Will Qualitative Data Analysis Software (QDAS) Keep Apace of Bigger Data Trends? Van Diemen's Land Convicts Bring NVivo to Trial

Kim Shaw

(University of Tasmania)

Abstract: *The introduction of Qualitative Data Analysis Software (QDAS) programs promised features that would enable analysis of larger volumes of data within a platform that supports and enables portable and mobile research practices. Reporting on a tailored engagement with the QDAS program NVivo, to manage a large historical dataset describing the bodily marks and features of convict men and women transported to Van Diemen's Land in the years 1803 to 1853, this article questions the capacity of NVivo to continue to offer these touted QDAS benefits in the face of a growing research trend engaging with increasingly larger volumes of data.*

Keywords: *qualitative data analysis software, QDAS, NVivo, data, big data, digital tools, qualitative research, history, convicts, Van Diemen's Land, Tasmania*

The inventive use of digital tools by qualitative researchers is on the rise. Examples include Davidson's Journal Project, where Qualitative Data Analysis Software (QDAS) are used to order and analyse personal writings, informing an art-based research with fibre-art outputs,[1] and Le Blanc's deployment of QDAS to create a "postmodern pastiche" challenging traditional narrative and research practices to demonstrate how such tools might "reintroduce" an element of "indeterminacy" and "ambiguity" to academic pursuits.[2] The following report focuses on a bespoke use of the QDAS program NVivo to address the conundrum of how to manage and become familiar with a dataset of approximately 40,000 archival records describing the bodily marks and features of convict men and women transported to Van Diemen's Land from 1803 to 1853.

This article follows others attempting to redress the lack of detail in both research papers and dissertations purporting the use of QDAS noted by Paulus, Woods, Atkins and Macklin,[3] and Jackson.[4] As Humble observes, researchers wanting to describe their use of QDAS programs in

---

[1] Judith Davidson, "The Journal Project: Research at the Boundaries Between Social Sciences and the Arts," *Qualitative Inquiry* 18, no. 1 (2012).

[2] Amana Marie Le Blanc, "Disruptive Meaning-Making: Qualitative Data Analysis Software and Postmodern Pastiche," *Qualitative Inquiry* 23, no. 10 (2017): 796.

[3] Megan Woods et al., "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* 34, no. 5 (2016).

[4] Kristi Jackson, "Where Qualitative Researchers and Technologies Meet: Lessons From Interactive Digital Art," *Qualitative Inquiry* 23, no. 10 (2017). See also, Austin G Oswald, "Improving outcomes with Qualitative Data Analysis Software: A reflective journey," *Qualitative Social Work* (2017); Cynthia S. Robins and Karla Eisen, "Strategies for the

any detail often find they need to publish in journals with a specific focus on qualitative methods as opposed to the key journals of their respective discipline, due to "page constraints" and a requirement to keep methods sections short.[5] Paulus et al. describe encountering a similar sentiment amongst authors.[6] This situation, it must be imagined, does little to support or encourage the broader application of digital tools across disciplines. Paulus, Lester and Britt have observed how "graduate student demand for informed conversation around the use of technology is growing".[7] Such conversations prove difficult to navigate, however, when discipline-specific approaches limit or encourage the specific use of particular tools that may actually have potential to be incorporated into a broad range of projects in increasingly unique ways. As Davidson, Paulus and Jackson write, "Creative combining of digital tools with different capacities may be an important option for getting what we want and need as researchers",[8] and as Salmona and Kaczynski observe, "Doctoral candidates are very much a part of this technological journey of change in qualitative research practice".[9] The current study and its quest to enlist NVivo to chart new waters in the navigation of a large historical dataset is one such example in this technological journey of change.

In ideal circumstances, digital tools are chosen with a clear understanding of their capabilities and suitability for the project in which they are to be deployed. The reality, however, is that software choices are often influenced by a range of factors:

> Often researchers simply use the program that is best known or most used in their geographical region, university, professional field, or academic discipline; that is available or supported at their institution; that is most successfully marketed by its originators; or for which training has been offered at an attended conference. These and other factors make it challenging to respond intentionally to ever-changing technological advancements.[10]

While there is some degree of intentionality in the choice of NVivo for the current project it would also be fair to acknowledge the role of determining factors including institutional investment in the software and available training. It may also be worth noting that an awareness of NVivo, and its most common use as a tool for managing and assisting the analysis of survey data, came from experience as a Research Librarian and not through any prior exposure in the Humanities. This gleaned insight initiated the inkling NVivo might be able to assist in the task of organising and becoming familiar with the unique project data at hand. As initial explorations of NVivo's suitability occurred early in the research process, there was never any stated commitment or pressure

---

Effective Use of NVivo in a Large-Scale Study: Qualitative Analysis and the Repeal of *Don't Ask, Don't Tell*," *Qualitative Inquiry* 23, no. 10 (2017).

[5] Áine M. Humble, "Qualitative Data Analysis Software: A Call for Understanding, Detail, Intentionality, and Thoughtfulness," *Journal of Family Theory & Review* 4, no. 2 (2012): 128.

[6] Trena Paulus et al., "The discourse of QDAS: reporting practices of ATLAS.ti and NVivo users with implications for best practices," *International Journal of Social Research Methodology* 20, no. 1 (2017): 44.

[7] Trena M. Paulus, Jessica Nina Lester, and Virginia G. Britt, "Constructing Hopes and Fears Around Technology: A Discourse Analysis of Introductory Qualitative Research Texts," *Qualitative Inquiry* 19, no. 9 (2013): 649.

[8] Judith Davidson, Trena Paulus, and Kristi Jackson, "Speculating on the Future of Digital Tools for Qualitative Research," *Qualitative Inquiry* 22, no. 7 (2016): 608.

[9] Michelle Salmona and Dan Kaczynski, "Don't Blame the Software: Using Qualitative Data Analysis Software Successfully in Doctoral Research," *Forum: Qualitative Social Research* 17, no. 3 (2016): para. 20.

[10] Humble, "Qualitative Data Analysis Software," 130.

to persist with the program if it did not perform as anticipated or offer anything tangibly enhancing to the research project.[11]

The following report details the way in which NVivo has been trialled and adopted as a tool, with some noted limitations, in the management and exploration of a unique dataset that well exceeds, by number of records (cases), the usual demands made on the software by more traditional forms of survey data. The implications for qualitative research, of a trend towards increasingly larger datasets, is considered in the light of the purported advantages of QDAS programs, developed to make "the analysis of large volumes of data more manageable and transparent".[12] As the NUD*IST and later NVivo designers remark in 1987, their software and the system they developed may be used "as a super filing cabinet, adapting methods of clerical sorting and classification to the new freedom from constraints on volume, complexity and detail of classification".[13] While successive versions of QDAS have, undoubtedly, enabled engagement with larger quantities of data and an increasing complexity in the scope of its management, it is argued there is a pressing need for QDAS developers to find ways for software applications to adapt and accommodate a current wave of research engaging with even larger volumes of data than previously conceived.

The Data

The data for the current project originates from handwritten nineteenth-century convict records which have been transcribed and added to the Founders and Survivors database.[14] Initially conceived in 2007, the Founders and Survivors project engages a multidisciplinary team of researchers with the stated intention of "building a longitudinal study of a key founder population of modern Australia: Tasmanian convicts and their descendants".[15] Over 72,000 convicts were transported to Tasmania between the years 1803 and 1853 (at the time, known as the British colony of Van Diemen's Land).[16] These men and women "were the subjects of intense documentation" and the records pertaining to them "are the most detailed nineteenth century source anywhere in the world for family history".[17] In acknowledgment of their "outstanding universal value" and "world significance", the Australian convict records were added to the UNESCO Memory of the World Register in 2007.[18]

Bearman has argued that in recent years there has been something of an "archival revolution".[19] As a result of the availability of digital technologies, "historians have been quietly building massive

---

[11] At the time of writing this report there were no QDAS options I could uncover that would be considered better for managing and processing large volumes of data.

[12] Rudolf R. Sinkovics and Eva A. Alfoldi, "Progressive Focusing and Trustworthiness in Qualitative Research," *Management International Review* 52, no. 6 (2012): 839.

[13] Lyn Richards and Tom Richards, "Qualitative Data Analysis: Can Computers Do It?," *The Australian and New Zealand Journal of Sociology* 23, no. 1 (1987): 31.

[14] https://foundersandsurvivors.com/

[15] James Bradley et al., "Research Note: The Founders and Survivors Project," *The History of the Family* 15, no. 4 (2010): 467.

[16] Alison Alexander, *Tasmania's Convicts: How Felons Built a Free Society* (Crows Nest: Allen & Unwin, 2014).

[17] Bradley et al., "Research Note," 467-68.

[18] Bradley et al., "Research Note," 468; UNESCO, "Memory of the World: The Convict Records of Australia," (United Nations Educational, Scientific and Cultural Organization, 2017).

[19] Peter Bearman, "Big Data and Historical Social Science," *Big Data & Society* 2, no. 2 (2015): 1.

archival data structures from the extant records of crucially important institutions and contexts".[20] A recent assessment of the Founders and Survivors database reports the project holding more than "1.3 million linked records pertaining to individuals who lived in Tasmania in the years to 1924", with plans apace to rapidly grow this number.[21] Perhaps more important than debates as to whether this kind of "administrative data"[22] might or might not be included under the still-to-be-settled label of "Big Data"[23] is the recognition that such projects form a component of a contemporary research revolution some have gone as far as to suggest may in fact be a Kuhnian "scientific revolution in the making".[24] The issue is important, primarily because debates and discussion around big data are the current site of reflections concerning contemporary approaches and attitudes towards data in general.

For qualitative researchers this has resulted in a call to reiterate the checks and added insights their respective disciplines might offer over a rising fear of domination by a quantitative turn. For example, Barnes provocatively declares, "It might be big data, but it is little history", while noting – "Correlation does not answer the big questions that social science most wants answering: why do things happen?".[25] Echoing Barns, Halford and Savage write, "Big data analytics needs sociology (as well as the other social sciences) to provide theoretical, methodological and empirical expertise to study the social world",[26] while Graham and Shelton argue for a "sustained inquiry into the question of whose interest big data practices and the big data meme ultimately serves".[27]

Most important perhaps, is the recognition that "no data, big or small, can be interpreted without an understanding of the process that generated them".[28] Interestingly, this observation links with early concerns around the use of QDAS programs creating a distance between researchers and their data,[29] and a worry that "raw data would not be revisited once the technology was used".[30] In the current project, which draws on handwritten nineteenth-century convict records, the need to view images of these original documents to check and clarify transcriptions is ever present. As Maxwell-Stewart explains:

> If 'Founders and Survivors' represents a departure from past quantitative explorations of convict transportation, it is through attempts to ensure that electronic transcripts and coding systems are linked to digitised images of the content from which they were derived.[31]

---

[20] Bearman, "Big Data," 1.

[21] Hamish Maxwell-Stewart, "Big Data and Australian History," *Australian Historical Studies* 47, no. 3 (2016): 362.

[22] Rob Kitchin, "Big Data and Official Statistics: Opportunities, Challenges and Risks," *Statistical Journal of the International Association of Official Statistics* 31, no. 3 (2015).

[23] Rob Kitchin and Gavin McArdle, "What makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets," *Big Data & Society* 3, no. 1 (2016).

[24] Patricia Bazeley, *Integrating analyses in mixed methods research* (London: SAGE, 2018), 176.

[25] Trevor J Barnes, "Big Data, Little History," *Dialogues in Human Geography* 3, no. 3 (2013): 298, 300.

[26] Susan Halford and Mike Savage, "Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research," *Sociology* 51, no. 6 (2017): 1145.

[27] Mark Graham and Taylor Shelton, "Geography and the Future of Big Data, Big Data and the Future of Geography," *Dialogues in Human Geography* 3 (2013): 259.

[28] Ryan Shaw, "Big Data and Reality," *Big Data & Society* 2 (2015): 1.

[29] Erika Goble et al., "Habits of Mind and the Split-Mind Effect: When Computer-Assisted Qualitative Data Analysis Software is Used in Phenomenological Research," *Forum: Qualitative Social Research* 13, no. 2 (2012).

[30] Humble, "Qualitative Data Analysis Software," 132.

[31] Maxwell-Stewart, "Big Data and Australian History," 362.

Consider, for example, the qualitative difference, in Figure 1, between the original record when compared with the same words transcribed as digital text. Only the image makes it clear that "Idiot" and "Died June 1827" were written at different times. Not being able to see the words, "Died June 1827", were added at a later time has the potential to significantly alter the meaning extracted from the text alone.
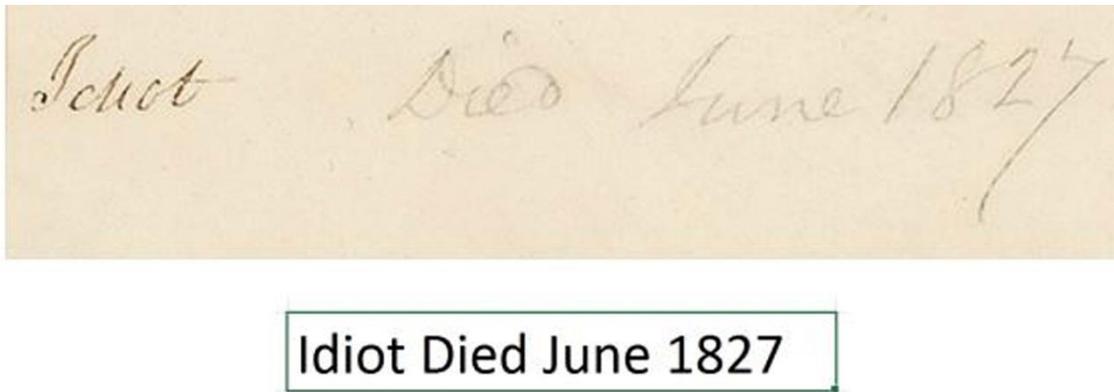
Figure 1. TAHO: CON 23

For an NVivo project, the size of the dataset engaged by the current study is decidedly large. To provide a comparison, Robins and Eisen report on their use of NVivo for a large-scale study - analysing data gathered prior to the repeal of the 1994 U.S. law prohibiting "openly gay individuals" serving in the army - and note how NVivo gave them "the firepower to code 4,000 open-ended comments", at the same time observing this is "a number practically unheard of in the world of qualitative research".[32] The open-ended comments detailing aspects of a convict's appearance that form the focus of the present project total 39,340 of which 10,864 cases refer to female convicts and 28,476 refer to male convicts. These transcribed comments, along with other identifying data including gender and a unique convict ID, have been exported from the Founders and Survivors database. An Excel spreadsheet has been prepared for importation to NVivo where each comment referencing a convict's appearance has been assigned a row under a designated "Marks" column and two further columns for the identifying data, gender and convict ID, have also been added.

How the Data Category (Node) System was Developed

The Convict Descriptions dataset is imported to NVivo 11 via the Survey Import Wizard. The approximate processing time for this step is around 20 minutes. Through this process 39,340 cases are being created – that is, one case for each of the 39,340 convicts and their records - so the lengthy wait time is not entirely surprising. In NVivo a case is a "core structural element" that "unites all the different components of qualitative and quantitative data you have about that entity,

---

[32] Robins and Eisen, "Strategies for the Effective," 777.

that unit of analysis in one place".[33] In this instance, a case would include all information related to a specific individual.

Salmona and Kaczynski note preliminary engagement with QDAS often begins with the question: "What does this software do that I can't do manually?".[34] In the case of the current study, it has been decided at the outset that classifying / coding 39,340 records, individually, by hand, is not a viable task or an efficient use of time. Experimenting with the Word Frequency Query function within NVivo suggested the possibility for a more automated approach that would still maintain the desired user integrity over decisions relating to classification and coding categories and rules.

Running a Word Frequency Query on the descriptive text, searching all words with a minimum length of one character, yielded a total of 4891 unique words and characters. While this is a large number, it is still considerably less than 39,340. The results from the Word Frequency Query have been exported from NVivo as an Excel spreadsheet. The plan is for this spreadsheet – containing a complete list of all the words and characters occurring in the field detailing appearance, as well as the frequency with which they occur – to become the master reference point for ascertaining all the possible categories and variations of marks on the convict body that have been recorded. Figure 2 is a word cloud created from the Word Frequency Query results within NVivo. The most frequently occurring words are the most prominent – in this case, "left", "arm", and "scar".

Figure 2. Word cloud

As handwritten documents are particularly prone to casting a range of variant spellings, misspelling and an array of individual shorthand styles for the expression of a single concept, a manual examination of the text under scrutiny is essential and unavoidable. This is where the master spreadsheet, obtained via NVivo's Word Frequency Query function, is especially useful. One example,

---

[33] Patricia Bazeley and Kristi Jackson, *Qualitative Data Analysis with NVivo*, 2 ed. (London: SAGE, 2013), 52.
[34] Salmona and Kaczynski, "Don't Blame the Software," para. 44.

attempting to identify all convicts with marks referencing potential smallpox infection, uncovered no less than 49 variant ways of noting "pockmarks" (see Table 1 for a list of these). Comparable issues with automated searching that highlight the need to closely scrutinize and become familiar with the data under analysis may also be witnessed in more traditional applications using NVivo. For example, Zamawe writes:

> in my study I wanted to find out if distance is an important factor to ANC attendance, so I searched for the word 'distance'. I got a few results which I coded to a node. But when I went through the transcripts in detail, I realised that other words like 'the hospital is very far', which also implies distance, were missed by the query search.[35]

Recognising "that data coded by QDAS is in many ways similar to data coded by hand may help alleviate concerns that QDAS is automating analysis in ways that take control away from the researcher".[36]

smallpox	pocpitted	pockpits	pockmark	pmarks
pscar	pocpd	pockpit	pockmakrs	pmarked
pptted	pocks	pockpiited	pockm	pmark
ppockd	pockptted	pockpd	pocked	pmar
ppitted	pockpttd	pockp	pockd	pm
ppited	pockpitts	pockmrk	pock	pitted
ppit	pockpitted-scaronrt	pockmd	pocikpitted	pcokmark
pox	pockpitted	pockmarks	pocd	p mark
posckpitted	pockpitte	pockmarked	poc	p marks
pocx	pockpitt	pockmarkd	po	

Table 1. 49 variant ways of noting 'pockmarks'.

The prospect of conducting further research, that may extend to collaborations with other researchers expressing an interest in the oral and dental health of convicts, led the decision to focus on all references relating to "teeth" in this phase of working with NVivo. The test run helped clarify the process for creating the data category (node system) this project will adhere to. The concept of nodes can be confusing and it may help to note that in NVivo "the term node is actually a synonym for what other QDAS packages call codes" and therefore not equivalent to the AT-LAS.ti conception of nodes denoting "objects in a network", representing "the centre points of relations coming together".[37] Assessing the results of the Word Frequency Count, it was determined there are five distinct words referring to teeth, including several misspellings. NVivo's Text Search Query function is next engaged to search for each of the five terms of interest to the present study. Each term will provide the basis for a single node. These will work as holding areas in NVivo for references to coded text. Each node will thus function as a storage container for everything

---

[35] F. C. Zamawe, "The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections," *Malawi Medical Journal* 27, no. 1 (2015): 15.
[36] Paulus et al., "The discourse of QDAS," 41.
[37] Jeanine C. Evers, "Current Issues in Qualitative Data Analysis Software (QDAS): A User and Developer Perspective," *The Qualitative Report* 23, no. 13 (2018): 64.

recorded in relation to an identified category.[38] The "Marks" node, created when the dataset was imported and containing all references belonging to the Marks field, is the source for each text search query. Thus, this deductive "Marks" node heads the inductive node system that follows. The results from each query are added to the project and saved under "Queries". The five text search results are also used to create five new nodes – "teeth", "tetth", "theeth", "tooth" and "tteth" (Figure 3). One additional node, named "Merged teeth references", is added and contains all references of interest with any duplication removed.

The merged teeth references are then sorted into four nodes which cover the variety of the types of teeth reference contained in the descriptions (Figure 3). These include "Broken Teeth", "Lost Teeth", "Teeth Appearance" and "Teeth Decay". These four nodes are developed using a combination of NVivo's Text Search Query function and manual coding of references within NVivo. For example, a Text Search Query for a key term such as "lost" is performed on the "Merged teeth references". Other references denoting lost teeth – e.g. missing teeth – not picked up in this search will be manually coded from the remaining references after all the Text Search Queries for all key terms has been performed. Additionally, all references assigned to a node via a Text Search Query are visually assessed and manually recoded if the allocation is deemed inappropriate.
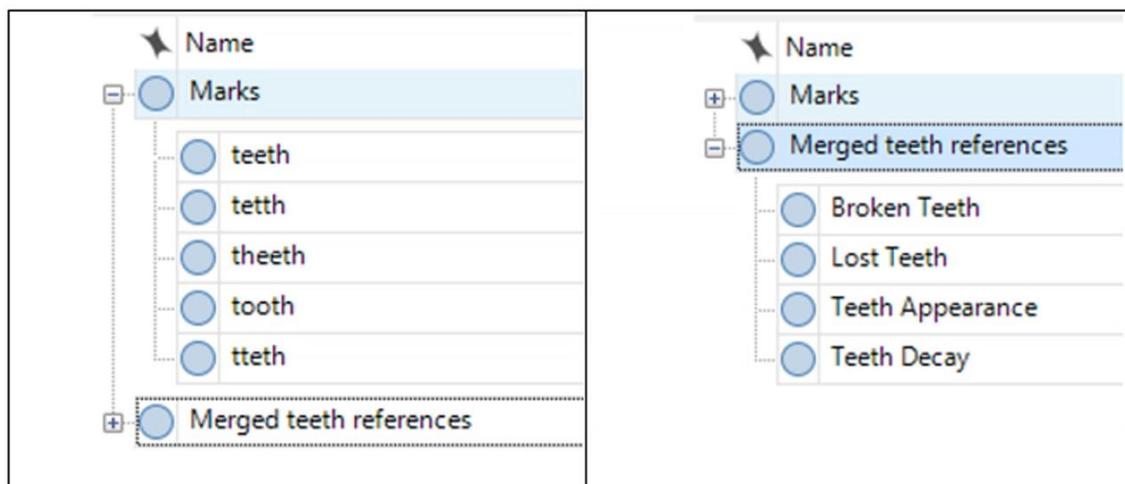

Figure 3. NVivo teeth nodes

Coding Rules and Coding Checks

While all coding rules for the current NVivo project, and the ensuing coding, have been written and performed by a single individual, the importance of creating rules and checks that have potential to open up team collaboration has not been overlooked. [39] As mentioned earlier, the list produced by the Word Frequency Query has become the key for keeping track of, and managing, all current and potential coding of material found in the "Marks" description field. This is due to the fact that when running a Text Search Query to create nodes, as described above, it is possible to

---

[38] Bazeley and Jackson, *Qualitative Data Analysis with NVivo*, 17.
[39] M. E. Fonteyn et al., "Developing a Codebook to Guide Content Analysis of Expressive Writing Transcripts," *Applied Nursing Research* 21, no. 3 (2008).

then confirm that the number of results, and when necessary, the content, match those captured in the initial Word Frequency results. This check was applied for all five Text Query nodes created in this project. The approach has led to the creation of coding rules that are by nature very specific, with little or no margin for alternative understanding or misinterpretation. Essentially, the primary coding rule for each node follows the pattern of "all descriptions in the Marks field that contain the word *teeth*". Of course, this changes slightly for node categories, such as "Lost Teeth", which are the offspring of a parent node containing all teeth references. For these nodes, the coding rules describe a concept or condition rather than a particular word. These child nodes represent a secondary level of classification that is contingent on an initial sorting of references based on the literal text they contain. Establishing an effective process for this first tier of classification or coding is crucial for the success of the project as a whole. Thus, the clear and easy to follow rules and structure for this coding process means that other potential researchers are able to integrate, without difficulty, into the project if so desired. The fact this method will be transferable and applicable for any aspect of the "Marks" description field wanting investigation is a positive outcome for the project at large.

In an earlier experiment, to identify a number of themes found in the descriptions of convicts, empty parent nodes describing broad categories were created, including "Tattoos", "Speech related remarks", "Scars, marks and injuries", and "Physical characteristics". Any remarks that do not appear to fit within these categories might be added to the free node labelled "Of interest – TBC". Free nodes are "used for storing deviant data" or "data that does not appear to fit within the categories being developed".[40] Using the same node creation and coding method described above, a cascading series of child nodes may be added under each category. While the appeal of coding the entire dataset this way is evident, it is not possible due to software issues and limitations encountered and detailed below.

NVivo Issues and Project Limitations

The greatest limitation encountered by the current project is the inability to perform an NVivo Matrix Coding Query on the full dataset. Matrix coding queries enable cross-tabulation of coded content. This function would be particularly useful for identifying and linking case IDs with references stored in nodes of interest. Uncovering the reason for the failure to successfully perform the query has been central in determining how best to define the project's engagement with NVivo software. The two main lines of inquiry were posed in the form of the following question - Was the reason for this failure due to limitations inherent in NVivo and the size of the dataset employed, or the processing power of the laptop being used to run NVivo?

The following information, found on the QSR NVivo 11 help pages, suggests that it is possible to work with a dataset much larger than the one used in the current project: "The maximum amount of data that can be imported into a single dataset is 256 fields (columns) and 1,048,576 records (rows)".[41] With only three columns and 39,340 rows it might be assumed the dataset being tested

---

[40] E Altmann, "Using NVivo Qualitative Research," ed. Maggie Walter, *Social Research Methods* (South Melbourne, Victoria: Oxford University Press, 2013). para. 13.

[41] QSR, "NVivo 11 for Windows Help," (2017). http://help-nv11.qsrinternational.com/desktop/welcome/welcome.htm.

is well within range. However, all attempts to run a Matrix Query either timed out or returned an "out of range" error message. During this testing period the version of NVivo used was updated to the current 11.4.1.1064 (64 bit) Pro Edition for Windows and the laptop used to run NVivo was upgraded to a Dell XPS 15 (Powered by 7th Gen Intel® CoreTM i7 Processor, 16GB memory, 512GB SSD). Nonetheless, the same issues occurred when attempting to run a Matrix Query.

After several exchanges with QSR Support it was agreed to supply them with the project so their technical team might further investigate. The QSR technical team were able to confirm the project did not contain any corrupt data and that the issue was in fact related to a limitation on the number of rows or columns for a dataset. It is interesting to note they appeared not to be aware of this limitation until they attempted to run the same query with the dataset and project provided them:

> The error is caused by the number of cases involved in the matrix code query. There seems to be a limitation of max 32,767 number of rows or columns of a matrix. Your project has 39,340 cases.[42]

This new information made it necessary to reconsider the overall approach to the project.

Overcoming NVivo Issues and Project Limitations

The necessity of splitting the project into smaller components is clear from the QSR advice received. Gender presents as the most obvious classification by which to divide the project. However, this is by no means an even split as the number of male cases far exceeds the number of female cases. Of the 39,340 cases in the current project, 28,476 are male and 10,864 are female. Given this disparity, a decision was made to focus on the male convict cases for the next phase of testing as it was deemed important to ensure NVivo could successfully run a Matrix Query with the larger of the two datasets. As with former attempts, the Excel dataset is imported into NVivo using the Survey Import Wizard. Importing the dataset proves much quicker this time round, taking only 7 minutes. This is likely due to a combination of fewer cases and the upgraded laptop processing power. The next part of the process involves an attempt to run the Matrix Coding Query that previously failed. While the query takes approximately 20 minutes to run, it is successful.

The results of the query, run to identify a list of cases coded to the various nodes containing teeth references, is then exported to Excel. The decision to move from NVivo to Excel for further processing and analysis of the data was determined by two factors. Firstly, while NVivo undoubtedly is a useful tool for identifying and coding references, the program struggles when running Matrix Queries - the usefulness of which has been mitigated by the need to divide the project. While there is certainly an attraction to having all themes and concepts in the entire dataset coded to nodes within a single NVivo project this is not currently possible. Instead, use of NVivo will be limited to identifying and coding cases for a particular theme or topic under investigation, for example, "burns", "scars", "injuries", "teeth" etc. Each of these investigations will necessarily entail two separate NVivo projects of their own - one for female convicts and one for male convicts.

---

[42] QSR Support response, 2017.

While this does not affect the reading and comprehension of the data, it does limit the usefulness of NVivo for exploring the dataset as a whole. For an overview of the bigger picture, it has been necessary to store the results of these multiple NVivo projects in a single Excel workbook. Secondly, as indicated earlier, the coding of this project data is of potential interest to a large interdisciplinary research team. Currently, familiarity and ease of access, as well as the capabilities of Excel, make it the most appropriate format for sharing results. An added benefit is the ease with which Excel spreadsheets can be imported to other software applications for further analysis. An example of how the NVivo results from the current project have been exported to Excel and then imported to a data visualisation program are presented in the illustration for qualitative inquiry below.

An Illustration for Qualitative Inquiry

The bar graphs in Figure 4 provide a visual overview of the results from coding for teeth references within NVivo.
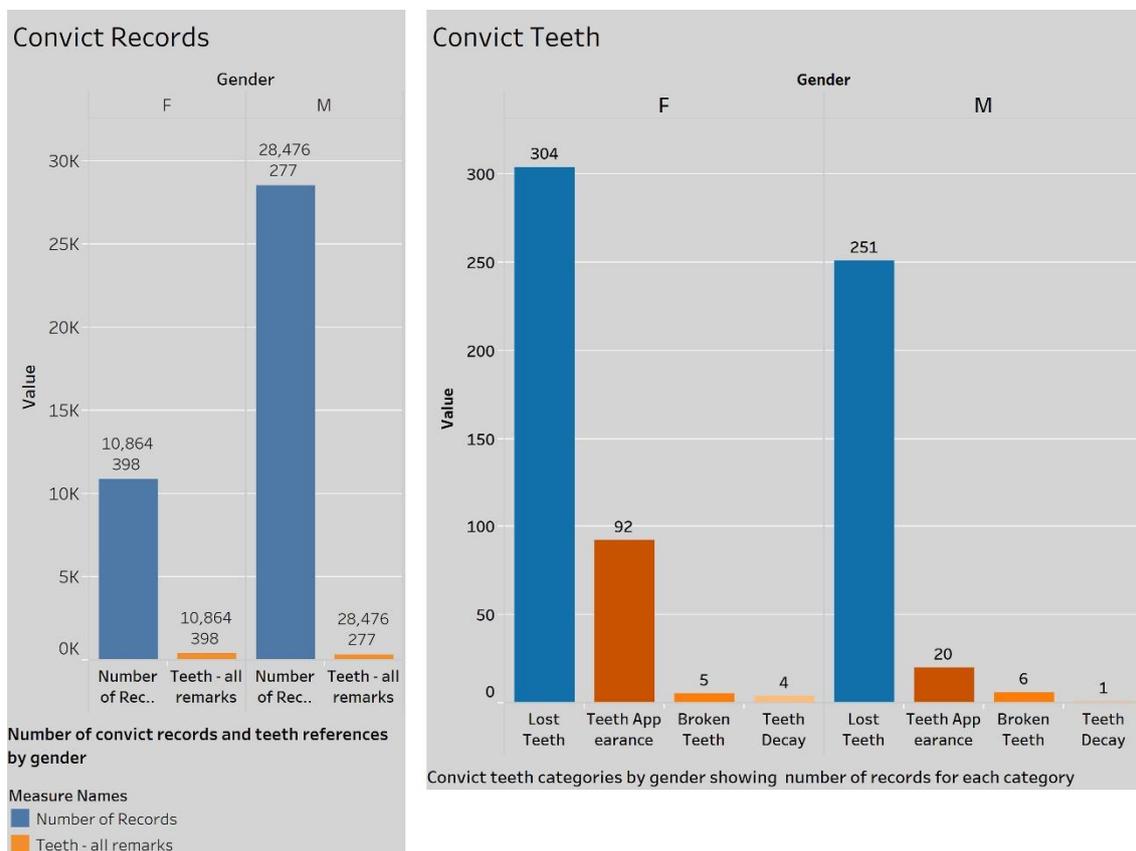

Figure 4. Number of convict records and teeth references by gender

A significant amount of concern has followed the introduction of QDAS as a tool to aid qualitative

analysis and much of this worry has focused on the actual or perceived reification and quantification of narrative text.[43] However, as Bazely points out, "qualitative analysis is necessarily interpretive", particularly when addressing open-ended narrative data.[44] While there is certainly scope for the noted occurrences resulting from this project's use of NVivo to be taken up in a study with a quantitative focus, the benefits for qualitative inquiry should not be underrated. For example, given the size of the dataset it would not have been possible to know that the larger number of teeth references belong to female convicts from a reading of the text alone. A fact that is even more striking given the number of female records is just over a third of the number of male records. While a focus on numbers might encourage the assumption female convicts lost more teeth than male convicts, refocusing on the category "teeth appearance" provides a qualitative check. Given the fact the appearance of convict women's teeth are more frequently recorded than those of male convicts, could the noting of a significantly larger proportion of lost teeth simply be a result of a greater focus on the appearance of women's teeth? The leading question here is – "Why were the teeth of convict women more of a focus than the teeth of convict men?"

Initial speculation suggests a focus on the suitability of female convicts for domestic service may be a reason for a general spotlight on appearance, including teeth. This is just one potential line of enquiry, and as mentioned earlier, further interdisciplinary research on this topic is intended. What is important to note is the role of NVivo in facilitating the initial exploratory analysis of the dataset and the potentially varied and unexpected lines of enquiry such explorations can generate.

Clearly, NVivo has been a useful tool and aid in the process of becoming familiar with the data for this research project investigating the recorded marks on the convict body. While many have raised the time spent "playing around"[45] with such programs and "the hassles one has to go through"[46] to become adept in the use of what is "recognized as a complicated software program")[47] as possible reasons to avoid QDAS, others, such as Altmann, view NVivo as "a relatively simple program to use".[48] Relative, is perhaps the key here and each researcher will need to weigh the pros and cons of QDAS use in light of the nature and aims of their specific projects. Selecting from the digital toolbox can be personal. For example, much, although by no means all, of what is achieved in NVivo by the current study could have been replicated in Excel, for example, text search queries. While not the only reason for choosing NVivo, a personal preference for what might be considered a more attractive and workable interface and alternative to Excel (at least for part of the research process) is definitely a consideration. In sharing this experience of engaging with NVivo it is hoped that others considering ways to tackle unique and large datasets might find the processes and outcomes here described useful in shaping or inspiring their own creative use of available digital tools.

---

[43] Kristi Jackson, Trena Paulus, and Nicholas H. Woolf, "The Walking Dead Genealogy: Unsubstantiated Criticisms of Qualitative Data Analysis Software (QDAS) and the Failure to Put Them to Rest," *The Qualitative Report* 23, no. 13 (2018); Winsome St John and Patricia Johnson, "The Pros and Cons of Data Analysis Software for Qualitative Research," *Journal of Nursing Scholarship* 32, no. 4 (2000).

[44] Patricia Bazeley, "Regulating Qualitative Coding Using QDAS?," *Sociological Methodology* 42, no. 1 (2012): 77.

[45] Humble, "Qualitative Data Analysis Software," 131.

[46] Zamawe, "The Implication of Using NVivo," 13.

[47] Salmona and Kaczynski, "Don't Blame the Software," 43.

[48] Altmann, "Using NVivo Qualitative Research," para. 5.

Looking Ahead

Encouragingly, a recent paper reports a commitment amongst QDAS developers "to work towards the goal of a common exchange format":

> If development of a common exchange format succeeds, it will be a major step forward for the research community at large. Such a standard will enable researchers to migrate between software packages … This can enhance the methodology of qualitative analysis as well, as researchers will no longer be locked into the confines of one particular package, instead they will have freedom to further their analysis by using another set of tools.[49]

While this is good news for increasing the range of potential digital tools available, the same paper's report on QDAS developers' views regarding data trends is more mixed:

> Developers clarified that the ease of collecting huge amounts of data via the internet is creating a demand for automated analysis tools in QDAS. However, they distinguish between 'a lot of data' and 'big data,' the latter being defined as too big for 'qualitative data analysis done by humans with tools.' To some developers, automated analysis really refers to data processing and management, not data analysis … Because it is based on frequencies and statistical procedures and cannot very well interpret text in context, as humans can, some developers feel automated analysis is antithetical to qualitative data analysis. Others feel the definition of qualitative data needs to be expanded to include big data.[50]

With a growing trend for research that engages increasingly larger datasets - headed by the "big data" revolution - and an acknowledged rise in "complex" and "large-scale interdisciplinary teams" engaged in qualitative research, it is crucial that QDAS developers keep apace of this trend and ensure their software has the capacity to assist analysis of these larger datasets.[51] This is perhaps even more important for weighing the balance given fears of a quantitative turn spawned by the growth of big data industries.

In relation to the noted "demand for automated analysis tools in QDAS" cited above, attempts to manage the large amounts of data in the current project would greatly benefit from a capacity to make every word from a Word Frequency Query into corresponding nodes in one step. Importantly, the nodes created should retain their associated references. While such a feature would ultimately benefit "data processing and management", data processing and management are arguably important and necessary steps to engaging in actual data analysis. On this point, a query was sent to QSR support to ascertain if there is a way to achieve this with current versions of NVivo. There is not. However, QSR do have an online tool called Canny where feature suggestions can be posted and shared. If the early intentions of NVivo developers still hold true, there is reason to remain positive. Likening researchers to composers at the invention of the piano, the 2002 decla-

---

[49] Evers, "Current Issues," 69-70.
[50] Evers, "Current Issues," 66.
[51] Judith Davidson, Shanna Thompson, and Andrew Harris, "Qualitative Data Analysis Software Practices in Complex Research Teams: Troubling the Assumptions About Transparency and Portability," *Qualitative Inquiry* 23, no. 10 (2017); Paulus, Lester, and Britt, "Constructing Hopes and Fears," 645.

ration of QSR founder and NVivo developer, Tom Richards, encapsulates the will and spirit desired:

> there is the continuing onus on developers such as ourselves to study what researchers do (and what they might want to do if they knew it could be done) and invent new tools, unimagined before, that help researchers do new work, also unimagined before. Let's keep inventing those new pianos![52]

In reference to data held in the Founders and Survivors database, it has been noted that despite the "increase in scale, the data analysis for this project can be conducted on laptops".[53] While this may be true, experience of the current project's engagement with NVivo would like to add a qualification of – "only just!" If software applications like NVivo develop to adapt to current research trends we may all (hopefully) continue to work effectively and comfortably from our laptops, bearing in mind that enabling mobile research has been a long-standing touted feature of QDAS.

---

[52] Tom Richards, "An intellectual history of NUD*IST and NVivo," *International Journal of Social Research Methodology* 5, no. 3 (2002): 214.

[53] Maxwell-Stewart, "Big Data and Australian History," 362.

Bibliography

Alexander, Alison. *Tasmania's Convicts: How Felons Built a Free Society.* Crows Nest, NSW: Allen & Unwin, 2014.

Alphabetical Registers of Male Convicts. CON23. TA60 CONVICT DEPARTMENT, Tasmanian Archive and Heritage Office.

Altmann, E. "Using NVivo Qualitative Research." In *Social Research Methods,* edited by Maggie Walter South Melbourne, Victoria: Oxford University Press, 2013.

Barnes, Trevor J. "Big Data, Little History." *Dialogues in Human Geography* 3, no. 3 (2013): 297-302.

Bazeley, Patricia. *Integrating Analyses in Mixed Methods Research.* London: SAGE, 2018.

Bazeley, Patricia. "Regulating Qualitative Coding Using QDAS?". *Sociological Methodology* 42, no. 1 (2012): 77-78.

Bazeley, Patricia, and Kristi Jackson. *Qualitative Data Analysis with NVivo.* 2 ed. London: SAGE, 2013.`

Bearman, Peter. "Big Data and Historical Social Science." *Big Data & Society* 2, no. 2 (2015): 1-5.

Bradley, James, Rebecca Kippen, Hamish Maxwell-Stewart, Janet McCalman, and Sandra Silcot. "Research Note: The Founders and Survivors Project." *The History of the Family* 15, no. 4 (2010): 467-77.

Davidson, Judith. "The Journal Project: Research at the Boundaries between Social Sciences and the Arts." *Qualitative Inquiry* 18, no. 1 (2012): 86-99.

Davidson, Judith, Trena Paulus, and Kristi Jackson. "Speculating on the Future of Digital Tools for Qualitative Research." *Qualitative Inquiry* 22, no. 7 (2016): 606-10.

Davidson, Judith, Shanna Thompson, and Andrew Harris. "Qualitative Data Analysis Software Practices in Complex Research Teams: Troubling the Assumptions About Transparency and Portability." *Qualitative Inquiry* 23, no. 10 (2017): 779-88.

Evers, Jeanine C. "Current Issues in Qualitative Data Analysis Software (QDAS): A User and Developer Perspective." *The Qualitative Report* 23, no. 13 (2018): 61-73.

Fonteyn, M. E., M. Vettese, D. R. Lancaster, and S. Bauer-Wu. "Developing a Codebook to Guide Content Analysis of Expressive Writing Transcripts." *Applied Nursing Research* 21, no. 3 (2008): 165-8.

Goble, Erika, Wendy Austin, Denise Larsen, Linda Kreitzer, and E. Sharon Brintnell. "Habits of Mind and the Split-Mind Effect: When Computer-Assisted Qualitative Data Analysis Software Is Used in Phenomenological Research." *Forum: Qualitative Social Research* 13, no. 2 (2012).

Graham, Mark, and Taylor Shelton. "Geography and the Future of Big Data, Big Data and the Future of Geography." *Dialogues in Human Geography* 3 (2013): 255-61.

Halford, Susan, and Mike Savage. "Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research." *Sociology* 51, no. 6 (2017): 1132-48.

Humble, Áine M. "Qualitative Data Analysis Software: A Call for Understanding, Detail, Intentionality, and Thoughtfulness." *Journal of Family Theory & Review* 4, no. 2 (2012): 122-37.

Jackson, Kristi. "Where Qualitative Researchers and Technologies Meet: Lessons from Interactive Digital Art." *Qualitative Inquiry* 23, no. 10 (2017): 818-26.

Jackson, Kristi, Trena Paulus, and Nicholas H. Woolf. "The Walking Dead Genealogy: Unsubstantiated Criticisms of Qualitative Data Analysis Software (QDAS) and the Failure to Put Them to Rest." *The Qualitative Report* 23, no. 13 (2018): 74-91.

John, Winsome St, and Patricia Johnson. "The Pros and Cons of Data Analysis Software for Qualitative Research." *Journal of Nursing Scholarship* 32, no. 4 (2000): 393-97.

Kitchin, Rob. "Big Data and Official Statistics: Opportunities, Challenges and Risks." *Statistical Journal of the International Association of Official Statistics* 31, no. 3 (2015): 471-81.

Kitchin, Rob, and Gavin McArdle. "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets." *Big Data & Society* 3, no. 1 (2016): 1-10.

Le Blanc, Amana Marie. "Disruptive Meaning-Making: Qualitative Data Analysis Software and Postmodern Pastiche." *Qualitative Inquiry* 23, no. 10 (2017): 789-98.

Maxwell-Stewart, Hamish. "Big Data and Australian History." *Australian Historical Studies* 47, no. 3 (2016): 359-64.

Oswald, Austin G. "Improving Outcomes with Qualitative Data Analysis Software: A Reflective Journey." *Qualitative Social Work* 18, no. 3(2017): 1-7.

Paulus, Trena M., Kristi Jackson, and Judith Davidson. "Digital Tools for Qualitative Research: Disruptions and Entanglements." *Qualitative Inquiry* 23, no. 10 (2017): 751-56.

Paulus, Trena M., Jessica Nina Lester, and Virginia G. Britt. "Constructing Hopes and Fears around Technology:A Discourse Analysis of Introductory Qualitative Research Texts." *Qualitative Inquiry* 19, no. 9 (2013): 639-51.

Paulus, Trena, Megan Woods, David P. Atkins, and Rob Macklin. "The Discourse of QDAS: Reporting Practices of ATLAS.ti and NVivo Users with Implications for Best Practices." *International Journal of Social Research Methodology* 20, no. 1 (2017): 35-47.

QSR. "NVivo 11 for Windows Help." 2017. http://help-nv11.qsrinternational.com/desktop/welcome/welcome.htm.

Richards, Lyn, and Tom Richards. "Qualitative Data Analysis: Can Computers Do It?". *The Australian and New Zealand Journal of Sociology* 23, no. 1 (1987): 23-35.

Richards, Tom. "An Intellectual History of NUD*IST and NVivo." *International Journal of Social Research Methodology* 5, no. 3 (2002): 199-214.

Robins, Cynthia S., and Karla Eisen. "Strategies for the Effective Use of NVivo in a Large-Scale Study: Qualitative Analysis and the Repeal of *Don't Ask, Don't Tell*." *Qualitative Inquiry* 23, no. 10 (2017): 768-78.

Salmona, Michelle, and Dan Kaczynski. "Don't Blame the Software: Using Qualitative Data Analysis Software Successfully in Doctoral Research." *Forum: Qualitative Social Research* 17, no. 3 (2016): 42-64.

Shaw, Ryan. "Big Data and Reality." *Big Data & Society* 2, no. 2 (2015): 1-4.

Sinkovics, Rudolf R., and Eva A. Alfoldi. "Progressive Focusing and Trustworthiness in Qualitative Research." *Management International Review* 52, no. 6 (2012): 817-45.

UNESCO. "Memory of the World: The Convict Records of Australia." United Nations Educational, Scientific and Cultural Organization, 2017. http://www.unesco.org/new/en/communication-and-information/memory-of-the-world/register/full-list-of-registered-heritage/registered-heritage-page-8/the-convict-records-of-australia/#c186408.

Woods, Megan, Trena Paulus, David P. Atkins, and Rob Macklin. "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* 34, no. 5 (2016): 597-617.

Zamawe, F. C. "The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections." *Malawi Medical Journal* 27, no. 1 (2015): 13-15.