Choosing a CAQDAS Package

A working paper by Ann Lewins & Christina Silver, 6th edition April 2009
CAQDAS Networking Project and Qualitative Innovations in CAQDAS Project. (QUIC)

See also the individual software reviews available from the CAQDAS Networking Project website, the referenced listed below and the following publication: Lewins & Silver (2007) **"Using Software for Qualitative Data Analysis : A step-by-step Guide"** Sage Publications, London

---

It is not always easy to visualise exactly what a CAQDAS package offers when exploring it for the first time yourself. Equally, when asking someone else for their opinion, it is not always easy to know which questions you should be asking. Most of the software packages we are aware of and discuss regularly are excellent products in several ways. Sometimes you choose the package that is already *in situ* and make good use of it – but if you have a choice about which software to purchase for your research project, you may be in some uncertainty about how to proceed. You may have basic understanding about what a CAQDAS package will do, but the differences between the tools offered in each package are subtly but significantly different.

What does this paper do…?

- This paper offers a **summary of types of software for managing textual or qualitative data as a starting point for thinking about which one may be most suited to the type of project and data** you are working with and the way you like to work.
- It provides more detailed information focused mainly on those traditionally categorised as CAQDAS packages. We cannot provide an exhaustive description or comparison of all available CAQDAS packages here, but we aim to highlight some key distinguishing elements in order to provide you with an early impression of each package.

    o Firstly we provide a description of the tools that these software packages have in common.
    o Secondly we provide links or references to the separate reviews available on a range of CAQDAS packages. Each review essentially builds on information here and they are available separate pdf files.

- It aims to address the most frequently asked questions that we at the CAQDAS Networking Project receive.
- It aims to assist you in your search for more information by strongly recommending that you also visit the Software Developer websites as part of your decision-making process, where you can access and download demonstration versions.
- It aims to review both commercially available software and open access and free or low-cost software.

In order to help you decide between software the CAQDAS Networking Project provide a series of (FREE) Software Planning Seminars [where we raise some of the issues you may need to consider, and discuss how some of the software programs differ in the way they handle such aspects.
http://caqdas.soc.surrey.ac.uk/training.html

What ranges and types of software support work with qualitative data?

In 1995 Miles and Weitzman[1] developed a typology concerning the handling and management of qualitative (and at that time, mostly textual) data. We make some reference to this typology, because it is still current in some respects, but increasingly with the advances of individual software described in this paper, the distinctions for instance between software formerly labelled Code and Retrieve and the more extensive functionality in Code-based Theory Builders have become blurred. Now very few software programs remain in the Code and Retrieve category. Similarly, some of the Code-based Theory Building software have taken on features more traditionally featured in Text retrievers, or Text based managers – e.g. *Content analysis* tools, Word frequencies, word indexing with *Key Word in Context retrieval* (KWIC) and complex text based searching tools. Conversely, by expanding to provide different *add-on* software to enable thematic coding, Text Retrievers and Textbased Managers are beginning to enable tasks originally only possible in Code-based Theory Building software.

Code-based Theory building software

Both these and the earlier Code and Retrieve packages assist the researcher in managing the analysis of qualitative data, to apply thematic coding to chunks of data, thereby enabling the reduction of data along thematic lines (retrieval), limited searching tools and probably good memoing facilities. Code-based theory building software packages build on those tools and extend the collection of *search* tools allowing the researcher to test relationships between issues, concepts, themes, to e.g. develop broader or higher order categories, or at the other extreme, to develop more detailed specific codes where certain conditions combine in the data. Some of the programs enable the graphic visualisation of connections and processes using mapping tools.

Text Retrievers, Textbase Managers

Tools provided in both these categories often provide more quantitative and 'actual content' based analysis of textual data. There are complex and sophisticated ways to search for text and language including the use of thesaurus tools to find words with similar meaning, to index all words contained in the text, to provide word frequency tables, to create active word lists, to provide easy key word/phrase in context retrieval (KWIC). However, broadly summarising them in this fashion masks the huge variety of differing ways in which they each handle the quantitative analysis of content, and the importance given to particular aspects in each software. Generally speaking, some automatic 'content' analysis often happens just as part of the process of importing data.

Some Textbase Managers have very sophisticated 'content analysis' functions; creation of keyword co-occurrence matrices across cases, creation of proximity plots for identification of related keywords, charting and graph building facilities etc. Textbase Managers tend to offer more functionality than Text Retrievers and more possibilities to manage huge datasets in varied ways, but it must be said again, that all these categories are becoming blurred, and we link to developer sites and software within these two categories without making any distinction between them.

The code-based theory building category itself is blurred by the increasing addition of some quantitative language based content analysis features (MAXqda, with add-on software MAXdictio with good KWIC), and to a lesser extent ATLAS.ti.

One or two of the Textbase Manager packages now incorporate 'thematic' qualitative coding functions which can be integrated with the range of content analysis tools e.g. the language based quantitative functions. So they offer a comprehensive range of both qualitative and quantitative approaches to data within one software, e.g. QDA Miner with the add-on Wordstat module, and CISAID. Where this is the case, we will try to provide information on the particular attributes of such software program's tools (from a qualitative standpoint) and will continue to add more comment and comparative information to this online resource as we undertake a review of each software.

---

[1] Weitzman, E. & Miles, M.; (1995) **A Software Source Book: Computer Programs for Qualitative Data Analysis** Thousand Oaks, Sage Publications.

Which software packages do we categorise as *CAQDAS?*

Software which falls under the CAQDAS 'umbrella' includes a wide range of packages but its general principles are concerned with taking a qualitative approach to qualitative data. A qualitative approach is one where there is a need to interpret data through the identification and linking or coding of themes, concepts, processes, contexts, etc., in order to build explanations or theories or to test or enlarge on a theory. Many approaches (including, for example, participant observation, action research, grounded theory, conversation analysis etc.) broadly make use of qualitative perspectives, though each may emphasise particular processes or differ in their sequence. Qualitative researchers may also use quantitative data, for instance, relating to the descriptive features of research participants – gender, age and so on – to help in the organisation of qualitative data. This assists in cross-referencing and comparing qualitative ideas across groups and subsets. Qualitative approaches remain distinct from 'content analysis' methodology, in which the presence of words, phrases and concepts within texts are analysed for their implicit meanings, relationships to other words and phrases and for the messages they convey. We include content analysis packages only where they also have a strong qualitative (thematic coding) element.

For more information on qualitative methodologies, including definitions and links to further reading and online resources and for a glossary of research terms see the OnlineQDA website http://onlineqda.hud.ac.uk

Definition

We use the term 'CAQDAS' to refer to software packages which include tools designed to facilitate a qualitative approach to qualitative data. Qualitative data includes texts, graphics, audio or video. CAQDAS packages may also enable the incorporation of quantitative (numeric) data and/or include tools for taking quantitative approaches to qualitative data. However, they must directly handle at least one type of qualitative data and include some – but not necessarily all – of the following tools for handling and analysing it:

- Content searching tools
- Linking tools
- Coding tools
- Query tools
- Writing and annotation tools
- Mapping or networking tools

The combination of tools within CAQDAS packages varies, with many providing additional options to those listed here. The relative sophistication and ease of use also varies and we aim to uncover some of these differences in our comparative reviews. Although it is sometimes purported that CAQDAS packages are reliant on or biased towards code-based approaches, we would not see coding tools as a pre-requisite for a software package being included under the CAQDAS umbrella.

In deciding which software packages to include in our comparative reviews we aim to both reflect the market in discussing those which are well-known and commonly used, and to raise awareness of newer and more methodologically specific options.

Which is the 'best' CAQDAS Package…?

This is perhaps the most frequently asked question we receive – however, it is impossible to answer! All the packages we work with and teach have their own range of tools to support you in the various stages of the analytic process. They each have their own advantages and disadvantages.

Whilst we would argue that **some software packages may be more suited to certain types of approach, their purpose is not to provide you with a methodological or analytic framework**. The tools available may support certain tasks differently and there is some debate as to whether a particular package may 'steer' the way you do the analysis . **However, as the researcher you should remain in control of the interpretive process and decide**

which of the available tools within a software can facilitate *your approach* to analysis most effectively. Whichever package you choose, you will usually be able to utilise a selection of tools which will help significantly with data management and analytic tasks .

Thinking about and using CAQDAS software should not necessarily be any different from other types of software package – just because a tool or function is available to you does not mean you will need or have to use it. You therefore need to think clearly about what it is you are looking to the software to help you with and we would caution against choosing a package simply because it seems the most sophisticated. Conversely, and this may have something to do with your longer term plans, you may feel a more ambitious selection of tools will serve you better over time.

The Basic Functionality: Key Elements in most CAQDAS Packages

Whilst there are many differences between CAQDAS packages, the key principles behind each of them in terms of facilitating the qualitative research process are similar in many ways; (the ways in which they differ are described in the software specific reviews). We also refer you to parts of Lewins & Silver 2007 and other publications where you can find more discussion about these aspects of using CAQDAS packages.

See also the Online QDA pages about what CAQDAS package do, including discussion of advantages and disadvantages. http://onlineqda.hud.ac.uk/Which_software/what_packages_are_available/index.php

Structure of work

Much of the data and information about the research project can be managed more efficiently because the user effectively needs to open only one file to have access to all these elements. The 'project file' that the user creates in the software acts as a container for or a connector to all the different data files within the current working project. Internal databases, *contain* the individual data files – i.e. when you *import* data it is a *copy* process. External databases *connect to* the data files – i.e. you *assign* data to the project, but the data remains in its original location. In either case, the opening of one project file enables immediate access to all components of the dataset. There are varying benefits and disadvantages to each system and some packages allow some files to be internal whilst others remain external. Internal databases are generally easier to move from computer to computer, whilst external databases may cope better with very large datasets and are more likely to directly handle a range of multimedia files. It is important to understand how the chosen software functions in this respect before work starts but this need not be a deciding factor between packages. *See Lewins & Silver 2007, chapter 1: Processes and Tasks in Using Qualitative Software.*

'Closeness to data' and interactivity

At the most basic level, the packages discussed here provide almost instantaneous access to all source data files once introduced into the 'project'. Whatever tools you subsequently use, 'live' contact to source data is always easy, which increases the researcher's closeness to data. It is sometimes contended that software acts as a barrier between the researcher and the data, but we argue that conversely, customised CAQDAS packages designed with the needs of the qualitative researcher in mind, actually enable much more reliable contact with source data than is possible when working manually. That said, using a CAQDAS package does not mean you need to do all your work behind a computer screen; print-outs and highlighter pens still have their place. *See Lewins & Silver 2007, chapter 13: Convergence, Closeness, Choice*

Explore the data

Many CAQDAS packages offer ways of exploring data without necessarily abstracting to the conceptual or coding level. Annotation tools may act like footnote tools in word processor applications enabling comments to be made about and anchored to specific pieces of data. Linking tools vary quite significantly between software but they can provide very flexible ways of manoeuvring around data to follow processes or to link pairs of statements. Alternatively 'text search' tools offer ways to search for one word or phrase, or even a collection of words around the same topic area. Such searches usually provide good access to parts of the documents where

those keywords appear – allowing a fairly instant retrieval of topic related material – sometimes abbreviated as KWIC (Key Words in Context). The sophistication of these tools vary quite significantly between packages but their use may act as a starting-point for exploring a large volume of textual data, or be used as a means of conducting a quantitative content analysis. *See Lewins & Silver 2007, chapter 4: Exploration and Text-level Work*

Code and Retrieve Functionality

All of the CAQDAS packages reviewed here offer code and retrieve functionalities. User-defined key-words and/or conceptual categories (codes) can be applied to selections of (overlapping, embedded) text and as many codes as required can be applied to the same text selection. Usually, the user has complete freedom concerning the structure of the coding scheme and the coding strategies employed – e.g. whether inductive, deductive or a combination of both. The software used cannot solve common dilemmas of coding. A thorough awareness of these dilemmas and methodological issues concerning codes and coding should be a prerequisite to using the coding tools in software packages. Seidel and Kelle (1995) illuminate the issues and challenges of reliability and validity when coding in different ways.

In general terms within software, code generation is both easy and flexible and the researcher is free to use a combination of strategies if desired and to refine coding if necessary. In all but one package reviewed here, the responsibility for thinking about each coding action rests entirely on the user. In Qualrus, however, the software (using *Artifical Intelligence*) learns from previous actions and makes suggestions (see Qualrus Review). It should be emphasised however that such technology in no way resolves coding dilemmas already referred to.

In all packages, coded textual data can be retrieved, re-coded and outputted with ease. Software differ in the way coded information is provided and how codes are visible in the data file itself, for instance in the margin area. While such views can be produced by most of the software reviewed here, it may be the ease with which such a view can be generated, the flexibility of what can be shown and the interactivity of the margin area which present key differences. *See Lewins & Silver 2007, chapter 5: Qualitative Coding in Software : Principles and Processes; chapter 6: Coding Schemes, Coding Frames; chapter 7: Coding Tasks in Software and chapter 8: Basic Retrieval of Coded Data. See Gibbs 2007, chapter 4: Thematic Coding and Categorizing*

Project Management and Data Organisation

Any files which can be converted into the format(s) supported by the given software constitute data as far as the software is concerned. This means that supporting information such as literature lists and abstracts can be managed, cross-referenced and coded within the software project. *For a detailed discussion about planning your research using CAQDAS packages see di Gregorio & Davidson 2008, in particular chapter 2 Designing and conducting qualitative research in a software environment and chapter 3 Research design in context: the E-Project and communities of research practice. See also Gibbs 2007 chapter 8: Getting started with computer-assisted qualitative data analysis.*

Additionally all these packages offer powerful means by which to organise data according to known 'facts', descriptive features and data types for later comparison and queries. Such project management elements mean that much work can be done before primary qualitative data are introduced to the software. Conversely additional descriptive information in the form of attributes or variables can be applied at any stage. Such an action might be prompted by a particular question the researcher wishes to ask of a newly perceived sub-set. *So CAQDAS packages both reflect and significantly facilitate the 'cyclical' nature which is characteristic of many qualitative research and analysis processes. Even when used in the most basic way, therefore, CAQDAS software packages significantly increase the researcher's access to the different elements of their project and sub-sets of data. See Lewins & Silver 2007, chapter 2: Data and their Preparation in CAQDAS Packages and chapter 11: Organising Data to Known Characteristics.*

Searching and interrogating the database

At any stage of the process all the packages offer means by which to interrogate the dataset. This includes searching the content of data – based on language used – or search for relationships between codes as they have been applied to data (for example, co-occurrence, proximity etc.). Searches may produce a secondary level

of analytic coding. Search tools also allow you to combine the coding (interpretive or conceptual) and organisational (descriptive) dimensions of your work. It is also usually possible to auto-code data for content or structure, in which case an awareness of the units of text that can be *auto coded* to is needed at data preparation stage. See Lewins & Silver 2007, chapter 2 *Data and their Preparation in CAQDAS Packages and chapter 12: Interrogating the Dataset.*

Writing tools

The process of qualitative data analysis is rarely linear and the various writing tools (for example memoing, commenting, annotating etc..) offered by CAQDAS packages provide ways to increase the researcher's contact with thoughts and processes – provided, of course, they are used in a systematic way. See Lewins & Silver 2007, chapter 9: *Managing Processes and Interpretations by Writing*

Output

All reviewed software packages have a fairly standard selection of reports (output) which allow the user to view material in hard copy or integrate into other applications, e.g. Word, Excel, SPSS. Those standard reports will usually include coded segments either from one code or a selection of codes. The latter option is not always as straightforward as it seems but when it happens, it is accompanied by varying amounts of identifying information (source document identifier and the coded segments themselves is usually the minimum, and sometimes but not always, the relevant code label is included above the segment).

Tabular output: usually simple tabular output is available providing breakdown of code frequencies which can be exported to Word, Excel or SPSS. Programs vary in the types of tables which can be generated. Output also varies in terms of the level of its interactivity with live source data. Results of searches can be viewed often in both output format or inside the software, integrated into the working project.

When the software supports the use of mapping or more graphic representations of coding schema etc., these can usually be exported and pasted or inserted in Word or Powerpoint files. The fine distinctions in the ways that output can be varied by the choices the user makes in each software are not reviewed comprehensively in the online reviews, though we point out particularly distinctive or interactive forms of output. *See Lewins & Silver 2007, various chapters, as relevant.*

Summary

The combination of these key aspects of CAQDAS packages mean that the researcher is able to keep in reliable and continuous contact with the different elements of the project and the analytic processes. When used systematically and within the individual researchers' 'comfort zone', CAQDAS packages can aid continuity, and increase 'transparency' and methodological rigour.

Deciding on which is the 'best' CAQDAS software package is necessarily a subjective judgement and will probably be based on reaching a balance between a number of factors.

Some general questions to ask when choosing a CAQDAS package

- What kind(s) and amount of data do you have, and how do you want to handle it?
- What is your preferred style of working?
- Do you primarily want better access to your data and good note/memo making tools?
- What is your theoretical approach to analysis and how well developed is it at the outset?
- Do you have a well defined methodology?
- Do you want a simple to use software which will mainly help you manage *your* thinking and *thematic* coding?
- Are you more concerned with the *language*, the terminology used in the data, the comparison and occurrence of words and phrases across cases or between different variables?
- Do you want both thematic and quantitative content information from the data?

- Do you want a multiplicity of tools (not quite so simple) enabling many ways of handling and interrogating data?
- How much time do you have to 'learn' the software? …
- ….How much analysis time has been built into the project?
- Are you working individually on the project or as part of a team?
- Is this just one phase of a larger project – do you already have quantitative data?
- Is there a package – and peer support – already available at your institution or place of work?

If you can readily answer some of these questions it may help you to focus on certain types of functionality in the reviews to discover which package will work well for you. The last question should be very influential in your decision making since local support and the free or cheap provision of software is a good starting point. Local support is often lacking however and therefore it would be a useful step to find out where online support can be accessed.

References

Note: The references in this document refer to publications which discuss CAQDAS packages generally or provide details of more than one such program. Software developer websites often list publications concerning their specific products (see http://caqdas.soc.surrey.ac.uk/links.html for links) and the CAQDAS Networking Project's Online Bibliography http://caqdas.soc.surrey.ac.uk/bibliography.html provides a comprehensive and frequently updated list of publications about computer software and qualitative research more generally.

Fielding, N. & Lee, R. (eds.) (1998) **Computer Analysis and Qualitative Research**, Sage Publications, London

Gibbs, G. Friese, S. & Mangabeira, W. (2002) **The Use of New Technology in Qualitative Research. Introduction to Issue 3(2) of** *FQS*

Gibbs, G. (*2007)* **Analysing Qualitative Data**. *London*: Sage. Part of the Qualitative Research Kit, ed. U. Flick.

di Gregorio, S. & Davidson, J. (2008) **Qualitative Research for Software Users**, McGraw Hill, Open University Press, UK

Kelle U., (ed.)(1995)**Computer Aided Qualitative Data Analysis: Theory Methods and Practice** Sage Publications, London.

Lewins, A. (2008) 'CAQDAS: Computer Assisted Qualitative Data Analysis' in (ed) N. Gilbert **Researching Social Life** (3rd ed), Sage London

Lewins, A. & Silver, C. (2007) **Using Software for Qualitative Data Analysis : A step-by-step Guide** Sage Publications, London

Seidel, J. & Kelle, U. (1995) 'Different Functions of Coding in the Analysis of Textual Data' in Kelle U., (2008)**Computer Aided Qualitative Data Analysis: Theory Methods and Practice** Sage publications, UK

Silver C & Fielding N (2008) 'Using Computer Packages in Qualitative Research', in Willig C & Stainton-Rogers, W. (eds.) **The Sage Handbook of Qualitative Research in Psychology**, Sage Publications, London.

Silver C & Lewins L (forthcoming) 'Computer Assisted Qualitative Data Analysis' in **International Encyclopedia of Education, 3rd Edition,** Elsevier, Oxford UK

Weitzman, E. & Miles, M.(1995) **A Software Source Book: Computer Programs for Qualitative Data Analysis** Thousand Oaks, Sage Publications.

Willig C & Stainton-Rogers, W. (eds.) (2008) **The Sage Handbook of Qualitative Research in Psychology**, Sage Publications London.