



Title **Qualitative and Qualitative Longitudinal Resources in Europe: Mapping the Field and Exploring Strategies for Development (D10.5b)**

Work Package WP10

Authors Libby Bishop and Bren Neale

Dissemination Level PU (Public)

Summary/abstract

This report forms part of the D10.5 “Audit report on CESSDA data collection strategies, harmonising strategies and virtual catalogue” deliverable.

The Bremen Workshop

Qualitative and Qualitative Longitudinal Resources in Europe: Mapping the Field and Exploring Strategies for Development

24 April 2009 – Bremen

Libby Bishop* and Bren Neale**

*The Timescapes Programme and Archive and UK Data Archive

**The Timescapes Programme and Archive

Workshop Report for CESSDA PPP

30 May 2010

Introduction

In April 2009 Timescapes, in collaboration with the Council of European Social Science Data Archives Preparatory Phase Project (CESSDA PPP) and the University of Bremen, hosted a residential workshop in Bremen to explore the nature of Qualitative (Q) and Qualitative Longitudinal (QL) research and resources across Europe. The workshop was funded by Timescapes with support from CESSDA PPP and hosted by the Archive for Life Course Research (ALLF) at the University of Bremen. It was attended by archivists and researchers from 14 countries, including 'transitional states' such as Belarus and Lithuania. The broad aim of the workshop was to map existing infrastructures for Q and QL data archiving among the participating countries, including the extent of archiving and the ethos of data sharing and re-use in different national contexts. We also explored strategies to develop infrastructure and to support Q and QL research and resources, including collaborative research across Europe and beyond.

Background and Context

The Bremen workshop can be seen as part of a much broader effort to co-ordinate research resources across Europe. The impetus for the workshop was provided through CESSDA PPP, which is funded by the European Commission (7th Framework Programme) under the European Strategy Forum on Research Infrastructures (ESFRI). The remit of ESFRI, codified in its 2008 Roadmap, is to support strategic policy making for research infrastructure in Europe and to facilitate the better use and development of these infrastructures at EU and international levels. CESSDA was funded to develop research infrastructure for social science data archiving across Europe, having been identified by ESFRI as a research network of excellence.

ESFRI defines infrastructure as “the facilities, resources or services of a unique nature that have been identified by pan European research communities to conduct top-level activities in all fields”. This definition covers major equipment or sets of instruments as well as knowledge-containing resources (e.g., collections, archives, and databases). These resources may be in one location, distributed across several sites, or networked (multiple organisations across sites) and they may take different formats, including electronic, paper-based or both.

CESSDA PPP is a distributed research infrastructure that provides access to European research data and supports their use. The CESSDA PPP remit is to develop infrastructure for all social scientific data and it currently supports the development of national level social science data archives in 20 European countries, through which access is provided to over 25,000 collections. The current focus in CESSDA PPP is on establishing a European infrastructure to address foundational issues such as: common authentication and access, cross-European resource discovery, secure data facilities, and adoption of interoperable metadata standards. To date, the majority of data available from the CESSDA portal are quantitative resources, including official government census data, social surveys, and quantitative longitudinal and cohort studies.

All the current infrastructure initiatives are vital regardless of data format; however, to date there has been limited work on any issues particular to Q or QL data. The Bremen workshop was designed to complement work undertaken in work package 10 of the CESSDA PPP by determining existing Q and QL resources and then exploring ways of building a European network of Q and QL researchers and archivists committed to preserving and organising qualitative data resources for sharing and re-use.

Human data, of course, are challenging because they are so vast: endlessly varied, fragmented, complex, dynamic, multilingual, and historically, politically and geographically situated. Preserving and disseminating the products of human culture and society is difficult and expensive, particularly for qualitative data. Even so, new digital resources, including software and e-networks are influencing the production of human records and how these are understood and communicated. The ESFRI Roadmap indicates the enormous potential of data—of all kinds—for understanding the profound social, cultural, political and economic life of Europe, including social continuity and change. The ESFRI Roadmap reminds us that the first step in developing such infrastructure is networking and co-operation and it was in this spirit that the Bremen workshop took shape.

The Bremen Workshop

The participants from the 14 countries were asked to produce a country report, setting out the nature of existing infrastructure for qualitative archiving, policies and ethos for data sharing, an overview of key resources and collections of Q and QL, and priorities for future development. The reports were tabled at the workshop and, for the purposes of presentation, were grouped into three broad categories (from most to least developed in terms of infrastructure). One representative from each group presented a brief overview of the group, pointing out areas of commonality across the group, and important distinctive circumstances for some countries. The afternoon break-out sessions mixed members from all three groups and was tightly focused on development planning and structured around these questions:

- What are the barriers and enablers of data sharing?
- How effective are existing models for sharing or archiving data?
- What are the pros and cons of having a range of archives and collections, centralised and devolved, generic and specialised?
- Is there a case for developing separate infrastructure for Q/QL data resources or for merging these resources with existing quantitative and/or longitudinal resources?
- What are the best ways of getting an archive started and what issues arise in developing and sustaining the resource?
- Would a European wide network for qualitative data archiving be beneficial and if so, how would you like to participate?

Group One – Finland, Ireland and the UK

This group has established national archives for social science data that include qualitative collections (Qualidata in the UK funded from 1994, The Irish Qualitative Data Archive (IQDA), from 2008, and the Finnish Social Science Data Archive in 2003). In each case the archives include primarily interview data (with focus groups and other textual sources). All three also have (or are planning to add) multimedia formats: text, sound images, moving images, and analytical files. Although funded as national resources, the three countries are characterised by patterns of decentralisation; in the UK, Qualidata, for example, is the most important but not the only hub in a vast network of independent and proliferating collections held by a wide range of organisations that are rarely co-ordinated. This ‘mixed’ infrastructure with specialist and generic resources existing alongside each other was seen as inevitable; though it may pose coordination challenges, there is also potential for innovation collaborations. QL research and resources are well represented across these three countries and in the UK a specialist Timescapes archive for QL data has recently been developed (funded from 2007).

At the national level, the three countries in this group have policies promoting data sharing. There was growing awareness of qualitative data as important research outputs, part of dissemination and, therefore, a greater focus on the production of high quality data outputs. Key national funding bodies in these countries all require data management, recommend archiving or made data sharing a condition of funding.

Despite these developments, however, support for data sharing in these countries was uneven. For example, in Finland there is no established culture of promoting qualitative data re-use and an assumption remains that primary researchers are the only ones to understand and use the data correctly. In the UK such views are becoming less acceptable but there are issues around balancing access with the need to protect confidential data. Moreover, archiving is seen predominantly as the end product of research, acceptable only when the researchers have ‘finished with’ the data. For QL research, where projects may run for many years and the data are never seen as complete, this may run counter to the drive to archive and re-use the resource.

Priorities for development included technical development of the archives to include multimedia data and to develop the specialist preservation needed for QL data. Despite the advances in these countries a need was identified in each case to build the culture of data sharing and re-use, and to strengthen policies and develop initiatives to support this aim, for example, through funding for secondary analysis of Q and QL data. A need was identified for greater co-ordination of data resources across the mixed infrastructure, so that specialist collections could more easily be identified, searched and accessed. Finally, funding was relatively fragile and there was a need to secure longer term funding to facilitate this work.

Group Two – Austria, Czech Republic, Denmark, Germany, Norway and Slovenia

Not surprisingly, this group was highly diverse with some members resembling Group One in many dimensions, but others being more like Group Three. Generally speaking, there is infrastructure in place for QN data archiving; all but the Czech Republic have existing National Archives. In most cases, some fledgling effort is underway for these predominantly quantitative-orientated institutions to begin handling QL data. Austria, for example, began archiving qualitative data at WISDOM in 2007 and the Danish Data Archive began handling qualitative data in 2009. NSD in Norway is planning to incorporate qualitative data and the Social Science Data Archive in Slovenia is in a similar situation. But these national infrastructures capture only a small amount of activity as there are numerous Q and QL resources widely distributed in smaller institutions, departments, and even individual projects. Many of these are attempting to archive Q and QL collections, and some are seeking to form alliances or collaborations with QN institutions, where they exist.

As with infrastructure, the situation regarding data sharing is also ambivalent. In terms of actual archive-mediated data sharing, levels of activity are rather low. But there is growing visibility of the issue and other indications of changing attitudes. Formal feasibility studies (for archiving Q data) were done in Austria, Denmark and Germany, revealing surprisingly positive attitudes toward both sharing data and using data collected by others. However, hurdles exist in translating these attitudes into more positive actions. Where archives do exist – in Denmark and Austria for example – few datasets have been deposited and the rate of new deposits is low. Major challenges remain in numerous areas: concerns about ethics and confidentiality; researchers’ belief in exclusive ownership of data; technological and financial resources constraints; and complex infrastructure models.

Development priorities reflected the national situations, but all pointed to the need for networking with other institutions and countries. Locating stable funding sources was also a

high priority, as was engaging in activities to bring about cultural acceptance of data sharing—finding exemplar cases and teaching methods for re-using data. There are, perhaps, at least some reasons to be optimistic – in Germany, the feasibility study, as well as publications and an annual workshop on secondary analysis, has encouraged more active debate about data archiving and sharing.

Group Three – Belarus, Hungary, Lithuania, Poland and Switzerland

Members of Group Three reported only minimal infrastructure for curating Q or QL data, though there was obvious enthusiasm for developing such infrastructure amongst a subset of the academic community. Other than the Lithuanian Humanities and Social Science Data Archive (LiDA), there are no national institutions for data archiving in this group. Where laws exist, these are general ones on archiving and data protection, but have not been made specific to qualitative data (Switzerland).

The culture of sharing is weak to non-existent, at least for Q and QL data. In Poland, there is “no academic tradition” of sharing qualitative data, perhaps partly because of a very strong prevailing positivist tradition in social research. In Hungary, there are some existing archives for particular surveys, but data sharing is not common, and the culture of re-using data is not widespread. In the case of Belarus, there is no national infrastructure for archiving. What data are retained is held by individual organisations. Secondary analysis is rare and occurs only after personal negotiations among primary and secondary researchers. In many cases, research data are not retained at all, even by primary researchers. The recent political climate has, in part, contributed to this situation. In contrast, Lithuania does have some national policies promoting sharing, and in addition to LiDA, there is now access to online research data via Electronic Information for Libraries (eIFL.net), but this focuses more on research outputs and not raw data. Despite progress, recent reports still suggest that Lithuania is lagging behind Slovenia and the Czech Republic.

As might be expected, the list of development priorities is long and wide-ranging. Basic work in establishing infrastructures is needed, with the concomitant requirements of appropriate technologies and financial resources. Practical examples of archiving policies and procedures would be highly beneficial, and even with the adaptations required for specific national conditions, could avoid a great deal of work being reinvented. Administrative advice is also needed such as how to staff, at what levels and specific skills needed. Specifics include collections strategies (deciding what to archive), and rights management (consent, anonymisation, access controls, IPR, etc.). In one area, however, there was strong unanimity in Group Three, and across all the groups for that matter: the desire and need for stronger international knowledge exchange, joint projects, and resource sharing.

Workshop Outcomes

The Bremen workshop produced an impressive collection of outcomes in three areas: short-term activities, agreed goals and objectives, and a strategic plan for future action. Some aspects of the strategy outlined below have emerged in subsequent communications among the workshop participants.

Short term activities

1. Create Bremen workshop webpage – available at:
<http://www.timescapes.leeds.ac.uk/about/bremen-workshop/>
2. Publish updated list of international data providers on ESDS website—this is in progress here: <http://www.esds.ac.uk/qualidata/access/internationaldata.asp>
3. Publish a list of all QL collections and resources provided in the country reports—this is in progress in a restricted area of the Timescapes website.
4. The complete set of revised country reports with an editorial introduction has been accepted for publication by *LASSIST Quarterly* (IQ) in late 2010.
5. Create communication lists—the existing email list is serving for communication amongst the group. We have plans for a new Methodspace list – Quali-Net – for QL researchers and archivists.
6. Investigate specific funding sources—several funding sources are actively being pursued (ESFRI, EU FP7, NCRM).
7. Organise another workshop oriented towards developing concrete steps towards meeting the challenges identified at Bremen, and towards seeking funding for this, including in the forthcoming round of calls under the FP7 Infrastructures programme.--IQDA (Ireland) offered some funding for this. Planning is continuing with a sub-group of participants who will be at IASSIST in June 2010.
8. Publish case studies from the most developed archives (IQDA, Finland, UK, and Germany)—we are attempting to seek funding to produce these reports.

Agreed goals and objectives

There was broad agreement on these as goals and objectives. Clearly, action in many of these areas is not specific to this group at Bremen, and it was further recognised that many of these objectives need national or international coordinated action. Nonetheless, the group felt it important to articulate explicitly how Q and QL archiving should become an integral part of these wider strategies.

1. Active networking, in some cases with better-resourced QN partners and institutions
2. Promote metadata standards, including developing ones for Q and QL data, especially for newer formats such as multimedia
3. Develop metrics for reuse and the technological systems to collect data for reuse
4. Lobby funders for specific policy changes:
 - a. implement mandated data deposit
 - b. ensure costs for preparing data for archiving are fully funded
 - c. ensure SA projects received equal treatment when being considered for funding
5. Change research reward systems to treat data creation more favourably
 - a. reference/citation credit when using archived data
 - b. data creators acknowledged as joint authors
 - c. ISBN assigned for data set creation

- d. Data deposit counted as part of formal research excellence review procedures (Research Excellence Framework and European equivalents).)
- 6. Promote activities to accelerate a cultural shift toward data sharing:
 - a. Work through professional associations
 - b. Teaching – especially post graduates and early career researchers
 - c. Direct engagement with ethical challenges to reusing data.

Strategies for Future Development

While all the above goals are vitally important, it was recognised that in most instances, these goals are not specific to Q or QL data. As noted above, CESSDA PPP (and later, CESSDA ERIC) and ESFRI are addressing areas of harmonised legal environments, multiple language thesaurus, secure access to ethically sensitive microdata, and much more. What this makes clear, then, is that the Bremen group is well positioned to define and address issues that are particular to Q and QL data.

When devising a strategic plan for archiving Q and QL data in Europe, the central question is: in what ways are Q and QL data the same, or broadly similar, to quantitative data, and therefore able to integrate into existing data infrastructures? And in what ways are they distinctive, and thus potentially in need of distinctive or customised treatment? Answers are emerging from several directions. The Timescapes Programme and Archive has built a QL specific archive, and in doing so, is uncovering special needs of QL data and also building a stakeholder model of researcher and archivist collaboration. The experience of the UK Data Archive is also relevant because it was a well-established archive for quantitative data and, in 2000, incorporated Qualidata into its existing infrastructure, proving that Q data can be processed in standardised ways. These experiences, along with related experiences in Ireland, Finland and Germany, point to similar lessons learned.

Broadly speaking, Q and QL data are distinctive from Quantitative data in three areas: metadata requirements, ethical considerations, and culture of reuse. In terms of the Open Archival Information System (OAIS) model, the middle processes of Data Management, Archival Storage, Preservation Planning and Administration are broadly similar regardless of data format. Of course, provisions need to be made for different formats, large video files being one challenge. However, the processes for handling all data are broadly similar. It is in the early and later phases of the data life cycle where Q and QL differences matter most. Two of these, metadata and ethics, lie in the Ingest (or pre-Ingest) phase while culture of reuse falls within Access. *By no means are these the only topics that could be chosen, and future strategic planning sessions may revise this list. However, the idea of defining distinctive aspects of QL and Q data, and the implications that follow for developing archiving infrastructures that both support Q and QL data and also integrate with ongoing initiatives, seems like a sensible way forward.*

The first challenges Q and QL data pose is for adequate metadata collection, in part because of the complex file formats involved. Data need more extensive metadata and contextual material to render them “independently understandable”, a requirement of the OAIS standard. Unlike most structured quantitative data with relatively standardised formats, qualitative research data and documentation are highly diverse. It is also generally accepted that qualitative data need

extensive contextual information to enable effective resource discovery and reuse. Much of this may fall into familiar metadata categories, but ideally context should also include information about the project background and the social and institutional conditions in the wider environment that might have shaped project design.

Ethics is the second area that distinguishes Q and QL data from Quantitative data. This is a particularly complex issue. On the one hand, the curation of much Q data presents few or no ethics challenges at all. Consent for sharing is readily obtained, typically with some form of anonymisation. However, ethical concerns remain a major factor in debates amongst researchers about the reuse of Q data and some aspect of this was raised by every participant at Bremen. Typical issues include: can consent be said to be informed when the topics of research for reuse cannot be known in advance? Are there risks to participants if reused data misrepresent their views? Are researchers exposed to unfair criticism when their work is made visible by archiving? These factors have the potential to limit the availability of data for archiving in the short term, even where consent has been obtained from research participants. In a QL context, this has implications for the way archivists work with researchers and suggests the need for involvement in the development of a research project from the outset to facilitate ethical archiving.

Despite rapid change in recent years, it is still the case that the culture of data reuse is weaker and less widely accepted for Q and QL data relative to QN data. This is decidedly the case in the Group Two and Three countries, as the country reports documented, and continues to be the case in Finland, Ireland and the UK, although to a lesser degree. For data archives, the resource implications are that more effort and resources are needed to promote reuse of Q and QL data. These range from preparation of focused outreach materials to the need for training and support, customised to distinct audiences. Nevertheless, successful Q and QL archiving is most important in this respect, because it plays a decisive ‘demonstrator’ role in alleviating researchers’ concerns and normalising the culture of archiving and re-use.

Possible next steps

The Bremen participants have stayed in regular communication since the workshop, primarily focused on revising articles for the special issue of *IASSIST Quarterly*. Informal meetings, usually conferences where a sub-group were attending, have taken place to discuss knowledge exchange and future funding options. The next such meeting will be held at IASSIST in June 2010 at Cornell University, when we will map a strategy for a more formal meeting later in the year at IQDA, which will be focused on developing an application for funding. The structure of any future bid has yet to be specified, but could well include work packages addressing particular Q and QL needs. Illustrative examples in the areas of metadata, ethics (including access), and promoting a culture of reuse might include:

- Q and QL metadata-several Bremen participants are members of the DDI Qualitative Data Working Group that will develop a DDI compliant schema for Q data.
- The Timescapes Project, as part of its mid-term review, has produced an Ethical Framework for QL Research and Archiving. This needs further development by being exchanged with others, and the addition of international case study examples. Additionally, the stakeholder model of archiving QL data, designed to encourage deposit

of longitudinal data during the life time of a project could be evaluated in a broader European context.

- Timescapes is currently developing infrastructure to support QL research and resources in an international context. This takes the form of a digital resource centre that includes a network of QL researchers and archivists, a database of ongoing and extant projects and a portal to link and enable access to QL datasets. Developing this resource is an important part of capacity building in QL research and resources, and its further development would be greatly enhanced through input from Quali-Net members. This may provide a focus around which some of the work of Quali-Net could cohere.
- Much technological development is still needed to create the complex access controls required for highly sensitive and confidential data. Fedora promises a more robust system and work is needed to assess existing projects and define next steps.
- Such work on access controls needs to remain aligned with ongoing work on secure data services (such as at the UK Data Archive) that are intended to enable sharing of potentially disclosive microdata.
- Capacity building in teaching a next generation of scholars about the benefits of data archiving and substantively grounded methodologies for conducting secondary analyses using Q and QL data.

Conclusion

The funding for the Bremen workshop was very welcome and has opened up a new and vital area for research archiving that is currently underdeveloped for the social sciences in Europe. The CESSDA PPP has highlighted the need to both recognise the unique situation of every archive, and also much shared intent over preservation, data management, and dissemination standards and practices. Extending this to encompass the full range of data across the spectrum of the social sciences, with initiatives to create connections across diverse datasets, would be a significant step forward. The creation of the fledgling Quali-Net, with a broad remit to put Q and QL archiving firmly on the map, is the first step towards this long term goal. A meeting of the network is planned for the Autumn of 2010 in Maynooth in which a mission statement for the group will be produced and plans for implementation drawn up. High on the list of priorities is to seek funding and support for the work of the network over the coming months and years. Different models for the growth of Q and QL archiving and data sharing will be needed, depending on the existing infrastructure and ethos within different nation states. But making such data 'count' in the world of archiving and secondary analysis will do much to enrich understandings of the social world.