Computer-Assisted Qualitative Data Analysis Software: An Illustration of Limitations and Advantages

Abdulaziz Ahmad
Manchester Business School, Abdulaziz.ahmad@postgrad.manchester.ac.uk

Mike Newman
Manchester Business School, MNewman@dom01.mbs.ac.uk

Follow this and additional works at: http://aisel.aisnet.org/amcis2010

Recommended Citation
http://aisel.aisnet.org/amcis2010/134

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Computer-Assisted Qualitative Data Analysis Software: An Illustration of Limitations and Advantages

Abdulaziz Ahmad
Manchester Business School
Abdulaziz.ahmad@postgrad.manchester.ac.uk

Mike Newman
Manchester Business School
MNewman@dom01.mbs.ac.uk

ABSTRACT
Computer-assisted qualitative data analysis software (CAQDAS) is a well-known acronym in qualitative research. Nowadays, qualitative researchers are inclined to apply such software for various purposes. Critiques of the use of CAQDAS in qualitative research focus on the term “analysis”, claiming that the tool does little to analyze data. According to these critiques, some users of CAQDAS advocate a positivist or quantitative approach using CAQDAS as a so-called “devil-tool” for science. In contrast, although the use of CAQDAS can be seen to do the mechanical part of the analysis (i.e. coding), it can never fulfill the conceptual part of the analysis which requires a human-touch. Thus the better use of CAQDAS is for coding, not for analysis. This research paper attempts to illustrate the challenges of manual versus computer assisted coding by identifying events that create gaps during an ISD change process using a grounded theory methodology (GTM).

Keywords
Grounded Theory Method (GTM), CAQDAS, PSIC Model, Computer assistance, Coding

INTRODUCTION
Computer-assisted qualitative data analysis software (CAQDAS) is a popular acronym in qualitative research. Nowadays, qualitative researchers are much more inclined to apply such software. Critiques over the use of CAQDAS advocate the positivist or quantitative approach using CAQDAS as a so-called “devil-tool” for science (Lee and Fielding, 1996). In contrast, although the use of CAQDAS can be useful in doing the mechanical part of the analysis (coding) it can never fulfill the conceptual part of the analysis which requires a human-touch (Thompson, 2002). Thus it is argued that the better use of CAQDAS is for coding, not analysis (Coffey, Holbrook and Atkinson, 1996).

The phrase grounded theory (GT) refers to the theory that is being developed inductively from data. Grounded theory method (GTM) thus dictates the approach being used in establishing this theory. In this research paper, the aim is to understand / explore how the use a CAQDAS namely NVivo8 partnered with the application of a GT approach enable us to understand the socio-technical process in ISD change.

CAQDAS provide more than just categorizing. It enables actual analysis of data such as automating coding using sections (Bringer, Johnston and Brackenridge, 2006) and querying of data inter alia. This research paper will focus on the use of CAQDAS as assistive software for qualitative data rather than advanced analysis software. In this context, assistive refers to the coding of data and seamless management of data during analysis of qualitative data.

We will also employ Lyytinen and Newman (2008) as our vehicle to discuss ISD change processes: “the major challenge is identifying events that create gaps, that is, how one can detect from a huge stream of changes, events that truly influence the system”. Thus our research question is:

RQ1: How can the combination of use of CAQDAS and GTM be used to deepen our understanding of an ISD change process?

The remainder of this research paper will be organized as follows: following this introduction, we present our understanding on GT and the application of computer-assisted qualitative data analysis software. In the subsequent section, we discuss the method applied in this paper. This will lead to the research findings and continue with the discussion section. This paper concludes with our conclusions and identification of future research.
CAQDAS AND GTM

It is claimed that computers do not and cannot analyze qualitative data (Roberts and Wilson, 2002). Although the use of computer assisted qualitative data analysis software (CAQDAS) can be seen as assisting the process of data collection, management, storage and retrieval of qualitative data, the gist of qualitative research analysis lies in understanding the meaning of text (Kelle, 1995). Thus the use of CAQDAS should only support rather than replace manual data analysis. It has, in one sense, encouraged prescriptive analytical methods which inhibit interpretation and creativity (Dey, 1993). The better approach would be to read, re-read and achieve familiarity with the data: i.e. the essence of qualitative data analysis (Strauss and Corbin, 1990).

The use of CAQDAS has become widely acknowledge by researchers in all areas. As usual there are those who support the use of CAQDAS identifying its benefits through time-saving (Lee and Fielding, 1991; Moseley et al., 1997), the ability of the researcher to analyze larger sets of data (Kelle and Laurie, 1995; Bowling, 1997; Webb 1997), making qualitative data more scientific (Kelle and Laurie, 1995; Webb, 1999). This enables researchers to concentrate on more creative and interpretive tasks (Morison and Moir, 1998) and exercise more substantive analysis (Moseley et al., 1997). As a result, it enhances the quality of the analysis (Lee and Fielding, 1991).

At the same time, there were those who criticized such usage through their concerns over the application of CAQDAS. One of the main concerns was that the use of CAQDAS has distanced the researcher from their data through poor screen display, segmentation of text and caused the researcher lose the context of their data (Bazeley, 2007; Morison and Moir, 1998). In one sense it has alienated the researcher from their data (Kelle, 1997; Webb, 1999). It is another concern of some researchers that the coding and retrieval methods offered by software packages dominate the qualitative analysis to the exclusion of other analytic activities (Bazeley, 2007; Coffey et al., 1996). Another concern is the fear that the use of computers will mechanize analysis and produce biases towards quantitative or positivist approaches (Bazeley, 2007). Thus it is claimed that CAQDAS has mechanized the process and marginalized the reflection of the researcher (Morison and Moir, 1998). There is also the misperception that computers support only GTM especially for the coding process (Bazeley, 2007; Webb, 1999), to the extent that some critics claimed that the developments of CAQDAS are solely focused on GT (Webb, 1999). According to Lonkila (1995), coding is necessary for all qualitative data analysis methods and not only the GT approach. In addition, it was found that only 30% of studies using CAQDAS are applying GT method and another 70% are using other methods of analysis (Lee and Fielding, 1996).

The use of CAQDAS has also limited the application of manual processes in qualitative data analysis. According to Kelle (1995), using word processors has greatly improved the analysis process compared to the traditional “cut and paste” method of coding and retrieving information. But it also creates a methodological problem since the text is removed from its context. But using CAQDAS, it has further removed the issues of manual coding (cutting, labeling and filing) and also the boundaries which limited paper-based marking and sorting of text. The use of CAQDAS has been criticized because segments of text are removed from the whole thus creating a loss of perspective. However the critics failed to identify that the use of paper (cut and paste) amounted to the same thing with a greater risk of not identifying the source of the segment (Bazeley, 2007).

As above, CAQDAS has been criticized to favor the GT approach (Coffey et al., 1996) but this was countered by Lonkila (1995) and Lee & Fielding (1996). In the GT approach the process of coding is the fundamental task that needs to be satisfied first before any further processes can be carried out. According to Strauss and Corbin, the first stages of analysis are the identification and the development of concepts or open coding (1990). Coding is a process of identifying a passage of text which the researcher interprets as having a particular meaning (Webb, 1999) or also known as ‘sign-posting’ by Kelle (1997) or as ‘referential’ by Richards & Richards (1995). In applying NVivo8 (a well-known example of CAQDAS), this can be achieved through creation of free nodes. These free nodes are then re-coded or attached to a parent code as a means of creating hierarchies which supports the GT approach. In GT approach, open codes (free nodes) are amalgamated in ‘higher’ levels of codes until a unifying core concept is identified (Webb, 1999).

Understanding these limitations and advantages of CAQDAS, many researchers use CAQDAS for the purpose of coding alone and going back to the manual methods for the theory building stage (Lee and Fielding, 1996) thus ensuring that the benefits of automation are reaped and its shortcomings are avoided.
RESEARCH METHODOLOGY

Qualitative Research in Information Systems

Qualitative research methods are designed to help researchers understand people and the social, political and cultural context within which they live. Interpretive studies attempt to understand IS phenomena through the meaning that people assign to them and the process where IS are influencing or have influenced the context (Walsham, 1993).

The qualitative research methodology is said to enable a detailed observation of and involvement of the researcher in the natural setting in which the study occurs (Kaplan and Duchon, 1988). For the purpose of this research, we are going to employ case study research. Yin (2002) (Yin, 2002) has defined the scope of a case study as an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident. A strong case study tradition already exists in the field of IS management (Myers, 1997; Robey and Newman, 1996; Orlikowski and Baroudi, 1990).

Case Description / Site Selection

This research paper aims to study the IS development in three case studies in Malaysian universities, Case 1, Case 2 and Case 3. Each of these cases has their uniqueness and similarities with the others. For the purpose of this paper we are going to focus on Case 3 alone.

This university was established in 2002, and is comprised of an amalgamation of nine institutes which were previously managed individually. The diverse operational processes among these previously autonomous institutes brought about a call for an integrated management system (IMS) similar to that adopted in another university (Case 2) with the intention of enhancing their operational efficiency. With thirteen thousand students supported by two thousand de-centralized staff members, the establishment of a single integrated database was envisioned to provide a more efficient management of resources. Initial data analysis in this two-year project showed that they have experienced the greatest number of problems encountered in IS development projects compared with the other two cases (Case 1 and 2). For our study we will focus on one component of an integrated management system, the integrated finance module.

Data Collection and Analysis

The vast literatures on grounded theory (GT) have created confusion for a novice researcher. The overly divergent understanding on the theory has made matters worse. GT concepts, features and procedures are intertwined during discussions (Soja and Paliwoda-Pekosz, 2009; Elharidy et al., 2008; Esteves et al., 2007; Strauss and Corbin, 1994). For the purpose of this research, we will apply a GTM as an iterative process (Elharidy et al., 2008) between open coding, axial coding and selective coding (Corbin and Strauss, 1990) or a constant comparative method (Glaser, 1965).

Interviews entail types of questioning that range from using open ended questions to closed questions. Semi structured interviews was conducted for the purpose of this research. The ability to listen, take notes and provide feedback requires skills which only be developed during the actual interview process itself. Applicable techniques such as the critical incident technique (Chell 2004) ensures a controlled interview session and methods such as replicating the interview process like a drama (Myers and Newman 2007) must be used to ensure efficient interview process. After the interview, a snowballing technique was used to ensure continuity of the story within the case under study. During these interviews, the researcher asked several main questions and probed further based on their responses in order to get an in-depth knowledge of the possible phenomena occurring.

To date, three phases of interviews have been carried out with at least twenty-two interviews at Case 3. The interview respondents for each of the phases consisted of multiple stakeholders involved in the project. It included the users, the vendor and the project managers.
Ahmad and Newman  CAQDAS and GTM


Table 1: Number of interview respondents by phase.

<table>
<thead>
<tr>
<th></th>
<th>Phase 1 (July 2008)</th>
<th>Phase 2 (February 2009)</th>
<th>Phase 3 (September 2009)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Vendor</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Project Managers</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>14</td>
<td>4</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 2: Ten-step data collection and analysis process

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Interview sessions (complete 2 phases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>Transcribing interviews (direct transcriptions)</td>
</tr>
<tr>
<td>Step 3</td>
<td>Translating transcripts (full English)</td>
</tr>
<tr>
<td>Step 4</td>
<td>Paragraphing (according to main ideas, supporting ideas and examples)</td>
</tr>
<tr>
<td>Step 5</td>
<td>Create general coding /initial coding process and categories</td>
</tr>
<tr>
<td>Step 6</td>
<td>Printing, cutting and pasting interview transcripts based on category / theme</td>
</tr>
<tr>
<td>Step 7</td>
<td>Establishing general / overall mind map of the transcript</td>
</tr>
<tr>
<td>Step 8</td>
<td>Dissect mind map according to 4 main stakeholders (vendor, top management, users, project managers)</td>
</tr>
<tr>
<td>Step 9</td>
<td>Re-establishing mind maps according to these stakeholders</td>
</tr>
<tr>
<td>Step 10</td>
<td>Re-analysing of data for identification of pro and con arguments</td>
</tr>
</tbody>
</table>

All of these interviews were digitally recorded and a double transcription was carried out, due to the bilingual nature of the interviews. It meant that a two level transcription process was conducted. This is actually part of the challenges faced during this data collection process. The use of bilingual interview process ensured that the respondents are able to better express their feelings and thoughts in relation to the development of the enterprise system. During the second transcribing process, from the Malay language to English language, the researcher attempted to find the most suitable words that reflected the respondent’s original thoughts. Completing the transcriptions was very time consuming but it enabled the researchers to really grasp the data.

The preparation of the interview transcripts consumes most of the time during the overall analysis process. In this process, the researcher is looking at each of the transcripts with a view to finding a story to tell. Therefore in order to understand the story line much more clearly, the researcher created a paragraph to differentiate between the different main ideas. Each paragraph of main ideas would have a sub paragraph of supporting ideas and followed with an example, if available. Once this process was completed, the researcher began the initial coding of the transcripts.

<table>
<thead>
<tr>
<th>Grounded Theory Method (GTM)</th>
<th>Manual</th>
<th>CAQDAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open coding</td>
<td>Creation of codes based by paragraph – description / summary of paragraph</td>
<td>Creation of codes as free nodes by sentence or paragraph – description / summary of sentences or paragraph</td>
</tr>
<tr>
<td>Axial coding</td>
<td>Re-reading of codes generated and re-arrangement according to theme / category – cutting and pasting</td>
<td>Re-reading of codes / free nodes and re-arrangement according to theme / categories / tree nodes – Creation of hierarchies by “drag” and “drop”</td>
</tr>
<tr>
<td>Selective coding</td>
<td>Re-reading of codes and categories and selection of category that most represent the cumulated categories.</td>
<td>Re-reading of codes and categories and selection of category that most represent the cumulated categories. Higher hierarchies of the tree nodes are established to show the selected codes.</td>
</tr>
</tbody>
</table>

Table 3: Summary of comparative analysis process – based on Strauss and Corbin (1990) GTM
The above table follows the GTM by Strauss and Corbin (1990). Generally, the manual process and the use of the CAQDAS process are similar and here CAQDAS has an advantage. All the benefits discussed earlier are being experienced by the researcher. For a novice researcher, the drawbacks lie on the understanding of GTM and the time taken to explore the CAQDAS functionalities. Once the researcher had managed to grasp both the method and the software’s functionalities the analysis was straightforward.

FINDINGS
The application of GTM in this particular study has enabled critical issues to be dissected and probed, sliced and diced. Together with the use of the NVivo8 software package, the process of coding was completed more efficiently. In this particular research paper, we are examining the implementation of enterprise system in a university which was recently being merged with nine other institutes. Therefore, the ultimate reason for such an endeavor was to streamline the operational processes between these institutes but the findings shows that the most critical process during the development was the requirements gathering process.

<table>
<thead>
<tr>
<th>1st level category</th>
<th>2nd level category</th>
<th>3rd level category</th>
<th>Concept</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business requirement session</td>
<td>Project Level</td>
<td>Communication</td>
<td>User’s limited communication with developers</td>
<td>We communicated with them during sessions and after sessions. But the only thing is that they are located so far away, we even asked them to locate themselves in our office so that anything unclear can be discussed. But they rejected. U5, phase 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limited channel of communication</td>
<td>At the user level is more difficult; we can only contact them during session since they are busy with their daily operations. And there’s no time to have an informal gathering. Plus there’s no other channel of communication. V1, phase 2</td>
<td></td>
</tr>
<tr>
<td>Control and power</td>
<td>Process requirement according to superiors preference</td>
<td></td>
<td>In Case 3, they are not sure what they do. The direction from their bosses keeps changing day to day. They live by the moment, they don’t have target. And their staff are not enthusiastic about their work. It should not happen during requirement session, but it was being transmitted to us. We see that this thing happens and they acknowledge that it is happening. V7, phase 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control over requirement session</td>
<td></td>
<td>…for me, sometimes it is difficult for them to accept my expression, so in the end… for it is better for a lesser people in the sessions. Lesser people lesser problem, it will not be a problem since there are less people to convince. U1, phase 1</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>User’s knowledge on accounting process</td>
<td></td>
<td>If they are comparing report from the same subsystem of course they will get a same figure. But now they are trying to compare one from transaction details (subsystem) and with customer ledger (GL) (require posting). Of course there will be a timing difference between these two reports. But they want it to be the same. Since their old system can give same figures. V1, phase 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>User’s limited knowledge on IT</td>
<td></td>
<td>I do not blame them but I think their IT knowledge is still limited. Since this is a new thing for them. I do not see anyone who knows how the system is being integrated. They should have an idea on how the systems should look like in the first place. They just want to see everything that was being done manually in the screen.</td>
<td></td>
</tr>
</tbody>
</table>
Since our system in a bit up to date, we have to customize to follow their requirement. V4, phase 2

Table 4: Samples of coding

The above table shows the analysis of data applying computer assistive software based on a manual analysis. The analysis followed the GTM advocated by Strauss and Corbin (1990). The interview transcripts were dissected into paragraph and coding process was conducted to produce concepts. This was followed by the “cutting” of these coded transcripts into stripes and “pasting” it according to categories / themes. The critiques have considered this process as the researchers is moving away from the source data that can cause loss of perspectives.

According to Strauss and Corbin (1998), concept is “a labeled phenomenon… an abstract representation of an event, object, or action/interaction that a researcher identifies as being significant in the data”. Further, concepts with similar or shared properties are identified as categories (Bazeley, 2007).

![Tree Nodes](image)

Figure 1: Image capture NVivo8 / Screenshot NVivo8

In NVivo8, coding is stored in nodes where nodes are created for each topic or concept. In each node, NVivo8 will store the reference to the segment of the data from the source and not the exact extraction (Bazeley 2007). This ensures that the source document will always remain intact for future use. During open coding, data were dissected into categories. These were done through creation of nodes: free nodes. These categories are a description of the data. These free nodes / categories were further re-analysed and combinations of these categories created a higher level category or in NVivo8 terms, tree nodes are created. These categories emerged from the combination of the initial categories itself. The naming of these categories was based on the issues that were being described. The beauty of using NVivo8 is that it assists the management of data seamlessly: just by using the “drag” and “drop” function.
In general, the results of analysis using manual process and CAQDAS are similar. During the business requirements sessions (BRS), there were several issues that emerged from the data. The first issue relates to project communication. According to the users, the communication with the vendors was limited as they were available during sessions only. The vendors also rejected the proposal for them to be stationed at the users’ office. In contrast, the vendors complained that the users could be contacted during sessions only and they (the users) ignored any informal gathering after office hours. From these statements, it shows here that there were high occurrences of communication breakdown in the project team itself.

The second issue relates to the decision making control and power structures during the BRS. Users were not able to make decisions concerning their own processes. Most of time, their understanding of the processes clashed with their superiors’ thus, they relied on their superior to make the final call on the processes and they tried to live with it. In another instances, during BRS, the superior only invited a small number of representatives from campuses in order to ensure less conflict.

The third issue corresponded with the level of knowledge during BRS. The users were requesting reports similar to their legacy systems outputs although they were not consistent with the nature of fundamental accounting principles. Due to their lack of knowledge of integrated systems, the users just requested to automate their previous manual processes rather than re-engineering these processes.

Applying Lyttinen and Newman (2008) and their PSIC model, antecedent conditions, outcomes, context and critical events plus at which level they occur, are best identified using GTM coupled with NVivo8. Due to limitation of space, please refer to Lyttinen & Newman (2008) for detailed construction of the PSIC model. Applying this model, BRS corresponds with the events project level of the ISD process and users can be identified as the people / actors elements in the Leavitt’s socio-technical model.

Figure 2: PSIC model (Lyttinen & Newman 2008)
DISCUSSION

We begin the discussion by revisiting our research question raised in the beginning of the paper:

**RQ1: How can the combination of use of CAQDAS and GTM be used to deepen our understanding of an ISD change process?**

In this research paper, we are attempting to understand one of the many aspects of our data through the application of GTM using interview data from multiple visits.

The two-year enterprise system implementation project ended with the vendor walking out of the project site. Based on the interviews conducted, there were many reasons why the vendor left the project site. These ranged from commercial to emotional reasoning. We were very fortunate to get continual access over the case under study and also access to the respondents especially the opinion leaders.

These applications enable critical events to emerge from the data. In a process based research, the most critical task is to identify events that can best depict a phenomenon. Using mere eye-balling techniques, although we are able to gain a certain degree of understanding of the data, it is usually on face-value alone. As mentioned by Lyytinen and Newman (2008), “the major challenge is identifying events that create gaps, that is, how one can detect from a huge stream of changes, events that truly influence the system”. The application of GTM proposed by Strauss & Corbin (1990) has helped in facing this challenge. Events emerged from the data through the coding processes.

Although initially, these were mere codes describing the data, further analysis of these codes (axial and selective coding processes), themes / categories emerged from the combination of these codes. Thus we could establish critical events based on the researchers’ theoretical understanding. In this research, we manage to identify issues concerning communication, control and power, and knowledge which are vital during the business requirement session (BRS) process. Based on this alone, the researcher was able to establish whether this process can be identified as an event. The application of CAQDAS in GTM is not only seen as capable in identifying events that creates gaps but is also able to establish the reason why such events occur.

The term “computer-assisted” in CAQDAS has really shown its advantages. During the analysis of the interview data, the use of Nvivo8 has assisted in managing vast data sets through creating and re-arranging codes or nodes. Codes / nodes are able to be viewed with the extract of source data. In addition, identification of critical events can be done through creation of tree nodes in the form of hierarchies. Applying the multilevel PSIC model, using Nvivo8, it can also assist in reallocating events according to level under study.

CONCLUSION

In this research paper, we are attempting to understand the complexities of an enterprise systems project through GTM as advocated by Strauss and Corbin (1990). The application of CAQDAS has minimised the burden of handling vast qualitative data. As a methodological research paper, the application of GTM has highlighted further interesting research areas plus it tries to mitigate the challenges of identifying critical events during ISD change process. Our early understanding of the method has created certain limitations for the analysis process but our theoretical approach has improved the interpretation of the data. Applying the GTM through an interpretive lens has improved the understanding of phenomena from the viewpoints of the participants (Elharidy et al., 2008). This research confirms the benefits of GT in interpretive information systems research, by developing context-based, process-oriented descriptions and explanation of phenomena (Myers 1997, Urquhart 2001).

Further research on this would include identification of a meso level of explanatory power during ISD change process further extending Lyytinen and Newman (2009)’s PSIC model.
REFERENCES