



# Data Service Infrastructure for the Social Sciences and Humanities

EC FP7

Grant Agreement Number: 283646

## **Deliverable Report**

Deliverable: D8.7

Deliverable Name: Qualitative Data Workshop Report

Deadline: M12

Nature: Public

Responsible: Jane Gray, National University of Ireland, Maynooth

Work Package Leader: Hanne Fersøe, University of Copenhagen

Contributing Partners and Editors: Jane Gray and Claire Byrne, National University of Ireland, Maynooth

# Contents

<b>1. Executive Summary</b>	<b>1</b>
<b>2. Background and Context</b>	<b>2</b>
<i>2.1 Background</i>	
<i>2.1 Context</i>	
<i>2.1 Organization of the workshop</i>	
<b>3. Data Archiving</b>	<b>4</b>
<i>3.1 Epistemological concerns</i>	
<i>3.2 Overcoming bottlenecks in data management for archiving</i>	
<b>4. Shared Data Access and Enrichment</b>	<b>5</b>
<i>4.1 Expanding the understanding of access</i>	
<i>4.2 Developing enhanced mechanisms for exploring, citing and linking qualitative data</i>	
<b>5. Legal and Ethical Issues</b>	<b>7</b>
<b>4. Conclusion</b>	<b>8</b>
<b>7 Appendices</b>	<b>10</b>
<i>7.1 Workshop programme</i>	
<i>7.2 List of participants</i>	
<i>7.3 Photograph of participants</i>	

## 1. Executive Summary

The Qualitative Data Workshop (Task 8.5) took place on 21<sup>st</sup> November 2012 at the National University of Ireland, Maynooth, Ireland. Co-hosted by the Irish Qualitative Data Archive and the Digital Repository of Ireland, the purpose of the workshop was to examine the contribution that qualitative social science researchers and archivists can make to DASISH, and to encourage their cooperation and involvement. Representatives from ten European countries were in attendance.

‘Qualitative social science data’ encompasses all non-numeric data generated in order to answer specific research questions. The workshop focused on issues associated with three of the DASISH work packages – WP4 Data Archiving, WP5 Shared Data Access and Enrichment and WP6 Legal and ethical issues – because these have previously been identified as topics of particular concern to the qualitative research community. Participants in the workshop identified a number of ways in which changes in the field of qualitative research have begun to overcome some of the mistrust around qualitative **data archiving**, including a trend towards increased scale in qualitative research projects, growing diversity in epistemological approaches and increased demand for archiving on the part of research funders and publishers. However, we also noted that the processes involved in preparing qualitative data for archiving are typically more demanding and time-consuming than for quantitative data. This may require attention and early planning to support the inclusion of qualitative data, on the part of data services.

Workshop participants argued that because the culture and norms surrounding the re-use of qualitative data are less robust than those surrounding quantitative data, improving **data access** requires that we expand its scope to encouragement and support for data *re-use*. A number of innovative and exciting projects oriented towards improving access to qualitative data through *enhanced delivery* were described. However, the time-consuming nature of preparing qualitative data for enhanced delivery was noted, and some participants articulated reservations about the appropriateness of such highly structured systems for qualitative data. The workshop identified a number of emerging **legal and ethical challenges** to making qualitative data available for re-use within existing data protection and rights management frameworks. These included limitations to the usefulness of anonymization as a strategy for preserving respondent confidentiality, doubts about the capacity of data services to defend rights management frameworks in some instances, and the negative consequences of insecure funding for maintaining trust in data services.

In summary, the workshop highlighted the extent to which many of the challenges facing qualitative data infrastructures are common across the SSH domains. However, qualitative data may require different – or additional – treatment both at the deposit and end-user stages, and it would be valuable to include consideration of these requirements at an early stage in order to integrate qualitative data to the developing community-based infrastructures.

## 2. Background and Context

### 2.1 Background

The Qualitative Data Workshop (Task 8.5) took place on 21<sup>st</sup> November 2012 at the National University of Ireland, Maynooth, Ireland. The workshop was organized by the Irish Qualitative Data Archive ([www.iqda.ie](http://www.iqda.ie)) and co-hosted by the Digital Repository of Ireland ([www.dri.ie](http://www.dri.ie)), on behalf of the EQUALAN network of European researchers and archivists committed to preserving and organising qualitative social science data resources for sharing and re-use. The purpose of the workshop was to examine the contribution that qualitative social science researchers and archivists can make to the DASISH project, and to encourage their cooperation and involvement. Representatives from ten European countries were in attendance.

DASISH aims to provide solutions to common issues experienced by the five ESFRI projects in social sciences and humanities (CESSDA, CLARIN, DARIAH, ESS and SHARE), and thereby to support and consolidate the process of building their research infrastructures, and to facilitate interdisciplinary activities. While the *quantitative* social science research community is increasingly well-served within the ESFRI framework, progress towards preserving and providing access to *qualitative* social science data is considerably less well advanced. This is despite the fact that qualitative approaches have a long and distinctive tradition within the social sciences, have a prominent role in interdisciplinary research across the humanities and social sciences, and are of increasing relevance for mixed-method and policy oriented research. DASISH thus provides an important opportunity to build on the work begun within the EQUALAN network to explore some of the shared and distinctive challenges associated with developing community-based solutions for qualitative data.

### 2.2 Context

By qualitative social science data, we mean all non-numeric data generated in order to answer specific research questions. Such data may typically include interview transcripts, images, audio and video files, and field notes. Qualitative data are particularly valuable for providing depth and context for our understanding of social life, for explaining innovation and social change and for the development of theory. Within the social sciences there is a growing interest in mixed-method study designs that link qualitative to quantitative data in order to enhance explanatory power. Qualitative data are also central to bridging research across social science and humanities disciplines, for example in fields such as comparative-historical research and cultural studies. The recent development of *qualitative longitudinal* approaches has considerably enhanced the scale, impact and potential for mixed-method and cross-disciplinary collaboration across the SSH research communities. In this context, meeting the distinctive challenges associated with preserving and providing access to qualitative data becomes more urgent.

Work towards a European approach to addressing those challenges began in 2009, with the establishment of EQUALAN at a workshop co-ordinated and part funded by ESRC Timescapes at the University of Leeds, United Kingdom, with additional funding from

CESSDA. The workshop was hosted by the Archive for Life Course Research (ALLF) at the University of Bremen, Germany, on 24<sup>th</sup> April. The context for the Bremen workshop was the CESSDA Preparatory Phase Project, which provided the foundation for a major upgrade of the CESSDA Research Infrastructure in preparation for the establishment of the CESSDA ERIC. The Bremen workshop was organized in recognition of the limited work that had been carried out to date on issues particular to qualitative and qualitative longitudinal data. It was designed to work in parallel with CESSDA PPP by determining existing qualitative resources and then exploring ways of building a European network of qualitative researchers and archivists. Representatives from fourteen European countries who are engaged in archiving a wide range of qualitative and qualitative longitudinal resources attended the Bremen workshop. Country reports mapping the current state of infrastructure for qualitative data archiving across Europe were produced from the workshop. These reports formed the basis of a special issue of the journal IASSIST Quarterly (IQ) <http://www.iassistdata.org/iq/issue/34/3>. A report on the workshop was submitted to the European Commission as Deliverable 10.5b for CESSDA PPP.

Following the Bremen workshop, which established EQUALAN and initiated a continuing process of network development and identification of qualitative resources in Europe, a second workshop was organized in Brussels on 8<sup>th</sup> October 2010. The Brussels workshop was co-ordinated by the Irish Qualitative Data Archive and co-funded by the National Institute for Regional and Spatial Analysis (NIRSA) and Research Office at NUI Maynooth, Ireland, and by ESRC Timescapes at the University of Leeds, United Kingdom. Participants from nine countries were in attendance at this workshop, which initiated an EQUALAN strategy oriented towards furthering the objective of integrating qualitative data resources within the European research area.

### *2.3 Organization of the workshop*

The report from the 'Bremen workshop' identified a number of common challenges faced by European data centres seeking to make qualitative social science data available to the research community. They included:

- Facilitating and coordinating access given the distributed nature of qualitative data resources
- Developing policies that promote trust and a culture of sharing and re-use amongst researchers
- Addressing legal and ethical concerns

Many of these challenges overlap with the issues currently being addressed by the DASISH project. In order to maximize our contribution to DASISH, the Qualitative workshop focused on issues relating to the following work packages:

- WP4 Data Archiving
- WP5 Shared Data Access and Enrichment
- WP6 Legal and Ethical Issues

### 3. Data Archiving

DASISH WP4 has the following objectives:

- a) Discuss the state of long term preservation in the SSH domain and work out suggestions to overcome bottlenecks and to establish trust in procedures
- b) Develop models for common deposit services which could be provided to all SSH researchers
- c) Work out policy rules, business and access models that can be taken up by a common data service layer providing long term preservation services

The qualitative data workshop identified a number of distinctive and shared challenges associated with depositing and preserving qualitative social science data.

#### 3.1 Epistemological concerns

In many cases, qualitative social science data are collected and analysed by researchers who adopt an *interpretive* epistemological stance in contrast to the *positivist* approach often associated with quantitative data. This has been associated with mistrust on the part of some qualitative researchers in data archiving. However, participants in the workshop identified a number of ways in which changes in the field of qualitative research have opened up new approaches to sharing and re-using data:

- a) There has been a trend towards increased scale in qualitative research projects, which often involves researchers working in teams, and/or collecting data over time from the same respondents. This increase in scale creates a momentum towards data management for sharing and preservation. For example, ESRC Timescapes is a qualitative longitudinal project at the University of Leeds, United Kingdom, that aimed to scale up qualitative longitudinal research through seven empirical research projects that explored changing family and personal relationships over time. In collaboration with the UK Data Archive, Timescapes has developed a stakeholder model of data archiving that serves the dual purpose of creating well-organized data for longitudinal analysis, which is also 'archive ready' for sharing and re-use.
- b) There is increasing diversity in epistemological approaches within qualitative social science, together with a growing awareness of local research cultures and disciplinary differences in approach across the SSH domain. Oral historians, for example, may have fewer – or different - concerns surrounding the archiving of in-depth interview data than qualitative sociologists do.
- c) Research funders and publishers increasingly require researchers to develop and implement data management plans that include strategies for archiving.

### *3.2 Overcoming bottlenecks in data management for archiving*

While participants in the workshop emphasized that the challenges associated with making data available for re-use are not unique to qualitative social science data, they noted that the processes involved may be more demanding. Typically, data services provide access to data through a combination of depositor and end-user licenses and anonymization to preserve respondent confidentiality. However, anonymization poses a particular challenge for qualitative data, especially with regard to 'new' data forms, such as images, video and audiofiles, and social media.

Participants noted that qualitative researchers may need early intervention and planning on the part of service providers to facilitate depositing data. Depending on the state of development of their data service, for some participants the challenge is to encourage and support researchers to anonymize their data in preparation for archiving, whereas others are seeking to move towards 'mixed economies' of moderated access to data which in some cases (or for some users) may be un-anonymized.

In summary, despite some traditional resistance to data archiving associated with the different epistemological approaches in qualitative social science, increasingly the challenges associated with archiving qualitative data are converging with those that pertain across all SSH domains. Emerging trends in qualitative research, together with the drive towards open data and the increasing prevalence of 'new' data forms other than the traditional in-depth interview, have created a momentum for including qualitative data in a common roadmap for SSH infrastructures. Nevertheless, managing the deposit and dissemination of qualitative data is complex and may require more attention and support on the part of data services.

## **4. Shared Data Access and Enrichment**

DASISH WP5 has the following objectives:

- a) To establish a trust environment and a joint data metadata domain based on well-established AAI principles
- b) To establish a widely used tool/services portal and setup a framework for testing, evaluating and commenting on tools/services to spread information about useful components
- c) To define typical processing chains, test them based on existing technology and spread the information about how to create and use them
- d) To create an interoperability framework based on typical use cases
- e) To create a data enrichment framework suitable for SSH researchers

Participants in the qualitative data workshop focused on two dimension of data access: (1) expanding the understanding of access to include not just technologies and infrastructures for data delivery, but also the services and support for re-use; (2) developing enhanced mechanisms for exploring, citing and linking qualitative data.

#### 4.1 Expanding the understanding of access

Because the culture and norms surrounding the re-use of qualitative social science data are not as robust as those surrounding quantitative data, improving *access* requires more support for *re-use*. For example, the UK Data Archive promotes re-use by providing training courses on secondary analysis, disseminating profiles of those who have re-used data for research and teaching, providing 'off-the-shelf' teaching datasets with 'how to' help, and nurturing collaborations with researchers – by providing reports to depositors on usage of their data, and pooling resources to digitize legacy data collections.

#### 4.2 Developing enhanced mechanisms for exploring, citing and linking

Access to, and reuse of, qualitative social science data has been limited by the absence of web-based exploration systems such as that provided by NESSTAR for quantitative data. A number of projects to develop enhanced qualitative data delivery were described at the workshop, including:

- The Digital Futures Project at the UK Data Archive aims to develop a 21<sup>st</sup> century system for delivering qualitative data that will enable federated catalogues across providers and borders, lower the barriers to non-academic users, and allow users to explore data in a journey that involves finding relevant extracts, examining contexts, and linking textual data to still and moving images and other related research outputs.
- The 'Voices of the 20<sup>th</sup> Century Archive and Research Group' at the Institute for Sociology, Centre for Social Sciences, Hungarian Academy of Sciences, aims to provide enhanced access, through keyword and full-text searching of the documentary, audio and audiovisual heritage of Hungarian sociology.

In order to deliver qualitative data to users, in ways that facilitate exploration within and across data collections analogous to those deployed in quantitative collections, the data must be highly structured and consistently marked up. However, marking up qualitative data and visualizing relationships across qualitative data objects currently entails a lot of time-consuming manual work.

Workshop participants also noted that some key principles of qualitative research – including the importance of context and inductive analysis – might be threatened by systems of enhanced delivery. Some participants argued that re-users should be encouraged to read whole interviews, rather than relying on retrieved segments, and that in highly structured data there was a risk of conflating keywords, or metadata, with concepts.

In summary, technological innovation and progress towards developing metadata standards for qualitative data mean that the infrastructural barriers to enhanced delivery – and to linking qualitative social science data to other kinds of SSH data – are breaking down. However, concerns remain about the appropriateness of such highly structured data delivery systems for qualitative data. Furthermore, especially with



respect to qualitative data, support for re-use must be considered as important for access as technical delivery.

## 5. Legal and Ethical Issues

DASISH WP6 has the following objectives:

- a) To identify legal and ethical issues, constraints and requirements for all data types occurring in the SSH domain as a result of the intended integration and linking
- b) To cope with the legal and ethical challenges imposed by new data types emerging in the social sciences,
- c) To look for professional long-run preservation strategies based on e-Infrastructures for data in the social sciences and humanities

Qualitative social science researchers often express greater fears about the ethical risks associated with data archiving than do quantitative researchers, or researchers in the humanities. This is partly attributable to the perception that, because qualitative data generation is a situated activity, involving close personal interaction between the researcher and the respondent, the researcher bears an exceptional degree of responsibility to protect the respondents with whom he or she co-constructs representations of the social world.

Data services that provide access to qualitative data for sharing and re-use generally seek to protect respondents and depositors by securing consent for archiving, anonymization to remove personal or sensitive information, and a rights management framework incorporating depositor and end-user agreements and licensing, access control and user restrictions.

Participants in the workshop identified a number of emerging challenges for this approach, none of which, however, is exclusive to qualitative data.

- 1) Anonymization may not always be a panacea for concerns about respondent confidentiality – for example when researchers wish to do follow-up research at a later date, to link data sources, or in the case of data formats that do not admit of ‘easy’ anonymisation. In this context, it may be necessary to implement solutions that involve privileged levels of access to special or expert users. Such an approach would have the benefit of lowering barriers to access for expert users, while also securing data protection.
- 2) In some cases the rights management framework implemented by data services are based on trust, and the capacity of those services to defend their licensing agreements remains legally untested. In the Irish case, for example, social science data services are university based, but Irish universities are currently

more oriented towards defending patents and intellectual property than data sharing.

- 3) The vindication of licensing agreements over time is threatened in many cases by the absence of secure funding. Many archives and data services depend on short-term project funding that compromise trust in their rights management procedures.

In summary, the legal and ethical landscape for qualitative data sharing is evolving in the context of new research practices, new kinds of data, the potential for integration across the social sciences and humanities, and the continuing threat posed by insecure funding.

## 6. Conclusion

The purpose of the Qualitative Data workshop was to examine the contribution that qualitative social science researchers and archivists can make to the DASISH project, and to encourage their cooperation and involvement. The workshop focused on issues associated with Work Package 4 (Data Archiving), Work Package 5 (Shared Data Access and Enrichment) and Work Package 6 (Legal and Ethical Issues), because these have previously been identified as topics of particular concern to the qualitative research community.

There were a number of shared themes arising from the workshop:

- Many of the challenges facing qualitative data infrastructures are common across the SSH domains. These include:
- Ensuring data protection in the context of new data forms and research practices
- Maintaining trust given the distributed nature of data resources and services and insecure funding

However, qualitative data may require different – or additional – treatment both at the deposit and end-user stage, and it would be valuable to include consideration of these requirements at an early stage in order to integrate qualitative data to the developing community-based infrastructures.

- Preparing qualitative data for deposit is time-consuming and costly, especially given the growing requirement for enhanced systems of delivery. Overcoming bottlenecks and establishing trust in common deposit models will require continuing active support for qualitative researchers and the communities they serve. Early intervention and planning will be essential to defray the costs associated with including qualitative data.
- Because the culture of data sharing and re-use is less well developed within the qualitative research community, the understanding of access must be expanded to include support for re-use

Finally, the inclusion of qualitative data represents an important opportunity in the context of DASISH's focus on the development of interdisciplinary 'cross-walks' between the humanities and social sciences. Qualitative social science research, with its focus on interpretive and situated data collection and analysis, reaches out both to the humanities and quantitative social sciences, but with its own distinctive conventions and traditions.

## 7. Appendices

Full details of the workshop, including copies of the presentations, are available here: <http://www.iqda.ie/content/dasish-qualitative-data-workshop-20th-21st-november-2012>

### 7.1 Workshop programme

- 9:00**      **Welcome & Registration**  
Sandra Collins, Digital Repository of Ireland.
- 9:30**      **Overview of SSH Infrastructures and introduction to the work of DASISH**
- Hans Jørgen Marker, Swedish National Data Archive –“Infrastructures for Social Science and Humanities”.
  - Hanne Fersøe, University of Copenhagen. “Qualitative research data for the Humanities and Social Sciences: CLARIN – a European ESFRI infrastructure and a stakeholder in DASISH”.
- 10:30**      **Coffee Break**
- 11:00**      **Data Archiving: Preservation, Policies & Challenges**
- Trond Kvamme, Norwegian Social Science Data Services. “Overview of DASISH WP4: Data Archiving in the SSH”.
  - Alen Vodopijevic, Institut Ruder Boskovic. “Potentials and infrastructures for establishment of social science data archives in WB countries and overview of SERSCIDA project goals”.
  - Piotr Filipowski, Institute of Philosophy & Sociology, Polish Academy of Sciences. “Oral History in Poland: A good base for qualitative Data Infrastructure?”
- 12:30**      **Shared Access to Qualitative Data: The challenges of MetaData**
- Louise Corti, UK Data Archive. “A ‘Nesstar’ for qualitative data: getting the building blocks and user needs right.”
  - Libby Bishop, UK Data Archive. “Open? Licensed? Secure? Challenges for enabling access to data”.
  - Judit Gárdos, Voices of the 20th century Archive and Research Group. “Qualitative data archiving: a case study from Hungary”.
- 1:30**      **Lunch**
- 2:30**      **Legal & Ethical Issues for Qualitative Data Archivists, Depositors and Users**
- Anne Sofie Fink Kjeldgaard, Danish Data Archive. “Addressing legal and ethical issues in dialogue with the designated communities”.
  - Aileen O’Carroll, IQDA and DRI. “To Share and Protect: Policy Challenges in Archiving Qualitative Data”.

**3:30 Discussion**

**4:00 Conclusion & Farewell**

*7.2 List of participants*

Name	Surname	Affiliation	Email
Judith	Gárdos	Voices of the 20th Century Archive & Research Group	gardos.judit.jgr@gmail.com
Piotr	Filipkowski	Institute of Philosophy & Sociology, Polish Academy of Sciences	pfilipkowski@ifispan.waw.pl
Louise	Corti	UK Data Archive	corti@essex.ac.uk
Jarkko	Päivärinta	Finnish Social Science Data Archive	jarkko.paivarinta@uta.fi
Hanne	Fersoe	University of Copenhagen	hannef@hum.ku.dk
Anne Sofie	Fink Kjeldgaard	Danish Data Archive	asf@sa.dk
Hans Jørgen	Marker	Swedish National Data Archive	h.j.marker@snd.gu.se
Ricca	Edmondson	National University of Ireland, Galway	Ricca.Edmondson@nuigalway.ie
Trond	Kvamme	Norwegian Social Science Data Services	trond.kvamme@nsd.uib.no
Alexandra	Stam	FORS, Swiss Foundation for Research in Social Sciences	alexandra.stam@fors.unil.ch
Libby	Bishop	UK Data Archive	ebishop@essex.ac.uk
Niamh	Curtin	University College Cork	niamh.a.curtin@gmail.com
Ann Marie	McCague	TV3	ammccague@tv3.ie
Alen	Vodopivec	Institut Ruder Boskovic	alen@irb.hr
Rebecca	Grant	Digital Repository of Ireland	r.grant@ria.ie
Natalie	Harrower	Digital Repository of Ireland	n.harrower@ria.ie
Siobhan	Fitzgerald	Trinity College Dublin	fitzges6@tcd.ie
Sandra	Collins	Digital Repository of Ireland	s.collins@ria.ie
Jane	Gray	Irish Qualitative Data Archive	jane.gray@nuim.ie
Aileen	O'Carroll	Irish Qualitative Data Archive	aileen.ocarroll@nuim.ie

Louise	Bolger	Irish Qualitative Data Archive	louise.bolger.2009@nuim.ie
Claire	Byrne	Irish Qualitative Data Archive	claire.byrne@nuim.ie
Cai	HuiHua	NUI Maynooth	huihuacai6@gmail.com
Sarah	O'Keefe	NUI Maynooth	sarakeeffe@gmail.com
Hilary	Darcy	NUI Maynooth	hilary.darcy@nuim.ie

### *7.3 Photograph of participants*

