



CAQDAS at a Crossroads: Affordances of Technology in an Online Environment

In: The SAGE Handbook of Online Research Methods

By: Christina Silver & Sarah L Bulloch

Edited by: Nigel G. Fielding, Raymond M. Lee & Grant Blank

Pub. Date: 2017

Access Date: May 21, 2020

Publishing Company: SAGE Publications Ltd

City: 55 City Road

Print ISBN: 9781473918788

Online ISBN: 9781473957992

DOI: <https://dx.doi.org/10.4135/9781473957992>

Print pages: 470-485

© 2017 SAGE Publications Ltd All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

CAQDAS at a Crossroads: Affordances of Technology in an Online Environment

Christina Silver Sarah L Bulloch

The History of CAQDAS and Current Developments

Software designed to facilitate the analysis of qualitative data, collectively known as Computer Assisted Qualitative Data Analysis (CAQDAS) packages, became available during the 1980s. Like other software supporting aspects of qualitative research, from document and bibliographic management, transcription, data collection, writing and visualisation (Paulus *et al.*, 2014), CAQDAS packages have since widely become seen as essential tools for researchers (Gibbs, 2014). Despite roots in the academic social sciences, their application extends far beyond to government, applied and commercial sectors.

The fact that separate groups developed CAQDAS packages at around the same time but in different countries and with different methodological impetus is significant. Tesch's (1990) seminal work relating analytic approaches to computing techniques highlighted from the outset of the CAQDAS trajectory that a range of analytic approaches results in the need for various software solutions (Fielding and Lee, 1991). It also explains the number of packages that have emerged and the differences between them (Silver and Lewins, 2013). The 1990s saw convergence in functionality, predominantly oriented around features designed to facilitate *qualitative* approaches to *qualitative* data, including data handling, content searching, code and retrieve, metadata organisation, Boolean and proximity querying and writing tools. Although several programs came to dominate the field, no global market-leader emerged.

Academic uptake was driven by those engaging with new technologies as part of their professional development – software developers who were themselves academics and methodologists across disciplines (Fielding and Lee, 2007). Discussion concerning the implications of CAQDAS has featured from the outset with distinct 'for' and 'against' camps. Some of the resistance to CAQDAS was ill-informed and has impacted upon the pace of uptake, although recent investigations of patterns of use have shown particular growth since 2000 (White *et al.*, 2012, Woods *et al.*, 2015, Gibbs, 2014).

This chapter focuses on current developments shaping the CAQDAS trajectory, particularly in relation to online methods. We consider the increasingly impactful interaction between technical, methodological and practical developments and the questions this raises. Technical developments include data digitization, technological normalization and the growth of online tools (both CAQDAS and others). Methodological developments include trends in online methods, mixed methods, social media and visual analysis and the use of bigger data sets. Developments that pose practical challenges include citizen research, collaborative and cross-disciplinary research and commercial engagement with qualitative research.

Several changes are observable in response to these developments, which we discuss using selected examples. First is the increase in data formats acceptable to CAQDAS packages, now ranging beyond textual formats to include visual (still/moving images), bibliographic (imports from Endnote, Refworks, etc.), and database material (including mixed data such as that derived from online surveys, social media and other Internet-harvested data). Second is a move to provide multi-user, server or online versions that facilitate team-working. Third, parallel CAQDAS versions are available for different operating platforms and App versions for tablets. In addition, new software tools are emerging that challenge the position of traditional CAQDAS programs. These developments benefit users by opening up the field, fostering cross-disciplinary and cross-sector collaboration and encouraging the appropriation of new methodologies and technologies. Fourth, CAQDAS packages are increasingly providing features designed to facilitate quantitative approaches to qualitative data and mixed approaches to mixed data as well as the core qualitative tools associated with their initial development. How these developments manifest in individual CAQDAS packages differs such that we are currently witnessing a new divergence between products.¹ Technological and analytic divergence and proliferation rather than convergence and standardisation will thus continue to frame the future of the field, raising pertinent challenges and opportunities for developers, methodologists, users and teachers.

CAQDAS and Trends in Online Data Collection Methods

Integrating multiple data forms within one study, whether textual accounts, non-textual materials, or qualitative and quantitative data, is a common analytic strategy. This section discusses the analytic tactics² required for, and made available by CAQDAS packages, for capturing, integrating and analysing such material, specifically when originally generated online.

Currently, no CAQDAS packages allow for direct³ analysis of 'live' online materials; they must first be 'harvested'. This is true for online applications, software installed on a server and locally on a hard drive. Working with online materials is therefore divided into two phases: data collection (sampling or harvesting) and data analysis. CAQDAS packages currently fall into two groups in this respect: those that address both phases and those that address only the analysis phase.

Some CAQDAS packages have recently developed 'add-ons' or components built for the purpose of capturing online materials. For example, NCapture is a web browser extension that supports NVivo users to work with online materials; enabling capturing of webpages, online PDFs and social media data (Facebook, Twitter and YouTube). There are different options for formatting and importing, including automatic transference of metadata into classifying attributes, options for including posted YouTube comments and Twitter Retweets and choices about the format of harvested material. Content harvested as PDF is highly functional within NVivo. For example, webpage front-end structure and the text in advertisements within the page are searchable and codeable, allowing for analysis beyond user-generated content. Images embedded within PDFs can be coded, although, as is the case in most other CAQDAS packages,⁴ it only allows

for the selection of rectangular areas and retrieval display is limited and somewhat distorted because the coded area is highlighted in colour. Imported material, whether in PDF or database form, retains the source active hyperlinks, allowing navigation from the harvested material back to the web to 'follow' leads and, where relevant, re-harvest altered content. Particularly useful with Twitter data is the option to 'merge matching social media databases' upon import via NCapture. This prevents cumulative harvesting resulting in duplication, thereby attending to challenges presented by working with the ever-evolving nature of online content.

Texifter, a cloud-based platform, has both the harvesting and analysis of online materials at its core. Whilst it is modular in its construction (with DiscoverText, CloudExplorer and FOIA Toolkit working off the SIFTER™ engine), its central premise is to enable processing of voluminous online materials. This includes harvesting and analysing large-scale online archives and merging data from various sources, including text files, email, open-ended answers to surveys, and online sources including Facebook, Google+, blogs, Tumblr, Disqus and Twitter. Timespans during which live web content is captured automatically can be set, as can the number of times within the timespan that capturing (termed 'fetching') takes place. It is possible to set fetches up to 12 months in advance. Additional functions include classification (machine coding), redaction and identification and clustering of duplicates. Data can be filtered by metadata, such as when and from where a tweet was posted and according to socio-demographic information. The full suite of tools therefore allows for both the capturing and the analysis of online materials.

These two developments are examples of what may come to constitute a shift in emphasis in the CAQDAS landscape, enabling researchers to engage in the act of capturing *and* incorporating online materials directly from one product. Amongst the advantages of such single-software solutions are that researchers (1) may be less likely to face a skills deficit and more likely to save time by not having to identify, learn about and download multiple software packages, and (2) that they thus avoid compatibility issues that can occur as proprietary software constantly develops.

Packages that have not developed such tools can be used in conjunction with a range of stand-alone 'middleware' (Fernandes, 2008) in order to harvest online materials. Because CAQDAS packages do not currently allow for direct analysis of live Web material, data collection involves the rendering 'static' of sampled Web material and converting it into a format legible by the chosen software. The technical requirements of harvesting are discussed in Welser *et al.* (2008), who refer to the processes of scraping, parsing and the use of structured query language databases as activities integral to the collection of online data, the technicalities of which are foreign to the vast majority of researchers. Fernandes (2008: 8) highlights further challenges in the collection of online materials, stating that 'although it is tempting to think of the Internet as a vast cornucopia of data riches simply there to be consumed, [...] existing infrastructural constraints obstruct ready and unimpeded access to those riches'. The technical and analytical skills required for both harvesting and analysing web materials may not co-exist within individuals. Human and technological challenges still abound. But what are the options for researchers using CAQDAS packages that have not integrated the capturing of online materials?

Capturing online materials begins with web browsers. Although they each work slightly differently, those most commonly used (Firefox, Chrome, Internet Explorer, Safari) interface with other software to allow harvesting in a variety of formats. Text-based and mixed material is commonly converted to PDF when harvested. This display format retains the visual structure of a webpage, including text and images, and, usually, transformed text remains machine readable. Although most CAQDAS packages can import PDF files, their nature varies depending on how and when they were created and certain limitations of the format affect the analysability of content. For example, conversion includes specific paragraph structures which usually cannot be changed. This affects the reliability with which results of text-search and coding exercises that rely on paragraph structure can be automated (see later), with often unexpected results in terms of the breadth of coding that can be applied. In addition, converting social media content into PDF sometimes limits what is displayed, such as truncating content rather than capturing its entirety. Despite these limitations, harvesting online material using a PDF converter constitutes a bridge to working within a wide variety of CAQDAS packages.

Other formats can also function as a bridge between online material and CAQDAS packages. Text-only (*.txt), Rich Text (*.rtf) and MS Word (*.doc, *.docx) files are flexible and fully functional in all CAQDAS packages, but spreadsheet formats (*.xlsx, *.xls and *.csv) function in different ways. Data can often be imported as the full database, displaying similarly to how they do in spreadsheet or statistical software packages. This can be cumbersome from a display perspective, as well as variously functional because code retrieval sometimes returns the full content of a partially coded cell. More flexible options, provided differently by Atlas.ti, MAXQDA and QDA Miner, convert databases into a series of text documents and categories upon importation, such that qualitative fields are extracted and treated as text and quantitative fields as numeric variables (see Fielding *et al.*, 2013, for in-depth discussion of this functionality).

Despite the emergence of bespoke CAQDAS harvesting tools and options for bridging, there is still a way to go before online material can be harvested in formats that are completely true to their web origins *and* completely functional in all CAQDAS packages.

Collaborative and Online Working with CAQDAS

Recent years have witnessed increased demand for collaborative and online rather than local working arrangements, of which developments such as Dropbox and Google+ are testimony. The operating platform has historically been an issue in the CAQDAS field, with most pioneer programs being initially developed for PCs.⁵ This has changed significantly with most now offering Mac versions, some also Linux and the beginnings of a move to online solutions. For individual researchers, the ability to access materials and undertake analysis online is attractive because it heightens mobility, allowing work from any Internet-enabled location. For research teams, the comparative affordances of working online may be even greater.

Collaborative working comes with the challenge of how to bring together progress made by different team members. With ever-increasing emphasis from funding bodies on large-scale multi-disciplinary multi-national consortia, many researchers find themselves working on bigger projects in larger teams. What are the options

open to teams needing to combine analysis and how is this related to online working?

Technically, there are three ways of working collaboratively using CAQDAS packages, two of which involve online activities (see Hindmarsh, this volume). First, each researcher may work separately on local copies of the project which are merged together at set time points. As this does not involve simultaneous or online working, care has to be taken in splitting and combining work. Most CAQDAS packages allow for the merging of stand-alone projects in this way (e.g. Atlas.ti, HyperRESEARCH, NVivo, MAXQDA, QDA Miner). Second, where multi-user versions exist (e.g. Transana and NVivo), projects are hosted on a networked server, which multiple researchers log into and access simultaneously. Third are online packages (e.g. Dedoose, Discovertext), which do not require local or server-based installation but are accessed by logging on via a web browser. Here we focus on server-based and online multi-user working because they are more recent developments and pertinent to our focus on online methods.

Transana-MU and NVivo for Teams are separate versions of their stand-alone equivalents. They are installed on a networked server and allow projects to be stored centrally and accessed simultaneously by multiple researchers. Their technical set-up is slightly different, as are the features provided for concurrent team-working, but both obviate the need for complex merging protocols to combine the work of individual researchers. Both include a centrally held team-work journal that can be used to detail project development across all contributing researchers. Amongst the specific benefits of Transana-MU are its cost-effectiveness; its availability for both Mac and PC users; the choice to set up one's own server or use the Transana Cloud Service; a chat window enabling real-time communication; and the local storage of media-files that significantly speeds up data transfer. For NVivo for Teams, each researcher needs a license for the stand-alone product and the server version. Central to its set-up is the NVivo for Teams Manager and Project owner(s); the latter granting users access to projects and the former managing the connections between them. At the time of writing, the server version is available for NVivo for Windows, but not NVivo for Mac. Amongst the advantages of using NVivo for Teams are that the project size is not limited to 10GB (as it is with stand-alone projects); and that permissions can be centrally managed, enhancing security and ethics.

Packages that allow online multi-user working include Dedoose and DiscoverText. Dedoose, released in 2009, was developed specifically to enable concurrent online multi-user work, and DiscoverText, available since 2011 as an online platform, allows for synchronous, multi-user working. Dedoose stores projects on the cloud and is designed for all modern Internet browsers. The web-based interface is platform-independent, and thus works equally well for Mac and PC users, and the ability to work concurrently without complicated server installations and set-ups is beneficial. Users can access projects from any web-enabled device, rather than being limited to machines that have specific software installed. Dedoose tracks the actions of different researchers and users can be added to projects at any point. Access works by logging-in using a password. Different levels of access can be given to different users and inter-coder reliability can be measured via code application and code weighting tests.

Although online working facilitates collaboration in various ways, important considerations and challenges remain. Online working leads to dependence on Internet connectivity, which can be particularly problematic

for mobile methods, participatory methods and projects undertaken in certain geographic areas. In addition, uploading data to a cloud or server hosted and maintained by a third party requires researchers to understand security protocols and to ensure they are in line with ethical requirements. In consortia where several institutions are involved, which may have different regulations, this requires planning. However, online applications obviate issues of software version and platform compatibility and the update protocols used by different team-members. Often overlooked, this issue is critical to successful, iterative team working and can cause problems for teams working with stand-alone packages.

Regardless of whether work is conducted using an online application, server version or stand-alone package, the issue of project management is pertinent. To ensure consistency across different researcher's contributions (as well as 'mergeability' where work is conducted locally), the team must develop, set-out and adhere to appropriate coding, annotating and memoing practices. When work progresses inductively, it is critical to develop standardised ways to feed additions back to colleagues. These issues must be planned for and pilot exercises are highly advisable.⁶ In addition is the issue of coding consistency. Most CAQDAS packages now incorporate tools for coder-reliability testing, although these vary in their premise and results. Even when they are not included, there are ways to manually compare. The relevance of the concept of coder-reliability testing may be on the rise given the trend towards bigger data and larger analytic teams. Qualitative research may be diversifying away from a prevalence of small samples towards an inclusion of larger data. With this trend, representativeness, reliability and replicability become more pressing issues. New affordances, including mixed methods, bigger data and automation (discussed later), give rise to new challenges such as larger teams of coders and, potentially, a greater emphasis on more statistically driven assessments of inter-coder reliability.

Opening up Qualitative Analysis

The capabilities of the Internet and increasing digital literacy greatly open up possibilities for citizens to undertake research, enabling wider engagement in practices hitherto the preserve of professionals (Smith *et al.*, this volume; Fielding, 2014). Web-based and non-bespoke tools offer accessible, open-source and cheap alternatives to professional CAQDAS, and are increasingly appropriated for research tasks. These developments are in line with broader normalisation of mobile digital technologies and their rapid developmental pace. Some CAQDAS packages are beginning to respond. Here we discuss the release of mobile app versions by two pioneer CAQDAS developers; the release of Quirkos, a program specifically developed to open-up computer-assisted analysis; and possibilities for using non-bespoke and Web 2.0 tools for analytic purposes.

Currently, app versions are available for Atlas.ti (since 2013 Atlas.ti Mobile for the iPad) and MAXQDA (since 2014 the MAXApp for iOS and Android). Providing mobile accompaniments to their corresponding full versions, initial developmental focus has been on enabling data collection, 'tagging' and note-taking tools. Apps access the image, audio and video capturing features of tablets or smart-phones, enabling

recorded data to be instantly available for 'on-the-go' reflection and initial conceptualisation via touch-screen technology. They also utilise embedded location technology to 'tag' captured data with geographic co-ordinates. Importation into the full package is straightforward and streamlined (via Dropbox or iTunes) when the full range of analytic capabilities are provided. App version capabilities blur data collection and analysis, which are often seen as distinct research phases, making the apps particularly attractive for researchers adopting 'grounded' approaches where these phases are expressly iterative; for researchers undertaking ethnographic studies where many different forms of materials are gathered; and also for mobile methodologies.

Although currently in their infancy, these apps have the potential to revolutionise the way CAQDAS-conversant researchers go about their profession. Similar to the impetus of Quirkos' development (see later), CAQDAS apps offer a means to enable non-professional researchers to contribute to data collection and analysis. Although at the time of writing there are no academically published articles concerning their affordances,⁷ they will no doubt develop considerably in forthcoming years and we can expect to see other CAQDAS packages following suit. There will likely always be limits to the extent to which app versions can fully support varied approaches to analysis as a result of both limited screen real-estate and technological capabilities of mobile technologies. Nevertheless, this constitutes a potential turning-point in the way CAQDAS develops.

Related is the addition in MAXQDA of Emoticons, a development that reflects the rise of 'emojis' (electronic graphic symbols, including 'smileys' that originated in Japan) since being integrated into most mobile technologies. Designed to transcend the limits of text-based coding, MAXQDA's coding system includes more than 300 emoticons and symbols, organised into thematic groups that can be used in combination with, or to the exclusion of, customary textual code-labels. Emoticon functionality is fully integrated into MAXQDA's full- and app-version features, such that symbols appear alongside coded data segments and are visualised in output representations resulting from matrix-based queries and other visual tools. This development offers an additional means of opening up analysis, particularly in citizen, collaborative, cross-national and multi-lingual contexts. For example, equivalence in application and interpretation of codes is a frequent issue in collaborative projects, compounded in cross-cultural studies when analysts may not have shared understandings of concepts. Expressing concepts using words when not all analysts have the same mother-tongue has the potential for confusion for cross-cultural conceptualisation. Symbols can of course also be interpreted and used differently, but as the proliferation of emojis inevitably continues, the potential for more universally understood applications increases. Their integration into Apple technology in 2011 was likely significant to this end, and although different proprietary mobile operators have developed and defined different variants, we might expect standardisation in emoji symbols to increase over time. It remains to be seen whether this development will have a significant impact on the way coding is undertaken within MAXQDA⁸ and whether other CAQDAS developers incorporate similar functionality. However, it is timely in the context of generations of digitally-native researchers that have grown up using emojis in mobile communication, engaging with CAQDAS technologies.

App versions and EmotICODES are examples of ways in which CAQDAS programs are harnessing wider technological developments and offering ways of widening participation in research. A recent development that goes beyond this by specifically focusing on involving non-professionals in every stage of a project is the development of Quirkos. Released in 2014, Quirkos represents a potentially significant step-change in the field, with its central idea being to 'make text analysis so easy, that anyone can do it'. Quirkos includes core qualitative text analysis features characteristic of all CAQDAS packages (Silver and Bulloch, 2015) but its look and feel is rather different. For instance, its layout is designed for tablet usage, encouraging touchscreen actions, and its default settings show non-linear, mind-map-like visualization of 'quirks' (or codes). There are many examples of research participants being involved in research projects; however, most are either at the data collection phase of a project and/or the seeking of feedback as part of interpretive validation. Quirkos is the first bespoke CAQDAS package that specifically aims to bridge the divide between professional practice and lay contribution to involve both parties in all stages of analysis by developing an interface and a suite of features that are as easy to conceptually understand as to technically operate.

A study of the impact of the 2014 referendum for Scottish independence (Turner, 2015) engaged non-specialist researchers to undertake participatory coding in order to explore the ease with which they could use the program for qualitative thematic coding. That these participants found the process easy, quicker than anticipated, interesting and enjoyable (Turner, 2015), suggests that the aim of developing a bespoke CAQDAS program that can be harnessed by non-professional analysts has been achieved. As its features are developed to become more equivalent to other programs, we can expect its use to increase and for reflective publications on its affordances to appear. The extent to which it will open up access to qualitative analysis for those who do not have a background in professional academic or applied research remains to be seen. However, it is anticipated that it will be well utilised for participatory approaches to qualitative research in which research participants can actively contribute to analysis without extensive time needing to be given over to learning the package (Silver and Bulloch, 2015).

A third example of how current changes in the field are opening up qualitative analysis is the development of Web 2.0 and other non-bespoke tools. The term Web 2.0 is used to describe the focus on enabling collaboration and the sharing of information online that is the feature of the second generation of the world-wide-web. Its orientation around users generating online content and analysing their own and others' engagement with it, using tools akin to those professional researchers use in CAQDAS packages to undertake analysis, has potentially profound implications for CAQDAS (di Gregorio, 2010). Although designed for alternative purposes, tools such as blogs, wikis, video/image sharing websites and social networking provide online spaces where researchers can not only communicate their research outputs but undertake forms of analysis. Blogs, in particular, are becoming increasingly omnipresent amongst academic and applied researchers as means of raising awareness about and gathering feedback on their work whilst it is ongoing, as well as disseminating results. This is occurring within a broader context of a move towards open access publishing, which we can expect to further encourage this trend.

Many of the tools available in Web 2.0 applications are analogous to those developed within CAQDAS

packages for bespoke qualitative analytic tasks, such as tagging (akin to coding) and commenting (akin to annotating). In addition are applications such as Onenote and Evernote, designed as general-purpose note-taking tools but which also include tools for generating data (for example, web-capture and clipping, Optical Character Recognition (OCR) capabilities and dictation), storing and managing data (creation of notebooks with any number of sub-levels), searching (word searching within and across notes), memoing (annotating notes with additional comments), coding (tagging of notes and paragraphs within notes), linking (hyperlinking from within notes) and collaboration (sharing of notes and chat functions). CAQDAS packages include a range of features that can be used for more sophisticated analytic purposes, but the growing availability of free, low-cost and/or open source tools that can be appropriated for analytic tasks enables researchers without access to customised CAQDAS packages to undertake analysis.

These developments bring with them challenges for CAQDAS as well as opportunities. The growing availability of a range of digital tools for social scientists (for an overview, see Paulus *et al.*, 2013) offer a plethora of options. Some CAQDAS developers have released free or 'lite' versions, opening up access to their tools to those with resource constraints. Others are increasing the number of proprietary formats that can be imported, including OneNote and Evernote as well as from a range of bibliographic tools (Endnote, RefWorks, Menderley, Zotero). We are therefore seeing a cross-fertilisation of features, with CAQDAS-like features appearing in other software programs and tools available in other applications increasingly appearing in CAQDAS packages.

Mixed Methods, Bigger Data and Automation

The term 'mixed methods' generally refers to the bringing together of some form of qualitative and quantitative data and there is much discussion about design-types for doing so (Creswell and Plano Clark, 2011; Tashakkori and Teddlie, 2010). However, in the context of CAQDAS, 'mixed methods' also relates to ways in which qualitative or quantitative *analyses* can be combined, therefore bringing into focus *how* mixing occurs as well as *what* is being mixed and *when* the mixing occurs, which has received less attention (Silver and Lewins, 2014). What affordances do CAQDAS packages have in terms of carrying out mixed analyses? Bazeley (2006) outlines that the use of software enables different approaches to be undertaken within a single analysis as well as different forms of data to be integrated during analysis in three key ways: (1) by combining text and numeric data, (2) by converting one form to another and (3) by combining and converting iteratively or in generating blended data for further analyses. Here we consider both data and approaches to analysis.

In terms of data, there are two considerations: type and amount. Most CAQDAS packages now accept a range of qualitative (text, audio, still/moving images), quantitative (numeric databases) and mixed data (combined numeric and alphanumeric databases). CAQDAS packages have long enabled the integration of quantitative information with associated qualitative records in the form of, for example, importing demographic or other numeric metadata pertaining to individuals and groups of respondents or other materials and linking them. This enables qualitative data to be interrogated according to those factors, for example comparing

how individuals with certain socio-demographic characteristics report an experience or discuss an attitude. Enabling the incorporation of mixed datasets, for example survey-type data that contains both numeric information from closed-questions and textual material from open-ended questions in the same file, is a more recent development, but most CAQDAS packages now enable it. As such, type of data is usually only a key driver for choosing between CAQDAS packages in certain circumstances, for example when working primarily with visual data.

Although most CAQDAS packages can handle thousands of files, processing speed is usually affected when working with larger datasets. Kitchin defines 'big data' as

huge in volume, consisting of terabytes or petabytes of data; high in velocity, being created in or near real-time; diverse in variety, being structured and unstructured in nature; exhaustive in scope, striving to capture entire populations or systems ($n=all$); fine-grained in resolution and uniquely indexical in identification; relational in nature, containing common fields that enable the conjoining of different data sets; flexible, holding the traits of extensionality (can add new fields easily) and scalability (can expand in size rapidly).

Kitchin, 2014: 262

Of the packages discussed here, and applying this definition, only DiscoverText and Provalis Research's Prosuite (QDA Miner, WordStat and SimStat) come anywhere close to handling 'big data'. Nevertheless, moves towards analysing bigger datasets using CAQDAS packages are observable, which raises questions about approaches to analysis, and thus the tools CAQDAS packages provide to analyse volumes of data that cannot reasonably be read and analysed in-depth by humans.

The CAQDAS landscape is complex in terms of analytic approaches, with two pertinent dimensions: (1) the types of results that are required, i.e. whether qualitative, quantitative or mixed and (2) the means by which to arrive at those results, i.e. the need for assistance with or automation of coding. Just as qualitative and quantitative approaches sit on a continuum with respect to these dimensions, so do CAQDAS packages. At one end, qualitative approaches, such as grounded theory and Interpretive Phenomenological Analysis (IPA), require a true adherence to fully grounded, inductive and iterative ways of working, typically including extensive data familiarisation and pre-coding. Such approaches lend themselves to smaller datasets and prioritize the role of the human interpreter, likely considering automated processes to be anathema. On the other end, approaches such as quantitative content analysis, text mining and data analytics, may require the application of inferential statistics, formal testing of hypotheses that follow the standard requirements of normal distributions, sampling theory, representativeness, etc. Such approaches lend themselves to larger datasets and likely necessitate assistance with, or automation of, qualitative coding.

Researching the social world on, through and in the Internet involves a host of automated processes. The very act of searching the web involves a number of these, as does the fact that browsers are constantly running protocols to determine what you see on the web, where and how. Working in a partially automated

environment, therefore, is second nature to the researcher conducting online research. CAQDAS packages engage with the issue of automation to differing degrees, and can be broadly categorized into three groups in terms of the types of results that can be generated and the means to arrive at them. Some have largely maintained their original qualitative roots, up to this point focusing development primarily on supporting qualitative approaches to qualitative data (e.g. Atlas.ti, Quirkos, Transana). Such packages include features for quantifying qualitative data, such as frequency information pertaining to code-application and co-occurrence, but have less-developed tools for other forms of mixed analysis and tend to include only rudimentary text-searching and auto-coding tools.

Second are packages originally developed as qualitative analysis software but have more recently added quantitative tools (e.g. HyperRESEARCH, MAXQDA, NVivo). Such packages now include a range of (rather different) ways of quantifying qualitative data and analyses of them. For example, MAXQDA's Quote Matrix generates a tabular joint display of qualitative and quantitative output, showing coded segments based on variable characteristics; its Typology Tables visualize variable-values based on previously created qualitative typologies; its Configuration Tables show coding patterns for sets of codes; and its Statistics of sub-codes create frequency tables and editable diagrams for sub-codes. MAXQDA has also recently released MAX Analytics Pro which incorporates the ability to undertake statistical analysis within MAXQDA. In addition, some offer ways to assist with and automate coding. For example, MAXQDA's Dictionary and MAXDictio Coder functions allow users to build categories of words and phrases that are used to search for and code text, and NVivo's pattern-based coding function compares passages of text to those already coded in order to facilitate researchers in identifying whether relevant data have been missed.

Finally are the applications which were specifically developed to provide mixed methods solutions (e.g. Dedoose, QDA Miner (with WordStat and SimStat) and DiscoverText), which currently have the most sophisticated options for conducting mixed analyses, although the emphasis of each is rather different. For example, Dedoose focuses on interactive data visualisations, containing a range of pre-configured charts that represent the intersections between the results of qualitative coding and various forms of quantification that can be explored and filtered in a variety of ways. It includes tools for measuring inter-rater reliability (code application and code weighting tests), but as yet has no functions for automating coding or undertaking statistical analysis. Provalis Research's suite of tools, however, stretches the boundaries of what is possible in CAQDAS packages in terms of mixing analyses because, when used together, they outstrip what is possible in other packages in relation to quantitative, mixed and statistical analysis (Silver, 2014). QDA Miner is Provalis Research's qualitative component with features equivalent to other CAQDAS packages; WordStat is a text-mining component with tools for identifying textual material for coding through the use of user-generated dictionaries; and SimStat offers a range of statistical analysis features. Used in combination, the possibilities for conducting mixed analyses of mixed data are greatly extended because tools for visualizing coding patterns and trends, exploring relationships in coding, and using inferential statistics to test hypotheses can be used together (Silver, 2014). DiscoverText has automation at its heart, enabling the mixing of human and computer training, thus enabling the development of bespoke analytic approaches by users (see Brent, this volume).

Assistance with accessing textual material in the form of word and phrase searching, with or without Key Word In Context (KWIC) functionality and resultant auto-coding, has been available in most CAQDAS packages since their inception. These features have a good fit with the full range of analytic approaches in that they can serve to facilitate data familiarisation, locate passages and narratives and contribute to ensuring all relevant data are captured. As discussed earlier, assistance with and automation of coding on the basis of text patterns is a key area of current CAQDAS development and in the context of big data analytics we can expect this to continue. Assistance and automation has historically presented a concern amongst parts of the qualitative methods community for fear that such features compromise the search for and interpretation of meaning, a task often considered that humans are best equipped to undertake. Packages that 'learn' from the analyst by enabling relevant hits to be prioritized and irrelevant ones dismissed in subsequent searches, strike a balance between automation and human interpretation (see Brent, this volume). They are therefore best placed to harness technological advances in the context of the needs of big data analytics and mixed methods approaches whilst preserving the value of qualitative interpretation.

What is the role of CAQDAS in relation to carrying out mixed methods in an online environment? The volume of material potentially gatherable via online in comparison to offline methods highlights the growing potential for CAQDAS use. Larger sample sizes, even in the context of qualitative materials, are possible and therefore drawing quantitative insights that require more representative samples becomes possible. This allows for meaningful use of descriptive statistics and potentially beyond to inferential statistics. Recent developments in the mixed methods capacities mean CAQDAS packages are a realistic way of connecting quantitative and qualitative insights, not only by enabling the integration of different forms of data, but also enabling the application of different analytic methods to that data. The challenge is that there are still few CAQDAS packages offering tools catering for the extremes of qualitative and quantitative approaches, particularly the latter, requiring researchers to use other tools to undertake statistical analysis. This is an example of where the opportunities provided by online methods are still ahead of the research community. There is a general trend towards more qualitative work in the qualitative/quantitative balance (see also Smith *et al.*, this volume), but although there are differences in methodological skills across countries and disciplines, there may be comparatively few researchers sufficiently skilled in the full range of analysis methods to be able to apply advanced quantitative methods as well as truly inductive qualitative methods.

Discussion: CAQDAS at a Crossroads

The developments discussed here raise two pertinent issues: (1) CAQDAS's role in the increasingly technological and online context of research methods and (2) how new generations of researchers can be equipped to harness them powerfully. These issues place the field at a crossroads; the way developers and researchers respond to the former and teachers respond to the latter will impact upon the future of CAQDAS.

Trends in data usage and analytic methods are changing. Although CAQDAS packages are responding, the use of non-bespoke tools is gaining ground; raising questions about which packages are considered

'CAQDAS' and the extent to which they will continue to maintain their place. Introduced by Fielding and Lee (1991) following the first conference bringing together the fields' methodological and technical pioneers,⁹ the term 'CAQDAS' emphasizes that software *assists* rather than *carries out* analysis. Nevertheless, critiques from those resistant to CAQDAS use have historically coalesced around concerns about software 'taking over' the interpretive processes involved in qualitative analysis, that, it is argued, only humans can do. Concerns involve contentions that software distances analysts from data; perceptions that the tools prioritise code-and-retrieve methods and thus homogenise analytic approaches; suggestions that their use mechanises analytic techniques and promotes quantification to the extent that the interpretivist foundations of qualitative research are undermined; criticisms about feature availability and misconceptions that CAQDAS packages only support certain types of analysis (Bazeley, 2006). CAQDAS users and teachers have countered such concerns, arguing that the analyst remains in control of the analytic process and that CAQDAS packages are tools that can be harnessed for various purposes (Silver and Lewins, 2014; Rivers and Bulloch, 2011). Coding-assistance and text-mining, however, are designed to contribute to the analytic process and their increasing inclusion in CAQDAS not only contributes to product divergence, but further blurs the parameters of what constitutes a CAQDAS package.

Given the roots of the acronym, it is logical to refer to the CADQAS Networking Project's definition, which states that CAQDAS refers to

packages which include tools designed to facilitate a qualitative approach to qualitative data. [...They] 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 them: Content searching tools; Linking tools; Coding tools; Query tools; Writing and annotation tools; Mapping or networking tools.¹⁰

This definition is broad enough to include all of the packages discussed here, and more besides.¹¹ However, the increasing move to include tools for integrating data types, analytic approaches and options for automation demands consideration as to whether the emphasis on qualitative data and qualitative analytic approaches remains accurate in terminological terms and desirable in analytic terms. Is it time to adjust the acronym to remove the explicit emphasis on the qualitative?

One danger is emphasizing mixed methods, and automation may further alienate purist qualitative researchers' resistance to software use. Although the software's existence neither necessitates their use nor excludes the ability or validity of working in purely qualitative ways, as rapid technological development continues the possibilities afforded by machine learning technology will inevitably be increasingly incorporated. The extent to which CAQDAS users harness the potential for increased analytic assistance or resist these developments remains to be seen. The rise of the use of bigger data, that by necessity requires automation, can only mean this issue becomes increasingly pressing.

The potential effect of concurrent development and penetration of non-bespoke analytic tools is also important. CAQDAS packages have developed significantly since their emergence and now have a firm place in professional research practice. The Internet enables bigger data-sets to be harvested, and CAQDAS enables bigger data-sets to be integrated and aspects of analysis to be automated. However, CAQDAS packages are not the only applications available in our arsenal of tools to understand the online world, and they will therefore need to keep up with the pace of technological, methodological and practical developments to remain at the cutting-edge.

So what do researchers want CAQDAS to be and do? As some packages move towards increasing diversification of tools, seemingly attempting to 'be all things to all people', others focus on specific analytic affordances, user-groups or solution-types. Do we want one package that does everything or do we prefer to utilize several applications, thus constructing our own bespoke suite of tools for particular purposes? No doubt developers and researchers have different opinions, but what about new generations of researchers?

Contemporary students have different expectations of technology than their predecessors because they have never really known a time without computers, the Internet or mobile devices (the exception being mature students). For the digitally conversant, it is obvious that technology can be used to facilitate analysis, and therefore debates must move beyond *whether* to use CAQDAS to detailed discussion about *how* to do so powerfully. Equipping students with the skills to harness technology appropriately for analytic tasks can no longer be side-stepped (Silver and Rivers, 2016). Research into how CAQDAS is taught is relatively scarce but continued demand for workshop-based learning illustrates that non-bespoke tools are not filling the gap. In addition, demand for training outside of degree programmes illustrates that learning about the technological tactics for undertaking analysis is frequently separated from learning about analytic strategies (Silver and Woolf, 2015; Carvajal, 2002; Gilbert *et al*, 2014).

Facilitating a step-change in CAQDAS teaching requires an accessible and adaptable pedagogy that transcends the specificities of products, applications and modes of instruction (Silver and Woolf, 2015). Detailed accounts of specific modules are valuable, and yet it is all too easy for lecturers to dismiss such accounts as irrelevant to their own contexts as a result of disciplinary, methodological or logistical differences. Some present a 'one-size-fits-all' method for CAQDAS use; however, this is inappropriate because of the variety in objectives, methodologies and analytic procedures. It is not surprising that researchers are resistant to using and lecturers are resistant to teaching CAQDAS when it is presented in a way that homogenizes the variety that is prized in qualitative and mixed methods research.

The rise of online methods adds a layer of complexity to the appropriate teaching and harnessing of CAQDAS packages because contemporary users seek flexible applications that support the whole research process, from problem formulation, through data collection, analysis and representation. CAQDAS packages have a huge potential in this regard, as illustrated by Schmieder's (2015) modularized curriculum for using CAQDAS packages for interview question analysis, Paulus and Bennet's (2015) discussion of integrating CAQDAS into a graduate research methods course and Silver and Woolf's (2015) exposition of Five-Level QDA as a CAQDAS pedagogy.

Although some CAQDAS packages are moving towards enabling the harvesting of online data and providing tools to facilitate a broader range of analytic approaches, the latter phases of the research process are less well supported. For example, there is a lack of flexibility in transforming the results of analytic work undertaken in CAQDAS packages into dynamic and accessible formats, in particular for generating non-linear representations, such as visual narratives, hypertexts or reportage (Silver and Patashnick, 2011). Although visualization tools are being developed in some CAQDAS packages, they are basic in comparison to what can be generated from big data analytics. In the context of online methods, these limitations are significant because those conducting research on, through or in the Internet require dynamic options for representation that reflect the multidimensionality of the Internet and data.

Conclusion

This chapter has provided an overview of the current state-of-the-art with regards customized CAQDAS packages in the context of online methods. It has detailed some of the significant moves towards providing effective solutions for harvesting online data, highlighting the divergence in practice with regards to software that is developing bespoke harvesting tools and that relies on researchers making use of non-CAQDAS tools for the harvesting phase.

The discussion of collaborative and online working touched on the recent advances in multi-platform working, as well as the significant opportunities and challenges posed by the proliferation in server- and online, multi-user-based working solutions across various packages.

Making the case that the field is seeing an opening up of analysis beyond the academic professions through the proliferation of web-based and non-bespoke tools that offer accessible, open-source and cheap alternatives to professional CAQDAS, the chapter also raised the question of the distinctiveness of CAQDAS packages in this fast-evolving space.

CAQDAS has responded to the mushrooming of material about the social world that is generated online by increasingly enabling the analysis of bigger data, automation, the integration of different forms of data, and also the application of different analytic methods to that data. What this means for the traditional identification of these packages with primary *qualitative* approaches remains to be seen.

These various discussions paint a picture of the field of CAQDAS at a crossroads. In order to maintain their place in professional practice and be adopted by non-professionals, software developers need to attend to the potential for CAQDAS packages to be utilized for all stages of the research process and to continue to respond to developments occurring in the sphere of online methods. This is likely to involve attending to the full triad of technical, methodological and practical concerns that researchers face today and may well result in increasing divergence in products.

Notes

1. See the CAQDAS Networking Project website for software reviews <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/support/choosing/index.htm> Accessed 1st October, 2015.
2. In referring to 'analytic strategies' and 'analytic tactics', we apply the definitions used by Woolf (2014) and Silver and Woolf (2015), in which strategies relate to *what you plan to do* in an analysis and are, in varying degrees, iterative and emergent in qualitative data analysis (QDA); and tactics relate to *how you plan to do it*, which when using CAQDAS packages comprise the use of cut-and-dried, pre-determined software tools.
3. Drawing on Silver and Lewins (2014), we distinguish between working 'directly' and 'indirectly' with source data. Analytic activities that are examples of direct working include annotating and coding parts of a data source. In contrast, activities that constitute indirect working include writing critical reflections about data sources.
4. Transana is an exception here, which has recently developed much more refined tools for coding and annotating still images (see Silver and Lewins, 2014).
5. Notable exceptions are HyperRESEARCH and Transana.
6. The CAQDAS Networking Project website provides generic and specific team-working protocols for merging separate CAQDAS projects when working with stand-alone versions (see <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/support/teamworking/index.htm>) Accessed 1st October, 2015.
7. See the ATLAS.ti blog (http://blog.atlasti.com/?_ga=1.181453735.1265748858.1432378242) and the MAXQDA blog (<http://www.maxqda.com/blog>) for informal accounts of the use of mobile apps by researchers. Accessed 15th October 2015.
8. At the time of writing there are no published articles reporting studies that have used MAXQDA emoticodes, but developers report interest in the functionality amongst their users and discussion of their potential was a feature at the 2014 CAQD conference, which brought together MAXQDA developers, trainers and users.
9. The Surrey Research Methods conference was held at the University of Surrey, UK in 1989 and resulted in Fielding and Lee (1991) and the setting up of the CAQDAS Networking Project in 1994.
10. http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/support/choosing/caqdas_definition.htm Accessed 15th October, 2015.
11. It is also relevant to the term Qualitative Data Analysis Software (QDAS), which is sometimes used to refer to this group of software packages.

References

- Bazeley, P.** (2006) 'The contribution of computer software to integrating qualitative and quantitative data and analyses', *Research in the Schools*, 13(1): 64–74.
- Carvajal, D.** (2002, revised 2007). The Artisan's Tools. Critical Issues When Teaching and Learning CAQDAS [47 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(2), Art. 14, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0202147>. (Accessed 15th October 2015)
- Creswell, J. W.**, and **Plano Clark, V. L.** (2011) *Designing and Conducting Mixed Methods Research*. (2nd ed). Los Angeles, CA: Sage Publications.
- di Gregorio, S.** (2010) 'Using Web 2.0 tools for qualitative analysis: an exploration'. Presented at the International Conference on System Sciences, 5–10 January, 2010, Hawaii.
- Fernandes, A. A. A.** (2008) 'Middleware for distributed data management'. In **N. G. Fielding, R. M. Lee** and **G. Blank** (eds.), *The SAGE Handbook of Online Research Methods*. London: Sage Publications, pp. 97–115.
- Fielding, N. G.** (2014) 'Qualitative research and our digital futures', *Qualitative Inquiry*, 20 (9): 1064–73.
- Fielding, N.** and **Lee, R. M.** (1991) *Using Computers in Qualitative Research*. London: Sage Publications.
- Fielding, N.** and **Lee, R. M.** (2007) 'Honouring the past, scoping the future'. Plenary paper presented at CAQDAS 2007: Advances in Qualitative Computing Conference, Royal Holloway, University of London, 18–20 April.
- Fielding, J., Fielding, N.** and **Hughes, G.** (2013) 'Opening up open-ended survey data using qualitative software', *Quality and Quantity*, 47(6).
- Gibbs, G.** (2014) 'Using software in qualitative research'. In **U. Flick** (ed.), *The Sage Handbook of Qualitative Data Analysis*. Los Angeles, CA: Sage Publications, pp. 277–94.
- Gilbert, L. S., Jackson, K.** and **di Gregorio, S.** (2014) 'Tools for analysing qualitative data: the history and relevance of qualitative data analysis software'. In **J.M Spector et al.** (eds.), *Handbook of Research on Educational Communications and Technology*. Springer Science and Business Media. New York. pp. 231–6
- Kitchin, R.** (2014) 'Big Data, new epistemologies and paradigm shifts', *Big Data and Society*, 1(1).
- Paulus, T. M.** and **Bennett, A. M.** (2015) 'I have a love–hate relationship with ATLAS.ti™: integrating qualitative data analysis software into a graduate research methods course'. *International Journal of Research and Method in Education*, 1–17.
- Paulus, T. M., Lester, J.** and **Dempster, P.** (2014) *Digital Tools for Qualitative Research*. 1st edn. Thousand

Oaks, CA: Sage Publications.

Rivers, C. and Bulloch, S. L. (2011) 'CAQDAS – A Contributor to Social Scientific Knowledge?' *NCRM Methods NEWS*, Spring: 2–3.

Schmieder, C. (2015) Interview question generation and qualitative data analysis software. A modularized curriculum. University of Wisconsin-Madison. Available at http://static1.squarespace.com/static/51e93258e4b09e55a0347903/t/557a598de4b03a9a1065c4f4/1434081677506/Curriculum_Q_Generation_June2015.pdf (Accessed 15th October 2015.)

Silver, C. (2014) Software Review: QDA Miner (with WordStat and SimStat) *Journal of Mixed Methods Research*, 9(4):386–7.

Silver, C. and Bulloch, S. L. (2015) 'QUIRKOS 1: distinguishing features and functions', The CAQDAS Networking Project. Available at <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/files/Quirkos%20-%20distinguishing%20features-July2015.pdf> (accessed 10th October 2015).

Silver, C. and Lewins, A. (2013) 'Computer Assisted Analysis of Qualitative Research: trends, potentials and cautions'. In **P. Leavy** (ed.), *The Oxford Handbook of Qualitative Research Methods*. Oxford: Oxford University Press, pp. 606–38.

Silver, C. and Lewins, A. (2014) *Using Software in Qualitative Research: A Step-By-Step Guide*. 2nd edn. London: Sage Publications.

Silver, C. and Patashnick, J. (2011) Finding Fidelity: Advancing Audiovisual Analysis Using Software [88 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 37, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101372>. Accessed 10th October 2015.

Silver, C. and Rivers, C. (2016) 'The CAQDAS Postgraduate Learning Model: An interplay between methodological awareness, analytic adeptness and technological proficiency', *International Journal of Social Research Methodology*, 19(5): 593–609.

Silver, C. and Woolf, N. (2015) 'From guided-instruction to facilitation of learning: the development of Five-level QDA as a CAQDAS pedagogy that explicates the practices of expert users', *International Journal of Social Research Methodology* 8(5):527–43.

Tashakkori, A. and Teddlie, C. (eds.) (2010) *Sage Handbook of Mixed Methods in Social and Behavioral Research*. 2nd edn. Los Angeles, CA: Sage Publications.

Tesch, R. (1990) *Qualitative Research: Analysis Types and Software Tools*. New York, NY: Falmer Press.

Turner, D. (2015) *Overview of a Qualitative Study on the Impact of the 2014 Referendum for Scottish Independence in Edinburgh, and Views of the Political Process*. Edinburgh: Quirkos.

Welser, H. T. and Smith, M. and Fisher, D. and Gleave, E. (2008) '*Distilling digital traces: computational social science approaches to studying the Internet*'. In **N. G. Fielding, R. M. Lee and G. Blank** (eds.), *The Sage Handbook of Online Research Methods*. London: Sage Publications, pp. 116–140.

White, M. J., Judd, M. D. and Poliandri, S. (2012) 'Illumination with a dim bulb? What do social scientists learn by employing qualitative data analysis software in the service of multimethod design?', *Sociological Methodology*, 42(1): 43–76.

Woods, M., Paulus, T., Atkins, D. P. and Macklin, R. (2015) 'Advancing qualitative research using Qualitative Data Analysis Software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994–2013', *Social Science Computer Review*.

Woolf, N. (2014) 'Analytic strategies and analytic tactics'. In **S. Friese and G. Ringmayr** (eds.), *ATLAS.ti User Conference 2013: Fostering Dialog on Qualitative Methods*. Berlin: University Press, Technical University Berlin, pp. 1–3.

<http://dx.doi.org/10.4135/9781473957992.n27>