



HAL
open science

Should we use NVivo or Excel for qualitative data analysis?

Marie Moncada

► **To cite this version:**

Marie Moncada. Should we use NVivo or Excel for qualitative data analysis?. Bulletin de Méthodologie Sociologique / Bulletin of Sociological Methodology, 2025, 10.1177/07591063251317084 . hal-04959772

HAL Id: hal-04959772

<https://hal.science/hal-04959772v1>

Submitted on 20 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Should we use NVivo or Excel for qualitative data analysis?

Marie Moncada

Sciences Po, Paris, France

Corresponding Author:

Marie Moncada, marie.moncada@hotmail.fr

Abstract

Coding is now a pervasive technique in qualitative data analysis, whether descriptive or theoretical, done manually by highlighting or rearranging text, or digitally using word processors or other software. However, coding relies on the researcher's reflection and does not think for itself. I compare two types of software here: one built for coding qualitative data and another originally developed for numerical calculations. Which offers more scientific freedom, and what biases come with each? NVivo excels at organising and visualising data within a single file, enabling researchers to view coded sections in context. Its flexibility supports various inferences—inductive, abductive, and deductive—and allows for critical research evaluations. It also makes it easy to create samples. In contrast, Excel is lightweight, easily transferable, compatible with other software, and better at maintaining data anonymity. Its lack of data import reduces hardware demands and errors, while filtering enhances triangulation and quantification. Excel's cell-based data reformulation aids narrative analysis. Lastly, neither tool proved clearly superior for data immersion ('Coding Fetishism') or transitioning to writing.

Résumé

Faut-il utiliser NVivo ou Excel pour l'analyse de données qualitatives ? Le codage est désormais une technique omniprésente dans l'analyse de données qualitatives, qu'il soit descriptif ou théorique, réalisé manuellement en surlignant ou en réorganisant le texte, ou numériquement à l'aide de traitements de texte ou d'autres logiciels. Le codage est cependant un processus guidé par la réflexion du chercheur et ne se pense pas par lui-même. Je compare ici deux types de logiciels : l'un fut spécifiquement conçu pour le codage de données qualitatives et l'autre fut créé initialement pour le calcul numérique. Quel logiciel offre-t-il la plus grande liberté scientifique et quels sont les biais méthodologiques qui leurs sont associés ? NVivo excelle dans l'organisation et la visualisation des données dans un fichier unique et permet notamment de replacer les sections codées dans leur contexte. Sa flexibilité s'adapte à diverses inférences – inductives abductives et déductives – et favorise l'évaluation critique des résultats. Ce logiciel encourage également la constitution d'échantillons. En revanche, Excel est léger, aisément transférable, compatible avec d'autres logiciels et permet d'anonymiser en partie les données. L'impossibilité d'y importer les données réduit les besoins en matériel et, de là, les erreurs informatiques. Par ailleurs, le filtrage améliore la triangulation et la quantification. De même, la reformulation des données dans les cellules facilite l'analyse des récits (« narratives »). Enfin, aucun des deux logiciels n'est préférable pour l'immersion dans les données (« coding fetishism ») ou la transition vers la rédaction.

Mots clé

Caqdas, codage, inférence, réflexivité, triangulation.

Key words

Caqdas, coding; inference; reflexivity, spreadsheet; triangulation.

Introduction

Like other social sciences before it, political science has faced criticism for its methodologies, particularly in qualitative analysis. Nonetheless, one promising method—coding—has emerged, as it enables multiple researchers to collaborate on the same corpus, fostering greater analytical rigor. Coding condenses data into a few words, transforming and aggregating raw data into units that describe relevant content (Holsti, 1969: 94). Coding is a subjective synthesis that can be applied to various formats—texts, PDFs, videos, audios, images, and spreadsheets—and used for materials such as interviews, questionnaires, ethnographic logbooks, or historical archives. Although useful, coding is an archaic technique, akin to basic highlighting or sorting. It cannot define research questions, formulate hypotheses, select materials, or replace the researcher's reflexivity regarding their field and problem. Nor can it develop abstract theories or identify causal relationships.

Artificial intelligence has introduced more advanced tools like autocoding and can assist researchers with corpus analysis (e.g., CoLoop from Genie Technology Inc.). However, these algorithms still lack the ability to critically distance themselves from the codes or the answers they generate. Coding does not think for itself. Despite AI advances, coding still relies on the researcher's reflection. Paradoxically, this archaic nature is its strength. Coding offers scientific freedom and avoids many methodological biases.

Coding has become a nearly universal technique for qualitative analysis and can be used in various ways. Codes can be descriptive or theoretical, created manually by highlighting or rearranging text, or digitally using word processing or more advanced software. In the following, I will compare two types of software: one designed specifically for coding qualitative data and another originally developed for numerical calculations.

First, CAQDAS programmes (Computer Assisted Qualitative Data Analysis Software) have been developed for coding qualitative data. Some of them, like NVivo (formerly known as NUD*IST for 'Non-Numerical Unstructured Data Indexing, Searching and Theory-Building'), refer to grounded theory since their creation. This social science methodology employs coding on three levels of abstraction: open coding, which captures initial ideas; axial coding, which establishes relationships between open codes; and selective coding, which groups and refines the main axial codes (Corbin & Strauss, 1990: 12-15).

In 2002, the software was renamed 'NVivo' to reflect 'in vivo' coding, derived directly from data, in contrast to 'in vitro' coding, which is developed from existing literature. Consequently, the software is positioned as a tool for inductive rather than deductive analysis (Carcary, 2011: 16; Elliott & Higgins, 2012: 5). Coding using CAQDAS is faster and more flexible compared to manual coding on paper or word-processing software (Tesch, 1990: 291; Fielding & Lee, 1991: 1; Coffey et al., 1996; Buston, 1997; Tak et al., 1999; Fielding & Lee, 2002: 214; Welsh, 2002; Fielding, 2002: 167; Park, 2002; Mangabeira et al., 2004; Park, 2004: 88; Wanlin, 2007: 258; Carcary, 2011: 15; Wickham & Woods, 2015). (For an opposite viewpoint, see MacMillan, 2005; Maher et al., 2018).

This efficiency provides researchers with additional time for analysis (Richards & Richards, 1991: 322-323; Morison & Moir, 1998: 109), thereby potentially enhancing the quantity and quality of results (Moseley & Murphy, 1997; Brent, 1984; Franzosi, 1989). Digital coding similarly facilitates the creation, modification, merging, and deletion of codes (Woods et al., 2016), thereby helping to control and eliminate potential errors (Evans, 2002: 73; Gilbert, 2002: 222). Additionally, contextualising manual coding is challenging because coded passages are cut out and pasted into a new document. CAQDAS allows for 'a more solidly grounded

analysis in empirical materials and more refined and robust results in terms of evidence’ (Rioufreyt, 2019: 99-100).

Secondly, since the early 2000s, another type of software has been utilised for coding qualitative data: spreadsheets. According to my knowledge, only four scientific articles highlight the qualities of Excel for qualitative analysis (Meyer & Avery, 2009; Amozurrutia & Servós, 2011; Bree & Gallagher, 2016; Ose, 2016). The ability to input substantial textual data into Excel (32,767 characters per cell) makes this software a valuable tool for qualitative analysis. These studies typically organise empirical data in the first column (sentences or paragraphs, questions and answers, document names, etc.) and codes in the first row (dates, hesitations, political orientations, etc.).

The screenshot below (see Figure 1) shows coding from an interview, where ‘I’ represents the interviewer and ‘R’ the respondents. The numbers in column C are codes (e.g., 1 for ‘Typical day’, 2 for ‘Time to take breaks’), with a legend in another worksheet. This simple, descriptive example reflects the preliminary coding stage, which will later become more complex.

Figure 1: Example of Coding in Excel 2021

	A	B	C
1	I	Can you describe a typical day at work?	1
2	R	Well, the days are very different. Some days we have a lot of patients and we run all day. Other days, with fewer or not very sick patients, we have time to take our breaks.	1
3	I	Do you often have to skip your breaks?	2
4	R	Yes, I often feel like a break but there is no time. I can't leave the ward at all.	2
5	I	Do you also skip your meal breaks?	3
6	R	Yes, very often. I'm often very hungry after work because I did not have time to eat my lunch.	3
7	I	Is this common among your colleagues as well?	3
8	R	Yes, but we often eat an apple or a banana when we observe a sick patient, but we do not eat any proper food outside the lunchroom.	3
9	I	How can this be improved?	4
10	R	Hmm, well, the only thing I can think of is that we would have to be better staffed during the busy periods.	4
11	I	Could you take lunch breaks in turns?	5
12	R	That's what we do on normal days.	5
13	I	So you don't sit down with your colleagues to have lunch?	6
14	R	No, never. I always eat alone.	6
15	I	Do your colleagues eat alone as well?	6
16	R	Yes, always. We take turns, and when the first is back, the second goes, and so on.	6
17	I	Who decides the order?	5
18	R	Well, some of us are hungry early, especially those with a long commute to work, and they must eat first.	5
19	I	And that works well?	5
20	R	Yes, some days. When it's very busy, we do not eat . . .	3

Source: Solveig Osborg Ose “Using Excel and Word to Structure Qualitative Data” p.153.

Excel’s numerous text functions—such as CONCAT¹, LOOKUP², IF/THEN³, and extraction⁴—empower researchers to manipulate their data effectively. Microsoft Word is sometimes employed both before importing text into Excel, for formatting purposes, and afterward, to reorganise coded content for writing (Ose, 2016). The different worksheets in the workbook also facilitate multiple coding of the same text to refine the code tree while maintaining

¹ This function combines the text of several cells into one cell.

² This search function has several variants (VLOOKUP, HLOOKUP and XLOOKUP). It aims to find the value of a cell.

³ This function assigns a cell the value or text of its choice (‘THEN’) according to the contents of another cell (‘IF’).

⁴ This function, whose main variants are threefold (‘LEFT’, ‘RIGHT’ and ‘MID’), extracts, in a new cell, the fragment of another cell.

a history of the analysis (Bree & Gallagher, 2016). These four articles unanimously support the advantages of Excel for qualitative analysis and generally overlook the software's limitations.

In terms of learning NVivo, there are abundant tutorials and forums available to quickly grasp its basics (Park, 2002; Cousins & McIntosh, 2005: 592-593; Matheson, 2005: 131). (For contrasting viewpoints, see Auld et al., 2007: 39; Kaefer et al., 2015; MacMillan, 2005; Matheson, 2005: 130; Park, 2012: 219; Salmona & Kaczynski, 2016; Schönfelder, 2011; Thompson, 2002). However, Excel is quicker to learn because researchers generally have familiarity with its basic functionalities. Tutorials and forums for Excel are also widely available online.

NVivo is expensive, and its two-week trial period may not always be sufficient for researchers to evaluate its full impact, positive or negative (Corti & Gregory, 2011; Humble, 2012: 129; Palys & Atchison, 2012: 361; Matheson, 2005: 130). For instance, the latest professional and perpetual version of NVivo (version 14, for Windows or Macintosh) costs approximately 1,000 euros for the academic community. The Excel license is less expensive: its professional and perpetual license costs 149 euros. It also offers a longer trial period of one month. As well, there are free spreadsheet alternatives that provide features comparable to Excel for qualitative data coding. In contrast, free equivalents to NVivo in the realm of CAQDAS are harder to find. Excel is thus favoured over NVivo due to its affordability and user-friendly interface. As well, NVivo's frequent updates hinder collaborative work, as different versions are often incompatible. These updates are also challenging to manage in long-term projects like a thesis. In contrast, Excel updates are usually compatible, making collaboration smoother.

Existing scientific literature does not thoroughly assess Excel's weaknesses for qualitative analysis nor compare the utility of CAQDAS and spreadsheets. Despite their different purposes and interfaces, what are the strengths and weaknesses of each tool for this type of analysis? Which software provides the most scientific freedom, and what methodological biases are linked to each tool? In other words, should we use NVivo 14 or Excel 2021 for qualitative data analysis? While NVivo and Excel do not encompass the entire spectrum of CAQDAS and spreadsheet options, comparing them is an initial step in this field.

Variables are explanatory elements with values that can be quantitative or qualitative (such as medical density in a district or political party). Codes can function as explanatory variables (or potential explanatory variables) or simply descriptive elements with limited or no explanatory potential, as shown by the example of (Ose, 2016). This might occur as well with codes determined, for instance, during an exploratory phase before a research question is formulated. Therefore, in this article, I will use the terms 'code' and 'variable' interchangeably, as my coding has consistently evolved in response to research questions. To maintain simplicity in vocabulary, I will use 'code' instead of 'node'—the latter being preferred by the company developing NVivo (Lumivero, formerly QSR International). Similarly, 'data' may be confused with 'documents' and 'interviews'.

I will summarise the strengths and weaknesses of NVivo 14 and Excel 2021 in 12 points to assess the degree of freedom each software offers. These points are not exhaustive, as not all functionalities of NVivo and Excel have been evaluated, and different researchers may use these tools in varying ways. These points, derived from my experience, have been selected for their significance in the three studies I conducted. Other subsidiary aspects, such as data visualisation, have been deliberately excluded to avoid overloading the article with minor considerations.

First, I will present the issues and codes of the three studies that form the basis of this article. Next, I will evaluate the strengths and weaknesses of the two software packages.

Overview of the three studies

My experience is that of a political science researcher. I received training in qualitative analysis and basic descriptive statistics. The main features of Excel and NVivo that I utilised were learned

at the PhD level and applied primarily in Public Policy and Media Studies. The main methodology employed during my doctoral and post-doctoral studies was discourse and narrative analysis using mixed methods. This section will present my political science thesis utilising NVivo, as well as French and a European research projects where Excel was employed. I chose NVivo for my thesis to maintain an inductive and qualitative approach and to avoid being confined to the two initial hypotheses. Conversely, Excel seemed more appropriate for the French project due to its quantitative approach. Finally, Excel was imposed for the European project because of the large number of variables involved. But the analyses from the three studies could have been partially conducted using other software. The thesis codes, being few, could have been transposed to Excel or created inductively, though this would have made re-reading time-consuming. However, searching for occurrences and selecting documents for coding would have been impossible. Likewise, NVivo's codes and functions (cases, attributes, etc.) could have replaced the variables in the French and European projects, but the quantitative analyses would have been difficult, if not impossible. To avoid any misunderstanding, I will always refer to codes created by the researcher, not the software: I have never used NVivo's artificial intelligence to generate codes automatically.

The Political Science Thesis Using NVivo

The thesis first aims to understand the stability of certain public policies over the long term. This immobility is indeed surprising considering the numerous reforms and controversies these policies undergo (Moncada, 2019). Two research hypotheses are proposed in the introduction to explore this stability and justify the chosen methodology: the cultural context of the country (Sabatier, 1988) and the influence of interest groups that obstruct political reform efforts (Pierson, 1994). Culture and interest groups are therefore identified as potential factors explaining the persistence of public policies. These two hypotheses are based on observations of the legal developments in public policies and the existing scientific literature on the topic.

Two types of data were selected: interviews and documents that explicitly take a stance on these policies. These documents originate from the legislative and executive branches, interest groups, and the print media. I gathered parliamentary reports and debates, reports from various French governmental bodies, press releases from interest groups, and newspaper articles.

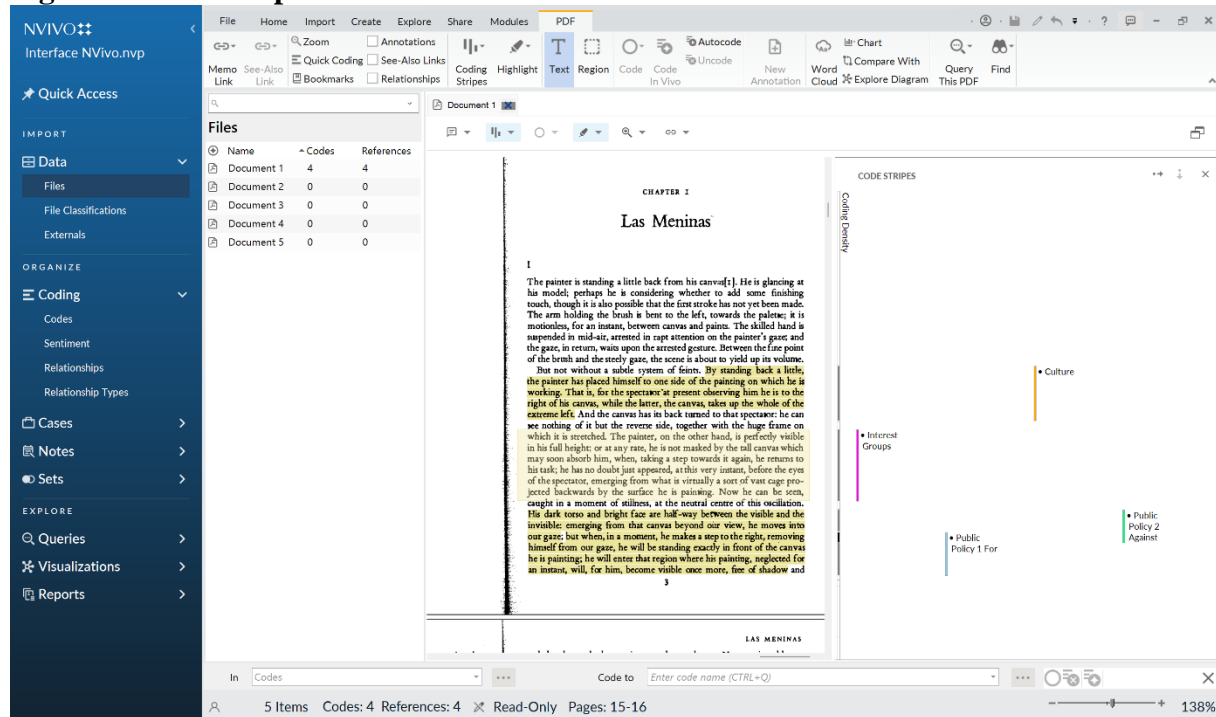
The hypothesis concerning interest groups is quantitatively tested through addressing two questions: where do mentions of interest groups appear in the documents, and which policies are associated with these mentions? This allows for measuring the influence of the associative narrative—typically supportive of maintaining public policies—within the media and political spheres. Two additional qualitative questions explore this hypothesis: how do other documents adopt the discourse of these groups, and how do interviews gauge the influence of these groups on public policy? The cultural hypothesis is primarily examined qualitatively based on the stance taken in the documents regarding the selected policies. Quantitative analysis refines these findings by examining the attention given to scientific knowledge and measuring the vocabulary used to describe the beneficiaries. This study is therefore both deductive, based on the two initial hypotheses, and inductive, given the flexibility provided by the codes ('Explanatory Variable X').

I organised the codes into three categories: stakeholders' positions on public policies ('Public Policy X: For/Against'), elements that support or challenge the initial hypotheses ('Culture' and 'Interest Groups'), and indicators suggesting other explanatory variables ('Explanatory Variable X').

The interface below illustrates how a specific document (Foucault, 1994: 3) can be coded using these categories. I selected this example for the article's anonymity and its poor optical character recognition, which was difficult to find in my thesis data and will be useful

for the remainder of this article. The text is coded by selecting an excerpt and then choosing the appropriate code. The code is selected at the bottom, in the 'Enter code name' bar. The codes listed in this drop-down menu are written in black. The stripe colours are automatically assigned by the software and are independent of this menu. NVivo offers two coding methods. The first involves highlighting text, which requires the PDF to have optical character recognition (dark yellow below). The second involves selecting a specific area of the image (light yellow). This second method is essential for PDF documents where the text cannot be highlighted due to the lack of optical character recognition.

Figure 2: Code Stripes in NVivo 14



Source: the author.

The French Research Project Using Excel

Regarding the French research project, a paper has been published (Moncada, 2021) that explores the innovative nature of proposals made by French parliamentarians, which frequently fail to be adopted. The study raises questions about whether parliamentarians function as policy entrepreneurs (Mintrom and Norman, 2009), actively seeking to reform French law. The data comprises parliamentary reports, debates in chambers, and committee discussions. An Excel file is structured from these testimonies, with each row representing a speaker on a given day. Consequently, the same speaker may appear multiple times if they speak on different days.

Variables are listed across the first row, primarily defining each speaker (first and last name, electoral district, profession, political party, status as opposition or majority, and any roles such as committee chair, rapporteur, minister, etc.). These variables also capture the exact quotes from speakers and indicate whether proposed policies may benefit their electoral district (if the speaker is a parliamentarian). The interface below illustrates the grid used to organise MPs' comments. This grid shows that the cells must be filled in manually, one by one.

Figure 3: Distribution of Data ('Quote') and Explanatory Variables in Excel 2021

	A	B	C	D	E	F	G	H	I	J	K	L
	Name	For/Against	Gender	Electoral District	Gain/Loss	Profession	Party	Majority/Opposition	Function	Chamber	Date	Quote
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												

Source: the author.

This primarily deductive study aimed to assess the willingness of parliamentarians using pre-selected variables and a theoretical framework. Only two variables were added during the analysis.

The European Research Project Using Excel

Finally, a report was written for a European project (Moncada, 2023) focusing on media coverage of a specific population. Three sources were analysed: print media, television news, and Twitter. Each row of data corresponds to a press article, television clip, or tweet. A television clip corresponds to a back and forth on the anchor. A single newscast is generally composed of about ten different clips, several of which may concern the same subject.

The coding grid is consistent for print and television, utilising 90 variables. However, the study on Twitter employs a different grid with only 33 variables. These variables were initially proposed by the work package leader—Marcello Maneri from the University of Milano-Bicocca in Italy—and were subsequently adjusted and interpreted.

For print and television, initial variables are descriptive (date, author, placement in the newspaper or program, word count, format, title, etc.). These variables also list quoted remarks (from politicians, individuals, etc.) and provide additional information about the speakers (gender, profession, political affiliation, etc.).

Visual elements in print media are similarly described through several variables dedicated to images, illustrations, maps, etc. In contrast, television clips are minutely coded second by second across separate Excel files—each tab corresponding to a clip and each file to a specific date. Each second is logged in the first column, followed by descriptions of the visual elements: camera angle, gender of individuals shown, and details of the scene.

Vocabulary associated with the target population is categorised into variables based on whether the population is active or passive (physical and mental behaviours and qualifiers used to describe them).

Additionally, interpretive variables synthesise the data, referencing narrative elements (Jones and McBeth, 2010: 329) that define plotlines, moral conclusions (policy solutions), and

causal relationships among actions and stereotyped characters (heroes, villains, etc.). These variables encapsulate the overall storyline, historical context, framing (typically described in one or two words), and setting (mainly objects and places). Salient expressions and emotional tones are also captured. Variables requiring rephrasing are often followed by exact quotes that prompted the adjustment—such as emotional expressions needing textual justification.

Twitter variables predominantly focus on description: quoted individuals (@), number of retweets, likes, comments, followers, hashtags, user type (journalist, NGO, politician, etc.), URL links, etc. Each tweet's content is recorded as well. Variables related to tweet type, approval of comments, framing, and emotional content require nuanced rephrasing.

Across all media, 'annotation' variables allow for subjective interpretations of content. These variables follow each group of descriptive variables and serve as final elements in the dataset as well. For instance, out of the 90 variables used for print and television, approximately 5 annotation variables exist. Table 1 below lists the main codes used in this survey. This study is both deductive, based on the concept of narrative, and inductive, incorporating a variety of other codes.

These three studies form the foundation for arguments developed in subsequent sections and enhance reader comprehension of examples provided in the article. They offer the advantage of incorporating diverse inferences and disciplines, including political science and media studies.

Software comparison

I will evaluate the strengths of the two software packages in organising qualitative data and ensuring transparency in the analysis. I will demonstrate how the flexibility of the coding influences the type of inference chosen and the researcher's degree of reflexivity. An additional evaluation criterion will be the contextualisation, or how the coded elements are presented within their original data. I will address the errors and limitations of the software in relation to the hardware. I will assess the ease of conducting searches for occurrences, particularly for constructing analysis samples. I will examine the impact of this software on triangulation and quantification of qualitative data. I will show how data reformulation facilitates the study of narratives. I will explore the relationship between this software and the phenomenon of 'coding fetishism'. Finally, I will discuss the ease of transitioning from coding to writing using these tools.

Data Organisation and Interface

The organisation of data in NVivo is highly flexible due to its capability to accommodate many levels in the folder tree structure (Matheson, 2005: 131). While the exact limit on the number of folders is unknown, it is possible to create at least 30 folders, each containing 30 levels of subfolders. Forum discussions often suggest that there is no practical limit on the number of folders and subfolders. This flexibility enables researchers to organise their data as they choose. While it was useful for classifying documents by public policy, I ultimately used only one hierarchical level.

Data such as text files, PDFs, and videos are duplicated when imported into the analysis file (or 'NVivo file'). The original data remains untouched and cannot be modified or coded. Data security measures such as using the computer's password and NVivo's password further ensure its protection. However, NVivo 14 does not currently offer data encryption, which remains a concern for securing sensitive data in CAQDAS (Palys & Atchison, 2012: 362). Compiling several thousand documents into a single NVivo file has proven invaluable for backup and portability, allowing seamless transfer of studies between locations without risking loss of original data (Roberts & Wilson, 2002; Di Gregorio & Davidson, 2009: 42; Lejeune, 2016: 191-192; Oswald, 2019: 438).

Although it is impossible to import files into Excel and code them directly, it is possible to copy specific data extracts into cells, as was done for the French project ('Quote'). However,

this procedure would have been impractical for the European project. This direct manipulation of materials increases the risk of accidental deletion or modification. Therefore, NVivo is regarded as more suitable for effectively organising data.

Transparency of the Analysis

Transparency refers to a study's ability to openly share how hypotheses were developed, the coding method, the data, the analysis, and the construction of results. This practice promotes two objectives: the study's reproducibility and its critique.

The scientific literature emphasises the transparency of coding with CAQDAS (Bringer et al., 2004; Carcary, 2011; Evers, 2011; Fielding, 2002; Gibbs et al., 2002; Hutchison et al., 2010; Kaefer et al., 2015; Paulus et al., 2017; Salmona & Kaczynski, 2016; Siccama & Penna, 2008; Sinkovics & Alfoldi, 2012; Wickham & Woods, 2015). This transparency was further bolstered by the introduction of the REFI-QDA Standard in 2019, facilitating project transfer across different CAQDAS platforms (such as ATLAS.ti, MAXQDA, QDAMiner, etc.) (Evers et al., 2020). However, this standard cannot convert all elements from an NVivo file (such as framework matrices, 'see also' links, etc.), and all participating CAQDAS require a fee.

I initially chose NVivo in my thesis to expedite the selection and coding of thousands of documents and to make my coding process transparent online. However, this idea of transparency was quickly abandoned for two reasons: anonymising my interviews would have been tedious, if not impossible, and the thousands of press articles I had collected were from paid databases.

Spreadsheets offer greater transparency because Excel is more easily transferable than NVivo. Workbooks can be readily shared with scientific journals since Excel files are lightweight and widely compatible. This challenges the notion that an NVivo file inherently enhances research transparency in peer review (Sinkovics & Alfoldi, 2012).

As well, unlike NVivo files that contain empirical data within the software, Excel requires data to be transmitted to a cloud service. If researchers prefer not to share their original potentially sensitive data, Excel may provide a minimum level of transparency compared to NVivo because the researcher can choose, for example, to copy only extracts from the corpus into the cells. As mentioned in the introduction, NVivo's frequent updates are also difficult to manage for collaborative or long-term projects. Thus, Excel is considered a software that promotes transparency and reproducibility in study research.

Coding Flexibility, Inference and Reflexivity

NVivo is flexible in coding. Firstly, the code tree is unrestricted, allowing for an unlimited number of codes to be created. NVivo also enables flexible operations such as creating, reformulating, deleting, merging, and prioritising codes. Moreover, coding is researcher-dependent and not software-dependent. It is important, however, to manage the depth of sub-codes cautiously to prevent potential software performance issues.

NVivo offers three additional prosopographical features: classifications (of documents or cases), attributes, and cases. Classifications are variables (e.g., person, place, organisation) whose value ('attribute') is specified by the researcher (e.g., gender/age, town/country, company/union). Cases are entities (e.g., Amélie, Paris, Total) that can be categorised based on these attributes. Additionally, it is possible to locate all elements associated with a case (e.g., Amélie's interview, Paris' photo, Total's video) with a single click.

Additionally, NVivo offers four other functions to enhance analysis. Memos are unrestricted notes that can only be linked to a single element (such as data or code), which some find overly restrictive (Schönfelder, 2011). Annotations provide comments on the content of data or codes. 'See also' links establish connections between different elements (like codes, data, or memos) based on similarities or contradictions. Finally, reports are standalone and are typically used to compile initial findings from the analysis.

If a researcher finds these tools too restrictive, they have the option to utilise sets. Sets allow researchers to link themes across data, codes, cases, memos, and analysis results (such as word frequency and code matrix). Multiple tools can belong to several sets simultaneously, enabling sets to offer a higher level of abstraction than individual tools. This flexibility provides researchers with the opportunity to create a second reading grid that can correspond to a new hypothesis or different research question.

These nine features in NVivo are labelled and organised at the discretion of the researcher. Many studies emphasise that the researcher bears sole responsibility for the quality of their research (Bringer et al., 2004: 249; Carcary, 2011: 15; Carvajal, 2002; Cousins & McIntosh, 2005: 597; Evers, 2011; Fielding, 2002: 162; Kaefer et al., 2015; Thompson, 2002). Thanks to this flexibility, inference can be inductive (creation of new codes during the study, for instance), deductive (pre-established codes) or abductive (pre-established codes and then creation of new codes, or even recoding of the old codes and data). Similarly, the researcher can be both close to and distant from their data (Bazeley & Jackson, 2013: 7). This flexibility allows for multiple interpretations, often enabled by the inductive approach. It also provides an escape from a potentially unproductive deductive approach by creating and cross-referencing new codes and features.

Old scientific articles express concerns that researchers might default to grounded theory because they believe it is the only method suitable for CAQDAS (Carvajal, 2002; Coffey et al., 1996; Lonkila, 1995: 50). However, I agree with Nigel Fielding's argument that '[i]t is most unlikely that qualitative software will prompt the invention of new methods or schools of thought' (Fielding, 2002: 173). Indeed, it would be restrictive to confine ourselves solely to the grounded theory method, which is very precise and restrictive, to use NVivo. The deductive nature of the two initial hypotheses in this thesis demonstrates that other approaches are possible. Furthermore, like many political scientists in France, I was never trained in grounded theory, even during my PhD.

The coding in Excel is also flexible and suitable for both literature review and data analysis. Researchers have the freedom to determine the names of the codes and the coded passages. Excel permits coding with finer detail compared to NVivo, where managing many codes can be challenging. In Excel, codes are fixed and can be quickly visualised using different display functionalities, such as colours for text or background. In contrast, NVivo's 'Enter code name' bar dynamically adjusts the order of codes based on the last coding action, and all codes appear uniformly in black. It is therefore not possible to change the colour or order of the codes in this bar. This lack of control makes it challenging for researchers to manage their coding process efficiently. When working with NVivo, researchers often must memorise their codes or manually type them into the search bar to locate them quickly, which becomes more cumbersome when dealing with many codes.

In Excel, it is straightforward to work with variables at the bottom of the tree structure. In contrast, NVivo presents all codes from the entire tree structure during coding, which can make the process laborious due to the multitude of codes offered by NVivo. Excel allows for faster information collection since copying and pasting suffice for most tasks. However, the code tree is inherently limited in Excel. Since coding is performed from variables at the bottom of the tree structure, grouping them into numerous higher categories can be difficult or even impractical. Moreover, having too many rows above these variables might interfere with Excel filters. Therefore, when dealing with large datasets, it is advisable to keep the tree structure of codes in Excel relatively constrained.

Similarly, while it is straightforward in Excel to create, rename, or delete codes, merging them or modifying their hierarchy is considerably more challenging. This limitation reduces the flexibility of coding in Excel. Consequently, this rigidity influences the type of inference favoured by spreadsheets. NVivo, in contrast, is capable of handling inductive, abductive, and deductive

analyses, whereas Excel primarily supports deductive analysis. Therefore, it is preferable to establish codes before coding in Excel.

The absence of a dedicated coding function in Excel necessitates repetitive copying and pasting of documents with optical character recognition. This repetitive task can be tedious for researchers due to its automated nature. This automatic process is more pronounced in deductive coding, which inherently discourages questioning the relevance of codes beforehand. In contrast, NVivo promotes researcher reflexivity by encouraging them to critically assess their actions and decisions during the coding process. For these reasons, NVivo surpasses Excel in terms of coding, inference, and fostering researcher reflexivity.

Contextualisation of Codes

It is possible, in NVivo, to access the content of a code with a single click. Researchers do not need to recall where these passages are located, enabling them to focus on analysing their data. This content is presented either alone or within the original document — on its page or in its original position within the entire document.

This contextualisation is crucial because isolated coded passages can sometimes be insufficient for complete understanding (Gibbs, Friese & Mangabeira, 2002; Roberts & Wilson, 2002; Leech & Onwuegbuzie, 2011: 82). NVivo's capability to reintegrate coded extracts into their original documents enables researchers to reevaluate and correct their coding during subsequent reviews if necessary.

Contextualisation in Excel occurs primarily through hyperlinks to online documents, allowing one-click access, though without highlighting coded sections. This method is restrictive compared to NVivo, making NVivo better suited for contextualisation.

Software Limitations and Errors

Before discussing the limitations and errors of NVivo, it is useful to note that my computer meets NVivo 14 requirements (see Table 3 in Appendix). The processor might be its only weakness. Having used NVivo on several computers, I can confirm that the software is highly sensitive to processor power.

An NVivo file on Windows is limited to 10 gigabytes (GB) compared to 512 GB on Macintosh (version 14). However, some NVivo features available on Windows are not accessible on Macintosh. To avoid crashing during a search or startup, this file must be under 10 GB. Yet, for my PhD thesis, the combined size of the four sample documents and interviews exceeded 10 GB using only word processing documents and PDF files (Moncada, 2019). Consequently, I had to split my work across approximately ten different NVivo files. Some functionalities (memo, annotation, 'see also' links, sets, etc.) could not be utilised to connect these separate NVivo files. Observations, tasks, and initial analytical conclusions were recorded in a thesis notebook or word processing document due to this fragmentation. NVivo primarily supports synchronic analysis as a result (Humble, 2012: 132).

An NVivo file can take 15 to 30 seconds to start up and may fail to start if it exceeds 8 or 9 GB. Similarly, importing documents is time-consuming (up to several minutes sometimes) and may eventually fail. For instance, importing several gigabytes can take tens of minutes. Coding can also take two to three seconds for large analysis files or data (Schönfelder, 2011). Additionally, the time for conducting a search varies. If at least three Boolean operators are used and the corpus has poor optical character recognition, the search can last several tens of hours. Moreover, a search may crash after 24 hours if the analysis file exceeds 9 GB (MacMillan, 2005).

Optical character recognition requires as well several hours in total for data that lack it. Scanning certain data (logbooks, photographs, etc.) can also be time-consuming (Franzosi et al., 2013: 3244; Matheson, 2005: 130).

A final drawback is the lower image quality of documents (especially PDFs) in NVivo, making it difficult to read data.

In comparison, Excel has the significant advantage of never crashing because the data does not burden the software. However, it is easier to make human errors in Excel. For example, one might unintentionally change or delete empirical material (if copied into the spreadsheet) or the contents of variables. Nevertheless, I believe that human errors are less restrictive than NVivo's numerous technical limitations. NVivo would significantly benefit from being less demanding on hardware.

Searching for Occurrences and Sampling

Using the text search query in NVivo helps the researcher to identify potentially relevant passages for the study (La Pelle, 2004: 100; Saldaña, 2013: 32-34) and to align coding with keywords. However, achieving accurate search results often required several attempts. NVivo employs its own Boolean operators⁵ (MacMillan, 2005) and is very sensitive, sometimes overly so, to punctuation choices (such as curved apostrophes).

The text search was instrumental in addressing key questions from my PhD thesis (Moncada, 2019): 'Where do occurrences of interest groups appear in the data?'; 'Which policies are associated with these hits?'; 'How is the target population labelled?' By searching for the names of interest groups, specific public policies, and the vocabulary used to describe the beneficiary population in my empirical data, I was able to cross-reference these occurrences. This analysis highlighted the policies that interest groups focused on and revealed the nuances in how different groups referred to the target population, indicating their positive or negative stance toward them.

Sampling in NVivo involves extracting raw results by placing all data on equal footing. My thesis's straightforward results illustrate this issue. To qualitatively analyse the thousands of collected documents, I consistently created samples based on document type. Four distinct samples were formed: interest groups' press releases, press articles, and documents from legislative and executive powers. These samples were constructed using two criteria: a significant number of hits related to the studied policies for each document type and a representative temporal distribution across all documents.

These samples were chosen quantitatively, sometimes leading to the exclusion of valuable documents. This sampling approach may inadvertently exclude data that could contribute to a more nuanced explanatory model. Similarly, this sampling may lead to an over-representation of the studied phenomenon, potentially overlooking new explanations found elsewhere. For instance, I suspect that European constraints, mentioned in general administrative documents, could provide an additional explanation for the budgetary restrictions on the public policies examined in this thesis. This negative aspect of sampling is, however, a methodological choice for which the software is not responsible.

In contrast, Excel lacks a hit search function, necessitating the construction of the corpus before analysis. This forces researchers to qualitatively select their data, a more time-consuming process but potentially preferable for specific research inquiries. NVivo's hit search adds value by facilitating sampling and harmonising coding, yet maintaining a qualitative perspective is crucial for constructing robust samples. Searching for hits in Excel is cumbersome because the software returns results by cell rather than within the cell itself. This means that if a cell contains thousands of words, researchers must read through the entire cell to locate specific words, making Excel unsuitable for efficient hit searches.

⁵ Boolean operators are the elements used to search for hits ('and', 'or', 'not', asterisk, brackets, inverted commas, etc.).

Qualitative Data Triangulation

In my thesis, the fragmentation of NVivo files inherently restricts data triangulation. In contrast, systematic coding in Excel facilitates visualising analysis on a single spreadsheet, which enhances triangulation.

This is particularly evident with annotation variables in Excel, like NVivo memos. They allow to freely interpret the different contents. These variables are positioned after each group (e.g., ‘emotion’, ‘favoured policy solution’) and at the end as a final variable (Moncada, 2023). Unlike memos that require separate opening, these annotation variables provide a brief view of all associated cells, facilitating data extraction and verification of coding coherence.

Excel’s filters further enable cross-referencing between variables. While NVivo also generates matrices crossing multiple variables (codes, classifications, data, etc.), file fragmentation limited to 10 GB in my PhD thesis (Moncada, 2019) hindered this capability. Excel’s filters are more user-friendly and efficient compared to NVivo’s matrices, making qualitative data triangulation more manageable in Excel.

Quantification of Qualitative Data

Recent studies have been generally positive about the potential for quantifying certain elements in NVivo, even if this feature is not used systematically. Quantification primarily applies to codes, allowing for measurements like saturation levels in interviews, focus groups, and scientific literature (Lowe et al., 2018). It can also be used to assign scores to various factors explaining the adoption of public policies (Banha et al., 2022).

I also used this possibility to answer the question ‘How do respondents measure the impact of interest groups on public policy?’ I thus compared the ‘Interest Groups’ code with other factors cited by my respondents, such as government political affiliation, new scientific discoveries, etc. (Moncada, 2019). Similarly, for the question ‘What is the position of the documents on the selected policies?’ (Moncada, 2019), it is straightforward to measure the proportion of ‘for’ and ‘against’ arguments while highlighting the most debated policies.

Like NVivo, Excel enables prioritisation of codes and their content. But NVivo lacks simple statistical tools, making Excel advantageous for quantification. Excel’s descriptive statistics yield visually accessible results and facilitate historical analysis by cross-referencing temporal variables with others using Excel filters. This function was useful, for instance, in tracking and analysing the historical development of the main frames across the different media I was studying.

Moreover, quantification and reformulation are crucial for second-by-second television coding in Excel: each second is recorded in the first column (one second per row), with subsequent columns describing the visuals (Moncada, 2023). While NVivo supports video coding (Evers, 2011) and quantifying certain types of images (visibility of political figures, institutions, target populations, etc.), it was impractical due to inability to load French National Audiovisual Institute content onto my computer and the unique nature of each video shot. Excel, however, allows for detailed initial coding of video content, which can later be simplified by aggregating codes. Although NVivo could potentially code video transcripts, the lack of timestamps in these transcripts and the complexity of coding images would have made the process lengthy or even unfeasible without creating new codes. A recent study highlighted NVivo’s limitations in analysing non-importable online videos, suggesting that current CAQDAS tools provide limited support for such analyses, with strategies often involving later-stage memo coding to describe these videos (Snelson et al., 2021).

In summary, Excel offers robust quantification capabilities compared to NVivo and is also valuable for analysing videos that cannot be imported into computer systems.

Data Reformulation and the Study of Narratives

An additional benefit of Excel is the researcher's capacity to reformulate their data. It permits reading and synthesising statements across multiple variables. Excel's reformulation enables summarising data concisely within a chosen variable, without needing to use an existing code or create a new one. This detailed level of abstraction, on a cell-by-cell basis, surpasses simple code formulation because the number of cells potentially describing the same document can exceed the number of codes. With Excel, the corpus description can be more detailed and complex, though it tends to be messier since there's no need to harmonise cell contents.

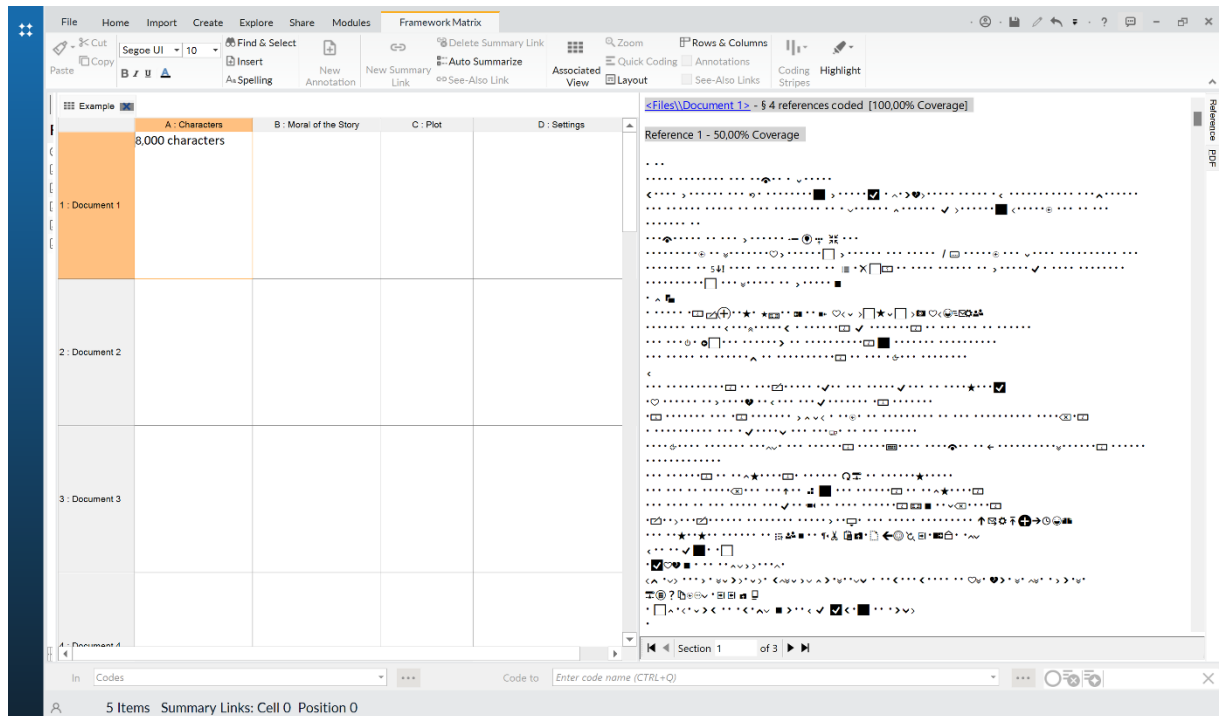
In narrative analysis, reformulation might involve capturing the overall emotion of data (for instance, as positive or negative), summarising key issues, or explaining why a political solution, explicitly or implicitly, is preferred. It also aids in data reorganisation. For instance, in analysing parliamentary debates, I consolidated various oral interventions by the same speaker on a single day into a single cell. This consolidation was then used to populate different variables (plot, settings, policy solutions, etc.). Similarly, it was convenient for compiling quotations attributed to individuals across different sections of a press article into a single cell. This allowed for the addition of attributes like emotion, political affiliation, qualifiers, and the journalist's tone to these political comments. Thus, I provide a nuanced perspective on a recent study that assesses the benefits of CAQDAS in narrative research (Consoli, 2021).

I encountered this challenge in a European project where I had the flexibility to use either Excel or NVivo to highlight the dominant narrative in a small dataset during an exploratory phase with a limited number of variables. Despite initially favouring NVivo for analysis, both the work package leader and I ultimately acknowledged, during feedback on the results, that we had both opted for Excel for all these reasons. We needed to reorganise our data, particularly for quantitative purposes, and to code them for each variable specific to the analysis of the narratives. It was also easier to freely rephrase the various emotions, even if it required later harmonizing them with word processing software.

In NVivo, reformulation can be accomplished using framework matrices linking cases and codes. For the media analysis I conducted (Moncada, 2023), each newspaper article and television clip would become a case, and each variable would become a code. NVivo allows for working with two windows: one for the empirical data and one for the table, which resembles a spreadsheet where cells can be manually filled. This setup offers two advantages. Firstly, clicking on a cell opens the corresponding document (press article or TV clip), facilitating faster coding and contextualisation. Secondly, each cell can accommodate up to 8000 characters, sufficient for many studies.

However, this table is cumbersome for three reasons. Firstly, it requires creating around 90 codes, which unnecessarily complicates coding in NVivo, as, as we have seen, using too many codes in this software is challenging. Secondly, handling this table becomes difficult with many variables: Excel's display features are more robust in comparison. Lastly, poor optical character recognition can undermine this process, as illustrated in my previous example (Foucault, 1994).

Figure 4: A Framework Matrix in NVivo 14 with a Poor Optical Character Recognition



In NVivo version 12, a similar grid could be created without the need to generate new codes, although the process was somewhat more intricate due to the requirement of creating classifications and attribute values.

Therefore, publications asserting the incompatibility between CAQDAS and narrative analysis must be nuanced (Cypress, 2019: 216; Franzosi et al., 2013; Humble, 2012: 132; Silver & Rivers, 2014: 14). It is also possible to create narrative codes in NVivo (e.g., a code for settings, a code for heroes, etc.). Basic narrative analysis is thus achievable in NVivo, but Excel's functionalities remain superior for this purpose.

Coding Fetishism

CAQDAS can create excessive proximity to empirical data – potentially overwhelming researchers with information. In the literature, this alienation is referred to as 'codomania' (Lejeune, 2016: 188), 'coding fetishism' (Marshall, 2002: 61-62; Humble, 2012: 132; Saldaña, 2013: 2; Woodset al., 2015: 390), and 'coding trap' (Gilbert, 2002: 218). On one hand, coding can mistakenly be equated with scientific rigor due to the quantification it provides (Maclaran & Catterall, 2002: 215; Lejeune, 2016: 188-191). On the other hand, focusing on coding a sample often limits the time available for writing, particularly for new users (Mangabeira et al., 2004: 174; Silver & Rivers, 2014; Lejeune, 2016).

My dissertation suffered from this alienation: it overlooked the transition between coding and obtaining results by minimising the iterative process between data, findings, and scientific literature (Moncada, 2019). Firstly, it could benefit from cross-referencing its findings with new interviews to uncover additional insights and refine initial conclusions. Secondly, deeper engagement with the scientific literature could help establish more nuanced connections between the results and existing research. Therefore, the iterative process between interviews and results, and between results and the scientific literature, remains insufficient.

Coding fetishism also exists in Excel. A large corpus or many variables can compel researchers to code their data quickly, sometimes without adequate time for reflection. This time constraint manifests in two ways: underutilisation of annotation variables and an overemphasis on quantitative results to compensate for qualitative shortcomings.

These issues were particularly evident in the European project where media had to be coded in a short time (Moncada, 2023). Annotation variables were underutilised despite their potential to provide analytical flexibility beyond the rigid variables imposed by the leader. The time constraint also impacted the writing of the final report which comprises approximately fifty tables and figures presenting quantitative representations of coding outcomes. Data interpretation relied heavily on interviews with journalists rather than direct media excerpts, posing two challenges. Firstly, interviews substituted for direct analysis despite their limited and inherently subjective nature. Secondly, a more nuanced integration of empirical data with researcher insights could have enhanced the objectivity of the analysis and strengthened the scientific rigor.

Thus, neither NVivo nor Excel offers immunity against coding fetishism, as this issue is inherent to the coding process itself, not the software used. This phenomenon highlights two inherent challenges of coding: difficulty in estimating time requirements and the temptation to adopt systematic approaches, often driven by quantitative goals. Thus, the literature debates whether coding is an analytical process or merely a preliminary stage to qualitative analysis.

Transition to Writing

How does one transition from coding to writing? In all three of my studies, the codes never directly formed the writing plan, but they were an integral part of it. My political science thesis is constructed in two parts, each corresponding to the two initial hypotheses. The subsections deal with different types of material. The article resulting from the French project is also divided into two parts. It begins with an observation—the existence of a parliamentary innovation—and then aims, in the second stage, to understand its existence. Finally, the report on the European project is divided into three parts, each corresponding to a case study. The subsections report on the dominant narrative in the French media and attempt to understand its existence.

The codes themselves include part of the phenomenon studied (in italics in Table 1) as well as the explanatory variables, just like the writing plan. The codes in the thesis address the existence of public policies that are controversial but stable over time ('For/Against'). The French project observes an innovative parliamentary position in comparison with other public actors ('For/Against'). Finally, the European project highlights the dominant media narrative of the three events ('Narrative'). The table below compares the writing plan with the codes in each of the studies.

Table 1: Codes and Writing Plan of the Three Studies

	Political Science Thesis (NVivo)	French Research Project (Excel)	European Research Project (Excel)
Codes	<p><i>Public Policy 1 (For/Against)</i> <i>Public Policy 2 (For/Against)</i> ... <i>Public Policy X (For/Against)</i></p> <p>Explanatory Variable 1 (Interest Group Hypothesis) Explanatory Variable 2 (Culture Hypothesis) ... Explanatory Variable X</p>	<p>Name <i>Public Policy (For/Against)</i> Gender Electoral District Gain/Loss Profession Party Majority/Opposition Function Chamber Date Quote</p>	<p>Description (Date, Author, Position in the Media, Number of Words, Format, Title, etc.) People Quoted (Gender, Profession, Party, etc.) Images (Plane, Gender, Details, etc.) Target Population (performing or undergoing an action) <i>Narrative (Plot, Policy Solution, Heroes, Villains, Victims, Frames, Settings, Emotions, etc.)</i> Tweet (Retweets, Likes, Comments, Followers, Hashtags, Type of User, etc.) Free Annotations Etc.</p>
Writing Plan	<p><u>Part 1: Interest Group Hypothesis</u> Section 1: Interest Groups Documents Section 2: Interest Groups Documents Section 3: Interest Groups Documents</p> <p><u>Part 2: Culture Hypothesis</u> Section 1: Written Press Documents Section 2: Legislative Documents Section 3: Executive Documents</p>	<p><u>Part 1: Parliamentary Innovation</u></p> <p><u>Part 2: Explanatory Variables</u></p>	<p><u>Part 1: Case Study 1</u> Section 1: Narrative Content Section 2: Explanatory Variables</p> <p><u>Part 2: Case Study 2</u> Section 1: Narrative Content Section 2: Explanatory Variables</p> <p><u>Part 3: Case Study 3</u> Section 1: Narrative Content Section 2: Explanatory Variables</p>

The codes let the data to ‘express itself’ through inductive openness. For the thesis, explanatory variables emerged as documents were read (‘Explanatory Variable X’). In the French project, the small sample size enabled adding new variables for testing (‘Opposition/Majority,’ for instance, was introduced during the article writing). In contrast, adding new variables in the European project was challenging due to the need to systematically fill in existing variables, their large number, and the substantial volume of documents. Thus, recoding for new variables would have been time-consuming. Consequently, few new variables were added, focusing on easily collectible information (qualifiers for target populations, television sequence start times, journalist names, etc.).

How are results identified and the writing plan constructed? In all three studies, sections and subsections are written in two parts: the phenomenon studied and the explanatory variables. The phenomenon studied involves simple synthesis (public policies for the first study,

parliamentary innovation for the second, and the dominant narrative for the third). Then, how to select the variables from the coding? For the thesis, these align with the initial hypotheses but also emerge inductively from documents and interviews (policy entrepreneurs, constitutional openness to external demands, etc.), mainly in the interest groups section. Fragmented analysis files made memos, annotations, 'see also' links, and sets unusable, so only the codes were used to highlight new variables.

While it is important to check the relevance of one's coding by rereading it at least once, coding too thoroughly is not necessary. Due to the difficulty of switching between codes during the process, it is better to use a simple tree structure in NVivo. A word processor suffices to rearrange and synthesise the codes more finely afterward. While NVivo can encode an image or a document without optical character recognition, it will render the encoded passage in encrypted form (e.g., '1:30,187 - 567,265'), making it harder to use in a word processor.

The variables selected for the French project come directly from the Excel file. Except for gender, which seemed unconvincing, all variables are used. The European project report employs almost all the Excel variables to test their significance, adding others such as the historical association of the target population with the case study, media cross-quoting, and the local or national dimension of the case study. The significance of the variables was always established quantitatively.

Once variables are tested and selected, how can they be incorporated into a general theory? This generalisation involves two elements: discussing significant variables with the scientific literature and initial hypotheses, and establishing a model significantly more abstract than the simple set of variables. This discussion aims to refine existing theories or construct new ones. Thus, spreadsheets and CAQDAS do not hinder the researcher from thinking (Fielding, 2002: 162).

This final abstraction should be distinguished from coding. An experienced user feels it is essential to differentiate between coding, analysis, and writing: 'I should be able to arrive at a cognitive map, but it is not there. The best way to use [a CAQDAS] is to organise data, as a text reduction exercise for large volumes of text. Then, leave the package and go somewhere else' (Mangabeira et al., 2004: 170). This user suggests performing the final analytical step in another software (Mangabeira et al., 2004: 171). A word processing program or a simple handwritten notebook can construct this cognitive map.

Excel's annotation variables, which can take the form of attribute values in NVivo, could hypothetically facilitate this transition to writing. They allow for less rigid abstraction than axial or selective coding (Corbin & Strauss, 1990: 12-15) because they are freely written. Axial or selective coding implies constructing homogeneous codes, which can be premature at certain research stages. These annotation variables can accompany research smoothly, written freely (like memos) or more homogeneously (like codes). Since annotation variables are easier to handle than attribute values, Excel might be superior to NVivo for moving on to writing. However, the hypothetical added value of Excel for transitioning from analysis to writing has not been empirically tested due to the underuse of annotation variables. The way I used these two software packages suggests they are on an equal footing: neither Excel nor NVivo proved superior in transitioning from coding to writing.

Conclusion

This article is based on three observations: Excel is useful for qualitative analysis (as supported by scientific literature and my own experience), it is easy to learn, and it is less expensive than NVivo. Given these points, which tool should be preferred for qualitative analysis: Excel or NVivo?

NVivo is preferable for organising and visualising data because it can import the data into a single analysis file. This allows coded passages to be viewed in their original context. The ease

of modifying codes in NVivo supports inductive, abductive, and deductive analysis ('Inference') and encourages researchers to critically evaluate their own protocols ('Reflexivity'). Additionally, importing data facilitates the search for occurrences, aiding in the construction of samples for coding.

However, Excel has several important advantages. Its format is lightweight, easily transferable, compatible with other software, and interchangeable with many free spreadsheets. It also ensures the anonymity of sensitive data. Excel promotes transparency and reproducibility. Because it does not import data, it places less demand on hardware, resulting in fewer errors. Its filtering capabilities make it easier to triangulate and quantify data. Furthermore, the ability to reformulate data within cells simplifies the study of narratives.

Finally, I found no superiority of one software package over the other in two areas: the risk of getting lost in the data ('Coding Fetishism') and the transition to writing. The table below summarises these observations.

Table 2: Strengths and Weaknesses of NVivo 14 and Excel 2021

	NVivo 14	Excel 2021
Data Organisation and Interface	+	
Transparency of the Analysis		+
Coding Flexibility, Inference and Reflexivity	+	
Contextualisation of Codes	+	
Software Limitations and Errors		+
Searching for Occurrences and Sampling	+	
Qualitative Data Triangulation		+
Quantification of Qualitative Data		+
Data Reformulation and the Study of Narratives		+
Coding Fetishism		=
Transition to Writing		=

I propose two extensions to this discussion. First, it is feasible to recode a dataset in Excel after sampling and clarifying the variables in NVivo. Conversely, it is also possible to import an Excel file into NVivo to encode the cell contents. However, these Excel files must be formatted correctly for the import to be successful, and it is not possible to modify cell contents in NVivo afterwards. This complementarity between NVivo and Excel warrants further exploration.

Second, the ability of Excel to summarise data across numerous variables offers a form of inductive analysis. While I have not tested this approach, it would be interesting to code a small sample using many variables in Excel. Such an exploratory study could help design a later pilot study with fewer variables or a larger-scale study with extensive data and fewer variables.

Declaration of conflicts of interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This publication has been funded by the European Union under the Horizon 2020 research and innovation programme under grant agreement no. 101,004,564 (project BRIDGES). Its contents are the sole responsibility of the authors and do not necessarily reflect the views of the European Union. The European Commission and the European Research Executive Agency are not responsible for any use that may be made of the information it contains.

Acknowledgments

I warmly thank Pablo Báez Guersi, Cécile Tron-Muratori, and Nathalie Chauvac for their valuable comments and suggestions, which helped improve and clarify this article. I also extend my gratitude to the two anonymous reviewers and the editorial team, particularly Ahmed Fouad El Haddad and Sophie Duchesne, for their meticulous and rigorous feedback, which has greatly enhanced this work.

References

- Amozurrutia J and Marcuello-Servós C (2011) Excel Spreadsheet as a Tool for Social Narrative Analysis. *Quality and Quantity* 45(4), 953-967.
- Auld G, Diker A, Bock A, Boushey C, Bruhn C, Cluskey M, Edlefsen M, Goldberg D, Misner S and Olson B (2007) Development of a Decision Tree to Determine Appropriateness of NVivo in Analyzing Qualitative Data Sets. *Journal of Nutrition Education and Behavior* 39(1), 37-47.
- Banha F, Flores A and Coelho LS (2022) Quantitizing Qualitative Data from Semi-Structured Interviews: A Methodological Contribution in the Context of Public Policy Decision-Making. *Mathematics* 10(19), 3597. Available at: <https://doi.org/10.3390/math10193597> (accessed 24 October 2024).
- Bazeley P and Jackson K (2013) *Qualitative Data Analysis with NVivo*. SAGE Publications.
- Bree RT and Gallagher G (2016) Using Microsoft Excel to Code and Thematically Analyse Qualitative Data: A Simple, Cost-Effective Approach. *All Ireland Journal of Higher Education* 8(2), 1-14.
- Brent E (1984) Qualitative Computing: Approaches and Issues. *Qualitative Sociology* 7(1), 34-60.
- Bringer JD, Johnston LH and Brackenridge CH (2004) Maximizing Transparency in a Doctoral Thesis: The Complexities of Writing about the Use of QSR*NVIVO within a Grounded Theory Study. *Qualitative Research* 4(2), 247-265.
- Buston K (1997) NUD*IST in Action: Its Use and Its Usefulness in a Study of Chronic Illness in Young People. *Sociological Research Online* 2(3), 39-51.
- Carcary M (2011) Evidence Analysis Using CAQDAS: Insights from a Qualitative Researcher. *Electronic Journal of Business Research Methods* 9(1), 10-24.
- Carvajal D (2002) The Artisans Tools. Critical Issues When Teaching and Learning CAQDAS. *Forum: Qualitative Social Research* 3(2). Available at: <https://doi.org/10.17169/fqs-3.2.853> (accessed 24 October 2024).
- Coffey A, Beverley H and Paul A (1996) Qualitative Data Analysis: Technologies and Representations. *Sociological Research Online* 1(1), 80-91.
- Consoli S (2021) Uncovering the Hidden Face of Narrative Analysis: A Reflexive Perspective through MAXQDA. *System* 102(611). Available at: <https://doi.org/10.1016/j.system.2021.102611> (accessed 24 October 2024).
- Corbin J and Strauss A (1990) Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology* 13(1), 3-21.
- Corti L and Gregory A (2011) CAQDAS Comparability. What about CAQDAS Data Exchange? *Forum: Qualitative Social Research* 12(1). Available at: <https://doi.org/10.17169/fqs-12.1.1634> (accessed 24 October 2024).
- Cousins K and McIntosh W (2005) More than Typewriters, More than Adding Machines: Integrating Information Technology into Political Research. *Quality and Quantity* 39(5), 581-614.
- Cypress B (2019) Data Analysis Software in Qualitative Research: Preconceptions, Expectations, and Adoption. *Dimensions of Critical Care Nursing* 38(4), 213-220.

Di Gregorio S and Davidson J (2009) *Qualitative Research Design for Software Users*. McGraw-Hill Education.

Elliott N and Higgins A (2012) Surviving Grounded Theory Research Method in an Academic World: Proposal Writing and Theoretical Frameworks. *Grounded Theory Review*, 11(2), 21-32.

Evans W (2002) Computer Environments for Content Analysis: Reconceptualizing the Roles of Humans and Computers. In Burton OV (ed.) *Computing in the Social Sciences and Humanities*, University of Illinois Press, 67-86.

Evers J (2011) From the Past into the Future. How Technological Developments Change Our Ways of Data Collection, Transcription and Analysis. *Forum: Qualitative Social Research* 12(1). Available at: <https://doi.org/10.17169/fqs-12.1.1636> (accessed 24 October 2024).

Evers J, Caprioli MU, Nöst S and Wiedemann G (2020) What is the REFI-QDA Standard: Experimenting With the Transfer of Analyzed Research Projects Between QDA Software. *Forum: Qualitative Social Research*, 21(2). Available at: <http://dx.doi.org/10.17169/fqs-21.2.3439> (accessed 24 October 2024).

Fielding N (2002) Automating the Ineffable: Qualitative Software and the Meaning of Qualitative Research. In May T (ed.) *Qualitative Research in Action*. SAGE Publications, 162-178.

Fielding N and Lee R (1991) *Using Computers in Qualitative Research*. SAGE Publications.

Fielding N and Lee R (2002) New Patterns in the Adoption and Use of Qualitative Software. *Field Methods* 14(2), 197-216.

Foucault M (1994) *The Order of Things: An Archaeology of the Human Sciences*. Vintage Books Edition.

Franzosi R (1989) From Words to Numbers: A Generalized and Linguistics-Based Coding Procedure for Collecting Textual Data. *Sociological Methodology* 19, 263-298.

Franzosi R, Doyle S, McClelland L, Putnam Rankin C and Vicari S (2013) Quantitative Narrative Analysis Software Options Compared: PC-ACE and CAQDAS (ATLAS.ti, MAXqda, and NVivo). *Quality and Quantity* 47(6), 3219-3247.

Gibbs G, Friese S and Mangabeira W (2002) The Use of New Technology in Qualitative Research. Introduction to Issue 3 (2) of FQS. *Forum: Qualitative Social Research* 3(2). Available at: <https://doi.org/10.17169/fqs-3.2.847> (accessed 24 October 2024).

Gilbert L (2002) Going the Distance: 'Closeness' in Qualitative Data Analysis Software. *International Journal of Social Research Methodology* 5(3), 215-228.

Grawitz M (2001) *Méthodes des Sciences Sociales*, Dalloz.

Holsti O (1969) *Content Analysis for the Social Sciences and Humanities*. Addison Wesley Publishing Company.

Humble Á (2012) Qualitative Data Analysis Software: A Call for Understanding, Detail, Intentionality, and Thoughtfulness. *Journal of Family Theory and Review* 4(2), 122-137.

Hutchison AJ, Johnston LH and Breckon JD (2010) Using QSR-NVivo to Facilitate the Development of a Grounded Theory Project: An Account of a Worked Example. *International Journal of Social Research Methodology* 13(4), 283-302.

Jones MD and McBeth MK (2010) A Narrative Policy Framework: Clear Enough to Be Wrong? *Policy Studies Journal* 38(2), 329-353.

Kaefer F, Roper J and Sinha P (2015) A Software-Assisted Qualitative Content Analysis of News Articles: Examples and Reflections. *Forum: Qualitative Social Research* 16(2). Available at: <https://doi.org/10.17169/fqs-16.2.2123> (accessed 24 October 2024).

La Pelle N (2004) Simplifying Qualitative Data Analysis Using General Purpose Software Tools. *Field Methods* 16(1), 85-108.

Leech N and Onwuegbuzie A (2011) Beyond Constant Comparison Qualitative Data Analysis: Using NVivo. *School Psychology Quarterly* 26(1), 70-84.

Lejeune C (2016) Logiciels d'analyse de données qualitatives ou d'analyse qualitative? Pièges, limites et questions liminaires. In Kivits J, Balard F, Fournier C et Winance M (eds) *Les recherches qualitatives en santé*, Armand Colin, 184-196

Lonkila M (1995) Grounded Theory as an Emerging Paradigm for Computer-Assisted Qualitative Data Analysis. In Kelle U (ed.) *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*, SAGE Publications, 41-51.

Lowe A, Norris AC, Farris AJ and Babbage DR (2018) Quantifying Thematic Saturation in Qualitative Data Analysis. *Field Methods* 30(3), 191-207.

Maclaran P and Catterall M (2002) Analysing Qualitative Data: Computer Software and the Market Research Practitioner. *Qualitative Market Research: An International Journal* 5(1), 28-39.

MacMillan K (2005) More than Just Coding? Evaluating CAQDAS in a Discourse Analysis of News Texts. *Forum: Qualitative Social Research* 6(3). Available at: <https://doi.org/10.17169/fqs-6.3.28> (accessed 24 October 2024).

Maher C, Hadfield M, Hutchings M and De Eyto A (2018) Ensuring Rigor in Qualitative Data Analysis: A Design Research Approach to Coding Combining NVivo with Traditional Material Methods. *International Journal of Qualitative Methods* 17(1). Available at: <https://doi.org/10.1177/1609406918786362> (accessed 24 October 2024).

Mangabeira W, Lee R and Fielding N (2004) Computers and Qualitative Research: Adoption, Use, and Representation. *Social Science Computer Review* 22(2), 167-178.

Marshall H (2002) What Do We Do When We Code Data? *Qualitative Research Journal* 2(1), 56-70.

Matheson J (2005) Computer-Aided Qualitative Data Analysis Software. In Sprenkle DH and Piercy FP (eds) *Research Methods in Family Therapy*, The Guilford Press, 119-135

Meyer D and Avery L (2009) Excel as a Qualitative Data Analysis Tool. *Field Methods* 21(1), 91-112.

Mintrom M and Norman P (2009) Policy Entrepreneurship and Policy Change. *Policy Studies Journal* 37(4), 649-667.

Moncada M (2019) L'homéostasie idéale. Comprendre la stabilité de l'accès aux soins des étrangers irréguliers en France et aux États-Unis (1970-2016). PhD Thesis, University Paris-Saclay. Available at: <https://theses.hal.science/tel-02177998/> (accessed 24 October 2024).

Moncada M (2021) Innovation parlementaire, entrepreneuriat et déserts médicaux en France. *Politique et Sociétés* 40(2):187-212.

Moncada M (2023) Migration Narratives in Media and Social Media. The Case of France. BRIDGES Working Papers 5. Available at: <https://doi.org/10.5281/zenodo.7694790> (accessed 24 October 2024).

Morison M and Moir J (1998) The Role of Computer Software in the Analysis of Qualitative Data: Efficient Clerk, Research Assistant or Trojan Horse? *Journal of Advanced Nursing* 28(1), 106-116.

Moseley L and Murphy FA (1997) Applying Lexical and Semantic Analysis to the Exploration of Free-Text Data. *Nurse Researcher* 4(3), 46-68.

Ose SO (2016) Using Excel and Word to Structure Qualitative Data. *Journal of Applied Social Science* 10(2), 147-162.

Oswald A (2019) Improving Outcomes with Qualitative Data Analysis Software: A Reflective Journey. *Qualitative Social Work* 18(3), 436-442.

Palys T and Atchison C (2012) Qualitative Research in the Digital Era: Obstacles and Opportunities. *International Journal of Qualitative Methods* 11(4), 352-367.

Park C (2002) Voice Type and QSR NUD*IST: Computer Aided Qualitative Data Entering and Analysis. *English Language Teaching* 14(2), 131-166.

Park C (2004) From Qualitative to Quantitative Continuum: Exploring the Possibility with the Aid of CAQDAS (NVivo 2). *English Teaching* 59(4), 71-91.

Park C (2012) Analyzing Qualitative Data with the Aid of Computer-Based Software Program (NVivo 9). *Journal of the Korea English Education Society* 11(1), 203-222.

Paulus T, Woods M, Atkins D and Macklin R (2017) The Discourse of QDAS: Reporting Practices of ATLAS.ti and NVivo Users with Implications for Best Practices. *International Journal of Social Research Methodology* 20(1), 35-47.

Pierson P (1994) *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge University Press.

Richards T and Richards L (1991) The NUDIST Qualitative Data Analysis System. *Qualitative Sociology* 14(4), 307-324.

Rioufreyt T (2019) Réanalyser des enquêtes qualitatives à l'aide de CAQDAS. *Bulletin de Méthodologie Sociologique* 143, 77-106.

Roberts K and Wilson R (2002) ICT and the Research Process: Issues around the Compatibility of Technology with Qualitative Data Analysis. *Forum: Qualitative Social Research* 3(2). Available at: <https://doi.org/10.17169/fqs-3.2.862> (accessed 24 October 2024).

Sabatier PA (1988) An Advocacy Coalition Framework of Policy Change and the Role of Policy-Oriented Learning Therein. *Policy Sciences* 21(2-3), 129-168.

Saldaña J (2013) *The Coding Manual for Qualitative Researchers*. SAGE Publications.

Salmona M and Kaczynski D (2016) Don't Blame the Software: Using Qualitative Data Analysis Software Successfully in Doctoral Research. *Forum: Qualitative Social Research* 17(3), 42-64. Available at: <https://doi.org/10.17169/fqs-17.3.2505> (accessed 24 October 2024).

Schönfelder W (2011) CAQDAS and Qualitative Syllogism Logic - NVivo 8 and MAXQDA 10 Compared. *Forum: Qualitative Social Research* 12(1). Available at: <https://doi.org/10.17169/fqs-12.1.1514> (accessed 24 October 2024).

Siccama C and Penna S (2008) Enhancing Validity of a Qualitative Dissertation Research Study by Using NVivo. *Qualitative Research Journal* 8(2), 91-103.

Silver C and Rivers C (2014) Learning from the Learners: The Role of Technology Acceptance and Adoption Theories in Understanding Researchers Early Experiences with CAQDAS Packages, Technische Universität Berlin. Available at: <http://nbn-resolving.de/urn:nbn:de:kobv:83-opus4-44300> (accessed 24 October 2024).

Sinkovics R and Alfoldi E (2012) Progressive Focusing and Trustworthiness in Qualitative Research. The Enabling Role of Computer-Assisted Qualitative Data Analysis Software (CAQDAS). *Management International Review* 52(6), 817-845.

Snelson C, Yang D and Temple T (2021) Addressing the Challenges of Online Video Analysis in Qualitative Studies: A Worked Example from Computational Thinking Research, *The Qualitative Report* 26(6), 1974-1988.

Tak SH, Nield M and Becker H (1999) Use of a Computer Software Program for Qualitative Analyses - Part 1: Introduction to NUD.IST. *Western Journal of Nursing Research* 21(1), 111-117.

Tesch R (1990) *Qualitative Research: Analysis Types and Software Tools*. The Falmer Press.

Thompson R (2002) Reporting the Results of Computer-Assisted Analysis of Qualitative Research Data. *Forum: Qualitative Social Research* 3(2). Available at: <https://doi.org/10.17169/fqs-3.2.864> (accessed 24 October 2024).

Wanlin P (2007) L'analyse de contenu comme méthode d'analyse qualitative d'entretiens : une comparaison entre les traitements manuels et l'utilisation de logiciels. *Recherches Qualitatives* 3(3), 243-272.

Welsh E (2002) Dealing with Data: Using NVivo in the Qualitative Data Analysis Process. *Forum: Qualitative Social Research* 3(2). Available at: <https://doi.org/10.17169/fqs-3.2.865> (accessed 24 October 2024).

Wickham M and Woods M (2015) Reflecting on the Strategic Use of CAQDAS to Manage and Report on the Qualitative Research Process. *The Qualitative Report* 10(4), 687-702.

Woods M, Macklin R and Lewis G (2016) Researcher Reflexivity: Exploring the Impacts of CAQDAS Use. *International Journal of Social Research Methodology* 19(4), 385-403.

Appendix

Table 3: NVivo Requirements and My Computer's Capabilities

Component	Minimum configuration under Windows	Recommended configuration under Windows	Dell G3 15 3500-156
	NVivo 14		
Processor	2.0 GHz (2 cores)	3 GHz (4 cores)	2.60 GHz (6 cores)
RAM	4 GB	8 GB	16 GB
Available hard-disk space	5 GB	8 GB	250 GB
Screen resolution	1680 × 1050	1920 × 1080	1920 × 1080
Windows	8.1	8.1	11