The Integration of Qualitative Data Analysis Software in Research Strategies: Resistances and Possibilities

Sylvain Bourdon

Abstract: Even though dedicated qualitative data analysis programs have been widely available for more than a decade, their use is still relatively limited compared with their full potential. The proportion of research projects relying on them appears to be steadily increasing but very few researchers are known to fully exploit their capabilities. After commenting on the dynamics of user's adoption of technological innovations, this article presents an instance where a qualitative analysis program suite (QSR NVivo and QSR NVivo Merge) was not only used as an ad-hoc appendage to a traditional strategy but fully integrated in the research project, insisting on the practical details pertaining to the use of the software's capabilities. It discusses how this integration can facilitate collaborative teamwork and open the exploration of analytic dimensions difficult to envision without it.

Key words: qualitative analysis, computer-aided analysis, methodological innovation, QSR Nvivo, QSR NVivo merge

1. Introduction

The use of software for qualitative data analysis is gaining grounds in many disciplines. In spite of this, it appears that the bulk of users consider these programs as ad-hoc means of efficiently managing large quantities of data (FIELDING & LEE, 1998). The influence software can have on the analysis process is either seen as mildly positive with regard to its time saving potential or as a threat to some kind of methodological purity, distancing the researcher from the data\(^1\) or imposing some rigid and foreign framework on the analytic process. The use of software is then kept under surveillance, a large fraction of the package's capabilities are underused and the potential for innovative applications and methodological advances is impaired. [1]

This paper opens with a comment on the dynamics of technological innovation adoption and moves on to outline an example of using QSR NVivo and QSR NVivo Merge in a team research

\(^1\) See RICHARDS (1998) for a critique of this often expressed worry over "closeness to data".
process, showing how the integration of the software from the project's onset impacts on certain methodological choices and facilitates the exploration of some sociological objects. It concludes with a discussion of some factors that could facilitate the integration of software in qualitative research processes. [2]

We should note that this contribution is not about describing some best way to analyze qualitative data or to use computer analysis tools, neither is it to compare different strategies. Rather it aims at documenting some uses, situations and objects that are made more accessible by actual qualitative analysis software. This proposition for renewed research strategies rests on the conviction that a fecund and dynamic multiplicity of approaches and techniques is needed to further our understanding of our plural social worlds while each object, situation and technique needs to coherently fit together in order to produce valuable knowledge. [3]

2. User’s Adoption of Technological Innovations

Many witnessed, as I did, the first uses of word processors by the clerical personnel of their universities. Those dedicated machines were first appreciated for their ability to produce much neater layouts of documents and to let one rework or make changes to a long document before final printing. One could then see many typist write a short letter to someone, review it, print it and then delete it from memory once it had been signed and sent. A very similar letter needing to be produced days or weeks later to someone else had to be typed and verified anew. I had very limited typing skills at the time, so this quickly appeared to me as a very inefficient way of doing things. One day, I asked an expert typist why she didn’t keep the first letter in storage for later edit and reprint as necessary. I was surprised to hear a casual dismissal of this proposition on the grounds that, since she typed so fast, it was easier and quicker for her to start afresh than it was to rummage through files in order to find an old letter to use as a template. She added that, to top it off, no two letters are exactly the same anyway and that the time gained would then be minimal. Considering the limitations of the dedicated technology at the time, this could well have been objectively as well as subjectively true and reason enough to use the computer not at its full capacity but, at best, as a glorified typewriter. [4]

But, soon enough, clerical workers became more familiar with word processors. As computers became more convivial, new ways of working emerged and, most important, a new attitude toward the features they offered. It is easy to imagine the same typist then, faced with the task of sending a lot of similar letters to a list of persons, wondering if there wasn’t a way to make the computer do part of the work in her place. Of course, by then, the mail merge function had already been available for some time, thought of by a well intended software programmer, but its simple existence hadn’t changed anything by itself. The whole technology had first to be accepted and its usefulness in doing things that were already done in ways that felt familiar but with significantly better results had to be acknowledged. Only then, as familiarity and trust developed, the attitude of making the computer do a larger part of the work, albeit in a significantly different way from what was usual, could come about. [5]

I will argue that a similar process has been—and is still—taking place with the use of software for qualitative data analysis. Many features exist in actual software that are very seldom used by most researchers who haven’t gone beyond utilizing qualitative software as souped-up filing cabinets. Part of the reason for this is that familiarity with qualitative software is only starting to be relatively widespread. After a phase of questioning the very pertinence of using computers to assist the research process the discussion is only starting to bear on the practical issues of software integration (RICHARDS & RICHARDS, 1999). [6]

2 The specifics of this discussion assume a familiarity with QSR NVivo and QSR NVivo Merge. For a general introduction to the use of NVivo in qualitative research the reader is referred to BAZELEY and RICHARDS (2000) and GIBBS (2002).
3. Faster or Differently?

It seems now to be pretty generally acknowledged that the use of computers for qualitative data management often results in a considerable shrinking of the time necessary to accomplish many operations. But it should not be forgotten that the impact of this time saving isn’t solely logistical in that it would allow some piece of analysis to be produced faster. There is not a single step of analytical process which is done with current software that can’t be done manually, albeit very often with intense, dedicated and boring work. Very clever, fine tuned “systems” relying on filing cabinets, piles of transcript extracts, memo cards, color schemes, that allowed researchers to process both data and concepts in very similar ways have existed for quite a while. One could then argue that computers and qualitative analysis software are not necessarily contributing qualitatively new ways of doing things, but mostly bring about means to process data faster, more precisely and with less floor space. [7]

Yet, faster can also mean differently. The immediacy and relative painless way of accumulating those data and concept manipulations is the very essence of the transformations brought about by the software. Production imperatives, combined with what can be gathered, for lack of a better concept, under the term "human nature", prevents most, if not all researchers, from spending an unlimited amount of time with their data. In our practically limited universe, the depth of the analysis will benefit most from methods, techniques and aids that accelerate the routine tasks related to data manipulation. No matter how dedicated and available a researcher is to her data, there is a definite encouragement to explore when a matrix display of the data, for example, can be produced within a few minutes or seconds instead of monopolizing the best part of a day’s—or a week’s work. With the appropriate software, there is no need for an immense amount of time or an army of research assistants each time a hunch has to be put to empirical test. Thus, time consuming operations that were relegated to some circumscribed phases of specific types of research, like the verification of inter-coder reliability, have the potential of becoming routinely used as exploratory tools in a wider variety of contexts. [8]

4. Team Work on a Mezzo-level Sociological Phenomenon

4.1 The challenges of team work on mezzo-level social object

This section describes a tentative integrating use of qualitative analysis software that made possible an otherwise very impractical research project. Before getting into the methodological aspects, a short description of the theoretical aims and practical challenges of the project needs be exposed. [9]

The research project started in 1999 and studied various aspects of work life in community organizations. From the start it was an interdisciplinary collaboration between a psychologist and a sociologist supported by a small team of research assistants. Noticing the fast expansion of the community associative sector, both in the provision of services somehow replacing the shrinking role of the state and as an employer of young college and university graduates, the project aimed at apprehending the phenomena from two perspectives. The first rested on the worker’s point of view and put the accent on the place and meaning of their work in the community sector in regard to their own personal trajectories, life situation and aspirations. The second perspective took the organization’s standpoint in assessing the impact of this sudden growth on their mission, culture, structure and overall organization. The breadth of these goals was compounded by the great variety of community organizations and the already known fact that some characteristics of those organizations could have a great influence on the individual and collective experiences. Instead of focusing the research on one particular category of

---

3 See BOURDON (2001) for a depiction of how the computer handling of inter-coder reliability verification can contribute to transform a specialized procedure into a more routine one and a discussion on the implications of that change on global research practices.
organization to avoid the problem, a widespread research practice, the choice was made to tackle it face on and create a methodology that could address the ambitious task of exploring the multiple dimensions of what could be labelled as a mezzo-level social phenomenon. [10]

The qualitative part of the project came after a first quantitative stage of research in which more than 1000 community groups were surveyed in order to map the global trends in workforce and work conditions. From the resulting database, 43 organizations were sampled (stratified by type of service, number of workers and urban/non-urban location) in order to try to cover the widest possible varieties of situations. All those sites were then visited and all available personnel (152) were interviewed and given a questionnaire. Field notes were also taken during the brief stays (half a day to a day) at every site. The first few visits were planned at wider intervals in order to fine-tune the semi-directive interview scheme, but time and logistical constraints were such that it was not possible to do much formal simultaneous analysis afterward in order to readjust and refocus the interviews. Yet, informal discussions leading to the production of memos were ongoing during the three month intensive data collection. [11]

Once these data were collected, transcribed and verified, the challenge of analyzing them came. The material was to be analyzed as a whole, to produce a general picture and to answer the main research questions, but it also was to be used to treat specific sub-questions, planned from the onset, but slightly divergent from the main project scheme. For example, one researcher would be trying to understand what brought so many women (about 80% of the workforce) to work in community organizations and what they got out of it. Another sub-project was focusing on issues of social participation in respect to paid employment within socially-oriented organizations and yet another planned to cover more specifically the school and work histories of the college and university graduates. Those sub-analyses focused on either a specific sub-sample of the interviews, on some specific parts of each interviews or on a combination of both, which is not to say that the researchers leading them didn’t need to have access to (and knowledge of) the rest of the corpus. [12]

Thus, the material needed to be analyzed as a whole and, also, be treated so as to facilitate these more specific analyses[4]. Obviously, this could be achieved with parallel and separate analysis. But working as a team with these data could mean finding ways so that each analyst’s work can benefit others and the global project as much possible. This is where optimal use of the software came into play. [13]

The following section describes some of the main characteristics of our analytic strategy along with the software capabilities used to implement them. Even though they are separated for the purpose of presentation clarity, these characteristics interact with each other during the process so that they cannot be conceived as distinct steps or phases but need to be thought of as a whole. [14]

4.2 Theme coding as a shared basis

A more traditional, inductive approach would have had each researcher start working on a blank set of data and more or less continue that way, sharing only hunches and results with the others during seminars or team meetings, until final report production. But, in order to attain the objective of maximizing the efficiency of team working, and considering that the fields of enquiry of every researcher crosscut each other while being fairly independent, it was decided that a pre-sorting of the material collected during the fieldwork along with the possibility of sharing some of the evolving data classifications would be useful. [15]

The pre-sorting started with a very broad, thematic coding of the collected material. The codes used were neither emergent nor grounded. They were predefined, remain quite descriptive and

[4] Those sub-analyses are related to the secondary analyses of qualitative data described by FIELDING (2000) with the difference that our sub-analysis were planned before data collection and, most important here, contributed to shape the computer analysis strategy.

Copyright © 2002 FQS http://www.qualitative-research.net/fqs/
in line with the research project’s main lines of inquiry. We called them "themes" to distinguish them from the myriad of other code conceptions. [16]

The themes used in this project covered the main aspects of the semi-directive interview guide like working conditions, previous work and school trajectories, organizational structure and finances (Illustration 1). They are neither totally inclusive of all material—though very little is left uncoded by them—nor mutually exclusive, certain passages being associated with more than one theme. This approach detracts both from in-time strategies like the constant comparative analysis put forward by the proponents of Grounded Theory and from quantitative post-coding schemes where categories need to cover all possibilities while being mutually exclusive. It also departs from the codebook format (MACQUEEN, McLELAN, KAY, & MILSTEIN, 1998) in that it is not meant as a way of generating building blocks for theory. Its goal is neither immediate conceptualization nor definitive classification. It is only seen as a broad sorting of the material aimed at isolating everything of relevance on a certain theme to ease further analysis.

Illustration 1: Hierarchic tree structure of themes and analysts coding [17]

Practically, the furthering of analysis relies on the ability offered by NVivo (and some other recent qualitative data analysis programs), to work with the content of a node⁵ (a code) almost the same way as if it were a new document. This way of working has furthered the software’s ability to destructure-restructure data (TESCH, 1990). With it, the node provides a bona fide new configuration of the data that may be handled just like a reconstructed original (Illustration 2), allowing selections of data within a node to be further investigated and assigned some other code by an operation Lyn RICHARDS labelled as "coding-on"⁶. Once the material is classified by themes, the live node browser makes it possible to focus the analytic work on a specific

---

⁵ The specific terminology used in this paper regarding the software and its different functions is the one proposed by the developers. For definitions of those terms or a more general description of the program, the reader is referred to the glossary in RICHARDS (1999).

⁶ Depending on the computer configuration (mainly processor speed and amount of RAM), browsing a node referencing a very large number of passages from many documents can be very slow or bring the program to a halt. In these cases, it is possible to "split" the node in as many parts as necessary to be able to work at an appropriate speed. To do this, first create a numerical document attribute called GROUP. Then assign the GROUP attribute a value of 1 to the first 30 documents (assuming groups of 30), a value of 2 to the next 30 documents and so on. Then, before working on a particularly hefty node, it can be split into parts by running a matrix intersection between the node and the GROUP attributes values. The coding-on of each resulting node will be much faster yet the end result will be equivalent.
theme and explore it in depth, recoding as necessary and using other tools provided by the software. Within this particular project, each analyst could develop a series of specific nodes—using analytical codes instead of thematic ones now—under their own part of the tree structure (e.g. Analysts/Fred in Illustration 1); these could eventually be compared and combined using merging tools (see Section 4.4).

Illustration 2: Reconstruction of original content in the node browser. Node X gathers all passages relevant to theme X and allows the analysts to view and manipulate them as if they constituted a new document. [18]

This kind of theme strategy seems appropriate to many research situations that are not purely inductive but start with more or less well defined questions and often need to work on rather large corpuses. Evaluative research is a good, but not unique example. It is also very useful for interdisciplinary teamwork where the preoccupations of the different researchers converge on a particular setting but cover different aspects, from different points of view. One can object that, in these cases, a good deal of depth and specificity in exploring a phenomenon can be lost during data collection by trying to accommodate different themes. This loss in specificity has to be weighted against the gain of scope and the richness brought about by the diversity of points of view. It allows for better exploration of patterns of differences amongst many cases and facilitates the integration—and confrontation—of multiple perspectives on a particular object. On the other hand, it will not allow as deep an examination of the richness of individual experiences. Each analysis will also lose some of its depth and specificity. [19]

4.3 Managing flat background data

Even though the qualitative portion of our research project relies partly on an inductive process, it has a clear focus on a relatively limited set of objects. Thus, a fair portion of what is taken into account as "data" is not meant to be examined—or deconstructed—as thoroughly as the rest and it has to be considered from the start as background data. For example, when labelling an individual as a male, one is implicitly opting not to explore the depth and intricacies of gender identification or the masculinity phenomenon, which are thus put outside the scope of exploration of the qualitative analysis; only the superficial signification of gender is retained in a classificatory scheme against which the foreground phenomenon can be understood". This labelling readily "flattens" some part of the possibly available information, discarding its complexity while making it simpler to manipulate. There is a difference between the intricacies, depth and scope of gender self-representation and the simple dichotomy of male/female, but there is also a difference between studying gender representation and any other phenomena that can or could be affected by gender as a "context". [20]

---

7 Of course, it is possible to use both the flat labels (male/female) and the qualitative material in order, for example, to explore the differences of gender identification between women and men.
In NVivo, these flat data are handled as attributes. In our project, attributes had classificatory function similar to themes. But as themes classify rich data (transcripts, notes) by their content, we used attributes to classify it by its context. Thus, the transcript of the interview with Mary, a worker at organization X, could be linked both to the attributes of Mary (women, 25 …) and those of her work setting (8 employee organization, urban setting …). With this possibility in mind at the start of the project, we collected a wide range of background flat data from each interviewed participant, both directly through a self-administered questionnaire including traditional socio-demographic variables and a few standardized instruments and indirectly from a preliminary survey from which it was possible to associate organizational characteristics to individual workers. These data, which had been input in a statistical package (SPSS) for quantitative analysis, were swiftly imported, through the Import Attribute function, into the NVivo project where they could be used to split the corpus into groups for analytical exploration. The fact that thematic coding is very broad, often including large chunks of text, makes these intersections much more likely to bring together substantial amounts of material than would more discrete categories. [21]

The analysis of the school and work histories of the college and university graduates, for example, could start by a search collecting all the material coded at the theme (node) "Schooling-training history" or at the theme (node) "Work history" which was also associated with the attributes DEGREE=College and DEGREE=University into a new node called "School and work grads". The restructuring of this material so that it could be brought together and isolated from the rest allows the researcher to focus on this particular topic. From there, further splitting can be made to compare the men and the women, for example, or the focus can be widened and comparison can be extended to the non graduates. [22]

Adequate handling of these data is essential in a mezzo-analysis resting on the heuristic potential of inter and intra group comparison. From the start of the project, our formal knowledge of the literature and our implicit knowledge of the terrain suggested many ways of organizing the grouping of community organizations (size of organizations, rural or urban setting, date of creation, staff composition …), but none was evidently more heuristic than the others. We thus needed to be able to shape different types of groups in order to assess what Gregory BATESON (1972, p.315) qualified as "a difference which makes a difference" (or, in our multidimensional scheme, the differences which makes the differences). [23]

Until the recent introduction of attribute tables in qualitative analysis software, researchers had been using various strategies to separate flat data (often socio-demographics) from rich ones (transcripts, field notes …), often through complicated procedures based on embedding keywords in documents, manually or automatically searching for them and coding the documents in which they were present to some "variable" node (e.g. gender/male). But the recent integration of spreadsheet-like facilities to handle those flat data is a definite evolution. It allows much easier separation between context data and the object of research while offering a lot of flexibility in the definition of comparison groups. For example, the ability to input the age of workers as a number (like 25) instead of a range (18-30) as was usually done when using codes to handle flat data in order to avoid the awkwardness of handling dozens of different age categories, greatly facilitates (and thus encourages, as we suggested before) the fine-tuning of the group ranges as the analysis goes along instead of freezing them once and for all at the onset. In the context of teamwork, it is then much easier to take advantage of a grouping scheme elaborated by a co-worker from her angle of analysis to explore the data from a different angle; merging projects (Section 4.4) makes this sharing even more efficient. [24]

### 4.4 Project merging for team synergy

Section 4.2 discussed the use of theme coding as a shared basis. The initial phase of analysis entailed the construction of a first generation Master Project including transcripts of all the interviews and field notes. Every passage of every document was then coded at the appropriate theme or themes and the attributes relating to each interviewee and their context imported and
linked to the related material. Copies of this master project were then made for each analyst who could then use the node browser to focus their work on the material pertinent to their specific interest. At this point, even though a gain in efficiency is actually achieved, it is somehow marginal. The real influence of team synergy is reached when the results of everyone’s work can be brought again together for the benefit of all. This is done through the QSR Merge software. [25]

Using Merge, all the specific coding, the memos and the attributes of two analyst’s projects can be brought together in a unified project. It gives the possibility to put together not only the end—or partial—results of two analyst’s work, but also the coding, memos and attributes they developed during their investigation. When two emergent categories come out as "generation clash" and "old and new culture", it is then possible to go back to the empirical traces (passages of documents coded at the nodes) that inspired the analysts and compare them side by side as easily as it is to compare their respective theoretical memos. When, after discussion, these can be combined in a new understanding, the respective nodes and their analytical memos can be merged and moved to a common part of the node structure. [26]

After each merging, the unified project can then be redistributed to become everyone’s basis for further investigation so the common knowledge base is co-constructed along with the individual ones. [27]

The potential offered by the ability to merge projects, mostly in the realm of much more integrated team work—a recognized but seldom met necessity these days—, is immense. DI GREGORIO (2001) even describes an example of a team working together at a distance. Though these kinds of collaborations are not totally new, what is offered here is a possibility to engage in collective—or at least thoroughly shared—analysis, breaking the traditionally isolated—or at best parallel—mode of qualitative enquiry. [28]

5. New Tools, Renewed Exploration

Qualitative analysis software like QSR NVivo and QSR NVivo Merge hold the potential to open new possibilities for collectively working on some particularly wide objects. For this, and other methodological innovations made possible by these tools to happen, their role has to shift from handy utensils to fully integrated parts of the very design of research projects. This paper has described some elements of an attempt to reach that integration. It is our conviction that more examples like this need to be brought to the attention of the research community. Yet, following our comment on user adoption of technological innovations, we would also suggest that the full impact of software on qualitative methodology is yet to be seen, at least in sociology, and that the familiarization phase is still going on for most users. [29]

Creative use of method—a part of sociological imagination (MILLS, 1959)—, is a key to good research in general and pertinent analysis in particular. It rests upon ample disciplinary and methodological culture and true mastery of the fundamentals. With the advent of qualitative analysis software, it also rests on the effort of developers to make their tools as convivial as possible and on a collective acceptance of the medium and its possibilities, which is an implicit goal of this contribution. [30]

Acknowledgements

The author would like to thank the Social Science and Humanities Research Council of Canada for its support in the production of this paper.
References


Copyright © 2002 FQS http://www.qualitative-research.net/fqs/
Author

Sylvain BOURDON is professor of education sociology and research methodology at Sherbrooke University (Quebec, Canada). His research focuses on the social aspects of the school—work relationship, particularly on youth’s transition from school to work. His current work tackles such aspects of this field as youth’s work in community organizations, marginalized youth and the written text and the evolution of training practices in organizations aiming at helping youth employment. He also acts as a consultant with many research teams on the integration of computer in the qualitative research process and he trains graduate students and researchers in the use of QSR Nud*Ist and NVivo software.

Contact:
Sylvain Bourdon
Faculté d’éducation
Université de Sherbrooke
Sherbrooke (Quebec) Canada J1K 2R1
Email: sbourdon@courrier.usherb.ca
URL:
http://callisto.si.usherb.ca:8080/cro/sbourdon.htm

Citation

Please cite this article as follows (and include paragraph numbers if necessary):