Qualitative Analysis Software Applied to IS Research - Developing a Coding Strategy

Brian O’Flaherty
University College Cork, boflaherty@afis.ucc.ie

Jason Whalley
University of Strathclyde, jason@mansci.strath.ac.uk

Follow this and additional works at: http://aisel.aisnet.org/ecis2004

Recommended Citation
http://aisel.aisnet.org/ecis2004/123

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
QUALITATIVE ANALYSIS SOFTWARE APPLIED TO IS RESEARCH - DEVELOPING A CODING STRATEGY

O’Flaherty, Brian, Business Information Systems, University College Cork, Ireland. Email: boflaherty@afis.ucc.ie

Whalley, Jason, Management Science, University of Strathclyde, Graham Hills Building, 40 George Street, Glasgow, Scotland. Email: jason@mansci.strath.ac.uk

Abstract

In the past two decades, the IS research community underwent an intensive debate and fought hard to legitimise interpretive and qualitative research paradigms. Now that the war is over it is ironic that despite a knowledge of technology the IS literature has been slow to embrace software in order to support qualitative data analysis. In the broader social science field, a range of software support tools have emerged offering diverse functionality and a developing critical mass of appropriate literature. This paper sets out to consider how IS research can embrace QAS.

Qualitative research has three distinct phases, namely data collection, data reduction and data display, with the later pair being most suitable for enhancement by QAS. Coding is central to QAS, but the IS field has been slow to develop rigorous coding schemas. Are there any frameworks within the IS literature that could be applied to such a task? The Qualitative Analysis Software must not be used without consideration for the research philosophy context, as a ‘package-led’ orthodoxy would regress the attempts to encourage quality research. Finally, the logical tests for measuring the quality of research (Yin, 1994) are reviewed and alternative tactics enabled by QAS are proposed.

Keywords: Qualitative analysis software, coding schemes.
1 INTRODUCTION

According to Markus (1997) the ‘War’ between quantitative and qualitative research is over in the IS field. Qualitative research has ‘won academic acceptance, both within the IS field and within the larger domain of academic management studies’ (Markus, 1997). Despite the successful paradigm shift the IS research community has been slow to adopt Qualitative Analysis Software (QAS) supporting the analysis phases of the every increasing number of interpretive studies being undertaken in the field.

This paper proposes a novel data analysis approach that is appropriate to IS research. Firstly, the nature of qualitative analysis is explored and coding, the common data reduction technique, is outlined. The coding techniques that have been adopted in the IS literature is grounded theory. The IS literature, in comparison to disciplines such as Education, has limited examples of coding schemas. The ‘context-process analysis’ framework proposed by Walsham (1993) is suggested as a basis for developing coding strategies that are suited to the requirements of IS research.

2 QUALITATIVE DATA ANALYSIS

Traditionally, the literature on qualitative research was weighted towards issues of data collection (Miles, 1994). Data analysis is defined as ‘the process of systematically searching and arranging the interview transcripts, field-notes, and other materials that you accumulate to increase your understanding of them and to enable you to present what you have discovered to others’ (Bogdan, 1984). Miles and Huberman (1994) describe data analysis as consisting of three activities, which occur concurrently. Firstly, data reduction refers to the process of selecting, simplifying, abstracting and transforming the raw case data. Secondly, data display refers to the organised assembly of information to enable the drawing of conclusions and conclusion drawing/verification involves drawing meaning from data and building a logical chain of evidence.

Data reduction is not separate from data analysis activity, but part of it, and is on going through out the analysis process. It can occur through the selection of a conceptual framework and is achieved via summaries, coding, teasing out themes, making clusters, making partitions and writing memos (Miles, 1994). A data display is defined as ‘an organised, compressed assembly of information that permits conclusion drawing and action (Miles, 1994)’. An extended text is ‘cumbersome’ for qualitative data display and the design of displays is an analytical activity in it’s own right, as well as, a form of data reduction (Miles, 1994). From the start of analysis the researcher is trying to determine what things mean by attempting to identify ‘regularities, patterns, explanations, possible configurations, causal flows, and propositions’ (Miles, 1994).

2.1 Coding Perspectives in IS literature

A central approach to data reduction and data display is the process of coding, which was first used by sociologists and appeared in the Chicago School in the late 1920’s. Around this time case descriptions were also emerging as a method of data reduction. In marketing research the use of coding for classifying responses to open-ended questions on surveys appears to have been in use since 1937 (Fielding and Lee, 1998). Rensis Lickert led the program, which was committed to open-ended interviewing and used a simplistic coding scheme to analyse the verbatim transcriptions. Miles and Huberman (1994) defines Codes as ‘tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study.’ Types of codes includes 1) descriptive codes, which involves little interpretation, 2) interpretative codes, representing ‘a backstage web of motive’, and 3) pattern codes represent an emergent pattern and are used in the later in the course of data collection (Miles and Huberman, 1994).
Bogdan (1984) outlines a coding category development approach, involving several steps. Typically you search through your data for regularities and patterns as well as for the topics that your data cover, and then you write down words and phrases to represent these topics and patterns. Bogdan and Biklen (1984) propose families of codes, which suggest classifications such that coding can be accomplished (Figure 1.). These coding categories are used in the education field and in particular as a means of data reduction in the qualitative data analysis process. There is no similar coding classification evident in the IS literature comparable to the above listed table of coding categories. This raises the question ‘is there a case for developing a coding categorisation specifically for the IS field?’

<table>
<thead>
<tr>
<th>Code Family</th>
<th>Code Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting/Context Codes</td>
<td>General information on the setting, topic, or subject can be sorted.</td>
</tr>
<tr>
<td>Definition of Situation Codes</td>
<td>Place units of data that tell you how the subjects define the settings or particular topics. The subjects world view.</td>
</tr>
<tr>
<td>Perspectives Held by Subjects</td>
<td>Shared rules and norms as well as some general points of view.</td>
</tr>
<tr>
<td>Subjects’ ways of thinking about</td>
<td>Subjects’ understanding of each other, of outsiders, and of the objects that makes up their world.</td>
</tr>
<tr>
<td>People and Objects</td>
<td>Coding words an phrases that facilitate categorising sequences of events, changes over time, passages from one type or kind of status to another.</td>
</tr>
<tr>
<td>Process Codes</td>
<td>Codes that are directed at regularly occurring kinds of behaviour.</td>
</tr>
<tr>
<td>Activity Codes</td>
<td>Directed at units of data that are related to specific activities that occur in the setting of in the lives of subjects who are interviewed.</td>
</tr>
<tr>
<td>Event Codes</td>
<td>Strategies refer to the tactics, methods, ways, techniques, manoeuvres, ploys, and other conscious ways people accomplish various things.</td>
</tr>
<tr>
<td>Strategy Codes</td>
<td>Regular patterns of behaviour among people not officially defined by the organisational chart.</td>
</tr>
<tr>
<td>Relationships and Social Codes</td>
<td>Isolates material pertinent to research procedures, problems, joys and dilemmas.</td>
</tr>
<tr>
<td>Methods Codes</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Developing Coding categories and code families. (Bogdan, 1984)

Bogdan (1984) outlines a coding category development approach, involving several steps. Typically you search through your data for regularities and patterns as well as for the topics that your data cover, and then you write down words and phrases to represent these topics and patterns. Bogdan and Biklen (1984) propose families of codes, which suggest classifications such that coding can be accomplished (Figure 1.). These coding categories are used in the education field and in particular as a means of data reduction in the qualitative data analysis process. There is no similar coding classification evident in the IS literature comparable to the above listed table of coding categories. This raises the question ‘is there a case for developing a coding categorisation specifically for the IS field?’

<table>
<thead>
<tr>
<th>Key Components of Change Framework</th>
<th>Associated Conceptual Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Organisations - products/Processes/ systems</td>
</tr>
<tr>
<td></td>
<td>Information Systems – hardware/software/systems</td>
</tr>
<tr>
<td>Social Context</td>
<td>Web Models - social relations/Infrastructure/history</td>
</tr>
<tr>
<td></td>
<td>Multi-level contexts</td>
</tr>
<tr>
<td>Social Process</td>
<td>Culture - subcultures/multiple meanings</td>
</tr>
<tr>
<td></td>
<td>Politics - control and autonomy/morality</td>
</tr>
<tr>
<td>Context/Process</td>
<td>Structuration theory - Action and Structure duality</td>
</tr>
<tr>
<td>Linkage</td>
<td>IS and modalities - embody co-ordination and control facilities - encapsulated norms</td>
</tr>
</tbody>
</table>

Figure 2. Synthesised analytical framework (Walsham, 1993)

The nearest is the IS literature is a synthesised framework by Walsham (1993) proposed as a basis for understanding organisational change called ‘Context/Process analysis’. This framework depicted in figure 2, contains four components i.e. content, social context, social process and context/process linkages and is based on the work of Andrew Pettigrew (1990). The content of an information system refers to a description of the nature of the IS and the particular organisation within which it is placed. The social context component involves web analysis to trace the social context of an information system and various levels of context (i.e. home, state, city) are acceptable if they improve the richness of understanding. Social process involves taking both a cultural and political perspective (Kling, 1992). A cultural perspective emphasises how the IS is related to the maintenance and change of subcultures and the multiple meanings that different groups attach to different meanings. The political perspective relates to control and domination. The final component provides the linkages between social context and social process (objective versus subjective) and the structuration theory (Giddens, 1984) concept of duality of structure, in particular its use of modalities, is used to provide a linkage
devices between the context and process. The role of the aforementioned theories have merit, but can the categories outlined by Walsham (1993) be used in a more conventional way, using qualitative coding techniques, which are outlined in the next section.

The primary coding approach adopted in IS research is grounded theory, which epistemologically emerged from symbolic interactionism, ‘where individuals enter their own experience only as an object, not as a subject, and that entry is a prediction of social relations and interactions’ (Urquhart, 2001). Grounded theory has also been proposed as an interpretive approach and Urquhart (2001) suggests that grounded theory ‘above all is a method’, which is used in the IS literature, irrespective of the research paradigm. Grounded theory has manifested itself in two distinct forms. Originally Glaser and Strauss collaborated on the original representation of the research approach (Glaser and Strauss, 1967). But following the publication of ‘basics of qualitative research’ by Strauss and Corbin (1996), a very public rift emerged with Glaser criticising the ‘forced, conceptual description’ rather than an ‘emergence’ through constant comparison (Urquhart, 2001). Glaser recommends that researchers do no refer to the literature, as this is a corruption of theoretical sensitivity (Glaser, 1978).

Grounded Theory is regarded as the basis for theory building and coding. It is also referred to as the prime technique for achieving the ‘constant comparative method’ (Strauss and Corbin, 1996). Comparative method is defined as comparative analysis to identify similarities and differences in the data. Coding involves refining the data into categories (i.e. conceptual constructs that emerge from the data and appear pivotal). This is organised into three categories namely; 1) Open coding (i.e. labelling the emergent concepts and grouping into categories), 2) Axial coding (i.e. identifying relationships between categories) and 3) Selective coding (i.e. developing theory to fit the data – leads to the emergence of core categories. The coding process, in turn, generates various questions, hypotheses captured in analytical memos, which help integrate the theory and refine further data collected reflectively.

Difficulties associated with coding in grounded theory have been identified and outlined as follows (Urquhart, 2001) 1) Can the codes be interpreted the same way by other people, 2) Difficult not to be influenced and resist applying concepts from elsewhere and 3) Difficult to decide on what level to ‘chunk’ the data.

Grounded theory has gained recognition in the IS literature, but more recently the emergence of software systems that support this mechanism have become evident. The following section outlines the functionality and characteristics of qualitative analysis software, as well as, the motivation for adopting this software in supporting the qualitative analysis process.

3 QUALITATIVE ANALYSIS SOFTWARE

In the early days, the last few generations of qualitative researchers did their analysis ‘by hand’, but software tools have emerged to support this process in particular Qualitative analysis software (QAS) is used to support the analysis process (Richards, 2002). The main advantages of using QAS are the ability to manage and organise large volumes of qualitative data, which can be bulky. Fielding and Lee (1998) contend that ‘the advent of software packages for qualitative analysis has been important in moving analysis issues to the forefront of concern.’ The advent of powerful and easy-to-use statistical packages undoubtedly enhanced levels of technical sophistication in quantitative research. Whether theoretical sophistication was by the same means enhanced is open to question (Fielding and Lee, 1998).

There are a range of packages available and this implies a selection criteria is required. With reference to selecting a program, Weitzman and Miles (1995) claim that

‘What’s the best program? There’s no answer in the abstract. Choosing the right software for you, depends on your own level of work with computers, on your time perspective, on the particular project you have in mind, and on the type of analysis you are expecting to do.’
The software systems can handle numerous data types, including digital pictures, sound files and various document types including web sites and transcribed interviews. Nvivo, a product developed by QSR Ltd, provides ‘a range of tools for handling rich data records and information about them for browsing and enriching text, coding it visually or by categories, annotating and gaining accessed data records accurately and swiftly’ (Richards, 2002). It facilitates coding of qualitative data, including ‘in-vivo’ coding i.e. selecting an appropriate word in the text as the code. There are different types of coding approaches, namely coding for retrieval or coding for process as outlined by Strauss and Corbin (1990). These can then be coded and categorised for retrieval purposes, as well as, theory development and model building.

3.1 Types of Qualitative Analysis Software

Weitzman and Miles (1995) suggest that software packages that have been explicitly developed for use by qualitative researchers fall into a number of categories. These include code-and-retrieve programs, code-based theory building programs and conceptual network builders. Miles and Huberman (1994) include word processors, word retrievers and text-base managers. Fielding and Lee (1995) present a classification of software systems, including the ‘generic and dedicated’ use of the system and a description of the application.

The selection of a software packages is dependent on the kind of analysis anticipated, relating to the general sense of style and approach that you are expecting. Exploratory research would benefit from a fast search and retrieval feature, while confirmatory would require a strong theory-building feature. Other issues that would have a bearing include coding schemes ‘firm (sic) at the start’ versus evolving, multiple versus single coding, iterative versus one pass, fineness of analysis, interest in context of data, intentions for displays and qualitative only, or numbers included (Miles and Huberman, 1994). The theory building category has evolved out of code and retrieve category of QAS (Fielding & Lee, 1995). The range of products available is quite varied and includes AtlasTi, Hyperreserch, Nud*ist, Nud*ist Nvivo, Decision Explorer, WinMax and Ethnograph (Kelle, 2000). Standard Software such as word Processors and text retrieval software do not have data management techniques needed to structure qualitative data, such as:

- The definition of pointers containing index words together with the ‘addresses’.
- The construction of electronic cross-references with the help of so called ‘hyperlinks’.
- Many of the contemporary systems have additional features such as;
- Facilities for string of researcher’s comments (memo’s), which can be linked to index words or text segments and for defining linkages between index words.
- The use of variables and filters so that the search for text segments can be restricted by certain limitations.
- Facilities for the retrieval of text with specified formal relationships (i.e. Distance apart)
- Facilities for the quantitative attributes of the database.
- Model building features (Graphical)

3.2 Motivation for adopting Qualitative Analysis Software

Qualitative data is notoriously ‘voluminous, unstructured, context specific and does not yield up its meaning easily’ (Fielding and Lee, 1998). Data can consist of interviews, observational field notes, videos, journals, memos, manuals, catalogues, and other forms of written or pictorial materials (Strauss and Corbin, 1998). With this in mind a key element of qualitative research is data management and the main issues are ensuring 1) high quality, accessible data, 2) documentation of just what analyses have been carried out, 3) retention of data and associated analyses after the study is complete (Miles and Huberman, 1984).

The claimed advantages of storing qualitative data on computer include, 1) having less paper to shuffle around, 2) decreased amount of time devoted to managing data, 3) the analysts becomes fatigued, 4) can devote more time and mental energy to the analysis itself, 5) computers allow the analyst to ‘play’
with the data, 6) facilitate ‘trying out’ new analytic approaches even though they may not work and 7) potentially increase creativity (Tesch, 1990).

These advantages are echoed by Kelle (2000), who highlights 1) mechanising tedious and cumbersome tasks of data organisation, 2) The use of software packages make research process more systematic, explicit, transparent and rigorous and 3) releasing the researcher from boring and cumbersome mechanical tasks, frees up time for more creative tasks (Kelle, 2000).

The main warnings associated with the use of computers in qualitative analysis centre on potential methodological dangers, where the computer could alienate the researcher from their data and facilitate strategies that go against the theoretical and methodological underpinnings of qualitative analysis (Kelle, 2000). There is a suggestion that the QAS literature leans towards the adoption of grounded theory (Lonkila, 1995). This can cause problems as grounded theory overemphasises coding and neglects other types of discourse analysis. The strong ties between ‘code and retrieve’ software and grounded theory may inspire a ‘new orthodoxy’ in qualitative research (Kelle, 2000). Others authors suggest that computers could take over qualitative analysis and alienate the researcher form the data (Fielding and Lee, 1995).

Figure 3. Qualitative analysis software (NVivo) showing interview transcript with coding stripes to the right of the screen

The categorisation of qualitative analysis software clearly shows that coding plays a central role in differentiating different packages and all of the dedicated software packages support coding. In open coding, when the name is taken directly from the text and used to code and uniquely identify it is referred to as ‘invivo’ coding (Corbin and Strauss, 1998). Some software systems allow ‘In-vivo’ coding with graphical highlight features, this is also referred to as visual coding (Richards, 2002). The In-vivo coding button is evident in the bottom left hand corner of figure 3, which also shows coding stripes on the left hand side.
The multiple sources of data features of QAS fulfils Yin’s (1994) first principle of data collection, but the issue being proposed now is can qualitative analysis software enhance the other principles by developing a case study database and maintaining a chain of evidence (Yin, 1994). Qualitative analysis software can be adopted as a means of addressing these principles by taking data from multiple sources, using the software system to maintain a database of evidence and facilitating the clear demonstration of a chain of evidence.

It is ironic that ‘commentary on QAS is almost totally missing from software magazines and technical reviews’ (Fielding, 1998). There is limited adoption within the IS literature, but the need for a well-organised case-study database is recognised and the potential of ‘specialised software tools’ are noted (Darke, 1998). The adoption of qualitative software in this study is an attempt to address the balance by exploring the QAS literature and proposing ways in which it could be applied to an IS examples.

Reviewing the types of software systems shows that the dedicated qualitative analysis software has coding as a central capability. The system demonstrated here is called Nvivo. A software system developed by QSR Ltd, the company who pioneered the NUDIST qualitative analysis system. Nvivo’s functionality includes document management (creation and management of documents), node management (category, free nodes, tree nodes and case nodes manipulation and linking) and a graphical modelling tool. Search tools and matrix display facilities are also available. (Richards, 2002).

A significant overhead, not readily referred to in the QAS literature, is the need to fully transcribe all interviews, which is a tedious activity that can take up to ten times the length of the interview.

![Diagram of hierarchical coding tree](image)

**Figure 4. Examples of hierarchical coding tree (Araujo, 1995)**

4 **HIERARCHICAL CODING DATA DISPLAY**

Miles and Huberman (1994) claim that better data displays are ‘a major avenue to qualitative analysis’ and they ‘urge a more inventive, self-conscious, iterative stance towards the generation and use of data displays.’ A key display mechanism in coding based analysis is the concept of relating categories to find ‘stories’, ‘themes’, ‘key linkages’ or ‘core categories’ are evident in the literature (Straus and Corbin, 1990). The goals of categorisation is referred to as ‘the discovery and ordering of ideas and themes, the storing of growing understanding, linking of ideas to data, cross-referencing, sorting and clarifying’ (Richards and Richards, 1995).
Qualitative researchers are ‘urged to use categories in particular ways’;
• To develop data-driven categories.
• To treat categories as being linked and structured.
• To exploit the category system as a flexible container for complex contents.
• Rich passages may relate to many categories, so it is possible to categorise richly and to code liberally.
• Critical examination and reporting the of indexing process is central to validation.(Richards and Richards, 1995)

Hierarchical categories (figure 4 & 5) are often called trees because of their one-way branching and the single category at the top of the hierarchy is called the root. The name of the root defines what the tree is about. Categories are classified into two main groups i.e. factual and referential. Hierarchical tree construction can be bottom-up and data-driven or top-down and theory-driven. Hierarchical coding are cited as ‘intermediary steps in mediating social actors’ and social settings (Araujo, 1995). Decision Explorer is a Network analysis and cognitive mapping tool, which is has an import/export link with Nvivo. Figure 6 shows a coding tree, developed in Nvivo, imported into Decision Explorer and presented as a hierarchical tree structure.

Figure 5. Example of hierarchical coding tree in Decision Explorer

5 PROPOSING AN ANALYSIS STRATEGY

Grounded theory can have a conceptual framework guiding the ‘constant comparison’ (Strauss & Corbin, 1998) or, if the ‘emergent’ school of GT was adopted then no previous assumptions are employed (Glaser, 1992). Coding schemes are not used in IS research, with no mechanism comparable to Bogdan & Biklen (1984). Is there a case for altering the scheme widely used in the education field? There is one IS oriented analytical framework that is comparable and that is the
The framework covers the ‘multi level contextual’ issues and there is an advantage of researchers studying the impact of IT on organisations by evaluating the impact at a micro level and at the macro-level. This is achieved in this study by including context in the case analysis. The technical aspects of the information systems are dealt with via ‘content’ codes, which are further sub-divided into, firstly, the organisation of the system including the roles and processes responsible for developing and running the systems and, secondly, the information system details that includes the hardware and software configurations. The organisational change codes refer to the information system as an intervention within the organisation and the resultant implications.

6 ENHANCING IS RESEARCH QUALITY

Qualitative analysis software is recommended in an attempt to systematic data immersion and internal consistency, add rigour and addresses the principles of data collection outlined by Yin (1994), namely, maintain a case study database and a chain of evidence. Yin(1994) claims that the quality of a research design can be judged by certain logical tests. The four tests commonly used in social science are listed as follows;

- **Construct validity:** establishing the correct operational measures for the concepts being studied.
- **Internal validity:** (for explanatory or causal studies only, and not for descriptive or explanatory studies); establishing a causal relationship, whereby certain conditions are shown to lead to other conditions, as distinguished from spurious relationships.
- **External validity:** establishing the domain to which a study’s findings can be generalised.
- **Reliability:** Demonstrating that the operations of a study – such as the data collection procedures can be repeated, with the same results.

Yin (1994) suggests tactics for each of these, but the approach adopted in this study gives more specific detail to the relevant phases of research. At varying stages of this study tactics are adopted which enhance the various validity tests. In light of the research issues depicted in Figure 7, construct validity is particularly important requiring appropriate measures required to enhance analysis. Yin(1994) suggests three tactics that enhance construct validity namely: multiple sources, chain of evidence and review with informants. In particular emphasis was placed on the first two tactics with multiple sources of data types sources from multiple informants. The adoption of qualitative analysis software strengthened the chain of evidence and demonstrates the emergence of theoretical constructs from beginning as an interview transcript right through to the final data display.
Central to internal validity are the issues relating to the analytical process of the research. Again the coding strategy and the use of qualitative analysis software enforces a consistent form of pattern-matching. The data display mechanisms of the software also provides an audit trail showing how explanations were built.

7 CONCLUSION

This paper is set in the context of the poor adoption of Qualitative Analysis Software (QAS) in the IS literature. The range of QAS software packages and the central nature of coding was outlined. Grounded theory is the coding technique that is most widely used (Baskerville, 1999 & Urqhart, 1997 & 2001), but here is no specific coding schema in the IS research. In response to this a novel coding strategy is proposed based on the synthesised ‘context/process’ framework of Geoff Walsham (1993). Qualitative Analysis software has clear advantages and can greatly enhance research quality as outlined above. But it must be set in the correct research epistemologies and researchers must guard against the emergence of a software package driven orthodoxy.

References


