CAQDAS Comparability. What about CAQDAS Data Exchange?

Louise Corti & Arofan Gregory

Abstract: This article seeks to address the theme of the comparability of *C*omputer *A*ssisted *Q*ualitative *D*ata *A*nalysi*S* (CAQDAS) packages through comparing current software exchangeability and portability. Our perspective is from a data sharing and archiving perspective and the need for open data exchange standards for qualitative data which will enable longer-term sustainability of both data collections and of annotations on these data. Descriptive metadata allow us to describe data robustly and using a common standard enables us to tap the common features of any complex collection. A set of "raw" research outputs (data) have common descriptive elements such as how the research project was funded and how the data were sampled, collected and analysed to form conclusions from that investigation.

Data kept for the longer term must ideally be software and platform independent. In this way, we can help future-proof data resources. Most CAQDAS packages use proprietary databases to manage their data and annotations, and very few enable export of annotated data. In this article we argue for an open descriptive standard that will enable description and interpretation of data for the longer term in data archives and to which proprietary software, such as all CAQDAS packages, can import and export. The use of the term "annotation" or "annotating" is taken to mean any action on the text—classifying, coding, memoing or relating. This meaning of the term is commonly used in the linguistic community, but less so by social scientists.

Table of Contents

1. Introduction

This article seeks to address the theme of the comparability of *C*omputer *A*ssisted *Q*ualitative *D*ata *A*nalysi*S* (CAQDAS) packages through comparing current software exchangeability and portability. [1]

Our perspective is from a data sharing and archiving perspective. The authors are engaged in helping set standards for data exchange and longer term social science data sharing and archiving. Louise CORTI is from the UK Data Archive which has been archiving and delivering data for over 40 years and qualitative data since 2000. Arofan GREGORY is from Metadata Technology which was set up in 2005 by the people that did the technical work on some key data standards —SDMX and DDI and the tools development, including the registry.[1] Both of these standards enable organisations to better manage their data and metadata. Gregory is involved in the work of the Open Data Foundation, a non-profit organisation dedicated to the adoption of global metadata standards and the development of open-source solutions promoting the use of statistical data. [2]

The partnership of the UK Data Archive and Metadata Technology has been instrumental in bringing together the data archiving community to help define open data exchange standards for qualitative data. This builds heavily on similar work for survey and aggregate data, and our belief is that any standard should not deviate too far from those used to describe these data types. This is important because, essentially, all data are structured in some way and have common features that include the following: they arise from a social science investigation, they contain a set of "raw" research outputs (data), they have common descriptive elements such as how the research project was funded and how the data were sampled, collected and analysed to form conclusions from that investigation. [3]

The "open" part means that we should try to provide solutions that do not require proprietary software to interpret details of the data. Data kept for the longer term must ideally be software and platform independent in order to help future-proof data resources. The use of CAQDAS packages in qualitative data management and analysis currently make it hard to keep or save any annotated data in a shareable format. For example, it is generally difficult to export annotations (e.g., data attached to codes). Some very new features of a small number of the CAQDAS packages enable import from each other's software, which is a very promising new step forward. Very few enable robust export to XML. ATLAS.ti and NVivo do, but neither use a common schema (or language) for doing so. ATLAS.ti was the first vendor to pioneer data exchange by exporting annotations

---

1   Statistical Data and Metadata Exchange (SDMX), an initiative to foster standards for the exchange of statistical information. SDMX Technical Standards Version 2.0 provide the technical specifications for the exchange of data and metadata based on a common information model. It is currently used by organisation such as the OECD for providing aggregate national data. The Data Documentation Initiative(DDI) is an effort to create an international standard for describing social science data. Expressed in XML, the DDI metadata specification now supports the entire life cycle of social science datasets. DDI metadata accompanies and enables data conceptualisation, collection, processing, distribution, discovery, analysis, repurposing, and archiving.

in XML (MUHR, 2000). Currently users are unable to export their annotated collection out of their chosen package to retain for archiving purposes or to display in open source tools or publish on the web, say in HTML. [4]

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. CARMICHAEL (2002) reports on how open source CAQDAS type tools could be useful [5]

2. How Exchangeable Is Current Data Analysis Software

Many social researchers undertaking qualitative research are making use of some form of data management software. This can be MS Word or MS Access but since the 1980s a number of dedicated packages, known as CAQDAS packages, have come on the market. The most popular CAQDAS packages are, at present, all proprietary which means that buying into one package essentially means locking up data into one technical solution. This is gradually changing as we see some recent versions of software packages (including NVivo, QDA Miner) emerging in late 2010 which are addressing import features from the market leading packages. [6]

On the whole, researchers tend to be trained in a single software—universities often have a single software license for teaching, and research groups tend to favour a single package. Also, licenses are expensive and there is a steep learning curve required to get acquainted with any one software. Buying into a single solution means that data are loaded into the software and classification and annotation of data is done "in situ". This value-added work cannot be exported in a way that preserves the relationship between data and annotations, unless a dedicated "export" feature exists for the software. As a simple example, codes are normally attached to a segment of text with a start and end point. Coding is a form of annotating. These reference points (e.g., character 1 to character 200) are usually stored in the software's database but the references of the segment to the code cannot be brought out of the package, or imported into a different software. The same goes for memos attached to a code or a segment of text, or a classification or variable e.g. gender, attached to a document. [7]

In the absence of any import or export features for *almost* all of the packages, with the exception of ATLAS.ti, this may cause a serious problem if either the software house closes, or the researcher or team decide that they wish to or need to utilise a different software, or if we want to archive data. [8]

As yet, there are no free open-source products which can compete with the functionality of the leading software packages, such as ATLAS.ti, NVivo or MAXQDA for example. Commercial packages have nice user-friendly interfaces and are very easy to install. Most are Windows based, whereas free open source products may be entirely web-based and very limited in terms of functions they provide. [9]

However, there is an increasing community demand for open source formats and standards. Almost all types of software—statistical software, databases and spreadsheets, text editors, HTML editing packages, programming packages and audio players e.g. iTunes have import and export features enabled. For example, for statistical software, while a range of proprietary packages are used, there has been a common portable exchange format since the 1970s, spss.por, which enables data to be moved around and shared for the longer-term. While there is usually some loss of detail on conversion during data transfer, the core mark-up remains stable (e.g. styles, bold, field names). Many data types have open descriptive standards:

• statistical software—SDMX and DDI standards;
• documents, spreadsheets and databases—some open formats, Sun OpenOffice (limited but baseline) and MS now store data in XML (e.g., Word):

> "OpenOffice.org 3 is easy to learn, and if you're already using another office software package, you'll take to OpenOffice.org 3 straight away. Our world-wide native-language community means that OpenOffice.org 3 is probably available and supported in your own language. And if you already have files from another office package—OpenOffice.org 3 will probably read them with no difficulty" (Open Office Marketing).

• Audio—FLAC now playable in iTunes:

> "Due to the design of iTunes, only Apple can add support for FLAC. And why wouldn't they? FLAC usage is accelerating, many bands like Pearl Jam, The Beatles, Phish, Dave Matthews Band, Metallica—the same hip, influential people whose fans Apple courts—are already distributing music in FLAC format, and users are clamoring for it in the iTunes forums" (FLAC website). [10]

So, in terms of exchangeability—are the various CAQDAS software packages comparable and how open are they? In terms of many of the core software functions that are shared they are comparable. Each package has its own unique style and functions but there are some core common features that include:

• structuring work—ability to access to all parts of a project immediately;
• staying "close to data"—instant access to source data files (e.g., transcripts);
• exploring data—tools to search text for words or phrases;
• code and retrieve functionality—create codes and retrieve the coded sections of text;
• project management and data organisation;
• searching and interrogating the database—searching for relationships between codes;
• writing tools—memos, comments and annotations;
• outputs—reports to view a hard copy or export to another package. [11]

For more of an overview on the principles of CAQDAS packages see LEWINS and SILVER (2007) and for a read on the original thinking and classic introduction, see FIELDING and LEE (1991). [12]

The heart of our position in this article is that if we all use a common standard to annotate data, or to export to, then more power is offered to our data for the longer term. [13]

3. Adding Value to Qualitative Data

In the 1990s, the UK the research community recognised the needs of qualitative researchers by supporting a qualitative data archive, Qualidata (CORTI & THOMPSON, 2004). Since then, certainly in the UK and spreading across the world, is an emerging culture of preserving and re-using qualitative data, and some classic research studies are now preserved for researchers to consult. ESDS Qualidata has over 160 collections of qualitative data, held in raw, uncoded format. Furthermore, since the mid 2000s, many funders of research have recognised the value of keeping research data for future use and have set up data sharing policies requiring researchers to document and share data from their own projects. [14]

Traditional "data archiving" models typically look at the raw data from any one research project. They assemble the data in the most logical way, guided by the research investigator, they ensure it is not breaching ethical or legal concerns and they find ways to describe this data by adding context, collating research instruments and so on. [15]

All data archives require common descriptive standards to describe, store and provide access to data. Examples of elements to describe that can use standardised fields and terms are:

- study description—design, methods, provenance;
- data types and files;
- relationships between files and parts of files;
- version of data (raw, edited, whose annotations);
- access conditions. [16]

Current ways of describing these collections is to use a paper-based archival approach known as the General International Standard Archival Description (ISAD-G) or an agreed descriptive standard for social science data—such as the DDI. These allow consistent descriptions of collections using formal fields. Various data archives that do have qualitative collections utilise the DDI in its crudest form. They use core fields such as study title, date of field work to provide an overview of the study such as a collection of raw interview transcripts in word format[2]. ESDS Qualidata also uses a structured "datalist" to accompany these

---

2   See a typical ESDS Qualidata study of qualitative interviews, such as those carried out with 52
    doctors from NHS health care settings: http://www.esds.ac.uk/findingData/snDescription.asp?

raw data which is constructed by the investigator and archive[3]. This list already alludes to some rough classifications of data, for example gender, location, and type of interview. Other useful classifications that may have been created in managing and analysing data using CAQDAS packages, such as codings, memos, variables or relationships, cannot be accessed outside of the package. [17]

3.1 Example of adding value to a text document

An example of how we might assign descriptors to an in-depth interview transcript might be

- adding metadata to an MS Word document via Properties, such as author, title, keywords, and version, see Figure 1:

kwalon fqs 10 Properties

General | Summary | Statistics | Contents | Custom

Title: WHAT ARE DATA ARCHIVES
Subject: FQS papwr
Author: s2591
Manager:
Company: University of Essex

Category: paper
Keywords: data, cadqas, archiving, software, standards
Comments:
Hyperlink base:
Template: Normal
☐ Save preview picture

OK | Cancel
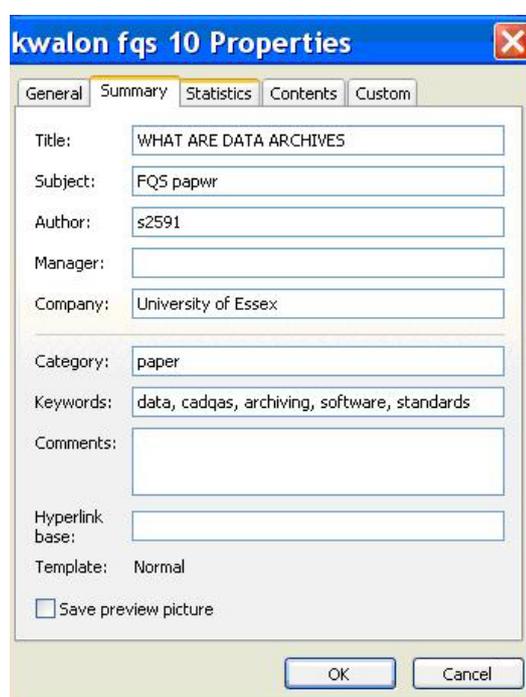
Figure 1: Example of MS Word Properties for capturing metadata about the document

- adding meta information, about an interview to the header of the Word document e.g. research study, name, ID, date of interview, interviewer, see Figure 2:

sn=6124&key=doctor [Accessed: January 18, 2011].

3   See a typical ESDS Qualidata data list: http://www.esds.ac.uk/doc/6124%5Cmrdoc%5Cpdf%5C6124ulist.pdf [Accessed: January 18, 2011].

Study Name: Interview number:
Depositor: Interview ID: Firstname Lastname
Interviewer: Date of interview:

Information about interviewee
Date of birth: Marital status:
Gender: Occupation:
Geographic region:

I: Just one or two factual details first of all before we go on to your health and
       that.... how old are you?

FL: I'm 58 in June.

I: What schools did you go to? Can you remember that far back!

Figure 2: Example of MS Word header information, taken from the UK Data Archive's guidance on transcription

- compiling a list of all interviews in a collection with summary details, as discussed earlier and used by ESDS Qualidata collections, see Figure 3:

Study Number 6124
Being a Doctor: a Sociological Analysis, 2005-2006
Nettleton, S

Interview ID	Gender	Description	Date of Interview	No of Pages	Text File Name
x001	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/09/2005	36	6124int001
x002	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	41	6124int002
x003	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	39	6124int003
x004	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	36	6124int004
x005	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	34	6124int005
x006	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	14/09/2005	50	6124int006
x007	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	31	6124int007
x008	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	35	6124int008
x009	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	33	6124int009
x010	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	23	6124int010
x011	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	36	6124int011
x012	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	41	6124int012
x013	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	21	6124int013
x014	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	30/09/2005	20	6124int014
x015	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	19	6124int015
x016	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	27	6124int016
x017	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/10/2005	27	6124int017
x018	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	17/10/2005	11	6124int018
x019	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	19/10/2005	33	6124int019
x020	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/11/2005	21	6124int020
z001	Male	Interview with Hospital Doctor in Northern Tourist City	19/07/2005	50	6124int021
z002	Male	Interview with Hospital Doctor in Northern Tourist City	10/08/2005	45	6124int022
z003	Male	Interview with Hospital Doctor in Northern Tourist City	17/08/2005	56	6124int023

Figure 3: Example of data listing from ESDS Qualidata

- adding metadata through manual or automated mark-up of XML documents:
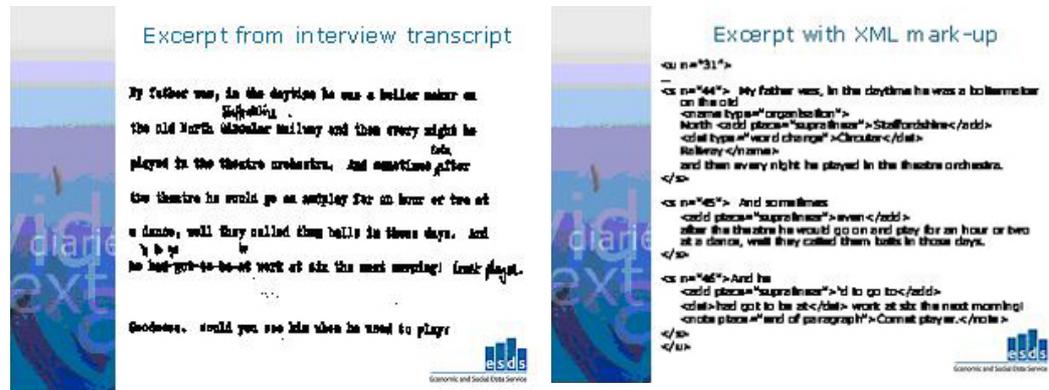
Figure 4a: Example taken from ESDS Qualidata for recording, formally an annotated qualitative interview using XML mark-up

Figure 4b: Example taken from ESDS Qualidata for an automated marked up qualitative interview using a beta text mining programme, not available (enlarge this figure here) (CORTI, 2006)

- annotating metadata about the interview in a CAQDAS package, in this case an older version ATLAS.ti, see Figure 5:
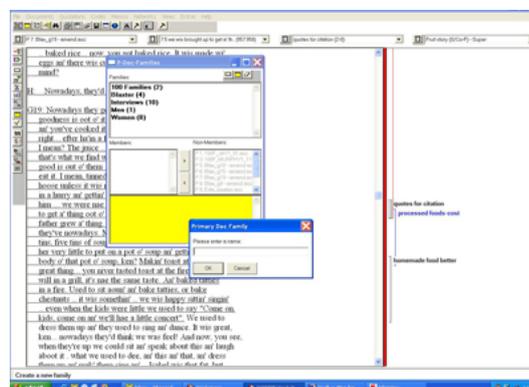
Figure 5: Example taken from ESDS Qualidata—transcripts assigned to groups and coded using ATLAS.ti (enlarge this figure here) [18]

3.2 Why is value-added material useful?

Archives are certainly very interested in exchange of all kinds of research data not only because they have a remit to preserve and share primary data through data sharing national policies, but also because, in the case of qualitative data, value-added "mark-up" offers more context for the re-user. Representation, coverage and context of research data are topics that have given rise to some heated debates within the qualitative data community. The fundamental issues of if and how someone else's raw data can be used were addressed by the UK qualitative data archive which began in 1994 (CORTI & THOMPSON, 2004). [19]

The basic argument lies with the belief that qualitative data cannot be used sensibly without the accumulated background knowledge and tacit understanding that the original investigator had acquired, which may typically not be written down formally, but held in the researcher's head. There have been a couple of vociferous critiques in the literature that consider the act of secondary analysis of qualitative data as both impractical and impossible (MAUTHNER, PARRY & BACKETT-MILBURN, 1998; PARRY & MAUTHNER, 2004). For example, the researcher's own deep engagement in the fieldwork and ongoing reflexivity enhances the raw data gathered and stimulates the formulation of new hypotheses in the field. In the process of analysing and coding data, researchers use their own personal knowledge and experiences as tools to make sense of the material that cannot be easily be explicated nor documented. If memos are used and are available to reflect this process, the analytic pathways leading to findings are made more transparent. [20]

However, the loss of context in archived data should not be seen as an insurmountable barrier to re-use. Indeed, there are very common and accepted instances where research data is used in a "second hand" sense by investigators themselves. For example, principal investigators working with data and writing up their final analyses and reports may not have been directly engaged in fieldwork, having employed research staff to collect the data with which they are working. These researchers rely upon sharing their own experiences of fieldwork and its context in order to make sense of the data gathered. In both instances, the analysers or authors must rely on fieldworkers and co-workers documenting detailed notes about the project and communicating them—through text, audio and video. Indeed, description of the research process can help recover a degree of context, and whilst it cannot compete with "being there", audio recordings, field notes, fieldwork diaries letters and memos can serve to help aid the original fieldwork experience. [21]

This value-added material created by researchers in helping to classify data (even though usually subjective) may be useful for a number of specific reasons:

- teaching with data where students can scrutinise or critique coding schemes and compare them against their own classifications;
- for getting data into an archiving or viewing system more easily, such as a research repository, where a research team might want to look at each other's data, or keep it in a more formal longer-term archive. Having some relationships between data already defined can be very useful. The use of social tagging is also becoming increasingly acceptable as we allow our own classifications to be shared;
- for exploring a very large collection, in, for example, a CAQDAS package, to show existing codings and interesting points;
- for providing additional context when working across collections by helping to gain insight into the data from researchers' reflections and memos. [22]

Providing a researcher's added-value codings, musings and other relationship annotations then offers additional and beneficial context. But, what is needed is the facility for original researchers to export the fruits of their labour to a format that would enable a future for their value-added data and for re-users to import older value-added data (context) into their own choice of software. [23]

One answer to the loss of annotation and context could be to use the same software package to iteratively add more. From a longer-term re-use perspective, this would be impractical as we are trying to advocate exporting and storing data outside of a given package. [24]

3.3 How can we capture these rich descriptions in a systematic way?

While raw qualitative data are typically available in Word, RTF, MP3, JPEG/TIFF, they are very easy to export to archive and share in open formats e.g. XML, FLAC, TIFF, PDF/A. Annotated data include derived data, linked data, annotated data, analysed output data, models and are currently harder to archive and share. [25]

However, all of the value-added annotations discussed above—be they descriptions or classifications, codes or memos, can be stored in a fairly simple structured metadata schema. A standard descriptive format for representing richly encoded qualitative data is useful because it ensures some degree of consistency across data collections (e.g., layout of a focus group transcript), supports the development of common web-based publishing and search tools, and facilitates data interchange and comparison. [26]

CAQDAS packages allow very rich annotation of data in a user-friendly way. Some packages allow export of a list of codes, annotations or memos, but the links to the underlying data or segments of data are usually lost. They are stored as an integral part of the software "project" which cannot typically be exported as a whole unit. ATLAS.ti is the only package to allow export of data and codes in a structured and connected way, using what is termed stand-off XML (where the XMLS is not embedded in the text or file). It uses its own XML schema which can very easily be translated into a common shared language. [27]

3.4 Import is important too

A CAQDAS import facility from a metadata standard for value-added data would enable users to utilise previously analysed or annotated data (by themselves or someone else), for example where:

- users are more familiar with one package but not another;
- an institution/department/team has a commitment to a single license for one software:
- a research group choose a common shared space to have access to recent versions of "analysed data"—the idea of "research archive boxes";
- the functions of particular software are unique. [28]

Many smaller archives and research groups are looking for "out the box" repository set ups for their projects' qualitative data. It is always preferable to configure upload of data using common metadata standards and templates. However, a simple one-size-fits-all set up doesn't (yet) exist although agreed basic standards can help us move a few steps closer. Good examples of where research groups are sharing qualitative data through a repository utilising structured metadata are Timescapes (BISHOP, 2007) and Ensemble (MARTINEZ, CARMICHAEL & CORTI, 2009). The heart of the matter for any project of this kind is to ensure that the defined basic relationships between data files and parts of data are specified or maintained. For team working shared rules and shared ontologies (classification and codes) are needed. Ideally, import of CAQDAS files from the different researchers to a common system should be possible. The idea of exportability does not undermine the commercial nature of software which wants a share of the CAQDAS market. Most researchers will choose a commercial CAQDAS solution for the "live phase" of their project. The SPSS.por model supports this—the use of R as an open source statistical analysis software has not taken a huge share from the SPSS or STATS giants. [29]

3.5 Agreeing a common data exchange standard—the QuDex schema

The development of a shared data exchange model took as its starting point two recent and quite roughly specified data models that had already been developed in the early 2000s. The first has been an ongoing work programme of ESDS Qualidata at the UK Data Archive who developed a draft but limited formal definition of a common XML vocabulary and Document Type Definition (DTD) based on the Text Encoding Initiative (TEI) for describing these structures (ESDS QUALIDATA, 2004). The Universities of Melbourne and Queensland had further developed a draft Qualitative Data Interchange Format for e-Social Science. [30]

Both initiatives shared early development ideas but the work took off in 2007 under a research and development project run by the UK Data Archive and Metadata Technology funded by the UK's JISC. The DExT project compared the key functionalities for the market-leading software packages (UK Data Archive 2008a). This comparison helped to distinguish the baseline, what may be thought

of as common denominator functions for coding or annotating data possessed by all software: coding, classifying, memoing, and relating. A comparison of existing and possibly relevant metadata schema was also undertaken. The core concepts identified helped defined the QuDEx schema, a draft qualitative data exchange model for the archiving and interchange of data and metadata between CAQDAS packages, termed QuDEx (UK DATA ARCHIVE, 2008b). The draft QuDEx standard/schema is essentially a software-neutral format for qualitative data that preserves annotations of and relationships between data and other related objects. See Table 1 for QuDEx elements and definitions.

Top level Elements	Sub elements	Definition
<qudex>	resourceCollection segmentCollection codeCollection memoCollection categoryCollection relationCollection	The root element; a "wrapper" for all other elements of the QuDEx Schema. Each top level element in QuDEx is defined as a "collection" and must appear in the order outlined below.
<resourceCollection>	sources memoSources documents	The *resourceCollection* section lists and locates all content available to the QuDEx file. A *source* points to the original location of the resource while each author working on the QuDEx file is assigned a surrogate *document* which points to the relevant *source*. The child elements *sources* and *memoSources* contain direct references to the files under analysis; the *documents* section contains their surrogates.
<segmentCollection>	Segment (sub elements text, audio, video, xml, image)	The parent element for all *segments*, which is a subset of a *document* (text, audio, video or image) under analysis defined in a manner appropriate to the format (text, audio, video, image or xml). *Segments* may overlap and multiple *memos* and *codes* may be assigned to a *segment.* Start and end points can be formally assigned to segments of text, and audio visual materials in other document.
<codeCollection>	code	The parent element for all *codes*. A code is a short alphanumeric string, usually a single word; may be assigned to a segment or document though assignment is not required. A code may optionally be taken from a controlled vocabulary defined under @authority.

Top level Elements	Sub elements	Definition
\<memoCollection\>	memo (sub elements memoDocumentRef, memoText)	The parent element for all *memos*; these may be pure text and embedded in the QuDEx file (inline memo) or may refer to external files. A memo is a text string internal to the document (inline memo) or an externally held document (external memo) which may be assigned to a segment, code, document, category or to another.
\<categoryCollection\>	category	The parent element for all categories. A category is an alphanumeric string (stored in @label) assigned to one or more documents. Categories may be hierarchically nested. Documents contained within a category are referenced using @documentRefs. Nested categories are referenced using @categoryRefs.
\<relationCollection\>	objectRelation	The parent element for all relationships between objects. For the purposes of a *relation* all of the following are considered to be "objects": A *document*: surrogate of a *source* or *memoSource* A *segment* within a *document* An assigned value: *code*, *memo*, *category*, *relation* A relation is a link between two objects in a QuDEx file. Each object is either the start or end point of a relation (source vs. target). Every relation may, optionally, have a name

Table 1: QuDEx elements and definitions [31]

A number of attributes are commonly used within the QuDEx standard, with standard attribute groups assigned designed to support the management of complex layers of analysis by multiple authors within a single QuDEx instance. These are:

@ cdate: the date and time the instance of the element was created

@ mdate: the last date and time the instance of the element was modified

@ creator: the original creator of the instance of the element or the author of the relevant resource

> @ label: a human readable string for the element in general or its specific contents
>
> @ displayLabel: a version of the label text appropriate for display, for example in a user interface
>
> @ language: this caters for describing the overall language of the study while permitting element level variations such as defining a segment, memo or code as being in a different language. [32]

All of the key CAQDAS vendors were consulted throughout the QuDex development process: ATLAS.ti, QSR's NVivo, NU*DIST, MAXQDA, QDA Miner, Qualrus, HyperRESEARCH, Tinderbox, Transana and WeftQDA. The draft V3 QuDEx schema, its UML model, accompanying documentation and XML instance files were released on the DExT Website. [33]

The schema accommodated most common file types used in qualitative research —text, audio, visual, image and also tried to make provision for most kinds of relationship between data, parts of data and code, classification or memos. It also included multiple forms of "referencing"; or "offsetting"—the method used to link a segment of data to a code, classification or memo, or other segment of data. A number of issues remain for discussion, such as the problem where an ingest or export system for a QuDEx file must take account of the fact that paths to source files may refer to a user's local machine, or needing to consider the possibility of defining underlying resources as inherently synchronised (e.g. an audio interview and its associated transcript). Finally, there may be some actions on data that a particular software package may use. These functions would not be exported to this schema, but the schema is extensible so could be added if they were generic enough. [34]

Transforming data from a proprietary data structure into the Qudex Schema is impossible without having access to the underlying data structure. The DExT project was not able to do this on a significant scale within the time frame, and a single leading software was chosen as the test case, ATLAS.ti—because it already exported data into basic XML. This enabled a mapping from this format (Atlas-XML) to the QuDex generic XML and back again to demonstrate proof of concept, which was important to show that the annotations could be defined in a way true to the original. A basic demonstrator of import and export utilities was developed in the last month of the project. [35]

Two simple demonstration tools were developed: the QuDex Viewer helped support proof of concept and was an open source tool whose objective is to facilitate the use and understanding of QuDEx XML files[4]. It was designed for simple browsing, transforming and viewing of core constructs such as code, segments, memo and their relationships and it also transformed ATLAS.ti XML to its native XML format. To use the software, any valid QuDEx XML file on a computer can be opened by using the application's menu at the top of the page. For all QuDEx XML documents, it provides basic functionalities such as a listing

---

4   The QuDEx Viewer requires the installation of tomcat server on to a local directory to run the application into a standard web browser such as Firefox or Internet Explorer. The QuDEx Viewer leverages the open sources packages, Yahoo! UI Library and Bubbling Library.

of all the codes, memos, segments, categories and their relationships and the ability to browse the document in either text or XML view. Sample QuDEx documents can be found on the DExT web site. [36]

The Transformer application allows transforming ATLAS.ti XML to QuDEx format by simply selecting the files for transformation. This feature can also be accessed using the top menu of the application. The files should be stored on a local drive and the application pointed to these files for transformation. Screenshots are shown in Figure 6. The tools, while basic, could be used as the basis for import/export and transformation.
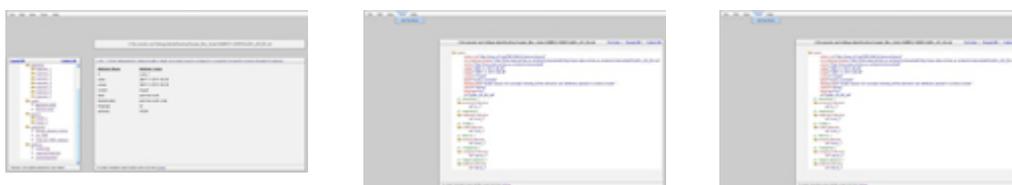
Figure 6: Three screen shots of QuDex Viewer (click on the figures to enlarge [37]

The very initial schema for the baseline concepts was presented by CORTI and BHAT (2007) to a significant number of the CAQDAS vendors at the CAQDAS 07 Conference: Advances in Qualitative Computing held in April in 2007 at the University of London. This was the first time the vendors had an opportunity, if only though curiosity, to come together to discuss what potentially might represent "competition" between them. While not all in agreement about the model and value of an exchange standard, the majority did agree that a discussion of which basic aspects of functionality might be common across packages would be worthwhile. [38]

4. Data Exchange—Conclusion

Testing of the robustness of this schema and conversion tools will be needed if take up is to be encouraged. Software vendors may need to take some initiative in developing their own project-based import and export tools if they wish to cater for open format translation. [39]

Following DExT staff attending an Open Data Foundation (ODAF) meeting in 2008 to discuss the DExT project and tools, the team was invited to propose a working group of the Data Documentation Initiative (DDI) alliance in an attempt to progress the standard. This was a major breakthrough in getting such a standard for qualitative data recognised. A presentation at the Association of Survey Computing (ASC) in September 2007 also gave an opportunity for feedback by wider social survey tools developers. In April 2008, after the project finished, another ODaF meeting was held, hosted by the UKDA, at which the DExT tools were discussed with unanimous support for the development under the ODaF umbrella of visualisation tools based on the QuDex schema. [40]

This draft standard has now been adopted, internationally, by the social science data sharing and archiving community as the basis for further development for a common standard. In 2009 a working group of the DDI on Qualitative Data has been assessing common needs for archiving qualitative data and work has begun on mapping DDI and QuDex and working with ingesting data into open source repository software such as FEDORA. The ENSEMBLE project (Semantic Technologies for the Enhancement of Case Based Learning) is building semantic web applications for case-based learning in education with reference to how Web 2.0 and Web 3.0 tools can be used for representing, visualising and communicating data. It is using DDI and QuDEx to look at data ingest. [41]

Ongoing support for such tools within our own community and by software vendors will ensure that they will be further teste and refined and our hope is for them to be fully embedded in everyday archiving and repository practice. If you are interested in supporting this initiative, please contact corti@essexac.uk. [42]

References

Bishop, Libby (2007). Moving data into and out of an institutional repository: Off the map and into the territory. *IASSIST Quarterly*, *Fall/Winter 2007*, http://iassistdata.org/downloads/iqvol313bishop.pdf [Accessed: November 28, 2010].

Carmichael, Patrick (2002). Extensible markup language and qualitative data analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *3*(2), Art. 13, http://nbn-resolving.de/urn:nbn:de:0114-fqs0202134 [Accessed: November 28, 2010].

Corti, Louise (2006). End of award report to ESRC for qualitative data tools (SQUAD). *Final report to the ESRC on SQUAD*, http://www.data-archive.ac.uk/media/1684/ESRC_SQUAD_finalreport.pdf [Accessed: November 28, 2010].

Corti, Louise & Bhat, Angad (2007). Qualitative data exchange: Methods and tools. *CAQDAS 07 Conference, L*ondon, UK, 19th April 2007, http://dext.data-archive.ac.uk/presentations/dext218apr07.pdf [Accessed: January 25, 2010].

Corti, Louise & Thompson, Paul (2004). The secondary analysis of archived qualitative data. In Clive Seale, Giampreto Gobo, Jaber Gubrium and David Silverman (Eds.), *Qualitative research practice* (pp.327-343). London: Sage.

ESDS Qualidata (2004). *XML application for qualitative data*, http://www.esds.ac.uk/qualidata/online/about/xmlapplication.asp [Accessed: November 28, 2010].

Fielding, Nigel & Lee, Ray (Eds.) (1991). *Using computers in qualitative research*. London: Sage.

Lewins, Anne & Silver, Christina (2007). *Using software in qualitative research: A step-by-step guide*. London: Sage.

Martinez, Agustina; Carmichael, Patrick & Corti, Louise (2009) The Ensemble project: Using the semanti*c* web for education. *Open Repositories Conference, 20th May 2009,* http://smartech.gatech.edu/bitstream/1853/28468/1/158-466-1-PB.ppt [Accessed: November 28, 2010].

Mauthner, Natasha; Parry, Odette & Backett-Milburn, Kate (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733-745.

Muhr, Thomas (2000). Increasing the reusability of qualitative data with XML. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(3), Art. 20, http://nbn-resolving.de/urn:nbn:de:0114-fqs0003202 [Accessed: November 28, 2010].

Parry, Odette & Mauthner, Natasha (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology*, *38*(1), 139-152.

UK Data Archive (2008a). *Comparison of CAQDAS packages key functionality*, http://dext.data-archive.ac.uk/Software2.xls [Accessed: November 28, 2010].

UK Data Archive (2008b). *QuDEx Schema Version 3*, http://dext.data-archive.ac.uk/schema/schema.asp [Accessed: November 28, 2010].

Authors

Louise CORTI is an Associate Director at the UK Data Archive where she heads the units of ESDS Qualidata, Research Data Management Support Services and Communications. She coordinates the international DDI Working Group on Qualitative Data.

Contact:

Louise Corti

UK Data Archive
University of Essex
Colchester
CO4 3SQ
UK

Tel.: +44 1206 872145

E-mail: corti@essex.ac.uk

Arofan GREGORY has undertaken technical work on the SDMX and DDI standards and tools development, including the registry. He works on technical implementation of these tools for a wide range of data organisations.

Contact:

Arofan Gregory

Metadata Technologies North America
00 Prosperity Drive
Knoxville, TN 37923
U.S.A-

Tel./Fax: +1 865 245 45 42

E-mail: arofan.gregory@earthlink.net

Citation

Corti, Louise & Gregory, Arofan (2011). CAQDAS Comparability. What about CAQDAS Data Exchange? [42 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(1), Art. 35, http://nbn-resolving.de/urn:nbn:de:0114-fqs1101352.