Two main aspects underpinned this issue and contributed to the difficulties experienced by the students in the stages of research design, reflection and interpretation of their results. The first of these was the background of the students and the lack of previous experiences in doing research. Across the students' interviews, it was identified that their own epistemologies drove the ways in which they approached their research and had an influence on their practices. There was a contrast between students who followed a positivist approach, and those who followed a more consistent inquiry process in which they had their own initial ideas but were aware of the importance of existing research. Some students perceived the existing research on their topic of interest as the 'absolute truth', therefore they were assuming that their findings would align with results published by previous studies rather than perceiving their own research as a contribution to new knowledge or findings. Student 1 for example, when asked about the process followed to design the research instruments, had already designed the interviews/questionnaire questions before exploring the available literature on the topic of interest:

"They were... [the research questions] I had some questions in mind and I had them written down but I did do my questions quite early. I think I got my questions done before my literature review was finished, um and luckily they do actually back up. My literature review was obviously changing and these do back them up."

It was then recognised, that while exploring the literature review, the research questions were being shaped to explore similar issues to the ones identified by those previous studies. Additionally, the selection of the data collection methods was performed at the very beginning of

the project, on the basis of the student's familiarity with particular methods rather than being informed by the existing research that was explored during the development of the research proposal: the literature review was not perceived as a mechanism for identifying those areas where potential research could be conducted, nor as a way of informing the methodological decisions taken. In this respect, Student 3 had similar perceptions to those of Student 1. When asked about what methods were selected, and the rationale for their decision, the student responded:

“Yeah [referring to the research aims changing]. Methodology as well, because I wasn't overly sure what I was going to use and what was going to be more beneficial to get the results I wanted really. And I think we had... I think we had a few lessons on it by that point, but it was sort of... I was more familiar with these methods so that's why I went with them” (Student 3)

However, other students (Student 2 and Student 4) had more consistent epistemological positions in that, while recognising the importance of their own ideas and initial research interests, they were able to identify new themes, or gaps in the area of interest during the literature review and their methodological decisions were underpinned by prior studies. Particularly of interest, is the case of the Student 4, who started with a general topic and then the literature review helped the student narrow down the area of research as well as identifying new themes to explore:

“Yeah. It was enough [the explored literature review]. The other thing that wasn't enough, it was directly children against parents, it was a lot. . . there's a lot of studies about children's attitudes towards PE, parents' attitudes towards PE but there wasn't actual putting them together. . . which made it kind of better. . . cause it made me doing the research.”

Despite Student 4 initially selecting a specific data collection method, based on prior knowledge rather than underpinned on existing studies, the student was able to reflect on which research instruments were more appropriate, given the nature of their research and research sample. In this respect, aspects such as the students' commitments and their background knowledge played an important role in their selection of research areas and approaches. For these students, their prior experiences during summer work placements in primary schools played an important role. They recognised that these experiences had been useful while developing their projects. Firstly, they had helped them to define their projects and secondly, those who had the prospect of becoming teachers, gained motivation and the ability to identify how developing their projects around specific areas could contribute to their future teaching. However, the approaches followed were of a pragmatic nature. Some of them perceived that the research outcomes of their projects would help them identifying the competencies required within teaching and learning environments.

“Yeah. Because I want to be a PE teacher, I kind of wanted to do a little bit of research in what people actually thought my job was. So I thought I could dive into that by asking teachers of other subjects what they actually thought the role of PE

was. Because then, I thought that would be directly related to what PE teachers were actually intended to do so... and I think I actually have discovered that again, with the three themes that I've discovered. Um, so, that's probably the motivation behind it, so to discover what the purpose of PE is and, as in my future career, what I'm *gonna* be expected to do basically." (Student 3)

In some cases students had developed learning practices to overcome previous difficulties as learners, and their level of commitment proved very useful for the development of their research projects. This was the case of Student 2 who recognised having had difficulties in the past with data organisation and maintaining focus during the development of their assignments. Consequently Student 2 used a mechanism to help her overcome these issues, which involved approaching the task as a whole and focusing on the individual sub-tasks involved. Student 2 then applied this approach when undertaking their research project. This approach facilitated, firstly, the organisation of the data collected and generated during the project and secondly, the interpretation of data and the linking of any findings with existing literature. This particular student, when exploring the existing literature, organised articles, journal papers, and book chapters by topic, and produced a summary record for each item consulted.

"So I print them all off and highlight them and then on the back of the piece of paper, I would write um the tit... [article title], the person who it is um what theme it is and, like, write a little summary on the back or... It's the same with books. I would get a blank piece of paper and write out what I want from it and make sure I reference it." (Student 2)

Secondly, another issue that the students highlighted was the impact of their background and curriculum structure and contents on the quality of their projects. While some students were familiar with research, i.e. from taught courses in which they were introduced briefly to research and conducted small surveys, for others the development of such type of projects, and the production of the associated dissertation, was new and challenging. In the course of their degree, students produced shorter, more reflective essays, in which the outcomes were focused on students' reflection on their own practices and experiences rather than discussion, examination and criticism of existing practices, or contributions to knowledge advancement. As part of the development of their projects, students attended introductory lectures concerning research foundations, research methodologies and underpinning theories. Additionally, they could attend a set of workshops to support them with specific tasks such as data analysis, in which they could analyse their own data or work in pairs utilising sample data provided by the lecturers. However, students recognised the importance of having had a richer background knowledge and how the lack of experience limited the quality of their results and their interpretation. While the latter was explicitly recognised by most of the interviewed students, some of them were able to identify other factors, such as the selected research sample and its size, that had an impact in the degree of detail and quality of their interpretation. Student 4 for example commented:

"I did, yeah, I did psychology which has got research methods in it. But it's not got... I understand this and that but it's not got like... doesn't go into it as much in de-

tail. It just basically goes over qualitative, quantitative and all the types of methods rather than actually going over more practical examples. I didn't struggle with analysis, I just thought if I had more, like knowledge of research then I could have done more tasks like... interpretation rather than just go for the means [the approach is quantitative]..." (Student 4)

Thirdly, some students saw their projects as an assessment mechanism rather than as a "real" research project where they could contribute new ideas, or knowledge advancement, and this contributed to difficulties in reflecting, critiquing and interpreting their results. They assumed that the results they generated would be similar to the ones in prior, published studies, and when these were different, students expressed having had difficulties interpreting them and were concerned as to how this might reflect on the quality, or reliability of their work. In contrast, other students were able to identify the literature review as a mechanism to support the design of their research questions and to gain information about methods used by other similar studies, and the benefits/disadvantages of the methods.

"So, first, I did my literature review, I got some themes. So I put the themes into... for the focus group but then obviously, in your actual... now you've got your data. You've gotta go back and think: 'is that relevant?' But then you gotta go back and think: 'well... actually, that is that' so you gotta think what's relevant, what's not. But then also, you might have, there's a really good point that you didn't even discuss in your literature review. It's a new theme that's come out!" (Student 2)

Other students expressed having had difficulties with research design. Student 1 and 3, for example, conducted short semi-structured interviews with some of the participants who completed questionnaires. Student 3 performed a preliminary analysis and then, based on that analysis, designed the interview schedule to explore the topics of interest. In contrast, Student 1 did not perform a preliminary analysis to inform the interview design. This was partially due to unforeseen issues, but also due to the original research design and the lack of previous experience.

"I collected all the questionnaires back... because that was in [place]. That's the only time I could get in so... because um yeah, [coordinator of the school] did email me a couple more but I didn't have time to do any analysing [analysis] before the interviews so I was just basically working on what I thought they would say in the questionnaires for my interview um... so I couldn't really elaborate, I didn't know whose questionnaire was who but I couldn't elaborate on their answers anyway, because it was anonymous"

Lastly, the guidance of teachers and tutors and their ability to refer to previous years' student dissertations proved useful when students were in the final stages of their projects. However, students were only able to refer to final products rather than contextualised research projects, and as a result the role of prior student projects was to provide a model of 'products' rather than insights into how other students had conceptualised and addressed research questions, challenges and dilemmas. The dissertation alone does not necessarily reflect the context of and the

processes followed by the students to construct reflections and knowledge. In this respect, working with archives could be useful to explore the research methods of other research investigations. The sampling methods used, data collection and fieldwork strategies, interview schedules of earlier research, or previous students' projects, can help new students in the design and development of their projects. The process of creating a research archive involves the documentation of the context of the research and includes both the collected and produced materials, as well as a more explicit representation of the inquiry processes followed. Involving students in the development of archives could help others to get a better understanding of research approaches, and the processes involved. Furthermore, the introduction of archiving processes within the development of such kinds of projects could facilitate locating students' individual projects in a broader community of inquiry. Students' generated collections could be integrated with others, and exposed for further study. One particularly informative example was provided by Student 4, who located an existing research instrument which proved useful not only in providing a means of gathering data, but also in framing their project more broadly, suggesting analytical themes and highlighting areas of difficulty. More significant, however, was the student's recognition that this instrument and the approaches that accompanied it would be potentially useful in their future teaching practices.

“it [the developed questionnaire] shows like, what aspects of PE girls like, and what they don't like and that. It's good but then at the same time it's hard because it's saying there's such varied opinions between girls and boys, it's like 'how can you plan mixed classes for PE' so... if the questionnaire itself was like... more adapted specifically to that actual, like, type of PE, then I think it could be useful.” (Student 4)

The project was recognised as more than a stand-alone assessment exercise, and had the potential to become the core of an extended inquiry process. The student's initially pragmatic, outcome-orientated approach (here, it was a research instrument that would gather the data for their dissertation) was elaborated and extended in the course of the project. Offering a means of archiving their research would allow students on this kind of learning trajectory a means of realising this change in orientation.

5.2 Researchers' Practices and Perspectives on Qualitative Archiving and Sharing

This section is structured as follows. Subsection 5.2.1 includes descriptions of projects in which the participants had recently participated. Aspects such as the nature of the project, the kinds of data produced, and participants' perspectives on archiving and data sharing for the selected projects are discussed. Subsection 5.2.2 discusses the participants' research practices, with a focus on their research approaches and the processes of data collection, organisation and analysis. The participants' perceptions of qualitative archiving and data sharing more generally

are also discussed in this section. Lastly, subsection 5.2.3 discusses the participants' teaching practices in undergraduate level, research-oriented modules with a focus on the kinds materials they use to support those practices and the issues/difficulties that students face.

5.2.1 Exploring Research Projects and Digital Archiving

The aim of the first project that was discussed (Researcher 1, first row in Table 5.2 overpage) was to evaluate the understanding of, and policy development on, creativity in the classroom, by studying a number of cohorts of recently qualified teachers. The research approach in their project involved action research conducted by the participants themselves (participants kept research diaries, and observation notes from their own activities in the classroom) and participatory research activities in which participants and researchers engaged in joint activities exploring participants' practices and teaching and learning activities.

Due to the multi-site and multi-institution nature of the project, the researchers used a VRE to archive the data generated. The VRE provided a space for students to upload and store their research diaries and collected materials, and for researchers to manage data and team communications (with participants, between participants and between researchers). The nature of the data collected by participants along with the different cohorts of participants necessitated the set up of a data management plan and ethics protocol in the very early stages of the project. The cohorts of participants ranged from final year teacher trainees to teacher training postgraduates in their first year of work placement. The data management plan covered aspects such as the type of materials collected, the intended uses of the materials, how they were going to be kept, and for how long. The researchers involved in the project (Researcher 1's project) also explored the implications of archiving the project's data and the possibility of enabling access to and sharing of research data with others. That exploration raised a number of issues. The first issue, the nature of the data collected, was recognised as a barrier not so much to archiving but to sharing. Data sharing, between the project participants and with the research community more broadly, could potentially affect participants in negative ways. Researcher 1 explained that the participants perceived the virtual environment, and the participatory activities as a 'private space' where they could freely discuss issues and experiences at work. There were concerns that such reflections, if made public, could have implications for them in their workplace. From the researchers' perspective, data sharing could potentially affect the researcher-participant relationship, and even bias the data collected.

Table 5.2: Research projects overview

| Participant | Project | Approach | Kinds of data |
|--------------------|--|--|--|
| <i>Researcher1</i> | The project was an evaluation of the understanding of implementation of state guidance and policy on 'Creativity' by aspiring and recently qualified teachers. | Action Research; qualitative methods. | <ul style="list-style-type: none"> • Participant-generated research data: research diaries, images, interview data, observation notes • Workshops data, including audio/video files, text notes • Interviews • Still images |
| <i>Researcher2</i> | Researcher 2 was involved in an international collaborative research project (multiple institutions involved) that aimed at the development of a course to inform teacher education programmes and to enhance the institutional capacity in international sectors. | Qualitative methods. | <ul style="list-style-type: none"> • Consultation, content-writing workshops • Summary reports • Meeting minutes, observation notes • Pilot course contents |
| <i>Researcher3</i> | This project was a single institution study aimed at recognising and describing the special nature of learning and teaching in a research-intensive, collegiate university. | Mixed methods; qualitative and quantitative methods. | <ul style="list-style-type: none"> • Questionnaires including qualitative and quantitative responses • Research diaries • Audio/video recordings • Still images • Observation notes • Focus groups • Interview data |
| <i>Researcher4</i> | This project was a doctoral ethnographic study exploring performance and local identities by looking at how participants' own identity shapes their interaction and participation in organised leisure activities. | Ethnography; qualitative methods. | <ul style="list-style-type: none"> • Fieldnotes • Still images • Qualitative questionnaires • Interview data: notes, recordings |

In addition to the above mentioned implications of data sharing for the project, Researcher 1 highlighted a second issue that was identified owing to the initial analysis of the requirements for archiving the project's data. This issue was the institutional and financial commitment required for supporting and maintaining the infrastructures for archiving beyond the life of the project. A specialised digital environment was used in the project, and therefore this required set up and maintenance. Additionally, ethics specifications required destroying the raw data after the specified time span once the project was officially completed. That is, transformed/analysed, and published data, would remain but all the raw data would disappear. Additionally, there were some concerns related to the nature of the technologies used in the project, and more general

ones about digital technologies (hard-copy versus digital data). For the participant, there were potential risks for data gathered in digital forms, especially in those cases where continuous support and maintenance are not available - again, directly related to institutional commitment:

“Yes, if [funding body] for instance were to say: ‘you don’t submit the report but you submit all the data for archiving’. That certainly would have built that into our ethical protocol and consent forms and we would then have a place to hand over the data and also a chance to retrieve ourselves as well as actually to... for other people to take advantage of it” (Researcher 1)

Despite the different issues and barriers that were highlighted, Researcher 1 recognised the importance of archiving. It would enable traceability of the generated research data, along with exploring new emergent themes, and it would also enable exploring the data further. For example, undergraduate or doctoral students could draw upon archived data rather than gathering primary data. Researcher 1 also pointed to other modules, about research methodologies, in which students are encouraged to go and explore existing archives.

The second project (Researcher 2, second row in Table 5.2) was a multi-institution research project that had two main aims. The first aim was to inform further teaching education programmes and the second was to enhance the institutional capacity of the collaborating institutions to work in international sectors. The outputs of the project were mostly publications, dissemination events and the development of a pilot teacher training course. In this particular case, the research approach was through participants’ involvement (project coordinators, teachers, students, external stakeholders), as part of a joint learning process. Data was gathered informally via consultation activities, invited speakers, or course design workshops, rather than following formal processes (consent, information sheets, formal interviewing). The outputs from those activities then became the ‘data’, some parts were not recorded, but still informed researchers’ understanding of issues. Particularly of interest was the design of the course materials which involved firstly an exploratory workshop to share participants’ understanding, ideas and teaching practices. This was followed by a set of content writing workshops drawing upon the data gathered from the previous activities. Then, a pilot course was run with students and lastly, feedback sessions with them reported positive feedback. The researcher’s epistemology and his understanding of qualitative research played a key part in those processes of course design and learning through experience and interaction with the participants in the different settings:

“... for a researcher, it’s um, you are being fed by the information around you without you even noticing it... so from that point of view, it has become much more sort of participatory ethnographic engagement in the field on a very regular basis, which has allowed us to learn a great deal and therefore it becomes very difficult to separate um from what you knew before you engaged in the field and what you know now” (Researcher 2)

Researcher 2 identified phenomenology, or narrative inquiry as the research approaches of their choice. For the participant, research was perceived as a personal, reflective process through

interactions with the research setting. It was described as a process that draws upon one's experiences and is constantly influenced by prior data gathered. The participant's perceptions of research had implications on his practices of data organisation and analysis, i.e. data collection and information gathering is sometimes implicit and occurs without the researcher even noticing it. There was an implicit learning/interpreting process taking place while gathering data, which made very difficult to describe, separate the interpretations from the actual data gathered. Likewise, the processes of gathering data, organising and describing those data, and analysing them were described as intertwined rather than a set of separate stages, e.g. sequential data gathering, organisation and analysis. This represents a challenge for archiving, particularly in those cases where data re-use with secondary analysis purposes is intended. However, contextual information, and detailed documentation of the interpretation processes followed, could be added to the archived materials. While this type of research differed notably from medium-large scale projects, for which archiving was a requirement and the outputs were less interpretive, there is still potential for personal archiving, that could be useful both for the original researcher (facilitating further exploration of the data) and for teaching purposes, providing insights about the interpretation processes, especially focusing on analysis and writing up.

The research project described by Researcher 3 (third row on Table 5.2) presented important similarities with the first case described (Researcher 1). Firstly, a range of participatory approaches were used. Mixed methods combined data gathered from the students participating in the project (including diaries compiling life experiences, pictures or audio-visual materials gathered) with focus groups to further explore students' perspectives around the topic of research. Secondly, the research team involved in the project used virtual research environment to manage, archive and visualise the different kinds of data gathered along with project activities. However, in this case data archiving was a requirement for the project, although only for its use internally within the university. Much discussion was held in the course of the project about data archiving and how to manage and present these data. It was recognised that the research approach generated large amounts of data, which was not anticipated in the beginning. The quantity of data, in combination with the different audiences to which the archived data were presented, was challenging and had an impact on the archive design. While the team of researchers involved in the project had similar backgrounds within the social sciences, the audience for the research was inter-disciplinary. Consequently, adapting the ways in which information was presented depending on the audience was important. The initial approach was to present the information in the form of brief case studies, based on Stenhouse's 'Case record' (Stenhouse, 1978) for the qualitative data, and statistical data was produced from the questionnaires. Those reports were brief, therefore the process of selecting the materials was also quite challenging, i.e. gaining an understanding of what was done and what data meant was described by Researcher 3 as difficult. In this respect, Researcher 3 explained that the use of the VRE helped solve most of those issues. Researchers' diaries in combination with a presentation structure, in which summaries of the findings around a number of topics were combined with interpreted accounts and the underlying raw data, helped not only to provide a better understanding of the

work of project but it also helped to present the data to the different audiences within the university. However, researchers had to perform most of the processes manually, i.e. the research environment provided a linear navigation through a set of web pages that included interpreted data with hyperlinks to accompanying raw data.

Lastly, Researcher 4 (last row in Table 5.2) conducted an ethnographic study exploring performance and local identities by looking at how participants' own identity shapes their interaction and participation in organised leisure activities. The nature of the data gathered in this study was similar to those gathered by Researcher 2. An important set of the data (participant observations, field notes) included not only descriptions of what was happening in the setting but they also included a preliminary analysis and sense-making of what was being observed. However, the research methods - participant observation and structured short interviews - produced still images, audio recorded and partially transcribed interviews that were supplemented with field notes. While the data was mostly paper-based, organised manually by extracted themes and dissertation structure, it was recognised that the use of technologies, for example to segment the audio recordings and their classification by coding scheme and/or research questions, could have been useful. The kinds of data gathered and the ways in which they were organised (processes), represent a good example for the use of archiving tools. Research notes, or analysis documents were classified by topic, additional note cards were used to record relevant quotes from the interview data and attached to the associated analysis document. This organisation by topic or 'dissertation chapter' formed the structure of the participant's set of data, which were used later on during the writing up stage.

5.2.2 Research Practices, Archiving and Data Re-use

The participants' areas of research ranged from mixed methods, participatory approaches to phenomenological and ethnographic approaches. Across all participants, the nature of the data gathered and the research practices had an impact on the kinds of data produced, and more importantly, on the organisation processes and archiving practices. For those approaches involving mixed methods, or participatory methods, the processes of working with the data and organising them were structured. For example, both Researcher 3 and Researcher 4 organised the data collected in similar ways: Researcher 3 worked mostly with digital data organised in a folder structure, and Researcher 4 used a similar approach but their materials were paper-based, supplemented with audio files from recorded interviews. Raw data were documented with descriptions, or margin notes. Researcher 3 also kept a digital research diary in a VRE, which referenced raw data from data collection so that their reflections and descriptions were linked to the relevant raw data. The importance of adding sufficient contextual information was highlighted, not only for the benefit of the researcher as it facilitates locating the relevant data and how analysis is presented, but also for other researchers. This was especially important in those cases where some form of collaboration was present so that researchers worked on the same data with different purposes. In this case the research diary was perceived as a mediating artefact, which added contextual information to the different research and data collection activities

that were taking place. It included some preliminary analysis so other researchers could look at them and obtain a better understanding of what was happening. In contrast, in those cases (e.g. Researcher 2) in which researchers perceived research as a process of interaction with the data, i.e. co-constructing meaning of the phenomenon, the research process was described as an abstract one in which the researcher's own epistemological underpinnings were injected within the interactions with the data. The latter implied less structured data organisation processes, that is, analysis and interpretations occurred at the time of writing rather than in the form of analytical notes, or descriptive summaries of the interpretive process. For example, Researcher 2, when asked about the materials organisation and analysis processes, responded:

“I suppose in my case, most of my own lived experiences have um [reflecting, pause] connections with the stories I collected in the field and therefore, in some ways, my own experience provided... validated or ratified these stories whereas the other end, the data which came in the form of narratives, informed my own lived experiences, it helped me to make sense of my own, you know, lived experiences. So um and I think it was also the thinking process rather than a technical process of doing things in a particular way, I suppose I did that much more in abstract form. Unstructured way inside my brain and therefore I was able to synthesise that information to be able to write about it [very reflexive on this].” (Researcher 2)

Similarly, for Researcher 4 whose preferred research approach was ethnography, initial analysis and interpretation were very often embedded in the data gathered (as described in the previous section). However, in relation to the participant's data organisation practices, the organisation process was more structured: data was organised by themes or research questions, analytical notes were added to different collected materials, therefore there were some “intermediate” materials that were then used when writing up. Such cases, especially Researcher 2, clearly represent a challenging scenario for archiving, and more importantly for data sharing. While archiving would be useful for the researcher's own benefit (personal use of the original data for further exploration), adding contextual information or documentation to provide other researchers with sufficient information about the research processes could potentially imply a change in practices for the original researchers. However, such issues should not be considered insurmountable barriers to archiving: contextual information at study level, such as the research design, methodology of the study, or limited information about the participants, in addition to carefully selected data collected, could still be provided.

The different kinds of data that researchers gathered, their organisation and archiving practices represented interesting scenarios with the potential for being enriched by using archiving tools that support semantic information linking. In the first research project (Researcher 1), the VRE that the researchers used provided a common space in which data gathered from the participants' experiences in teaching environments were combined with the research assistant's observations and diaries from the research settings. These data were then used to generate a correlation of what the participants were expressing and what the researcher observed in the field. In this case, the VRE enabled to store those data together but the correlation was performed ex-

ternally. With respect to Researcher 3, their research approach was participatory data collection in technology-oriented projects in which prototypes or online tools were used in teaching and learning activities. The outputs for these activities were then presented as case studies or *vignettes*, which drew on carefully selected materials from the data gathered. Those outputs were typically used to present a summarised description of what took place in the different research settings accompanied by selected verbatim quotes, as well as diagrams, or images related to what was produced.

With respect to data sharing, with secondary analysis purposes, those researchers engaged in action research or participatory approaches to research identified less barriers to data sharing and open archiving than those participants engaged in other research approaches. For example, Researcher 3's prior experiences showed the possibilities for secondary analysis in those scenarios where the same data were used in reports discussing different themes, or when looking at data that were coded by other researchers, and then asking new questions. Furthermore, one of the recognised advantages of secondary analysis was, when looking at larger-scale data, the possibility of analysis generalisation from a more diverse number of fields, without requiring new data collection. However, the risks of data mis-interpretation in those situations where not enough contextual information was present, or there was an absence of the *whole* perspective on the data if producing incomplete datasets, were highlighted.

“research should be shared as widely as possible. I think you would need to be careful about the level of analysis that you present because things could be interpreted [pause] in a negative way. If you just present the raw data, or if you just presented it half-formed and it doesn't have the right context... so I would probably would have quite a lot of labelling and contextual information related to the data...” (Researcher 3)

Similarly, Researcher 1 participated in a number of projects investigating trainee teachers' practices, in which participants were actively engaged in data collection activities. Participants were interviewing each other regarding their own practices. Those interviews were recorded and transcribed by the research assistant of the project. What was interesting here was that the research assistant was not involved in the data collection and was analysing participants' collected data without having access to all the contextual information or the events during those interviews, and yet this was not identified as an issue. The activities with participants were identified as being quite useful, both to the researchers and the participants themselves. On the one hand these activities facilitated the building of a good rapport between researchers and participants and on the other, they allowed for an exchange of ideas that may not have been possible within a traditional researcher-participant power dynamic. Additionally, the importance of participants' voices was highlighted, as an emergent issue, when reflecting upon such kinds of projects. Researcher 1 recognised that the accounts from the exploration of such interactive experiences with participants, along with the possibilities for archiving activities and data sharing could be enablers of new research processes in which participants would have a different role. Partici-

pants would be co-researchers, involved in developing research questions as well as analysing the data and co-writing with researchers.

Lastly, publications, either in journal or at conference presentations, were the most common output for all the participants. Research data sharing and semantic technologies are contributing to the construction of “enhanced publications”, publication together with a subset of underlying research data. Participants were asked about the concept, and while not all of them were familiar with the term, most of them recognised the value of including original data accompanying the research publications. For example, when asked about enhanced publications, Researcher 3 commented:

“I have heard of the concept [enhanced publications] and I’ve been pretty interested in it, particularly as in my master’s research, I used some video-based research methods and I felt it was a shame that I had to transcribe the videos in order to analyse them because my project write up had to be word-based so I was interested, in the future, if it would be possible to present video as your analysed data but I haven’t yet... I would prefer to do that, I think it would be more valid in a way to present the data in that way but there would be ethical issues I think. I would have had to know that I planned to do that from the beginning so to get consent from participants” (Researcher 3)

For some participants (Researcher 1, 3), it was perceived as a more valid way of presenting the data, so that other researchers could get a better understanding of the original research process and how the conclusions were reached based on the data collected. Moreover, Researcher 4 argued that the nature of qualitative research quite often leads to only showing one perspective, that is, the one of the researcher. And then, with respect to enhanced publications commented that having access to the original data alongside the publication would be useful since it may facilitate looking at the original research from different perspectives. However, emergent ethical issues were identified. Firstly, the ways in which outputs were presented would have required to be known in advance, in order to specify and define appropriate research protocols, reflecting and informing participants about further uses of the data, not only concerning the archiving element but also the subsequent uses within publications. Secondly, the use of data alongside publications could potentially require further consideration on the part of the researcher about the level of analysis and how it is presented. The lack of contextual information could bring up different issues such as mis-interpretation of the results, or the actual research performed.

To conclude, Table 5.3 provides a summarised view of the participants’ research approaches, the barriers for and potential benefits of qualitative data archiving and sharing that have been discussed in this section.

Table 5.3: Barriers to and benefits of qualitative archiving and data sharing

| Participant | Approach | Barriers to Archiving/Sharing | Benefits of Archiving/Sharing |
|--------------------|--|---|---|
| <i>Researcher1</i> | Participatory approaches; re-action search. | <ul style="list-style-type: none"> Complexity of the ethics protocol, and degree of data openness - who has access to what data. Impact of sharing/archiving on participants, i.e. participants being able to access other participants' data. Participant-researcher relationship. Participants often considered the research activities as a personal space where they could discuss personal or sensitive issues related to their work practices. | <ul style="list-style-type: none"> New research processes in which participants could contribute to the archive as co-researchers. New research questions could be developed, and new analyses performed. Traceability of the generated research data, and the exploration of new themes or further exploration on existing data. Use of archived materials to support teaching and learning activities. |
| <i>Researcher2</i> | Narrative inquiry. Phenomenology. | <ul style="list-style-type: none"> Less-structured data collection and organisation processes: the description and analysis of data was intertwined rather than a sequential process (e.g. data collection, then analysis). Production of 'informal' sets of data, from a wide variety of events through participants involvement. Not necessarily recorded but that informed the researcher's understanding of issues. | <ul style="list-style-type: none"> Personal use of the original data for re-visiting or further exploration. Use of carefully selected archived materials to support teaching and learning activities. |
| <i>Researcher3</i> | Participatory approaches; qualitative methods. | <ul style="list-style-type: none"> Institutional related issues: privacy and confidentiality. Institutional internal use of the data rather than being shared with the wider research community. Risk of misinterpreting the data. There is a strong perception of data sharing which could potentially lead to data misinterpretation: other researchers or institutions could interpret the data in unexpected and unwanted ways | <ul style="list-style-type: none"> Personal use of the original data for re-visiting or further exploration, and for others provided that the materials are carefully selected and include sufficient contextual information. Better understanding of the original research process and how the conclusions were reached based on the data collected. Exploration of large-scale data at different levels (raw data, interpreted data and publications which could be linked to the supporting raw data), and from a more diverse number of fields, without requiring new data collection. |

Table 5.3: (continued)

| Participant | Approach | Barriers to Archiving/Sharing | Benefits of Archiving/Sharing |
|--------------------|-----------------------------------|--|---|
| <i>Researcher4</i> | Ethnography; qualitative methods. | <ul style="list-style-type: none"> The nature of the collected data, which was highly interpreted. The data included the researcher's reflections on and perceptions of the observed phenomenon or participant. This is challenging in terms of the documentation of data and provision of contextual information The notion of 'personal data' and data ownership. This concern was not so much about others exploring the data but more so about its usefulness in terms of how the data could be understood and made use of. This concern was due to the data being very tied to the researcher's interpretation. | <ul style="list-style-type: none"> Personal use of the original data for re-visiting or further exploration, especially in the case where data was not explored or analysed owing to time constraints in the primary research. Exploration of the original research from different perspectives: publications often show the perspective of the original researcher. If the publication was accompanied by selected raw data, others could interpret the data from another perspective. |

5.2.3 Teaching Practices: Supporting Undergraduate Research

This section focuses on describing the interviewed participants' practices when teaching research-related modules to undergraduates, the types of supporting research materials and how they were used within teaching activities, along with the issues/difficulties students experienced. All the participants recognised the degree of difficulty and the challenges of teaching research methods modules to undergraduate students, and therefore a wide range of approaches needed to be used to help students grasp the concepts and the process of research. The interviewed participants were involved in a range of curricular activities involving undergraduate, postgraduate and doctoral students. They were at one point involved in teaching modules in which undergraduate students developed small-scale research projects, in the final year of their Education Studies degree. Additionally, some of the participants were involved in teaching action research and visual methods to undergraduate and masters' level students¹.

The differences between school-based research projects (action research) and more general inquiry projects were quite interesting. In the latter, the research approaches typically involve developing the ability to test, question and critique knowledge, rather than more pragmatic approaches, in which the focus is on developing research instruments relevant to their career path (i.e. becoming a primary school teacher) and reflecting on their own practices. Researcher 2 recognised the importance of encouraging students' awareness of the importance of questioning knowledge, whether acquired in the classroom, or on a daily basis: "If we know something, it is important to ask how do we know, so how do I... you know, if the government is saying that teaching phonics is the only way forward if we would like to improve literacy of our children, then as a teacher you would need to ask or we would need to ask 'where is the evidence?, why..., is it true?'" , rather than simply preparing students for practice. In relation to pragmatic approaches, the participants recognised that students had issues with identifying the relevance of research methods in relation to their career prospects and therefore, supporting them so that they are able to question their methods and practices, reflect on how they do things and the rationale for those decisions, was identified as a key objective for such modules.

For those students involved in action research, most of them had prior experiences of analytical writing, conducting a literature review of more generic topics, and/or a subject of interest within teaching practices. They developed reflective writing practices, both on their individual teaching and their teaching profile, so that students reflected on what they learnt, and how it related to the literature review. These processes were supported by the use of online tools (e-portfolios), to keep a research diary, as a reporting mechanism which was part of the assessment. To support them in the research processes, students participated in a series of short activities to help them personalise and apply the different elements involved in the research process. Researcher 1 used exemplar, anonymised materials from previous students to support students' understanding the different elements that are part of the research process, for example, if looking at developing one's research focus, students were shown concept maps, or when discussing results presentation, they looked at different ways of tabulating or presenting data.

In contrast, those students not involved in school-based research studies took an introductory taught module to research approaches and research methods, prior to the development of their research projects. This module aimed to support students writing their research proposal, as the basis for the development of their own projects². The backgrounds of the students were quite different, and while some of them had prior basic research experience from other courses where they designed questionnaires, the objectives of the students differed considerably. Prior to the development of their research projects students were presented with some theoretical perspectives about research in which epistemologies were simplified so that students were encouraged to distinguish between qualitative and quantitative approaches. The kinds of materials to support those activities ranged from research textbooks (in combination with diagrams or visual representations of research methods) to exemplary materials for a number of sources, or sample materials from the lecturers' own research. Additionally, students attended a number of sessions in which they had group discussions around research articles in relevant fields or even around exemplary materials, in the form of prepared data, including pieces of analysis, supporting discussions around different research approaches, why they were appropriate; to help them linking between theory and practice. Table 5.4 summarises the kinds of materials that the interviewed researchers used to support the teaching and learning activities.

Table 5.4: Kinds of materials used in teaching activities

| Participant | Materials supporting teaching activities |
|--------------|---|
| Researcher 1 | <ul style="list-style-type: none"> • Exemplary anonymised materials from previous students. • Prepared original research data samples: interview excerpts; observation notes. |
| Researcher 2 | <ul style="list-style-type: none"> • Exemplary research instruments and designs with small case studies prepared by the researcher. • Online electronic case studies. • Invited guest speakers who present their research to engage students in discussions about the different research approaches. • Exemplary dissertations from previous years' students. |
| Researcher 3 | <ul style="list-style-type: none"> • Research Textbooks, diagrams and visual representations of research methods. • Exemplary materials from the researcher, lightly edited to show particular approaches. • Exemplary dissertations from previous years' students. |
| Researcher 4 | <ul style="list-style-type: none"> • Small case studies. • Prepared materials that combine snippets, real and fictitious examples, of different research approaches. • Exemplary materials from the researcher, adapted for their use in the lectures. For example, during a lesson on Ethnography students were shown excerpts of the lecturer's own research diary. |

One of the most common issues students faced was their lack of familiarity with the theories

behind research, which aligns with the students' perceptions discussed previously (in section 5.1). The ways in which students approached their research were based on prior experiences in other modules. Students were familiar with producing short reflective dissertations but they were not aware of the research process and its associated outcomes, that is, production of new knowledge. Additionally, the development of the research project was located in the final year, its weight was considerably higher than any other prior module, and more importantly, they had to approach the research entirely on their own for the first time. Due to the lack of research background, students had difficulties understanding what was expected from them, and how the research process led to the dissertation writing. Students had to choose a topic of interest, reflect on the most suitable approaches, write it up and then, it was assessed. To support students with the latter, most participants used similar approaches. The use of exemplary dissertations from previous years' students, for example, was identified as useful to help students with the structure of their dissertation (what were good/bad examples). However, the latter approach gave them insights into what to produce, rather than how to get to that stage. Attempts to solve this included guiding students through a very structured process in combination with early formative sessions with supervisors for feedback and support. Before writing the proposal they filled in a form expressing the areas of interest, aims and outcomes for their research, which was followed by the proposal writing, where they had to specify an initial literature review, methods of choice and so forth. However, there was an issue with them understanding the concept of methodology, literature review, how analysis is performed, i.e. understanding the sorts of more practical expertise.

With respect to practical expertise, another key issue for students was the analysis of their data after data collection, thus the participants conducted a series of practical workshops on research methods in which students brought some data of their own, or they were given exemplary data so that they could familiarise themselves with analysis, either individually or in small groups. Additionally, most of the participants drew upon existing qualitative small archives - online case studies, prepared small case studies - to support students with data analysis and interpretation. The use of existing archives and previous students' work was identified as useful to help them with their own research, identifying different research approaches, and those methods that are appropriate:

“In a session on doing questionnaires or doing interviews or doing observations, in which case I drew on actual data, sometimes of my own, or I have used data from the educational evaluations archive [online archive] to show them interview transcripts so: interviews with researchers to see them reflecting on why they used a particular approach, um or images that were produced as research data that are useful...” (Researcher 3)

The use of existing small archives can play an important role to support students' understanding of the research process. The students and researchers that were interviewed could appreciate how it would support research design, data collection and analysis, and writing up. In

contrast, other participants who have used similar approaches expressed that sometimes students thought that showing them research with which they were not familiar, or research that differed from their topics of interest, or background, was not relevant.

“[using existing online archives] I think that [would] be great to show to students but it's not subject specific, even though that they would get a lot out of it, they sometimes put up this mental block when they go 'this is anything to do with tourism or events'. But actually they might not, they might actually, you know, see the point of what I'm showing them. For example tomorrow, I'm talking about visual research, I've got a video from an anthropological institute which is about getting the participants involved in the filming, and it's a tribe in Africa... There's no way they [students] are ever gonna go and do their research in a tribe in Africa, but they should be able to understand...” (Researcher 4)

In this case, presenting them with work and small archives of previous years' projects, along with small case studies drawing on lecturers' own research could alleviate the latter issue since the study area would be similar. Teachers archiving their own materials, preparing a small set of data covering research design, research questions, and exemplary data collection materials linking to bits of analysis, especially if they are visualised in ways that show how the different materials relate to each other, could help students gaining a better understanding of the notion of research by bringing authenticity to the classroom. Additionally, it could help to familiarise students with the processes of organising and selecting materials from their small research projects to produce a research archive. However, such approaches of exploring archived materials, whether sourced from students' projects or prepared by the teachers, require teachers' mediation to support the learning activities. The creation of exemplary collections, with the support of archiving systems, such as the toolkit implemented in the present study, should be embedded within teaching practices in order to be useful and to add pedagogical value. Therefore, for archiving to be a pedagogical practice, it should be integrated into the work of both teachers and students, and be a focus of the pedagogical discourse. Getting teachers to self-archive their research and modelling the approach represents an important pre-requisite and serves as a 'model' for students developing their own projects. Students need to be walked through the existing archives in order to be able to understand not only the process but the epistemologies involved in the knowledge and solution construction.

Whilst participants recognised the benefits of using existing online resources or small archives to support teaching activities, they highlighted issues around overwhelming students with lots of information and the effort required on the part of the lecturer to prepare the materials, that is, exploring existing archives in order to present subsets of relevant materials. Moreover, a number of participants highlighted that there were curricular changes taking place which moved the orientation and content of the module from research. This was attributed to time constraints and as a partial response to the difficulties, or challenges that students faced when developing their research projects. In this respect, the self-archiving toolkit developed in this study seeks to provide them with a means of producing collections of exemplary materials, by following pro-

cesses which do not differ substantially from their current practices of preparing and presenting exemplary research materials. This is discussed more in detail in section 5.4.

5.3 Exemplary Case Studies Digital Archives

The last section of this chapter describes two research projects which led to the development of multimedia case study archives. As introduced in Chapter 3, the first archive was a multimedia case study exploring a school-based research setting, which was developed to support masters' level students undertaking action research. The second archive was a set of digitally archived case studies that was developed in the course of a large-scale research and development project.

5.3.1 Multimedia Case Study of a School

The first example was a multimedia (MM) case study in an archived form that was developed to support masters' level students in developing their own research inquiries. The multimedia case study was included within a Master's distance course (aimed at primary education teachers) as part of an introductory module on action research. However, a significant minority of the students took the course for other reasons: the background of the students was quite varied - nursing, management, and computing. The structure of the course was based on three different sections. The first section, 'Inside the classroom', focused on supporting students performing research in their own classrooms by performing a set of small research tasks. In the second section, students used the multimedia case study to perform research tasks. Lastly, the third section, 'Inside the teacher', was about getting teachers to reflect on their own teaching practices and motivations. The multimedia case study was based on a prior case study which was the output of a doctoral thesis studying teaching practices in school-based settings (Groundwater-Smith, 1989). The focus of this interview was on the second section of the course, in which the multimedia materials were used as the basis for a set of research tasks that students had to perform. The topics explored covered aspects such as:

- the design processes prior to the archive's implementation, and the key issues faced during those, including technology-related ones,
- the actual archiving processes: archive organisation, materials selection, and data compilation, and
- teaching and learning activities supported by the use of the archived materials.

With respect to the design of the MM case study, Researcher 5 described the process as being challenging owing to the lack of experience, within the team, in putting together all the sorts of materials in an electronic form. To overcome this and develop the structure of the archive, a project team including members with different backgrounds (researchers, academics, multimedia producers) was constituted. The understanding of what was going to be produced, from the technological perspective, came from exploring external multimedia materials, mostly from

outside education. The team members brought in examples of materials they found of interest, with an emphasis on interactivity. Such kinds of materials differed from the educational multimedia materials available at the time, which were more structured, linear and with less interactive elements.

“Another multimedia project we looked at was project ‘Perseus’ which is this classical material that guy worked on, which it’s been through a whole lot of different versions but the idea is to put all the resources you need to do classical research onto a desktop. And it’s now web-based, so it’s architectural science, classical science, the scholarly literature and stories, you know, everything you can think of related to classical studies.” (Researcher 5)

The design process of the multimedia version consisted of a number of iterations in which different materials, some of them from new settings, were added. The main steps of the redesign of the materials into a multimedia form were to revisit the original materials, perform additional visits to school-based settings, and collect new materials, both qualitative and quantitative. These materials were then incorporated into the archive. This iterative process, back and forth to the different settings collecting new materials, was interesting and posed a number of challenges. Firstly, the number of emergent themes increased owing to the variety of the materials collected (video interviews, classroom observations, census data) and therefore decisions around material selection and when to stop the process needed to be made. Secondly, the case study was organised based on Stenhouse’s idea about ‘case study’ and ‘case record’ (Stenhouse, 1978; Walker, 2002). Researcher 5 expressed having had difficulties with the archive’s organisation process, that is in identifying the boundaries of the case record, in terms of material selection. The initial idea was to compile ‘lightly edited’ case records including a description and minimal interpretation of the primary selected materials, while at the same time ensuring the provision of sufficient contextual information. Those records were then organised based on the original case study’s organisation: small sets of materials were clustered around a central artefact, e.g. a video lesson around a specific topic. The selection criteria was based on the perceived relevance of the materials for future use. However, as more data was being collected, new ideas, topics and research questions, were arising. Those led to the development of a new classification/organisation of the materials in addition to the initial case record structure, that is, the materials were categorised around broader themes.

“We came with those headings which were on the front screen [looking at the on-line version of the archive], which... I am trying to remember now... curriculum, classroom, um history, school history, system-wide things, you know, those broad categories, which we thought of as sort of chapters, for the material so we did kind of edit, as Lawrence [Stenhouse] would say, to fit those categories. They seem to be general enough categories that they worked out just fine.” (Researcher 5)

The categorisation into broader themes helped to identify other kinds of materials that were relevant to the context of the school, such as statistical data. This combination of existing data

with external data which came from multiple sources is interesting here. It links to the concept of ‘linked data’ which was introduced in Chapter 2, and how materials can be used to develop new research questions. Researcher 5 explained that the inclusion of both statistical and qualitative materials facilitated students asking interesting questions that were ‘outside of the setting environment’, and gave them a better understanding of what was going on inside the setting by also being aware of what was happening outside:

“One of the big categories was community, so it was around what, where did this school sit in this community, which led us to collecting lots of census data, which is one of the features of the materials. I was very keen on using statistical data. I think it was around that time the Australian census [statistics centre] started creating local census data [public archive] so you could buy a print out of all the census data by post code um and post code pretty much overlaps with the sort of catchment area of the school which was lucky... um so we can get a picture of the community and then you start to get questions” (Researcher 5)

Alongside the archive design and implementation processes followed by the researchers, other key aspects explored in the interview were the ways in which the multimedia case study was used to support teaching and learning environments. While the archived case study was redesigned throughout different runs of the course, the pedagogical uses across the different versions were very similar. The case study was organised based on themes and case records, around which students had to perform a set of small research tasks. Some students said that they had difficulties identifying what parts of the case study were relevant, owing to the navigation across the archive, and their lack of familiarity with the materials. The navigation across the materials was quite linear and based on sections with different theme-based headers. However, the researcher was concerned with ‘over programming’ how the activities were conducted by simply ‘pointing students to the relevant set of materials’. Instead, the researcher, and the team, wanted the activity to be more self-exploratory. With the support of the available technologies at the time, the multimedia case study was then re-designed to include a word-based search functionality as an attempt to help students navigating the materials. This enabled scenarios in which one could be looking at a specific material, identify a topic of interest, and perform a general search to find a subset that references that emergent topic.

What was interesting here was the approach with which the case studies were used to support students’ tasks. Previously, in the section describing the interviews with the teacher participants, one of the identified uses of archived materials was as ‘prepared case study’ presentations, in which materials are selected and edited for their presentation so that students can get a better understanding of different research approaches by looking at existing studies. In contrast here, the role of the multimedia case study was more about providing students, and new researchers, with an environment to ‘do a case study’. To support such an approach, students needed to be presented with sufficient raw materials so that they could explore and contribute to existing materials. Teacher trainee students were interested in authentic, realistic tasks around the case study materials. This required them to explore as many materials from the case study

as they could, while at the same time being able to keep track of those exploration processes. Technologies³ allowed, both teachers and students, to keep track of the navigation through the materials. Students could print a record of the different materials explored in a session as well as the sequence. The feature was also useful for the teachers since they could then explore the ways in which students interacted with the materials, what were the most used materials, and so on. Additionally, students could share their 'paths', or itineraries through the data, so that others could follow the same path or start from somewhere else. This idea was perceived as very interesting and served as the basis for students producing what was called 'guided tours'. Students explored a topic of interest such as gender or language issues, and then selected the relevant materials from the case study for others to look at, or to develop further. In this respect, the researcher highlighted one particular example from a student who had an interest in a particular topic (school architecture) but discovered the lack of information around it within the archive:

"I remember there was a nurse, with one of the early versions [of the multimedia archive], that said 'the school's got this really interesting architecture but I can't find an interview with the architect...' so she said that she'll invent one: 'here is my fictional interview with him...'" (Researcher 5)

The student identified existing gaps in the data and reflected on those questions for which she could not find an answer, thus developing their own research instruments to collect more data on the topic of interest. The use of the case study facilitated students' reflection and stimulated them to formulate new questions for the data as well as identify ways of contributing to it. Whilst the case study was thought of as an instrument to encourage 'students as inquirers', designed for students to explore and contribute to it, some students perceived it as a secondary source. The original research was performed, then the case study was compiled and students explored it as a secondary source with which they replicated the tasks performed previously, but making use of the new materials. Either way, the value that working with archived materials added to the teaching materials was highly recognised. It provided the opportunity to present students authentic data, and more importantly, it provided the space for working and extending the existing data, in the form of case studies, rather than simply learning about them:

"I wanted to find ways of working with case study that were not just read this case study but do something with it...and I thought... I've been involved quite a lot in case studies evaluation and one of the realisations there, was that it's great fun to produce case studies but they are not always so interesting to read [laughs]. So I thought I *gotta* find a way of giving people access to the experience of doing the case study, not just read it... and that's why the case record is directly useful cause you could say 'here is a case record, your job is to produce a case study later' so you actually have to work with the data and it's not quite as good as doing the interviews yourself or all that sort of thing, and doing the observations but if you include bits of video, and stuff like that, is close." (Researcher 5)

Likewise, the different technologies used were perceived as enablers for a range of new possibilities that opened up discussions. New ideas emerged for new designs and uses for the ma-

terials. The affordances of the technologies that the researchers used, in combination with an appropriate organisation structure for the materials allowed for the enhancement of previous materials, and the ways in which they were used. Lastly, different types of evaluations involving students and external academics were conducted. Students were asked to report about their progress and the issues they encountered when making use of the materials, what was easy, what was difficult, their thoughts about the tasks and the materials. External academics came and explored the materials, or evaluated the course as a whole. Researcher 5 explained that those evaluations provided positive feedback but they also highlighted the need for a balance between designing highly exploratory tasks and providing students with more structured and guided ones, so that they are more confident and able to find their ways through the materials. This, again, is quite similar to the experiences of using archived materials with undergraduate students that were described earlier, for which some support and guidance on the part of the teachers is key to support students' interpreting the tasks more expansively and engaging with the materials.

5.3.2 BrainAble Multimedia Case Studies

This subsection presents a descriptive account of the development of a set of multimedia case studies as part of a large-scale research and development project⁴. The project in which Researcher 6 was involved was a research and development project that explored, designed and validated an ICT-based HCI based on BNCI sensors combined with affective computing to control 'smart home'⁵ services and virtual environments. Testing sessions of the prototype developed were conducted with disabled and non-disabled participants in different locations. Alongside the technological component of the project, the researcher interviewed was involved on performing qualitative research and providing feedback (from the testing sessions with participants) so that the developers could improve the system as required. As part of the qualitative research, the researcher produced a set of case studies that were included in an archive. The archive included a wide variety of data - test data from the device, policy documents, or qualitative data from the interviews conducted with the prototype's users.

There were a number of aspects that made this archive such an interesting example of a complex archive. Firstly, the multidisciplinary nature of the project required a flexible design of the archive so that the different stakeholders could explore data in ways that were useful to them. The researchers involved in the project had to deal with very varied results, i.e. the data could be visualised by different teams such as technical developers specialised in different areas, or social sciences researchers, all of them with different backgrounds and responsibilities within the project. For example, Researcher 6 explained that the prototype developers were interested in the technical data gathered from the testing sessions with the participants, whilst the researcher was more interested in analysing participants' interactions with the system as well as their perceptions and feedback. Researcher 6 also highlighted that providing mechanisms for displaying data in different and interactive ways, alongside producing interactive reports, were some of the

motivations for the development of the archive. They were also highly important aspects that drove the initial archive design.

Secondly, the wide variety of data gathered along with their highly sensitive nature placed a number of challenges during the initial design of the archive. Researcher 6 had to develop a very detailed data management plan in the very early stages of the project, before conducting any research, as part of the ethics protocol. The plan covered aspects such as the types of data generated, description of the intended uses of the data that were going to be included in the archive, how long were the data going to remain archived and accessible to others, and whether there was intention of reusing the data, with what purposes and in what ways. All the information included in the data management proposal was reflected in all the consent forms to ensure that the participants were fully aware about the kinds of data that were going to be produced, sourced from testing sessions, and qualitative interviews, and the ways in which these data were going to be used. In addition to this, informal discussions with the participants took place, firstly to find out how they felt about archiving and making the data gathered available for others to explore, and secondly to ensure that they had a complete understanding about how these data were going to be used and with what purposes. None of the participants identified any issues related to data archiving and re-use. Moreover, they acknowledged their willingness to be recognised as participants of the project and they recognised the value of sharing their experiences with the community. One of the most common arguments among social sciences researchers for not archiving data was described in Chapter 2. This argument is that making the data available for others to re-use may potentially affect the relationship and interaction between participant and researcher. Researchers collecting qualitative data, especially when exploring sensitive issues, often assume that participants would not accept the idea of archiving, and tend to be over-protective. However, numerous studies exploring participants' perceptions of research interaction and the potential issues of archiving qualitative data showed that in most of the cases research participants did not share some of the researchers' preconceived arguments (Kuula, 2005; Graham et al., 2007). In this particular example, the researcher emphasised the importance of discussing with the participants, in every session, what data were collected and how they were going to be used, even though the participants consented for archiving and re-use.

Interestingly, the issues around confidentiality and data anonymity came from internal and external partners of the project rather than from the participants themselves. Data anonymity issues were anticipated since there were clients such as hospitals with an interest in using the device, for which anonymity and privacy were absolutely required, i.e. they were concerned about the public release of the case study data. Additionally, data confidentiality was also a key issue across the project since the technical partners required that different sets of the data remained totally protected, at least until the prototypes were ready and fully operational. These issues required a balance, when developing the archive, between explaining the research done in the course of the project and ensuring that the plans for the prototype were kept fully protected. The establishment of appropriate data management plans, if archiving and data sharing/re-use

were considered; or anticipating what data would be generated and the intended uses, are key aspects to design and produce research archives that could be used and explored by different audiences and with different purposes. Researcher 6 recognised that the initial design and discussions among the project team and participants were key in facilitating the development of such an archive.

The researcher designed and developed the archive jointly with an external collaborator that provided the knowledge and expertise on the different technologies available and how they could be used to produce an archive of such characteristics. The researcher used different data collection methods (see Figure 5.1), which generated varied kinds of materials:

- a literature review on existing studies in the areas of Assistive technologies, ICT-based HCI, and Brain-Computer Interface (BCI) was conducted;
- external larger scale surveys based on people involved in short tests in Spain, which provided basic medical data and test results;
- detailed biographical data and surveys of existing use of technologies (assistive technologies and BCI interfaces from a smaller sample of participants in the UK;
- technical data from testing sessions with the participants in the UK. Data was collected during a number of iterations: each participant contributed about 6 to 8 sets of data, and then was involved in up to 6 sets of technical data from the system. Those sets included data from the interactions with the prototype, which was at the second by second level of interaction, attention and brain activity data, and;
- researcher's observation data from the testing session, combined with informal interviews with participants after those sessions.

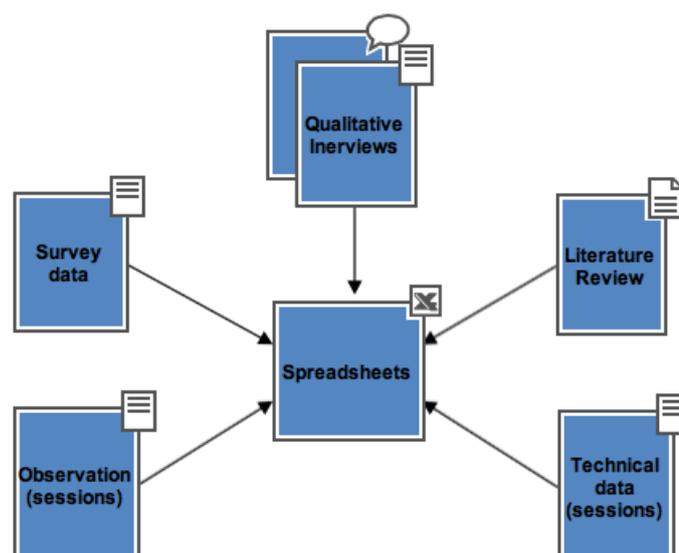


Figure 5.1: Data collection methods employed

The process of selecting the materials was quite interesting: heterogeneous kinds of data were collated together into Excel spreadsheets, over a number of iterations. However, these data were mostly interpreted data owing to the sensitive nature of the data gathered. Researcher 6 described the process as follows:

“Yeah, raw data is kept separate, I mean, and then there is a ‘barriers and affordances’ spreadsheet where everything that’s been said is thematically analysed against things that were identified earlier in the project as “barrier and affordances” to the prototype and these have been... that’s the analysis. [pause] And there’s a ... you know, there’s demo... and we’re also looking all the time for any other factors that could affect: motivation to using the prototype, um demographic factors or drugs that they are taking that could affect to the prototype results...”

This was an incremental design, in which data gathering activities were performed alongside analysis tasks. In the initial stages of the design there was a piloting stage with a small set of data. This stage was useful, firstly to explore the kinds of data that could be displayed together and how they could be linked, and secondly to identify appropriate ways of documenting these data. Additionally, the researcher’s exploration of this small dataset using the archive’s visualisation tool proved useful for subsequent analysis, since it allowed for exploration of the different ways in which data could be visualised and related, therefore making the links between the different information explicit.

The design and implementation activities were complemented by sessions with the project team. In those sessions the researcher ran *demos* of the archive and obtained useful feedback from the different stakeholders involved in the project. The research approach of the project was based on case studies from the participants, therefore it was agreed that the best way of organising the different data available was around the notion of ‘persona cards’ representing individual case studies, around which different kinds of related data were linked. Qualitative data gathered from interviews with participants, observation notes, technical data from the testing sessions, or related external policy documents are examples of the types of resources that were linked to each case study. With respect to the archive’s underlying technologies, a combination of digital repository with visualisation web-based frameworks were used:

- *Persona cards* were created using Cascade Style Sheet (CSS) styling for their display online. Each of them represented a record or participant profile compiling information about the participant disability, about their aspirations or expectations from the participation in the project, and about their use of computers,
- *A Digital repository* to securely store technical data, information concerning the technical outputs of the project or confidential data from the testing sessions with participants, and
- *A web archive interface*, which makes use of Exhibit semantic web framework to implement a filtered, faceted browsing interface, was used to navigate across a number of integrated data sources from the project. Those input datasets were transformed into a set

of spreadsheets that integrated qualitative data from participants' sessions with technical data, and external relevant information.

The use of digital repositories allowed solving the issues around data protection and access control. However, the materials that were incorporated into the digital case studies were transformed into spreadsheets and included selected 'snippets' from technical files, qualitative interviews, as well as analytical data and the researcher's interpretations. The web pages for visualising the case studies, which made use of Exhibit framework, facilitated to organise and display the various datasets in multiple ways.

"Yeah yeah. It's organised deliberately in multiple ways, by the subject, um by what they were tested for, and then their own results are compared internally so their results are compared with their own results so they are also compared against other results with the other subjects... with the same condition, with different condition so...so there's different ways" (Researcher 6)

The use of 'linked data' tools such as Exhibit enabled organisation of the data in different ways, by subject/participant, or by 'what the participants were tested for', and by results comparison over different testing iterations. Secondly, the different datasets included in the archive could be visualised in integrated ways owing to the semantic capabilities of the visualisation framework used. Snippets from interviews with participants were linked to test session data and also contrasted with original statements from earlier stages of the project, including evidence from prior research in the area, related disability information, evaluation data, and so on. Whilst Researcher 6 perceived the use of these technologies as an enabler rather than imposing constraints or barriers, it was recognised that significant effort and amounts of time were required for data entry, which had to be done in parallel to the research activities. However, the researcher highlighted that analysis tasks were enhanced owing to the navigation through the archive and the ability to link different sets of information.

Researcher 6 recognised that the use of the Exhibit framework, more concretely the data filtering and faceted browsing by data features and data relationships, provided new ways of looking at the data, supporting enhanced analyses so that new links and relationships between data were made explicit: "allowing the researcher to think about things in different ways and display data in different ways" (Researcher 6). However, some of the project members said that they experienced some difficulties when using the tools owing to their lack of familiarity with and understanding of some of the functionalities, that is, filtering and faceted browsing. Feedback sessions and demonstrations of the pilot archive work proved useful to help users understand how the tools work and how they could be used to assist them exploring the datasets available. Moreover, project partners expressed their interest in using the archive, both for research purposes and to support developers. The potential for future completion so that a more complete set of the technical data files could be linked to qualitative, participant related data was also highlighted.

The set of archived case studies that Researcher 6 developed was a prototype rather than a complete and fully integrated archive. The archive was used and tested by project participants and some of the project external collaborating organisms. However, secondary analysis and data re-use were some of the central expected uses of the archive, both for the project members benefit (exploration of the archive for further analysis, supporting publications as outputs from the project) and the wider BNCI community, and specialists working in the areas of assistive technologies to provide advice to the different assistive technologies communities. In this respect, Researcher 6 recognised that working through case studies was potentially useful to researchers with an interest on BCI/BNCI technologies as well as to developers, helping them to develop new ideas and further research.

“[expected uses of the case studies] to give professionals in assistive technologies field such as IT, assessors the ability to be able to see what kind of people this kind of system might work for. . . an example of it working through case studies to allow people to actually learn about BCI and BNCI [Brain/Neuronal Computer Interface] and to developers to gain ideas and insights and also for future researchers to get ideas and insights about how you can best communicate to developers.” (Researcher 6)

To conclude, the archive compiled a number of case studies incorporating heterogeneous data from the different project deliverables. The way in which the archive was organised, and more importantly the ways in which users can navigate and search across the data generated by the project, assisted with analysis tasks, supporting both technical and research elements of the project, and proved to be a rich resource for supporting further exploration and contribution, not only internally within the project team but also within the wider community.

5.3.3 Implications for the Self-Archiving Toolkit

The two case studies that were described in Sections 5.3.1 and 5.3.2 represent complex examples of digital archives. Both archives, especially the set of archived multimedia case studies (Section 5.3.2), included heterogeneous materials gathered using different data collection methods and added to the archive over time. This is a very good example of archives for which the accurate documentation of the study, and the provision of sufficient contextual information are central, especially if those archives are intended to support data re-use and secondary analysis. Without these accompanying information, subsequent users may have difficulty navigating the archive, but more importantly, there is the risk that those users may misinterpret the data.

The design process by which the first archive (Section 5.3.1) consisted of a number of iterations in which different materials, some of them collected from new research settings, were added. The archive was initially organised based on the previously mentioned idea of ‘case records’. For example, a video interview was supplemented with a description of the research setting and a limited interpretation of the primary material. For subsequent materials collected from new school-based settings, their inclusion in the archive was determined by the researchers’ perceived relevance of such data to future use in learning environments. These newly added

materials were categorised around broader themes, hence changing its organisation from one consisting solely of case records to a mixed collection.

As with the first archive, the design and implementation of the second archive (5.3.2) was performed incrementally. Data collection, and subsequent data analysis, were performed in multiple iterations, in which new, interpreted data was added to the archive. This incremental archiving process was described by the participant (Researcher 6) as useful, firstly, to explore the kinds of data that could be displayed together and how they could be linked to each other, and secondly, to identify appropriate ways of documenting the data. This archive used 'linked data' visualisation tools (more precisely, the Exhibit framework) which allowed data to be organised in different ways. The semantic capabilities of these visualisation tools also facilitated integrated visualisation of the different materials included in the archive.

The exploration of these two archives contributed to the design of the visualisation tools of the self-archiving toolkit developed in this study. Two features in particular resulted: search functionalities (including filtered search) and a user search history facility. The search functionality allows users to perform general searches across the case study to find subsets of materials that are related to a particular material, or topic of interest. This functionality was incorporated into the 'Collection Explorer' interface, as described in Chapter 4 (Section 4.4.1), and enables text-based searches across the entire collection, or archive. Additionally, a 'keywords' filter was incorporated into the 'Collection Explorer' interface, so that the user can filter data according to specific topics. In this way the user can visualise a reduced set of results related to selected topics. The search history facility that was incorporated into the visualisation tools keeps track of the materials that have been explored by the users of the archive. This feature allows users, once they have explored a topic of interest and selected the relevant materials from the archive, to store these results for future use. In this respect, the collection explorer can keep track of user navigation, by generating 'snapshots' of the search results that are generated once the user has applied one or more filters to the archive. This enables the user to save the results of searches, which are then permanently accessible by simply using a URL in a web browser, as described in Chapter 4 (Section 4.4.1).

The first archive also proved to be a suitable test case of the data model and the archiving processes of the self-archiving toolkit. The participant (Researcher 5), who was involved in the development of that archive provided the author with a web archive that contained a variety of, mostly qualitative, visual data from the original multimedia case study. The types of materials contained in the web archive included still images, video interviews and their associated transcripts, census data, and relevant publications. The author explored this web archive and used the toolkit to generate a "test archive". The design process of this test archive involved a number of steps, as described in detail in Chapter 4 (Section 4.4), are summarised here as follows. Firstly, the archive organisation was defined by basing it on the theme-based organisation of the original multimedia case study. Following the organisation process, the materials to be included

in the test archive were documented. This documentation was performed both at study and data level, that is, methodological information such as a summary of the case study, data collection methods employed, and sample design were added to the collection-level spreadsheets. Data-level documentation involved the creation of file-level spreadsheets, in which each file to be included in the test archive was described. Lastly, once all the spreadsheets were created, the data was stored in the digital repository with the support of the self-archiving toolkit's 'Collection Manager' application.

Compared to the original web archive, the navigation of the test archive was improved in a number of ways due to the data model and the visualisation interfaces implemented in the self-archiving toolkit. Navigation of the archive can be performed in exploratory ways, rather than by linear means. The toolkit's underlying data model allowed one to see how the different materials (images, transcripts, audio files) relate to each other. Additionally, more complete methodological documentation was added by using metadata terms from the different vocabularies that are supported in the toolkit. Figure 5.2 shows a sub-set of the data included in the test archive along with examples of the types of relationships that link the different resources. Lastly, the use of semantic information such as relationships for representing the collection structure, or linking the different data elements, allows the archive's data to be displayed in more interactive ways. For example, a particular resource, such as an interview transcript, could be visualised in an RDF graph along with all the resources, or analytical information associated with it.

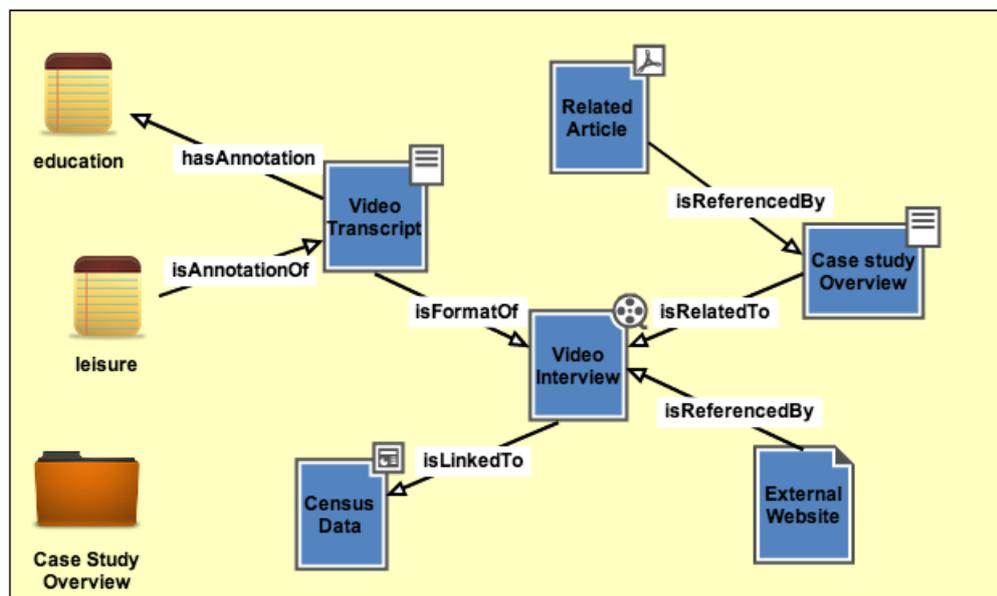


Figure 5.2: Self-Archiving toolkit test archive: sample resources

5.4 The Self-Archiving Toolkit's Support of Research Practices

This section presents together three different cases of research and archiving practices and provides a descriptive account of how the exploration of such cases informed/aligned with the design and implementation of the self-archiving toolkit developed in the course of this study. The subjects of these cases ranged from less experienced researchers (undergraduate students), researchers involved in individual research projects (also involved in teaching research methods), to more experienced researchers who were familiar with the design and development of research archives. The section is structured as follows. Firstly, the different cases of research data organisation and archiving practices from the three groups of participants are described and summarised with the support of comparative tables. These tables compile both data organisation/archiving practices and the kinds of research materials produced/archived, by each group of participants. Lastly, the conclusions drawn from the explored cases are presented along with a table that summarises the implications that the research and archiving practices analysed had on the design of the self-archiving toolkit.

The interviews conducted with students (the first cohort, as introduced in Chapter 3) provided important insights into their practices and perceptions of research in the context of their small-scale research studies. Furthermore, these interviews provided a suitable setting to explore how the self-archiving toolkit might be of use to undergraduate students conducting research. A common issue that was identified was that students experienced more issues during the research design and writing-up stages of their projects rather than during analysis and interpretation. In terms of the data organisation practices of the participants, these cases provided rich descriptions of the processes that the students followed to organise and structure different kinds of data and how they analysed and interpreted that data during the writing-up stages of their projects. Table 5.5 summarises the data organisation processes that the students followed and the kinds of materials they produced.

Table 5.5: Student data organisation practices and materials

| Participant | Data Organisation description | Overview | Materials (format) |
|-------------|---|---|--|
| Student 1 | All materials, apart from the collected questionnaires, were kept electronically in a single-level, folder-based structure. The student developed a coding system for the analysis of the questionnaires that was colour-coded, and linked to specific questions within the questionnaire. For the qualitative interviews, all files were electronic and included analysed transcripts and audio files. Additionally, the student developed a 'themes' document listing the themes explored in the interviews. For the dissertation, the student did not use version-control, or separate documents for individual sections of the dissertation, but rather, it was developed as a single Word document. | <ul style="list-style-type: none"> • Single-level, folder-based (electronic) | <ul style="list-style-type: none"> • Interview schedule (MS Word) • Consent form (MS Word) • Questionnaire results (MS Excel) • Analysed interview files (MS Word) • 'Themes' document (MS Word) • Dissertation (MS Word) |
| Student 2 | All materials were kept electronically in a folder-based structure. The organisation of this folder was based on the different sections of the dissertation (introduction, literature review, methodology, discussion, and conclusion) and included a 'resources' folder for collected data, which was organised by type of material (audio files, images, and transcripts). For the literature review, the student produced a summary document for each article consulted; and for those articles in electronic form, the student highlighted relevant information that was used in the literature review. For data analysis, the student produced tables which included the themes analysed in the data with relevant quotes. The dissertation was developed as individual documents, one per section, which were then incorporated into a single, final dissertation document. | <ul style="list-style-type: none"> • Multiple-level, folder-based structure (electronic) | <ul style="list-style-type: none"> • Journal articles (PDF) • Consent forms (MS Word) • Cue cards (images) • Focus group transcripts (MS Word) • 'Codes' document (MS Word) • Analysis tables (MS Word) • Dissertation section drafts (MS Word) • Dissertation (MS Word) |

Table 5.5: (continued)

| Participant | Data Organisation description | Overview | Materials (format) |
|-------------|--|--|---|
| Student 3 | Most materials were paper-based, with the exception of the dissertation and the analytical materials produced. For the literature review, printouts of the articles consulted were kept in folders with dividers and classified by theme and subject. The materials produced by the student were organised by codes and kept in another folder. For analysis, electronic transcripts were produced from the interviews, with that raw transcripts kept separate from those that were analysed. The student produced a number of Excel files with the results of the questionnaires, containing graphs and charts. Additionally, a table summarising the results of the questionnaires was incorporated into the dissertation. Lastly, the dissertation was developed as a single document rather than as separate documents. | <ul style="list-style-type: none"> • Single-level, folder-based (electronic) • Folder-based, organised by themes/codes (hard copy) | <ul style="list-style-type: none"> • Journal articles (PDF) • Consent forms (MS Word) • Questionnaire results (MS Excel and Word) • Transcripts (MS Word) • Dissertation (MS Word) |
| Student 4 | Apart from the collected materials (the questionnaires), all their materials were stored electronically, in a single folder. Most of the articles consulted were electronic versions from online journals, which were not organised in any particular way. The student developed a coding system for analysing the questionnaires, which was kept in Excel files, along with the results of the questionnaires, with one Excel file per cohort of participants. Lastly, the dissertation was developed as a single document rather than as separate documents. | <ul style="list-style-type: none"> • Single-level, folder-based (electronic) | <ul style="list-style-type: none"> • Journal articles (PDF) • Consent forms (MS Word) • Questionnaire results (MS Excel) • Questionnaire coding system (MS Excel) • Dissertation (MS Word) |

The general linearity of the development of their projects, which was introduced earlier in the chapter, was due to the structure of the research module itself. For most of the students, the actual process of managing the different kinds of data generated during their empirical work also followed a linear approach. This approach consisted of a number of steps in which different data were produced and incorporated into the final product, namely the dissertation. Additionally, one or more iterations of individual steps (analysis, interpretation and dissertation writing) were required within this linear process. The various steps of this process can be summarised as follows:

- *Literature Review.* Students consulted a set of articles during the literature exploration, most of which were sourced from online journals. These articles were organised by topic or subject and stored as both electronic and hard copies. Additionally, some of the students

produced summaries of these articles and highlighted relevant information to assist in subsequent dissertation writing.

- *Data Collection.* Some students chose a single data collection method (either through surveys or focus groups). Those who used mixed-method approaches conducted surveys first, followed by interviews. The data collected from surveys was paper-based, with none of the students keeping electronic copies. In contrast, interview data was kept electronically.
- *Data Analysis.* Students who used a single data collection method, produced all of the analytical data after data collection. In contrast, students who used mixed-methods performed further iteration of both the research design and data collection stages after a preliminary analysis had been made of the data collected from surveys.
- *Dissertation writing.* When writing their dissertation, the majority of the students worked on and stored their dissertation in a single file. Only Student 4 maintained separate files for each section of their dissertation, which were then integrated to produce a final, single document.

For the most part, the different practices of the students aligned well with the data organisation processes required of users of the self-archiving toolkit, especially those concerned with the organisation and documentation of the collected/generated data. Firstly, the toolkit's data model was designed to be flexible in terms of how the user can organise their data, so that practices such as the ones described here could be supported. The toolkit also supports the two data organisation approaches used by the students, which are shown in Table 5.5 and can be summarised as: 1) all the data is grouped together (similar to a single-level folder structure); 2) the data is classified according to the stage of the process. Additionally, the toolkit allows for the organisation and documentation of the data to be performed at any stage of the research process.

More importantly, the high degree of similarity between the current data organisation practices of the students and those that would be required to use the toolkit, means that issues often associated with the introduction of new technologies are kept to a minimum. Such potential issues might include changes in practices, or the efforts required to learn and make use of new tools. For example, the data of Student 3 was mostly paper-based and organised by theme, subject and a coding scheme developed by the student. In this case, the introduction of digital archiving activities would clearly imply a change to their practices. However, they could still organise their data according to themes or coding schemes plus they would gain the ability to be able to visualise how their data has been organised.

Lastly, the case of those students who chose mixed-methods were of particular interest due to the additional iteration of research instrument design, data collection and then analysis. This process of analysing data so as to inform the design of a second data collection instrument was

often either not described in the final dissertation, nor included in the appendices that accompany the dissertation. The ability to document this iterative approach is an important feature that was incorporated into the design of the toolkit so that the research design for a particular data collection instrument can be described, related to the exemplary data which informed its design, and then visualised together, with a description of the process that was followed. While the students did not show any initial inclination towards archiving, the value of having future access to their research instruments, data or written outcomes for future professional development or inquiry was recognised by all. At present, the expected outcomes of the research module in which the students participate means that their projects do not lend themselves to being archived. The student projects' output is a final dissertation in which descriptions, if any, of the processes followed by the students and the exemplary data that they collected, are included in appendices that the student submits in addition to the dissertation itself. Additionally, as described in Section 5.1, students were shown exemplary dissertations from previous years. Apart from the students being able to use the exemplary dissertations as a guideline for how to organise their own dissertations, this activity provided them with very little insight into the research processes that had been followed by previous students. A better approach might be to have students produce small sets of materials accompanied by descriptions of the processes they followed and visualised in ways that show how the different materials are related, so as to help future students to gain a better understanding of the processes involved in producing a dissertation.

Sets of exemplary materials which are currently used to help students gain a better understanding of research, or exemplary collected data that are currently included in the appendices of a student's dissertation could be enriched in the following ways by using the toolkit:

- methodological information and documentation could be added to the materials collected and produced by using the standard vocabularies included in the toolkit,
- the different kinds of data could be organised according to the different stages of the research processes, and users could identify “what data was generated in what stage”, and
- basic analytical data, and a description of the analytical frameworks used, could be included with the exemplary materials.

With regards to the researchers cohort, some of their research projects did not naturally lend themselves to archiving for the purpose of sharing. In these cases, this was due mostly to ethical and confidentiality issues. However, it was recognised that some form of archiving could have potential benefit, either for the researchers' personal re-use (in the case of Researchers 2 and 4), or for sharing with others (Researchers 1 and 3). Topics or themes that emerged in the course of the original research, but that were, at the time, deemed to be out of scope, could be explored at a later stage. Table 5.6 summarises the data organisation processes followed by the researcher participants, and the kinds of materials they produced.

Table 5.6: Researcher data organisation practices and materials

| Participant | Data Organisation description | Overview | Materials (format) |
|--------------|---|---|---|
| Researcher 1 | The activities undertaken by Researcher 1 are often process-led. This involves initial discussions about the research and documentation of activities that the team is undertaking. The data collected/produced is organised by the type of activity and presented in a blog site with both still and moving images accompanied by text and commentary. The primary data is stored mostly electronically, and normally each data file is labelled accordingly, e.g. 'analysis-of-interview-XX'. Additionally, text files with written decoding of the primary data are produced. The relationship between analytical data and raw data is recorded by using appropriate labelling of the materials and also owing to the way they are stored. | <ul style="list-style-type: none"> • Multiple-level, folder-based (electronic) | <ul style="list-style-type: none"> • Visual data (moving and still images) • Summaries, descriptions of visual data (text) • Audio-recorded interviews • Analysed interview files (MS Word) • Drafts of working papers |
| Researcher 2 | Researcher 2 employs phenomenological, narrative inquiry research approaches. The data collection tools are narrative writing tasks and qualitative interviewing. The collected data is stored both electronically and as hard copies. Interviews are audio-recorded and transcripts are produced electronically. The data gathered from participants is kept as hard copies. The main research instrument used is reflective writing and the analysis process takes place mostly in an abstract form: reflective, interpretive process in which the information is synthesised and then reflected in the researcher's own writing. | <ul style="list-style-type: none"> • Multiple-level, folder-based structure (electronic) | <ul style="list-style-type: none"> • Narrative texts from participants (paper-based) • Interview audio files • Interview transcripts (MS Word) • Analytic documents (MS Word) • Drafts of working papers |

Table 5.6: (continued)

| Participant | Data Organisation description | Overview | Materials (format) |
|--------------|---|---|---|
| Researcher 3 | <p>Researcher 3 uses participatory approaches, in which data collection mainly involves collaborative discussions with participants and the production of 'something' (e.g. software tools) rather than the researcher collecting data. Additionally, data is interpreted in a participatory way rather the researcher being the sole interpreter of the data. The central material of data organisation is the researcher's reflective diary. The primary data is then organised a very structured way, e.g. files that are produced are labelled by participant name and stored in folders according to event type, or research setting. The research diary is also used to keep track of links between the researcher's descriptions of activities and the data gathered during those activities. Lastly, the sorts of materials produced after analysis include publications that draw on case studies, which present summarised descriptions of what took place in the settings, accompanied by selected verbatim quotes, and diagrams or relevant images.</p> | <ul style="list-style-type: none"> Folder-based, organised by research setting/type of activity (electronically) | <ul style="list-style-type: none"> Research diary (web-based) Interview audio files Interview transcripts (MS Word) Case studies (MS Word and web-based) Drafts of working papers |
| Researcher 4 | <p>Researcher 4 employed an ethnographic research approach for their doctoral studies, which involved participant observation, qualitative questionnaires and interviews. The data gathered was organised by theme and stored in file boxes, into which note cards with relevant quotes or analytical notes were added. Data collection and analysis was performed over multiple iterations. The researcher's field notes included not only descriptions of what was observed but also preliminary analysis. The researcher's diary also held descriptions of the events attended and reflections on preliminary analysis in order to facilitate understanding of what had been observed. The researcher structured their writing in multiple separate sections, which were then integrated into the final dissertation.</p> | <ul style="list-style-type: none"> Multiple-level, folder-based (paper-based), organised by dissertation chapter | <ul style="list-style-type: none"> Research diary (electronic) Observation notes (hard copy and electronic) Interview audio files Interview transcripts (MS Word) Dissertation sections drafts (MS Word) Dissertation (MS Word) |

Most of the participants organised their data in structured ways, similar to the students. The

data generated by researcher participants was organised into folders (electronic or paper-based) according to research stage (e.g. data collection, analysis), subject or theme, or type of event (e.g. focus groups, interviews). In contrast, the type of research in which Researcher 2 (described in Section 5.2.2) is involved often leads to highly interpretive data, and its main outputs are publications. The practices of this researcher are challenging for data archiving, and sharing, in terms of the provision of contextual information and the presentation of collected data due to ethical and confidentiality issues. Unlike the other participants, their analysis process was described as being abstract, rather than a technical process in which a set of materials is produced. That is, the meaning extracted from the data collected is co-constructed along with the researcher's own experiences, and then reflected in their writing. In this case, the benefits of archiving that were perceived by this participant extended only to the presentation of carefully selected exemplary materials in support of research-oriented teaching activities.

The cases in which participants used research diaries were particularly interesting in relation to the toolkit. Research diaries included descriptions of the research settings, which were accompanied with relevant, selected primary data. Those research diaries that were web-based, that is a VRE was used, included electronic links to interpreted and primary data. This was perceived by the participants as extremely important not only for their own benefit, but for other researchers so that they could gain a better understanding of the research that took place. In these cases, the research environment provided a linear navigation through a set of web pages that included interpreted data with hyperlinks to accompanying raw data. Additionally, Researcher 3 selected and prepared small sets of exemplary research data, in the form of case studies, to support presentations of their research, or to be shared with others. Such presentations are supported by and can be enhanced by use of the self-archiving toolkit. The toolkit's use of semantic technologies, particularly the storage of relationship information by using relationship ontologies, facilitates the linking of analysed data with their original sources in ways that are machine-processable. As introduced in Chapter 2, this is key to enable data sharing. At the same time, the toolkit's visualisation interfaces allow for an exploratory navigation of project data, rather than a linear or more structured navigation. These visualisation interfaces (faceted browser and graph explorer) were described in detail in Chapter 4. These scenarios have highlighted the importance of incorporating into the toolkit, mechanisms to support contextual documentation, linking interpreted data to primary data and visualising how different data are related. At the same time, the toolkit was designed to be flexible in order to support those practices that are less structured in terms of the types of data that are produced.

To conclude, the main goal of the self-archiving toolkit is to provide a means by which students and researchers alike can create data collections in a simple manner and by following processes with which they are already familiar. In this respect, the design approach followed proved useful in the development of a self-archiving toolkit, underpinned by a data model and archiving processes, which supports more traditional archiving practices of research archivists as well as those of researchers with varying levels of expertise, such as the ones who participated in this

study. The initial design of the data model, and the archiving processes, was informed by current practices of qualitative research archives (as introduced in Chapter 3 and described in Chapter 4). This contributed to the design of a flexible data model which supports: a) a wide range of different research materials, b) management of basic analytical information, and c) an explicit definition of how the different research materials relate to each other. However, the toolkit's initial model of archiving required users to perform activities such as the creation and storage of research datasets almost independently from their research activities. That is, while data organisation and documentation activities could be performed at any stage of the research process, the actual creation of an archive had to be performed in the final stages of the research rather than in a progressive manner, throughout the research process. This process is summarised as follows:

1. Definition of the collection structure. The structure of the research archive, and its associated spreadsheet templates for data description and documentation can be designed and extended at any stage of the research process.
2. Data documentation. The data spreadsheets, both to document and add contextual information for both the research study and qualitative materials can be filled in or modified progressively along with research activities.
3. Archive creation. The archiving process consists of a number of operations that are performed once, that is, once the selected materials are stored in the digital repository with the support of the toolkit the contents of the collection cannot be modified or extended.

The different empirical cases outlined in this section provided, firstly suitable tests, of both the data model and archiving process that underpin the self-archiving toolkit. Secondly, the case studies provided detailed descriptions of research and archiving practices, as well as the kinds of materials produced/generated in the course of the explored research projects. The analysis of the kinds of materials generated in the course of research projects served as a test of the toolkit's data model in order to ensure that the documentation of such kinds of materials was fully supported. Table 5.7 compiles those aspects from the analysis of research practices that informed subsequent enhancement/refinement of different elements of the self-archiving toolkit, i.e. the archiving processes that underpin the toolkit, as well as functionalities that were incorporated into the visualisation tools. These aspects are organised by group of cases (cohort of participants).

Table 5.7: Summary of the contribution of the empirical cases to the toolkit design and development

| Cases | Design and Development Input |
|--------------------|---|
| Students | <ul style="list-style-type: none"> • 'Work-in-progress' data collections: very structured data organisation and archiving approach, which often helped with data analysis, interpretation and dissertation writing. • Final compilation of supplementary materials: less structured approach in which data were organised and documented at the final stage of the project. |
| Researchers | <ul style="list-style-type: none"> • 'Work in progress' data collections, in which research documentation is accompanied by relevant, selected subsets of primary data. • Exploratory navigation of research project data. • Presentation of exemplary selected materials to support the presentation of research outputs. |
| Senior Researchers | <ul style="list-style-type: none"> • Archiving process intertwined with research practices and consisting of a number of iterations in which different research materials are added incrementally. • Exploration of the kinds of data that can be visualised together and how they can be linked to each other. • Ability to keep track of the archive's search results so that subsequent archive visitors can explore materials previously selected by others. |

The creation of a final archive as the 'final step' once the investigation is complete was recognised as useful across all the different cases. For this archiving model, the central element would be a publication, or interpretive writing accompanied by contextual information at study level (research design, methodology, and data collection instruments) as well as selected supporting materials. In this respect, the original archiving model implemented in the toolkit aligned well with current research practices. However, across the three different empirical cases, the ability to create pilot collections or subsets of qualitative materials was identified as a key requirement and an aspect that needed to be fully supported. The participants perceived data organisation activities and the creation of 'work in progress' data collections as highly important, and as mechanisms to support them with the research process: the exploration of their own archived materials helped them with analysis and interpretation of data collected/produced. Additionally, these archived materials could also be used to support the presentation of preliminary research outputs. In order to support these archiving scenarios, the toolkit applications concerned with the creation and storage of qualitative materials were adapted so that they incorporated mechanisms to provide support for a more flexible archiving model, in which the creation of subsets or pilot collections can be performed progressively, and in parallel with research activities. This way the generated data can be structured and organised by the researcher alongside the steps followed during the study, thus making it possible to document and relate data to relevant, previously produced materials. Therefore, the archival process becomes an integral part of the research process rather than as a stand-alone activity.

Notes

- ¹Students from the Primary Education degree who develop school-based research projects. They undertake course modules in action research and inquiry-based methodologies to support them in research tasks, ranging from preparing ethics protocols to grasping the different methodological paradigms underpinning school-based research.
- ²This introductory module (Research methods in Tourism and Events) is included in the second year of the Events Management degree. A similar module is taught in the different specialities of the education degrees: Early Childhood Studies, Education Studies and Early Years, Education Studies and PE, Education Studies with Special and Inclusive Needs.
- ³An important feature of one of the versions of the case study, CD-ROM based, was the ability to keep track of how students explored materials. This works in a similar fashion to the history feature of an internet browser. The software used, namely Supercard, is able to keep track of the sequence in which students accessed materials in a particular iteration through the archive, as well as printing the visited sections.
- ⁴BrainAble, FP7 (Seventh Framework Programme for research and technology development) funded project. Project website: <http://www.brainable.org>
- ⁵Smart homes can most easily be described as "a collective term for information and communication technology in homes where components communicate through a local network" (Cheek, 2005)

CONCLUSIONS

The design and implementation of the self-archiving toolkit developed in this thesis has taken into account various aspects from the areas of digital archiving, data documentation and exchange, and qualitative data archives. The toolkit's supporting technologies, namely Fedora repository and Mulgara database, are open-source, widely adopted and actively developed within the digital archiving community. These are important features since they facilitate integration with other technologies and provide the possibility of extending the toolkit's current functionalities. Data documentation standards are a central component within information systems: they provide a means to accurately describe data in digital forms, and this in turn facilitates their dissemination and discovery across the web, and thereby enables data sharing and exchange between different information systems. The literature review highlighted that, while a number of widely adopted standards and documentation vocabularies exist for describing quantitative data, these are lacking when it comes to describing qualitative data. The work on the [DDI Qualitative Data Exchange WG](#), and the exploration of a set of use cases from various research archives (presented in the [WG](#)) supported the findings of the literature exploration and also provided key input for the design of the toolkit's data model. The key issue encountered within most of the use cases explored was the limited support of existing archiving systems to accurately describe qualitative data, and more importantly, the relationships between the generated materials for a qualitative study. The underpinning data models for those particular archives provide very detailed study-level information, and file listings for the materials included in a study. However, they do not include mechanisms for expressing the different relationships between the qualitative materials, nor do they handle the inclusion of more specific contextual information associated with data collection. The toolkit developed in this study attempts to provide a solution to the above-mentioned issues by implementing an underlying data model that firstly, includes support for managing basic analytical units such as categories and codes that can be attached to any data file included in the collection. Secondly, the data model uses a set of common relationships to express how the different elements within a collection relate to each other, while at

the same time, maintaining descriptive documentation at the study level.

The qualitative data exchange model, which was the main output of the working group, was introduced in Chapter 2 and described in detail in Chapter 4. This model is underpinned by research processes such as those explored with the different stakeholders involved in the [WG](#) rather than sole practices of archivists. That is to say, it combines research methodological aspects with existing archiving practices, and enables contextual information to be associated with archived materials. This way, more complete datasets, accompanied by research results and interpretation, can be presented. Additionally, this qualitative model attempts to represent generic research processes within qualitative studies, supporting complex scenarios such as large-scale research projects in which mixed methods are used and collaboration might be a requirement. The complexity of the processes of data documentation and research description that are required for the use of such a model can be quite challenging for researchers, especially in scenarios such as the ones explored in the research settings of this study: students conducting research, or researchers involved in small-scale or individual research projects. Firstly, detailed documentation of the data and research processes introduces a further level of complexity and requires additional efforts and time-consuming tasks on the part of the researcher. Secondly, while the introduction of archiving elements within researchers' practices is essential for data sharing and re-use, it can also introduce procedures that may potentially require a change to the research practices themselves. Researchers may have to perform tasks in a different manner. In contrast, the self-archiving toolkit provides researchers with a basic framework to create and store small-scale research data collections, without the need to change their research practices. Users of the self-archiving toolkit can archive small-scale qualitative data collections by simply:

- organising their data in similar ways to their actual practices, such as the ones explored in the interviews with students and researchers, and
- creating data spreadsheets that contain resource descriptions annotated using appropriate metadata terms which provide sufficient contextual information and also facilitate data exchange and re-use.

Although the nature of the research projects that have been the subject of this study, and the use of [CAQDAS](#) tools with analysis purposes was not a requirement owing to the scale of their projects, the inclusion of mechanisms in the toolkit to support further analysis has proven useful and a key feature for facilitating data sharing and re-use. In addition to documenting and storing project data in the repository system, the entire collection is represented as a [QuDEx XML](#) file that is stored internally in the archiving system. This is highly important for the purpose of data sharing and re-use since it allows for re-analysis or further analysis of existing data.

The main goal is to provide a means to create data collections in simple ways and by following processes with which students and researchers are already familiar. In this respect, the model process of structuring and constructing collections with the support of the [QuDEx](#) toolkit aligns well with the kinds of research activities in which participating researchers and students

were involved. The toolkit supports the various research processes followed by the participants interviewed, ranging from more interpretive approaches such as phenomenology, or ethnography to participatory approaches, or case studies. For example, the students followed typically linear research processes in which generated data were organised according to the stage of the research process (e.g. research proposal, literature review, analysis, dissertation writing), and elements such as analytical materials were kept separate from those gathered during data collection. For those more experienced researchers involved in participatory approaches, the processes that were followed led to a variety of output(s), such as analysis of quantitative data and qualitative research which drew on small case studies. Those output(s) were described as *vignettes* that presented a summarised description of what took place in the different research settings, accompanied by selected verbatim quotes, as well as diagrams, or images related to what was produced. Those scenarios can also be supported by the toolkit. The functionalities of the toolkit that enable the explicit description of the relationships between the heterogeneous data produced, along with an integrated visualisation of those data, can enhance the representation of the kinds of collections created. Data can also be classified in multiple ways, e.g. by source (researcher, participant), by production stage (research design, data collection, analysis), annotated by themes, and related data linked so that they can be filtered and visualised together with the support of the 'Collection Explorer'. The participants who were interviewed could appreciate how archiving activities could support research design, data collection and analysis, writing and future re-use of their work. But what the digital repository architecture and the QuDEx toolkit also allow are different kinds of representations of research projects in which the final dissertation (student projects), or publication (researchers), is only one element. While it would be possible for subsequent visitors to archives generated by this toolkit to simply read project research outputs, or students' dissertations, they would also be able to explore the data on which such projects drew and to replicate, or develop, parts of a selected research project.

The process of building a digital archive of qualitative data when the aims include data re-use presents not only technical challenges but also barriers to its adoption by practitioners of particular research practices. The attitudes of the researchers interviewed were mainly positive to digital archiving as they perceived its potential benefits for the organisation and personal re-use of their data. However, some were reluctant to share their data with others, so that they could conduct secondary analysis. This reluctance mostly relates to the researchers' epistemologies and their own perceptions of qualitative research. The interviews with lecturers and researchers highlighted that the type of research undertaken - phenomenology, or ethnography vs. participatory approaches, or case-based research - is what determines their attitude towards data sharing and archiving. Researchers that were interested in phenomenological and other more interpretative approaches were, to a greater or lesser extent, opposed to the idea of making available their data for the purpose of secondary analysis. Their reluctance to share data was due mainly to the nature of the data itself. For example, it may not be appropriate to share highly sensitive data with a wider audience. Furthermore, for those researchers, the process of organising and documenting their data was described as being less structured than would be the case with other

research approaches. That is, analysis and interpretations occurred at the time of writing rather than in the form of analytical notes accompanying the data gathered, or descriptive summaries of the interpretive process.

In contrast, for researchers involved in participatory approaches, or case study research, the data and results of their research is of a less subjective nature. Additionally, their research practices involved quite structured data organisation processes, which combined highly contextual documentation of their data and presentation of their results with supporting data. Subsequently, these participants recognised the importance of presenting complete datasets, if the aim is data sharing, so that the risks of data misrepresentation, and research misinterpretation, are minimised. The above discussion serves to make clear that the design of archiving tools, and their underlying data models, must be informed by the existing practices of both archivists and researchers alike, while also taking into account the best ways to document qualitative data so as to address the above-mentioned issues. These two factors were at the forefront in the design of the self-archiving toolkit developed in this study. The toolkit supports an archiving process that can be performed in parallel with research activities, whereby the generated data is structured and organised by the researcher alongside the steps followed during the study. To be more precise, data can be documented and related to relevant, previously produced materials. The user can also create subsets or pilot collections progressively, alongside the research processes. Lastly, in those cases where the data organisation is less structured, the researcher could still create an archive with the support of the toolkit as ‘the final step’ once the investigation is complete. In this case the central element of the archive would be the publication, or interpretive writing, accompanied by contextual information at study level (research design, study methodology, and data collection instruments) as well as carefully selected supporting materials.

Whilst digital qualitative archiving for the purpose of secondary analysis has highlighted a number of issues, the benefits of using qualitative archives for teaching and learning purposes were recognised by all the participants interviewed. Complete, existing archived studies are mostly used with and by students at masters’ and doctoral level undertaking research methods courses and modules. In this situation, archives are used as secondary sources, to reflect on existing practices, or as a source for primary research where new questions can be explored, or new instruments developed. This was the case of the scenario presented in subsection 5.3.1 in which multimedia archived case studies were used to support masters’ level students undertaking an action research course. The archived multimedia case study, in which materials were selected and edited for their presentation, provided students with an environment to ‘do a case study’. The participant recognised that the use of the archived materials facilitated students’ reflection, and that they were able to formulate new questions and identify ways of contributing to the archive. These students had a more solid background in research approaches, and the processes involved in research more generally, as opposed to the undergraduate students who participated in this study. In the case of the latter, their understanding of research design and the processes involved in conducting a research study were emergent. The interviewed participants

that were involved in teaching research-oriented modules recognised the complexity of teaching these modules at undergraduate level. They were already using a wide variety of materials, ranging from previous years' dissertations to exemplary materials, either from their own research or with prepared materials from existing studies. The main purpose of using these kinds of materials was to support students' understanding of the underlying theoretical concepts behind different research methodologies as well as more practical aspects related to the development of their projects. While the participants identified the benefits of using exemplary case studies or small archives and realised how they could help students with research design, and data collection and analysis, they raised two important issues:

- While the use of previous years' dissertations proved useful to help students with the structure of their dissertation, it provided little insight into the processes involved in the research.
- Some students who were presented with sample research materials whose content did not directly relate to their topic of interest failed to fully appreciate the research processes involved.

In this respect, the self-archiving toolkit has the potential to support research-oriented teaching and learning activities. It may also provide partial solutions to some of the issues raised in relation to both the students' projects and the teaching and learning practices of those participants involved in undergraduate research modules. Two scenarios for the use of the toolkit to support such activities have been identified. The first of these involves students and teachers archiving their own materials with the support of the [QuDEx](#) toolkit. The tools allow students to organise the data associated with each stage of the research project, for example at the research design stage, during the literature review, data collection, analysis and dissertation writing. The analysed data can then be linked to either specific sections of the dissertation, or other relevant materials, such as analysis notes or literature. In this way, students would be able to track what data has been documented and stored at each stage of the project. Additionally, students' reflections and ideas about their inquiry could be combined with actual data generated during their research. With regards to teachers, the preparation of small archives covering research design, research questions, and collected materials linking to analysis, could help students gain a better understanding of the notion of research and the processes involved, especially if these materials are visualised in ways that show how they relate to each other. The second scenario builds on the potential for the development of shared archives. If sets of students' collections were made available within a single archive, the Exhibit-based 'Collections Explorer' in particular, would provide flexible ways to search, filter and visualise existing collections, allowing teachers and students not only to read finished dissertations, but to see how other projects had been designed and undertaken, how particular methods had been used, how ethical protocols had been framed and how data had been presented. Collections of 'supporting materials', rather than being consigned to the appendices of dissertations, or disappearing from view, would contribute to a continuing and cumulative process.

Approaches to exploring archived materials, whether sourced from students' projects or prepared by the teachers, require teachers' mediation to support the learning activities. The creation of exemplary collections, with the support of the self-archiving toolkit has to be embedded within teaching practices in order to be useful and to add pedagogical value. Therefore, for archiving to be a pedagogical practice, it has to be integrated into the work of both teachers and students, and be a focus of the pedagogical discourse. In this respect, getting teachers self-archiving their research and modelling the research approach represents an important prerequisite and could serve as a 'model' for students developing their own projects. Students need to be walked through existing archives in order to be able to understand not only the process but the epistemologies involved in the knowledge and solution construction.

Self-archiving tools that draw upon semantic web technologies and archiving can help students and researchers, more generally, to expand the output of their research projects from a single dissertation/publication to the production of shareable knowledge on which they and others can build. The semantic web has, at its core, a commitment to escaping the constraints of 'documents', prioritising instead the exchange of information in order to facilitate knowledge creation and sharing, with the support of tools such as the self-archiving toolkit described here. There are opportunities for students to become producers of knowledge that go beyond the current constraints of assessed writing; and for researchers to contribute to the research community, not only with publications as research outputs but also with well organised, well documented, preserved and shared data, thereby increasing the opportunities for learning and innovation.

While the technological developments, frameworks and standards outlined throughout the thesis are key in relation to the development and exposure of data archives and have direct impact in the discovery, search and retrieval of available datasets, many individuals and small research archives face numerous limitations in terms of having access to and using appropriate archiving infrastructures and resources. Amongst these, one of the most critical is providing a means by which individuals can structure, describe and add their own collections to existing archives. In this respect, the toolkit implemented in this study provides researchers and students with an accessible set of tools supporting such processes, which are enhanced by combining linked data technologies with well-known documentation standards. This combination allows the researcher to not only present a collection of materials but also express how the different data relates, keep track of the processes followed during the research, and more importantly, share and re-use their data.

6.1 Suggestions for Further Work

There are a number of technological areas in which the first version of the toolkit could be enhanced in the future. At present, the toolkit is suitable for small-scale research projects, in which the number of elements included in a research collection is restricted. The collection size limitation concerns the 'Collection Explorer' application in that only one collection is displayed

at a time for its exploration rather than all the collections in the repository being displayed in an integrated manner. This issue is due to a limitation of the visualisation tools used, the Exhibit framework, which can handle the display of up to a thousand items owing to its underlying technologies being client-side. However, such a limitation could be solved in the future by migrating to the new version of the framework (Exhibit v3¹) which introduces the use of [RDF](#) databases to handle bigger server-based data visualisation rather than relying on in-browser data display. Another aspect of the toolkit that could be enhanced is the level of data documentation. The toolkit provides a basic set of standard metadata terms, facilitating data re-use, dissemination and discovery, for example, basic terms from [DDI2](#) or Dublin Core terms are used to describe qualitative data and can be shared owing to the [RDF](#) database underpinning the toolkit. However, the level of detail for describing is relatively low in comparison with complex data models, such as the Qualitative Data Exchange schema that was described in Chapter 4, which incorporates a larger set of terms for the description of research-related aspects, such as research data collection instruments, analysis process, or research events, in great level of detail. The rationale for including a smaller set of documentation terms has been to make the toolkit more accessible, and simplifying the associated archiving processes. There had to be a balance between the complexity of the documentation and archiving processes, and the levels of detail when describing, documenting the research data.

Lastly, a key area that has the potential for further work is concerned with data analysis capabilities. It may be possible to integrate existing annotation tools with the toolkit's current archiving model by extending it to include support for handling segment-level data (and their documentation) along with a set of annotations (and appropriate relationships) that can be attached to segment-level data. This first version of the toolkit provides basic support for analysis tasks. [CAQDAS](#)-based analytical terms such as categories, codes and memos are included in the data model, and in the [QuDEx](#) instance produced per collection. However, they can only be attached or related to *whole* documents or materials, e.g. analysis categories or codes can be attached to an interview transcript but not to segments of it. The current support for analytical tasks might not be sufficient to accomplish complex analysis tasks. The rationale for this design decision was firstly, the lack of open-source annotation tools, and documentation standards for qualitative analytical data, in the early stages of the project. Secondly, existing specialist software for qualitative data analysis ([CAQDAS](#)) could be used in those cases where supporting complex data analysis tasks was a requirement. Recent exploration of existing open-source annotation tools has highlighted a number of new developments that could potentially be used to extend the self-archiving toolkit. Particularly of interest is a browser-based online tool using linked data approaches, namely [AutoKitty](#), which was the main output of a Joint Information Systems Committee ([JISC](#)) funded project as part of the [OER Rapid Innovation Strand](#) ([Litherland et al., 2012](#)). The tool allows users to non-destructively aggregate a variety of open online content into 'presentations', or structured narratives centred around playback of one piece of audiovisual media. However, it could be easily extended so that it could include support for adding analytical annotations to research data. The integration of such tools, alongside including support for

segment management within the archiving model, would provide mechanisms for performing enhanced analytical tasks within archiving. This may be a potential benefit for students using the tools in that they could be trained for analytical tasks, such as initial coding, annotation of audio/video data and so forth. One of the issues that participants highlighted with respect to the use of CAQDAS packages to support analysis activities was the complexity of such tools. While most of the participants were familiar with CAQDAS tools, and some of them had used them for their research, most of them recognised that other tools were more suitable to their analytical activities, especially when developing small-scale research projects. In this respect, more accessible and easy-to-use tools could lower the bar to their adoption, especially by less experienced researchers, such as students. Moreover, archived data presentation with the support of the toolkit could be enhanced to incorporate more detailed analytical documentation. Additionally, the combined use of such annotation tools with archiving ones could provide students with an initial training that would serve as the basis for future use of more complex QDA tools, such as CAQDAS packages.

Self-archiving tools drawing on semantic web technologies and appropriate data documentation frameworks, and designed following principles that align well with existing practices of both archivists and researchers alike can lower the bar to their adoption by users less familiar with research archiving, such as the ones explored in the research settings of this study. Moreover, such tools can help to make real the notion of 'students as producers', and to support the inclusion in the curriculum of more inquiry-oriented teaching and learning activities. The positioning of the archival process as an essential element, not only of student inquiry, but also of the pedagogical discourses that it accompanies, presents a range of challenges. The immediate follow-on from this study will be to put the self-archiving toolkit to use within research settings such as the Independent Project module of LJMU. In doing so, significant work will be required in the first stages of the archiving process supported by the toolkit, that is, getting students to articulate their research designs and how their data, and the research tasks performed, might need to be described. This could be addressed by including concepts of data documentation and archiving of research data, as well as their implications on research ethics in undergraduate introductory modules on research methods. While this approach is similar to research archives' existing approaches to more advanced training on data management of research data, it will have to be adapted to the context of the development of small-scale research projects. Additionally, despite the fact that the toolkit has been designed to be sufficiently accessible to users with varying levels of research experience, training sessions to support the use of the toolkit may be required. In such sessions, teachers and students will be supported in creating sample archives, following exactly the processes that are required by users of the self-archiving toolkit, and which could serve as 'pilot collections' prior to the development of their research projects. Lastly, an evaluation of the use of the toolkit in such environments, following participatory approaches involving teachers and students, could help to identify aspects of the toolkit that require enhancement, or new features to be incorporated into subsequent versions of the toolkit.

Treating the archiving process, with the support of the toolkit, simply as a ‘reporting’ activity or as a ‘deposit’ of students’ and researchers’ data, could highly constrain important outcomes such as the engagement with existing research data or further development of existing studies. That is, users might never think of some of the data that has been collected, or how it might be re-used. However, the indications from this study suggest that a combination of appropriate underlying infrastructures, user-oriented tools and commitments to engage students/researchers in inquiry activities, supported by self-archiving tools, such as the toolkit developed in this study, make this a means by which students/researchers may be fully aware of the potentials and benefits of engaging with existing archived data.

Notes

¹<http://www.simile-widgets.org/exhibit3/>

APPENDICES

QUDEX REPOSITORY TOOLS v0.1 – TECHNICAL DOCUMENTATION

The QuDEX Repository is a toolkit designed to help users with a requirement for archiving, searching, and disseminating collections of non-quantitative data, regardless of the type of media. These tools provides researchers, such as research students, early career researchers, or more senior researchers with a means to organise, archive and present their research data. While the tools will work well with the files making up studies conducted with CAQDAS packages, it makes no assumptions about the software used to generate a collection. Many qualitative researchers do not use any dedicated software for marking up their qualitative data - support for research of this type is also provided.

QuDEX Repository is based on a number of existing open-source tools and open standards that are in common use in data archives and repositories. These include Fedora (a digital repository); Lucene (a popular search engine which has been integrated into Fedora); Mulgara (an RDF database used to publish 'linked data' as a SPARQL end-point); the Exhibit browser; the popular Data Documentation Initiative (DDI) standard (in its 2.1 (and earlier) version); Dublin Core (a popular metadata standard for citations and related metadata); and QuDEX (a standard published by the UK Data Archive for describing CAQDAS metadata and similar qualitative data collections). It also uses a number of popular RDF vocabularies for disseminating the collections via Mulgara.

A.1 General Overview

In its first release, QuDEX Repository is a basic set of tools, intended to meet the needs of data archives and libraries and small-scale projects which have no immediate solution for archiving, searching, and disseminating qualitative collections and studies. Main deliverables include:

- a spreadsheet-based tool for capturing metadata about qualitative studies and files, used for ingest into the Fedora repository,
- a tool for creating the collection structure in the repository, and uploading the all the relevant files associated with the collection,
- a tool for visualising and exploring the qualitative collections, displaying the metadata and associated resources via an online interface based on Exhibit framework. The collections' data are sourced directly from the [RDF](#) database, which had been populated with the data from the Fedora repository after ingest, and
- an automatically populated and managed Mulgara triplestore, mirroring the contents of the [Fedora](#) repository and exposing the contents as both [RDF](#) in a [SPARQL](#) end-point, and also as Exhibit-compatible JavaScript Object Notation ([JSON](#)).

With this basic set of tools, it is expected that archives, libraries, and small-scale research projects can begin to better manage, search, and disseminate their qualitative data collections. It is hoped that these tools will be extended in future, working on this basic framework. These tools are developed under the Apache License 2.0, entirely developed in Java language and using frameworks like Spring¹. They can be easily extended and adapted depending on different needs.

The implemented set of tools are online tools that need to be deployed in an application server like Apache Tomcat (Template builder tool, QuDEx Repository management application and Exhibit-based visualisation application). Furthermore, they rely on and communicate with two different external systems that need to be installed and configured before installing these tools. The two systems are Fedora Commons repository and Mulgara Semantic Triplestore . The general architecture of the systems and tools is shown more in detail in Figure A.1.

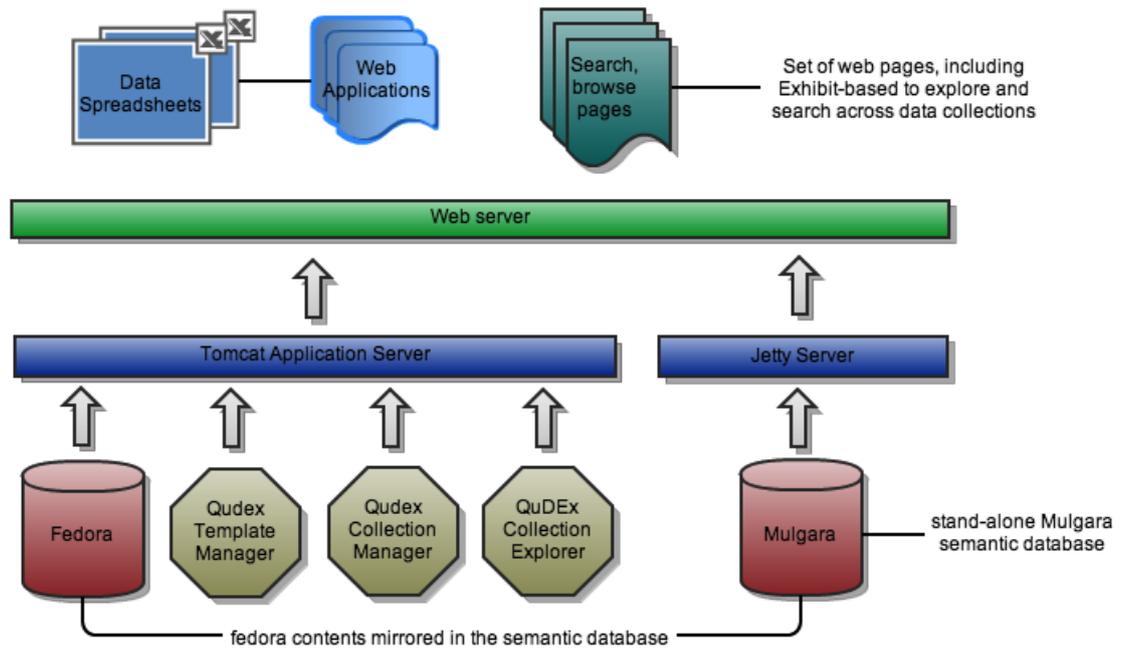


Figure A.1: General overview of tools' architecture

A.2 Technology Background

There are a large number of archival systems being developed to enable structuring, managing and storing large collections of heterogeneous data available in digital forms as well as facilitating access to those resources. Some of them provide a full solution to manage and store large collections of data, i.e. they implement complete and user friendly applications such as web and desktop data managers, advanced search interfaces and collection visualization tools. However, there are others that are community developed and their current implementation state is more experimental.

Despite the latter, these experimental systems provide flexible and configurable digital repositories, which are 'semantic web ready' by offering interfaces such as [OAI-PMH](#) or allowing close coupling with semantic triplestores that in turn can be accessed typically via [SPARQL](#) gateways. This is the case of [Fedora Commons](#) digital repository and that is why, after following evaluations of the functionalities of several digital repositories, it has been selected as the basis of the implemented [QuDEx Repository Tools](#). [Fedora](#) offers many useful features such as:

- A digital object model that allows objects to contain multiple data representations. This way images can be stored at different resolutions or in different formats.
- Support for metadata annotations under different schemas, so bibliographical records might use [Dublin Core](#), [MARC/MODS](#) records and other well-known schemas and even data already converted into semantic-web ready formats such as [RDF-XML](#) or [N3](#).

- Provide access to an external [RDF](#) database (Mulgara) that allows to store all the metadata annotations of the described resources so that they can be accessed and retrieved in combination with other resources coming from external sources.

Fedora also provides a powerful set of tools for data ingest, description and management, although they are more oriented to system administrators or technical personnel. Therefore data, and metadata associated with them, from the digital repository can be combined with data from other sources, rules and inferencing applied and query interfaces used to populate web interfaces or lightweight data visualization frameworks like Exhibit from the [SIMILE](#) Project. The [QuDEx](#) Tools rely on these functionalities and allow to use and combine resources relationship information (expressed using primarily Dublin Core, DDI2 and [QuDEx](#) schemas) with external data description information coming from other sources. This is performed by storing all the metadata records describing the annotated qualitative resources in Mulgara, which is closely coupled with Fedora repository.

A.3 Technical Documentation

The [QuDEx](#) Repository Tools are entirely developed using Java programming language and can be installed and executed in multiple platforms. They have been tested primarily in Mac OS X and Windows (XP and 7). They should also work in Unix and Linux based systems.

The toolkit is composed of a number of applications that implement the core functionalities, including the repository management logic and web interfaces. These applications rely on multiple independent Java libraries implementing separate groups of functionalities. The first set of libraries implements the functionalities related to connection and communication management with the digital repository (Fedora) and the [RDF](#) database (Mulgara). The second set implements all the classes required to manage both [QuDEx](#) and Fedora Object XML ([FOXML](#))² documents. Lastly, the third set of libraries implements additional metadata schemas management (DDI2, Dublin Core schemas). [Figure A.2](#) shows the groups of libraries (colour-coded in the figure) which comprise the whole system.

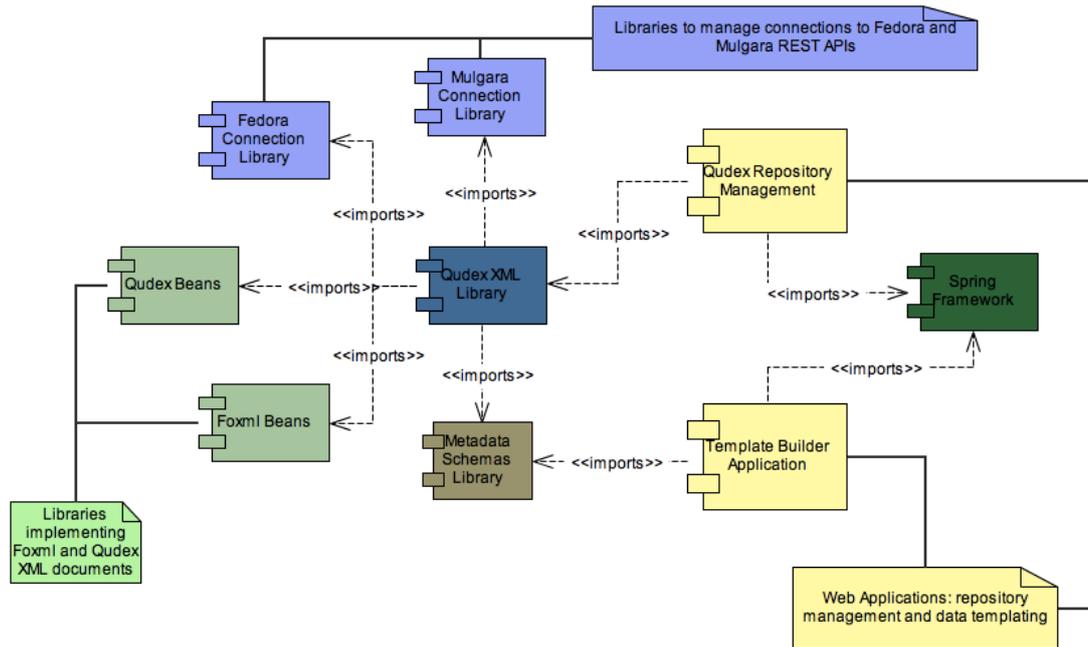


Figure A.2: Components diagram

A.3.1 Java Libraries

As described above, the system comprises multiple independent Java libraries grouped by functionality. The following groups are identified, which will be described in detail in the following sections and sub-sections:

- Set of libraries to manage connections and operations on Fedora repository and Mulgara triplestore (Fedora and Mulgara connection libraries in Figure A.2). The two systems provide web REST-like interfaces to allow access and perform different operations.
- Set of libraries implementing XML data management functionalities (FOXML Beans and QuDEX Beans in Figure A.2). The general systems work primarily with three different types of data: XML documents representing fedora objects (FOXML schema-based), XML documents representing QuDEX instances and finally, Excel or Comma Separated Values (CSV) spreadsheets to perform ingestion and update operations across the repository.
- Set of libraries implementing other Metadata schemas management (Metadata Schemas library in Figure A.2). The applications use metadata elements from the Dublin Core schema and DDI2 Lite. The management of these data is performed by a separate library.
- Library implementing core management of QuDEX Repository Toolkit (QuDEX XML library in Figure A.2). It uses the other groups of libraries to implement all the operations needed to perform collection creation, ingestion and updating in the repository.

The following sub-sections describe individually each of the above-mentioned groups of libraries. The documentation of each of the components includes general overview; description

of the main implemented functionalities and detailed diagramming using the standard Unified Modelling Language (UML)³ version 1.4. Additionally, each of the implemented libraries includes code documentation (Javadocs).

A.3.1.1 Fedora and Mulgara Connection Libraries

The repository tools interact both with Fedora repository and with Mulgara triplestore. Both systems provide a set of APIs to access and manage them from different applications. Mulgara provides a Java API and a web interface (REST-like service) to enable Create, Read, Update and Delete (CRUD) operations on the triplestore. In Fedora's case a HTTP REST interface is provided, which enables to perform CRUD operations and object relationships management.

With respect to the Fedora connection library, Fedora provides a Web Application Description Language (WADL) document of their REST HTTP interface. This has constituted the base of this Java library. Open source tools to extract and generate Java classes from a WADL document have been used to generate the base connection classes forming this library. A Java interface, implementing the most common operations for accessing and managing the repository via external applications, lies on top of these. There are two interfaces within the library: the first of these (FedoraOperationsAPIA) implements all the operations available in Fedora APIA (access API for read-only operations); the second interface (FedoraOperationsAPIM) implements all the operations available in Fedora APIM (management API, read/write operations).

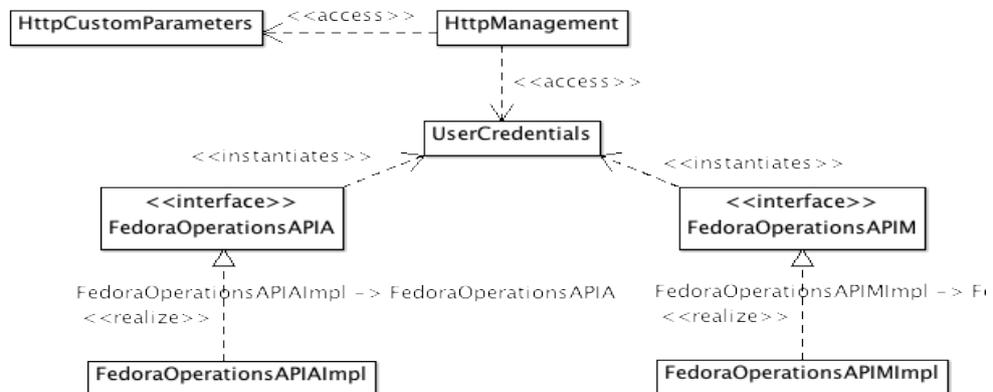


Figure A.3: Fedora Connection classes diagram

The library implements two interfaces and three auxiliary classes, instantiated by the main interfaces. Each of the interfaces provides access to a different set of fedora operations, as described in their APIs. Fedora APIM operations are CRUD operations in the repository that primarily enable to create, update and delete digital objects as well as managing their relationships by using the relationships operations, included in this API. Fedora APIA includes query and search operations over the repository. The list of operations available for each of the interfaces

is described in the following two sections.

FedoraOperationsAPIA Interface

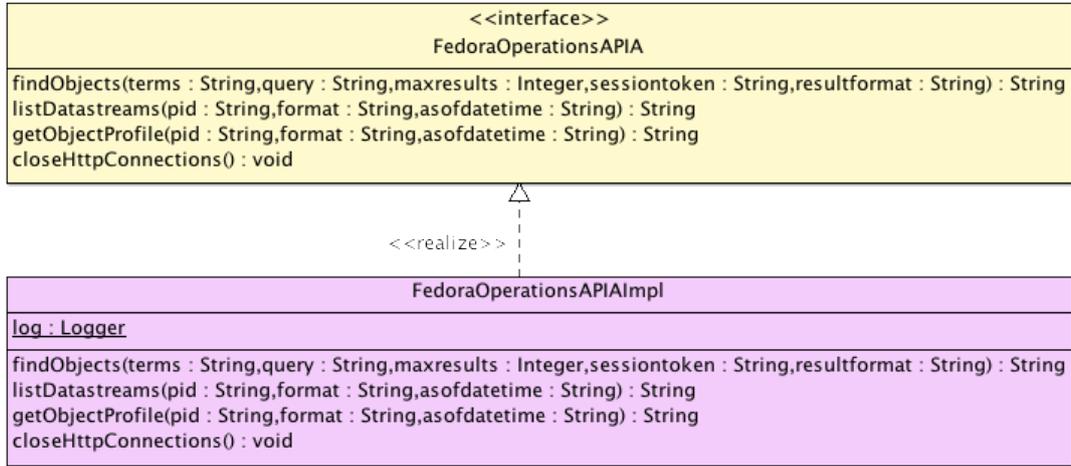


Figure A.4: FedoraOperationsAPIA classes diagram

The class presents the following operations:

- findObjects. Search across the repository to find digital objects given a query command.
- listDatastreams. List all the datastreams available for a given digital object
- getObjectProfile. Returns profile (XML or HTML formats) of a particular digital object

FedoraOperationsAPIM Interface



Figure A.5: FedoraOperationsAPIM classes diagram

The class presents the following operations, grouped by object type⁴:

Operations applied to objects

- getNextPID(). Obtain the next unique not assigned object identifier (pid).
- createObject(). Create a new object in the repository.
- deleteObject(). Delete an existing object from the repository.
- getObjectXML(). Get object profile information in XML format.
- modifyObject(). Modify an existing object in the repository.
- validate(). Validate FOXML document for a given object.

Operations applied to datastreams

- addDatastream(). Add a new datastream to an existing object in the repository.
- deleteDatastream(). Delete a given datastream from an existing object in the repository.
- modifyDatastream(). Modify a given datastream from an existing object in the repository.
- getDatastream(). Get the profile information for a given datastream of an existing object in the repository.
- getDatastreamContent(). Get the contents of a given datastream of an existing object in the repository.

Operations applied to relationships between objects

- addRelationship(). Add a new relationship between two objects in the repository.
- getRelationships(). Get all the relationships of a given object.
- purgeRelationship(). Delete a particular relationship from a given object in the repository.

In relation to the Mulgara Connection Library, Mulgara triplestore implements various mechanisms to provide external applications access to the database. The most useful ones are a set of java classes that implement connection and querying capabilities and a [HTTP REST](#)-like interface. The implemented Mulgara Connection library implements a set of interfaces to connect with the triplestore and relies in this [HTTP](#) interface.

This library provides one interface that implements a set of query functions to obtain data from the triplestore. It also implements a class with predefined queries that are used by the repository tools to obtain collections and files information ([QuDEx](#) collections) from the triplestore. The following classes diagram shows the most relevant classes implemented in the library.

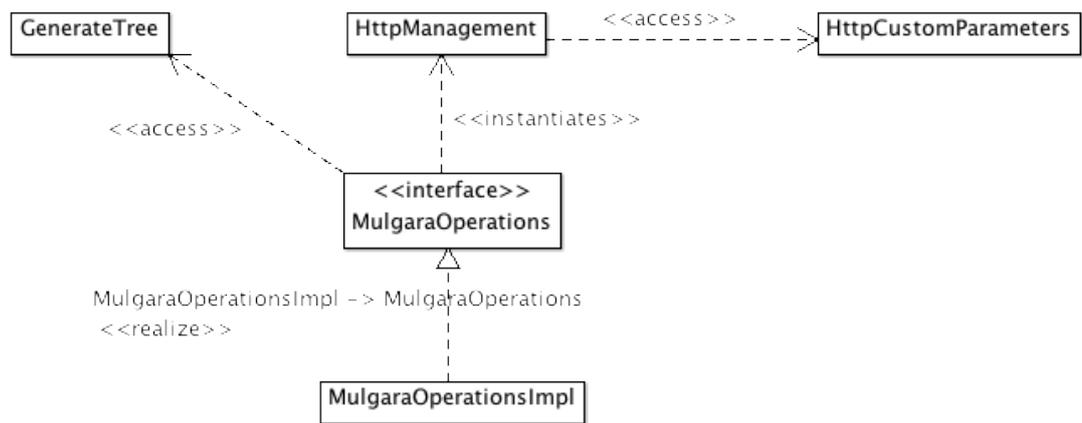


Figure A.6: Mulgara library main classes

The following section describes in details all the operations implemented in the main connection interface (MulgaraOperations).

MulgaraOperations Interface

This is the main interface implemented in the library. It implements a set of methods to connect - via [HTTP](#) - to mulgara triplestore and to query its database. The database contains information about the [QuDEx](#) collections stored in Fedora repository. While these methods are designed specifically for connecting to the Fedora-Mulgara graph (graph of the triplestore storing the information about the [QuDEx](#) collections in the repository), it also provides with regular GET/POST [HTTP](#) methods to query the repository via Java classes.

Methods for querying repository collections

- `getRootCollections()`. This method queries the triplestore to obtain a list of all the root collections in the repository. A root collection is container collection that has no parent.
- `getRootCollectionsTitle()`. This method queries the triplestore to obtain a list of pairs containing root collection identifier and its Dublin Core title annotation.
- `getNumberOfChildren()`. It returns a list with the pids of all the children of a given collection.
- `getCollectionInfo()`. Given an object PID it returns all its metadata records.
- `getResourcesInfo()`. When displaying collection information a list of the most recent resources associated with it is displayed. These information is obtained by invoking this method.
- `getCollectionChildren()`. Returns all the children identifiers associated with a sub-collection (it is not a container collection or 'root' collection).

- `getCollectionChildrenTitle()`. Returns all the children `dc:title` values associated with a sub-collection (it is not a container collection or 'root' collection).
- `getCollectionChildrenTitleQuDEx()`. This function is the same as 'getCollectionChildrenTitle' but applies specifically to QuDEx collections; the collections membership is specified using different types of Fedora relationships, therefore the queries differ.
- `hasMembers()`. Returns true or false depending of whether the given collections has members or not.
- `hasMembersQuDEx()`. The same as 'hasMembers' but specifically applied to QuDEx collections.
- `getRecentCollections()`. Returns the identifiers of the most recent collections stored in the repository/
- `getRecentCollectionsTitle()`. Returns pairs of identifiers-titles of the most recent collections stored in the repository.
- `getRecentResources()`. Returns the identifiers of the most recent resources associated with a given collection.
- `getCollections()`. Returns the identifiers of all the collection objects stored in the repository.

Methods for generating the repository collections tree

These methods are used to query Mulgara triplestore and construct a hierarchical tree of the collections stored in the repository.

- `callRecursiveCollectionsTree()`. This is the main method to start the generation of the collections tree.
- `recursiveCollectionsTree()`. This recursive method generates the tree.
- `nodeExists()`. This method checks whether a given collection node is already present in the tree.
- `isCollection()`. This method determines whether a given object identifier represents a collection object in the repository.
- `initTree()`. This method initialises the collections tree.
- `resetTree()`. This method re-sets or updates the collections tree after having performed collections or objects updates.



Figure A.7: MulgaraOperations class diagram

HttpManagement Class

This class is a singleton class that implements the low-level http connection functions to query via HTTP the triplestore. It relies in the external Apache library `HttpClient`⁵. The main methods of this class are the following:

- `getInstance()`. This method allows to get the instance of the class

- `getHttpClient`. This method allows to access the `httpClient` instance associated with the class.
- `httpSparqlQuery()`. This method makes Sparql HTTP requests to the triplestore.
- `httpItqlQuery()`. This method makes iTQL⁶ HTTP requests to the triplestore.
- `httpPostMulgaraRequest()`. This method makes POST requests to the triplestore; it is used to create and load new graphs in the triplestore.
- `destroyConnectionManager()`. Destroys the connection manager object associated to the `httpClient` instance.

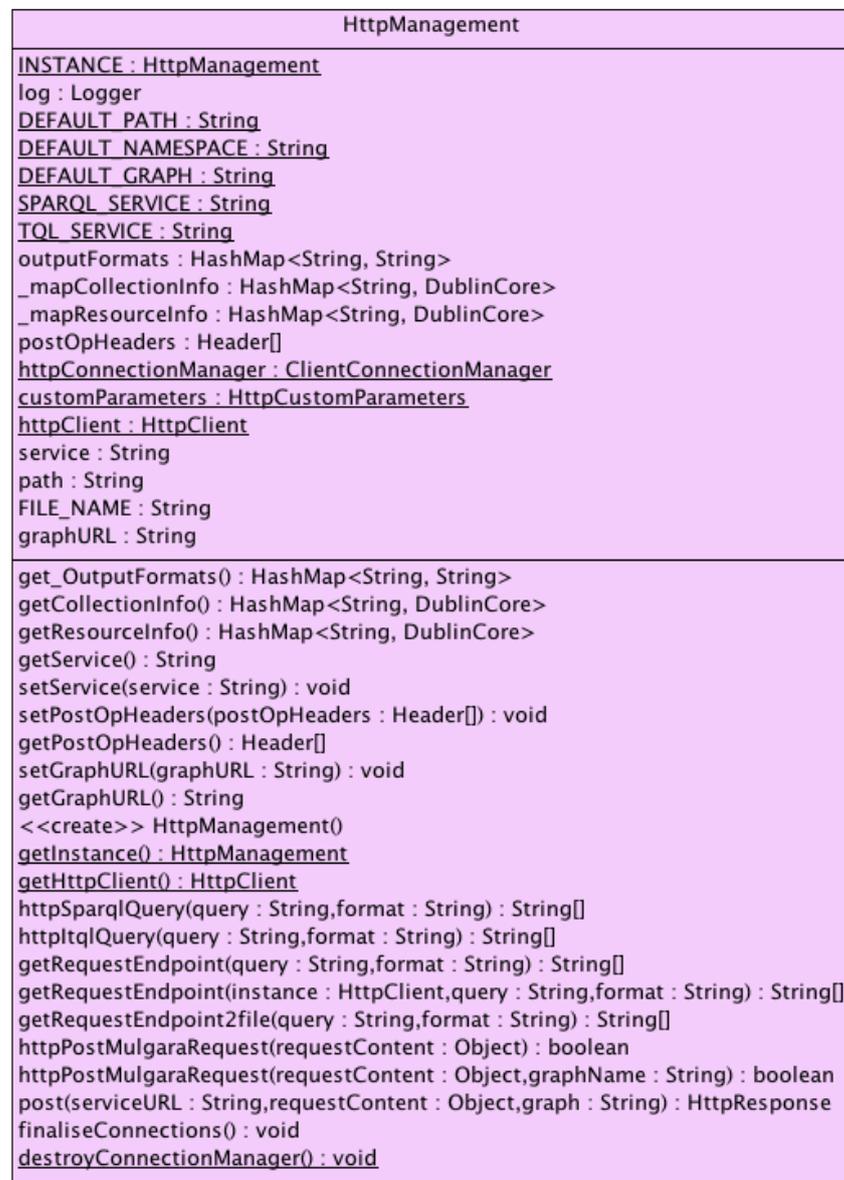


Figure A.8: HttpManagement class diagram

A.4 Toolkit Web Applications

The process of creating QuDEX-based collections in the repository involves two separate processes; the first one is related to preparing the data files that will represent the collections and their associated data files and the second one is related to defining collections structure and store/ingest them in the repository. For these purposes two different web applications have been developed.

A.4.1 QuDEX Template Builder

This is a Java-based web application that is designed to be deployed in Apache Tomcat application server although it could be deployed in other environments that are Java-based as well, like JBOSS. This application uses the Spring MVC framework to implement the logic and services of the application, and Spring Webflow to implement all the user interaction processes (different visual interfaces to perform all the steps of the data templates creation).

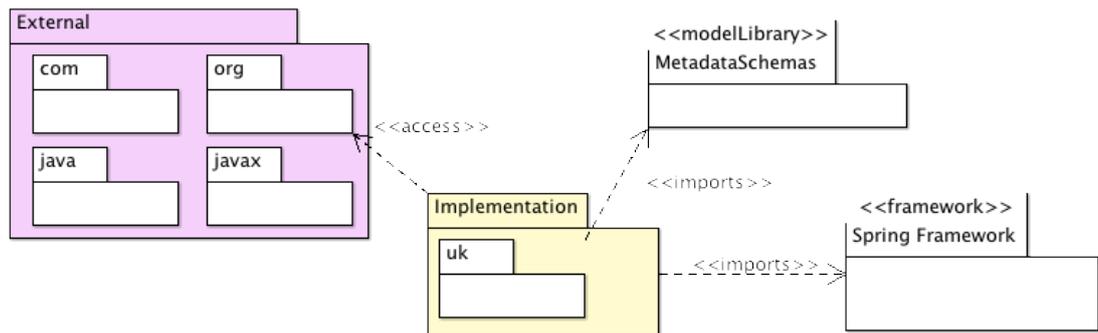


Figure A.9: Components Diagram

The web application uses external Java libraries in its classes implementation and uses the library MetadataSchemas which implements all the functionalities to retrieve and manage the information of the different metadata schemas used in the application (DDI2, DC and QuDEX).

A.4.1.1 General Use cases

This section describes the application different use cases, which represent application processes and user interactions with the system. At a general level two different processes are identified: generate collections or sub-collections spreadsheet templates and generate files spreadsheet templates.

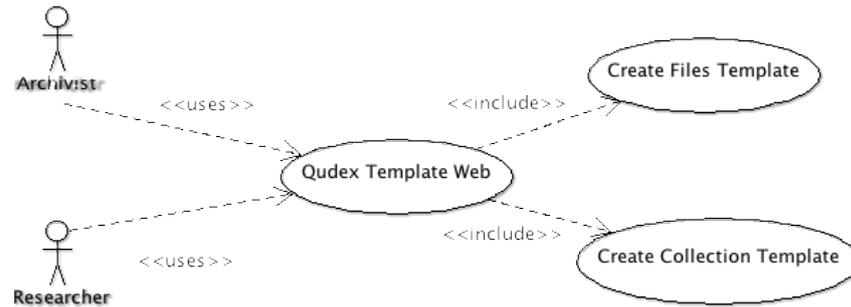


Figure A.10: General use case for the web application

The first process enables to create a spreadsheet template to describe what is called “QuDEX collection”. A QuDEX collection is a container for data files within a study. A collection is described using Dublin Core schema elements, DDI2 Lite elements and QuDEX elements. This object does not have physical files attached to it apart from the spreadsheet that was used to generate, so in Fedora terms, this will be represented as an object which contains metadata records attached to it and a file which represents the spreadsheet used to generate this object. This allows not only to store the collection object but also offers a roll-back mechanism in case one wants to either update or delete this collection.

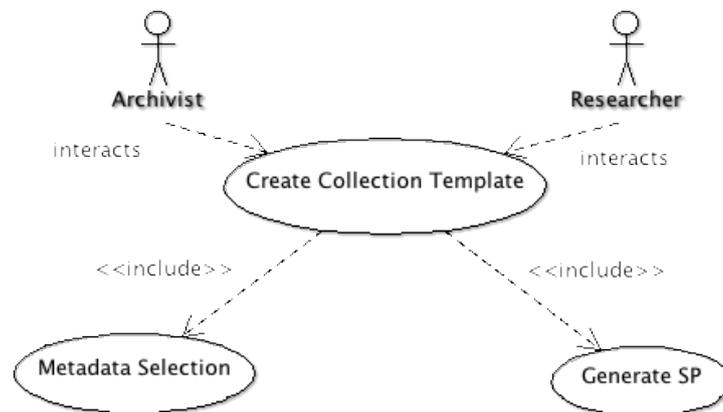


Figure A.11: Create Collection Template use case

The use case “create collection” represents the scenario of creating a collection/sub-collection template and includes two sub-scenarios or separate activities within the process.

The use case “metadata selection” represents the activity of selecting which elements from the available schemas (DDI2, QuDEX, DC) are going to be included in the template. There are compulsory and optional elements, therefore enabling the user to select/de-select elements from the schema is a necessary step within the process. However, this activity is optional since all the needed elements are preselected from the beginning so if the user wanted to include all the elements he could not perform this step and jump directly into the “generate spreadsheet”

activity.

The use case “generate spreadsheet” represents the activity of validating all the elements included in the template and finally generate the spreadsheet template to create collections/sub-collections. Optionally the user could attach the DDI2 information present in an external DDI2 instance XML document. The application incorporates a parser that allows parsing a DDI2 XML instance and incorporating the relevant DDI2 elements into the generated spreadsheet template. The use case “review schemas” enables to go back to the metadata selection process in case any changes to the metadata elements are needed.

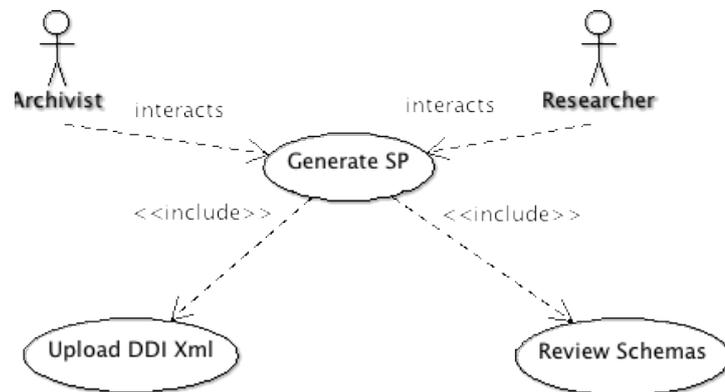


Figure A.12: Generate Collection Template use case

The second process enables to create a file spreadsheet template. A file object is any data (image, document, report, video, audio, etc.) that is associated with an existing collection or sub-collection. Files objects are stored in the repository and present the following different types of data:

- Metadata record describing the object, which includes elements from the following metadata schemas: [QuDEx](#), [DC](#), Dublin Core terms, GEO vocabulary and Simple Knowledge Object System ([SKOS](#)) vocabulary.
- Metadata elements holding relationship information (elements included in Fedora relationships ontology) that specifies all the relationships between this object and its collection parent or between this object and other files objects present in the repository. This information is stored in a special datastream within the fedora object (RELS-EXT datastream).
- File physical data file. The data file associated with this object is stored in a separate datastream within the fedora object and it could be physically present in the repository or be an external reference to the contents, which are stored somewhere else (for example in a web server, other repository, etc.).

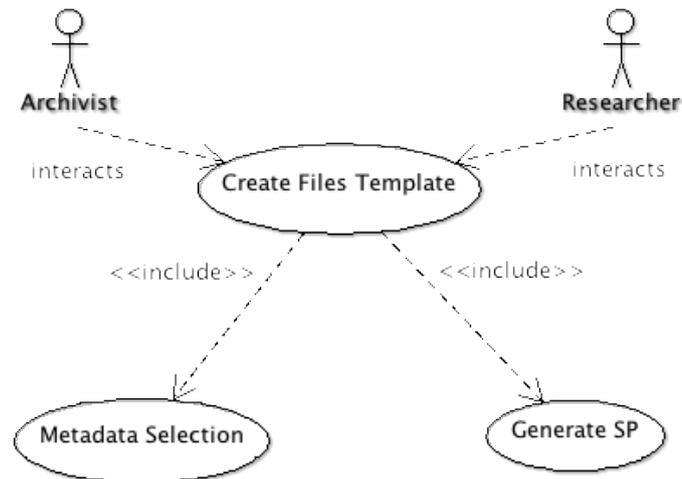


Figure A.13: Create Files Template use case

The use case “generate spreadsheet” in this activity is different since a files spreadsheet does not incorporate DDI2 elements, therefore the use case “upload DDI XML” is not present. However, the user can attach physical files (grouped in zip files) to the spreadsheet template. These files will be used later to upload them in the repository by using the repository management web application. Multiple zip files can be attached to a spreadsheet files template and the application will parse these files and auto-incorporate their information into the template being generated. This is very useful when managing large amount of files since the user does not have to do this process manually in the spreadsheet.

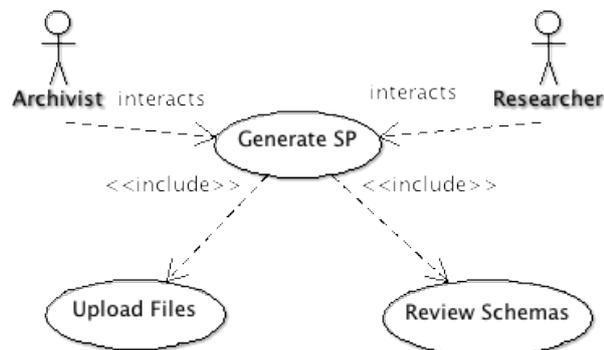


Figure A.14: Generate Files Template use case

A.4.1.2 Application Flows description

As described in the introduction to the web application, it uses Spring Webflow to design and implement all the interaction flows in the application (user web interfaces are design by using flows depending on the activities being performed). The application comprises the following flows (from Spring perspective):

- Metadata Flow. This flow represents the process of choosing the different metadata schemas

that will be available to include in the spreadsheet template.

- Metadata-elements Flow. This is a sub-flow of Metadata-flow and it is used to select, add or delete elements from a particular schema and included in the template spreadsheet.
- Upload Flow. This flow manages the physical files uploads.
- Generate Flow. This flow manages the process of generating the spreadsheet templates and including any physical attached files to the spreadsheet template.

Metadata Flow

This is the initial flow of the web application. It is composed by different view states, depending on the stage within the process and it calls an additional sub-flow (Metadata-elements) to perform operations in the metadata elements. Its end state calls a new flow (Generate Flow) to continue with the template generation process. It presents the following views:

- Select type of spreadsheet. Initially the user has to select the type of template there are going to build: collection/sub-collection template or files template
- Select metadata schemas. Depending on the previous selection different metadata schemas are available. From this view, one can select any of the schemas in case the elements present need to be modified.
- Add Metadata elements. This view displays all the elements available for a given metadata schema. The user can select/deselect any of the elements.

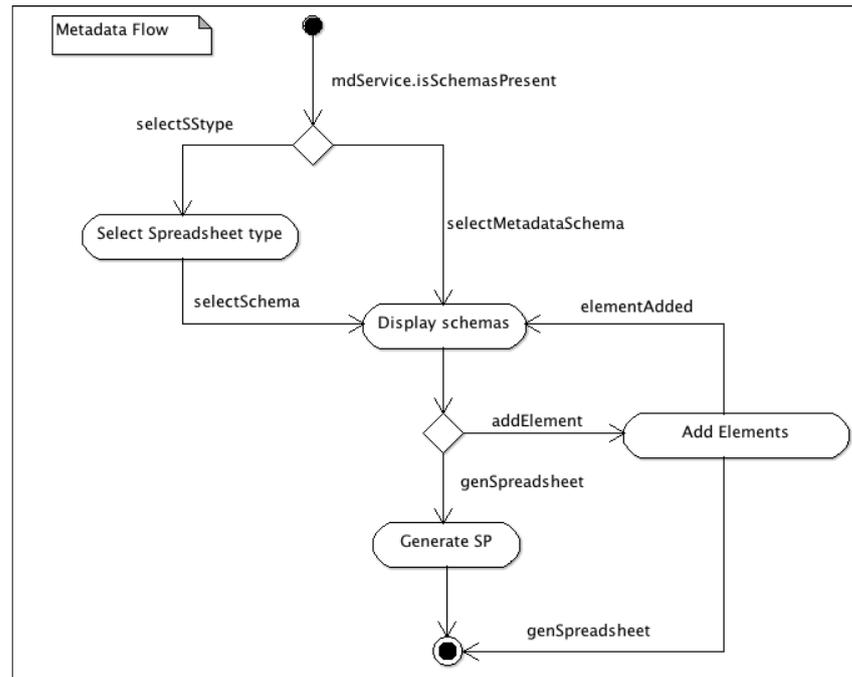


Figure A.15: Metadata Flow activity diagram

Generate Spreadsheet Flow

This flow describes the final process of generating a spreadsheet template. It is reused in either of the templates scenarios (file and collection templates). Depending on the type of spreadsheet template only certain views/states are present although there are shared views between the two scenarios. The views that are dependent on the type of spreadsheet being generated are "File upload" and "Upload DDI XML" and they are represented in the Figure A.16 below (fork element, black bifurcation).

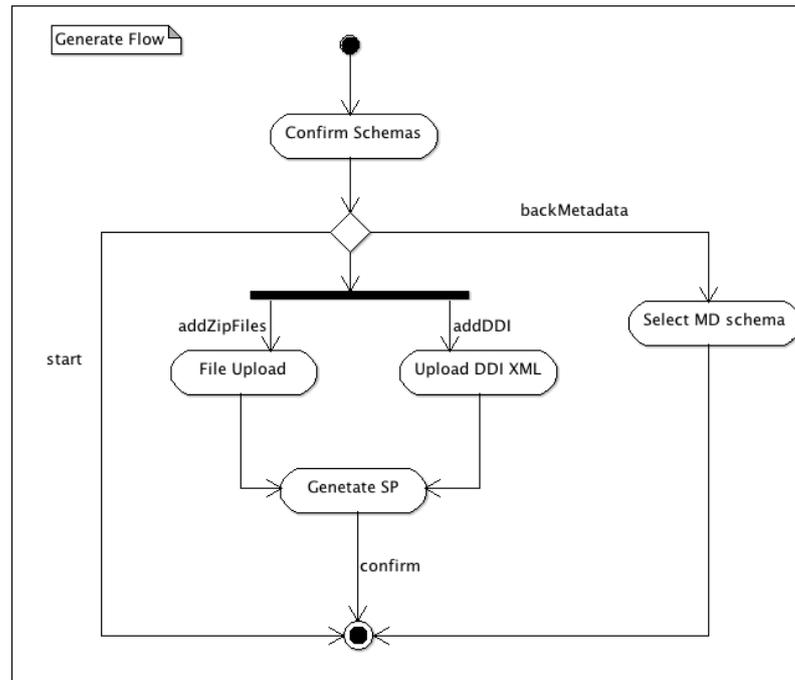


Figure A.16: Generate Flow activity diagram

Upload Flow

This flow represents the process of uploading ZIP files associated with a spreadsheet template. This flow is only present if the user is creating a file template since this type is the only type that accepts descriptions of physical files associated with each file being described in the template. It represents a sequential process of uploading ZIP files one at a time.

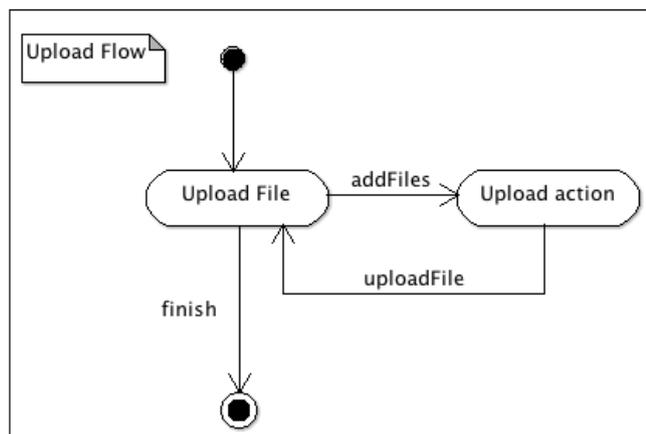


Figure A.17: Upload Flow activity diagram

A.4.1.3 Application sequencing and collaboration

The logic of the web application is implemented combining spring web flows, services and action classes. The services are java interfaces that make available a set of methods to perform the application operations. The action classes are a special type of class available from Spring MVC framework that implements methods that can be called from web flows. They provide the same functionality as a service although they are classes specifically designed for working with web flows. The application implements one service and three action classes to perform all the functionalities needed in the web flows. These will be described in detail in the “application components” section.

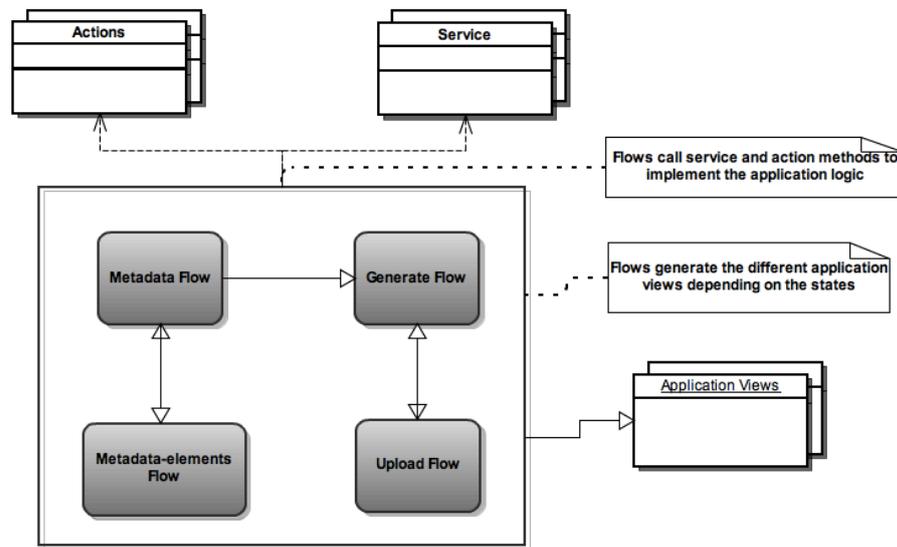


Figure A.18: Flows and components interaction

The application presents two different scenarios that are managed by the two main flows (Metadata Flow and Generate Spreadsheet Flow). The first scenario represents the process of designing all the metadata elements that will be included in the generated spreadsheet templates. The steps of these processes are the same for both types of spreadsheet templates and the only difference is that the schemas and elements differ. The second scenario represents the process of validating the selected schema elements and generating the spreadsheet template. Depending on the type of spreadsheet being generated the steps and options presented to the user differ. Both processes are orchestrated by the same main flow (Generate Spreadsheet) but depending on the type of template being generated, different secondary flows are invoked.

The first scenario involves selecting all the elements that will be included in the final generated spreadsheet template. Initially all the possible elements, depending on the type of spreadsheet, are pre-selected by default. This covers optional and compulsory elements. The user has to check manually each of the different schemas to select/deselect the elements that will be included. From a technical perspective, this process is performed by two flows: Metadata Flow and Metadata-elements flow. Metadata flow directs the process of selecting the different schemas

that will be included and Metadata-elements flow directs the management of the elements of a given metadata schema. This process is described in Figure A.19, which represents the sequence of steps and methods invocations involved in the process of designing the metadata elements to be included in the template.

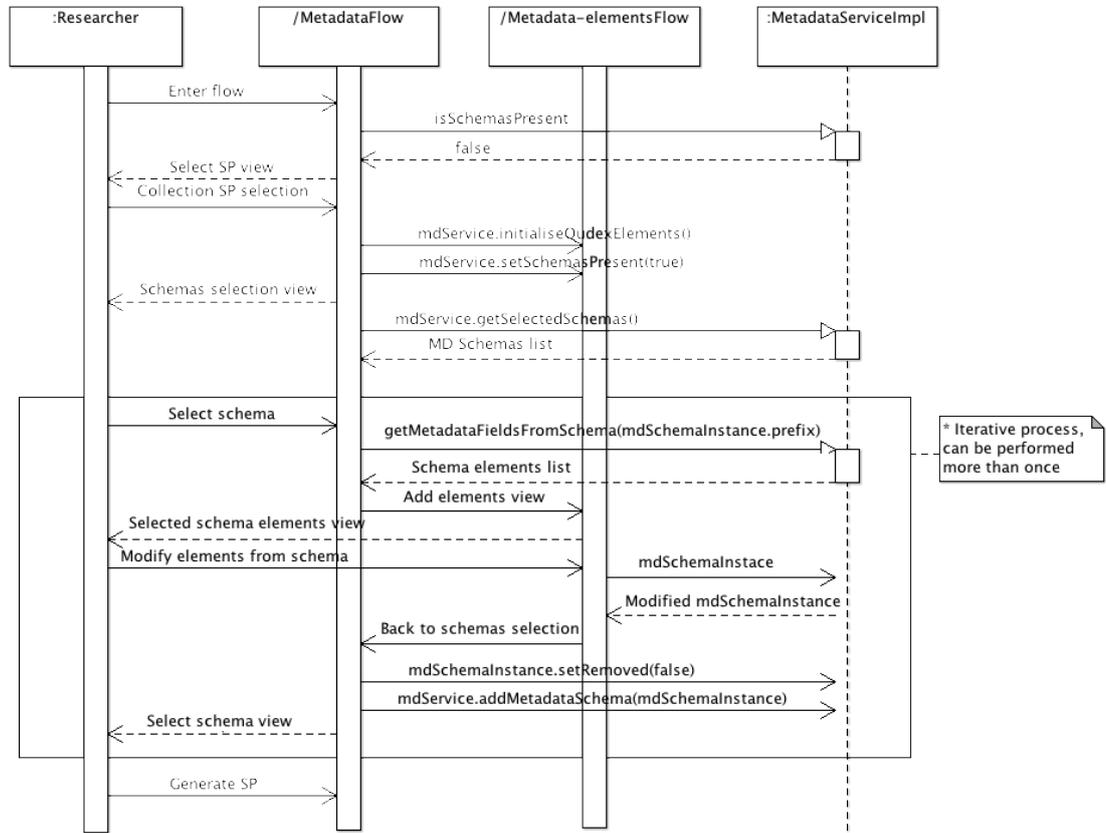


Figure A.19: Metadata sequencing diagram (scenario of selecting the included metadata elements and then proceed to spreadsheet generation)

When the user enters the application, the default controller (Spring web application dispatcher) directs the user to the introduction page. When the user hits start process then the dispatcher brings the control to the defined flows, in this case metadata flow. If this was a current session, the initialization of the metadata schemas (depending on the type of spreadsheet template) has been already performed. This means that the user is directly presented with the “Metadata Schemas design” view. Otherwise, the user is presented with the “Selection type of spreadsheet” view. This is the case represented in the diagram above: when the user enters the flow it calls the service (MetadataService) method “isSchemasPresent” which returns a Boolean expressing whether the array of schemas has been generated. If false then the user has to select the type of spreadsheet and then this will generate the appropriate schemas array (this performed by invoking the methods “initialiseQuDEXElements” and “setSchemasPresent”). Once the schemas are displayed the user can either start the iterative process of editing the elements selected in each of the schemas or could go directly to review schemas and generate the spreadsheet template since, by default, all the necessary elements are preselected. The scenario repre-

sents the first case. In this case, when the user selects a particular schema from the list and clicks the “add elements” button then the sub-flow “metadata-elements” starts and all the elements of the selected schema are displayed (“getMetadataFieldsFromSchema”). Once the schema has been modified, it is added to the array of schemas and the application returns to the “schema selection” view.

Finally, once the user has performed all the changes in the different metadata schemas and to exit the flow, he has to hit “generate spreadsheet” button that will start the “Generate flow”, which is in charge of the review and spreadsheet template generation processes.

Collection spreadsheet template generation

When generating a collection spreadsheet template three different actions can be performed:

- Review the metadata elements for each of the schemas and, if needed, the user can go back to the metadata design stage.
- Upload a DDI2 instance file which will be parsed and incorporated to the spreadsheet template.
- Generate the template spreadsheet.

Figure A.20 represents the process of collection spreadsheet generation with the previous upload and parsing of a DDI2 instance file and finally spreadsheet generation. In this case the components used are the Generate Flow, MetadataService, domain class “Files” (in charge of the management of the uploaded DDI2 instance file) and the Action classes “checkSpreadsheetAction” and “generateSpreadsheetAction” in charge of validating the selected elements and generating the spreadsheet template respectively. Finally, “parseDDIaction” is used to parse the uploaded DDI2 instance and generating the appropriate data structure for its later inclusion in the spreadsheet template (see diagram below).

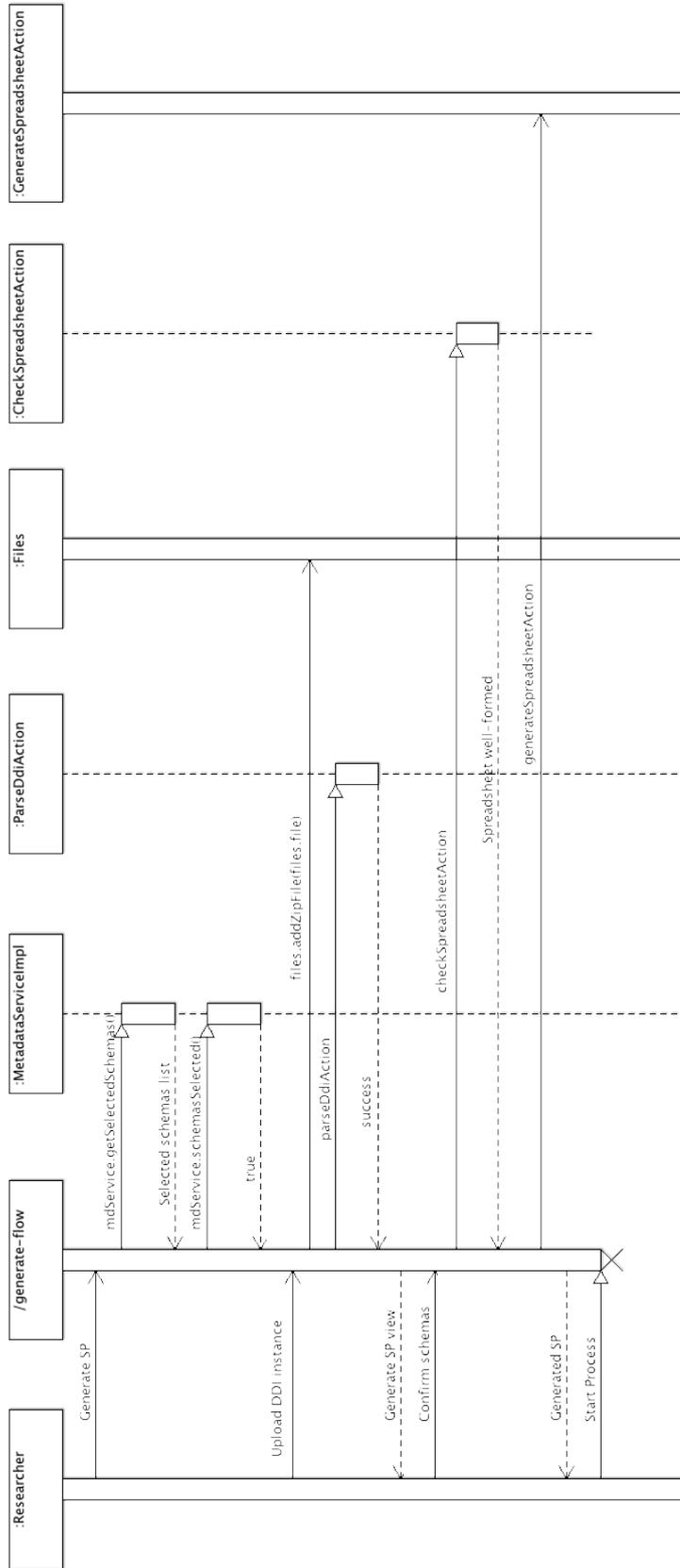


Figure A.20: Sequence Diagram of Collection Spreadsheet generation

When the user enters the Generate Flow, an array of the schemas associated with the template and the selected items from each of these is displayed (invoking “getSelectedSchemas”). The next step of the scenario is to upload the XML file containing the DDI2 instance and then “parseDDIaction” is invoked to validate and parse the file. If this the validation is successful then the data structure containing the DDI elements to include in the template is generated and then the user has to be click on “generate spreadsheet” which will validate that the selected elements conform to the collection template schema and then the spreadsheet will be returned to the user for its download. After this, the user can either finish the process or start another spreadsheet template generation.

Files spreadsheet template generation

This scenario uses the same components than the Collection template generation with the difference of calling the Upload sub-flow instead of the DDI generation process (a files template does not include DDI elements but includes the possibility of including data files descriptions in the template). The scenario in the sequence diagram below represents the components interactions during the process of generating a spreadsheet template with contains data files associated with the documents being described in the template. This will return a template with the relevant metadata fields auto-filled in.

When the user enters the “Generate flow”, again visualises an overview of the selected metadata elements from the schemas included in a files template. He could choose to review them, which will bring him back to the metadata design stage or could start the ZIP(s) file(s) upload process, only in case there were files attached to the template. The upload process is an iterative process that can be performed as many times as needed and the components involved here are the “Upload Flow”, the Files domain object that manages the all the uploaded files for its later parsing and metadata generation.

Once all the files have been uploaded the spreadsheet template is validated to check the conformance of the selected elements with the files template schema (by invoking the checkSpreadsheetAction class) and then the generation of the spreadsheet is performed by the action class ‘generateSpreadsheetAction’, which in this case parses compressed zip file. All the files included are analysed and then the spreadsheet is filled in and returned back to the user for download.

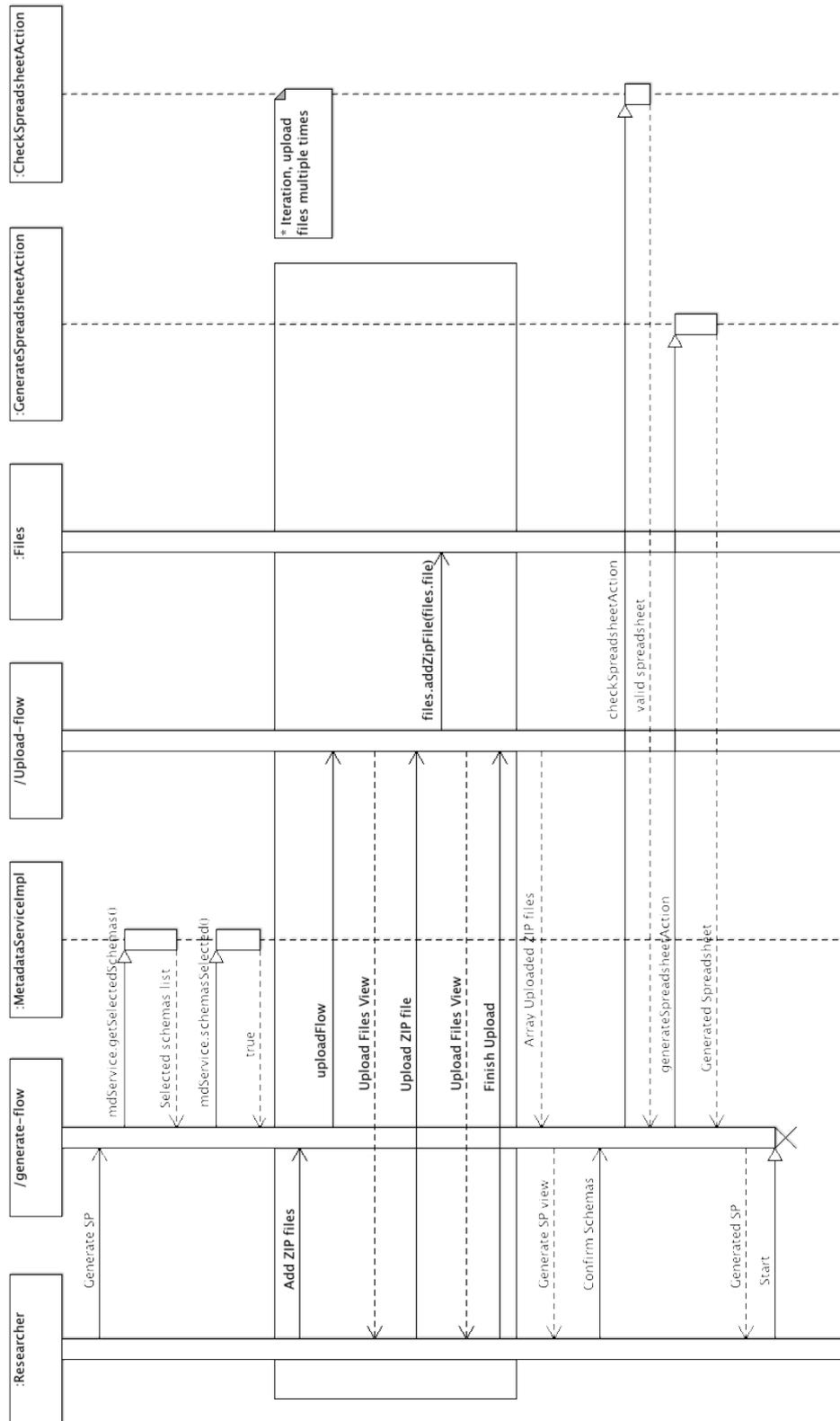


Figure A.21: Files Spreadsheet Template sequence diagram

A.4.1.4 Application Packages and Classes description

Actions Package

This package implements the Action classes. These classes extend from the “AbstractAction” class (included in the Spring MVC framework) and implement a set of methods used in the application web flows. The following classes are included in this package:

- CheckSpreadsheetAction
- GenerateSpreadsheetAction
- ParseDDIaction

CheckSpreadsheetAction class implements the ‘doExecute’ method that performs a validation of the elements included in the spreadsheet.

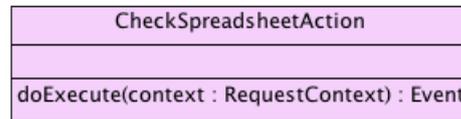


Figure A.22: CheckSpreadsheetAction class diagram

GenerateSpreadsheetAction class implements the ‘doExecute’ method that generates the spreadsheet template and returns it back to the user for download.

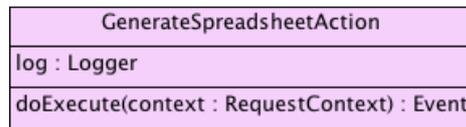


Figure A.23: GenerateSpreadsheetAction class diagram

ParseDDIAction class implements two methods to manage the process of parsing and generating the DDI elements data structure that is used to auto-fill in the template spreadsheets being generated. The first method is ‘doExecute’, which takes the uploaded files and invokes the private method “parseDdiAction” that implements the logic of parsing each of the uploaded files and create the record containing the metatata descriptions that will be included in the spreadsheet template.

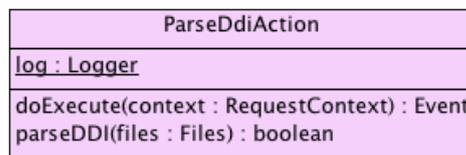


Figure A.24: ParseDdiAction class diagram

Domain Package

This package contains the classes implementing the different data models used in the web application (Spring MVC environment). The application manages three different types of data: spreadsheet elements validation (defined in XML files that are parsed into the application), elements from the different metadata schemas (primarily Dublin Core and QuDEx schemas) and DDI2 elements. The next sections describe in details the functionalities implemented in each of the model classes as well as the methods defined.

Elements and Element classes

These two classes are used for validation purposes. The application uses two different XML files (depending on the type of template being generated: files or collection) that define what elements and their multiplicity are permitted in each of the two types of templates. These files are serialised into the two classes being described now: Elements and element. The Element class is used to load one validation element from the validation XML files. The attributes defined for validation purposes are the following: element name, element namespace (dc for Dublin Core, qudex for QuDEx schema and so on), element permitted multiplicity and optional (whether the element is optional or compulsory).

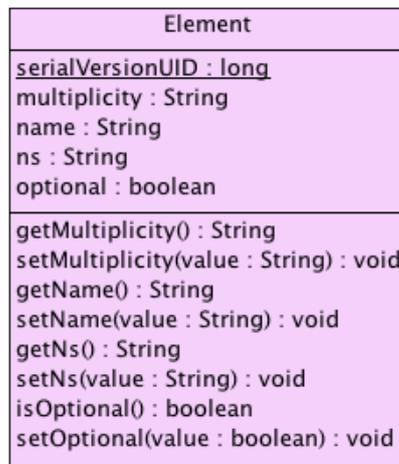


Figure A.25: Element class diagram

This class is used to manage validation metadata elements and it only provides with getter and setter methods to access the values of a particular record.

Elements class

This class defines implements the data structure used for the validation of the elements of a given spreadsheet template type (files or collection). It defines a list of individual metadata element validation records (instances of the Element class).

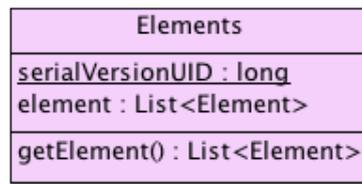


Figure A.26: Elements class diagram

This class provides one getter method to access and initialise the array of validation elements. The structure is generated from the serialisation of the template validation XML file.

VocabularyElement class

This class implements the data structure needed to describe a metadata element with independence from its schema precedence. This is used to define the metadata elements used in the templates and that are managed in each type of spreadsheet template. The previous two classes are used only with validation purposes whereas an instance of the “VocabularyElement” class provides descriptive information from each metadata element being used. This information includes the name of the element and the comment attribute, both coming live with the online versions of the metadata schema being used. The following attributes are defined to describe each element record:

- The attribute “name” represents the name of a specific metadata element (comes from the [RDF](#) schema).
- The attribute “label” represents the human-readable name of a specific metadata element (comes from the [RDF](#) schema)
- The attribute “type” represents the [RDF](#) type of element (property, resource). It is defined in the [RDF](#) schema.
- The attribute “comment” represents the description of the element (its value comes from the [RDF](#) schema).
- The attribute “hidden” represents a Boolean field that specifies whether the element has to be displayed or not. This is used when working with compulsory elements that cannot be deselected by the user.
- The attribute “selected” represents a Boolean field that specifies whether the user selected this particular element.
- The attribute “optional” represents a Boolean field obtained from the validation schemas to determine which elements are optional or compulsory.

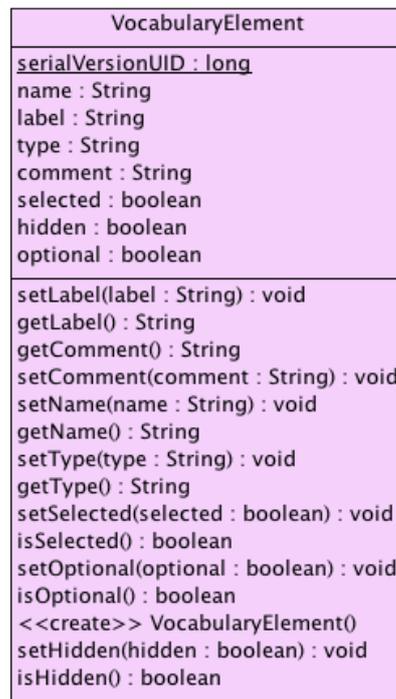


Figure A.27: VocabularyElement class diagram

All the methods defined in this class are getters and setters used to assign and obtain the values of the attributes defined in the class.

Files Class

This class implements the logic to manage the files that are uploaded onto the application. Two different types of files are managed: compressed files containing resources associated with files spreadsheet templates and DDI instance XML files associated with collection templates.

- The attribute “file” is used to manage the DDI2 XML file.
- The attribute “zipFiles” is used to manage the compressed files uploaded.
- The attribute “ddiObjectList” is used to store the list of DDI elements parsed from the uploaded XML instance.

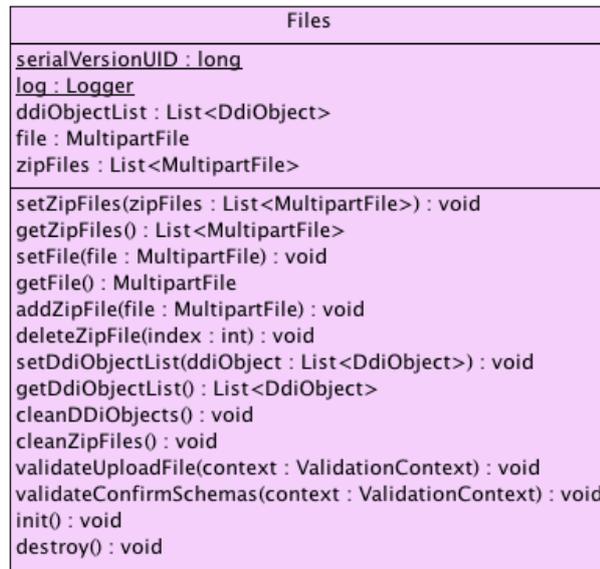


Figure A.28: File class diagram

This class provides the setters and getters of the objects used for managing the files upload as well as those methods for the validation of these files.

- The method “validateUploadFile” validates whether the compressed uploaded file presents the correct format.
- The method “validateConfirmSchemas” is used when working with collection templates and validates that the user has uploaded an XML file and it is a valid DDI2 instance.
- The methods “cleanDDiObjects” and “cleanZipFiles” dispose all the objects used for the management of the uploaded files once they have been processed.

MdSchema Class

This class implements the representation and management of a particular metadata schema used in the application. Two different schemas are used: **QuDEx** and Dublin Core. Both schemas are represented using **RDF-XML** format. Dublin Core presents an online version which is accessed live by the application and then loaded into an instance of a Jena model object (Jena **RDF** framework). The **QuDEx XML** schema does not present an official **RDF** version (it is defined using **XML** stylesheets) but a light **RDF** version has been developed for this toolkit which includes the elements used in the Collections model and, that are expressed using **RDF** properties. Again, this **RDF** file is accessible from the web application and then loaded into a Jena model object. The model objects are used to obtain the metadata elements and their description fields. This information is used to construct the dynamic list of selected metadata elements.

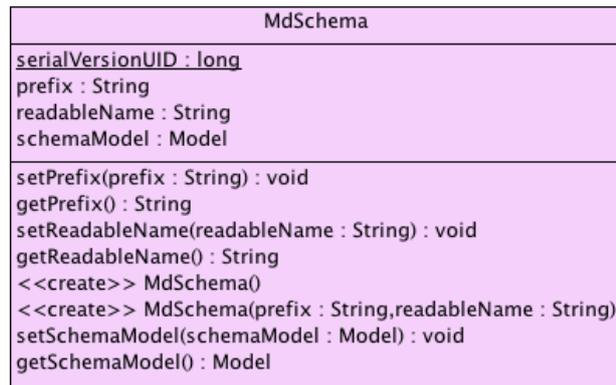


Figure A.29: MdSchema class

The methods implemented in the class are the getters and setters to provide access to the RDF model of the metadata schemas used in the application.

MdSchemaInstance Class

This class implements the live management of the elements associated with a particular metadata schema and that are being used by the user during a session in the web application. When the user works with a particular schema, either when creating a files or a collections template, this class holds the information of each of the elements being used within a particular schema. The attributes of the class are the following:

- The attribute “removed” represents a Boolean to determine whether the user has removed a particular metadata schema from the spreadsheet template.
- The attribute “selectedSchemaElements” represents a dynamic list of “VocabularyElement” objects which represents the current schema elements that are being used in the working spreadsheet template.

The methods implemented in the class provide getters and setters to access to all the attributes of the class as well as search methods (`getElementFromSelectedSchema`) that provide access to particular elements within the dynamic list of all the selected elements from a given metadata schema.

| MdSchemaInstance |
|--|
| <pre> serialVersionUID : long removed : boolean selectedSchemaElements : List<VocabularyElement> </pre> |
| <pre> setSelectedSchemaElements(selectedSchemaElements : List<VocabularyElement>) : void getSelectedSchemaElements() : List<VocabularyElement> <<create>> MdSchemaInstance(prefix : String,readableName : String) <<create>> MdSchemaInstance() validateSelectMetadataElements(context : ValidationContext) : void getElementFromSelectedSchemas(elementName : String) : VocabularyElement setRemoved(removed : boolean) : void isRemoved() : boolean </pre> |

Figure A.30: MdSchemaInstance class

Controllers Package

This package implements the different controllers used by the web application to handle all the different requests from the user.

Spring's web MVC framework is request-driven, designed around a central servlet (in the application's case this is the DispatcherServlet) that dispatches requests to controllers and offers other functionality that facilitates the development of web applications. This servlet is in charge of receiving the requests from the user and then direct them to the appropriate controller. The web application is implemented combining Spring MVC and Spring Web flow. Each of the flows implements internally their own controller, therefore it was only needed to implement their associated handlers to deal with flow termination and exception management. The handler classes for each of the web flows are the following:

- GenerateFlowHandler
- MetadataFlowHandler
- MetadataElementsFlowHandler

These classes extend from Spring's AbstractFlowHandler class and they implement two different types of method: handleExecutionOutcome (when the flow reaches a termination state this method checks whether it returns any objects and handles the next redirection) and handleException (method to manage any exception occurred during the flow execution). Each of the classes listed above implement these two methods. Additionally, the class GenerateFlowHandler implements additional exception handling methods to deal with exceptions that are thrown by business logic methods (handleOtherException method). To avoid duplicities only one diagram is provided since all of the classes implement the same methods.

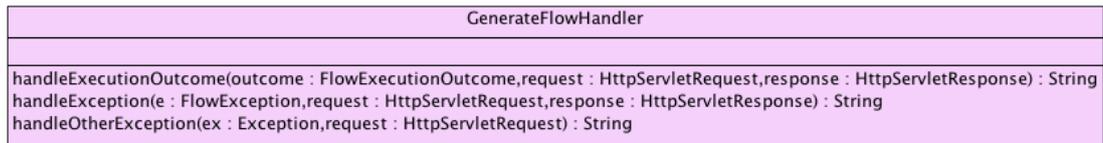


Figure A.31: GenerateFlowHandlerClass class diagram

On the other hand, as mentioned above, the defined flows implement internally their own controllers and it is only needed to extend the handler classes to manage flow termination and exception handling. The application needs to implement additional controllers to handle any requests that are not included within a web flow scenario. In this case the application implements a default controller which handles requests outside of the scope of the defined flows, like for example, access to the home page or access to the online web application documentation. The class implementing this controller is `DefaultController`, which uses Spring's annotation mechanisms to inform the dispatcher that this class represents a Controller class (annotation `@Controller`). `DefaultController` class implements the following attributes and methods:

- The attribute "zipFile" represents a `MultipartFile` object which is used for managing file uploads into the application.
- The attribute "msService" represents a singleton instance of the service `MetadataServiceImpl` which provides the implementation of most of the methods in charge of the application's business logic.
- The attribute "messageSource" represents an instance of the class `MessageSource` that is used to handle all the information messages displayed to the user. These messages are defined in a properties file and then used dynamically during the application execution.

This class implements methods to handle particular requests:

- The "intro" method handles a request to go to the application's main page.
- The "restart" method handles a termination request, which means that the user has generated the spreadsheet templates and wants to finish the process.
- The "download" method handles how to return a spreadsheet template once this has been generated.

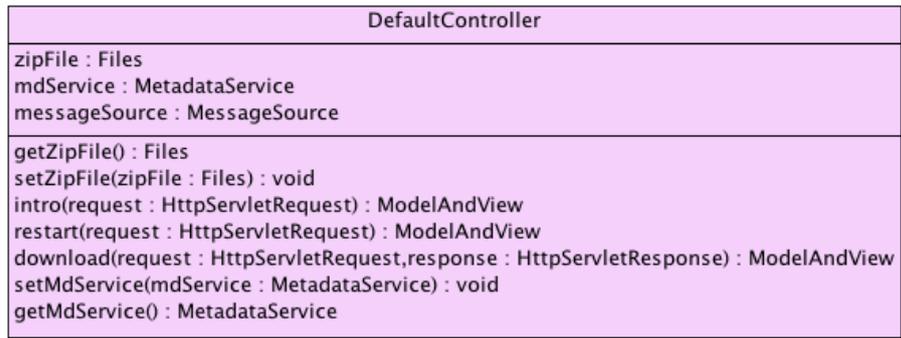


Figure A.32: DefaultController class diagram

Services Package

This package includes the service implementation that manages all the business logic of the application. It is implemented as an interface (MetadataService) and its associated implementation class (MetadataServiceImpl).

This service is loaded as a singleton instance (defined in the application Spring's configuration file) once the application is started in an Applications Server like Tomcat and is auto-wired using the Spring mechanisms so that all the relevant classes in the web application, like web flows, request controllers can access it and invoke its methods. As a general overview, it provides all the methods to manage all the metadata schemas used during spreadsheet templates generation, initialization methods to access to the metadata models that hold all the elements included in each of the schemas used and termination and clean-up methods to dispose all the objects and instances used in the generation processes and finally, implements template validation methods based on the XML files defined in the application. It uses the domain classes to generate and manage the data structures that represent the metadata elements handled during user sessions.

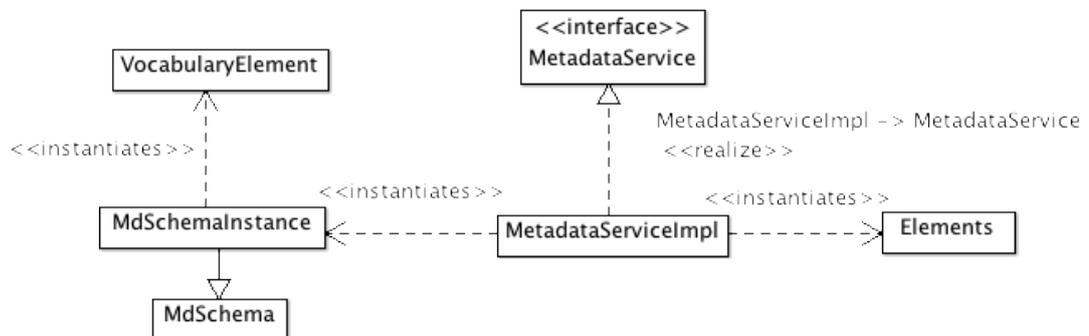


Figure A.33: MetadataService classes diagram

The class instantiates the class Elements used to handle the template validation and instan-

tiates `MdSchemaInstance` class to handle all the metadata elements lists and their status during the template construction processes. The class defines the following attributes or properties:

- The attribute “typeSS” is a Boolean field used to specify the current type of template being generated (collections or files templates).
- The attribute “qudexElements” is an instance of the `Elements` class and is used to hold the permitted elements of the working spreadsheet template.
- The attribute “mdSchemaInstanceContainer” is a dynamic list of `MdSchemaInstance` object used to handle the metadata schemas being used by a given working spreadsheet template.

The class implements the following methods, grouped by functionality:

Template Validation Methods

- `initialiseQuDEXspreadsheetElements()`. It constructs the validation metadata elements list depending on the type of template being generated. It loads and serialises the validation XML files into Java objects.
- `initialiseQuDEXElements()`. It generates the pre-selected metadata elements of a given schema used in the spreadsheet template based on the elements specification provided by the schema validation object constructed in the method above.
- `addQuDEXInfoToElements()`.

Metadata Schemas management methods

- `initialiseSchemas()`. This method initialises all the models of the managed schemas. It uses the class `MdSchema` which contains the RDF model of a given schema. This method is only called when first initialising the service instance.
- `getSelectedSchemas()`. It returns the list of currently used metadata schema instances.
- `getMetadataFieldsFromSchema()`. This method returns all the elements of a given metadata schema. It obtains the elements from either from the static Java class that represents this metadata schema or from the Jena [RDF](#) model object. It is used when the metadata schema is not available online and needs to be accessed offline and when the schema presents an online RDF version, like the Dublin Core schema.
- `addMetadataSchema()`. It adds a new metadata schema instance to the dynamic list of used schemas.
- `anyElementSelected()`. It returns a Boolean depending on whether a particular schema instance presents selected elements.

- `deleteMetadataSchema()`. It deletes a metadata schema instance from the dynamic list of metadata schemas.
- `createMdSchemaInstance()`. It returns a new instance of a `MdSchemaInstance` object.
- `elimGroup()`. It deletes all the non-compulsory elements from a given metadata schema that is being used in the template generation.
- `findSchemaInstance()`. It finds a particular schema instance in the list of used metadata schemas.
- `schemasSelected()`. It returns the full list of selected metadata schemas.
- `isSchemasPresent()`. It returns a Boolean specifying whether the current template being generated presents selected schemas.

Clean-up and disposal methods

- `clean()`. It clears all the metadata schemas lists and validation lists constructed during the process of generating a spreadsheet template.

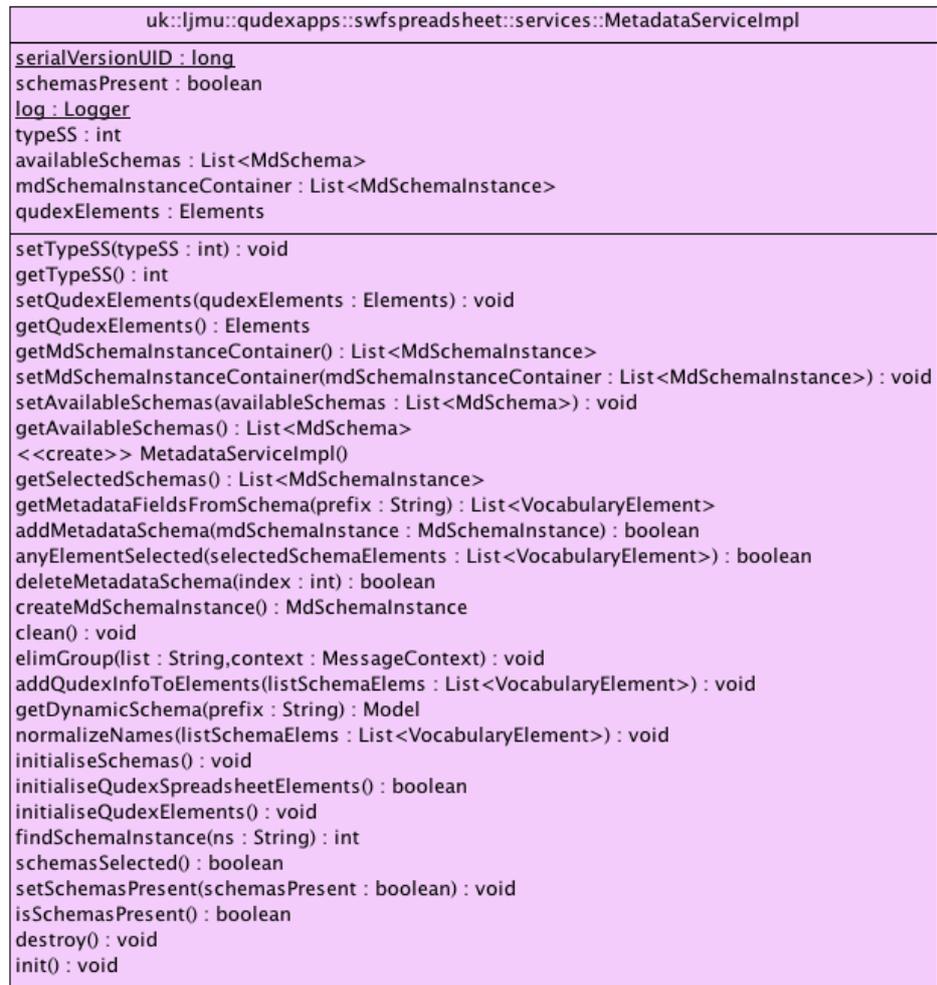


Figure A.34: MetadataServiceImpl class diagram

Utilities Package

This package includes auxiliary classes that are used globally within the web application. It includes classes to manage and create excel spreadsheets, constant classes and RDF management classes, used to manage the RDF models of the metadata schemas used in the application.

Constants Class

This is a static class which defines all the constants specifying spreadsheet header names and metadata schemas names and prefixes.

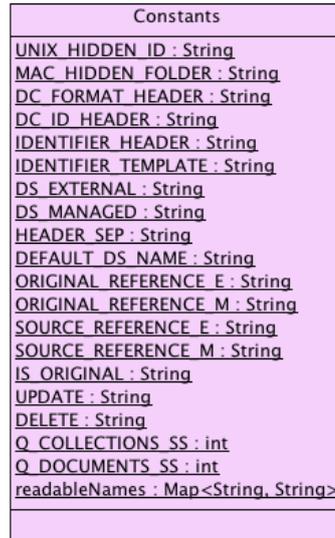


Figure A.35: Constants class diagram

RdfFunctions class

This class uses the Jena framework to manage data in **RDF** through Java objects. The web application service (MetadataService) works with the data coming from the **RDF** models of the used metadata schemas and this class implements the method needed to create **RDF** models from the schemas, read the models and access to specific elements from the schemas.

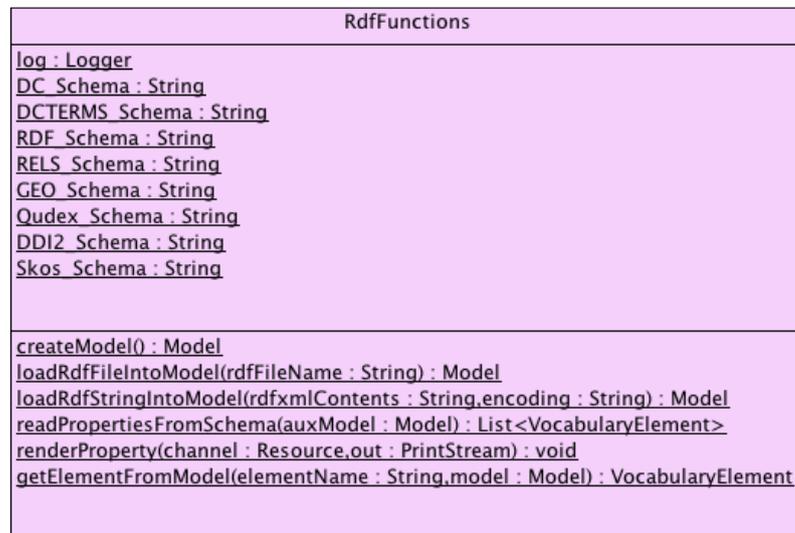


Figure A.36: RdfFunctions class diagram

This class defines a set of constants that specify the different metadata schemas used within the web application. This schemas are: **DC**, Dublin Core Terms extension, **RDF** schema, customised version of **DDI2** Lite, customised version of **QuDEX** schema, **GEO** vocabulary and **Skos** elements. The implemented methods are the following:

- `createModel()`. It returns a Jena [RDF](#) Model object which will be used to manage a given metadata schema model.
- `loadRdfFileIntoModel()`. It loads a set of [RDF](#) statements coming from a file into an existing Jena Model object.
- `loadRdfStringIntoModel()`. It loads a set of [RDF](#) statements coming from a text string into an existing Jena Model object.
- `renderPropertiesFromSchema()`. This methods visualises all the properties defined in a given Jena Model object.
- `getElementFromModel()`. This method enables to retrieve a specific element from a Jena Model object.

ExcelManagement Class

This class is a static class that implements all the methods to create and manage the generated excel spreadsheet templates. It uses the open source Apache POI library to manage Excel spreadsheets through Java interfaces. The class defines the following attributes:

- The attribute “fileSeparator” is used to determine the file separator used by the operating system hosting the web application. Its value is obtained using the Java System properties object.
- The attribute “filesList” is a list of string to hold the names of the uploaded files into the web application (included in compressed files).

The class implements the following methods:

- `createExcelFile()`. This method is the main method invoked to generate the spreadsheet template. It uses all the data structures (metadata elements, [DDI2 XML](#) elements and uploaded files) to create and auto-fill in the spreadsheet. This method does not handle the uploaded ZIP files.
- `autoFillDDI()`. This method uses the parsed [DDI2](#) elements and populates their values in the generated spreadsheet.
- `readZipFiles()`. This method decompresses an uploaded ZIP file and reads its contents.
- `indIsOriginal()`. This method finds in the spreadsheet the values for the qudex field “isOriginal”.
- `createExcelFileZipManagement()`. This method performs the same operations as “createExcelFiles” but with the difference, that it manages uploaded ZIP files and auto-populates their properties into the generated spreadsheet template.

- decompressZipFile(). This is an auxiliary method invoked in “readZipFiles” to decompress a ZIP file.
- processDirectory(). This recursive method reads all the files in a given directory.
- findExcelElement(). This method finds a specific column within the template spreadsheet.
- findDatastreamColumn(). This method finds a datastream type column in a given spreadsheet template.
- findUpdateDelete(). This method returns the column index of the qudex fields for updating and deleting spreadsheet elements from the spreadsheet template.
- findOriginalReferenceColumn(). This method returns the column index of the qudex field “originalReference” from the spreadsheet template.
- findSourceReferenceColumn(). This method returns the column index of the qudex field “sourceReference” from the spreadsheet template.

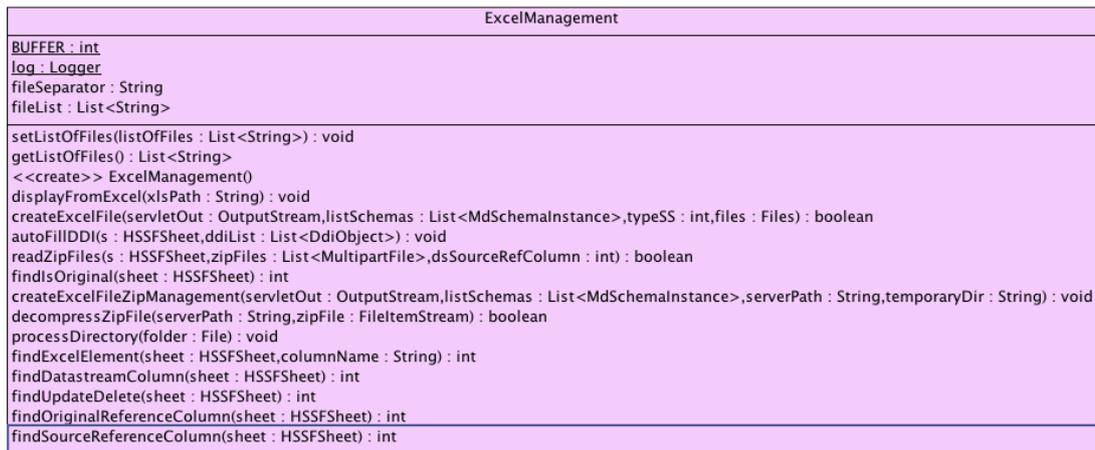


Figure A.37: ExcelManagement class diagram

A.4.2 QuDEX Collection Manager

QuDEX collection manager is a web application implemented using Java framework, Spring MVC and Spring Webflow frameworks. It is designed to be deployed in an Applications Server like Tomcat or Weblogic. This web application is used in combination with the Template Builder (described in the previous section) to create, update and manage QuDEX-based Fedora digital collections containing mostly qualitative data resources although it does not impose any restriction on the type of materials being stored in the digital repository. It implements web processes to upload collection spreadsheets which will create the collection structure and logical organization objects as well as to upload files spreadsheets which contain the actual data resources that will be stored in the repository. Since this application provides with all the mechanisms and functions needed to perform management of repository collections, it relies in the different

Java libraries described previously in this documentation. Figure A.38 shows all the components used by this application to provide repository management functionalities and collections visualisation using the metadata information stored in Mulgara triplestore.

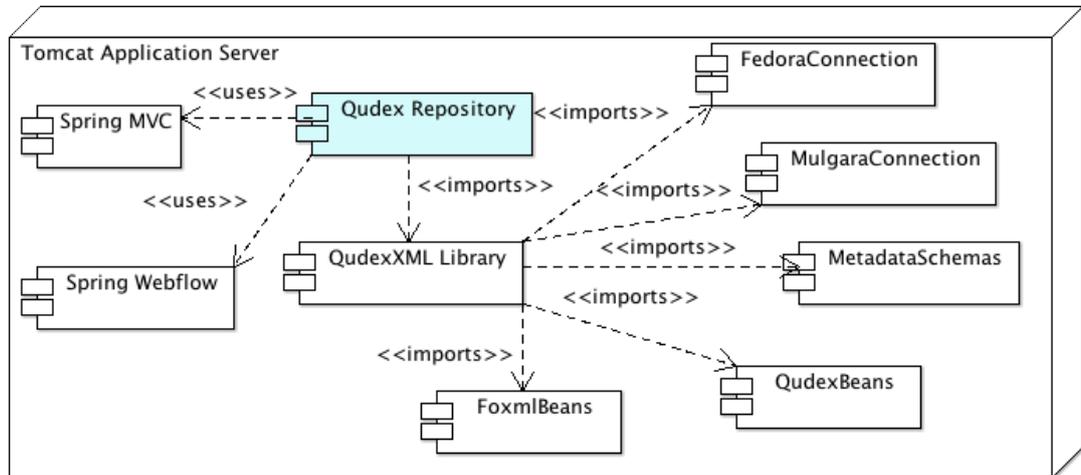


Figure A.38: QuDEX Repository deployment diagram

The core library to implement all the repository functionalities and metadata schemas management is **QuDEX XML** library. This library imports the auxiliary libraries that implement different groups of functionalities: communication with repository and triplestore; metadata schemas parsing; **FOXML** and **QuDEX** Java beans.

The **QuDEX** Repository web application is based in a **MVC** model and implements data management, business logic and interfaces or views separately. To implements the business logic it uses a combination of web flows and services implementations (Java interfaces). The data models used by the application are implemented in multiple Java beans and finally, the different views are designed and composed using Apache Velocity templates and Apache Tiles libraries.

A.4.2.1 General Use cases

This section describes the application different uses cases, which represent application processes and user interactions with the system. At a general level two different processes are identified: creation and upload of collections/sub-collections spreadsheets and creation and upload of files spreadsheets.

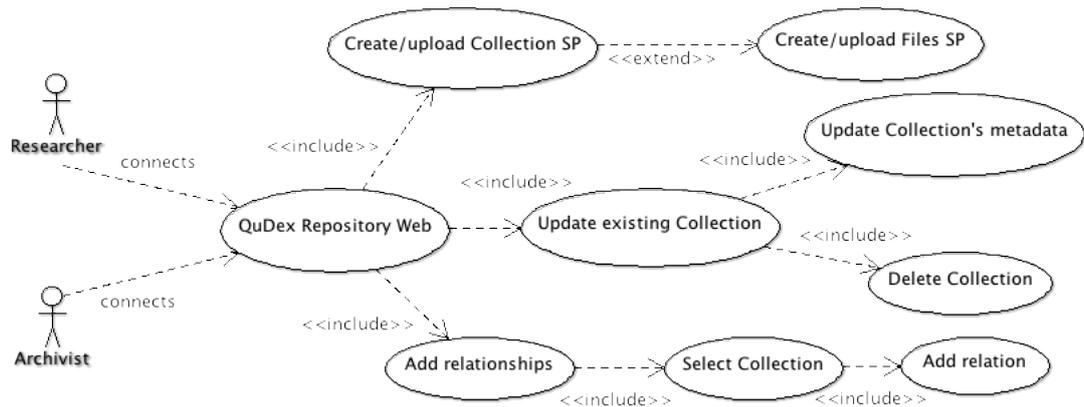


Figure A.39: General use case for the repository application

The application processes are sequential and need to be performed in a particular order. The web application enables to create/upload new **QuDEX** collections into an instance of Fedora repository and to modify and delete collections. The current version only supports updating the **QuDEX** collections metadata records (does not apply to Files documents) and the delete operations apply to whole collections (in this version it is not possible to delete sub-collections and then regenerate **QuDEX** Instance documents).

When the user enters the **QuDEX** application he has to select the operation to perform: start new collection creation process or update an existing collection. The first process comprises creating the collection structure and uploading the different collections spreadsheets, beginning with the higher-level collection components and finishing with their dependent sub-collections. Once the collections/sub-collections are created, the user can then start adding and uploading the associated files documents spreadsheets (if any). The process of creating a new collection and associating files with it is transactional, in other words, the user has to complete both processes to successfully create a **QuDEX** collection in the repository.

The second process (update an existing collection) enables the user to either modify the metadata records associated with one or more collections which are already present in the Fedora repository or delete complete collections. This is performed by uploading a collections spreadsheet which differs slightly from a normal collection spreadsheet template. This template includes the Fedora Permanent Identifiers (**PIDs**) of the collection objects that the user wants to modify or update. The following sub-sections explain more in detail the different sub-use cases contained in the general use case diagram shown above.

Collection operation selection use case

This scenario represents the process of starting the creation of a **QuDEX** collection. The user could select to start the collection creation process from the beginning, which means creating an entirely new collection, or could select to continue creating an existing collection. This second

scenario is only possible when working with collection that do not present any files associated with them, in other words, this is the case of a previously started collection creation without the association of files spreadsheets to the current collection.

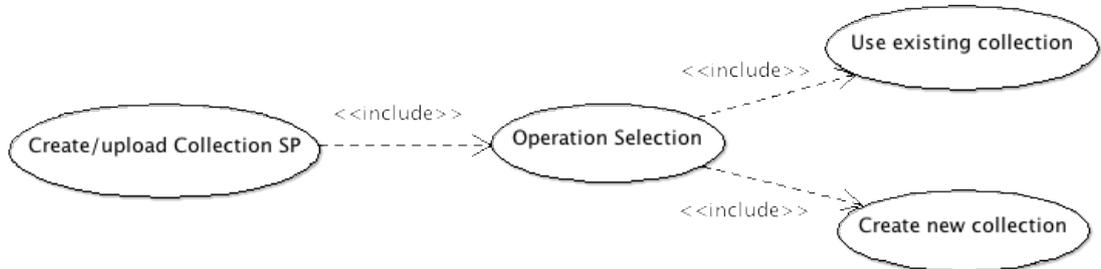


Figure A.40: Collection Operation Selection use case diagram

New collection creation use case

This scenario represents the process of creating an entirely new collection in the repository. The user creates first the structure of the collection in terms of hierarchy of collection objects and optionally, uploads the files spreadsheets containing the associated documents, if any.

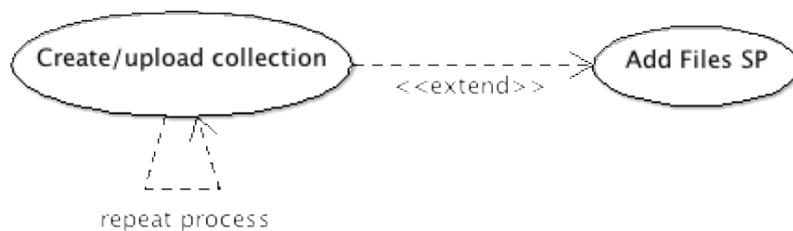


Figure A.41: New Collection use case

Continue collection creation from existing collection use case

This scenario represents the process of continuing working with an existing collection which does not have files associated with it. In this case is presented with a list of collections without files associated with them and he selects the one he would like to work with. Once this is performed, the user either could continue adding sub-collections to the pre-selected one or could directly start the process of adding files to the collection. This option is only available when working with an existing collection. If the user was creating a new collection, he has to create the complete collection structure and then he could add new files spreadsheets.

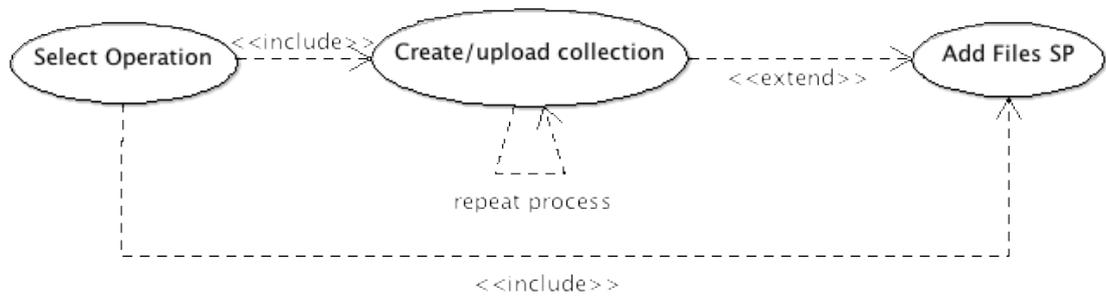


Figure A.42: Existing Collection use case

Uploading Files spreadsheets use case

Once the user has created the whole collection structure, he could then start uploading Files spreadsheets associated with any of the collections/sub-collections created in the previous processes. In this case, the user either could upload self-contained Files spreadsheets (they contain external links to the data being uploaded and accompanying the metadata records described in the spreadsheet) or upload complementary compressed files which contain the resources that are described and referenced in the Files template being uploaded.



Figure A.43: Upload Files Spreadsheet use case

A.4.2.2 Application Flows Description

The repository web application also uses Spring MVC and Spring Webflow frameworks. This application implements a main flow which manages the whole process of creating/updating collections into the Fedora repository. Additionally, it calls an additional sub-flow which manages the process of uploading complementary ZIP files containing data resources when working with files spreadsheets. All the scenarios are managed by one main flow due to the transactional nature of the application. The application presents a linear and constrained process of creating QuDEx collections to avoid the user performing operations in the wrong order. Therefore, only one main flow is implemented which presents with sequences of user views and operations to perform. For simplicity, the flow is described in terms of groups of operations. Each group presents a set of view-states and action-states that enable to perform the different application activities.

Collection update

This group of operations enable to perform metadata modifications over existing collection objects as well as deleting collections from the repository. This part of the flow presents a single view that enables the user to upload update collection spreadsheets. This type of spreadsheet is a collection spreadsheet, which presents specific qudex fields to perform update/delete operations over existing collections/sub-collections (see section 1 of Figure A.44). Once all the update operations have been performed then the process (and the flow) terminates.

Collection creation/modification

This group of operations within the flow enables either to perform a [QuDEX](#) collection creation (uploads of collections/sub-collections and associated files or documents) or to continue with a previous collection creation process (this refers only to collections previously stored that do not present an associated [QuDEX](#) instance document). In the latter case, the user could only associate files to the collection that he wants to update or could add new sub-collections associated with the selected one and after that include files in the collection (see section 2 of Figure A.44).

Files upload

This is part of the collection creation/modification processes. Once the collection structure has been created, one can associate files or documents objects to it. This group of operations enable to upload both files spreadsheets templates and compressed files with the associated data resources (see section 4 of Figure A.44). The upload of compressed files is optional and can be performed more that once. To perform the latter, the main flow calls the Upload sub-flow that manages the upload and operations over compressed resources files.

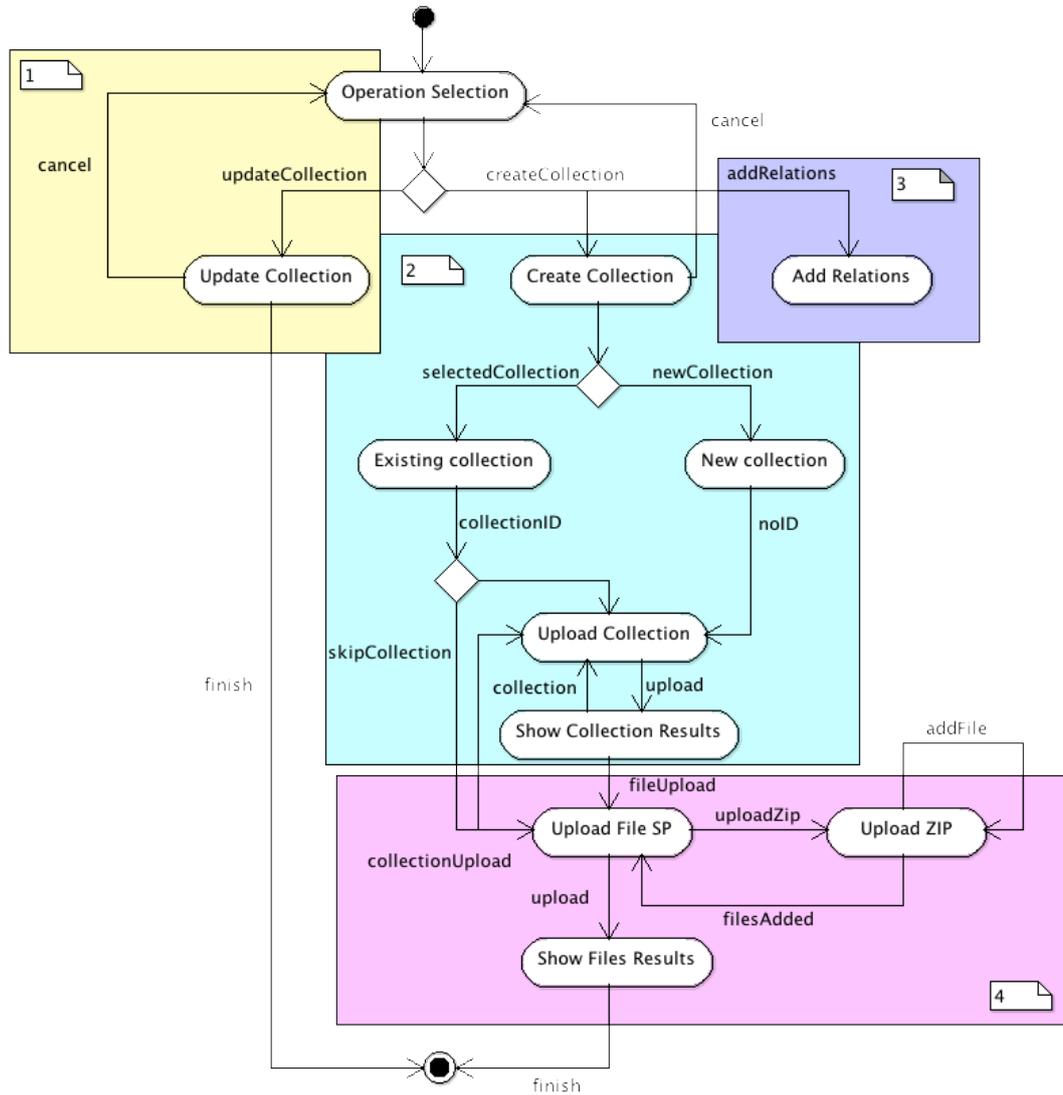


Figure A.44: Process Flow activity diagram

Upload Sub-flow

This flow represents the process of uploading ZIP files associated with a spreadsheet template. This flow is only presented if the user is uploading a files spreadsheet since this type is the only type that accepts physical files associated with each file being described in the template. It represents a sequential process of uploading ZIP files one at a time.

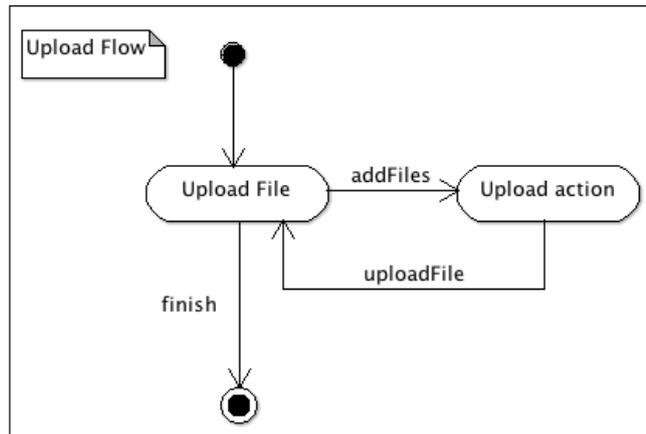


Figure A.45: Upload Flow activity diagram

Adding relationships to existing collection sub-flow

This sub-flow represents the process of adding relationships between files that belong to an existing collection. Once the user enters the flow the next step is to select the collection to which the new relationships are going to be added. The next step would be to select the resource that is the source of the relationship, then the type of relationship is selected and finally the target resource is selected. This process can be repeated as many times as relations to add. Once the user finishes the process, then the application returns to the initial state for selecting the different operations (Figure A.46).

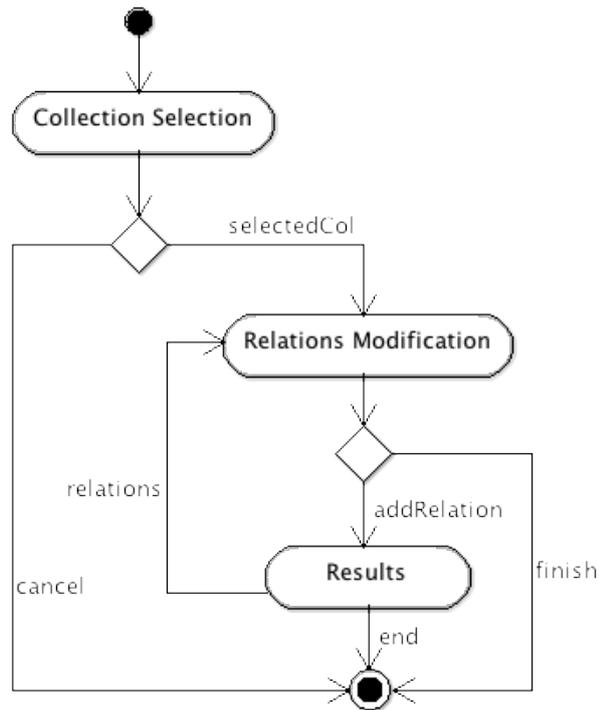


Figure A.46: Add Relations Flow activity diagram

A.4.2.3 Application Sequencing and Collaboration

The logic of the web application is implemented combining flows, services and action classes. The services are java interfaces that make available a set of methods to perform the application operations. The action classes are a special type of class available from Spring MVC framework that implements methods that can be called from web flows. They provide the same functionality as a service although they are classes specifically designed for their use within web flows. The repository application implements one service and three action classes to perform all the functionalities needed in the web flows. These will be described in detail in the “Application packages and classes description” section.

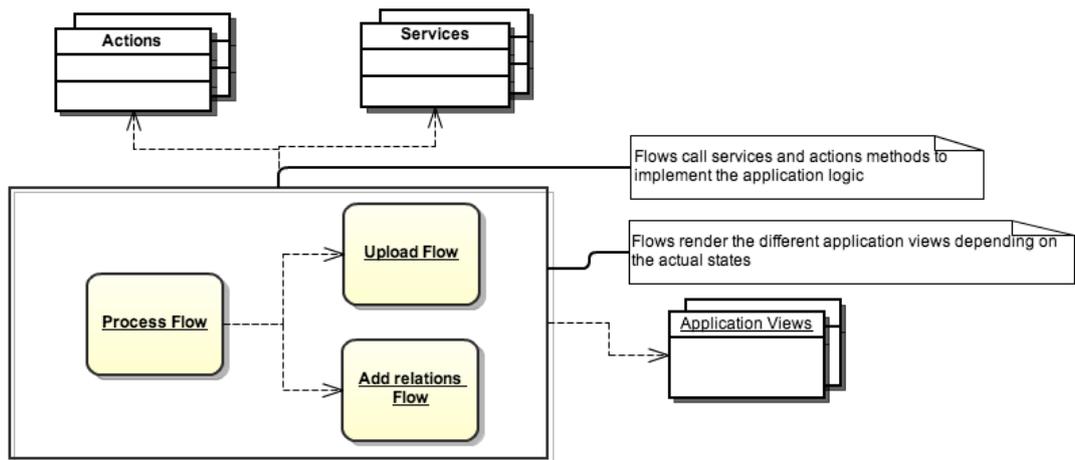


Figure A.47: Flows collaboration diagram

The application presents three different main scenarios that are managed by the main flow (Process Flow). The first scenario represents the process of updating or deleting existing QuDEX collections. This only involves the upload of modification spreadsheets (collection spreadsheet with update/delete fields). The second scenario represents the process of creating and uploading QuDEX collections into the repository. Depending on whether the collection creation is a new process or is a continuation process, the options presented to the user differ. Lastly, the third scenario represents the process of adding relationships between files that belong to an existing collection. The three processes are orchestrated by the same main flow.

Before entering any of the application's flows the user is presented with an introduction page that presents an overview of the different operations that can be performed via the web application. This page also provides access to more extended documentation pages and additional information about the metadata standards and technologies used in the application. These sections are managed by the spring main controller (implemented in the Spring MVC framework). When the user selects "Start process" then the control of the application is passed from the main controller into the Flows controller and then the application enters the main flow (Process Flow). The first view displayed is the "Operation selection" view. This interface asks the user to select one of the two options displayed: update existing collections or start a new collection creation process. In this scenario, multiple application classes are involved and the flow interacts both with model objects (FilesModel class) and Service objects in charge of the application logic.

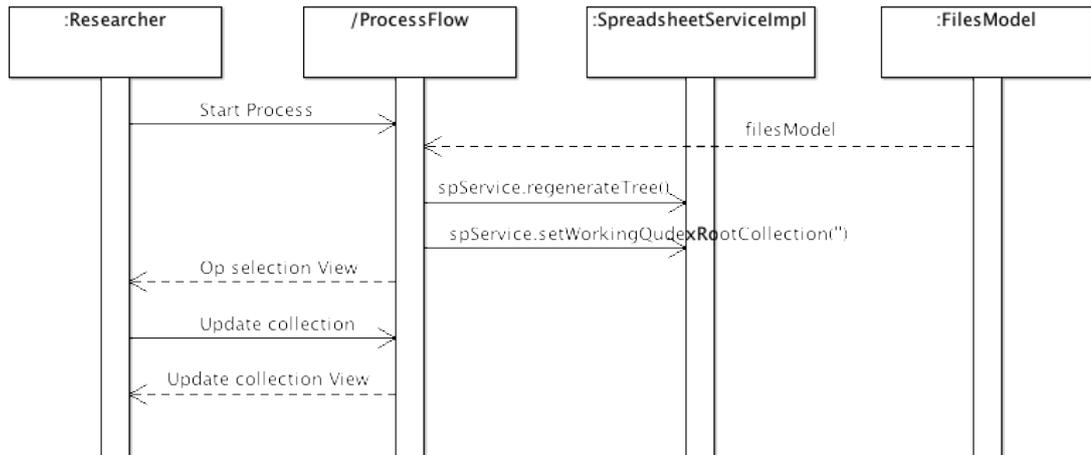


Figure A.48: Operation Selection sequence diagram

This is the first view launched when the user enters the flow and some update operations are performed before displaying the operation selection interface: initialise spreadsheet uploads management (FilesModel domain class), generate collections tree from Mulgara (regenerateTree method from the service class). In Figure A.48, the user has selected the “update collection” operation. Therefore, the next view presented is the one associated with update/delete operations over existing collections in the repository.

In the update/delete scenario the main flow performs operations over the repository by using mainly the following classes: the main service invokes the methods to perform operations in the repository (updates or deletes over collection objects); the FilesModel domain object is used to manage the upload of the modification collections spreadsheet so that it can be later used by the services to perform operations in the repository collections; and finally, the emailActions instance manages sending email confirmation with the results of the operations being performed.

In this scenario, the user is presented with the upload view, which enables to select a modification spreadsheet and upload it into the application. Once this is performed, the service invokes the methods needed to perform the operations in the repository and then a results view is displayed. From this view, the user can select either to receive a results email or to continue with more update/delete operations.

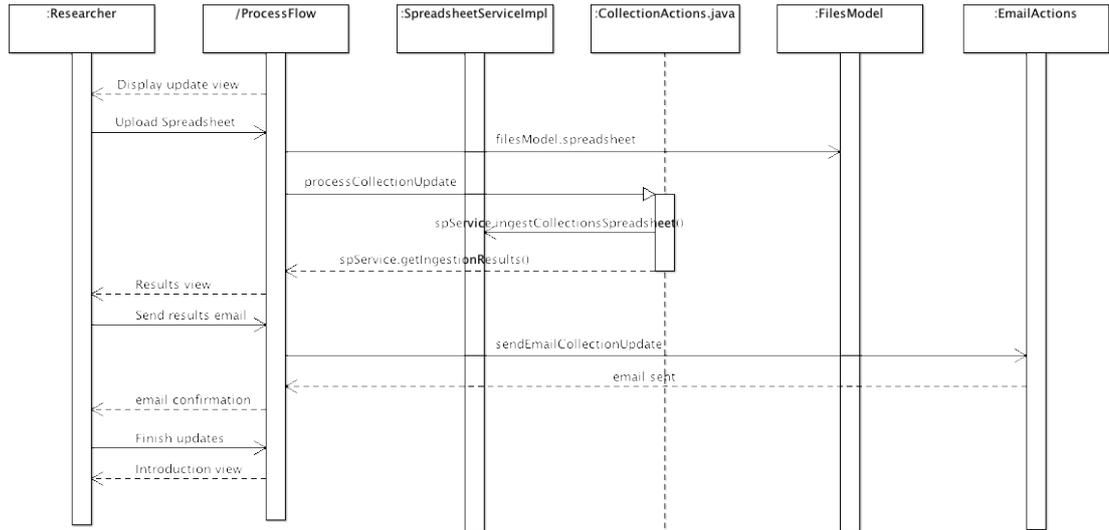


Figure A.49: Collection Update/Select sequence diagram

The second scenario includes all the collection creation operations. Depending on whether the user is editing an existing collection or creation a new one from scratch, the options presented and application flows differ slightly. These two different scenarios are described separately and they present their associated sequence diagram.

The third scenario includes the operations performed when adding relationships between two files within an existing collection. Firstly, the user is presented with the collection selection view. A list of all the collections included in the repository is presented and once the user selects the appropriate collection the add relationship view is then presented. The operations and services invoked by the flow to perform the operation for adding a relationship is shown in Figure A.50.

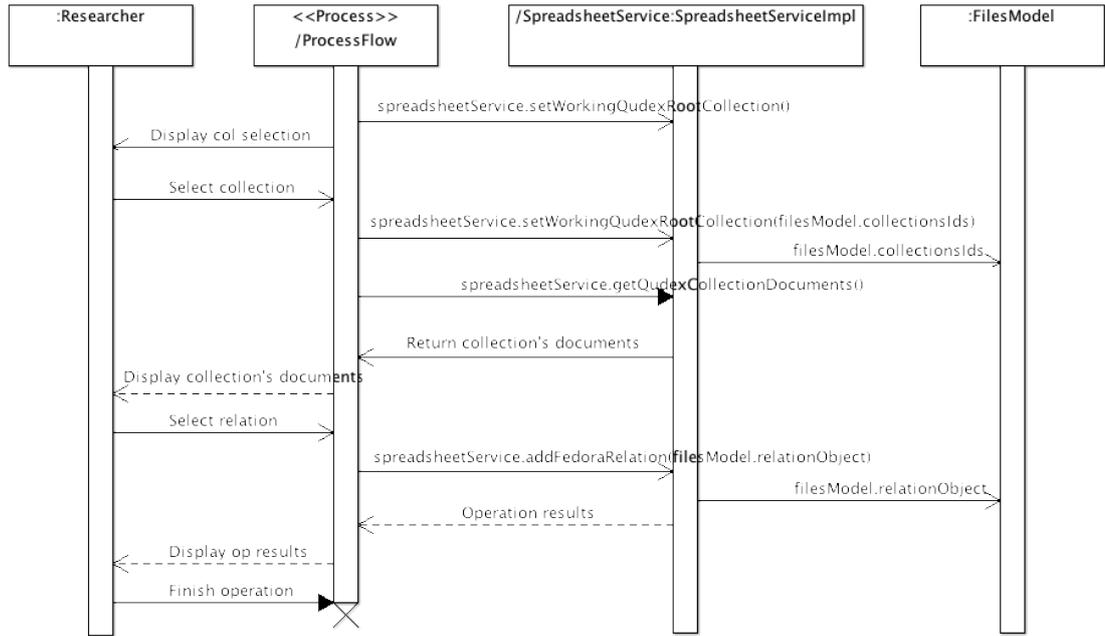


Figure A.50: Adding Collection Relationships sequence diagram

Modify and extend an existing collection

This scenario only applies to QuDEX collections stored previously but which do not include a generated QuDEX Instance. This means that the user previously has only created a collection/sub-collections structure in the repository and now wants to continue with this process either to add more collections/sub-collections or to associate files objects with any of the existing collections in the previously created QuDEX collection.

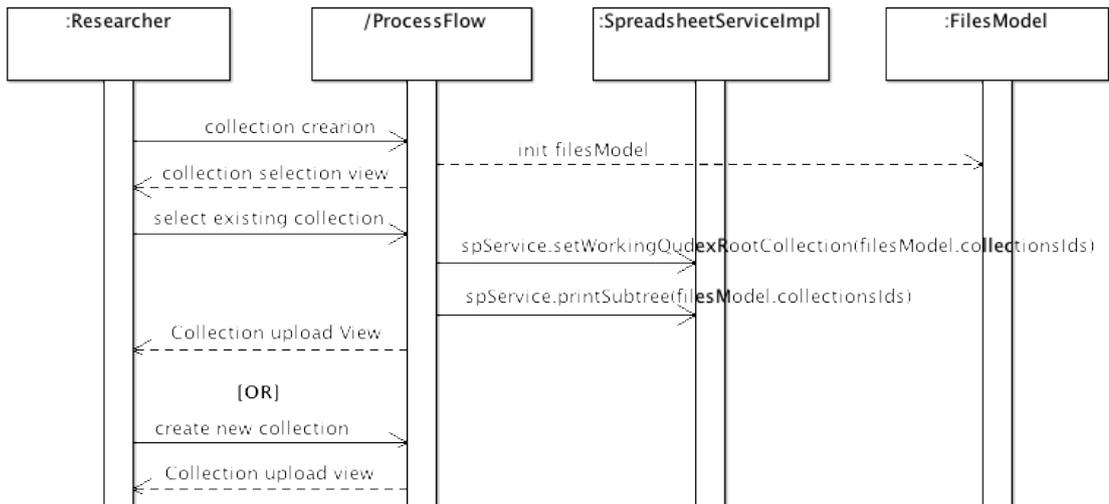


Figure A.51: Select Type of Collection Creation sequence diagram

The diagram above (see A.51) is common for both types of scenarios the generated view (Col-

lection upload view) is the same but with the difference that in the case of working with an existing collection, the user visualises the collection structure sub-tree and also has the possibility of skipping adding more collections, that is starting the process of uploading files spreadsheets.

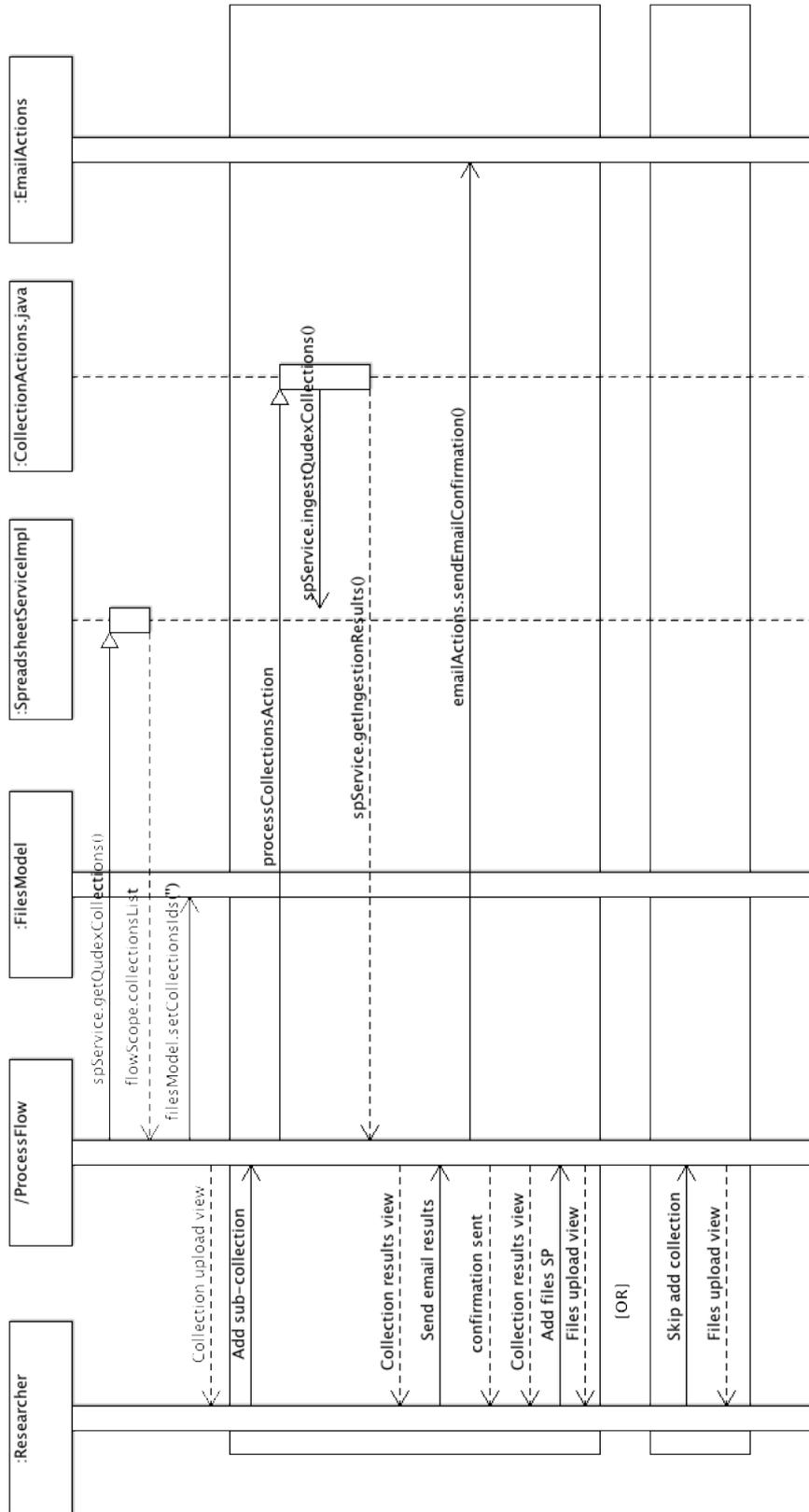


Figure A.52: Existing Collection Modification sequence diagram

When modifying an existing collection, the user has to select the collection container from the list of repository collections that do not have an associated QuDEX instance. Once the collection has been selected, the user is then presented with the selected collection's tree (collection structure or hierarchy in the repository) and he can start working with the new collection structure by adding new sub-collections to the existing hierarchy. Adding new collections is in this case optional; therefore, the user could opt for adding files spreadsheets directly. These two different situations are represented in Figure A.52 by an “OR” clause and two differentiated squared sections, since the initial methods invocations are the same in both situations.

When the user starts a collection creation process from scratch, he is redirected to the “collection upload” view and initially he can only upload a collection containers spreadsheet. This means that the spreadsheet will only contain one or more container collections, which will hold any sub-collections or files directly associated with them. Once the spreadsheet has been parsed and the collections generated and stored in the repository, the “Collection results” view is displayed where the user can opt to receive a confirmation email with the results and then go back to the collections upload process. In this case, he will not be redirected to the “Upload collection” view directly since he needs to select the collection container with which he will be working (the previously uploaded collections spreadsheet could contain one or more container collections, therefore he will be redirected to the collection selection view).

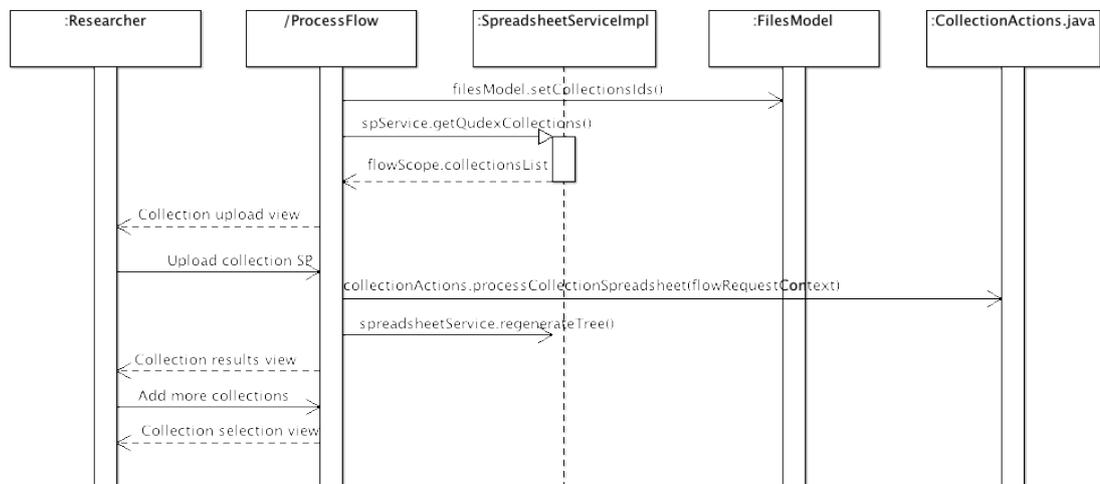


Figure A.53: New Collection Creation sequence diagram

A.4.2.4 Application Packages and Classes description

Actions Package

This package implements the Action classes. These classes extend from the “AbstractAction” class (included in the Spring MVC framework) and implement a set of methods used in the application web flows. The following classes are included in this package:

- CollectionActions
- EmailActions
- FileActions

CollectionActions class implements the “processCollectionSpreadsheet” method that performs the process of generating QuDEx collection objects and storing them in the Fedora repository instance. It also implements the method “checkSelectedCollections” which validates that the spreadsheet uploaded meets the right spreadsheet format and whether necessary, checks that there has been provided a parent collection to which the new collections will be associated.

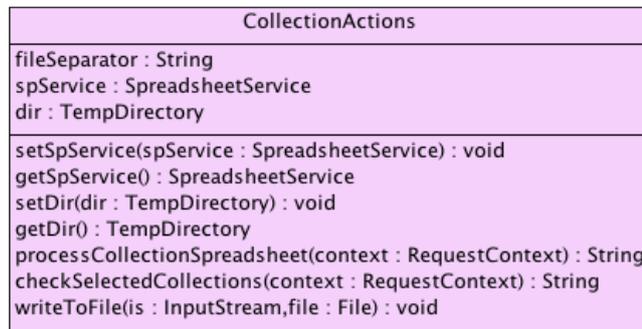


Figure A.54: CollectionActions class diagram

EmailActions class implements the “sendConfirmationEmail” method that generates a confirmation email that is sent to the user with the ingestion results and it provides the spreadsheet used as an attachment so that the user can then have records of the collection being created and the ingestion results.



Figure A.55: EmailActions class diagram

FilesActions class implements two methods that perform the processes of validating, parsing and generating Files objects associated with a given collection in the repository. These methods are: “processFileSpreadsheet” and “processFileSpreadsheetWithContents”. The first receives as input only the uploaded spreadsheet since it does not have physical data files associated with it. The second method manages the upload and parsing of both the files spreadsheet and the compressed files associated with it. In this case the spreadsheet contains references to the data files stored in compressed files, and therefore, they need to be uploaded and processed in the web server.

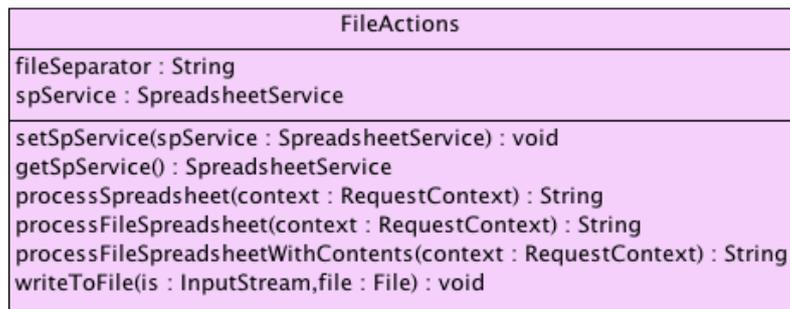


Figure A.56: FilesActions class diagram

Domain Package

This package contains the classes implementing the different data models used in the web application (Spring MVC environment). The application manages one type of object, which manages the different data types used in the web application, which are mainly files and arrays of files (spreadsheets uploaded and ZIP files when working with external data files). The next section describes in detail the functionalities implemented in the “FilesModel” class that implements the management of all the data files uploaded in the application. The class presents the following attributes:

- The attribute “option” is an integer, which represents the type of spreadsheet with which the application is working (0 for collections spreadsheet and 1 for files spreadsheet).
- The attribute “tempDirName” is a string, which represents the name of the directory used for temporary files uploads and processing.
- The attribute “collectionsIds” is a string, which represents a list of collections identifiers. These identifiers are the fedora identifiers of the collections that will be used as parent collections of the new uploaded collections or qudex documents.
- The attribute “modifiedSpreadsheet” is an inputstream, which holds the modified files spreadsheet once the external data files have been parsed. This spreadsheet is a files spreadsheet that includes the web paths of the associated files which have been upload in the web server.
- The attribute “spreadsheet” is a commons multipart-file object that holds the uploaded spreadsheet file.
- The attribute “zipFiles” is a list of commons multipart-file objects that holds all the different uploaded data ZIP files.
- The attribute “zipFile” is a commons multipart-file object used to handle individual zip files uploads. The process of uploading external data ZIP files is performed by uploaded

one file at a time, which is then included in the array of multipart-files (attribute “zip-Files”).

The class presents the following methods:

- All the getters and setters for all the attributes defined in the class.
- All the validation methods needed to validate correctness of each of the data types being handled in the class.
 - `validateCollectionUpload()`. It validates that the correct type of collections spreadsheet has been uploaded.
 - `validateSelectZipFile()`. It validates that the correct type of ZIP file containing external resources has been uploaded.
 - `validateShowResultsCollection()`. This method is used when sending email confirmations after having performed any management operation in the repository. It validates whether the provided email address is a valid email address.
 - `validateFileUpload()`. It validates that the correct type of files spreadsheet has been uploaded.
- The “init” method initialises all the attributes before loading the singleton instance of this class.
- The “destroy” method clears all the instances used by the class when the application is being unloaded from the applications server.
- The method “deleteZipFile” deletes a specific uploaded ZIP file from the array of ZIP files.
- The method “cleanZipFiles” cleans all the objects associated with ZIP files that have been uploaded into the application during the current session.
- The method “deleteTemporaryFiles” deleted all the files uploaded into the application’s temporary folder once they have been processed by the application.



Figure A.57: FilesModel class diagram

Controllers Package

This package implements the different controllers used by the web application to handle all the different requests from the user. Spring's web MVC framework is request-driven, designed around a central servlet (in the application's case this is the DispatcherServlet) that dispatches requests to controllers and offers other functionality that facilitates the development of web applications. This servlet is in charge of receiving the requests from the user and then direct them

to the appropriate controller. The web application is implemented combining Spring MVC and Spring Web flow. Each of the flows implements internally their own controller therefore; it was only needed to implement their associated handlers to deal with flow termination and exception management. The application only implements one controller handler class: `ProcessFlowController`.

This class extends from Spring's `AbstractFlowHandler` class and they implement two different types of method: `handleExecutionOutcome` (when the flow reaches a termination state this method checks whether it returns any objects and handles the next redirection) and `handleException` (method to manage any exception occurred during the flow execution). The `ProcessFlowController` class implements these two methods. Additionally, it implements additional exception handling methods to deal with exceptions that are thrown by business logic methods: “`handleFoxmlException`” (to handle exception handling when executing methods that deal with repository operations) and “`handleMulgaraException`” (to handle exceptions thrown by methods accessing and operating in the Mulgara triplestore).

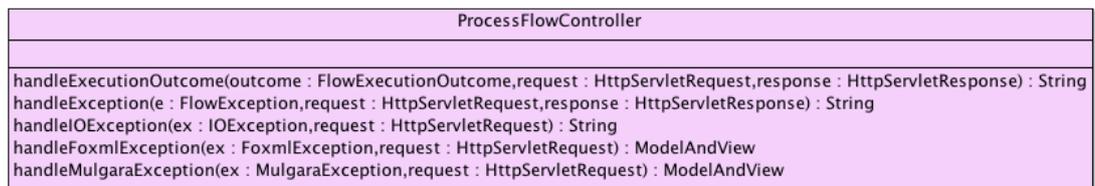


Figure A.58: `ProcessFlowController` class diagram

Listeners Package

This package contains application listener implementation classes. They are implementations of standard interfaces provided by both the Servlets Java libraries and Spring MVC framework. The application uses three different listeners:

- `ApplicationContextProvider`
- `ApplicationServletContextListener`
- `CustomServletContextAware`

Each of the classes is described individually below.

ApplicationContextProvider Class

This class provides an application-wide access to the Spring `ApplicationContext`! The `ApplicationContext` is injected in a static method of the class “`AppContext`”. This class is used to get access to all of the Spring Beans defined in the web application.

It implements the static method “AppContext.getAppContext()” which provides the context object that will allow to access all the declared beans in the application context.

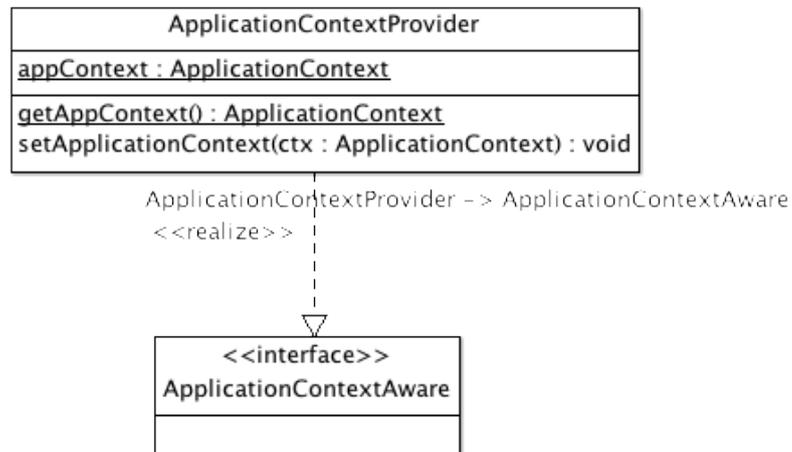


Figure A.59: ApplicationContextProvider class diagram

ApplicationServletContextListener Class

This class implements the interface “ServletContextListener”. Implementations of this interface receive notifications about changes to the servlet context of the web application they are part of. To receive notification events, the implementation class must be configured in the deployment descriptor for the web application. More concretely, this class provides implementations of the methods: “contextDestroyed” and “contextInitialised”. When the application is being reloaded or unloaded from the application’s server in where the application is deployed (Tomcat container for example) there are a few cleaning operations that need to be performed before the application is unloaded or reloaded. This methods are invoked from “contextDestroyed” which is fired just before the application destroys the context.

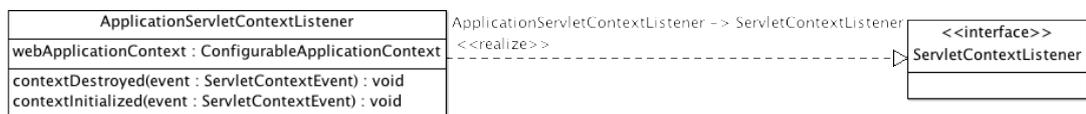


Figure A.60: ApplicationServletContextListener class diagram

CustomServletContextProvider Class

This class implements the interface “ServletContextAware” provided by the Spring MVC framework. This interface has to be implemented by any object that wishes to be notified of the ServletContext (typically determined by the WebApplicationContext) that it runs in.



Figure A.61: CustomServletContextProvider class diagram

Servlets Package

This package includes the servlets implementation handled by the web application. In this case, the application only uses one servlet implementation: `FileServlet` class. This class implements a servlet that is used to provide access to all the external data files that have been uploaded into the application. This servlet manages the temporary folder in the web application which stores all the decompressed files and that are associated with the current files spreadsheet being managed by the web application.

This servlet implements the following attributes:

- The attribute “tempDir” is a file object that represents the temporary directory used for managing the external files being uploaded into the application.
- The attribute “fileSeparator” is used to obtain the operating system internal file path separator, which normally differs depending on the operating system in which the application is running.
- The attribute “fileList” is a string array, which represents all the file names of the files included in the temporary folder.

The servlet class implements the following methods:

- All the getters and setters associated with the class defined attributes.
- `doGet()`. This method handles all the get requests received by the servlet. This method receives file retrieve requests. It searches in the temporary for the requested file and if found, it will return the outputstream associated with the requested file.
- `doPost()`. This method redirects a received post request to the “doGet” method which is charge of manage all the requests received in the servlet.
- `findFileInDirectory()`
- `processDirectory()`
- `getBytesFromFile()`

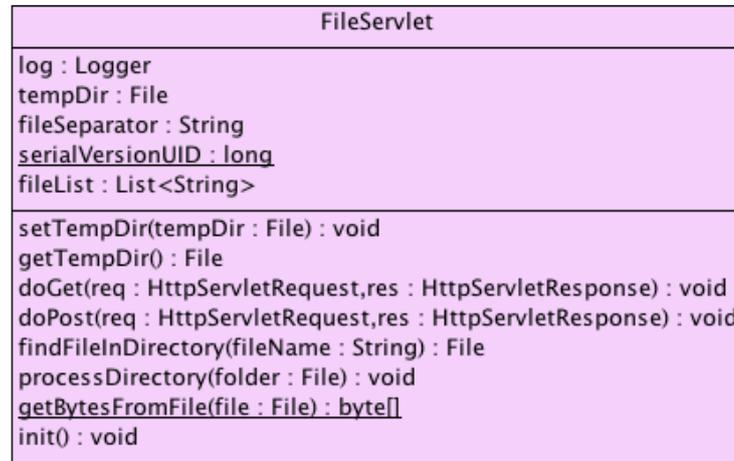


Figure A.62: FileServlet class diagram

Services Package

This package includes the service implementations that manage all the business logic of the application. It is implemented as interfaces and their associated implementations classes. The application implements two different services: “SpreadsheetService” and “DWRService”.

Both services are loaded as singleton instances (defined in the application Spring’s configuration file) once the application is started in an applications server like Tomcat and is auto-wired using the Spring mechanisms so that all the relevant classes in the web application, like web flows, request controllers can access them and invoke their methods. Figure A.63 shows the services classes and their relationships with other components included in the application (Service classes diagram).

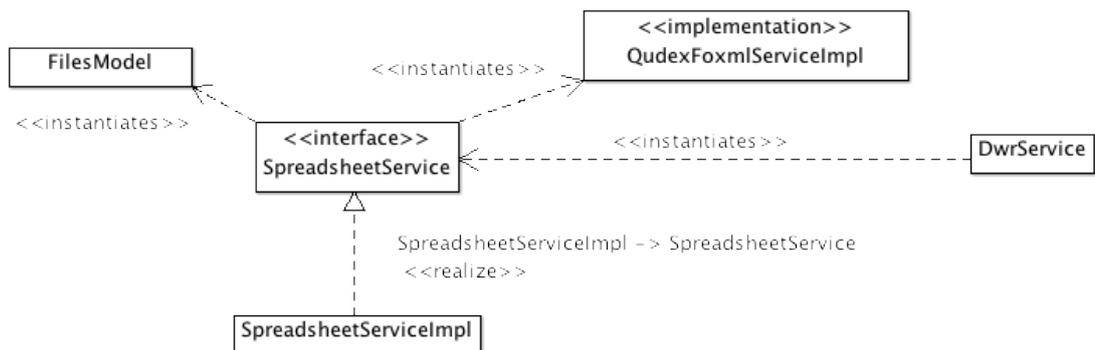


Figure A.63: Service classes diagram

The first service implementation (Spreadsheet service) provides all the methods to manage the repository operations (collection modifications and collection creation processes). It uses

the domain classes to generate and manage the data structures holding the files uploaded into the application. The class implements the following methods:

- `createFilesModel()`. This method creates and returns an instance of `FilesModel` domain class that will be used by this service to manage all the files being uploaded into the application.
- `ingestCollectionsSpreadsheet()`. This method parses a collections spreadsheet and performs all the operations needed to create the included collections in the Fedora repository being accessed by the application.
- `ingestFilesSpreadsheet()`. This method parses a files spreadsheet and performs all the operations needed to create the included files in the Fedora repository being accessed by the application.
- `getResults()`. This method returns the array of results objects from the operations performed in the repository.
- `getProgress()`. This method returns the percentage of progress of the current set of operations in the repository.
- `printSubtree()`. This method return the javascript code needed to print the collections tree in the current application view.
- `traverseTree()`. This recursive method generates the sub-collections tree
- `regenerateTree()`. This method regenerates the collections tree after new operations in the repository have been performed.
- `cleanDisplay()`. This method clears the array containing of the collections in the actual collections tree being displayed.
- `init()`. This method calls all the initialisation methods before loading the singleton instance of the class.
- `destroy()`. This method clears all the instances used by the class before unloading the application from the applications container.



Figure A.64: SpreadsheetServiceImpl class diagram

Utilities Package

This package implements all the static classes, which include utility methods and defined constants, which are accessible globally to the application. Each of the classes included in this package are described individually in the following sub-sections.

Constants Class

The class implements the following constants:

- The constant PID_HEADER is a string constant that specifies the spreadsheet column header of a fedora object identifier (collections or files objects).
- The constant DC_FORMAT_HEADER is a string constant that specifies the spreadsheet column header containing the Dublin Core “Format” metadata element.

- The constant DS_INLINE is a string constant that specifies the spreadsheet column header containing the metadata element used to represent the path to a data file associated with a particular file object being described in the spreadsheet and that will be stored within the fedora object as an inline datastream .
- The constant DS_MANAGED is a string constant that specifies the spreadsheet column header containing the metadata element used to represent the path of an external data file associated with a particular file object being described in the spreadsheet and that will be stored internally in the repository.
- The constant DS_EXTERNAL is a string constant that specifies the spreadsheet column header containing the metadata element used to represent the URL to an external data resource that will associated with the file object being described in the spreadsheet.
- The constant LABEL_HEADER is a string constant that specifies the spreadsheet column header containing the label associated with the object being described in the spreadsheet (this field applies to both files and collections)
- The constant EMAIL_DEFAULT_ADDRESS is a string constant that contains the default email address used to configure the email confirmation service implemented in the web application.
- The constant SUBJECT_EMAIL is a string constant that specifies the default subject field used when sending confirmation email after having performed operations in the Fedora repository.

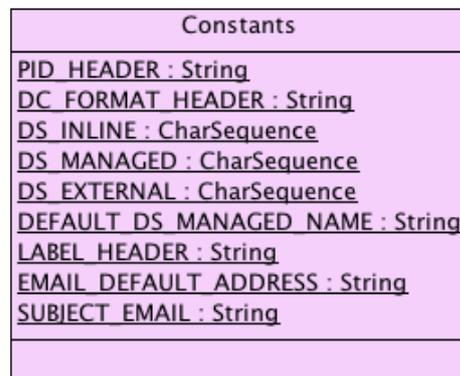


Figure A.65: Constants class diagram

ExcelManagement Class

This class implements all the auxiliary functions to manage Excel spreadsheets and it is used globally within the web application. The class implements the following attributes:

- The attribute “fileSeparator” specifies the file path separator used by the operating system in which the web application is running. It uses the Java system property “file.separator”.
- The attribute “fileList” represents a list containing all the names of the files used by the application and that are stored in the application’s temporary folder.

The class implements the following methods:

- fillExcelFile(). This method is used when working with files spreadsheet. It parses the spreadsheet and adds all the information associated with the external data files which accompany the uploaded spreadsheet.
- autoFillSpreadsheet(). This method is invoked by “fillExcelFile” and it implements the process of filling the spreadsheet with the information associated with the files accompanying the uploaded spreadsheet.
- findFormatElementExcel(). This method parses the spreadsheet looking for the Dublin Core “Format” column header.
- findDatastreamColumnExcel(). This method parses the spreadsheet looking for all the column headers that refer to types of datastream objects.
- findFileInDirectory(). This method is used to find a particular file within a file directory.
- processDirectory(). This recursive function processes all the files in a file directory.

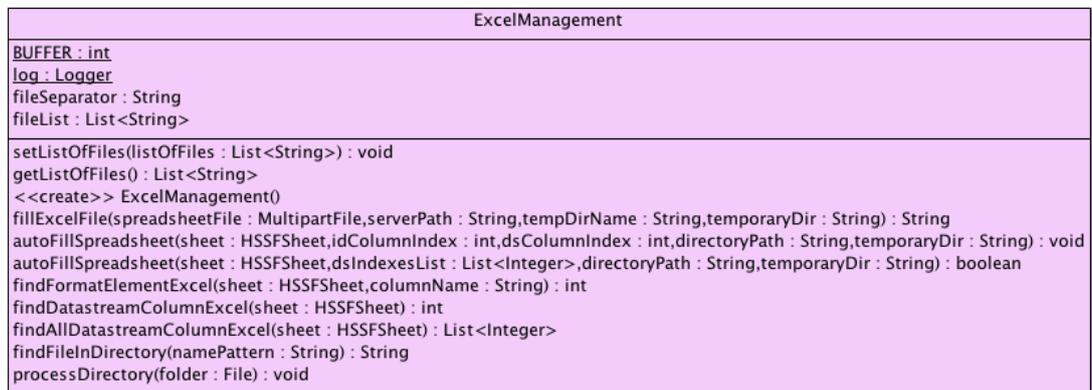


Figure A.66: ExcelManagement class diagram

GeneralUtilities class

This class implements general-purpose functions, which are used globally within the web application. The set of static methods implemented in the class are:

- createFile(). This method creates and returns a file object from a given file inputstream.

- `copyFile()`. This method returns a copy of the input file specified as parameter.
- `decompressZipFile()`. This method decompresses a given ZIP file.
- `deleteDirectory()`. This method deletes all the files included in a given file directory.

A.5 Toolkit's QuDEx Support

The QuDEx repository toolkit implements a data model that incorporates elements from two different schemas: Fedora's Digital Object Model and the QuDEx Schema (as described in Chapter 4).

In addition to the use of the QuDEx schema with data archiving purposes, the original main aim of the QuDEx schema was to provide an exchange model between the different packages for QDA or CAQDAS packages so that data analysis activities performed with a particular package (and the resulting data) could be imported in other packages. This is a key functionality for data sharing and re-use, therefore the toolkit generates a full QuDEx XML instance for each individual collection stored in the digital repository, in addition to its basic support for analytical functionalities. This XML instance is stored in the Fedora object which represents a qualitative collection, that is, the 'root' or high-level collection element from a particular archive. This XML file is accessible online, therefore in order to access and re-use it in CAQDAS packages that support QuDEx as an import capability, the user only needs to import this XML file into the QDA package.

The QuDEx instance shown in the next subsection represents a sample collection that contains 4 resources, represented both as "sources" and "documents" in QuDEx terms. Those documents have "categories" attached to them, and also "memos" (both analytical elements). Finally, the relationships between the resources in the collection are expressed by using QuDEx relationships:

- *isRelatedTo*, which is a 'document to document' relationship, expresses how a resource is related to another,
- *isTextualRepresentationOf*, which is a 'document to document' relationship, expresses how a transcription can be related to its associates audio/video file,
- *isDescriptionOf*, which is a 'memo to document' relationship, expresses how a memo annotation can be attached to a resource, and
- *hasDocumentation*, which is a 'document to document' relationship, expresses how a topics guide (study documentation) can be related to an interview schedule (data collection instrument).

A.5.1 Sample QuDEx Instance

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <qudex:qudex xmlns:qudex="http://www.data-archive.ac.uk/dext/schema/draft"
3   creator="AMG"
4   displayLabel="The Edwardians"
5   id="qudex-7b8ca9f0-eab3-46b7-bc9c-e5bd8b26f68a"
6   label="SN 2000"
7   language="en-GB"
8   status="open">
9   <qudex:resourceCollection
10     id="qudex-012a3674-030d-48dd-a47f-c2edd8d9851a">
11     <qudex:sources id="qudex-1a39938f-de70-4bae-8abd-1e18e90dca67">
12       <qudex:source checksumType="md-5" creator="AMG"
13         displayLabel="Interview Transcript with person 001"
14         id="qudex-6e50142c-9828-4789-9845-98f72a60e73c"
15         label="2000Int001"
16         language="en-GB"
17         location="2000Int001.xml"
18         locType="other"
19         mimeType="application/xml"
20         otherLocType="File"
21         otherResourceType="Interview transcript"
22         resourceType="other"/>
23       <qudex:source checksumType="md-5" creator="AMG"
24         displayLabel="Image of agriculture in Edwardian era"
25         id="qudex-dd961fd5-6ee7-4c04-86b4-0c3ec4a5456a"
26         label="2000_agriculture001"
27         location="2000_agriculture001.jpg"
28         locType="other"
29         mimeType="image/jpeg"
30         otherLocType="File"
31         otherResourceType="Visual data"
32         resourceType="other"/>
33       <qudex:source checksumType="md-5" creator="AMG"
34         displayLabel="Audio clip of interview for person 001"
35         id="qudex-5d4a43ae-4059-446f-b551-fc7854407a75"
36         label="2000_Int22a_clip1"
37         language="en-GB"
38         location="2000_Int22a_clip1.mp3"
39         locType="other"
```

```
40     otherLocType="File"
41     otherResourceType="Interview recording"
42     resourceType="other"/>
43 <qudex:source checksumType="md-5" creator="AMG"
44     displayLabel="SN 2000 user guide"
45     id="qudex-44d2c154-277d-489f-8c6d-d0b2491e2466"
46     label="SN 2000 user guide"
47     language="en-GB"
48     location="http://www.esds.ac.uk/doc/pdf/q5404uguide.pdf"
49     locType="url"
50     mimeType="application/pdf"
51     otherResourceType="User guide"
52     resourceType="other"/>
53 </qudex:sources>
54 <qudex:documents id="qudex-5ea7cde1-1045-4be4-949b-a5b82a82c215">
55     <qudex:document creator="AMG"
56         displayLabel="Interview Transcript with person 001"
57         documentType="source"
58         id="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
59         label="Interview Transcript with person 001"
60         language="en-GB"
61         resourceRef="qudex-6e50142c-9828-4789-9845-98f72a60e73c"/>
62     <qudex:document creator="AMG"
63         displayLabel="Image of agriculture in Edwardian era"
64         documentType="source"
65         id="qudex-6a595762-f358-4cfb-9bc9-aaa995594ade"
66         label="Image of agriculture in Edwardian era"
67         resourceRef="qudex-dd961fd5-6ee7-4c04-86b4-0c3ec4a5456a"/>
68     <qudex:document creator="AMG"
69         displayLabel="Audio clip of interview for person 001"
70         documentType="source"
71         id="qudex-da951f1e-3864-43fd-a0c2-cf1269b10a8a"
72         label="Audio clip of interview for person 001"
73         language="en-GB"
74         resourceRef="qudex-5d4a43ae-4059-446f-b551-fc7854407a75"/>
75     <qudex:document displayLabel="SN 2000 User guide"
76         documentType="source"
77         id="qudex-0fef1de5-071a-4077-8389-b7155c88647a"
78         label="SN 2000 User guide"
79         language="en-GB"
80         resourceRef="qudex-44d2c154-277d-489f-8c6d-d0b2491e2466"/>
```

```

81     </qudex:documents>
82 </qudex:resourceCollection>
83 <qudex:memoCollection id="qudex-30d3c700-9aef-4768-80f0-fff2eb2922d2">
84     <qudex:memo creator="AMG"
85         displayLabel="Image description"
86         id="qudex-54594c58-bf13-403c-b88e-35b5f80ab520"
87         label="Image description"
88         language="en-GB">
89     <qudex:memoText id="qudex-734def9b-87ba-47dd-883a-243b2e88b659">
90         Three generations of a family pose for a photograph in
91         front of the family home and business c.1904. Behind
92         them are various works in progress for these wheelwrights
93         including an urban district council wheelbarrow.
94     </qudex:memoText>
95 </qudex:memo>
96 <qudex:memo displayLabel="Image description"
97     id="qudex-6ae236ab-aac8-4128-9128-33171d529e47"
98     label="Image description"
99     language="en-GB">
100 <qudex:memoText id="qudex-4ab6c0dc-4563-4c48-a99c-901699030213">
101     Informal photograph taken outside Bicknacre post office
102 </qudex:memoText>
103 </qudex:memo>
104 <qudex:memo displayLabel="Historical note"
105     id="qudex-804d38f6-77d7-4c7d-b728-a793f71ffe60"
106     label="Historical note" language="en-GB">
107 <qudex:memoText id="qudex-3947cdc0-a65e-407d-8302-cfd508e94c22">
108     The fragility of the rural economy increased after
109     industrialisation as improvements in transportation
110     meant more produce was imported from abroad
111     rather than being produced internally. One response
112     can be seen in this retail outlet which acts not only
113     as a local post office but also draper, tea and
114     tobacco merchant and general grocery shop.
115 </qudex:memoText>
116 </qudex:memo>
117 </qudex:memoCollection>
118 <qudex:categoryCollection
119     id="qudex-e5b34fdf-784b-4b03-b682-9a08f4fbe3ac">
120 <qudex:category creator="AMG"
121     displayLabel="2000Int001"

```

```
122     documentRefs="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
123     id="qudex-14fdd873-fa3d-4ca2-b257-d79a99d07e08"
124     label="2000Int001"
125     language="en-GB"
126     categoryScheme="idno - CV"
127     categoryType="primary"/>
128 <qudex:category creator="AMG"
129     displayLabel="male"
130     documentRefs="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
131     id="qudex-80d10f87-5822-4bf1-b6fb-611b129a6ded"
132     label="male"
133     categoryScheme="sex - CV"
134     categoryType="primary"/>
135 <qudex:category creator="AMG"
136     displayLabel="Mr"
137     documentRefs="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
138     id="qudex-0b6d1e83-a356-4d33-96c5-c2fda6480766"
139     label="Mr"
140     language="en-GB"
141     categoryScheme="Title - CV"
142     categoryType="secondary"/>
143 </qudex:categoryCollection>
144 <qudex:relationCollection
145     id="qudex-7738a90d-751b-405e-8898-66e236c7309d">
146     <qudex:objectRelation
147         displayLabel="Relationship between resources"
148         id="qudex-da38a366-3cbf-4882-9b31-0606f9d9b35c"
149         label="Relationship between resources"
150         language="en-GB"
151         objectSource="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
152         objectTarget="qudex-6a595762-f358-4cfb-9bc9-aaa995594ade"
153         objectType="documentDocument"
154         relationName="isRelatedTo"/>
155     <qudex:objectRelation
156         displayLabel="Transcription to audio clip"
157         id="qudex-f6d77ff1-acbf-4a65-9bfd-f4c04a508e27"
158         label="Transcription to audio clip"
159         language="en-GB"
160         objectSource="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
161         objectTarget="qudex-da951f1e-3864-43fd-a0c2-cf1269b10a8a"
162         objectType="documentDocument"
```

```
163     otherRelationName="isTextualRepresentationOf"
164     relationName="other"/>
165 <qudex:objectRelation
166     displayLabel="Relation with documentation"
167     id="qudex-23da7c59-dd5f-498a-b1c7-280a4c9d12d2"
168     label="Relation with documentation"
169     language="en-GB"
170     objectSource="qudex-52ee85c4-9454-4bb3-854b-b0c0ff9b3806"
171     objectTarget="qudex-0fef1de5-071a-4077-8389-b7155c88647a"
172     objectType="documentDocument"
173     otherRelationName="hasDocumentation"
174     relationName="other"/>
175 <qudex:objectRelation
176     displayLabel="Memo to document"
177     id="qudex-445d8d46-710c-4c0d-a870-8d81edeb8098"
178     label="Memo to document"
179     language="en-GB"
180     objectSource="qudex-54594c58-bf13-403c-b88e-35b5f80ab520"
181     objectTarget="qudex-6a595762-f358-4cfb-9bc9-aaa995594ade"
182     objectType="memoDocument"
183     otherRelationName="isDescriptionOf"
184     relationName="other"/>
185 </qudex:relationCollection>
186 </qudex:qudex>
```

Notes

¹<http://www.springsource.org/>

²FOXML is one of the various XML schemas that Fedora supports to represent digital objects in the repository. These schemas are used for ingestion, update operations and export/import of objects in the repository. More information on the schema: <https://wiki.duraspace.org/pages/viewpage.action?pageId=30221134>

³More information on the standard: <http://www.uml.org/>

⁴For more detailed information about operations and parameters refer to the Fedora API: <https://wiki.duraspace.org/display/FEDORA36/REST+API>

⁵Apache HttpClient Library: <http://hc.apache.org/httpclient-3.x/>

⁶ITQL query language is the Mulgara proprietary language to query the triplestore. It is similar to SPARQL but implements additional functions. More information about ITQL available at: <http://code.mulgara.org/projects/mulgara/wiki/TQLUserGuide>

CADQAS DESIGN APPROACHES

Early CAQDAS packages were specifically developed to store, organise and manage large amounts of qualitative data (mostly textual, at the time) with analytic purposes. Developments in information technology like storage systems and databases, data formats and standards, and internet-based information systems, enabled an increase in the range of software packages and, similarly, the range of functionalities within each software tool. These packages were classified into two major groups (Lewins, 2001):

- ‘Text retrievers’ and ‘Text-based Managers’, which were mainly concerned with quantitative aspects derived from qualitative data, such as automatic generation of word/phrase indexes, statistical information on word frequency and the retrieval of text in context, and
- ‘Code and Retrieve’ and ‘Code-based Theory Builders’ which were oriented towards thematic analysis and interpretation of textual data.

Traditionally, the latter group of QDA tools - as introduced in Chapter 2 - were the product of attempting to solve individual research issues or enhance certain aspects of qualitative data analysis, subsequently their design was mainly driven by specific research practices or approaches (see last two columns of Table B.1 below). Table B.1 summarises the functionalities, design, and research methodological approaches of the main CAQDAS packages. Different analytical approaches involve different steps and make use of different analytical data, consequently the data model for capturing the analysis process, what information is gathered and what elements are involved, differs considerably among QDA tools. The analytical approach of choice has an impact on the application design, both in terms of the functionalities and more importantly on the data model underpinning the software. The NVivo package, formerly N*DIST (Table B.1, row 6), provides a good example of this in that the underpinning methodological approach (Grounded Theory) requires a clear separation between the data generated from the analysis process and the primary data. The tool’s data model had to be designed in

such a way that it enabled researchers to keep track of every analytical step and the data it generated. Methodologies such as Grounded Theory (GT), which describe the analysis process in great detail, can facilitate an accurate implementation of analytical functionalities that align with researchers' practices of QDA. Consequently, those methodologies were often chosen by tools' designers to underpin CAQDAS technological developments. Moreover, many developers of software for supporting qualitative analysis, who are searching for a methodological underpinning, usually draw on the methodology of GT as one of the most well-known and most explicit approaches in qualitative analysis (Kelle, 1997) since it provides a very rich description of the analysis process (Strauss and Corbin, 1998). Here the primary method of analysis is a continuous coding process (Creswell, 2007, p.67). Firstly, 'open coding' where the information is categorised and its properties and dimensions are examined. This is followed by 'axial coding', which consists of interconnecting the main categories that have emerged from open coding of the data. Finally 'selective coding' is applied to the data to identify a story line and write an account that integrates the categories from the axial coding. GT involves the discovery of theory through the systematic analysis of data. Categories and concepts created from the source data are inspected, annotated with comments and re-structured in multiple ways, allowing the creation of 'snapshots' of the different stages of the analysis; as opposed to other approaches, such as narrative inquiry or conversation analysis, in which the researcher chooses a theoretical framework that is applied to the phenomenon to be studied. In the latter approaches, the analysis is more of an interpretive process where one focuses on the meaning emerging from the data: the coding process is an instrument to track how an interpretation emerges from the original data and evolves into a finding through several levels of aggregation and abstraction. The purpose of codes is to create annotations of the qualitative materials that are aggregated and evaluated within the context of the case study, that is, the documents from which the codes have emerged, rather than aggregating and evaluating them by their separation from their documentary and case study context.

Table B.1: CAQDAS comparison

| CAQDAS | Functionalities | Import/Export | Design principles | Methodology |
|-----------------|--|---|---|---|
| <i>Atlas.ti</i> | <ul style="list-style-type: none"> • 'Hermeneutic Unit': all project information handled in a single file. • Data files (documents) kept in an external database. • Handles multimedia data: audiovisual, audio. • Central work around documents, memos, codes, quotations. | <ul style="list-style-type: none"> • Full XML export • Data lists exported into common formats such as PDF, Excel, or CSV. | <p>The package was originally developed in the context of an inter-disciplinary project to support textual interpretation and analysis. Its original design was also heavily influenced by knowledge acquisition tools and approaches such as KADS¹. With regards to the data model, it is based on a network model, where every object included can be connected to any other object: e.g. connecting audio with a transcript, or memos with document segments. The user can define relationships and use them to link data within the project.</p> | <p>The package was originally designed to support highly hermeneutical approaches in which the core functionalities were around text interpretation. In more recent versions its use has been extended to other fields of research such as software engineering methodologies or Knowledge Management Systems (KMSS).</p> |
| <i>Dedoose</i> | <ul style="list-style-type: none"> • Project organised in customisable 'context specific' workspaces, e.g. topical, by type of media. • Codes Panel (central feature): creation, import and management of code lists. • Textual and multimedia data types support as well as basic support for surveys. | <ul style="list-style-type: none"> • All items (incl. visualisations) can be exported into common formats: PDF; Word or Excel. • Proprietary metadata vocabularies used. • Lack of full project export. Only output reports. | <p>The first version was specifically designed to support the concurrent analysis of large amounts of mixed data collected by teams of geographically dispersed researchers working collaboratively.</p> | <p>The support for mixed methods approaches was one of the original aims while developing early versions of the software (Lieber et al., 2003). There was also an emphasis on research collaboration, therefore the package was online-based to enable multiple researchers accessing to and working with the same project. Additionally, code-based approaches were also supported although, unlike most of the current CAQDAS packages, the coding functionalities are more oriented towards quantitative approaches.</p> |

¹<http://www.commonkads.uva.nl/frameset-commonkads.html>

Table B.1: (continued)

| CAQDAS | Functionalities | Import/Export | Design principles | Methodology |
|-----------------|---|--|---|--|
| <i>Transana</i> | <ul style="list-style-type: none"> External DB structure (3 components): the application, the audio/video files and the database (transcripts, clips, codes). Transcription tools: multi time-coded transcripts linked with multiple videos or clips. Collection organisation based on the audiovisual materials: collection, series, episodes. Data interrogation keyword-based searches. Searches can be saved and used to create collections or materials. | <ul style="list-style-type: none"> Reports, maps and graphs based on collection items. Lack of full project export; only output reports. | <p>Transana was created specifically with conversation analysis in mind, however it has features that make it useful for other forms of qualitative interactional and language-based analysis. It supports interpretation of data from multiple perspectives; focus on interaction and non-verbal communication.</p> <p>The package has a multi-user version to support team work and projects involving large amounts of data. Multiple researchers can be analysing the same or different data concurrently: for example, its support for handling multiple transcripts around the same video material facilitates collaborative work. However, there is no mechanism to track researchers' interactions and 'who has done what'.</p> | <p>Transana original design was based on education research approaches involving the use of audiovisual materials to build case studies, and document educational practices.</p> <p>The main approach is visual methods (video is central here), and therefore, the support for its combination with other sorts of materials was not a requirement. Indeed, even with recent versions, the support for this is still limited.</p> |

Table B.1: (continued)

| CAQDAS | Functionalities | Import/Export | Design principles | Methodology |
|---------------|--|---|--|--|
| <i>MaxQDA</i> | <ul style="list-style-type: none"> Internal Database (DB) system where data files are kept together with project and analysis information. Support for both qualitative and quantitative data analysis. Coding system is central to the application. Annotation tools to assign memos and codes to every data object. Support for GeOLINKs: data segments or codes can be linked to Google Earth positions. | <ul style="list-style-type: none"> Export of segments into TXT, XLS or HTML. Partial export of project elements into XML or spreadsheet. Combined reports including coded text segments, memos, images from visualisation tools. | <p>The package was originally designed to support analysis of qualitative data. Focuses on large amounts of interviews - open ended interviews. The core functionalities of the software are attributes function - quantitative variables is the terminology used - rather than 'coding and retrieve'.</p> | <p>Unlike most of the CAQDAS packages, its design was not developed on the background of GT. Instead, it was designed to work with both qualitative and quantitative data and mixed-methods research approaches.</p> <p>The methodological background of the tool is providing support for the combined use of qualitative and quantitative data, highlighting three approaches:</p> <ul style="list-style-type: none"> Triangulation - combination of methodologies in a study, Mixed Methods - different types of research designs involving different methods, and Methodological Integration - a methodological framework that gives orientation for the selection of methods and techniques of data collection and analysis (Kelle, 2001). |

Table B.1: (continued)

| CAQDAS | Functionalities | Import/Export | Design principles | Methodology |
|------------|--|--|--|--|
| <i>DRS</i> | <ul style="list-style-type: none"> External DB in which analytical information is stored and links between the files - stored externally - are created. Multiple sets of analyses can be stored for a single project (longitudinal projects). Transcription tools including time-coded embedded editor or the possibility of imported transcripts from other packages. Data formats comprise both textual and multimedia: txt or html, audio/video and still images. | <ul style="list-style-type: none"> Matrix or tabular export including text, code and timing information reusable in applications such as Excel or SPSS. Proprietary metadata that can only be used internally and it is not indexed for interrogation purposes. Intentions of using meta-data standards to add richer descriptive information to the outputs of export. | <p>The package was developed as a research tool to explore heterogeneous data (focusing on audio-visual materials, still images, textual data). Its design and implementation are underpinned by the use of grid computing technologies applied to qualitative datasets.</p> <p>While grid computing is primarily focused on high performance computing, therefore leading more naturally to quantitative data, it raises new possibilities for Social Sciences (SS) research.</p> | <p>The package was designed for e-social science approaches, and it combines the affordances of grid computing with more traditional qualitative research methods. As a result of this, the research data collection instruments are underpinned by the concept of 'digital records' (Crabtree et al., 2006), which combines external records such as video, field notes, photographs gathered by the qualitative researcher; and internal records which include the set of digital media (text messages, voice mails, or emails) generated from users within electronic media environments,</p> |

Table B.1: (continued)

| CAQDAS | Functionalities | Import/Export | Design principles | Methodology |
|--------------|--|---|--|--|
| <i>NVivo</i> | <ul style="list-style-type: none"> Internal or external DB either holding both project information and data files, or only project information and links to external data files. Data is organised into units or 'cases' which include different sets of data. Data within or from different cases can be grouped in different ways. Support for hierarchical/non-hierarchical coding schemas, which can be organised in different ways (e.g. thematic, based on relationships, or according to data features). Data can be linked using different mechanisms: memo linking, text segments, external hyperlinks or relationships between data items. | <ul style="list-style-type: none"> Lack of full project export; only output reports. Output reports only exportable into Microsoft applications: Word, Excel. | <p>The first version, NUDIST (Non-numerical Unstructured Data Indexing Searching and Theorising), was primarily designed to support the analysis of unstructured data.</p> <p>The underlying model is based on three components: document system - processing and management of the research data, indexing system - holds references to text units within the research data or documents and analysis system - central part that manipulates the indexing system to process and create categories designed to help researchers defining and exploring research ideas</p> <p>A key design principle of the database is to keep separate DBs for the documentary data and the actual data themselves, rather than keeping individual records containing the textual and analysis data on individual documents. The indexing system of the DB attempts to represent the 'researcher's thinking about the project' so that that 'image' is available for exploration and interrogation.</p> | <p>The design of the package aimed to support for GT approaches, based on an early research project which was set out to support a range of methods for the analysis of unstructured data.</p> <p>The 'code and retrieve' capabilities differ from other packages in that there is a focus on creating theory from data rather than testing existing theories. This process follows a particular analysis approach in which there is a separation between the data produced from analysis and the original research data collected.</p> <p>There is a shift from "marking and finding text" approaches to an approach where the inspection of categories and concepts emerging from the analysis are central. Those categories and concepts can be annotated with comments and re-structured in multiple ways.</p> |

Sources: (CAQDAS Networking Project, 2012; Muhr, 1991; Lieber et al., 2003; Halverson et al., 2012)

Most of the approaches to qualitative data analysis share the processes of collecting and documenting data, organising/categorising these data into concepts and connecting data to show how concepts influence or relate to others (Schutt, 2012), thus some sort of coding/annotation procedure is used during analysis. This is one of the reasons why “coding and retrieval” is the feature that most of the CAQDAS packages have in common. However, the affordances of new technologies along with new research methodologies have facilitated an increasing degree of sophistication of the functionalities of the different tools. Whilst coding and support for annotations is a central feature, features such as those providing support for working with a wide variety of formats other than textual; performing analysis using different methods (qualitative and quantitative research instruments) and visualising results in integrated ways; and collaborative and multidisciplinary work, have started to be a requirement.

The growing number of CAQDAS adopters and diversity of types of users have also contributed importantly to an evolution of CAQDAS more generally, and have had an impact on how the tools are designed. There has been a shift from developers leading the design of the software into a type of design in which anticipating the user requirements and understanding the social processes involved in facilitating/inhibiting CAQDAS tools adoption becomes considerably more difficult owing to the diversity of users, new applications of the software and their methodological implications. In this respect, two quite important studies were conducted in the 1990s (Mangabeira et al., 2004; Fielding and Lee, 2002) in the UK, one studying an academic setting (Higher Education Institution) and exploring how user generation, in terms of age, level of computer literacy, and experience in qualitative research, affects the take-up and use of CAQDAS. A second one explored the growth of software use by researchers engaged in qualitative work in non-academic settings, to provide a typology of the different users by comparing their views about CAQDAS capabilities - who use CAQDAS and what functionalities they use. The typologies of users, the patterns of use and the associated social dynamics play a very important role on the directions of the software design and development but it also raises critical issues around research methodologies and researchers’ practices. Mangabeira et al. (2004) identified three main groups of academic users adopting CAQDAS tools: early career researchers who are less experienced in qualitative research but who are familiar with the use of software tools; experienced researchers that have been using QDA software in the past; and experienced researchers, less computer literate, that are more familiar with manual analysis of qualitative data.

For the first group of users, the dissemination of CAQDAS packages by some of the developers but more importantly, the social and cultural processes surrounding existing use of the tools - graduate student networks or informal exchanges between academics - were the grounds for software adoption. However, the latter interfered with wider research methodological considerations. For this type of users, the use of the tools was justified in terms of community usage, the software capabilities - speed up manual processes and work with large amounts of data - and the transparency and robustness added to the analytical processes. This was perceived as a mechanism to prove the validity of their work, rather than in terms of methodological con-

siderations, such as theoretical sampling or the rationale for the size of data linked to the overall research design. This type of users presented a lack of a critical stance towards the advantages/disadvantages of using a specific software package.

In contrast, researchers with previous experience on using CAQDAS packages had a critical ability that was related to methodological procedures and the affordances of the software. These users were able to perform a selective usage of the different functionalities included in the packages. More importantly, they were aware of epistemological and methodological issues present within the use of particular software tools, such as methodological claims of the tools differing from what the users perceived as required of qualitative data analysis. Additionally, this selective use of the tools reflected a criticism of the greater and increasing range of functionalities of specific software packages, which was, and still is, one of the trends of development within CAQDAS packages. To them, what was required was the use of a selection of features from a set of packages, like for example, using one tool for the organisation and filtering of large amounts of data and other tools for performing exploratory work and interrogating the data rather than using most of the functionalities of a single tool. Lastly, the third user group (researchers familiar with manual analysis of qualitative data) provided an important critical and reflective awareness about the use and design of different packages, despite their lack of familiarity with software use. Their richer knowledge of different research methodologies and analytic experience enabled them to identify how the use of some of the QDA tools could constrain analytical tasks owing to the way the software was designed. The ways in which analysis tasks were designed within the software could impose specific data structures, like organising data in hierarchical ways, which could drive how researchers work with their data and conduct their analyses.

Although CAQDAS packages had their origins primarily in the academic community, there has been an increase on their use by non-academic users, highlighting those engaged in applied research and those involved in research whose focus is not social science (Fielding and Lee, 2002; Baugh et al., 2010). This newer group of users, users inexperienced both as researchers and QDA software users, raises a number of important issues. Some of these issues are similar to those present within a more academic use of CAQDAS in that users might adopt a particular approach to analysis without being fully aware of the range of approaches available. This is partially due to them having a limited background in qualitative methods, but also because the particular package they have chosen points them in a particular direction. Other issues that arise are the product of the analytical and procedural differences between applied research and more pure qualitative research. In applied research, the approaches to, and procedures for, analysis can be less complex than in qualitative research. The latter, in combination with the need for working with vast amounts of data and producing research results in relatively short periods of time, leads to the use of CAQDAS packages with data management purposes rather than with more analytical and conceptual ones (Fielding and Lee, 2002).

The increasing spectrum of CAQDAS users (academic communities and users in applied research); an increasing use of multiple method studies that require more complex analysis of larger amounts of data; or Internet-based research in which the analysis of traditional primary data is combined with data available online and from multiple sources, widens the spectrum of CAQDAS features and places new requirements on their design. New CAQDAS developments are moving away from designs based on specific research methodologies (as originally) to designs which provide program versatility as opposed to specific use of the software. These design approaches aim to support both the more academic community, which bring into the design the knowledge of qualitative research methodologies and analytical practices, and the applied research community, which provides new uses of the software and that also requires support for different research practices.

RESEARCH DESIGN

C.1 Study Topics

Key areas related to digital archiving of qualitative research data have been explored in-depth in the course of this study. Figure C.6 shows a concept map in which the most relevant aspects for each of the areas that have been explored are summarised. The main areas explored are the following:

- *Digital archiving technologies.* Data documentation and access, digital repositories development and data exchange are covered. These areas are key in relation to qualitative data archiving and the development of accessible tools that are open-source so that they can be re-used and extended. Open descriptive standards, which are part of data documentation, cover those vocabularies or terms that are suitable and widely adopted to describe research data.
- *Qualitative Data Analysis Software.* There is a cross-over between the areas of digital archiving and developments within QDA software, concretely those related to data exchange formats and open descriptive standards. The key current software packages for QDA have been explored, with a focus on the functionalities (particularly those related to import/export of analysis data), and the research methodological approaches that underpinned the design of those software tools.
- *Research Data Management.* Data management practices, and research data lifecycle, are a central element of digital archiving. They cover aspects, such as archiving planning, ethics and confidentiality, data sharing and re-use, and implications for researchers, that need to be considered when implementing archiving tools, especially when the nature of the data (as in the case of qualitative data) presents challenges for their re-use and sharing.

- *Qualitative Inquiry and Research Practices*. With a focus on secondary analysis, data sharing and re-use; and covering those aspects that are related to research methodologies and research practices within qualitative inquiry to gain a better understanding of the present barriers/enablers of digital archiving for qualitative research.

Figures C.2, C.3, C.4 and C.5 expand the topics covered in the different identified sub-areas.

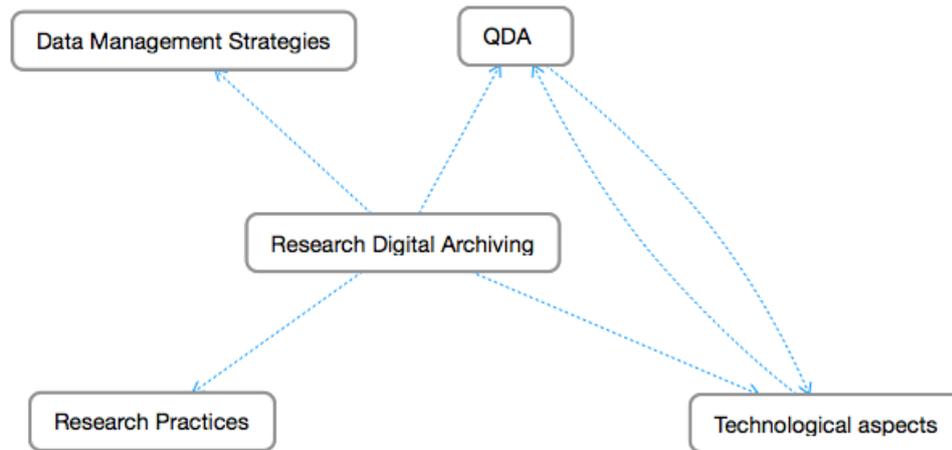


Figure C.1: Digital archiving, key areas

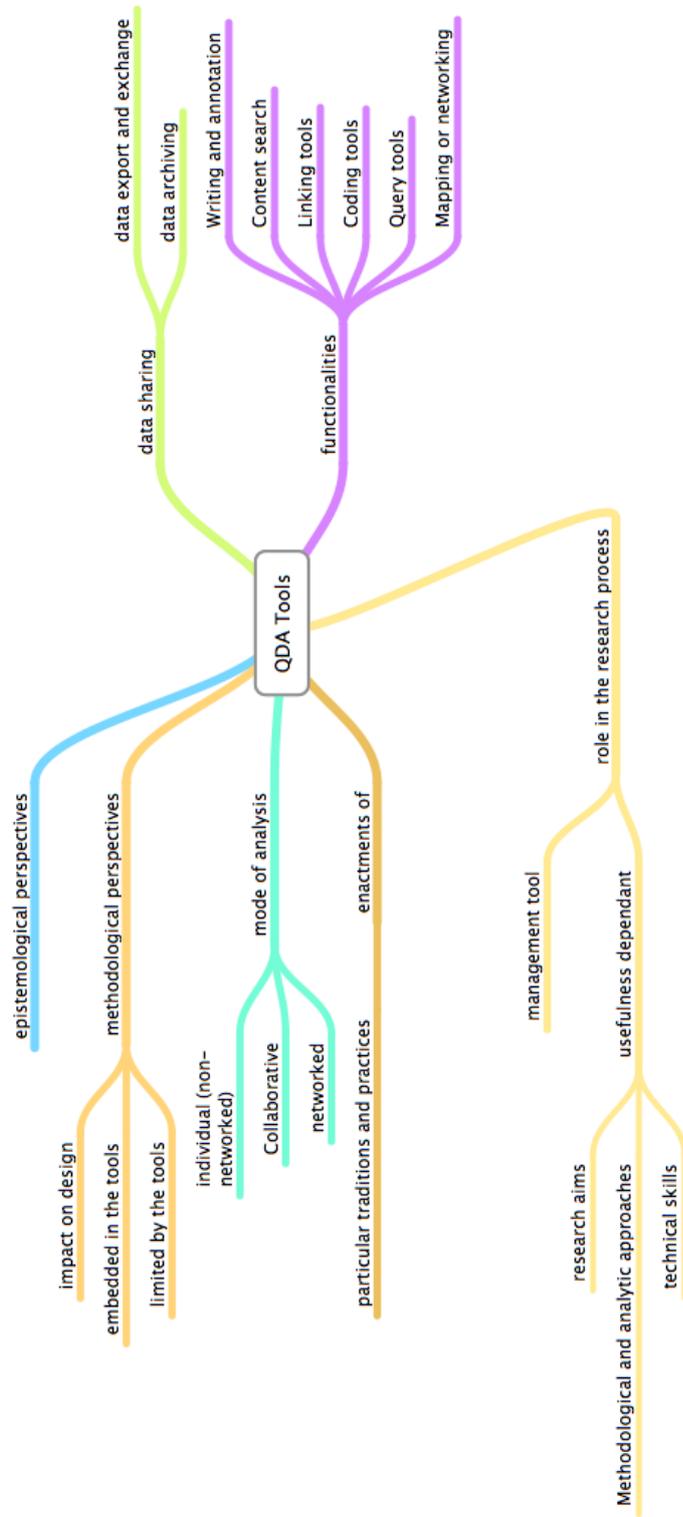


Figure C.2: Digital archiving, sub-area QDA

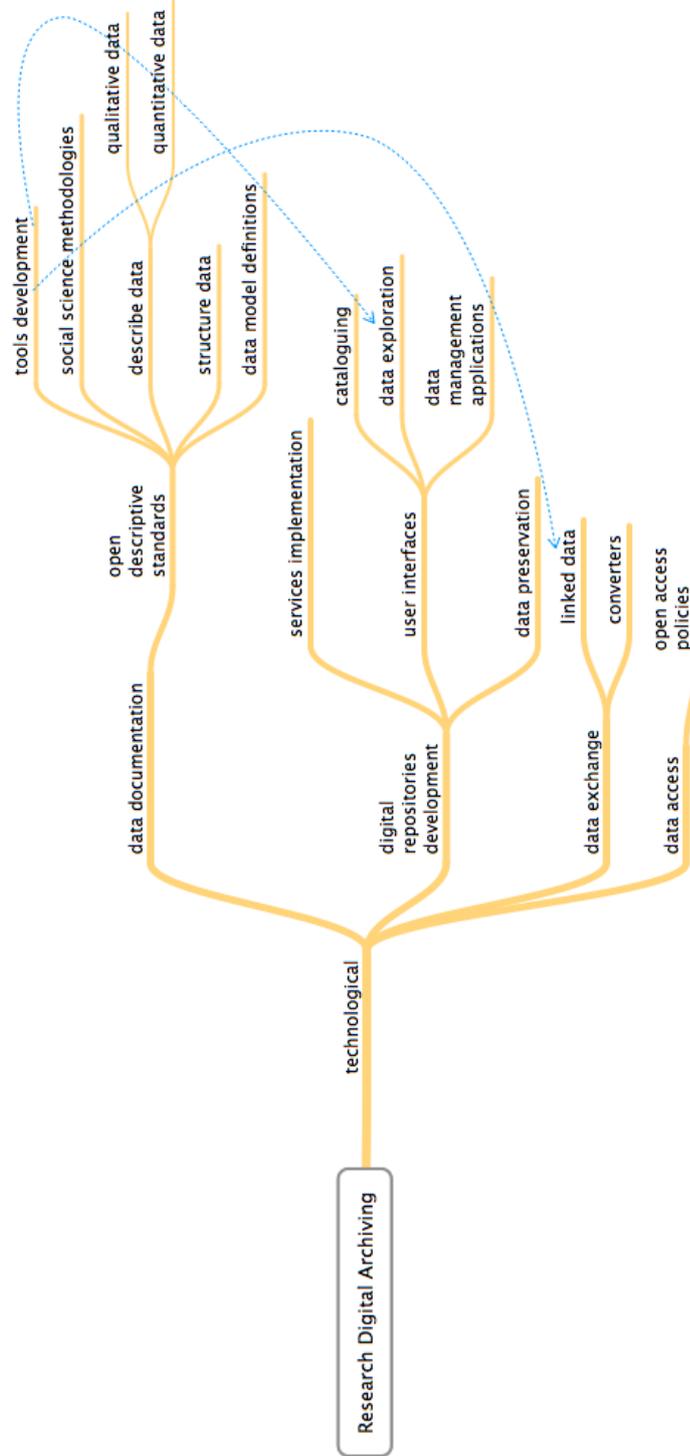


Figure C.3: Digital archiving, sub-area Technological Themes

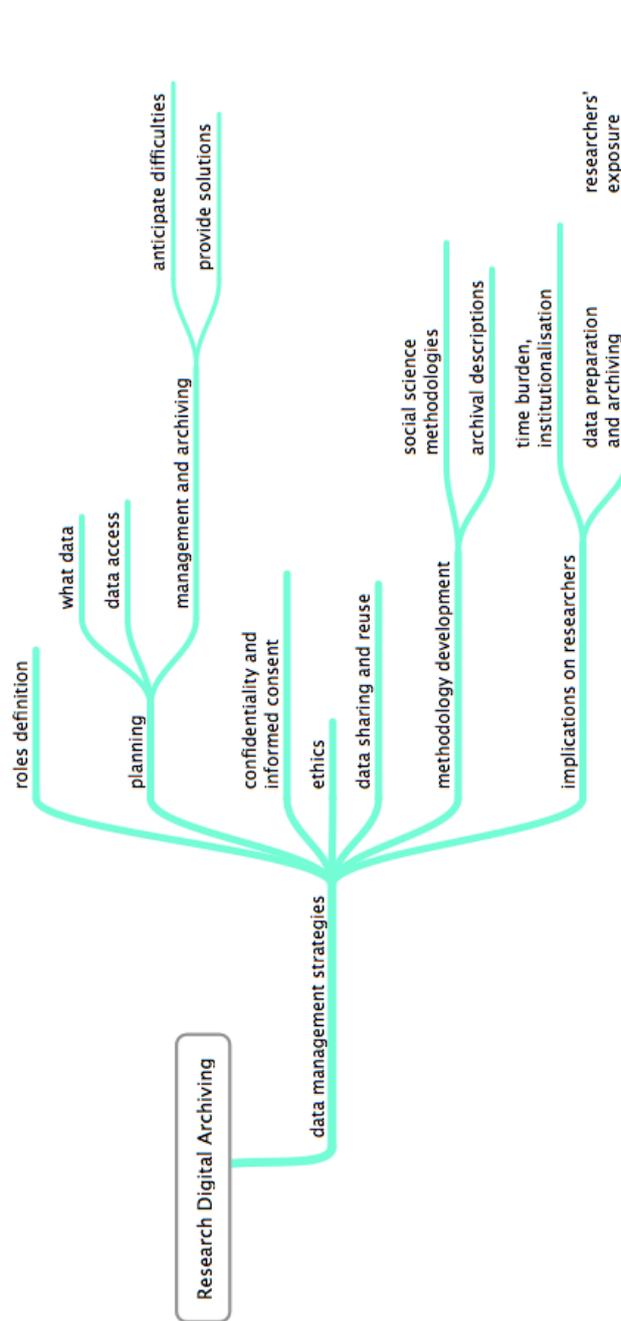


Figure C.4: Digital archiving, sub-area Data Management Strategies

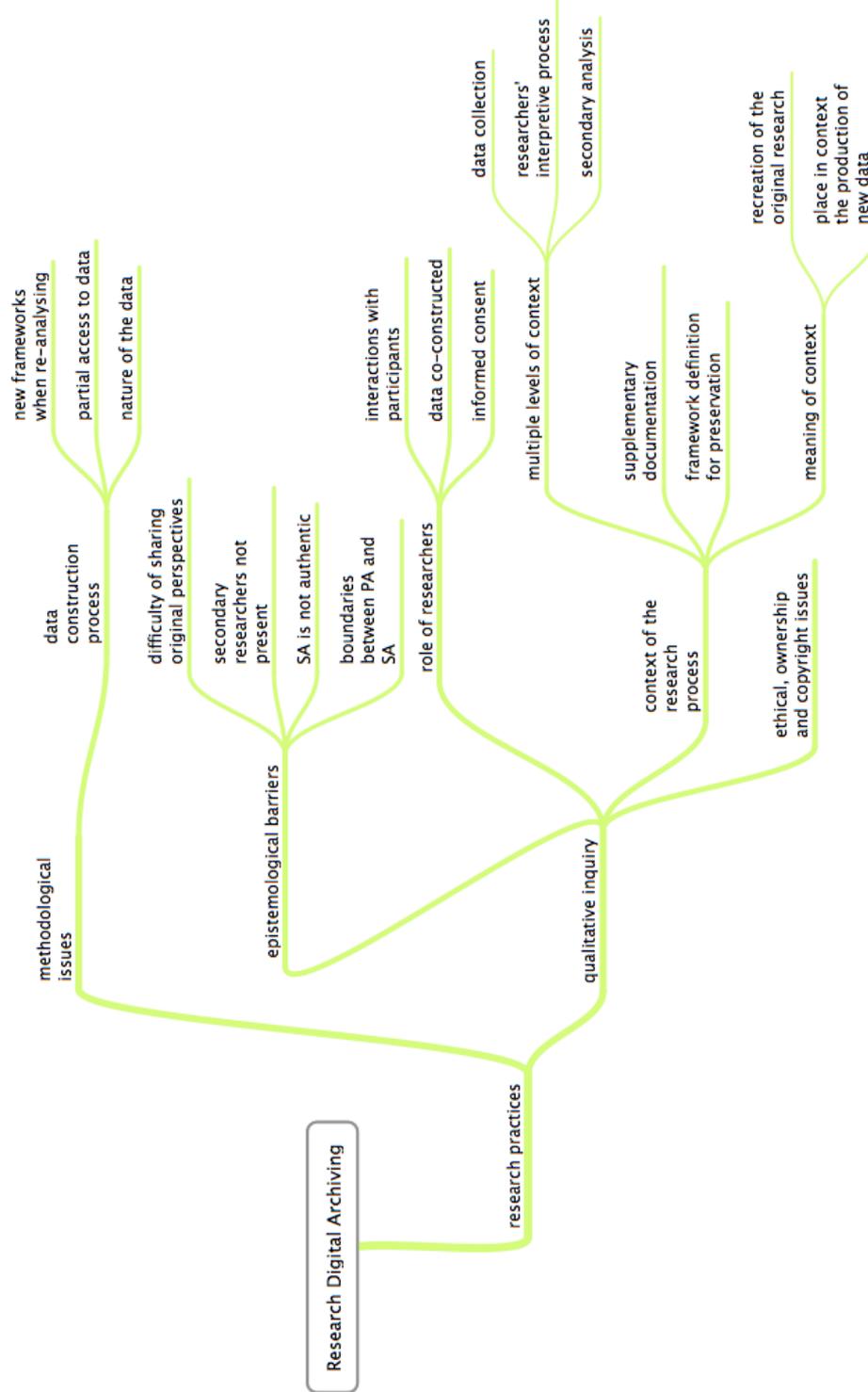


Figure C.5: Digital archiving, sub-area Research Practices

C.2 Participant Information Sheet

LIVERPOOL JOHN MOORES UNIVERSITY PARTICIPANT INFORMATION SHEET



TITLE OF PROJECT *DEVELOPMENT OF SEMANTIC TECHNOLOGY TOOLS TO SUPPORT ARCHIVING, ANALYSIS AND REUSE OF COMPLEX QUALITATIVE DATA*

RESEARCHER: *AGUSTINA MARTINEZ GARCIA, FACULTY OF EDUCATION, COMMUNITY AND LEISURE*

You are being invited to take part in a research study, before you decide whether to take part it is important to understand why the research is being done and what it involves. Please take time to read the following information. Ask me if there is anything that is not clear and you would like more information. Take time to decide whether you would like to take part.

1. WHAT IS THE PURPOSE OF THE STUDY?

This is a research and development study, which will explore and analyse researchers' practices, archiving and reuse of qualitative data on one hand, and on the other will develop a set of tools for documenting qualitative materials, archiving the documented studies in a digital repository and searching across the archived data in enhanced ways. The study of researchers' practices, focusing especially on the archiving and analysis of qualitative data, will provide a better understanding of the key stages, problems that may arise and decisions made to overcome those problems and will guide the development of a set of tools to support/assist researchers.

2. DO I HAVE TO TAKE PART?

No. It is entirely up to you whether or not to take part. If you do, you will be given this information sheet and asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw will not affect your legal rights.

3. WHAT WILL HAPPEN TO ME IF I TAKE PART?

If you decide to take part in this study, you will be involved in a series of 1-2 interviews of about 30 minutes each. The interviews will be scheduled within the duration of your course module, although you may be contacted later on, if additional information is needed. This contact will be made by email. The interview will be audio recorded and pictures and/or copies of your work might be taken. The actual content of these documents will not be reproduced, this is for the purpose of understanding the issues in its archiving.

4. WILL MY TAKING PART IN THE STUDY BE KEPT CONFIDENTIAL?

Research data will be used by the research student, her supervisory team and eventually archived in an access-controlled digital repository and/or password protected hard disks. The data collected will be archived for a minimum period of three years and up to 5 years.

Your institution and course shall be identified, recordings will be transcribed and sections may be quoted in reports, but confidentiality will be maintained regarding the identities of all individuals who feature. All the data gathered, i.e. copies of fieldwork, original surveys or interview schedules, data sheets; will be anonymised in order to protect whatever ethical framework the participant themselves has established for their own research.

The research outcomes, such as the research dissertation and associated publications, based on this recorded material, will not indicate your name or personal details. Additionally, if you wish so, you could be consulted prior to publication, to provide additional information or comment on the interpretations derived from the data collected.

CONTACT DETAILS OF RESEARCHER

Agustina Martinez-Garcia
Faculty of Education, Community and Leisure
Liverpool John Moores University
B107, Barkhill building
I.M. Marsh Campus
Barkhill road
L17 6BD
Liverpool

Email: a.martinez-garcia@ljamu.ac.uk
Phone: 0151 231 4605

NOTE: A COPY OF THE PARTICIPANT INFORMATION SHEET SHOULD BE RETAINED BY THE PARTICIPANT WITH A COPY OF THE SIGNED CONSENT FORM.

C.3 Consent Form

LIVERPOOL JOHN MOORES UNIVERSITY CONSENT FORM



Study title: Development of self-archiving tools to support archiving, analysis and reuse of qualitative data

Researcher: Agustina Martinez Garcia, Faculty of Education, Community and Leisure

I would like to obtain your consent to use an audio recording of this interview and collect copies of paperwork from your research: i.e. field notes, original interview schedules or survey/questionnaires, as well as taking pictures to clarify the contents of this interview. The research group involved in this study (PhD student and her supervisory team) will use these electronic data to study research practices, archiving and analysis of qualitative data.

1. I confirm that I have read and understand the information provided for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and that this will not affect my legal rights
3. I understand that any personal information collected during the study will be anonymised and remain confidential
4. I agree to take part in the interview for the above study
5. I understand that parts of our conversation may be used verbatim in future publications or presentations but that such quotes will be anonymised

Name of Participant _____ Date _____ Signature _____

Name of Participant _____ Date _____ Signature _____

Name of the person taking consent
(if different from researcher) _____ Date _____ Signature _____

Note: When completed 1 copy for participant and 1 copy for researcher

C.4 Correspondence with Participants

Informal conversations were maintained with the participants to explore the nature and scope of their involvement and more detailed information (participant information sheets) were provided electronically via email. An example of the distributed email to recruit participants is shown below.

From: [REDACTED]
 Sent: 18 October 2012 14:14
 To: [REDACTED] Subject: Interview request from PhD student

Dear all,

I am forwarding on a request from Agustina Martinez who is carrying out her PhD at ECL. She is looking for lecturers to interview about the materials and approaches they have used in teaching research methods. Her PhD is developing a digital archive to support the storage and re-use of qualitative and mixed-methods data.

I have agreed to be interviewed and I also suggested you as potential interview participants based on your experiences of teaching on independent research modules and supervision of student research projects. Please see below for Agustina's project information sheet and further details.

Kind Regards

[REDACTED]

Dear all,

My name is Agustina Martinez and I am doctoral student in the Faculty working with [REDACTED].

My study is about research practices and digital archiving, focusing on qualitative and mixed methods studies. I am interested in two different strands, the first of these being to find out about the different processes followed when doing research, looking at the kinds of materials you work with and the role of archiving in those processes (how your data is organised, stored, issues you encounter) and also to discover a bit more about your perspective on secondary analysis on and reuse of existing data (your own research data or other's). The second strand related to existing practices when teaching research related modules, focusing on the types of research materials and how they are used within the teaching and learning activities and also what issues/difficulties students experience.

I will be running short informal interviews and I would really appreciate it if you could participate. It would initially involve one 30 minute interview and, if necessary, further contact by email. If you do decide to participate, I would ideally like to run the interviews on the following dates: Friday 19th, Monday 22nd to Friday 26th October. But if these dates do not suit the following dates would also be available: 5th-9th of November.

I have attached more detailed information about my project in the form of a participant information sheet.

Thanks a lot and I am looking forward to hearing from you,

Agustina

Figure C.6: Participant recruitment email

C.5 Interview Schedules

C.5.1 Interview Schedule - Students

1. I would like you to introduce yourself briefly (tell me about your course, what year you are in...)
2. Could you describe in detail what your study is about?

3. How did you choose your topic? (topic of interest, continuation from recent studies in the course...)
4. When writing your proposal, for your initial literature review, what sources did you choose? (online journal, Internet, paper-based in the library...)
5. How important do you think it was all of this background information?
 - 5.1 Check whether it has changed initial questions, topics
6. What have been the most difficult part when writing the initial proposal?

[DATA COLLECTION]

7. What methods have you used for data collection?
8. Why have you decided to use these methods?
9. How have you designed your research instruments? Tell me about the process.
 - 9.1 Sequence process: research proposal, then methods design, and so forth
 - 9.2 Iterative process: back and forth to the literature review
10. How have you selected the participants?
 - 10.1 Discuss ethics.
11. Can you describe the process of data collection?
12. Did you find any difficulties or key issues when doing data collection?
13. After data collection, what materials have you gathered?

[DATA ANALYSIS]

15. How have you organised the data gathered?
 - 15.1 Manual (paper-based)
 - 15.2 Electronically: folder, additional document listing all the sources...
16. Could you describe the analysis process?
 - 16.1 How are materials constructed
17. What materials have you produced from your analysis?
 - 17.1 Have you kept original materials, like raw data, separate?
18. Have you found any remarkable difficulties in the analysis process?

[DISSERTATION]

21. How have you structured your dissertation?
22. How have you decided what pieces of analysis were the most relevant?
 - 22.1 Explore the process (back to analytic materials)
23. Show research process diagram, more details of the participant's research process
24. Regarding your study and your findings, do you see yourself going back to it and exploring it further?
 - 24.1 How have you kept all the materials used in the study?
 - 24.2 How are they organised?

C.5.1.1 Interview Supporting Materials

In order to obtain more detailed insights about the participants' perceptions of the research process, during the interviews, they were shown a diagram exemplifying the traditional research process flow:

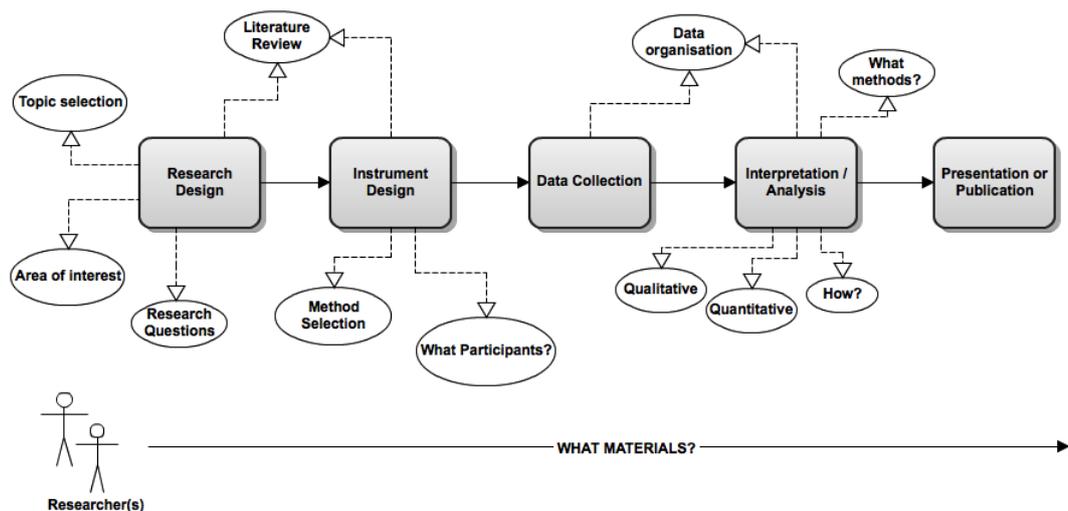


Figure C.7: Research process flow

C.5.2 Interview Schedule - Researchers/Lecturers

1. Research introduction
 - 1.1 Role in the faculty (teaching)
 - 1.2 More specifically about experience on teaching research related modules briefly areas of research

[RESEARCH PRACTICES]

2. Areas of research and the types of projects in which the researcher has participated

3. Participation on medium/big scale research projects? (special interest on data management plans)
 - 3.1 YES. Talk about data management plans
 - 3.2 NO. Talk about the researcher's personal practices for data management
4. Experiences with collaborative work (with other researchers)
 - 4.1 Is there any data sharing involved?
 - 4.2 Its impact on how materials are produced/organised [data sharing]
 - 4.3 Has the nature of the data been a barrier to sharing? Description of any major difficulties.
5. What are the kinds of data collected?
6. What data collection methods?
7. How are the research materials organised?
 - 7.1 Electronically, paper-based?
 - 7.2 How are these materials described for subsequent analysis?
 - 7.3 Difficulties when organising the data gathered
 - 7.4 How is the organisation process performed?
 - 7.5 Integrated with research vs. all materials collected first and then preparation for analysis?

[DATA ANALYSIS]

8. Could you describe the analysis process?
 - 8.1 What materials have you produced from your analysis?
 - 8.2 Has it been mostly manual/paper-based?
 - 8.3 How are the analysis materials constructed: raw data separated, how are analytic materials linked to the primary data?
 - 8.4 Do you use any software to support your analysis? CAQDAS
9. Research outputs, findings presentation and in what forms
 - 9.1 Mostly publications?
 - 9.2 Levels of inclusion of research data. Mostly own interpretations and edited from the raw data, interview excerpts
 - 9.3 What are your perceptions about the inclusion of the collected materials to support the publication or article?

- 9.4 Mention briefly enhanced publications, explore perceptions of articles with selected primary data

[DATA ARCHIVING]

- 10. Data re-use (own data): further primary research/secondary analysis
 - 10.1 Explore whether it would be useful to archive their data and then use online tools to search across and bring relevant materials: topics, codes, categories
 - 10.2 Ultimate purpose and audience of archive content
 - 1) Whether it is for them, for future secondary analysis, other researchers, for teaching, to support publications
- 11. Familiarity with secondary sources and literature exploration: what are the sources...
 - 11.1 Looking at mostly publications
 - 1) Access to more data that the included in the publications: full interview transcripts, interviews, observations...
 - 11.2 Access to existing online archives
 - 1) What are the perceptions about exploring other researchers' data

[TEACHING PRACTICES]

- 12. Brief introduction on the experiences on teaching research-related modules
 - 12.1 Teaching research methods (e.g. introduction to research)
 - 12.2 Students developing their own research projects
 - 12.3 Supervising students carrying out small research projects
- 13. Description more in detail about those modules related to research
 - 13.1 What's the students' initial background, as in research experience prior to the module?
 - 13.2 What exemplary materials have been used in those modules? Any materials to help them developing their own projects?
 - 13.3 Researcher's experience (in teaching): what are the most remarkable issues students face during these modules or the development of their own projects?
 - 13.4 How have students been supported to overcome these issues?
 - 1) Issues related with research design
 - 2) Issues related with analysis and interpretation
 - 3) Issues related with the lack of experience doing research
 - 13.5 Given the feedback provided by the students during their projects development, are there any modifications in the course contents/ teaching approach?

1) Course structure, materials used

13.6 Are there any activities within the course that encourage students sharing, discussing their work?

1) Research e-portfolios development [reflection on their work or practices, sharing with other students]

C.5.3 Interview Schedule - Senior Researchers

1. Introduction to the participant's research and background of the archive project
2. What are the main motivations for these projects? (with respect to the archive production)
 - 2.1 Inquiry learning
 - 2.2 Secondary analysis/ further exploration of primary data
3. How have they been used in the teaching/learning context? (if used in T&L)
 - 3.1 Research basis first and then secondary development of case studies
 - 3.2 Primary research

[DATA COLLECTION]

6. What methods have you used for data collection?
 - 6.1 Brief description of the data collection process.
7. What are the kinds of data generated during data collection?
8. Describe the process of data collection. (also key issues)

[DATA ORGANISATION]

9. After data collection, how have has the data gathered been organised?
 - 9.1 Manually, electronically (folder-based)
10. How has the data been described?
 - 10.1 Contextual information
11. What have been the most relevant issues when organising the data gathered?
12. How can data be explored?
 - 12.1 Limitations technology-related

[ARCHIVE DESIGN]

13. How has data archiving been contemplated?

- 13.1 Was any specific data management plan designed in the early stages? How was the design approached?
 - 1) Expected difficulties
 - 2) Data anticipated
- 13.2 Perceptions of archiving. Degree of integration within the research process.
- 13.3 What are the most relevant issues?
- 13.4 What are the expected uses of the archive? [impact on design]
14. Design process for the archive production
 - 14.1 When it started?
 - 14.2 Anticipated issues
 - 14.3 Who was involved?
 - 14.4 Data selection criteria
15. What were the main aims of the archive?
 - 15.1 Secondary analysis
 - 15.2 Data sharing
 - 15.3 Has the nature of the data been a barrier to sharing? Describe any major difficulties.
16. What has been the criteria to decide what data goes into the archive?
 - 16.1 Does it raise important issues regarding what data stays private and what is shared and in what forms? What are the issues when preparing private/public data?
 - 16.2 Ethical issues? If any, how have these been addressed?
17. What are the audiences of the archive?
 - 17.1 Why archiving? What can it offer?
18. What is the role that the available technologies have played during the design and construction of the case studies?
 - 18.1 Has this design been driven by the existing functionalities provided by the technologies?
 - 18.2 Have they placed any limitations/barriers?
19. Walk-through the archive

C.6 Interview Transcripts

This section includes sample transcripts from the interviews conducted in the course of the study as part of the empirical work. The interviews have been anonymised and confidential information such as institutions, research projects, places or personal information have been removed (white or black boxes in the text). The transcripts from the three different cohorts of participants that have been included are the following:

- Students conducting small-scale research projects (4 participants). One interview transcript is appended.
- Researchers involved in undergraduate research (4 participants). One interview transcript is appended.
- Senior researchers involved in archived research (2 participants). The two interview transcripts are appended.

C.6.1 Interview Transcript Student 4

- 00:01 [I] Well, first of all I'd like you to introduce a bit yourself, as in what course you are in...
- 00:09 [S4] Yeah, well my name is _____ . I am in third year doing Education Studies and PE and next year I'll be doing the PGCE in Secondary PE.
- 00:20 [I] Ok very good. Could you describe a bit in detail your project, your independent project? Just an overview.
- 00:26 [S4] Yeah. Well I decided to look at parental attitudes in regards to physical activity because at the minute the issues is gender stereotypes for sport, like males football and girls dance and that... and the underlying factor of that is... it starts with the parents attitudes affecting the children. So I wanted to investigate what sort of impact that has.
- 00:53 [I] Ok. When you started...well why did you decide, why was this topic of interest to you? Have you worked with something similar before?
- 01:02 [S4] Yeah...well it was just...when I did a placement, I found that out teaching... speaking to the teacher and that, and I knew that like, lads don't wanna do dance and girls don't wanna do football but I wanted to see whether it was actually parents that was giving it that attitude or whether it was environment that was deciding whether they liked it or not.
- 01:25 [I] Ok. That's good. Um I guess you started with your research proposal first. So how did you approach it? What were the things you were doing first when you started?
- 01:36 [S4] Um I was looking at whether research has been taken up before and looking at the recent agendas from the government, what they were saying on stereotypes, things like that. And then I looked at how they conducted the research, like how do they get around it.
- 01:52 [I] So was that for your initial literature review?
- 01:56 [S4] Um my literature review was more um like about attitudes in PE and why... how children found attitudes towards physical activity.
- 02:06 [I] Ok, what...when you were doing this literature review for your proposal um what sources were you using? Were online mostly?
- 02:13 [S4] Yeah, mostly journals online. But I was trying to find more UK rather than American so it could be more directed...yeah. Because there was a lot on American but I wanted to focus in the UK educational system.
- 02:29 [I] Ok. When you started your research, as in when the proposal, I guess in your case you already had a clear idea about the topic and then you used the literature review more for...
- 02:40 [S4] Yeah
- 02:40 [I] Like for writing your research questions or it was more like exploratory there. As in, you used it like, you started with a general idea, like topic, and then used it for refining?
- 02:51 [S4] Yeah, yeah. I started with the attitude... parental attitudes influence and then, the more I researched the more I found similar journals talking about similar things, directing me more.
- 03:04 [I] Did you find any changes there? As in when you were doing the literature review, did you find like new perspectives or things you didn't think in the first place, that would have been interesting? Or it was just basically on your topic there?
- 03:16 [S4] Yeah. It was more just on the topic, more than anything.
- 03:19 [I] So how... did you find like rich enough journals there? As in was enough for your research?
- 03:23 [S4] Yeah. It was enough. The other thing that wasn't enough it was directly children

- against parents, it was all about... there's a lot of studies about children's attitudes towards PE, parents' attitudes towards PE but there wasn't actual putting them together.
- 03:39 [I] Ah ok
- 03:39 [S4] Which made it kind of better... cause it made me doing the research.
- 03:43 [I] As in... in the end it was like your research basically, part of your research. Ok, did you find any issues in these first stages there?
- 03:51 [S4] Um No, I did it all right actually. I didn't find troubles, no.
- 03:56 [I] Ok. Um, data collection. So what methods have you used for your research?
- 04:04 [S4] Um I used a published questionnaire. When I was doing my research I found that in quite a few of the studies. This questionnaire has been used.
- 04:14 [I] The same one? Ok
- 04:15 [S4] Yeah. And it's been dev... like from 1960s and it's been developed and developed and then it kept coming up so I found the questionnaire and that's what I've used to conduct my research.
- 04:26 [I] Ok. So then, did you use the last version of it in the end?
- 04:29 [S4] Yes
- 04:29 [I] Because you were saying that it's been kind of like changed...over time.
- 04:33 [S4] Yeah... it's been developed over time.
- 04:34 [I] Ah ok. Um did you... because in the first place um when you were doing the literature review did you have any methods in mind? Like, something different or you were looking for questionnaires?
- 04:45 [S4] Well, I was going to use questionnaires as the main thing. But I was going like, to do my own questionnaire.
- 04:53 [I] Why did you decide...because it was informed by the literature, the thing with questionnaires or you had it in mind from the beginning? So you thought...
- 05:01 [S4] Um a bit of both. Well I wanted interviews and that but... I think it was more ethical than anything with interviews [his research]... interviewing children.
- 05:09 [I] Is... you are interviewing... than interviewing children so... ah ok, because... any issues? How did you design your... you were using that questionnaire. Did you change it, did you design something else apart from that?
- 05:23 [S4] Um no. I just used that. Yeah.
- 05:25 [I] directly... ok. What participants did you select? How did you select the participants?
- 05:35 [S4] Well I chose out of the secondary schools from year 7 to year 9, that was the focus group because in that year you decide what subjects you want to choose for your GCSEs so I found that they've got um... cause if you choose year 7s they'd just agree with whatever you ask them cause they are still young and year 11s are just into it so it already happened so I think... I chose year 9 cause that's when... they've got like a fixed idea on whether they like PE or they don't so I used that. And then I just randomly selected um three forms from the year group in the school and this should be it.
- 06:14 [I] Ok. How many participants did you have in the end?
- 06:19 [S4] I think it was 32 that replied.
- 06:22 [I] Any issues there? How was that data collection process?
- 06:25 [S4] It was ok. But it was the fact that the children had to take the questionnaires home to their parents cause they had to fill one in and the parents did and it was the fact that having to go from the school and back to the parents and back into school... that's why I didn't get as many back.
- 06:42 [I] Ok, so 32 in the end. Do you reckon... has it been sufficient, has it been ok for your

research?

- 06:48 [S4]] It's been enough but if I had more time and that, I would have done a lot more cause I would have been able to interpret the results better.
- 06:58 [I] Ok, so at this point just after your data collection, what materials did you have already there? Like, as in for your research. Um did you start... I guess at that point you already had the literature review and things like that so, what sort of data did you have?
- 07:15 [S4] Um. When I got the questionnaires back in... I'll put it up [we are looking at his computer, he has the materials electronically] The questionnaire split up into 7 aspects of PE... so the first one is 'doing PE to meet new people', which clusters social growth and then, as you can see, there are 5 questions where they've got to mark with good being the best answer and bad being the worst. So for each question they've got this scale, so... as you can see, for each question... The only one that's different is this question which is 'health and fitness'. Which had ten cause I had to split up to like social and that. So then, I collected the results... I'll show you, it's in here [we are now looking at the excel documents with his analysis from the questionnaires].
- 08:09 [I] Oh yeah, it's mostly quantitative? Or do you have any like qualitative questions there or not?
- 08:14 [S4] No, it was all... it was all quantitative. But then as you can see from this like... we use this one, participant 5 for the social question which was either good or bad they've all... good, five is the highest mark so they've all 'good good... quite good' and then 'unsure, unsure'.
- 08:37 [I] Ah ok.
- 08:37 [S4] So I did that for each question.
- 08:39 [I] So how did you... because did you have, when designing the way you were going to do your analysis in the end, how did you translated from the questionnaires into the actual excel that you are using here? Do you have codes?
- 08:53 [S4] Yeah, this is the um first question for the child, so that's like... this is social continuity, which is... this [showing a specific question in the sample questionnaire], which is that [pointing at the equivalent column in the spreadsheet in Excel] and then... so SC1 represents my question, SC2 represents that and if they put good they have a 5 cause it's 5 out of 5. So that's how comes in and then that's the mother of that, and then that's the father and in the end, what I've got is the means and then compared so it was quite... time structured.
- 09:32 [I] Ok, because you have it here. Since you have mothers, you have like parents and then the children, so I guess... how have you analysed the results like together? How have you done, like, I don't know... cross-tables or? Because having such different perspectives there so how have you approached that?
- 09:53 [S4] Yeah, I have swayed up into the 7 categories and then worked out the mean of the child, the mean of the mother, the mean of the father. Then I've compared the means and then I wanted to correlate it but because it was only a small sample and such... it was quite hard to find correlation so I was mainly going off means. I've got against female child against father and female child and mother, I wanted to see if there was a gender difference.
- 10:24 [I] Ok, so then in the end for your analysis, you ended up with... you have your physical questionnaires there and you had like, when you were doing your excel files. Have you used anything else? Have you used like um, word documents for writing notes and things you could use for your writing after?

- 10:42 [S4] No, not really. I've just been... because it's more quantitative, I've just gone for some visual graphs.
- 10:47 [I] Ah ok. What has been the most difficult part when doing your analysis? Any issues you found?
- 10:55 [S4] Um just trying to pick out like, certain connections in there. Because sometimes the mean doesn't directly say evidence so I tried to work out whether I could find any other way of like...
- 11:09 [I] Um how have you decided what's like, I guess you've used in the end to present your results. Have you used graphs? Or, what sort of graphs?
- 11:18 [S4] Yeah, graphs.
- 11:18 [I] So how have you decided those ones were like the most suitable for this?
- 11:23 [S4] Um it just... it was easy to... from the graph using the bar chart type of graph it's easy to just say the differences straight away... from the themes... like that, straight away it shows that female children and the mothers feel that aesthetic approaches, graceful movements [we are looking at one of the graphs from the results] it's more important than males think. So just from that you find a clear gender difference.
- 11:53 [I] Ok. A gender difference there. Now, that clearly represents a...
- 11:57 [S4] And well, I've also compared later on in my results I've compared different aspects so like that's aesthetics and esthetics, the other way around with the males is higher so I've put them on one graph and then compared them.
- 12:14 [I] So any issues there? So I guess this is... is this the first time you do such a project?
- 12:21 [S4] Yeah.
- 12:21 [I] Of research so what sort of... did you have any background before in doing statistics?
- 12:27 [S4] No, I think that's why... it's a problem like, I found more detail like I've got friends that were on sports psychology courses where they use SPSS and all that. So I thought I might need to use that but having no experience in it at all...
- 12:45 [I] Because in your A levels did you do something related?
- 12:49 [S4] I did, yeah, I did psychology which's got research methods in it. But it's not got... I understand this and that but it's not got like... doesn't go into it as much in detail. It just basically goes over qualitative, quantitative and all the types of methods rather than actually doing research.
- 13:06 [I] Ok so that's you reckon... that's maybe one of the reasons... have you struggled a lot with this? You've found in the end the analysis kind of OK?
- 13:14 [S4] Yeah, I didn't struggle I just thought if I had more, like knowledge of research then I could have done more tasks like... interpretation rather than just go for the means...
- 13:27 [I] Ah ok. So... regarding the writing, how have you structured your dissertation in the end?
- 13:37 [S4] Um I did my lit review at the beginning, then method and then my results and then interpretation and discussion and then conclusion and recommendations.
- 13:50 [I] Ok. How has been the process... yeah, when you did your literature review and then you went into your methods, and then did the analysis... with your analysis did you find like you needed to go back and change...
- 14:06 [S4] Yeah, a little bit.
- 14:07 [I] So how was that? It was because of your results or?
- 14:12 [S4] Um, just what I found in the research some of the literature didn't really like... was relevant to what I've actually done so there's like 'do I keep it or do I not' type thing

- but... the methodology like... the methodology helped me to work through my thinking so I did quite structured in terms of participants, what my design instrument was and all that. So I could just like use that as my guidelines to do my work.
- 14:43 [I] Ah ok, so you reckon the method was obviously something closed at that point.
- 14:46 [S4] Yeah
- 14:46 [I] So nothing of that really didn't change, you went and proofread it and do changes but... in terms of the actual content that was it and that was closed. Um yeah, because you were saying some of the things in your lit review they really weren't matching what you found so then, how did you deal with that? Did you put it in your conclusions as in new findings...?
- 15:10 [S4]] I put in my conclusions and recommendations at the end. If I'd do it again what I'd do differently so that's...
- 15:16 [I] So when it comes into writing your analysis part in your dissertation there, um how easy did you find to go back to your results and take the bits and pieces that were relevant to that?
- 15:29 [S4] Yeah... it was all... it was cause like some of the factors clearly stated that there was more to the results so I could describe them quite a lot. But there were some I could have analysed as well but... it was all right I think.
- 15:48 [I] Why... yeah, how was the process of actually deciding what was really important? Did you find any issues there... as in, how did you approach it?
- 15:59 [S4] It took some time cause I needed to... like get all the actual results and put them along each other. I actually spent time just actually thinking: 'right is there a link between them...' and 'there isn't a link between that and that...'. So it took quite a while to find the links and then discuss them.
- 16:17 [I] When doing your writing, in general, in your project, how have you been keeping the materials? What sort of materials you ended up with? Do you have all of them mostly electronic?
- 16:30 [S4] Yeah, I've got my questionnaires at home in paper.
- 16:37 [I] And apart from that, everything else has been there like electronically?
- 16:42 [S4] Yeah, all my results... yeah it's basically all of my excel files and then written up in my discussion and analysis.
- 16:56 [I] Have you been keeping versions of the different things? For example, with your excel, have you been working all the time with the same one?
- 17:05 [S4] No, I've got like I've got a big folder. It's got like... [we are looking at the computer folders with his work] basically all I've used...
- 17:14 [I] Ok, so you have them separate.
- 17:15 [S4] Yeah separate... and then put together in the... [dissertation]
- 17:19 [I] And with your writing have you done it as well? Like that, you have different drafts or different sections?
- 17:23 [S4] No, I had drafts but I kept it on into one file.
- 17:27 [I] In the same one, ok. So, you reckon having this structure here you reckon it has helped when it comes into writing your analysis? Was it like structured or clear enough for you to go back and forth?
- 17:43 [S4] Yeah
- 17:43 [I] And look at the different data you have there?
- 17:45 [S4] Yeah... rather than being on one thing it's easy just to scan through...
- 17:50 [I] Did you find easy to use excel?

- 17:53 [S4] Yeah I've used quite a while ago cause I did ICT in school...so I've used excel before but it took a couple of days getting it back through my mind to do things...
- 18:08 [I] Ok
- 18:08 [S4] So yeah, I could do the formulas and stuff like that.
- 18:11 [I] So, finding this questionnaire... it was really good there... do you think it's been really useful for your purposes, your research?
- 18:22 [S4] Yeah, yeah.
- 18:22 [I] Um because, overall, as if you had to think overall about your research project, what has been the most difficult part?
- 18:32 [S4] Um...it was just finding more direct relevant literature at the start I think. Cause that took quite a long time but... as soon as I got that then, it kind of just opened up and was able to just flow through it.
- 18:49 [I] Why do you reckon it was that with the literature review? Because in your case, you were saying before it was more like you used the literature more to refine what you were looking at after... or why you reckon that has been an issue...focusing?
- 19:04 [S4] I think it was finding the actual... what I wanted to actually find...rather than finding bits around it.
- 19:11 [I] Do you reckon it's because it was a very wide... you had like a lot of literature around so for you it was more like... ok 'filtering here... I just want to focus and look at what's more relevant here'.
- 19:24 [S4] Yeah.
- 19:24 [I] Ok, so that has been like the most difficult part there, what about like writing? How has that been?
- 19:32 [S4] Yeah... it wasn't too bad. The... I think the hardest bit was the actual discussing it, discussing the results so...
- 19:41 [I] So you reckon that has been... it's that the most difficult part from the writing?
- 19:46 [S4] Yeah.
- 19:46 [I] Because before in the degree do you have similar dissertations?
- 19:50 [S4] No, this is the first one.
- 19:53 [I] Um... a little bit tricky question, yeah, what are the things you think you could have done... now that you've finished and you've done it... what would you have done differently?
- 20:07 [S4] Yeah, um I would have tried to... [reflecting, pausing] I think it was my time more than anything... I was end up rushing things so I thought if I could move my timescale it would have been fine and I don't know really. It was all right, it was just. I think I would have done more... not more research but used more... like try to rely on online journals... stuff like that, I think for my research at the start... I got around it but I think if I've done my research better earlier, I'd had more time on other aspects.
- 20:49 [I] Ok, so far, were you happy with your results? Did you think they were sufficient there?
- 20:55 [S4] Yeah, I was happy with my results!
- 20:56 [I] Because I was just thinking what plans...do you plan to go and teach after the degree? And, do you reckon this has been useful for for... just thinking about what you are going to do next. Do you reckon it's been useful, as in for teaching?
- 21:12 [S4] Yeah, yeah because it's like...giving more an insight of actually children, children actually feel towards PE and their parents are having an influence on what they are doing. Cause they are trying to change it in schools that it's more PE for all the... males

and dance and girls could do what they want and it's... it does help and along with that it's like helping with... what I found is that... like this sort of instrument could be used to design like part of the curricula and that.

21:47 [I] Ah, that's very good...

21:48 [S4] That's like...it shows like what aspects of PE girls like, and what they don't like and that. It's good but then at the same time it's hard because it's saying there's such varied opinions between girls and boys, it's like how can you plan mixed classes for PE so...

22:07 [I] So do you reckon it actually could have an impact on teaching practices?

22:13 [S4] Yeah! If it was like... if the questionnaire itself was like research more adapted more specifically to that actual like type of PE then I think it could be useful.

22:27 [I] No, it's because this questionnaire does it aim to a specific year? As in is targeting specific...

22:34 [S4] No, that's why I think if it's just adapted slightly more...

22:37 [I] Right... to meet more the...

22:39 [S4] Yeah!

22:39 [I] No, that's very good. So you reckon it's something you probably would be extending in the future?

22:45 [S4] Yeah! I could, if I was to... [teach], yeah.

22:48 [I] Ok, that's very good! I think like... anything else... yeah because you were saying you were struggling a little bit in the beginning but so far appeals like your process has gone smoothly there.

23:01 [S4] Yeah.

23:01 [I] Any other remarkable issues? Along the whole project.

23:06 [S4] Um... I can't really think of any... no it was... I found out what I thought I was going to find out and the most thing was about the PE, like when I thought about it more and that.

23:23 [I] Did you have any pre-conceptions in the beginning? As in, in the end, I think you maybe had ideas in the beginning of what, maybe what you were going to find out. And then when you did it and actually your results...was it kind of similar enough? Or you really found things... unexpected things you never thought about.

23:44 [S4] No, it was kind of similar but I wasn't expecting such differences between... there was a bigger gender difference than there was with the parents. There was a big difference between males in general, like there was a big relationship between the males and their fathers and then the girls and their mums rather than... that, that was the underlying relationship.

24:17 [I] Ok, that's a bit unusual isn't it? Because you would expect it maybe you know, if you were looking at gender there, you would expect it more like gender, not really between the girls and their mums.

24:27 [S4] Yeah, yeah.

24:27 [I] Ok, that's interesting. I think that's pretty much it! That's been useful, thanks a lot!

24:34 [S4] No, it's fine!

C.6.2 Interview Transcript Researcher 3

- 00:01 [INT] Ok. Hello, thanks a lot for participating um I'd like you to introduce yourself briefly and tell me about your role within the faculty.
- 00:12 [T3] Ok, I am [REDACTED] I'm a senior lecturer in Education studies, which is a degree in the centre for Education studies and early childhood studies. I teach at levels 4, 5 and 6 in the Education studies modules and also some early childhood studies modules as well.
- 00:34 [INT] Ok, do you teach any more research-oriented modules?
- 00:37 [T3] Yes, the third year module or level 6 module, I should call it, independent writing and independent project. Both modules, I've been module leader for this year and last year, and I taught on it in the previous year as well.
- 00:56 [INT] Ok.
- 00:56 [T3] I've got a few years experience of teaching on that um I've done a couple of workshops for people at other universities about um teaching research methods, one is at City University and one at the University of Cambridge as well.
- 01:18 [INT] What was the audience for those?
- 01:19 [T3] Um they were masters level, but they were just single workshops where I was talking more actually about my own research and discussing the methodology and methods that were used.
- 01:30 [INT] Ok, um I'd like you to describe a little bit in detail your areas of research and some types of projects you've participated in?
- 01:37 [T3] Aha. So um my own research background is in Technology Enhanced Learning and um I have used a participatory approach in my research, from the beginning I suppose. I started in Plant Sciences, because I have a plant sciences background and so I was conducting research on using online technologies to enhance teaching and learning in plant sciences at the University of Cambridge and I did that for a couple of years and then started to move into more disciplinary field, inter-disciplinary field, and look at how to use some of the same methods I used at plant sciences in some other faculties and disciplines at the University of Cambridge. At that time I did masters in Educational Research, part-time.
- 02:33 [INT] That's quite a shift.
- 02:34 [T3] which gave me more of a background in teaching Educational research and methodology, and methods and so on and I just sort of continued in that vein, um as a postdoctoral researcher until I became a senior lecturer at the university.
- 02:50 [INT] Ok, um would you like to describe a bit in detail any like large-scale project you've participated in. As in, in terms of, like projects which generate a lot of data, research data.
- 03:02 [T3] Yeah, well, probably the first really large-scale project was the Learning Landscapes project at Cambridge where I came in sort of mid-way into the project and mixed-methods had been used, some of them quite innovative um but they had a large scale questionnaire which was about um ... they had dat... daily life or life experience documentation. Students took photographs and wrote diaries about their day and then, there would be focus groups and interviews and there was all of this data that was in a virtual research environment.
- 03:40 [INT] Ok.
- 03:40 [T3] It was my job to turn that into a report on various different themes.
- 03:43 [INT] Ok, so you said you started it when it was half-way so all the data was previously

collected, you would say?

03:50 [T3] Nearly all of the data

03:51 [AM]Or you had to do some primary collection?

03:53 [T3] We still did a couple of interviews and focus groups um to draw more from the data that was previously collected but the majority of the... the actual research methods had been conducted already.

04:07 [INT] Was it multi-disciplinary? Was different researchers with different backgrounds working together?

04:13 [T3] Um...[thinking]

04:14 [INT] Or it was more education based...

04:15 [T3] In a way, it was researchers who were based at CARET (Centre for Applied Research in Educational Technologies) and so although they did have different original disciplinary backgrounds, they had been working as researchers in inter-disciplinary work for a few years, so it was more of that sort of social sciences field that they were working in but actually the audience for the research was very inter-disciplinary and that was quite difficult to be able to give them the information in ways that will keep them happy.

04:52 [INT] So did it then impact the outputs from the project, as in they way you thought about what outputs to generate, so what were the main barriers there?

05:05 [T3] Well, we did have problems with presenting the qualitative information to some of the audience because we had critical friends from various different disciplines and some of the more scientific disciplines like chemistry, physics or mathematics so they were not comfortable with us presenting case studies.

05:29 [INT] Sorry, they were or weren't?

05:29 [T3] They were not comfortable with us presenting case studies. They would have considered a day in the life experience sort of say, two or three students, they didn't feel that would be representative of a body of students that WORD thousands so it was hard to argue for the value of a case study in that situation. So it was useful that we used mixed-methods because we could then present to them the questionnaire which was more quantitative and more generalisable um but even then, the kind of statistics that we did was more from a social sciences background, so it didn't necessarily match up to their expectations from the scientific field.

06:13 [INT] Ok, because how was then that specified from the beginning? It was like it wasn't approached correctly from the design, or it was something that came up after more?

06:27 [T3] The problem was that there was a group of about 25 advisors who were all from different disciplines and there would be these meetings where they would not all attend necessarily, and so you get two or three of them being particularly WORD in one meeting and they would steer the project in one direction, and then you would do what they advised and then present the data back in the next meeting, and there would be different people there, who would be from a more WORD, from a different discipline [laughs] so actually, it was a very difficult project to um try and get an outcome from... in a way.

07:03 [INT] Ok.

07:04 [T3] um and what we ended up having to do was present these reports that had information from all of the different methods that we used so we will present the questionnaire data and a short case piece, um along with other data and information, be very exact about 'this represents two students or this one...'

- 07:27 [INT] Ok, so, was then the report including raw data as well?
- 07:30 [T3] Um... We were using a structure similar of that of Stenhouse's raw data case record and then a case study because, because the raw data was already in the VRE. We took trouble to try and do a level of analysis of that data which will lead to a case record that was still stored in the VRE but keeping the links there so, when producing the case studies and the report, we can then refer back to the original source.
- 08:02 [INT] Ok, was it original or you would say lightly edited?
- 08:08 [T3] We acknowledged that it was analysed and but it was clear where that original data were coming from and also what that analysis had been because we had a sort of middle tier of analysis that had been conducted.
- 08:24 [INT] That maybe then adding more contextual information to the actual data there. Was this designed, agreed in the beginning of the project, you would be producing these outputs? I'm asking you that because u... thinking about archiving as in, was that one of the aims, how were you expecting the data to be used after, if you were.
- 08:47 [T3] Um... we actually had a lot of discussions about that during the project [laughs]
- 08:50 [INT] Um ok!
- 08:50 [T3] And as I said, I came in sort of at the middle and these discussions were still very much happening at that stage [laughs]. So, I'd say... I don't think they had realised at the beginning that there would had been so much data that there would have been a need for mixed methods in the way they were, and that the output would be in the form of these reports, not quite sure that that had been set out so when that became apparent sort of midway and there was so much data, and then they realised it was a lot of data so we were having a lot of discussions about how would we structure it and keep the links, um so... and then, how it was gonna be presented at the end and the Stenhouse's paper was particularly useful in conceptualising that and I think actually, we did produce a front page to the VRE which was a graphical representation of raw data, and we had let's say questionnaires, day experience, focus groups and interviews, which you could click on that and it will take you to the right folder and then we had a second stage which will take you to the case record which were analysis of questionnaires and more... so you could click into it there and then we had the reports at the top level so you could just go to those.
- 10:10 [INT] Sort of like a top-down, or bottom-top, you could go actually both directions.
- 10:15 [T3] Yes. In that it was that would represent a form of database that somebody in the future could come and do... and use some of that. I think for practical, ethical and employment reasons um I don't think that has been given open access to the university let's say, or... I'm not sure there's somebody who now keeps an eye on that database, it's not me I can't access it anymore [laughs] so...
- 10:44 [INT] Ok, so it's actually not public at all, is within the university?
- 10:51 [T3] It's within the university, it was hosted by CARET, in the Sakai environment and so I'm note sure.
- 10:56 [T3] Did you participate in this selection process... of selection of the materials to be included in... Would you like to talk a bit about that?
- 11:04 [T3] I did... yes. Because I was... I wrote two of the reports so um I'm not actually sure how the themes came out but there were 5 or 6 themes that arose from the data collection. One of which was small group teaching, and I also did one on transferrable skills and so it was my job to look through that raw data and also to the level of analyse data, to put that into a short-ish sort of 5/6 pages report on what we have found.

[laughs]

11:41 [INT] Ok, that a pretty condensed... [laughs]

11:42 [T3] ... data um, which was tough to actually understand what had been done and what the data actually meant, when you saw it in that form, just in a VRE... I needed to collect one time, but I suppose I did it one at a time in that I first looked at the questionnaires, I read about how the questionnaires had been conducted, looked at what analysis had been done, if I felt I needed to do more analysis, I did this and did that as well. Um and, I summarised it in my notes and then I went onto the next method and so I looked at what did the data experience find out about um small group teaching let's say and I looked at, maybe it has already been analysed and coded, in which case I looked for small group teaching within those codes or something related to it. If that hadn't happened, I had to go and do that analysis myself from the raw data and then, put that into my notes in that sort of summary form, and then went onto: 'did anyone interview, anyone about this topic...' or 'did it come out of any of the interviews?' So, some things, had already been through a process where they'd been coded up and some things were still quite raw.

12:54 [INT] Ok. So, was there any collaboration as in that was... the analysis was a collaborative process between the different researchers or?

13:03 [T3] You could say that in that some things had already been started to be analysed. Not particularly, necessarily for my purposes, but for maybe another report that somebody else had written um and so I was building upon that analysis, and adding to it. So anything that I added code-wise, I added to the case record they took and store it in the VRE, as well. So it was collaborative but it was quite strange because um the researchers had left already, who had started doing the research. So it wasn't like we were having face-to-face conversations about it although we did occasionally meet again and we had talks but I was collaborating via the data which was in the VRE.

13:50 [INT] Ok. No... and the other thing is because I thought... was it an evaluation project? Evaluation of current approaches or it was...

13:58 [T3] Not in so many words. I think it was called the learning landscapes project because it was to give a view of what interesting things were happening or what might be an issue. We had to be very careful in the language that we use not to say that something had to be done.

14:18 [INT] Ok.

14:18 [T3] And actually, in the end, what we did was that we filtered the data right down to about ten issues that we then worded as carefully as possible as not being issues and push it to our critical friends group to say 'do you think this is something that needs to be actioned so that we do something about it' or 'do you think it's just a fact and we are perfectly happy with that' or 'do you actually think we've presented that incorrectly and we were wrong' [laughs]

14:48 [INT] Oh right [laughs], ok...

14:48 [T3] So we gave them those options and we had a discussion in a workshop based on that.

14:55 [INT] Ok, no because I was wondering, because this project it was more about sharing or describing what has been done. I was wondering why it was never open in the end and shared with the community, because it could be of value.

15:08 [T3] Yeah... and I really think it should have been. I suppose the thing was that there may have been some people who thought they were ... [reflects. thinks carefully how to

word this] controversial findings that came from it, a couple of the things may have been seen as negative reflections on the university, um they didn't want a journalist to get hold of it and interpret it in the wrong way.

15:34 [INT] Ok.

15:34 [T3] And I think they'd been very careful about where the data lies.

15:40 [INT] Ok, so then definitely, I would say then that archiving and opening it up for the public it wasn't really an aim originally.

15:50 [T3] No... I don't know about originally, maybe it had been an aim originally... [laughs] but by the middle to the end it was ruled out.

16:01 [INT] Ok, the way it was going actually, and the findings, it was kind of more leading towards not actually opening it up.

16:11 [T3] Yeah...

16:11 [INT] Ok. Let's talk now a little bit more about your data, as a researcher, the types of data you work with, and what collection methods, that sort of things.

16:20 [T3] Yes. Um, well, it's tricky to talk about data in relation to the kinds of research that I do because a lot of it it's very participatory, in which case is less about me collecting data and such, and more about those collaborative discussions and production of something, probably technology based that can be used in higher education generally. But I do tend to use methods like observation, um initially, and interviews and um maybe more ethnographic approaches to actually participating in the learning activities in higher education um When it comes to seeing how the technology is used, then sometimes I might use data tracking um.

17:18 [INT] Ok.

17:18 [T3] of virtual learning environments. I've done that before in um using data tracking and then holding a focus group to present the data to student users or staff users and asking them to talk to me about the interpretation of that data and things like that.

17:38 [INT] Is this the same students that originally participated in the first stages or it's different audiences there so you show them some of the data and then you discuss about it?

17:49 [T3] Um I would say it's probably the same users although I think I have done it presenting student use of data to staff and asking them to think about it or consider how that should affect their teaching practices in some ways so it's not necessarily the same groups. But it's just adding at that level of interpretation of the data, in a participatory way rather than me being the sole interpreter of the user data [smiles].

18:22 [INT] Ok. So, when it comes into once you have all of these data, how do you normally organise it? Well, is these data it's mostly digital data? Or you have really both, your notes as in paper-based?

18:35 [T3] Some of it will start as paper-based but I would tend to type that up um into a reflective diary or a research diary of some kind. If I've done interviews, I might transcribe them as well um some although it might start as data, I do try to turn it into electronic format of some kind um, transcriptions would be a word document, which'll be saved with the data and the name of the person who is interviewed, and then put it into a folders. I try to be quite structured with the folders [laughs].

19:08 [INT] Ok, so do you tend to organise it in a way that facilitates your later analysis or?

19:13 [T3] Yes, yeah. It's very important to be able to do that otherwise you lose where you are so um, say with the Ensemble research, we had several settings, um first of all I had Ensemble folder, and then um I'd have research settings folder and then each research

setting had its own folder within that. And then I would have interviews, or observations, audio recordings, and transcriptions which would be within that.

19:47 [INT] Ok.

19:47 [T3] Added to that, for Ensemble we had the VRE where I would upload some of those files to the folders in there and then in my research diary, when I was explaining to people what I'd been doing, and reflecting upon it I would then make links to those different files.

20:05 [INT] Yeah... so what sort of... do you add any additional contextual information um to that data?, once it's organised? So it's in the form of..., for Ensemble maybe you add some contextual data to your research diary there when you are talking about specific data? What sort of things?

20:23 [T3] Yes, that's what then happens with the research diary in that, that would be for various different days, and I would be explaining about the context and what I have done um... Before putting in a link to the audio recording or the transcript so that gave it some sense. Originally did it, so that other people would be able to read it and see what was happening, but after that it was really useful for myself to be able to remember and find the files and what were the current places. But to be honest, if I hadn't had that, I would have been desperate saving the files and I would then have thought to write at the top of the transcript a little description for myself about what I was doing [laughs].

21:09 [INT] Ok, when it comes into analysis, so once you've organised and you have all of your data there, what sort of materials do you produce for your analysis normally?

21:20 [T3] Um so far, most of the papers that I've written have drawn on case studies or vignettes so um, they would present a sort of summarised description of what had taken place in a setting in a particular time and maybe using some selected verbatim quotes to show what had been discussed um possibly putting in diagrams or images relating to what had been produced, what had taken place um [pause, reflecting] I haven't really presented video for that kind of dissemination but we have done presentations at conferences, maybe might have been a bit more multimedia and presented information, yeah.

22:18 [INT] Ok, um are you familiar, do you use any software for doing your analysis, like computer based software for analysis?

22:29 [T3] Myself, I like to transcribe using Transana which is just a cheap simple transcription software and then I do code things within transana as well, um sometimes, I've done hand-based coding if it's just a small amount or a bit simpler um but for larger and comparative interviews trying to get themes then I would use Transana rather than NVivo. More because I haven't had access to that previously, it's probably more complex than I would need it to be.

23:10 [INT] Because have you ever used this more for organising your data as well? Because also, some of these software they allow you to organise, some people indeed only use it for organising the data.

23:21 [T3] Ah, I don't know. I wasn't really aware that you could do that [laughs].

23:23 [INT] [laughs] Just in a similar fashion to what you do with your folders basically.

23:29 [T3] Ok. Yeah, um and sometimes I've got questionnaire data. I tend to deal with that in Excel. In the past I've done it with SPSS, if I was dealing with large numbers but I prefer to use Excel if I can. With smaller numbers I do use Excel.

23:47 [INT] Have you ever shared any of these data with other researchers, researchers you were collaborating with?

- 23:54 [T3] I've done that with the large scale questionnaires, yes and with SPSS, I've worked with... one other researcher was also analysing the same data.
- 24:06 [INT] Ok, how did you find that? Like, was the process easy, how did you find working with other researchers using these tools?
- 24:13 [T3] Well, it was ok cause actually I was analysing the numerical data but there had been some free text questions on the questionnaire um so the other researcher came to do the analysis in a more qualitative way on those free text questions afterwards. So we've been actually dealing with different sections of the data.
- 24:36 [INT] Ah ok, so it wasn't really working on the same data, collaborative working but... yeah ok.
- 24:42 [T3] Yeah, same questionnaire data, but not the same data.
- 24:46 [INT] Ok, so...um let me see [look at my notes] Related to the outputs of your research, what sort of information do you include there as in from your data, is mostly in the form of publications.
- 25:05 [T3] Yes I would say, journal articles, papers um presentations at conferences sometimes um, which may present a case study but a very very brief, which probably uses more imagery, like technology than data.
- 25:24 [INT] Because I was wondering what were the levels of inclusion of data, like data accompanying your publication... not really much...
- 25:33 [T3] No, it's probably highly analysed and filtered by the time it gets into kind of publication or presentation.
- 25:40 [INT] No because, I was wondering are you aware of the enhanced publications? You know the concept?
- 25:46 [T3] I have heard of the concept and I've been pretty interested in it um particularly as my master's research um I used some video-based research methods um and I felt it was a shame that I had to transcribe the videos in order to analyse them because my project write up had to be word-based so I was interested in the future, would be possible to present video as your analysed data but I haven't yet.
- 26:22 [INT] So what are your thoughts about that. Would you feel comfortable about providing some of your data accompanying your actual publication?
- 26:29 [T3] I would prefer to do that, I think it would be more valid in a way to present the data in that way but there would be ethical issues I think. I would have had to know that I planned to do that from the beginning so to get consent from participants that if they were going actually to be presented in video to the public [laughs]. But that's a different thing to consent to rather than of verbatim quotes from a transcript.
- 26:57 [T3] So in your particular experience, of the type of research you do, do you see any barriers? As in would you have any ethical problems when it comes into sharing data or presenting it like for example using this enhanced publication. Would you have there important barriers?
- 27:15 [T3] There may be um ethical issues um I suppose it would depend who had funded the research, how happy they were for those data to be shared afterwards, sometimes it's actually a requirement of the funding body.
- 27:33 [INT] Yes... ESRC for example does that sometimes.
- 27:34 [T3] But it's also possible that it might not had been like for example the Learning Landscapes project was very much in-house of the university and they wanted to keep the data for themselves.
- 27:46 [INT] Ok, but sometimes, in general, if it was possible, you see... you think about it

positively.

27:52 [T3] Yes! I have no personal problems um with doing that. I think, yes research should be shared as widely as possible um I think you would need to be careful about the level of analysis that you present because things could be interpreted um [pause] in a negative way. If you just present the raw data, or if you just presented it half-formed and it doesn't have the right context, that's what I mean um so I would probably would have quite a lot of labelling and contextual information related to the data before just...

28:32 [INT] Before going and share it. Ah ok. Because I was wondering what are your perceptions about secondary analysis and working either with your own data, how do you feel about archiving. Do you see advantages/disadvantages? What are your feelings about that?

28:49 [T3] Well, I can certainly see the advantages for... data can be used in multiple ways and as I saw with the learning landscapes project the same data we used for different themed reports. So one person had already done the analysis but they were thinking about um notions of space, so I can go through it and code it up differently if thinking about small group teaching so I can see that being um done with larger scale set of data and it would add value to generalise more within your analysis from lots of different fields without having to go out and collect the data yourself, so you could get more sort of meta-style analysis um report. So that's a positive. Like a negative might be as I was saying it would be possible to interpret the data incorrectly if you hadn't got the right context for it um and maybe if you were trying to prove something in a negative way. You could also pick and choose certain aspects of the data without thinking about how it really relates to the rest of it.

30:01 [INT] Ok, umm that's interesting. If you were to archive your data, what kind of... what would be the purposes and what audiences would you envision using it?

30:14 [T3] I suppose what jumps to mind is other academic researcher audience. Mostly researchers in higher education so that'd be other people who would be interested in higher education and researching it, in relation to technology enhanced learning but um also in management of university, um pedagogic related to higher education um. What else? [laughs]

30:43 [INT] No, that's ok.

30:45 [T3] I'm sure there's many different things.

30:48 [T3] Because I was wondering, do you have any experience with secondary analysis, either, it could be your own data or other's data. Have you used any archive, digital archive?

30:57 [T3] Um...[thinking] I've witnessed it but I don't think I've actually got my hands into it. I mean... for the Ensemble project we were making Exhibits of data that was representing that data in various different ways, so in a way, that was doing secondary analysis of different forms of data and often it was open data that had been metadata tagged and in those ways that allowed us to present it in different forms but actually, cause I'm not so technologically advanced [laughs]. Most of what I made and presented was um from data that I inputted myself so I'm not very good with dealing with databases originally but I did see other people do it, like the live earthquake data that was being streamed from RSS feeds or from the museum of Archaeology and Anthropology database, that draw on the data from that and presented in that and graphs and that sort of things.

32:15 [INT] Ok, no no, that's very interesting, that's good. I was meaning, a bit more maybe,

when it comes into your research, you are doing like your literature review there whether actually you've accessed any archives, not only the publications but maybe go and look into qualitative archives.

32:34 [T3] For literature?

32:34 [INT] aha [nods], for example, yeah

32:34 [T3] Um, I suppose, I mean I do literature searches within databases. These days I do the LJMU university database which draws on sort of the databases they are subscribed to and I'm not really very aware of what databases they are subscribed to...

32:56 [INT] So it's mostly publications really rather than go and explore qualitative digital archives. It's more publications based isn't it?

33:03 [T3] Yes... yeah

33:04 [INT] Ok, that's ok. I'd like to move now into more about teaching, especially research modules. Could you describe a little bit some of the modules, research oriented modules you teach?

33:19 [T3] The main one is the Independent project and Independent writing modules that I teach for the third year and this is for undergraduate students who have not necessarily done any research before um. These days they don't even have a second year module that supports it, so it is actually the first time that they are thinking about research so... we take it from nothing... [laughs]

33:47 [INT] Was it like that in previous years or...?

33:51 [T3] The very first year I did it, they had a module in the second year that talked about approaches to research and methodology and asked them to produce a research proposal before they got to the third year. Then last year, they changed that second year modules, it didn't really talk about um...[laughs]

34:16 [INT] Ok...

34:16 [T3] research... [pause and carefully words it] approaches, they were two lectures that just talked about methodology.

34:23 [INT] That's interesting because I would have expected to be kind of like the other way around. So you said actually, regarding the students backgrounds, actually some of them they have no research background... not even before university?

34:38 [T3] Not necessarily, no no. I mean some of them do say that they've done small-scale project before, but it's not something that I can assume they'd have done.

34:50 [INT] Ok, like in general, most of them. What are the sort of materials you use for the teaching of these modules and the organisation of it as well?

34:59 [T3] Well um, I think, first of all we start off with presenting theoretical perspectives so the epistemology in a simplified form which is quantitative and qualitative approaches um which is described based on some of the key texts. So I'm quite dependant on the Crotty um foundations of social sciences diagram that shows how these things link together um I present that in a discussion um where we talk about the sort of philosophy behind the two approaches and then we look at some papers that have taken different approaches and have a look and try to have a critical discussion about the kinds of information that they found using those different approaches. And in that case I can draw upon the couple of lectures that they've have in the second year that do similar things, I remind them...

36:05 [INT] Ah ok.

36:06 [T3] ...about that, but we would use general article papers that have researched in education and in fields familiar to them so that they can draw on some of the

educational knowledge of theory that they've got and see how it's been researched in practice to make those links there. I sometimes draw on my own research to present them how my title has changed through doing some research if we are talking through a title.

36:39 [INT] Ok, so do you use like your own real research data support teaching there?

36:43 [T3] Yeah I'd say not as... I do use data um but it's not just data, it's also the process of being a researcher that I present to them so I show them my concept map that I drew of the research when I was first starting it to show them how to do a concept map because to get them to do it.

37:05 [INT] Ok, that's interesting.

37:07 [T3] How my title developed while I do my research, so they can see while we are doing a title development or lecture... that it would change, and there are better ways of doing it. Try and make it a bit more real for them and show them um that this is done in practice, and in an educational setting.

37:26 [INT] I guess there is a part about research methods as well there.

37:30 [T3] Yes, well first of all, this year we started off doing very general lectures about um theoretical perspective, then methodology then we split so half of the group are just doing writing and so they've gone with Gill Forrester to talk more about how to critique published research and then the other group have come to sessions with me where we are gonna talk practically about how to conduct various different types of research.

38:03 [INT] Oh, so I guess those are the ones doing the independent projects.

38:07 [T3] They are the ones doing the fieldwork and those types of things um and so I would say have a session on doing questionnaires or doing interviews or doing observations in which case I drew on actual data um sometimes of my own um... I have used data from the educational evaluations archive to show them interview transcripts so interviews with researchers to see them reflecting on why they used a particular approach um or images that were produced as research data that are useful um. I used clips from youtube um of interviews being conducted. They were actually not research interviews but it was the idea of how to do interviews and power relationships and that sort of thing.

39:05 [INT] Ok.

39:07 [T3] Um I've used my own interview question schedule and handed those out and I've used um a published instrument words a ... questionnaire, that we gave them to sort of critique and look at and maybe try and use it. So little bits here and there.

39:30 [INT] Ok. How do you think using these data and the archive for example as well, do you think it has helped them to get a notion of the whole research process or?

39:42 [T3] Yes I think it really does and particularly to give them a view on how they can practically go about doing something, what something really looks like because I know that they don't have any of their own experience so when I say a questionnaire, I don't know what they are thinking about [laughs] So it helps for me to know that they've seen an actual research questionnaire, two different forms of it, so they can draw on those in their own decision so they make about how to make a questionnaire.

40:14 [INT] According to your experience with this course um what are the main... major barriers or issues that they experience when developing this project?

40:22 [T3] Um in the past, because they didn't have that research background, um they were very concerned that they didn't know what was expected from them and what it was that they were supposed to produce. They are very used to writing an essay but this is a

report with different chapters, headings in it so they don't they are not familiar with it and what the expectations are. So, um subsequent years I've been careful to bring in exemplars early on and also to make it clear that actually the structure of the project report matches up to what's in a journal article paper and so to get them to engage with general article papers more so they get to know what it is that is expected from them. So there's an issue with them understanding what is a methodology, what is a literature review, and how will they analyse their data and that sort of expertise.

41:18 [INT] I guess, when they were showed, for example, with the exemplary publications or previous years um reports, that might get them a view about what they are expected to produce but doesn't really give them any insights about the whole process and how they arrive there.

41:37 [T3] No, not to start of with um we have formative feedback as we go through. So first we ask them to fill in a topic form which makes them consider the topic and the aims, and the outcomes and then they get formal feedback from the supervisors on that. And they fill in the proposal form, that also talks about the methods, methodology and a short literature review and they get feedback from that. So we try and structure the process but at the beginning we know that they are not going to be entirely clear about it.

42:08 [INT] Has this come from previous years, feedback from them. What has been the sort of feedback you've got from students?

42:15 [T3] It has come from previous years. I've only be doing it for 3 years but the first year that I did it, I went in assuming that they knew what a research project was and that they'd already had some background to what research methods and methodology, and I was sort of orchestrating with um 'this is your methodology section, this is your literature review section bla bla bla...' so I battled on through it and I got a shocked audience who were upset... [laughs]

42:47 [INT] Ok...[laughs]

42:48 [T3] And then, you know, I had to backtrack on everything and break it down again in the next lectures so I've learnt from that now. First of all, to let them know that they will be uncomfortable and not to worry about that because it is a new thing and so to let them know I know that they haven't got an account... that I will gonna take them through one stage at a time, that they'll get feedback along the way. This is really where the supervisor and his support comes into play a lot as well so on other modules they won't have one to one access to a supervisor, which is what they get in this module.

43:30 [INT] Ok. Because the other thing I was wondering about is... do they have, when they are in the part of the research methods or you are showing them um things about analysis, are there any collaborative activities, group activities where they look at data together, and then they discuss or they do some analysis? Sort of workshop, you know, manner kind of thing.

43:51 [T3] Yeah, well, again we did that differently for writing or fieldwork, I just talk about the fieldwork one. We do that in semester 2, when we have one in qualitative analysis workshop and one in quantitative analysis and that's actually when they can, either bring in their own data or we've got some data for them to analyse, and they do that either in groups or sole, it depends um on how they feel really with the data. But that's something that we really do at the stage where they probably already got some data of their own, not something we do before they've done the research, which thinking about it... it might be helpful for them to actually look at data before they've produced any

data... um it's just sort of practical time constraints and things like that. But I think, trying to get more, might be an aside [laughs]. But more of the background to research into the second year so we can be more...

44:58 [INT] So change the way the curriculum is at the moment with that, ok. No, because I think, having to face that in the last year, without any background before...

45:09 [T3] Um, aha.

45:09 [INT] can be tough, challenging.

45:14 [T3] So I think the response to that "challenginess" so far, it's been the weighting of that module and to move more away from being research, to put less emphasis on research so it doesn't matter as much that they are not comfortable and they don't produce pretty good output [laughs...]. Actually, I think it would be better to move it back to put more emphasis earlier on, on to support research so we get higher quality outputs from the third year.

45:46 [INT] How do students feel about research themselves? Do they see any value um in it? As in...

45:55 [T3] Yes.

45:55 [INT] either for their future teaching or for themselves.

46:01 [T3] Yes, it's quite variable actually um some students do really fear the fact that it is open for them to think about research and they have very little idea about what interests them, what they want to know about it, there are always those students but actually, quite a lot of the students enjoy the fact that they can look at something in more detail that they found interesting long away, and a lot of them have enjoyed a particular module or have an interest from placement, something they've observed and they want to do in the future, that they want to look into further and so they are enthusiastic about things and research on that. Particularly, I found they've had some good voluntary work experiences or if their placement went well and they got on well with the people, actually wanted to continue in that discipline, or in that profession in the future, then they would be enthusiastic about doing some research practically in that sense. Um... some students do just feel that they are practically pushed for time and would rather do a literature based piece of research but then they enjoy the opportunity to focus on a particular topic more of their own interest but as I say, there will always be those students that are not happy with being given less guidance and that.

47:25 [INT] Right, ok, I was wondering I guess most of them they intend to be teachers after..

47:31 [T3] Quite a lot of them

47:32 [INT] Quite a lot of them, and I was wondering whether they are actually able to make the connection between how research can be useful, even for their... pragmatically for their teaching practices after.

47:45 [T3] Yes and no. Um [laughs] I think they see as far as that it would look good on their CV for applying for a PGCE um but there isn't enough emphasis really in this undergraduate module because is not a PGCE and it's not about teaching practice. We leave that for the PGCE people really, to get them to start thinking about action research and reflecting their own teaching practices um because it wouldn't be fair on the general student body to make that assumption that they all wanted to be teaching professionals or active, practical in the future I guess. But the process of reflection is something that could be put into all the modules and relate to future professional practice but it won't necessarily be doing research methods like action research style um approaches that would not necessarily reflect on their own practices, right. There is another module on

early childhood studies, which is called the reflective practitioner which is for doing research that is action research and they have a placement which is related to that module so it's all for doing action research.

49:12 [INT] What are the outputs they are expected to produce for that, are they using any reflection tools like, I don't know port-folios or something like that, when they actually talk about?

49:20 [T3] I think they are doing observations and research diaries, but still it's going to be report produced at the end, rather than a portfolio I think.

49:32 [INT] Ok. It's more like to be assessed rather than be shared between all the students and discussed about it.

49:40 [T3] Yes, it's just more for assessment, individual assessment.

49:43 [INT] Ok. I think, yeah, that's pretty much it, yeah, very good!

49:49 [T3] Ok, cool!

49:50 [INT] Thank you very much, yeah very useful.

49:54 [T3] No probs!

C.6.3 Exemplary Archives: Sample Transcript Interview with Researcher 5

- 00:21 [RES1] The history of it was, that it was a distance education course for um mainly primary school teachers and it was based like in a kind of Open University model and... it had been running since the early 1980s and in the mid 80s I was... I took over this course and I made a lot of changes to it. I re-designed it in that way.
- 00:53 [INT] So when you say changes, you mean changes in the curriculum in the way... the content of the course?
- 00:57 [RES1] Completely rewrite the course, yes, yeah. I found ways of completely rewrite it... I didn't really like the original course and so... the original design goes back to the mid 1980s and it consisted... it was a year long course for teachers to study in their on time. Um, and there were around three hundred people a year that took the course and there all over Australia and a few outside Australia but we never got to make them, we never got to save? them [the materials produced by the students were never saved in electronic forms]. Um it was all done, at that time, it was through written materials and written responses. It was pre-internet um.
- 01:46 [INT] Was mostly research-based so they would be researching something, even their own practices?
- 01:51 [RES1] Yeah, very much that.
- 01:51 [INT] So what was... That was the main aim?
- 01:54 [RES1] Yeah, it was in three sections. The first section was really giving them some ideas about how to do research in their own classrooms so to give you some examples, um, one of the things I have them to do was take photographs of themselves in their own classrooms and then they exchange some of them with people in the course and the other person had to write a commentary...
- 02:18 [INT] About?
- 02:18 [RES1] About what they saw in the photo and that was very interesting and productive. I had them record, audio record, bits of their lessons and transcribe them. I have them to do interviews um with people in school and people outside so... One of the interviews was to talk to somebody about something they've found very difficult to do and how did they lent to it. So it's a little research task like that and they spent three, four months doing this task. And then, the next section of the course, the had the Hathaway material as a case study of a school and they had tasks to do based on that material.
- 03:08 [INT] And I guess this was introduced when you re-designed the contents of the course.
- 03:12 [RES1] Yes, yes.
- 03:14 [INT] So, what were the motivations for using the Hathaway? Like, what made you use that particular case study?
- 03:23 [RES1] Um, there were case studies in the previous material but I think anybody really knew how to use them. They were just given to students to read. Um, what led me to using Hathaway was a PhD thesis. I was an examiner for a PhD thesis which... the person who wrote the thesis had produced three case studies of primary schools for um student teachers to use as the basis for their Education studies, part of their degree. And I was really impressed by the case studies, I was impressed by the way she used them and I really adopted the case studies from there. One of those case studies was the first version of Hathaway.
- 04:16 [INT] Ah... ok.
- 04:18 [RES1] But it was... at that point it was nearly all text with a few photographs and it consisted of um interviews with the teachers before the lesson, an a account of the

lesson and then the interviews after. With just some background materials on the school. Um, but that was... we started by using that material. Um, and they got some research money to do work... for me to work with that person to develop a whole lot more materials. That's why when we went to the school with a video crew and collected a whole lot more material which forms the basis of the CD.

05:03 [INT] Ok

05:03 [RES1] Um, the first versions of the Hathaway case study we used in the course which are on here [he points at two volumes of printed materials and to a video tape] so... it was this video material here.

05:20 [INT] So this was added afterwards.

05:23 [RES1] Yes, this was version 2. The first version was text and this was the second version. Um, and there was some text material with it but we begun to realise then, this was sort of 1993-4-5, in that period, that there was potential for using CD-ROM. It was CD-ROM just becoming available around that time, as I remember, um but very few people had access to it um, so the university was thinking 'oh it's all very well thinking about CD-ROM but nobody can use it', yeah. But we got a small development grant to produce the first CD-ROM and as we produced it, schools were just beginning to get computers with a CD-ROM, so people who were doing a course would have to go to school out of hours or...

06:19 [INT] Ok

06:19 [RES1] during lunchtime and get a whole of a one computer that had a CD drive um, and they could then access the material. But within 3 or 4 years um most people had them at home, I mean, that happened very quickly. So we were just lucky that we chose a technology that was emerging...

06:37 [INT] At that time...

06:37 [RES1] We didn't choose one that didn't go anywhere but we were just lucky.

06:41 [INT] Um

06:42 [RES1] And then it run for at least 10 years I suppose.

06:47 [INT] That particular version?

06:48 [RES1] Yes.

06:48 [INT] Right, so I am interested on... how were the materials been used in the teaching and learning context? How were the students... so what was the structure of that module when they started using the Hathaway materials?

07:01 [RES1] The sort of pedagogical structure remained the same from that version in that we gave them a task to do and said 'here is the material' and as far as possible, what we said was um... this is working with a case study, with case study material...it was that case record idea...

07:26 [INT] Ok

07:26 [RES1] Case record-case study, which has been in Susan's PhD as well, it goes back that far and... so we are giving you a case record and we want you to do various tasks with it. I think I overestimated... underestimated how difficult that task was.

07:46 [INT] Ok

07:46 [RES1] Because faced with a CD you've got those headings and so on... you can look at um... but you got no idea what's there and even when you start looking at it, you don't know what's relevant, what isn't... how do you decide...

07:59 [INT] That's what I thought... it's very rich...

08:01 [RES1] Yeah...[laughs]

08:01 [INT] As in... lots of materials there...so

08:02 [RES1] So, we didn't want to over program it, we didn't want to say: 'go to this, go to that, go to something else'... so we tried to find ways of giving people clues: 'start here and see what else you could find that's relevant'. And that's why the word search was so important in the first version because you could start looking into a lesson and you could see something about... I don't know... the curricula so you can go to the index and look at the curricula or you could look at history, look at English or whatever. And you could see where there were other references to that and that's why I said there was a lot really in this version [pointing to the version online, not in the CD] that we didn't have that function anymore um. The task that we gave people to do was... the original task was at the time we were collecting the material, we didn't realise until... very late on that the school had been to a school review process just recently. And the principal of the school gave me a copy of the report that reviewers... a bit like an Ofsted report.

09:13 [INT] Yeah

09:13 [RES1] But it was more open than that, gave a copy of the report and some other documents related to that so we said to people... the task was something like 'here is a copy of the report on the school' um 'you've been asked to write a response' so you write a response to the report and here is the evidence you can use. Because one of the things about those reports is that they don't use much evidence, they are very selective in what they comply to.

09:44 [INT] Right, ok

09:44 [RES1] So it allows you to kind of respond to statements that were written into the report um and then... I think that the later task that we developed after that one, was something like 'you've been asked to run a training course for people doing reviews' I think that's what it was, I can't remember the exact words now. So... you've got this case materials, 'how would you use that to develop this course for teachers'. We found it was quite hard coming out with tasks that were... they had to be realistic cause teachers reject them if they are not, you know, if it's a funny task. They have to be real education task or believable task.

10:32 [INT] Um

10:32 [RES1] Which requires them to actually make as much use as they can of the materials.

10:39 [INT] Right, so is then sort of... in a way kind of do a particular... they go through the materials, you know, do they keep track of the different places they've been going through, what materials they've selected and then was that kind of like, recorded or written in some form?

11:01 [RES1] That's a really good question. So, in Supercard [special software MAC only] it automatically kept a track of where you've been and you could print that off. You could see where you've been and then you could go back to anyone of those pages.

11:13 [INT] Did you actually then use any those functionalities to keep track of the way students were actually going through the different materials?

11:20 [RES1] Yeah

11:22 [INT] Did you do that?

11:22 [RES1] So now you can use a web browser to get back through the history to get back the pages but... I can't remember exactly but I think in the earlier versions of... I can't remember, we weren't using explorer, we were using something else.

11:36 [INT] Maybe Opera or something else...

11:39 [RES1] Yeah, they didn't have histories that were easy to find... so it was... took a while

for that to kind of catch up with the...WORD... we encourage people to do that and um...and in fact, the multimedia producer that we worked with to put all this together, he recently did his PhD in the University of Maryland and they have a um... they produced materials in Classics um, that's produced mainly by librarians so he's been trained in archiving and multimedia.

12:22 [INT] Ok.

12:22 [RES1] In using material from Archaeology and classical literature and all that sort of stuff. but with people who are primarily librarians and so he was very keen on that um. What you could do with the track through the material and he had this idea of 'if I went through it you could see my history, where I've gone through...'. You could then show that to somebody else and they could either follow your track...

12:55 [INT] Or start from somewhere else...

12:55 [RES1] I think somewhere... that's where the idea of the tour came from [he refers to the guided tours included in the online version of the CD-ROM].

13:01 [INT] Ok.

13:01 [RES1] Because we realised that you could get um you could do guided tours where you could say 'here is someone who has a special interest in language issues' and get them to produce a guided tour... or he is someone who has an interest in, I don't know, gender issues or whatever it is and get a specialist guided tour. Um, so we started doing that but we really didn't have time to develop it.

13:30 [INT] Was this intended from the beginning? As in, were you expecting people with different backgrounds to look at the material? Or it was aimed at a particular discipline.

13:42 [RES1] It had to be fairly open because, like I said, we had 250 students a year, most of whom were primary school teachers but a significant minority took the course for other reasons. So maybe, 20 or 30 people a year. We got nurses, um people working with computer software and, particularly people interested in content um we had management people and then other people from the WORD system who were interested and they were often the most interesting students so we had to adapt our tasks.

14:25 [INT] Aha

14:25 [RES1] Um and at first, they would say like things like 'I don't really know anything about schools...' and we said 'all the better, because it's good, it's an advantage' [laughs]

14:33 [INT] Ok

14:33 [RES1] Um I remember there was a nurse, with one of the very early versions, she said 'the school's got this really interesting architecture but I can't find interview with the architect...' so she said that she'll invent one: 'here it's my fictional interview with him'...

14:51 [INT] Ok

14:51 [RES1] So then we went and did a real interview with the architect so we could say '... and here is what he thought' um but she alerted us to the fact that the architecture was important.

15:03 [INT] Um ok

15:03 [RES1] Yeah and...

15:06 [INT] So I guess this is their inputs, it brings in lots of different themes so you could...

15:13 [RES1] Yeah

15:13 [INT] look at it from different perspectives there. From the teaching context, um how was the case study seen: was it seen, from the student's perspective I mean, was it seen as: you've done the research first, you produced the materials and then it was used as a

secondary source. Or it was more to promote students as inquirers so they could actually contribute with their own research to the existing materials.

15:42 [RES1] Very much the second...

15:42 [INT] So you see it more like the second?

15:44 [RES1] We saw it as the second but sometimes students would read it as being the first so they would say: in the first part of the course you take them through all of this research exercises and these were very defined and specific and they would actually say, they would have time limits, we'll say 'take 15 mins to do this'.

16:03 [INT] Ok

16:03 [RES1] Um and then you give them this case study so they assume what they have to do is go back through the tasks they've just done, using this material. And we had to say no, that won't work this is completely a different approach. This is starting at the other end, you will find it difficult, it's full of ambiguity, you will get lost in it um but we said, that's... a real school is like that, a real school is confusing and it's ambiguous and they accepted that for the most part. Um... I think... there was all that evaluation material I threw away [laughs]

16:40 [INT] [laughs]

16:41 [RES1] There were a small number of people, maybe 25 out of 250 who really liked it and they said 'this is different from anything I've done before but it's a real opportunity and I feel I've got the freedom to express myself through this material'. There were probably another 20/25 who said 'I found this really confusing, just tell me what I have to do, what's the minimum I have to do to pass this' um and then most people kind of found a way through it, they did OK, they um they've discovered things that..., you know, they found it difficult because it was so different to anything else they've ever done and the struggled with it but they did OK, yeah.

17:26 [INT] OK. So yeah...

17:29 [RES1] And then we realised we kind of had to pull more people into that first 25, you know, so what kind of help we can give to get people more involved in the material um without over-programming what we are asking them to do.

17:43 [INT] Right ok

17:44 [RES1] That was the key question [after that initial evaluation].

17:45 [INT] What sort of evaluations did you do? Did you do any evaluations after with the students? So um you get an idea of, you know, all of these things?

17:55 [RES1] Yeah, we did, we did

17:55 [INT] The problems they found...

17:58 [RES1] We had a lot of sort of formative evaluation written in so, like, every few weeks we asked them to send us um a report on where they got to, what was easy, what was difficult, what were they thinking um and that was the material I had on these boxes [laughs]

18:12 [INT] Ah ok

18:15 [RES1] But we also commissioned quite a lot of external evaluations because we used to get people who want to come and visit and look at the materials and so on... so we tried and get them to and ask them to do and we had three or four people to do evaluations of the course as a whole so there was this guy from the Open University, somebody from Germany um another, um somebody... another Open University person who came on study leave and we were looking at things for him to do and so.

18:48 [INT] Ok

- 18:50 [RES1] So we came with that task some of... a school inspector from England who came and did some really helpful telephone interviews with people and he really got me thinking about that issue of 'if you don't get people enough structure they get confused' in that they interpret the task more narrowly rather than more expansively so how can you give people a confidence that kind of engage with the materials and of course the other thing was, as it became more familiar to have CD-ROMs, a lot of the early CD-ROMs that people produced commercially were very programmed.
- 19:32 [INT] Ok, like very structured as in...
- 19:34 [RES1] Yes
- 19:34 [INT] you know, they would take you through the materials in a very linear way, is that what you mean?
- 19:39 [RES1] Yes, yes. One of the things...when we were working on it, we had... we tried to involve people from different backgrounds so we had like a video producer, this multimedia guy, um a text editor, two text editors, an audio producer as well as academics. We used to meet every couple of weeks.
- 19:58 [INT] So this is with regard to the design of the materials...
- 20:00 [RES1] During the period we were designing it and one of the things that group did was that they started bringing along um some of the CD-based material they discovered that they thought provided helpful models. So one I remember, one early one was on the Beatles.
- 20:19 [INT] Ok.
- 20:19 [RES1] One that influenced me a lot was by... I'm trying to remember his name, he's um sort of world music guy um he produces one WORD and he's taken a series of books called 'The Griffin and Sabine' books, do you know ever came across these?
- 20:39 [INT] Um no no, I've got no idea
- 20:39 [RES1] They are very well known in graphic design, it's three books and it's called the 'Griffin and Sabine' story I think it is. It consists of letters between a man and a woman and they are all illustrated, they are beautifully produced books, very heavily illustrated and the story basically is he sees this woman on a tube and somehow they start a correspondence and it becomes more and more um fantastic, it ends up on a desert island somehow [laughs] I am trying to remember the story
- 21:16 [INT] [laughs] ok
- 21:17 [RES1] But, you never really know if it's a real person he's talking to, he's writing to, or if it's a figment of his imagination cause there's sort of hints that he's been through a period of psychiatric treatment and is he writing to a fictional person or is he writing to somebody? But... and is he writing the replies or are the replies written?
- 21:40 [INT] Right
- 21:40 [RES1] It's one of these things where it's very ambiguous to follow. But it was produced as a CD-ROM and it's brilliant, I mean the whole thing is brilliant because you get the first letter and you only get the reply through... you have to work through these cues.
- 22:01 [INT] Oh, I'll have a look to see if I can find it.
- 22:01 [RES1] I don't know if I've still got a copy... um and I don't know if it would work still...
- 22:07 [INT] Well, might be online versions, you know, you can try in amazon or something like that.
- 22:13 [RES1] So there's one sequence where you get a letter and it ends up with a goldfish swimming in a bowl and the only way you can move to the next one is, you have to click, I forgot how many times, on the goldfish bowl (it was like ten times) and something

in the glass shutters and the fish swims out and then you get to the next letter. So it's all kinds of things, and each one is different, and you have to decode the cue.

22:39 [INT] So it has different elements of interaction there...

22:40 [RES1] So it doesn't just say next [laughs] you have to work out how to get at it and it gets more and more difficult to get through the sequence. Um... I can't remember his name... he is a very well-known guy who produced it and it's got WORD voices, there are voices, the voice of the girl is Isabella Rossellini and the man is, I forgot who, but a well-known actor um just got this brilliant cast and it's beautifully produced and we all thought that was great.

23:17 [INT] Ok, so to kind of inspire you to develop the structure you were looking at existing multimedia materials there.

23:23 [RES1] Yes, yeah, and the most interesting ones at that time were not in Education...

23:28 [INT] Right

23:28 [RES1] That's probably still true...

23:30 [INT] Ok

23:30 [RES1] And the other big thing that we looked at, I think, it was project 'Perseus' which is this classical material that guy worked on which it's been through a whole lot of different versions but it's the idea is to put all the resources you need to do classical research onto a desktop.

23:51 [INT] Ok

23:51 [RES1] And it's now web-based, not sure. So it's architectural science, classical science, the scholarly literature and stories, you know, everything you can think of related to classical studies and it's all there.

24:09 [INT] Ok... so before moving into, more in detail into design of the materials... what data was used as the basis for the development of the multimedia research? What research data did you look at?

24:26 [RES1] We started with the video that we already had and I intended to go back and re-edit the original video but I never got around to it. Um so we started with that and then started to add material to it, so we added some more video, we collected a lot of documentation from the school, some more interviews...

24:51 [INT] Ok, so that was going back to the research setting at the school, ok.

24:53 [RES1] Yes, and it was a little bit difficult cause we were based near Melbourne and the school was in Sidney, and... miles away, so it was expensive to go, you know, three/four people and spend time in the school.

25:08 [INT] Was it all done before actually developing the multimedia materials or actually, you... while doing the materials did you go back to collect more data?

25:21 [RES1] The first version used the material we had...

25:26 [INT] Aha

25:26 [RES1] And one of the good things about that was we had quite a lot of stuff sort of in filing cabinet drawers that we haven't been able to put in here [pointing to the previous printed version of the case study], then suddenly we could make available um so we kind of went back though archive that we had, and added material and as we were doing it um new things started to kept coming in because: 'have we got time to do another version?' or 'I've found this'.

25:57 [INT] Right, ok, did you have any major issues as in... because since you are working with materials already archived but you are actually going back and bringing new materials, and especially because you are working with children so... what were the

major issues to... if any, to actually include those materials? What I mean is like ethical issues there, maybe privacy? So I guess all of this was taken care of in advance.

26:24 [RES1] Yes and no. Um we had better consent process at this than it was normal at the time, so like I said, we got these letters from the parents that we used. And we've chosen a class of children to include that were the top end of the school so they were grade 6 which is the last year they are at the school. Grade 7 is high school in Sidney or it was...um so we knew by the time we produced the material that children were no longer going to be at the school so even if you could contact the school, you wouldn't be able to trace the children and that was important because there's been... it's one of these crazy chances but there had been an issue at the university were they produced a booklet, a bit like this one, with a photograph at the front with a primary school class and the school has given consent but the parents haven't and one of the children in the classroom was from a family where there'd been domestic abuse, the mother had moved away to another area to protect herself and the children from the father.

27:38 [INT] Wow

27:38 [RES1] Um the father was a student of the course...who identified his child and traced them back to the school, I mean, it's like a chance in a million of that could happen...

27:50 [INT] ...

27:54 [RES1] But the University got very nervous and thought 'we've gotta get consent from everyone'...so that was before it became normal to get consent. There is an issue about 'is this research?' because there are different procedures for research or publication of materials production um. I have a friend who works in the health service producing films, basically training films, and he doesn't have to go through any of the consent that researchers have to go...

28:20 [INT] Right

28:20 [RES1] through... because they don't call it research, they use some other category...

28:29 [INT] Well I guess it was the case here...?

28:31 [RES1] Yeah, it comes under the same sort of tradition of consent as you would have when publishing which is much more about copyright, much more concern about copyright...

28:41 [INT] rather than participants...

28:41 [RES1] yes

28:56 [INT] Right ok, I didn't know that. Thats very interesting...um

28:45 [RES1] Um, so what we did on the original version, I don't know if it works on this [pointing at the online version on the computer screen]. With each page, there is a little... you can call up a little picture, a little frame which tells you the date this was produced, and the copyright...

29:05 [INT] That's important yeah because...

29:06 [RES1] ... and who the um researcher was...

29:09 [INT] We can try and look but I don't think I've seen it [referring to the online version of the Hathaway].

29:15 [RES1] I think this is one of the things that got lost...

29:15 [INT] Is that part... because that brings me into the case record concept and the way of actually structuring the raw data on one hand and what you called before the 'case record'... because I guess... it doesn't include much of interpretation. Would you like to describe a little bit more about the way it was structured, um

29:38 [RES1] Yeah

29:38 [INT] on the basis of the case record... so what would you include in the case record of these materials there?

29:48 [RES1] Yes, so the Stenhouse distinction of case study and case record, um to be honest, it probably was just things we liked, things we just found interesting, so it was us... we selected things on the basis of what we thought it was interesting but it was clustered around um I suppose the core material which is the video so the more and more we begun to think about, we've got these video selections of lessons so we had the religious studies lesson, we had the dinosaur lesson, I've forgotten about other ones but um and we started to build stuff around that. So with the dinosaur stuff we started, we collected all the material that the teacher had used in putting that lesson together and then we found there were some photographs of... they got this visiting artist come to the school with some inflatable live-style dinosaur.

30:51 [INT] laughs

30:51 [RES1] So we put the photographs in, obviously, and we started to build around that material and then there were other things that um we just thought we are interested to include so at one point, we put photographs from the notice boards, staff room and we put them on. Um we got... as we were collecting the material, things arose as questions so, after we talked to the principal of the school, we realised that what went through the office, the front office, was important so I asked the people in the office if they could keep a record of the phone calls in the morning, all the faxes that came in, everyone who came to the desk so we've included those. They kept diaries for us, I mean, people were really helpful and we've included those um and then out of that, somebody said a important source was the incident book. The incident book is where they have to record all the accidents that happened or any illness among children and that intern kind of led back into the multicultural stuff. Some of the incidents were about cultural communication so there was one instance where a chinese girl, the teacher noticed she had this weird marks on her arms, and then when they looked, they actually spread across her back and they got concerned about was this some infectious condition? It turn out that it was that the child had not been well and... you know that chinese healing technique with glasses?

32:34 [INT] Yeah, yeah, I've heard about it.

32:34 [RES1] The mother has been treating the child and she'd...I think she's done something like put coins in a microwave and put them on the child skin, they actually have damaged, damaged the skin so... how do you deal with that kind of incident...

32:48 [INT] Right, ok

32:48 [RES1] So something would lead to another, you get this sort of China things, and we sort of included things but it wasn't systematic really: 'we have to have this, we have to have that...'.
33:00 [INT] Right, so... what kind of structure... what would be the structure or what descriptive information would you include in the case record? Because I guess that, you know, for different materials or compilations of things it would have a similar structure I guess... or not? When it comes into describing what materials were included in the case record, what sort of information would you include there?

33:23 [RES1] We came with those headings which were on the front screen, which...
33:28 [INT] As in: date, a few details about...

33:31 [RES1] I am trying to remember now...curriculum, classroom, um history, school history, system-wide things (administration), you know, those broad categories, which we

thought of as sort of chapters.

33:48 [INT] Ok

33:48 [RES1] for the material um so we, yeah, we did kind of lightly edit as Lawrence would say, to fit those categories. And they seem to be general enough categories that they worked out just fine.

34:04 [INT] So did you then use themes to kind of organise the materials in themes?

34:08 [RES1] Yeah, yeah Ah, one of the big categories is community so it's around what... where did this school sit in its community, which led us into collecting lots of census data which is one of the features of the materials. Um...

34:28 [INT] No because, that was going to be one of my questions, as in, were the data mostly qualitative or you included as well any survey data... quantitative data? So it's a mix really...

34:37 [RES1] yeah I was very keen on using statistical data

34:37 [INT] so you have some quantitative um

34:39 [RES1] I think it was around that time the Australian census started making a local census data so you could buy a print out of all the census data by post code um and post code pretty much overlaps with the sort of catchment area of the school which is lucky...

35:02 [INT] Um

35:02 [RES1] Um so we can get a picture of the community and then you start to get questions. The school was 78% Lebanese/Muslim children

35:07 [INT] Right ok...

35:18 [RES1] Um but the community is 30% Lebanese/Muslim so why does that happen? And then you start to find out, well, the children, the population of children is different to overall population so there are more children from that background but also there was a Catholic school in the community so the Greek children went to the Catholic school, very few Greek children here so, you know, we thought it was interesting how you get statistics about the community...

35:50 [INT] So it's outside the actual environment...

35:51 [RES1] 'How could you explain it?' And we thought... so we said to the students 'try and explain this' [laughs].

36:01 [INT] Ah no, that is very interesting

36:02 [RES1] Because there are clues...

36:02 [INT] Because you are trying to get them... to give them a better understanding of what's going on inside but you have as well the outside... what's going on outside...

36:10 [RES1] Yes

36:11 [INT] component and how that affects actually to what's going on inside. Ok, that's very good.

36:16 [RES1] And somewhere in the interviews, I can remember whether it was with the principal or people in the office, they said um because the census data is 10 years out of date, because it takes that long um or it did then um, and it's true that 10 years ago this was the status of the community, but since then, you know, we have the Somali people, the Vietnamese, we have whoever whoever moving into the area and so, the current community structure is different from the statistics... but these statistics are the ones that they use in planning decisions so we were going to all sorts of things about statistical evidence and what... how can you treat it, what are its problems...

37:03 [INT] Did you develop that further or that was really out of scope here?

- 37:08 [RES1] Um, we didn't so much with this course but one of the things that came to light as we produced it, people were asking us 'I'd like to use this with my PhD students' or 'I'd like to use this within my masters course' um and it was in those situations that you started to get those more special questions emerging.
- 37:30 [INT] Ah right, ok
- 37:32 [RES1] So someone in... not in Education but in sort of Social policy and started working on these things about um community data... oh, Criminology that's the one. So he started out by saying if we look at crime statistics and map them, they are always reported by the place where the crime was committed not by the place where the person who committed the crime lives so... in some ways it's, you know, you have to be careful about how you read the maps.
- 38:09 [INT] Um
- 38:09 [RES1] Um and so he got interested in some of the questions about community so... it was, I was interested in how it spread its usage in different directions.
- 38:21 [INT] Did you keep track of the different uses there?
- 38:20 [RES1] I tried to, I tried to but... who knows really, I gave away a couple of hundred copies of the disc and once it was on the Internet who knows how it's been used...
- 38:31 [INT] Ah right. No because I was wondering whether you kept track of whether it was extended, or if it was extended, you know, how it was used.
- 38:41 [RES1] Yeah
- 38:41 [INT] And things like that, ah no, that's...
- 38:44 [RES1] Because the other thing was, we gave 300 hundred copies a year to the students, and I don't know what...
- 38:50 [INT] Right
- 38:50 [RES1] Every so often someone would say things like, 'oh I share it to my husband' you know, cause he wanted to know what I was doing and he's a journalist or whatever... they had different interests on them... Oh! There was a head teacher of a school who kept saying to me things like "this course is rubbish and it's all stuff I know and why do I have to do it". It turned out that he was told that he had to do the course to get a promotion um and he said about the CD "I thought it was so stupid" he said, 'that I took it to the staff room and showed it to the teachers'.
- 39:30 [INT] Right
- 39:30 [RES1] And they thought it was really interesting [laughs] 'Can we have a copy' [laughs]
- 39:34 [INT] [laughs]
- 39:34 [RES1] So yeah, things like that happen.
- 39:38 [INT] OK um so that's really things... were you expecting that sort of uses in advance or not really?
- 39:46 [RES1] Um... If I stop to think about it, I could probably why people would want to use it, um but it got used much more than I thought and I showed it in a few conferences and people were just amazed by it and like, um, I showed it to... I was asked... I took it to a conference and someone in the conference said 'can you come and show it at my University' and his University had a law school and they thinking on setting up a um like a legal practice. Because a law degree doesn't qualify to practice, it's a academic study of the law. So they were setting up a year of legal practice for people who've done the degree and they were panicking about how they could organise placement for these people in the law offices and with people in law offices, really want to teach people the sort of routines. And they were trying to think about other ways of doing it... and the said

'you could have a virtual law office, you could learn all the basic stuff and it could look like this...' you know, and I don't know if they ever did it... but it's just things completely unexpected. I just couldn't think about them. I suppose in a way was a bit like Exhibit, you have this basic idea, and then you keep discovering things that might fit.

41:27 [AMG] Ok, yeah yeah. But then, do you reckon... because well, I'll go back to when we talked about design, who was involved in the design of the multimedia materials?

41:24 [RES1] There were two of us who produced the... who had the original idea, who produced this stuff [point at the printed handbook and video] um and then we were able to call in help from people so we called in help from people who did video, for instance, we didn't do it ourselves. But I cultivated a relationship with the video producer, the camera man, um because we had a video unit in the university cause it was a distance programme and he had a very... he started out with a very conventional view about videoing, you have to write a shooting script and work out all your shots, you don't start filming until you got everything written down and over a period of time, I found ways of working with him, where we would just go and film stuff [laughs]

[INT] Ok [laughs] That's good so

42:19 [RES1] And he thought of that to be really good, it was really good actually, flow the
42:20 whole documentary stuff, and we did a whole series of things like that, this was just one...

[INT] Could you just describe a bit more in detail...

42:30 [RES1] So literally I would train people into the idea and they become enthusiasts... and
42:31 then he would say, somebody from psychology would say... we go to him and say 'we want to produce this stuff in psychology' and he would say, 'well, can I have a look at this first, because this might be a more interesting way of doing it than what you are talking about'.

[INT] Ok...

42:51 [RES1] So it was influential yeah...

42:53 [INT] Ok, so in general, how was this design process? And, could you describe a little bit
42:54 more in detail all the design process you followed? So how was that structured as in... how were you working with each other? Did you have different roles for different people?

[RES1] Yeah... I'm trying to remember the order in which things happened cause it
43:25 goes back a long while but I think one of the things was... [pause] This course and some others had sort of a set of readings, you know, like read these books for students. We thought it would be helpful for students to have interviews with the authors of the books.

[INT] Ah ok...

43:52 [RES1] So I started doing audio interviews with... I'd go and find where the author of the
43:52 book was [laughs] and get him to talk about how he wrote the book and just sort of try and see what sort of people they were and there was one person particularly liked, I thought so... I continued doing it, as you could also do video so we did some video ones as well. Um so that idea of having an interview with the author was a sort of early consideration of design, as I said, I wanted to find ways of working with case study that were not just read this case study but do something with it... and I thought... I've been involved quite a lot in case studies evaluation and one of the realisations there was that it's great fun to produce case studies but they are not always so interesting to read [laughs] So I thought I gotta find a way of giving people access to the experience of

doing the case study not just read it...

[INT] But they could then develop...maybe develop their own...

45:06 [RES1] Yeah

45:09 [INT] contribute with their own way there...

45:09 [RES1] That's it and that's why the case record thing is directly useful cause you could
45:11 say 'here is a case record, your job is to produce a case study later' so you actually
have to work with the data and it's not quite as good as doing the interviews yourself or
all that sort of thing, and doing the observations but if you include bits of video, and stuff
like that is close, you know.

[INT] How much interpretations or analysis... Did the case record include any
45:34 interpretations there or analysis? Or it was more a way to work through the raw data?
How were interpretations and analysis linked with the actual case record and raw data?
Were they connected somehow?

[RES1] Only includes interpretation in the sense of 'we selected it' so there is
45:54 interpretation in the selection and some people picked up on that um... I'm trying to
think what the arguments were [long reflexive pause] It was something you actually
mentioned which was the ethics, one of the ethical issues that I ran into unexpectedly, I
should have known but... Put it at its crudest, what you do with a case study somebody
writes that's racist...somehow...

[INT] Ah ok...

46:29 [RES1] So, because of the nature of this material where it's about children from... about
46:29 migrant children, emigrant children, people would sometimes write things that... very
deprecating of the children: 'these children you know, they shouldn't be allowed in the
country, we need tighter immigration laws so this sort of thing doesn't happen', which I
think is a complete misreading of everything um but we can't think, how do you counter
that? and is... is that allowable? If we say, it's for you to produce the case study, can
you allow people to produce a case study of that kind so I begun to get quite worried
about, concerned about some of these things even though it was a very small number
of people who said those things um it was still important.

[INT] Yeah

47:22 [RES1] Um... yeah, so that became a design question but that was later. I think it was
47:22 more a reaction against standard distance education, because standard distance
education as it was in the early nineties was: here is some text, here is some stuff to
read bla to miss that.

[INT] Right, so I guess, using the Hathaway, using the case study within the course, did
47:48 it have an impact in the teachers? As in, did it change their practices at all, by using this,
while introducing this multimedia case study?

[RES1] We don't really know but unless people tell us, and there were a few occasions
48:07 where people told us um but they were more influenced by the research tasks that
proceeded.

[INT] Ok.

48:25 [RES1] So we had one... one of the little research tasks that we had was um 'sit in front
48:25 of the tape recorder without any clues to help you um, recite the names of the children
in the class' And it's a interesting exercise because there's some swedish research at
the time which said that if you asked people to do that, you get pauses in the list and the
pauses mark groups, related to some construct in their head so typically, you know, all
of the nicest/brightest children come first and then there's a pause and then you get the

kind of another group or you get all the trouble makers where... you know the pattern varies but you do get the groups.

[INT] Ok.

49:09 [RES1] Nobody ever goes through and then lists the names, you know, without any
49:09 pauses and then there's the phase where they say 'I can't say more, I can't remember, my mind's got blank' and they start to panic, you know: 'isn't it terrible, I can't remember' [laughs]

[INT] That's interesting...[laughs]

49:23 [RES1] But we got them to do that and then to think about what the groups were, were
49:24 marked by the pauses um and people found that really interesting and so they would start getting their colleagues to do it or other people, they try on other people. Or they even... some said 'oh I tried it out with my class, I got them to list the names of the children in the class and see whether they had groups' and things like that. So, the research tasks had an impact um there was one very moving one, I remember, where we ask... the one we are asking them to talk to someone about something they found very difficult to learn and this woman talked to her colleague about um, there's been a death in the school, of one of the children in the class died in a swimming accident and it was, how do you cope with that, how do you manage your class when the children are aware that somebody's died and she said we decided we were gonna tell the children everything and the children came to the funeral and the school didn't really want to do that, they wanted just to forget about it.

[INT] Umm

50:46 [RES1] And she wrote about all the stuff and she said 'I haven't been able to talk to
50:46 anyone else about this, this is the first time' um I think it was important for her and but that was kind of really exceptional case. Um, the other... there was another bit of the course where they had to choose a book and read it and keep a diary um one of the books was, I forgot what it was called now, but I think it was a little review of gender studies in education, nice, quite a nice book, a very kind of readable book and people said things like um 'I should change my relationship with my family, you know, I suddenly realised that I do everything that I can at school to treat the children the same' but she just said "but with my two boys at home, I treat them quite differently, I never expect them to wash their own clothes or...", everything's changing [laughs]

[INT] [laughs]

51:45 [RES1] But the case study... yeah, I think the aim was to make people realise that
51:45 schools are complex and we've been aware that a lot of these primary school teachers didn't think much about the school as a whole, their classroom was their world and they tended not to kind of let things outside the classroom to interfere or they felt like, you know, they felt confident in this contained space.

[INT] Ok

52:28 [RES1] And started to think about the school as a whole: school policy things, and
52:28 relationships with the community, all that, I think it did change their view of teaching in many cases. But I think it made it harder for them, not easier so...

[INT] Um ok, so it was a bit more about giving them enough contextual information. So
52:46 how was that contextual information expressed within the case study?

[RES1] Not very much, I don't think. It was more when I spoke to people or um in the
53:00 evaluation comments people would say those sort of things, yeah. It wasn't a straight forward 'once you've done this, you would do things differently'.

[INT] Yeah, ok.

53:20 [RES1] They did comment sometimes on other bits of the course, they've said 'I have
53:20 done a module on school leadership and I thought it was quite straight forward but now I
realise it's not, it's a difficult thing to do, to manage a school'. Some of my colleagues
too said 'oh I've got some students in my course that just have done yours and they are
quite different, they've got a different attitude'. So I had some of that, um, yeah, so it
made some changes but I'm not quite sure why.

[INT] Um, you don't know really...ok um

54:00 [RES1] It's an ... really, that's what it is.

54:00 [INT] Ok. So back to the design quickly. How were the materials selected, what was the
54:04 selection process for the different materials included in the case study, how did you put
the boundaries there? Because I guess you kept collecting materials and you have lots
of previous materials so, what made you decide what bits to include and what bits to
leave outside?

[RES1] We didn't reject very much, it was more what we didn't choose to collect, so we
54:32 could have collected other things and sometimes we think 'oh if only we've done this or
that... that would have been a good thing to do'. Um, it... yeah, it was pretty much, you
know, a bit like a lot of qualitative research I suppose, it's just what seems important at
the time...

[INT] Ok

54:58 [RES1] Or...people tell you things and you think 'oh I must do that', yeah.

54:58 [INT] So but then when you were collecting the materials like let's say going back to the
55:04 school and do the videos or collect, as you said, all the faxes from the day, and things
like that, did you have already in mind the actual materials, in the design of the
materials, was that driving a little bit what materials you included and which ones you
didn't?

[RES1] No, I think it was more that we had a structure that allowed us to add things. So
55:26 one of the things I've forgotten about but soon after we've been in the school a
television company went to the school and made a little video about playground
relationships and we thought 'oh we haven't really got very much on that so...' so we
added stuff on the playground and um... and we had a structure that allowed us to do
that. So, there was a point when we thought where do we stop and you could actually
put everything in the world in [smiles] and it will connect somehow [laughs]...

[INT] Ok, so then you were really adding things as you were actually designing the
56:05 different... ok

[RES1] Yeah, it's just... somebody I used to work with used to use this phrase, which is
56:10 a quote from a poem, or it's a paraphrase... the original quote was something like 'a
poem is never finished, it's only abandoned' and it adapted that to saying 'case studies
are never finished, they're just left...' [laughs] At some point you just stop...

[INT] haha, ok

56:33 [RES1] But you could go on forever um yeah.

56:34 [INT] Ok, what was the role of technologies within this design process? Did you find any
56:39 limitations? Things you wanted to do but because of the technology, you couldn't? So,
what was this influence of the technology?

[RES1] Well, there were, I mentioned to you about people's access to CD-ROM readers
56:54 being a constraint and I've forgotten when I said that. We actually did a survey, we
asked somebody to survey the students to see whether they had access, and we found

that although they didn't have the machines they could always find a way to access, like at school or something... um we used that in the university to argue for the resources to make these things um. We tended to see the technology more as enabling all kinds of new possibilities rather than restriction. We weren't nearly so aware of restrictions, it was more... because at that time producing multimedia CD-ROM was kind of a like a new thing, so what can we do, you know. Or somebody would say 'I just realise we could do this' and the guy who did the programming um he'd only recently been appointed to the university and he worked in industry or somewhere before and done very highly specific sort of programming, you know, materials, and he got really interested in, really his design, for the material design and he was the one who sort of realised you could do these things with the running transcript [context commentary about the videos with running transcriptions]. Um

[INT] Ah right...

58:26 [RES1] And he was like 'you know, I've just discovered you could do this' um which just
58:26 led onto it, which was great um and we wrote the script in that did the transcribing but... he set up the frames to make possible.

[INT] Ah right, so would you like to go a little bit through the different materials...

58:42 [RES1] Yes... it's been a long time since I looked at it.

58:47 [INT] [laughs] I was very interested actually in the different ways people can interact
58:47 with... you were mentioning before it has an indexing, as in you can search by keywords and things like that...

[RES1] yes

59:00 [INT] Maybe not in this version

59:00 [RES1] It's only the previous version, the Supercard version I think. So it's got these like
59:02 sort of like chapter headings and some of them have to do with the school, some right this is a bit sensitive [complaining about the mouse] um that's about the unit itself, where you find the assignment.

[INT] Ok, ah I see

59:22 [RES1] Um

59:22 [INT] So that's basically module related, that one there...

59:25 [RES1] Yeah...it's a bit playing games [referring to the CD] but they turn here as well, so
59:27 the thing about the disc, oh what's going on? Um try again. Um so that's about Hathaway, there's the guided tours where we were trying to experiment with finding ways through it um this is the name of the module so that's the module stuff. And then there's the school's stuff. So it's under children, teachers, lessons, parents, community, school history, school architecture, school documents, um curriculum, school organisation...

[INT] Ok

60:13 [RES1] System documents, I mean the educational system, 'other resources' it's like
60:13 additional reading.

[INT] Ok

60:21 [RES1] And then, there's a section on case study methods um which was a bit of an
60:21 after thought but people, some people asked about it.

[INT] So was that added afterwards?

60:29 [RES1] Um later on, later on. Yeah the first things we did were um the school things:
60:34 children, teachers, parents, community... um so the community one, it gives you a little description of what's in the section and then you can choose pages from the section, so

this is all that statistical stuff.

[INT] Ok

60:58 [RES1] So you can... some of it is in maps, something called the Sidney Social Atlas.

60:58 [INT] Ok, so is it somehow related to the more qualitative materials there? Are the links
61:03 made explicit? or... you know...

[RES1] No, they are not. But I think maybe in one of the guided tours, maybe just did
61:11 that. Um, some of it it's school stuff, so 'language back in our home' is school statistics.
So that's all statistical stuff there, but it's amazing how much detail there is in the
census.

[INT] Yeah, well you have even newspaper articles there.

61:39 [RES1] Yeah, when you get down, further down, you have this newspaper article, um
61:42 awards to the schools um student letters... things like that. And, I thought there was
some other things in there [looking throughout the sections in the CD] em.

[INT] Do you reckon given this... the technology at the time, did it impose a kind of
62:08 linear, like linear navigation though the materials rather than a more free kind of
exploratory way of looking at the materials?

[RES1] No, I don't think that was a problem, maybe the links was more of a problem,
62:22 you are not actually saying 'nabble up to this go and try looking at that' which might
have been a good way to go. I thought um I can't remember which section it's in. I
thought there were some other things in here, um but they might be in a different
section and I've forgotten uh maybe in one of the guided tours it might be. So this is
were we ask people to kind of sample a disc so there's my tour of the school and this is
from Louise, she's the academic I worked with and this is actually one of the students.

[INT] Um ok

63:16 [RES1] We thought, you could actually have the students' tour so I haven't look at this
63:16 for years so I'm not sure... so she talks a bit about some...

[INT] About her background there...

63:28 [RES1] Yeah, what she's done... um but I don't think she's actually got um a tour
63:28 herself, I think the tours we did, Im trying to remember now, are more like 'you go from
this page to this page'... so... maybe not, some of my ideas about the school but...
maybe it's on this version is the problem. In the Supercard version, it was actually link to
link to link...

[INT] Ah, you have actually like the track you followed through the materials, like step by
64:05 step there...what materials you've been looking at.

[RES1] That's right, yeah.

64:12 [INT] But, no, in this one doesn't look like. But that's an important part um.

64:15 [RES1] This was stuff on architecture, so um we got aerial photographs of the area of
64:21 the school so you could go from ah, the census data which has stuff about the number
of dwellings, number of people per dwelling, you could start looking at what's like on the
ground. Um, so we got, this... what we did was when we realised we got a lot of
statistical stuff on the neighbourhood um we actually just, we did a drive around the
school area.

[INT] Ok

64:59 [RES1] So it just shows you what the...

64:59 [INT] Yeah, how the area...

65:04 [RES1] And this is one of the examples where we had to reduce the size of the

65:04 document, we turned it into stills, the original was a video here. Some video on there.

[INT] Oh yes, right, it's kind of like a still image, as you were saying before.

65:18 [RES1] We had some high-school students work through early prototype, to see how
65:25 easy it was to navigate. One of them um looking through this said, it was clicking away
and said 'but there is no sound', we gotta have music [laughs]

[INT] Ok

65:38 [RES1] commentary or something... so there is actually local radio station plays in the
65:38 background... but we haven't got the sound there at the moment...for some reason. Um

[INT] No but, as I told you before, I was very interested in like in, when you have the
65:49 video and you have the transcripts going on there, yeah, like... because as you said,
that would be very useful to be able to search across different bits and pieces from the
transcriptions there. So you can access directly to the video.

[RES1] Right, right. [looking through the materials] This was, if I can get to it, visually it's
66:07 not very good quality, it's one of these things you can do much better now. But it's a plan
of the school roughly, and the school is um is built around this pods as they call them,
which is four classrooms um with a more area in between, so um there's like a year
group in each one and then sort of, there's that resource centre at the middle, so it's
quite an interesting design and then, I got the camera man to carry his videocamera
through the school at child height so the red dots mark stills from the video which show
you where you are looking at here, you're looking into this common area between the
classrooms and then it should actually take you in and walk around the school.

[INT] Ahhh

67:30 [RES1] and you can see the dots moving...

67:30 [INT] That's nice

67:33 [RES1] So.. it's a bit quick to follow you know but you can't stop it. Um I think.

67:33 [INT] Ah no, but given the time, the resources available, it's a very good job there.

67:40 [RES1] Yeah, yeah. So again, with sort of modern technology you could do that with
67:43 much better quality but the idea was...

[INT] It still gives you the idea anyway, so it works I think...

67:51 [RES1] Yeah, and then we got to know this guy, I forgot how now, but he was an
67:57 architecture student and he took the school plan and produced this 3D thing, which
again, it's standard now, but at the time it was quite...

[INT] Yeah

68:13 [RES1] so you could see this is one classroom, that's the common area... [looking at
68:13 the 3D map of the school], this is where the camera walked through...

[INT] It was pretty much open plan this one, wasn't it?

68:18 [RES1] Yeah, it's a mixture of open plan and ... this is the common area, some of them
68:20 work better than other I think. But it's trying to give somebody and idea of what it looks
like and... the architect has all these ideas he talks about in his interview...

[INT] Oh yeah, you can see you have an interview there Chris Johnson

68:43 [RES1] One of his ideas was he wanted to create a space where any children had
68:43 access to so above this classrooms, this brown area [pointing at the place in the school
map] is what he called a loft, which he deliberately made to stare up to it difficult for an
adult [laughs]

[INT] [laughs] That's pretty interesting

69:03 [RES1] But there was a little window from up there so the children could look down at
69:04 the classroom and see what was going on...

[INT] So only children were allowed there [laughs]

69:10 [RES1] That was the idea, the school didn't always found it easy to work with, yeah, but
69:13 there was... this is the architect, and it's quite a long interview um and this is done in
stills... oh at the time I was much younger [he is seeing himself in stills interviewing the
architect] [laughs] I think it was 1993 so... there is him somewhere, I don't know where
the sound is um. But this is that running transcript that goes with the image and that...
apparently when I said we've done these interviews with authors of books, this was kind
of like an equivalent, the interview with the architect of the school. It's very, very
interesting guy. He was the senior government architect and he had all these ideas of
what he wanted the school to be like and he's actually gone and talk to the school after
to see how it went. It was actually quite funny talking to him. He said he realised these
lofts um weren't such a clever idea when he was walking underneath one and the
children were spitting at him [laughs]

[INT] [laughs]

70:24 [RES1] Yeah so it's the interview, there's those 3D things, the interview with him, they'd
70:30 won the design award and it used to be here the original design award but um
something it's got...reduced, ah no this is it.

[INT] Oh right

70:49 [RES1] 1992 seems like a long time ago... an article about... from the paper about the
70:53 award, a letter from the state premiere. So that's all the architecture stuff that's there.
Then there's... um I don't know what we got about teachers... um we got.

[INT] So you've got these guided tours coming from different perspectives so you could
71:18 give examples to people, to users of ways in which they could actually interact with the
materials

[RES1] Yes

71:28 [INT] Ah ok

71:28 [RES1] Because if you are really interested about how teachers planned lessons in that
71:29 kind of open plan area, here's a video of the teachers planning big meeting and here's
the principal talking about um school and the deputy principal as well. Um... the
languages teacher um, and he was a very interesting guy, I later got him to tell me his
life-story, which is this. He grown up at Lebanon, he was training to be a lawyer, he had
to leave to Australia during the war... he was working on the trains as a guard and he
saw this advert saying they wanted arab-speaking teachers so he applied to be trained
as a teacher.

[INT] Ah, it is very interesting

72:27 [RES1] Yeah... so he is... there's his story there and then there's these other things,
72:28 these are the things from the noticeboards, weekly diaries... I mean, they've almost got
some historic interest now because there are things that are handwritten, not type-
written, so you don't see that anymore [laughs]

[INT] Probably not, yeah, no that's very good

72:59 [RES1] But you could see how you just start looking at things and the next things
73:00 follows, 'let's look at this, let's look at that'

[INT] um it's kind of very free, sometimes could even be free flowing as in you could just
73:06 start and look at different materials

[RES1] One of the things, when people said 'where do I start?' One of the things we
73:14 used to send them to was... we got the children to say who they were, so each one said
a little bit about themselves and some of them are quite interesting themselves, you can
also kind of go through them and just get some idea of the variation in the children, like

a gallery, and we deliberately chose the school that was in a city multicultural city, cause a lot of our students were rural monocultural

[INT] Monocultural?

73:58 [RES1] Yes [laughs]

73:58 [INT] Right...just to place them in a different environment...

73:58 [RES1] you know, here's a different sort of school, if you had lots of children like that would that make a difference to how you teach?

74:00 [INT] Right... yeah ok... that's very good, well I think that's good enough, that's great

74:09 [RES1] Sorry must have to stop talking about these things [laughs]

74:15 [INT] Ah no, no it's brilliant, thank you very much and I definitely think it's been very helpful, so no, it was very interesting

74:18

C.6.4 Exemplary Archives: Sample Transcript Interview with Researcher 6

- 00:04 [INT] Ok. First of all I'd like you to introduce yourself a bit and tell me a little bit about your area of research.
- 00:10 [RES2] Right ok. I'm [redacted] working on the EU Strap [redacted] which is funded under a FP7 funding and it's working on a brain computer interface. It's for people with word syndrome and my role is to testing of the system with the disabled people in order to feedback to the developers so they make changes to the system to improve it and ensure it works for people, in a sensitive and responsive way (laughs).
- 00:40 [INT] Ok. um do you have previous experiences with research archiving, so have you been archiving research data before?
- 00:50 [RES2] Um I have done research for... normally it's just paper-based or... um paper-based archiving, not networked.
- 01:01 [INT] Not networked?
- 01:02 [RES2] No.
- 01:02 [INT] Ok. For this particular project, did you have any initial data management plan? Did you design anything?
- 01:11 [RES2] Yeah. Um, yes we did in that we knew the results would be very varied and viewed by different teams so we needed to do something that um met the needs of the different teams. In addition to that, there was also requirement in the description of work, which said that we would be producing multimedia reports so that kind of made us to think about using web-based tools.
- 01:40 [INT] Who was involved in this initial data planning?
- 01:44 [RES2] No one else.
- 01:46 [INT] No one was involved?
- 01:47 [RES2] No, just the project in that we talked about it at the big meetings, you know, what people li... want: do they want separate reports, do they want to look at different bits of information and there was an agreement that, yeah, there probably would be needs by different teams.
- 02:03 [INT] Ok, so was that then established like, in the very early stages of the project?
- 02:06 [RES2] Yeah.
- 02:06 [INT] Um...did you anticipate, I guess then, the different types of data you were going to be outputting, as research outputs?
- 02:15 [RES2] Yeah, there was a very clear description of what was required us to do, some detailed case studies and case-based research to do, um generalised testing of the prototype and also develop what's called 'personas' um so there was that requirement.
- 02:34 [INT] Yeah, I'd like to talk about the personas more in detail later. Um, do you think this process... when you were producing the data, this... let's say this kind of curation or archiving, has it been integrated with the research practices?
- 02:53 [RES2] Um...
- 02:53 [INT] ...as you were going, so it was something that was going along your research?
- 02:58 [RES2] In a sense that we also had to think about the ethics, we had to go to full NHS ethics, we did get asked about our archiving practices. Also, UK commissions of guidelines on archiving as well. So we did think about what we are gonna do with project data, how long is it going to remain live for and, you know, would we reuse it afterwards and we do want to reuse the data in different ways.
- 03:28 [INT] Aha
- 03:28 [RES2] So we included that in our consent and that was done early in the project so we

included that in the NHS.

03:35 [INT] Ok, so what were the expected use of this, for after? Like after the project. the expected uses of the data.

03:44 [RES2] To probably use it for the BNCI projects so they can look at results just to see what kind of people we tested it with and what was accurately tested um so that's of interest universally. Um

03:57 [INT] Ok

03:57 [RES2] And then also for specialists working with assistive technology to look at how this kind of tools might... the tools we are developing are actually gonna help with people with disabilities.

04:13 [INT] Ok, since this is clearly a multidisciplinary project, a, so I guess

04:14 [RES2] Yeah

04:14 [INT] So I guess different people there would be expecting to have different uses of the data.

04:20 [RES2] Absolutely.

04:20 [INT] So... did you experience any barriers or relevant issues there, when it came into the archive of the data? Because I guess, due to the different... the wide variety of the data, people want to do different things with it so, was that causing any problems, how have you dealt with that?

04:39 [RES2] I mean, the only issues are anonymity, which is something that's expected and anticipated by most. Our clients, who participate in the project, don't necessarily feel they require anonymity and they are quite happy to be acknowledged and even recognised for the role they've played [laughs]

04:59 [INT] [laughs] ah ok.

05:01 [RES2] But there is a WORD, we do have project partners who are also getting results and they are in um hospitals so their ethics board absolutely requires that their data remain protected and um subjects and that... are anonymised so that's the only issue because if we are developing a case studies, then in some ways is quite nice to um use real people and if they are willing...

05:24 [INT] Ok, what about the technical people? Because I believe that in this project were you working with private companies as well?

05:31 [RES2] Yes.

05:33 [INT] So... what about their confidentiality?

05:37 [RES2] Oh... well I mean, there's real confidentiality issues over all the project... Data Bank is protected um so yes, there's a real anxiety that any of the plans about the Brainable prototype are not released before they are ready and other people shouldn't really know much about our prototypes so there's a balance between explaining our research and... actually keeping the plans for the actual prototype protected.

06:07 [INT] Ok, this is specifically for your research, as a researcher. Um when you were thinking on the archive um were you thinking about secondary analysis maybe after? Like going back to these data and do some more work on them?

06:21 [RES2] Yes, definitely. Um yeah, we were talking about that, and now even trying to obtain a new funding stream to put the case studies um up so that other people can look at it, maybe even publicly.

06:37 [INT] ...

06:41 [RES2] So... but there was protective issues about the prototype.

06:43 [INT] Ah ok, so now moving into data collection so... what data collection methods have

you used yourself?

06:50 [RES2] Um literally is paper-based which is then transferred to spreadsheets and we've used... done audio, video...

07:04 [INT] So that's for interviews I guess.

07:04 [RES2] Yeah

07:04 [INT] Do you have as well any survey data there?

07:07 [RES2] Yeah, surveys and things like that... All the data gets transferred into the spreadsheet where some of them are... WORDS here

07:27 [INT] What about the test data like the one coming from the device? I guess that's mostly numerical data.

07:19 [RES2] There's different.... yeah, some of it is fed back to the partner through a server, so it's just FTP data um we send that... the results of the... produced by the system.

07:37 [INT] Ok, so once you have all the data collected, well, have you gone through multiple iterations of the project...

07:44 [RES2] Yeah

07:44 [INT] You have different batches of data at the moment.

07:46 [RES2] Yeah

07:46 [INT] So after having collected all the data, um have you organised them in any particular ways or?

07:55 [RES2] Yeah yeah. It's organised deliberately um in multiple ways um by the subject, um by what they were tested for, um and then their own results are compared internally so their results are compared with their own results so they are also compared against other results with the other subjects... with the same condition, with different condition so...um and yeah, so there's different ways.

08:26 [INT] Have you actually, like interacted with them at all, or they've done... they have their data and do their analysis and then you have yours, the one you are interested in and then you do your own analysis or there's been actually... where you look at the same data together?

08:43 [RES2] Well, you mean as the partner ?

08:43 [INT] Yes.

08:43 [RES2] Um, yeah well, we give them the data um when we've got a problem so if we have a problem that the person can't access the system we don't think we don't think is attributable to anything else... so once we've resolved any other problems then we look at their data files, that are generated by the system, to see if there's um problems. Because the CNG's and things like that... we can look into the data and see... what...

09:08 [INT] Ok, because umm because regarding your part of the research, I guess you used mostly the data gathered in the interviews and video.

09:21 [RES2] That's from the testing sessions yeah.

09:22 [INT] um So, in the end, just after your analysis, what kind of outputs do you get from that? Because, I guess, have you kept obviously raw data separated from your analysis?

09:33 [RES2] Yeah, raw data is kept separate, I mean, and then there is a "barriers and affordances" spreadsheet where everything that's been said is thematically analysed against things that were identified earlier in the project as "barrier and affordances" to the prototype and these have been... that's the analysis. [pause] And there's a ... you know, there's demo... and we're also looking all the time for any other factors that could affect: motivation to using the prototype, um demographic factors or drugs that they are

taking that could affect to the prototype results...

10:11 [INT] Ah ok. Um, so then how have you classified all the data after your analysis?

10:20 [RES2] Oh, under the 'barriers and affordances', mainly, yeah.

10:22 [INT] Any major issues due to the nature of the data? So what have been the most difficult parts when working with these data like for organising everything and doing the analysis?

10:36 [RES2] Oh Well, one of the things is CSS, making things look nice on the screen so everyone can view differently. One is people does not understand instantly how things work so they don't bother to follow it on...

10:52 [INT] Ohh... [laughs]

10:54 [RES2] Um... and yeah, different people have different needs and it's a different, you know, have different ways of wanting to view the data: some of the developers, to inspire them, it's kind of better to look at the persona cards that we have created. For the scientists in the _____, they like looking at the charts so we have actually got other results shown through charts and things like that.

11:19 [INT] Ok, so when did you really start the design of the archive then?

11:24 [RES2] Well, the design of the archive was done with Patrick Carmichael, who is outside our project and it was a relationship that I've got so yeah, he helped designed the spreadsheet and everything based on semantic web tools from Exhibit.

11:41 [INT] Ok, so it was after all the data collection I guess...

11:43 [RES2] No, before...

11:45 [INT] It was once you had the data... ah no before?

11:45 [RES2] Yeah, before we did the data we knew what we wanted to collect, we already had an idea of how to put it um and then once we, obviously as we were collecting it, we then started looking at how we would lay it out.

11:57 [INT] So do you reckon, like, having done the design first has helped actually?

12:03 [RES2] Yeah.

12:03 [INT] Because once you knew like ways of displaying the data it could actually help you while you were collecting the data.

12:08 [RES2] Yeah that's true. Um yeah, at the moment for example, today I've accounted um problem that one of the disabilities is facing when using our system then links to medical data so now I'm gonna look at medical data and then bring that in as well so the developers could really see the nature of the problem that we are dealing with for this person to access the system so yeah... I suppose I wouldn't be thinking about that if I didn't know the way that these tools can work like that.

12:39 [INT] Ok, no that's very good. So what are the main aims in general, you reckon, of the archive?

12:44 [RES2] Um to give professionals in assistive technologies field such as IT, such as assessors um the ability to be able to see what kind of people this kind of system might work for um... an example of it working through case studies um to allow people to actually learn about um BCI and BNCI and to developers to gain ideas and insights and also for future researchers to get ideas and insights about how you can best communicate to developers.

13:20 [INT] Um, ok. Because I was thinking, given the nature of these particular data, is very sensitive data and you have obviously privacy and confidentiality issues, was data sharing in mind? Well, you were saying in the beginning that is was part of the project.

13:35 [RES2] Um... is in mind with the participants in that they know they've consented to

- have their data reused, I mean, it's implied to us each time we reuse it and tell them in what ways. They just say 'don't worry, we don't care, just use it in any way you want' but I still think it's polite.
- 13:52 [INT] Ah ok, yeah, definitely.
- 13:53 [RES2] And then we've got um issues with the actual... with... the technical partners. There are issues around releasing too much data. At the moment everything is under .htaccess. Only a very certain number of people are allowed to access the system and before we open it up we are going to collect all the data, archive it, make sure they can see it, make sure we get consent forms from everyone and then we can talk about how we really open it up.
- 14:22 [INT] Ok, so what are you going to be using for the archiving? I guess digital repositories maybe.
- 14:28 [RES2] Fedora, um yeah we are using a digital repository and we are using client side Exhibit. Nothing too heavyweight, there's not that many results there'll be about, I don't know, 20 sets of results which generate, I don't know, hundreds of records I would say. So it's pretty big.
- 14:52 [INT] It is...I mean
- 14:53 [RES2] Really, in the end it ends up being quite big, you think 'oh well there isn't that many participants' but then you know, each one does lots of runs in the system and there's also additional information collected so it ends up being quite substantial.
- 15:06 [INT] In your case, with your data, how have you described your data so it's actually descriptive enough to be displayed using.
- 15:17 [RES2] It's varied how it has been described...
- 15:19 [INT] As in, what I mean is, from the archiving point of view, have you used like any documentation standards like Dublin Core or something like that or...
- 15:29 [RES2] Yeah...
- 15:29 [INT] classification systems...
- 15:31 [RES2] We've used some... some to do with the ICF and how that's described there. Um, and yeah... I don't think we've used DC because it doesn't really work. We've used RSS obviously and... [laughs] we've used our own labels.
- 15:57 [INT] Ok, no no [laughs] that's ok!
- 15:59 [RES2] [laughs] Yeah I think most of it [she is looking at her spreadsheets describing the data] it's sort of... I don't know it's nothing there that...
- 16:10 [INT] Ah, do you mind if we do our walk-through the archive? So you show me a little bit the structure of it and the elements included in there [setting up screencast].
- 16:21 [RES2] I'm just trying to find the different...[looking at the screen], because if you go to home, I think...
- 16:24 [INT] Yeah, we are there I think, yeah
- 16:24 [RES2] Yeah, well, um hang on, there should be another page before you get to this one...
- 16:31 [INT] Ah...if you just go maybe there [going to the main site of the project]...
- 16:35 [RES2] Yeah
- 16:35 [INT] Yeah, wait where do we go...
- 16:39 [RES2] Right, if we go to 'Pilots and resources', this explains some of the... frameworks we've used...
- 16:48 [INT] Ah ok, yeah so the different prototypes they are just there...
- 16:49 [RES2] This for the ICF, we use this for the... turtle, JSON, and this are the ones we've

used for...

- 16:56 [INT] Ah ok, because that's linked data, so enables to share it... to mix it, sorry, with other sources. Yeah, do you have any... do you include here any external sources and access to other relevant data?
- 17:09 [RES2] We will be... linking to the medical data and stuff like that.
- 17:13 [INT] Um... ok. Do you want to give it a go? and then you show me the different parts there [the archive]
- 17:20 [RES2] Yeah, I also have all the spreadsheets for you if you want!
- 17:24 [INT] Ahh, that's good, umm
- 17:30 [RES2] I don't think anyone can relax with this music, do you? [laughs, referring to the background music...]
- 17:32 [INT] [laughs]
- 17:34 [RES2] If I'm doing Physio, I wouldn't want this!
- 17:36 [INT] haha... that's it there... [setting up the archive recording]
- 18:01 [RES2] Right ok...
- 18:02 [INT] Ok yes so where would you like to start? um
- 18:06 [RES2] I don't know um err yeah, ok... I mean... if we just go to the Brainable project data bank.
- 18:21 [INT] Ok....
- 18:24 [RES2] Um if we then go to maybe the participants, is a good one.
- 18:28 [INT] List of participants? or...?
- 18:30 [RES2] Yeah, the participants... they use this top menus.
- 18:33 [INT] Ah ok, that one there or?
- 18:36 [RES2] That's ok and then if you go um now to each ID you can actually see kind of... we based this on FISA who were articles about using persona cards and so our idea was to use, well my idea was to use persona card made out of CSS. The idea being the developers can print these off and then have them beside them.
- 19:02 [INT] Ok, so it's like a profile, so...
- 19:06 [RES2] yeah
- 19:06 [INT] So what sort of data then do you include for each record?
- 19:08 [RES2] Some of it it's just... we've collected data about their disability, about the function, about the participation um about er their aspirations and about their use of computers so that data goes into the profile and these are arbitrarily, randomly selected some bits... which is...
- 19:32 [INT] Ok
- 19:34 [RES2] Just to give people an idea of sort of things. Then, we've talked about... yeah then you got the technology use... this is for the MPT which is a very set way of... tabulated way of explaining the disability...
- 19:50 [INT] Ok.
- 19:50 [RES2] Yeah... so... and then there are also... because we... motivation has a very key part... um P300 attention, we also collect details about their computer self-advocacy and their attitudes to technology...
- 20:12 [INT] Ok, no because definitely given the nature of the project, it's actually a very important factor there, how comfortable they feel with the use of technologies...
- 20:20 [RES2] Yeah and now we've got BNCI test data itself and this is from the first iteration and just to explain um... [looking at a particular record] this guy is building... well this is all very technical... kind of stuff... and then any problems that they encounter,

researcher observations about why possibly things haven't reached.

20:42 [INT] So are you actually including access to the actual data?

20:46 [RES2] Yeah, the original data is actually included as well.

20:52 [INT] Ah ok

20:53 [RES2] Yeah

20:53 [INT] Ah so... how... could you then search or filter?

21:02 [RES2] Well yeah um set it up so that it can be filtered on their disability... well this are just the ones we've chosen at the moment we probably will change these [referring to the facets]. These were partly to demonstrate to people how it could work because they don't seem to have imagination [laughs]

21:19 [INT] [laughs]

21:19 [RES2] These are the ones they ended up using so things like disable... we've got controls in and we've got disabled people so that's obviously the first, look at the results for the people that were disabled, look at the results with the people that are not disabled, um look at the different sessions, whether it was a controlled session or not...

21:39 [INT] Ok

21:39 [RES2] Um different disabilities, different kinds of run that we do on the system, and then the different modes of the system. Then we got very specific BNCI related data and in fact it's P300 data and accuracy of the system so quite a variety of data. We will be actually putting more and different data in for the these situations...

22:04 [INT] So have you had any feedback from the actual users of the archive so you can...

22:09 [RES2] Yeah, two of the... well three of the... we had... it's funny cause a lot of people of the team in fact don't use it but [redacted] who are the main project partners the lead of that really liked it and maybe wants to start using it for himself for his company. He has found it useful and enters it, and so does his researcher. So our technical partner have found it useful and we are working ways of maybe we put in even the really technical files so that they can look at those as well um and then in top of that, one of the reviewers for the research project really liked it and actually said that he saw it as a possibility for the whole future BNCI!

22:57 [INT] Wow, that's a very good outcome!

22:59 [RES2] But nothing further has been said about that but it was... at the time that's what he said.

23:05 [INT] Ok, so what further plans do you have in mind?

23:12 [RES2] For use, we have planned to complete it, to make sure that all the data is there and we need to keep it there for a couple of years so that people who are maybe still writing articles and things can access it. And then, in addition, to develop it so we can transfer it across to ability.net which is my... the charity that I work for and see if we can use it in case... just case studies with people with MND [Motor Neron Diseases] and you know, make it much more disability related and then for professionals to then use as examples.

23:47 [INT] So then do you see it as in it's kind of going to be integrating other case studies as well.

23:55 [RES2] I think it will illuminate how you can do some things so it's possibly that they could add to it yeah.

24:04 [INT] So it could be not only searching across this particular project but you could actually be comparing different kinds of...

24:08 [RES2] Yeah, different kinds of assistive technologies as well, you know, so this is BNCI

- so if you can't use BNCI or is not necessary to use it, you know, you could use different ones.
- 24:22 [INT] Ok, have you had... do you reckon technologies have been sometimes a barrier when it comes to you doing your research, was it limiting some things you wanted to do? or not?
- 24:36 [RES2] Um...[long pause] Not really, it's just a problem with time... I mean, getting you know, when you are doing the research, getting the data in it's time consuming. And it's really difficult in a research project where you are having to do quick feedback.
- 24:59 [INT] Right, ok.
- 25:02 [RES2] So you need to be organised... [laughs]
- 25:03 [INT] So... ok, so you reckon then it's...?
- 25:05 [RES2] So... a form interface would have been good... I mean a form interface that allows you to put the results straight in.
- 25:21 [INT] So you reckon then... has it put a lot of time... efforts there? Apart from doing your actual research.
- 25:28 [RES2] Does and doesn't. I mean, because you've got the ability to cross...um enables more analysis than was previously possible.
- 25:36 [INT] Oh, ok, that's good.
- 25:36 [RES2] and analysis in different ways. And then, in fact, it's a big time saver but the actual kind of manual task... getting stuff onto it spreadsheets out of paper data or the things that I used to collect, it's actually time consuming...
- 25:53 [INT] But, once then, once the data is in, you reckon actually it's been very useful even for your analysis.
- 26:00 [RES2] Yeah, it saves time and allows you to think about things in different ways and display data in different ways. That enables you to um then it really saves time [laughs].
- 26:13 [INT] Ok, yes. Ok, I think that's pretty much it... so what has been the most difficult part when working with the archive?
- 26:24 [RES2] [laughs] getting people to use it.
- 26:27 [INT] So you haven't got yet that many people...
- 26:30 [RES2] We have people using it but not...
- 26:31 [INT] Not as many as you were like kind of anticipating...
- 26:35 [RES2] Well if you look at people's technology use that's very typical, to be honest. Yeah, people are very reluctant to use different interfaces, interfaces that they are not familiar with...
- 26:45 [INT] But even, you reckon even people that actually they are doing some kind of technical research.
- 26:52 [RES2] They are even more high banned? sometimes [laughs]
- 26:54 [INT] Ok, wow that's something I wouldn't... yeah I wouldn't be expecting that actually.
- 26:57 [RES2] I think sometimes those people have... even less tolerance to different interfaces, sometimes.
- 27:06 [INT] Ok, because you were saying before um it's intended for different audiences, you have technical people working with it, um
- 27:15 [RES2] and non-technical
- 27:16 [INT] non-technical as well but then, you know, how have you done, when it comes into the design, because I guess you have to come up with a kind of combined design so it actually needs both...
- 27:26 [RES2] Yeah, the filtering is bad because, the filtering um has to be explained before

anyone uses it, and then to take the filters off cause people forget to take the filter off so they then end up with quite a refined research and then, you know, they don't really understand it... so you have to explain that and the moment you go into explaining things then people think it's difficult... or it's going to be difficult...or a hassle to use...

27:52 [INT] Yeah but I guess, you know, it's good actually that you give them some initial training and then, but once you know it's not very complicated to use so once they get the training on it...

28:03 [RES2] Yeah, but they are not, they are not used to it, they are not used to a search like this so it's still a new technology to people...but when they understand it then they are really glad of it.

28:16 [INT] And then you get positive feedback actually yeah, that's good.

28:18 [RES2] But it's that initial understanding and also people having to work in different languages, so it would be good if we... it was translated um you know like google translates words on it, or all or something like that.

28:32 [INT] Ah, that shouldn't be technically challenging or ... it should be kind of straight away. Ok, thank you very much, that's been... it's been very helpful.

28:42 [RES2] Yeah? Good! You can switch off! [laughs]

BIBLIOGRAPHY

- ARL Digital Repository Issues Task Force (2009). The research library's role in digital repository services. *Research on Institutional Repositories: Articles and Presentations*. Paper 34. <http://digitalcommons.bepress.com/repository-research/34>.
- Baugh, J. B., Hallcom, A. S., and Harris, M. E. (2010). Computer assisted qualitative data analysis software: a practical perspective for applied research. *Revista del Instituto Internacional de Costos*, 6:69–81.
- Baxter, P. and Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*, 13(14):544–559.
- Bentz, V. M. and Shapiro, J. J. (1998). *Mindful enquiry in social research*. Thousand Oaks, CA: Sage.
- Berners-Lee, T. (2000). Semantic web on XML. Retrieved from <http://www.w3.org/2000/Talks/1206-xml2k-tbl>.
- Berners-Lee, T. (2009). The next Web. In *TED 2009*. Retrieved from http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific Am.*, 284(5):34–43.
- Bishop, L. (2005). Protecting respondents and enabling data sharing: Reply to Parry and Mauthner. *Sociology*, 39(2):333–336.
- Bishop, L. (2006). A proposal for archiving context for secondary analysis. *Methodological Innovations Online*, 1(2):10–20.
- Bishop, L. (2007). A reflexive account of reusing qualitative data: Beyond primary/secondary dualism. *Sociological Research Online*, 12(3):Art. 2.
- Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *The Australian Journal of Social Issues*, 44(3):255–272.
- Bishop, L. (2012). Using archived qualitative data for teaching: practical and ethical considerations. *International Journal of Social Research Methodology*, 15(4):341–350.

- Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked data - the story so far. *Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22.
- Borgman, C. L. (1999). What are digital libraries? Competing visions. *Inf. Process. Manage.*, 35(3):227–243.
- Borgman, C. L., Bates, M. J., Cloonan, M. V., Efthimiadis, E., Gilliland-Swetland, A., Kafai, Y., Leazer, G. L., and Maddox, A. (1996). Social aspects of Digital Libraries. Technical report, Final report to the National Science Foundation. Retrieved from http://is.gseis.ucla.edu/research/dig_libraries/UCLA_DL_Report.html.
- Britain, S. and Liber, O. (2006). A Framework for the Pedagogical Evaluation of Virtual Learning Environments. Report, JISC (Joint Information Systems Committee). Updated version of the report published in 1999, <http://jisc.ac.uk/publications/reports/1999/pedagogicalvlefina1.aspx>.
- British Library (2012). British National Bibliography (BNB) linked data platform. <http://bnb.data.bl.uk>.
- Broom, A., Cheshire, L., and Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, 43(6):1163–1180.
- CAQDAS Networking Project (2012). Choosing an appropriate caqdas package. Retrieved from <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/support/choosing/>.
- Carmichael, P. (2002). Extensible markup language and qualitative data analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(2):Art. 13.
- Carmichael, P. (2008). *The Handbook of Online Research Methods*, chapter Secondary Qualitative Analysis using Internet Resources, pages 385–402. London, UK: Sage Publications Ltd.
- Carmichael, P. and Procter, R. (2006). IT for education research: using new technology to enhance a complex research programme. Teaching and Learning Research Programme Research Briefing No.16. Retrieved from http://www.tlrp.org/pub/documents/no16_procter.pdf.
- Carusi, A. and Reimer, T. (2010). Virtual research environment collaborative lanscape study. Research project report, Oxford e-Research Centre, University of Oxford and Centre for e-Research, King's College London.
- Castells, M. (2000). *The Rise of The Network Society: The Information Age: Economy, Society and Culture*. Oxford, UK: Wiley, 2nd edition.
- Castells, M. (2005). *The Network Society: From Knowledge to Policy*, chapter The Network Society: From Knowledge to Policy. Washington, DC: Johns Hopkins Center for Transatlantic Relations.

- CENDARI (2012). Collaborative European Digital/Archival Infrastructure. <http://www.cendari.eu/>.
- Cheek, P. (2005). Aging well with smart technology. *Nursing Administration Quarterly*, 29(4):329–338.
- Cheshire, L. (2009). Archiving qualitative data: Prospects and challenges of data preservation and sharing among australian qualitative researchers. Retrieved from http://assda.anu.edu.au/forms/AQuAQualitativeArchiving_DiscussionPaper_FinalNov09.pdf.
- Cleveland, H. (1982). Information as resource. *The Futurist*, pages 34–39.
- Cohort [Def. 1] (n.d.). Cambridge Dictionaries Online. Retrieved from <http://dictionary.cambridge.org/dictionary/british/cohort>.
- Consultative Committee for Space Data Systems [CCSDS] (2012). *Reference model for an Open Archival Information System (OAIS)*. CCSDS 650.0-M-2, Magenta Book. Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>.
- Corti, L. (2005). Qualitative archiving and data sharing: Extending the reach and impact of qualitative data. *IASSIST Quarterly*.
- Corti, L. (2006). Editorial. *Methodological Innovations Online [Special Issue: Defining context for qualitative data]*, 1(2):1–9.
- Corti, L. (2007). Re-using archived qualitative data - where, how, why? *Journal of Archival Sciences*, 7:37–54.
- Corti, L. (2008). Data exchange tools and utilities (DEXT) Repositories and Preservation Tools Programme. Project final report, UK Data Archive.
- Corti, L., Bhat, A., and L'Hours, H. (2007). Data exchange and conversion utilities and tools (DEXT). CAQDAS Conference 2007, Egham, Surrey.
- Corti, L. and Bishop, L. (2005). Strategies in teaching secondary analysis of qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(1):Art. 47.
- Corti, L. and Blackhouse, G. (2000). Qualitative archives: Short descriptions - qualitative data archival research. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social [Online Journal]*, 1(3):Art. 31. Retrieved from <http://www.qualitative-research.net/fqs-texte/3-00/archive/qualidatae.htm>.
- Corti, L., Foster, J., and Thompson, P. (1995). Archiving qualitative research data. *Social Research Update*, 10.
- Corti, L. and Gregory, A. (2011a). Caqdas comparability. what about caqdas data exchange? *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1):Art. 35.

- Corti, L. and Gregory, A. (2011b). Report back from the DDI Qualitative Working Group. In *IASSIST 2011*, Vancouver, Canada.
- Corti, L. and Thompson, P. (1998). Are you sitting on your qualitative data? Qualidata's mission. *International Journal of Social Research Methodology*, 1(1):85–89.
- Corti, L. and Thompson, P. (2004). *Qualitative Research Practice*, chapter Secondary Analysis Of Archived Data. London, UK: Sage.
- Corti, L., Witzel, A., and Bishop, L. (2005). On the Potentials and Problems of Secondary Analysis. An Introduction to the FQS Special Issue on Secondary Analysis of Qualitative Data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(1):Art. 49.
- Crabtree, A., Greenhalgh, C., French, A., Rodden, T., and Benford, S. (2006). Exploiting Digital Records: New Resources and Tools for Qualitative Research in Contemporary Social Science. In *Words Instead of Numbers: The Status of Software in the Qualitative Research World*. Association for Survey Computing.
- Creswell, J. W. (2007). *Qualitative Inquiry and Research Design Choosing Among Five Approaches*. Thousand Oaks, CA: Sage Publications Inc., 2nd edition.
- Creswell, J. W. (2013). *Qualitative Inquiry and Research Design Choosing Among Five Approaches*. Sage Publications Inc.: Thousand Oaks, California, 3rd edition.
- Dutton, W. H. (2004). *Social Transformation in an Information Society: Rethinking Access to You and the World*, chapter ICTs and society: the evolution of different perspectives. Number 3. United Nations Educational, Scientific and Cultural Organization.
- ESRC (2012). Enter the World of Datasets. <http://www.esrc.ac.uk/impacts-and-findings/features-casestudies/features/19231/enter-the-world-of-datasets.aspx>.
- E&T 2010 Programme (2010). Learning, Innovation and ICT. Programme report, EU Knowledge System for Lifelong Learning.
- Fielding, N. (2000a). The shared fate of two innovations in qualitative methodology: The relationship of qualitative software and secondary analysis of archived qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3):Art. 22.
- Fielding, N. (2004). Qualitative research and e-social science: appraising the potential. Report, ESRC.
- Fielding, N. and Cisneros-Puebla, C. A. (2009). CAQDAS-GIS convergence: Toward a New Integrated Mixed Method Research Practice? *Journal of Mixed Methods Research*, 3(4):349–370.
- Fielding, N. and Lee, R. M. (2002). New patterns in the adoption and use of qualitative software. *Field Methods*, 14(2):197–216.

- Fielding, N. G. and Fielding, J. L. (2000). Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance. *Sociology*, 34(4):671–689.
- Fielding, R. (2000b). *Architectural Styles and The Design of Network-based Software Architectures*. PhD thesis, University of California, IRVINE. Chapter *Representational State Transfer (REST)*. http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm.
- Fink, A. (2000). The role of the researcher in the qualitative research process. A potential barrier to archiving qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3):Art. 4.
- Gartner, R. (2008). Metadata for Digital Libraries: state of the art and future directions. Technology & standards watch report, JISC (Joint Information Systems Committee).
- Graham, J., Grewal, I., and Lewis, J. (2007). Ethics in social research: the views of research participants. Prepared for the Government Social Research Unit, Research Unit by National Centre for Social Research (NatCen).
- Groundwater-Smith, S. (1989). *A process of critical enquiry: the evolution of curriculum knowledge in preservice teacher education*. Thesis. University of Sydney.
- Hakkarainen, K., Palonen, T., Paavola, S., and Lehtinen, E. (2004). *Communities of Networked Expertise: Professional and Educational Perspectives (Advances in Learning and Instruction)*. London, UK: Elsevier Ltd., 1st edition.
- Halverson, E. R., Bass, M., and Woods, D. (2012). The process of creation: a novel methodology for analyzing multimodal data. *The Qualitative Report*, 17(21):1–27.
- Hammersley, M. (1997). Qualitative Data Archiving: some reflections on its prospects and problems. *Sociology*, 31(1):131–142.
- Hammersley, M. (2010). Can we re-use qualitative data via secondary analysis? notes on some terminological and substantive issues. *Sociological Research Online*, 15(1):5.
- Harnad, S. (2007). Mandates and Metrics: how open repositories enable universities to manage, measure and maximise their research assets.
- Harnad, S. (2012). What is Open Access? <http://www.eprints.org/openaccess/>.
- Harris, S. (2012). Moving towards an open access future: the role of academic libraries. Retrieved from <http://www.uk.sagepub.com/repository/binaries/pdf/Library-OAReport.pdf>.
- Harvey, P. (2010). Applying social systems thinking and community informatics thinking in education. In Rudestam, K. and Schoenholtz-Read, J., editors, *Handbook of online learning*, pages 91–128. Thousand Oaks, CA: Sage Publications, Inc.

- Hayes, H. (2005). Digital Repositories: helping universities and colleges. JISC briefing paper. Retrieved from <http://www.jisc.ac.uk/media/documents/publications/repositorybpfe.pdf>.
- Healey, M. and Jenkins, A. (2009). Developing undergraduate research and inquiry. Retrieved from http://www.heacademy.ac.uk/assets/documents/resources/publications/DevelopingUndergraduate_Final.pdf.
- Heaton, J. (1998). Secondary analysis of qualitative data. *Social Research Update*, (22).
- Heaton, J. (2004a). *Reworking Qualitative Data*, chapter What is Secondary Analysis?, pages 2–16. London, UK: Sage Publications Ltd.
- Heaton, J. (2004b). *Reworking Qualitative Data*. London, UK: Sage Publications Ltd.
- Hedges, M. and Blanke, T. (2008). Federated access management in digital repositories. Workshop report, JISC (Joint Information Systems Committee).
- Heery, R. and Anderson, S. (2005). Digital repositories review. Technical report, UKOLN and Arts and Humanities Data Service. Retrieved from http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf.
- HM Government (2012). Open data white paper unleashing the potential. Retrieved from http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf.
- Holstein, J. and Gubrium, J. (2004). *Qualitative Research Practice*, chapter Context: working it up, down, and across. London, UK: Sage Publications Ltd.
- Hoyle, L., Corti, L., Gregory, A., Martínez-García, A., Wackerow, J., Alvar, E., Betancort-Cabrera, N., Gallagher, D., Gebel, T., Haumataki, J., Kuula, A., McEachern, S., and Zuell, C. (2013). A Qualitative Model for DDI. *DDI Working Papers Series*. doi:10.3886/DDIWorkingPaper05.
- Huynh, D., Karger, D., and Miller, R. (2007). Exhibit: lightweight structured data publishing. In *Proceedings of the Sixteenth International World Wide Web Conference (WWW2007)*, Banff, Alberta, Canada.
- Hycner, R. H. (1999). *Qualitative Research*, volume 3, chapter Some guidelines for the phenomenological analysis of interview data, pages 143–164. London, UK: Sage Publications Ltd.
- International Institute for Educational Planning (2007). Building knowledge societies: technology and education. *IIEP Newsletter*, XXV(2). Retrieved from http://www.iiep.unesco.org/fileadmin/user_upload/pdf/apre07.pdf.
- Irwin, S., Bornat, J., and Winerton, M. (2012). Timescapes secondary analysis: comparison, context and working across data sets. *Sage Qualitative Research*, 12(1):66–80.
- Jacob, N. (2009). A project to identify successful models for embedding repositories in research management systems and processes within higher education institutions. Technical report, JISC (Joint Information Systems Committee).

- JISC (2009). Virtual Research Environment Programme. <http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>.
- Kelle, U. (1997). Computer-assisted analysis of qualitative data. *LSE Methodology Institute*. Paper prepared for the Discussion paper series of the LSE Methodology Institute.
- Kelle, U. (2001). Sociological explanations between micro and macro and the integration of qualitative and quantitative methods. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 2(1):Art. 5.
- Key Perspectives Ltd. (2010). Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. Retrieved from <http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP%20SYNTHESIS.pdf>.
- Kiecolt, K. J. and Nathan, L. E. (1985). *Secondary Analysis of Survey Data: Introduction*. Quantitative Applications in the Social Sciences. London, UK: Sage Publications Ltd.
- Kinney, J. (2012). RISK - A student-centred pedagogical model for Teaching and Learning at George Brown College. Retrieved from <http://tinyurl.com/pw6unl7>.
- Kuckartz, U. (2010). Realizing Mixed-Methods Approaches with MAXQDA. Retrieved from <http://www.maxqda.com/download/MixMethMAXQDA-Nov01-2010.pdf>.
- Kuula, A. (2000). Making qualitative data fit the "data documentation initiative" or vice versa? *FQS*, 1(3):Art. 19.
- Kuula, A. (2005). Ethics reconsidered. Secondary Analysis in Qualitative Research. Utopia and Perspectives. International Symposium. MHS-Alpes Grenoble.
- Kuula, A. (2010). Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly*, 34(3):12–17.
- Kuula, A. and Borg, S. (2008). Open access to and re-use of research data – the state of the art in finland. Technical report, Finnish Social Science Data Archive.
- Kvale, S. (2007). *Doing Interviews*. London, UK: Sage Publications Inc.
- Lagoze, C., Payette, S., Shin, E., and Wilper, C. (2006). Fedora: an architecture for complex objects and their relationships. *Int. J. Digit. Libr.*, 6(2):124–138.
- Lakkala, M., Muukkonen, H., Paavola, S., and Hakkarainen, K. (2008). Designing pedagogical infrastructures in university courses for technology-enhanced collaborative inquiry. *Research and Practice in Technology Enhanced Learning*, 3(1):33–64.
- Laterza, V., Carmichael, P., and Procter, R. (2007). The doubtful guest? a Virtual Research Environment for education. *Technology, Pedagogy and Education*, 16(3):249–267.
- Laurillard, D. (2002). Rethinking teaching for the knowledge society. *Educause Review*, 37(1):16–25.

- Lave, J. and Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lehtinen, E., Hakkarainen, K., Lipponen, L., Rahikainen, M., and Muukkonen, H. (1999). Computer supported collaborative learning: A review of research and development. CL-Net Project. The J. H. G. I. Giesbers Reports on Education (number 10), Department of Educational Sciences, University of Nijmegen.
- Lewins, A. (2001). *Researching Social Life*, chapter Computer-Assisted Qualitative Data Analysis, pages 302–323. London, UK: Sage Publications Ltd.
- Lewins, A. and Silver, C. (2009). Choosing a CAQDAS Package. Retrieved from <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/files/2009ChoosingaCAQDASPackage.pdf>.
- Lieber, E., Weisner, T. S., and Presley, M. (2003). Ethnonotes: An internet-based field note management tool. *Field Methods*, 15(4):405–425.
- Litherland, K., Morris, S., and Martinez-Garcia, A. (2012). Open educational resources: Autokitty. Online Project Blog. <http://ensembleljmu.wordpress.com/oers/>.
- Magnan, F., Habel, G., and Fournier, N. (2007). Building knowledge ecosystems in web 3.0. In *The 4th Annual International Conference on User Centered Knowledge Environments: from theory to practice (I2LOR 2007)*, Montreal, Canada.
- Mangabeira, W. C., Lee, R. M., and Fielding, N. G. (2004). Computers and qualitative research: Adoption, use, and representation. *Social Science Computer Review*, 22(2):167–178.
- Mason, J. (2007). 'Re-using' qualitative data: on the merits of an investigative epistemology. *Sociological Research Online*, 12(3):Art. 3.
- Mauthner, N. and Parry, O. (2009). Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3):291–307.
- Mauthner, N., Parry, O., and Backett-Milburn, K. (1998). The data are out there, or are they? implications for archiving and revisiting qualitative data. *Sociology*, 32(4):733–745.
- Mazzochi, S., Garland, S., and Lee, R. (2005). SIMILE: Practical Metadata for the Semantic Web.
- McLellan, E., MacQueen, K. M., and Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field Methods*, 15(1):63–84.
- Miles, M. and Huberman, A. (1994a). *Qualitative Data Analysis, an expanded sourcebook*, chapter Introduction, pages 1–15. Thousand Oaks, CA: Sage Publications Inc., 2nd edition.
- Miles, M. and Huberman, A. (1994b). *Qualitative Data Analysis, an expanded sourcebook*, chapter Focusing and binding the collection of data: further design issues., pages 40–89. Thousand Oaks, CA: Sage Publications Inc., 2nd edition.

- Moore, N. (2006). The contexts of context: Broadening perspectives in the (re)use of qualitative data. *Methodological Innovations Online*, 1(2):21–32.
- Moore, N. (2007). (re)using qualitative data? *Sociological Research Online*, 12(3):Art. 1.
- Muhr, T. (1991). Atlas/ti — a prototype for the support of text interpretation. *Qualitative Sociology*, 14:349–371.
- National Committee of Inquiry into Higher Education (1997). Higher Education in the Learning Society. (Dearing Report, Summary) Retrieved from <https://bei.leeds.ac.uk/partners/ncihe/>. Norwich: HMSO.
- Neary, M. and Winn, J. (2009). *The future of higher education: policy, pedagogy and the student experience*, chapter The student as producer: reinventing the student experience in higher education, pages 192–210. London, UK: Continuum.
- O'Reilly, T. (2006). Web 2.0 compact definition: Trying again. Retrieved from <http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html>.
- Paavola, S. and Hakkarainen, K. (2005). The knowledge creation metaphor: an emergent epistemological approach to learning. *Science & Education*, 14:535–557.
- Paavola, S., Lipponen, L., and Hakkarainen, K. (2002). Epistemological foundations for CSCL: a comparison of three models of innovative knowledge communities. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, CSCL '02, pages 24–32. International Society of the Learning Sciences.
- Parry, O. and Mauthner, N. (2005). Back to Basics: Who Re-uses Qualitative Data and Why? *Sociology*, 39(2):337–342.
- Parry, O. and Mauthner, N. S. (2004). Whose data are they anyway? *Sociology*, 38(1):139–152.
- Peters, M. (2007). *Knowledge economy, development and the future of higher education*. Rotterdam and Taipei: Sense Publishers.
- Peters, M., Liu, T.-C., and Ondercin, D. (2012). Open Learning Systems: The Next Evolution of Education. In *The Pedagogy of the Open Society*, volume 1 of *Open Education*, pages 55–65. SensePublishers.
- Richards, L. (1998). Closeness to Data: the changing goals of qualitative data handling. *Qualitative Health Research*, 8(3):319–328.
- Richards, T. and Richards, L. (1991). The NUDIST Qualitative Data Analysis System. *Qualitative Sociology*, 14:307–324.
- Schutt, R. K. (2012). *Investigating the social world*, chapter Qualitative Data Analysis, pages 320–357. Thousand Oaks, CA: Sage Publications Inc., 7th edition.

- Schwandt, T. A. (2007). *The SAGE Dictionary of Qualitative Inquiry*. Thousand Oaks, CA: Sage Publications Inc., 3rd edition.
- Seidel, J. V. (1998). Qualitative Data Analysis. Originally published as Qualitative Data Analysis, in *The Ethnograph v5.0: A Users Guide*, Appendix E, 1998, Colorado Springs, Colorado: Qualis Research.
- Shadbolt, N., Berners-Lee, T., and Hall, W. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96–101.
- Shaohua, H. and Peilin, W. (2008). Web 2.0 and social learning in a digital economy. In *Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. IEEE International Symposium on*, pages 1121–1124.
- Silver, C. (2012). Qualitative software planning seminar. a comparative overview of various software packages which assist in the analysis of qualitative (textual or multimedia) data and/or facilitate the analysis of mixed methods projects. [Seminar slides] Retrieved from <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/files/SWP%202012.pdf>.
- Slavnic, Z. (2013). Towards qualitative data preservation and re-use—policy trends and academic controversies in uk and sweden. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(2):Art. 10.
- Slotta, J. D. (2010). *Classroom of the Future. Orchestrating collaborative spaces*, chapter Evolving the classrooms of the future: The interplay of pedagogy, technology and community, pages 215–242. Rotterdam, The Netherlands: Sense Publishers.
- Sobal, J. (1981). Teaching with secondary data. *Teaching Sociology*, 8(2):149–170.
- Stenhouse, L. (1978). Case study and case records: Towards a contemporary history of education. *British Educational Research Journal*, 4(2):21–39.
- Strauss, A. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage, 2nd edition.
- TEI (1994). TEI: Text Encoding Initiative. <http://www.tei-c.org/index.xml>.
- The Europeana Foundation (2010). Europeana. <http://en.wikipedia.org/wiki/Europeana>.
- Thompson, P. (2004). Pioneering the life story method. *International Journal of Social Research Methodology*, 7(1):81–84.
- Triplestore (2013). In *Wikipedia*. Retrieved from <http://en.wikipedia.org/wiki/Triplestore>.
- UKDA (2013). Transcription. <http://ukdataservice.ac.uk/manage-data/format/transcription.aspx>.

- United Nations Educational, Scientific and Cultural Organization (2005). *Towards Knowledge Societies. UNESCO World Report*. UNESCO Publishing.
- Vallima, J. and Hoffman, D. (2008). Knowledge society discourse and higher education. *Higher Education*, 56(3):265–285.
- Van den Eynden, V., Bishop, L., Horton, L., and Corti, L. (2010). Managing and sharing data. Retrieved from http://www.data-archive.ac.uk/media/203597/datamanagement_socialsciences.pdf.
- Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., and Horton, L. (2011). Managing and sharing data. Retrieved from <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>.
- Vanderfeesten, M. (2011). Use Case Enhanced Publications. http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_Enhanced_Publications.
- Vardigan, M., Heus, P., and Thomas, W. (2008). Data documentation initiative: Toward a standard for the social sciences. *International Journal of Digital Curation*, 3(1):107–113.
- Voss, A. and Procter, R. (2009). Virtual Research Environments in scholarly work and communications. *Library Hi Tech*, 27(2):174–190.
- W3C (2004). Resource description framework (RDF): Concepts and abstract syntax. W3C recommendation. <http://www.w3.org/TR/rdf-concepts/>.
- W3C (2004). RDF Primer Recommendation. <http://www.w3.org/TR/REC-rdf-syntax/>.
- W3C (2013). Linked Data. <http://www.w3.org/wiki/LinkedData>.
- Walker, R. (2002). Case study, case records and multimedia. *Cambridge Journal of Education*, 32(1):109–127.
- Weert, T. (2006). Education of the twenty-first century: New professionalism in lifelong learning, knowledge development and knowledge sharing. *Education and Information Technologies*, 11(3-4):217–237.
- Welman, C., Kruger, F., and Mitchell, B. (2006). *Research Methodology*. Southern Africa, SA: Oxford University Press Southern Africa, 3rd edition.
- Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.
- Wenger, E. (2009). *Learning in the network society and the digitized school*, chapter Preface: Educating World Citizens: Leveraging the Potential of ICT. New York, NY: Nova Science Publishers, Inc.

- Wilson, A., Rimpiläinen, S., Skinner, D., Cassidy, C., Christie, D., Coutts, N., and Sinclair, C. (2007). Using a virtual research environment to support new models of collaborative and participative research in scottish education. *Technology, Pedagogy and Education*, 16(3):289–304.
- Winn, W. (2002). Research into practice: Current trends in educational technology research: The study of learning environments. *Educational Psychology Review*, 14(3):331–351.
- Yin, R. K. (2009). *Case Study Research, Design and Methods*. Sage Publications Inc.: Thousand Oaks, California, 4th edition.