Using databases for qualitative analysis

John McIntyre

Published in J.Higgs (ed). *Writing Qualitative Research*, Centre for Professional Education Advancement Series. Sydney: Hampden Press.

Commercial software has become available to researchers in recent years with the potential to make the task of qualitative data analysis easier, more thorough and accountable. However, not all researchers want to invest the time in learning to use specific software applications despite the longer-term benefits from doing so, including being able to claim this competence in their research skills repertoire. Thus there is a question of the alternatives to be found in generic spreadsheet and database software.

The aim of the workshop is to explore the applications of database software to problems in managing an interpretive analysis of qualitative data. I will often refer to 'data management' and by this I will mean how software can be used to carry out and assist interpretive work. I will be referring to Claris Filemaker Pro Version 3 and not to qualitative research software such as NuDIST. I would characterise this approach as an 'unstructured' use of computer software compared to applications like NuDIST which are 'structured' in that they provide a technology which embodies both procedures and assumptions about qualitative analysis. I do not claim that the use of a generic database is better than the specific application, but it is different in the demands placed on the researcher.

One theme of the paper is that interpretive work is facilitated by the sophistication and ease of use of database software. The progress of the analysis is assisted by and recorded in the developing database. In the simplest terms, database analysis refers such procedures as entering 'chunks' of data in fields, setting up other fields in which interpretative comment can be added, coding these interpretations and sorting them, finding instances of particular kinds of interpretations, exporting these selected records into a word-processing file in a given order and so on. If the business of qualitative analysis is systematically conducting a complex textual analysis and producing an account based on instantiation through quotation of text, then this can be made less laborious than it used to be.

The paper falls into several sections:

- a brief discussion of some prior considerations arising from the problematic meanings of qualitative research which are raised by the technical emphasis of the paper
- a description of some database concepts that are necessary to understand the following discussion of database development and data management
- a discussion of typical procedures which might be found in the use of a database
- two examples from research conducted using this approach.

Prior considerations

The paper might seem to be a description of way such software can be applied to data analysis, and certainly it will often focus on technical concerns. Moreover, research

texts can give the impression that the use of computers is to be regarded as a technical adjunct to the process of analysis (eg Glesne and Peshkin, 1992:141-146).However, solutions to technical questions embody consciously or otherwise a theoretical position or positions on interpretive research, and this is particularly true of proprietary software such as NuDIST or The Ethnograph. Software creates a technology of interpretation which makes particular assumptions about the meanings of qualitative research, meanings which are far from agreed.

Software might seem to provide ready 'answers' to methodological problems of data analysis but in practice it will point back to what is taken as problematic by the research, to the categories of data and interpretation created, to the researcher's presuppositions about analytical procedures and to their notions of qualitative research. As I argue in my other paper in this seminar, the beauty of qualitative work is its creative and intuitive use of interpretive procedures (in Garfinkel's sense), and this can be reduced to computerised analysis but only by sacrificing something. The use of a proprietary software in a thesis, in my view, has to be problematised, it does not warrant 'methodological adequacy'. Methodological adequacy has to be made out in terms both of technique and the creative and critical deployment of researcher understandings.

What drives qualitative analysis is its interpretive interest and not the technical and instrumental aspects of data-management. It is of course perfectly possible to create empiricist forms of qualitative research - for example, where the reduction of data to coding categories and their quantification take precedence over the uncovering of complexities in textual analysis and the problematisation of the categories by which an analysis is accomplished. The point is what kind of interpretive analysis is being claimed under the heading 'qualitative research' when it is 'computerised' in a software package? A researcher should be able to answer such questions and critique the constructs which are assumed.

It is interesting that the current generation of software no longer supports a hard and fast distinction between qualitative and quantitative analysis. A decade ago, quantitative analysis required an esoteric program like SPSS (Statistical Package for the Social Sciences) with its own syntax, which excluded any kind of textual analysis, so that specific software was needed to perform 'qualitative' work. Contemporary spreadsheet and database software has recently reached the point where it is so powerful and flexible that it can be adapted either to quantitative or to qualitative analysis. Excel 5 or Filemaker is indifferent to whether information is numeric or textual, since spreadsheets have complete database functions for managing text information and database software incorporates spreadsheet functions to manage numeric information, and both have full statistical functions. Hence the question is what kinds of 'qualities' are being claimed by the term 'qualitative', what kinds of interpretive practice are being carried out, whether they use software or not.

In saying this, it is something of a paradox that as the deconstructive critique of professional research has been deepened and extended, computers have been applied to technologising the kind of textual analysis this critique recommends. This sets up some interesting tensions around the way interpretive practices themselves are constructed, conducted, described and justified in contemporary research. These tensions, I believe, need to be acknowledged in doing interpretive work. Or is the software package to be the last refuge of the epistemologically challenged?

These remarks no doubt are signalling my research prejudices which are of a social phenomenological or ethnomethodological kind. I would hope that these theoretical interests can be traced in my understandings of computer-assisted data management, because I argue that there is no qualitative analysis which is above and beyond some

construction of interpretive method.

Database procedures

It may be helpful to outline some database concepts. Familiarity with these concepts is a necessary part of developing research expertise in applications of database management to interpretive work.

A key concept is that of the 'record' which can be thought of as a form divided up into areas into which are placed specific types of information. These locations are known as 'fields'. There can be any number of records in a database. Each record could correspond to a person (as in medical records) but equally can correspond to an 'individual case' of a chunk of information such as a response to one interview question or an episode from a transcript (see examples below).

A range of operations can be performed on a set of such records, a process resembling filling in, shuffling and sorting a set of file cards. These operations include sorting records into one or another order, finding records of a given type, classifying and coding the information in a given field, finding instances of a given type of information, counting instances and so on. In addition, electronic databases allow cutting and pasting of information into fields, the 'import' of records from other sources and the 'export' of data into word-processing formats. Further, the appeal of current databases like Filemaker is that information can graphically displayed in a what are known (in Filemaker) as different 'layouts' tailored for different purposes. Whereas a set of file-cards records information in one and only one form, an electronic database can display that information in a whatever ways one chooses, limited mainly by practical considerations. The major advantage of Filemaker is the ability to show information and adapt its display.

The application of the database is clear. If qualitative analysis is identifying and analysing 'chunks' of text, then these 'chunks' can be put into a field and analysed. The result of an analysis of a given fragment of text (readings, interpretations, glosses) can be entered in another field or fields. These fields can be searched for keywords and the records obtained coded in some way in yet another field. This procedure can thus be applied in conventional terms to developing first and second order constructs derived from the understandings of participants evidenced in transcripts. Clearly, it can support very close textual analysis of various kinds, including deconstruction or conversational analysis or interactional analysis. It can also support conventional analysis of responses to interview questions.

Current software provides many resources for this kind of work. The graphical display capability of software means that the same information can be arrayed in whatever way is convenient for doing the work. There can be many displays of the same text, different arrays of the same set of fields, for different purposes. Thus it is possible to carry out analytical operations such as finding and sorting records of different types and to display the information in 'customised' ways that take account of visual literacy and ease of use over the many hours researchers often spend looking at their data. In current databases such as Filemaker, the 'fields' might often be text fields and these can be as large as the user wants them to be, though practical terms, there are limits to how easy is to see all of the text in a field at once.

Some applications of databases to qualitative analysis might include the following, of which there are many possible variations:

- a record may contain one response to an interview topic or question. The responses of all participants to this question can then be displayed in summary form, where the purpose is to compare and classify these responses and perhaps code them.

This could be referred to as the ‘response summary’ layout (see Figure 1 below).

- a record may contain an individual chunk of text for close analysis in one 'text field' that is surrounded by several other fields in which is entered a range of interpretive information (readings, glosses, classifications, codings, orderings, numberings and on). This array of interpretations can then used in further stages of analysis. This could be referred to as the ‘text and commentary’ layout (Figure 2 below).

- a record may contain details of the characteristics of participants or text fragments or some other thing, and these can be displayed in a similar summary layout.

These and other features of database software increase the ease and power of interpretive work beyond the older manual systems. These include the facility with which the ‘prospective’ work of outlining an analysis can be done, the ability to generate (export) instances, display classifications, check codings and classifications and so on.

A concept of database development

A useful concept is that of database development, which refers to a process by which information is entered and its analysis progressively built up within the database. The results of the interpretive work are recorded alongside the matter being analysed. This is one way in which computerised analysis differs from the older manual systems where the texts produced by the analysis might have been kept separate from the records themselves on another set of notes or cards. Though researchers have their own approaches to building databases, experience suggests that there are a number of stages involved. These might include:

- *design*, which involves the definition of data fields and the design of layouts, and the progressive creation of data records. This is a stage where a steep learning curve is experienced as the researcher may experiencing the frustration of half-knowing the capabilities of the application while labouring to enter the data.

- *development*, where the interest shifts to working with the data, creating additional fields which categorise, comment upon, typify or enumerate the data fields, and which enable the sorting, searching and summarising of interpretations. This may be experienced as less frustrating or laborious than design, and the researcher may be able to improve the display of data and interpretations. More is learned both about the capabilities of the software and the corpus of data, increasing the range and depth of the analysis. The first outputs of the database can be printed out and studied and the results fed back into the database.

- *refinement*, a mature stage where data entry is complete and the interpretive fields are being finalised. Procedures for sorting and searching the data are well established. Categorisation and coding procedures can be ‘scripted’ and performed through keyboard commands or display ‘buttons’ which make rapid movement from display to display possible. Selected records or parts of records which exemplify a given type of response or construct or interpretation can be exported and pasted into a word-processing format, repaying the labour of data entry and enabling ‘sourcing’ of the extract to its location in the data base.

In the following I refer to some typical procedures which might be employed in developing a database for qualitative analysis. I do not represent these as simply ‘technical’ matters, because these techniques are being used to manage ‘interpretive’ decisions which go right to the purposes of the analysis. There is always the question: what are the qualities being explored, the categories and constructs in play in the

researcher's understanding?

(a) defining a record and its fields

The first stage of database design is defining a record by the fields which make it up. A record comprises several fields of information, but to what should these fields refer? One organising principle has already been suggested - one field may hold the 'data' or 'information' or 'text' to be analysed, while other fields are created to hold a range of information to be added as analysis proceeds. Designing a database entails decisions about what these fields should be, though fields can be added, removed or renamed as analysis develops. The definition of a record can be troubling for students, who may not be clear about what fields they want, or what the unit of analysis should be. Should the record refer to a person or a chunk of data?

(b) deciding what data is to be entered

Entering data is straightforward for interview responses where the fields may be defined for each question or interview topic. In this case it is easier to set up the database first and type directly into the fields. Unstructured transcript data can be cut and pasted into records, though this often poses another problem, since the body of data needs to be sectioned, with each section pasted into a field in new record eg the field 'incident' in Figure 2) . The section can be identified in various ways including its location in the corpus (eg field 'source' in Figure 2). One of my students working with dialogue transcript sectioned it by identifying 'episodes' and making each episode a record, resulting in over 300 records. Each episode could then be analysed in a variety of ways, which depend on what kind of interpretations are looked for. It may be enough to select extracts from a large corpus.

(c) designing layouts and data displays

Graphically-based software has made it possible to design the visual form of the database record in whatever way the researcher pleases. Fields may be large or small and contain text, numbers, calculations, objects or other content. For example, a field could contain an audio or video fragment five seconds in length, with a close transcription in another field. A record may contain a great many fields, though there are practical and aesthetic limits to how many can be shown in a 'layout'. Thus a record can have many layouts displaying different sets of fields, or the same set of fields can be arranged in different 'displays' - for example, showing all of a text extract, or only part. Again, it is helpful to refer to the examples in the figures. Much more can be said about the display capabilities of current software.

(d) developing and entering interpretations

Reiterating a key point, some fields can be created to hold the researcher's interpretations of 'data' entered from some source. These interpretive fields can be given any name, since it is what they contain that matters, not their title (which may change as ideas are clarified). At this point, the researcher can no longer postpone confronting their real interpretive purposes - the question of what they are doing. Thus a field can be named 'themes' to signify that kind of meanings is being picked out, or 'questions' to mark ways of interrogating text, 'discourses' to name constructions at work in the text and so on. These fields thus define the scope and extent of the work, and reflect the theoretical interests driving interpretation. This analysis may be superficial with few fields or highly elaborated, with many fields related to each other.

These interpretive or analytical fields record the progressive elaboration of the analysis - elaborated both in terms of new fields of comment added, and in the detail and richness of interpretive comment. The question how much analysis is 'enough' can only be answered in terms of what theoretical and other purposes are being satisfied

by the inquiry. It is entirely open to argument, not to any a priori methodological prescription.

(e) systematising an analysis

Beyond this stage, further kinds of analysis are possible depending on how much emphasis is given to systematising and developing operations on the database, such as classifying, searching, sorting and displaying results. A question arises as to how the researcher can know what has been developed in the database. As the analysis becomes more elaborate, keeping track of what has been added becomes more problematic. Procedures are needed to locate particular examples of text and interpretation. The point is that a database permits a large amount of information to be managed, and realising this potential means systematising the analysis. Thus first order interpretations can themselves be interrogated, classified and coded so that the retrieval of sets of interpretations is possible, since the researcher wants to be able to 'exhibit' or 'account' for the analysis at some point, to make out what has been done. At this point, questions arise as to criteria for the adequacy of the analysis, where the researcher may appeal to such qualities as the consistency, explicitness, methodicality, verifiability, comprehensiveness or reflexiveness of procedures.

(e) producing an account of the analysis exhibiting interpretive work

At some point the researcher will need to produce an account of the analysis by writing out the understandings achieved. The database has enabled them to arrive at this point. How can the interpretive work which has been built into the analysis now be 'exhibited' in writing an account of the inquiry? Here the database now repays the labour of its construction. First , the structure of the database is an artefact of the activities which have produced it and can be discussed in terms of the interpretive problems faced and resolved through those activities. Much more could be said about the problems in 'exhibiting' the interpretive work embedded in the structure. Second, the database will now provide much of the text material that is needed to write an account of the analysis. For example, it may be that a thematic analysis has been carried out and the researcher now wants to discuss and exemplify a given theme, giving instances of its occurrence from transcript 'evidence'. The relevant records can be found in the database, the relevant fields (the transcript extract, associated interpretations, location or other data) can be selected and these records then 'exported' as a text file. From there, this text (which represents part of the database) can be sorted and rearranged as the analysis is written and further discussion is interpolated. Third, the writing of the account can refer to the qualities of the analysis as accomplished through the activities carried out. Writing can examine these activities in a reflexive way, for example, by problematising the deployment of transcript 'evidence' and calling into question the validity of classifications. The database provides a resource for such discussions of methodological integrity, consistency, validity, trustworthiness and so on.

To conclude this discussion, two examples of research using database analysis are provided. These are homely rather than sophisticated examples, produced under pressure, but this workability is a reason to recommend their use given what has been said about interpretive methods.

Case study: youth worker legal support

The brief of the 'youth worker legal support' project was to determine the training needs of a range of people working with children and young people involved with law either as victims or offenders. Legal support work refers to such activities as acting as advocates or adviser to young offenders or victims, educating youth about legal procedures, providing information about resources and so on. Workshops were

conducted in city and country locations to explore the experiences of youth workers and to draw out the main areas of competence that seemed to make up legal support work typically carried out by youth workers. The workshops were taped and transcripts produced. The data management problem was to analyse in a short time over 150 pages of this data and generate an account of the main dimensions of legal support work (Flowers McIntyre and Loughman 1992).

The report produced a preliminary account of the competence of legal support workers as a basis for developing training courses. This analysis of legal support was workshopped with practitioners before being written up in a report which was illustrated with quotations sourced back to the transcripts. Later, critical incidents illustrating the application of this expertise were drawn out of the transcripts and entered in the database (see Figure 1).

Figures 1 and 2 show that the database is relatively undeveloped. Figure 1 is a summary layout which provides an overview of all the records. The fields contain information that is mainly classifying or identifying interpretations. They refer to such things as: 'Source', a place in a particular transcript (eg Taree WS:35); 'Statement', a summary of the type of competence implied or stated at this point in the transcript; 'Worker', the type of youth worker making the comment (eg school counsellor); and 'Classification' of the competence statement into one of ten types (eg advocacy). 'Situation' refers to a situation described by a respondent in relation to the exercise of their legal support skills and relates to 'critical incidents' which were requested from those attending the workshops.

This database was not used for an elaborate textual analysis of this interesting material due to limitations of the project. However, at a later stage, some extracts from the transcripts were cut and pasted from a text file into a field called 'Incident' in a special layout called the 'Critical Incident Page' (Figure 2). Here a whole screen is taken to display this information which forms part of one record. It is possible to use such a layout to carry out discourse analysis in a field or fields added for that purpose. In yet other fields these comments could be classified or coded in order to identify particular instances of a given reading of the text. These records could be retrieved and the comments exported.

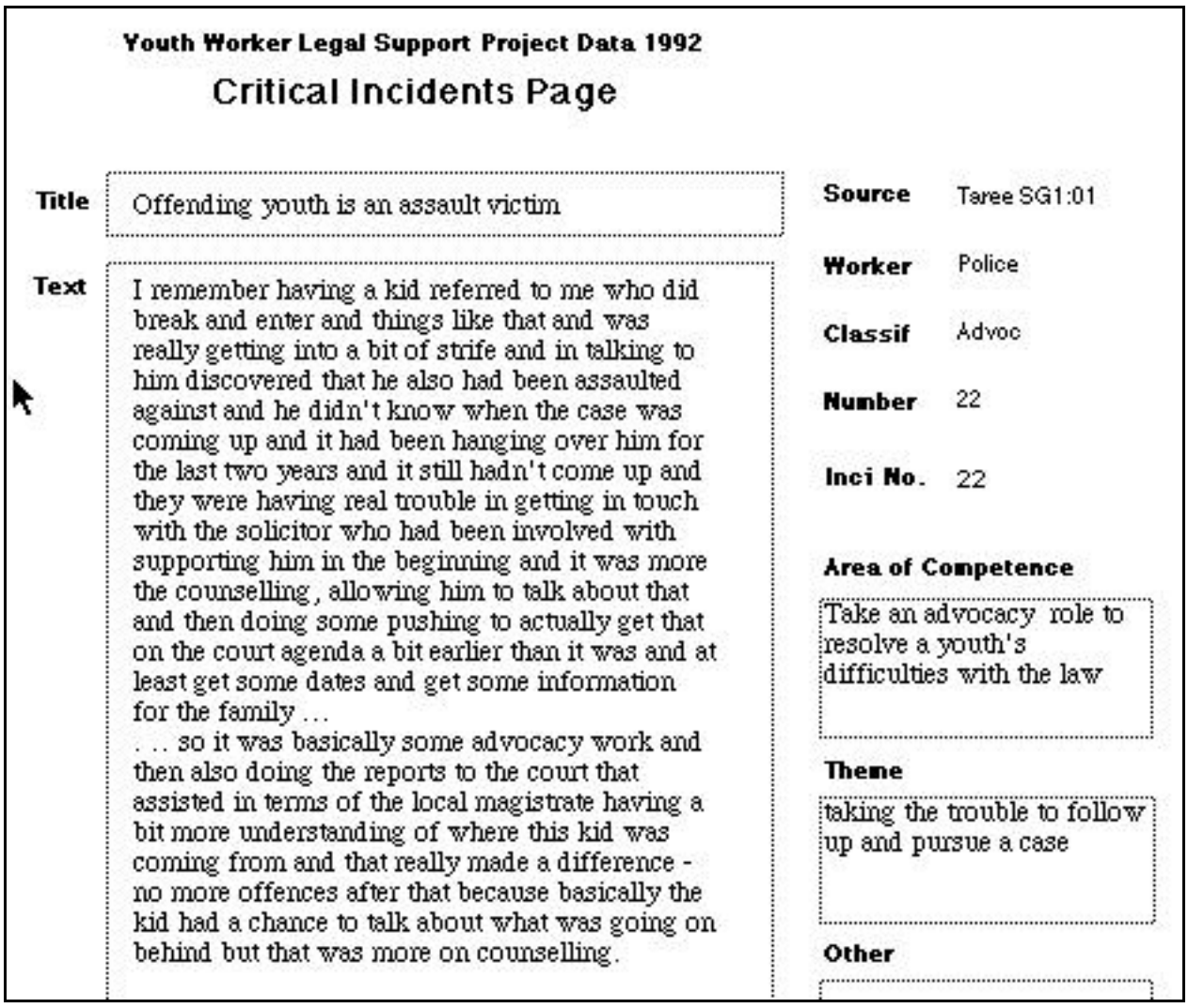
Figure 1. Summary page layout

Youth Worker Legal Support Project 1992

Summary Page

	Statement	Source	Worker	Classif	Inc
1	Make a decision to use legal processes to intervene in a situation	CitySG1:1	DO	Intervene	
2	Explain what will happen in court, likely consequences	CitySG1:11	DO	Advise - adv	
3	Explain the law and legal processes to the community 19	Bourke 19	Police	Advise - educ	
4	Help young people to know their legal rights and entitlements	CityWS:01	Police	Advise - educ	18
5	Educate youth about the legal consequences of criminal behaviour	CityWS:10	YthWkr	Advise - educ	
6	Teach children about access to legal support. understand their need to know	CityWS:15	YthWkr	Advise - educ	
7	Inform, teach facilitate learning in a variety of educational contexts	ParraSG1:01	Several	Advise - educ	
8	Advise a young adult on the difference between Courts for children and adults	ParraSG1:02	YthWorker	Advise - educ	
9	Informing people (family members) of legal outcomes and their rights in crisis	Taree SG1:02	DO	Advise - educ	
10	Advise and help prepare refuge kids for court appearances	Taree SG2:02	RefW	Advise - educ	
11	Brief and advise a solicitor in child protection matter coming before a	Taree WS:01	DO	Advise - educ	

Figure 2. Text and commentary layout



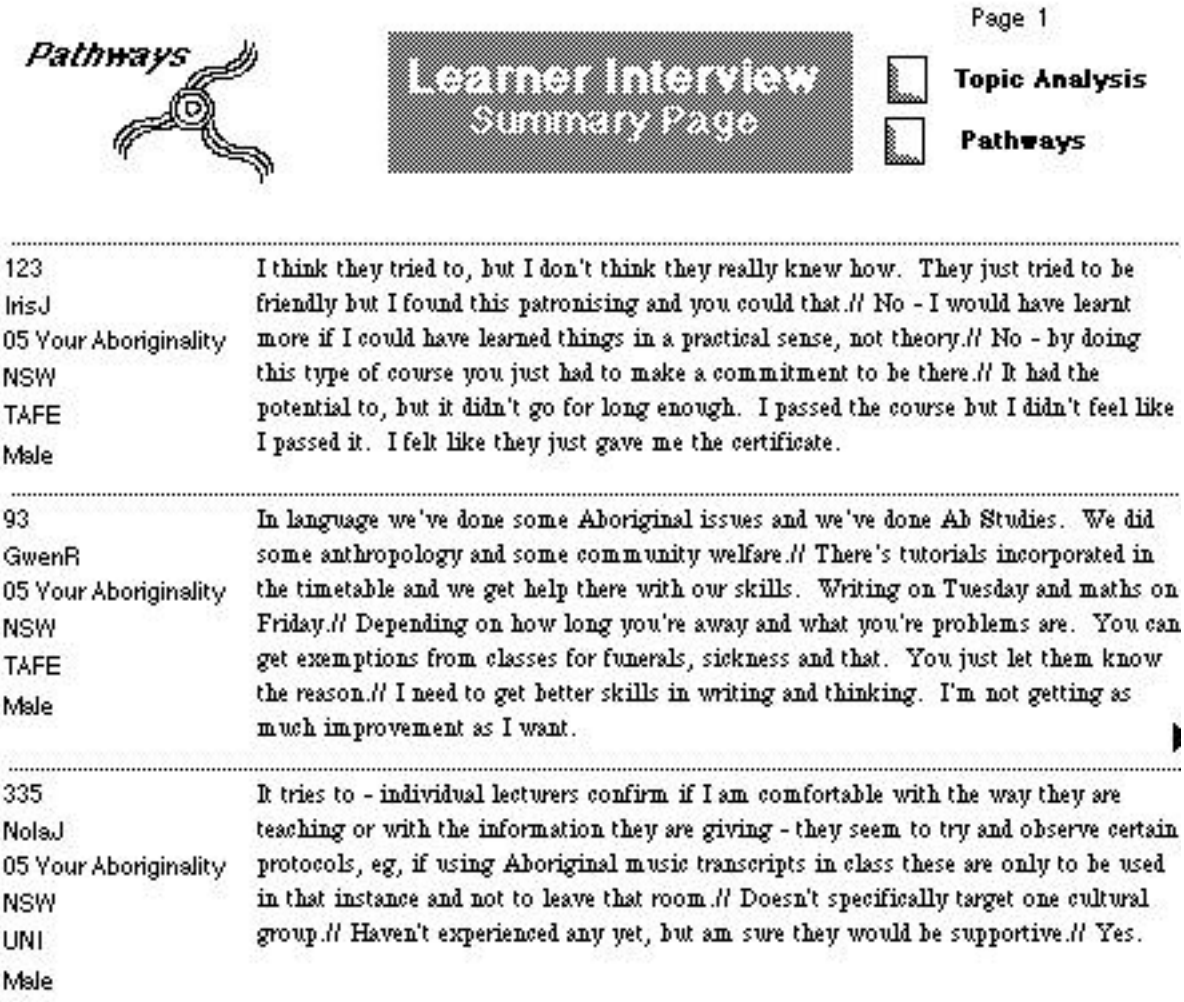
Case study: Analysing interviews

My other paper for this seminar describes the interpretive approach of an ANTA-funded project investigating factors affecting outcomes of VET for Aboriginal and Torres Strait Islander people. This project took this approach in order to hear Aboriginal 'voices' and bring out the cultural issues. We worked with about eight Aboriginal researchers (mostly mature age students) to explore the experiences of Aboriginal students in vocational education and training, starting with a reflection on their 'educational journey'. The researchers used a thematic interview form with open-ended questions. The researchers wrote up the interviews as notes on a form. Interviews with institutions were also involved (McIntyre et al 1995).

The problem was how to manage this interview data so that it could be displayed and sorted by members of the team, and so that selected field notes could be extracted and written into an analysis of learner perspectives which was to be structured around themes or topics. Should each interview be a record? or each topic within an interview? We decided to make each topic in the interview, one record, so that a 'topic' refers to one page of the interview where the interviewer recorded responses.

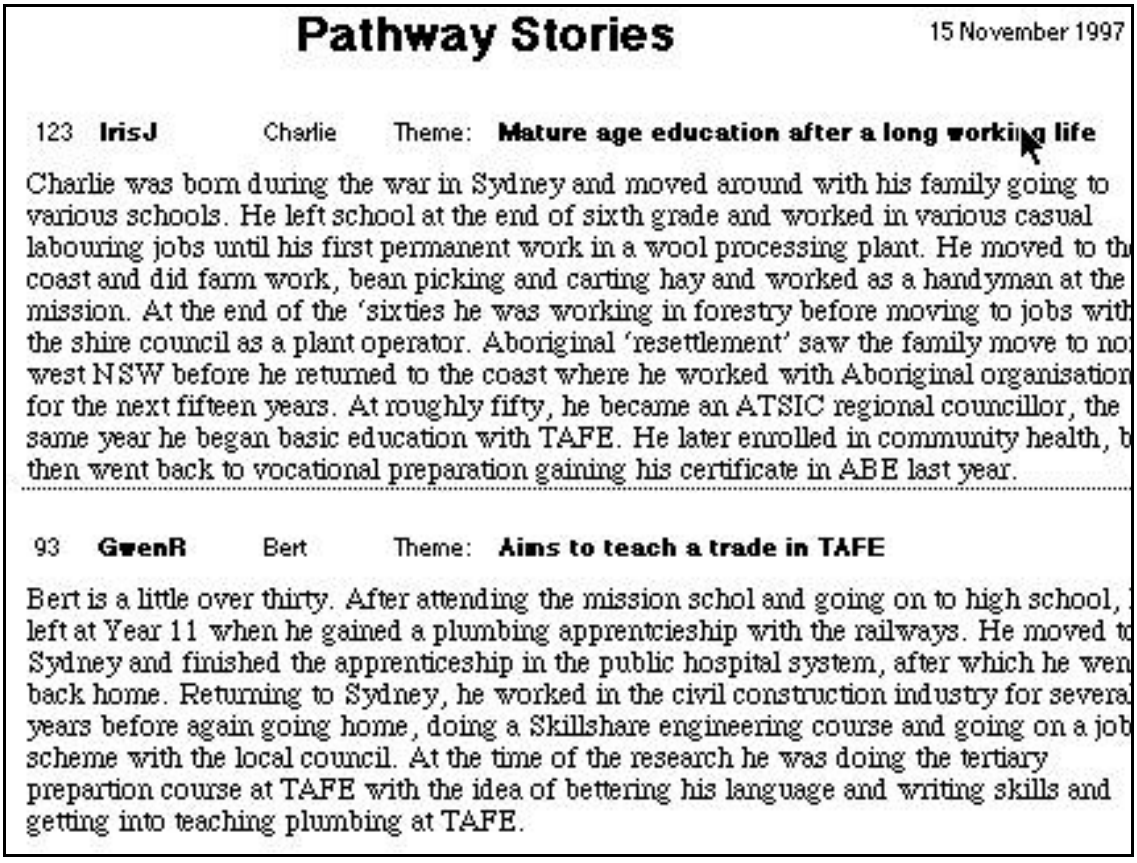
Thus each interview would yield ten topic-records, which could then be displayed or sorted in various ways - by person, by topic, by theme or other feature.

Figure 3. Interview topic summary



In Figure 3, the main field is text of the field notes, and those to the left of the text identify the record number, the interview (by codename 'NolaJ') and the topic name, the state, the institution attended and the student's gender. Missing from this layout is any 'commentary' field recording interpretations of the text. The function of this layout is to display some or all of the 70 responses to this topic.

Figure 4. Pathway story and index layouts



Pathway Stories

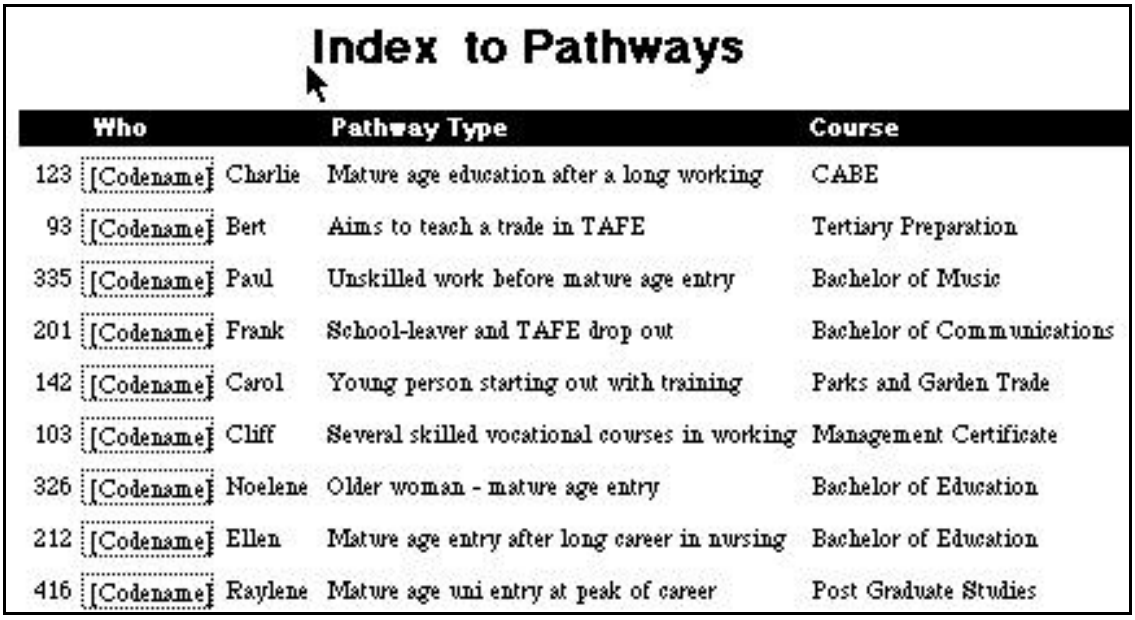
15 November 1997

123 **IrisJ** Charlie Theme: **Mature age education after a long working life**

Charlie was born during the war in Sydney and moved around with his family going to various schools. He left school at the end of sixth grade and worked in various casual labouring jobs until his first permanent work in a wool processing plant. He moved to the coast and did farm work, bean picking and carting hay and worked as a handyman at the mission. At the end of the 'sixties he was working in forestry before moving to jobs with the shire council as a plant operator. Aboriginal 'resettlement' saw the family move to north west NSW before he returned to the coast where he worked with Aboriginal organisations for the next fifteen years. At roughly fifty, he became an ATSIC regional councillor, the same year he began basic education with TAFE. He later enrolled in community health, then went back to vocational preparation gaining his certificate in ABE last year.

93 **GwenR** Bert Theme: **Aims to teach a trade in TAFE**

Bert is a little over thirty. After attending the mission schol and going on to high school, left at Year 11 when he gained a plumbing apprentcieship with the railways. He moved to Sydney and finished the apprenticeship in the public hospital system, after which he went back home. Returning to Sydney, he worked in the civil construction industry for several years before again going home, doing a Skillshare engineering course and going on a job scheme with the local council. At the time of the research he was doing the tertiary prepartion course at TAFE with the idea of bettering his language and writing skills and getting into teaching plumbing at TAFE.



Index to Pathways

	Who		Pathway Type	Course
123	[Codename]	Charlie	Mature age education after a long working	CABE
93	[Codename]	Bert	Aims to teach a trade in TAFE	Tertiary Preparation
335	[Codename]	Paul	Unskilled work before mature age entry	Bachelor of Music
201	[Codename]	Frank	School-leaver and TAFE drop out	Bachelor of Communications
142	[Codename]	Carol	Young person starting out with training	Parks and Garden Trade
103	[Codename]	Cliff	Several skilled vocational courses in working	Management Certificate
326	[Codename]	Noelene	Older woman - mature age entry	Bachelor of Education
212	[Codename]	Ellen	Mature age entry after long career in nursing	Bachelor of Education
416	[Codename]	Raylene	Mature age uni entry at peak of career	Post Graduate Studies

Conclusion

This paper has attempted to show that generic database software can be easily adapted to qualitative analysis, by outlining a number of concepts including database development. It has argued that this technology assists but in principle cannot do the interpretive work of researcher. The design and development of the database will help the researcher to confront and formulate their real interpretive purposes. It can assist

in elaborating and managing a complex textual analysis, the results of which can be used as a resource in writing out the understandings gained, in exhibiting the analysis and discussing its procedures and assumptions. The paper has provided a number of simple examples from research practice to illustrate some of the concepts and principles discussed.

References

Claris Filemaker Pro Version 3.

Glesne and Peshkin. (1992). *Becoming Qualitative Researchers.* New York: Longman.

Flowers R, McIntyre J and Loughman J. (1992). *Working Together.* Report on the training needs of youth legal support workers. Marrickville Legal Centre and the Law Foundation of NSW. 5 vols.

McIntyre J, Solomon N, Ardler W, Morley-Warner T and Spindler L (1996).*Culture Matters*: Factors Affecting the Outcomes of participation of Aboriginal and Torres Strait Islander students in Vocational Education and Training. Research report for the Research Advisory Council of ANTA. Research Centre for Vocational Education and Training. Sydney: University of Technology, Sydney.