Computer-aided Archiving of Qualitative Data with the Database System "QBiQ"

Susann Kluge & Diane Opitz

Abstract: In Germany up until now qualitative data material has rarely been comprehensively and systematically archived. At The Special Collaborative Centre 186 (Sfb 186), we responded to this situation by developing a computer-aided archiving concept for qualitative longitudinal data that also contains suitable strategies for anonymisation and detailed conditions for access and transmitting of data. We needed new computer technologies as it is very time-consuming to archive large amounts of qualitative material in paper form. The computer readable format offers several advantages; it reduces the amount of storage required, allows quick copying of the data and makes it feasible to selectively search the data for particular research questions.

In addition, we are developing the database system "QBiQ" (pronounced "cubic") which not only archives qualitative and quantitative data in one system but it also provides main functions of common text analysis like coding, carrying out of text retrieval etc. Furthermore, it offers some useful interfaces for text analysis systems in order to facilitate the data exchange between different programs.

Table of Contents

1. Introduction
2. Advantages of Computer-assisted Archiving
   2.1 Reduction of storage requirements and fast transmitting
   2.2 Establishing qualitative research
   2.3 Application of text analysis systems
   2.4 Documentation and reanalysis
3. The Concept of Computer-assisted Archiving of Sfb-Data
   3.1 Integrative database system
   3.2 Integration of qualitative and quantitative data
   3.3 The database system "QBiQ"
4. Data Protection and Ethical Aspects
   4.1 Anonymisation
   4.2 Consent of interviewees
   4.3 Use for research purposes only
   4.4 Further regulations
   4.5 Conditions for access and transmitting of data
5. Summary
References
Authors
Citation
1. Introduction

The Special Collaborative Centre 186 "Status Passages and Risks in the Life Course" (Sonderforschungsbereich 186, Sfb 186)\(^1\) has been sponsored by the German Research Council (DFG) since 1988. It had to break new ground in many research fields in order to investigate relationships between social structures on one hand and individual action strategies on the other hand. As far as methods are concerned, we developed for example new ways to integrate qualitative and quantitative methods. Because most of our projects combine qualitative and quantitative methods in longitudinal panel studies, unique verbal and standardised data sets are available with an observation window of up to ten years.

<table>
<thead>
<tr>
<th>Selected Sfb-Projects</th>
<th>Qualitative Studies</th>
<th>Quantitative Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupational Training Careers and the Family in the Life Course, Walter R. Heine (A1)</td>
<td>Number of Waves</td>
<td>Number of Waves</td>
</tr>
<tr>
<td>Occupational Careers and Delinquency, Karl P. Schumann (A3)</td>
<td>Number of Waves</td>
<td>Number of Waves</td>
</tr>
<tr>
<td>Occupational Trajectories in Social Change, Ansgar Wegmann (A4)</td>
<td>Number of Waves</td>
<td>Number of Waves</td>
</tr>
<tr>
<td>Role of Occupations in the Structuring of the Female Life Course, Helga Kangas (B1)</td>
<td>Number of Waves</td>
<td>Number of Waves</td>
</tr>
<tr>
<td>Social Assistance Careers: Between Life-Course Policy and Social Change, Stephan Lehnkühler (D3)</td>
<td>Number of Waves</td>
<td>Number of Waves</td>
</tr>
</tbody>
</table>

Fig.1: Selected projects of the Special Collaborative Centre "Status Passages and Risks in the Life Course" [1]

It is a well (and long) established practice for quantitative researchers in Germany to transfer their data to the Central Archive in Cologne and also to receive data sets from there. As in Great Britain where the "Qualitative Data Archival Resource Centre" (Qualidata) was founded at the University of Essex\(^2\) in 1994, there is no central place for archiving verbal data in Germany. Archiving of qualitative data is not only instrumental in verifying research results through reanalysis but also in carrying out secondary analysis. On one hand, the density of information and therefore the analysis potential of qualitative data is normally very high and on the

---

1. See: [http://www.sfb186.uni-bremen.de](http://www.sfb186.uni-bremen.de)
2. For further information on the archive see CORTI and THOMPSON 1998 as well as the homepage of the archive: [http://www.essex.ac.uk/qualidata](http://www.essex.ac.uk/qualidata)
other hand the analysis and preparation of verbal data is very extensive and time-consuming (especially the transcription). [2]

In the German-speaking countries, however, there exist only a few archives for qualitative data up to now. These archives were based on the commitment of individual researchers only, and moreover, they specialised in very different subject areas³. Furthermore, only very few archives provide their data digitally: Most interviews are not available in a computer-readable format. Usually, they were written on a typewriter (mostly before the introduction of computers) and it is very time-consuming and expensive to scan these transcripts. [3]

In order to archive this data material, it is a great advantage for our work that all interviews of the projects of the Sfb 186 are available in a machine-readable format. It facilitates the access to these unique data and the researchers can easily apply computer-assisted methods of text analysis. However, there are no general rules up to now for a computer-assisted archiving of qualitative data material which also consider the requirements of legal data protection and ethical aspects of research. Because qualitative interviews mostly contain many personal details of the life of the interviewee, the risk of identification is relatively high. Therefore, we elaborated a new concept for the computer-assisted archiving of qualitative data which contains suitable anonymisation strategies and detailed transmitting conditions for sensitive data. [4]

Before presenting our concept in general, we point out the advantages of computer-assisted archiving. Finally, we discuss the data protection and ethical aspects of archiving qualitative data material. [5]

2. Advantages of Computer-assisted Archiving

The advantages of computer-assisted archiving refer to four different fields:

1. Firstly, the storage requirements are reduced and the copying and transportation of qualitative data are considerably improved. The filed data can be passed fast and simply to other researchers.
2. It does not only simplify the carrying out of secondary analysis and reanalysis, but it also allows the use of already collected and prepared material for university teaching or for graduating students.
3. Text analysis systems can be used without any further preparation steps, if qualitative data are available already in machine-readable format.
4. Computer-assisted archiving facilitates not only the administration and analysis of the extensive data material but also the documentation of the research process and allows for verification of the research results. [6]

³ See the table of contents.
2.1 Reduction of storage requirements and fast transmitting

It is very expensive to archive and transmit qualitative data in account of its large quantity. As long as qualitative data material is only available on paper, the copying costs much time and work and the transportation is very extensive and—with regard to data protection—insecure. This is why the researchers normally go to the archive place in order to screen the material. [7]

New computer technologies facilitate the handling of these amounts of data significantly: They make it possible to reduce the storage requirements of the voluminous verbal data material considerably as the data is stored in machine-readable format. Furthermore, the interviews can be saved on a portable data medium—like every quantitative data set—so that they can be transmitted quickly and easily to interested researchers. Therefore, it is no longer necessary for researchers to go to the archive, rather the data can be sent by email or by post. Finally, secondary analysis and reanalysis are much easier due to the reduction in storage requirements and the improved transportation possibilities. [8]

2.2 Establishing qualitative research

The simplified sharing of qualitative data is of particular interest for the methodological curriculum at universities and for researchers who are writing their qualification papers. Because the analysis of verbal data is very comprehensive, time-consuming and expensive, students and young researchers have no possibilities to investigate greater samples—for instance more than 20 persons—and to transcribe all these interviews. Therefore, it would be an added advantage, if they could use qualitative data material which is already collected, transcribed and available in a machine-readable format. [9]

Qualitative research methods still lead a shadowy existence, because they do not belong to the official curriculum at German universities up to now. The procurement of these methods normally depends on the individual commitment of the lecturers, who mostly tend to work with qualitative research methods themselves. In order to get the scientific junior staff acquainted with qualitative methods and to consolidate the position of qualitative research, it is indispensable that qualitative data is available for the education in sociology. The computer-assisted archiving opens excellent possibilities for interested researchers, because they can quickly get the already collected and prepared data (see also GLÄSER & LAUDEL in this issue). [10]

2.3 Application of text analysis systems

If the interview texts are already available in machine-readable format, text analysis systems like winMAX, ATLAS/ti or NUD.IST can be used without further preparation of the data material (see KELLE 1995; WEITZMAN & MILES 1995; FIELDING & LEE 1998; ALEXA & ZÜLL 1999; KUCKARTZ 1999), because all texts in ASCII-format can immediately be imported into every text analysis system. [11]
These programs facilitate the administration and analysis of qualitative data considerably, because all interviews are integrated in one dataset. The individual analysis steps can be carried out for all interviews simultaneously and it is possible to accelerate the retrieval of coded text passages or search for significant terms. [12]

In addition, these text analysis systems create good preconditions for archiving the analysed data material, because not only the interviews but also the codes, codings, variables or memos can be stored by an export function on a floppy-disk or CD-ROM. Thus, this supplementary data can be archived too, and later be imported into another computer for further research. [13]

2.4 Documentation and reanalysis

If qualitative data are analysed and archived computer-assisted, the research process can be documented very well. The thematic aspects for the analysis of the data material must be defined precisely in the form of codes and these codes must be explicitly assigned to particular text passages. Interpretations which are based on several text passages can be checked very quickly by these codings. It is not necessary to search tediously for these passages in a large amount of transcripts, but they can be found with special retrieval functions very quickly. [14]

Recently, several scientists have insisted on the importance of the documentation of the research process and the archiving of the empirical data material in order to control the research results. In responding to this demand, the German Research Council (DFG) recommended keeping primary data on more durable and secured media for ten years (DFG 1998, p.12). Other scientists require (KAASE 1998, p.95; WAGNER 1999, p.14) that data must also be available to other researchers, if they have a well founded professional interest in reanalysis. Thus, the forgoing of data and research results can be better prevented. Leading American journals in social and economic sciences accept contributions, only, if the data material is also available for reanalysis. [15]

3. The Concept of Computer-assisted Archiving of Sfb-Data

For the computer-assisted archiving of the Sfb-data, in addition to data protection conditions, two further aspects have to be considered:

- On one hand, many projects have already used common text analysis systems like winMAX or NUD.IST in order to analyse the extensive data material systematically—but with different programs and versions. Therefore, the question came up on how should the data be entered in an integrative database system so that it simplifies both the administration and the transmitting of the material.
- On the other hand, it should be considered that most projects collected qualitative and quantitative data. Therefore, the standardised data should also be at hand in order to analyse the verbal data in an appropriate manner. [16]
3.1 Integrative database system

Because many projects of the Sfb 186 have managed and analysed their qualitative data material with text analysis systems (see also KÜHN & WITZEL in this issue), numerous supplementary material is available like codes, codings or memos. This additional material is not only important for reanalysis within the Special Collaborative Centre 186 but also very useful for secondary analysis in other research projects and for teaching purposes at universities. [17]

But up to now qualitative data can only be imported into the same program with which the data was also produced and edited before. This is because these programs hardly have any interfaces to other programs (see also MUHR in this issue). In order to use the codes and codings of other projects the researchers has to work either with the original program or they have to enter the codings, memos and variables again. This procedure is not only very extensive but it is also redundant, because the data is already available but not in the appropriate format. Often it does not make sense for different reasons to work with the text analysis program of the data providing project:

- "Old" DOS programs often cause problems when working with Windows.
- Previous versions of qualitative programs mostly offer fewer functions and are not as user-friendly as modern Windows versions.
- The newer versions of text analysis programs offer special functions—like the building of coding families or the creating of network structures in ATLAS/ti—and
- offer the possibility to combine different methods. [18]

In general, it is much too complicated for an archive to handle the whole variety of different programs. In order to be independent of the different text analysis systems and their versions, we decided to develop an integrative database system.
3.2 Integration of qualitative and quantitative data

While creating a program for a suitable database system, we also had to consider that many Sfb-projects combine qualitative and quantitative data within one study. Because of long empirical experience, our researchers know that this integration of methods provides various advantages.\(^4\)

\(^4\) Concerning the combination of qualitative and quantitative methods see: PREIN, KELLE and KLUGE 1993; ERZBERGER and PREIN 1997; ERZBERGER 1998; KELLE et al. 1998; KELLE and ERZBERGER 1999; PREIN and ERZBERGER 2000.
To give an example: The projects use the results of the first standardised inquiry to develop a sampling design for the qualitative sample. Because of the relatively small extent of qualitative samples in general, the aim is not to take a random sample but to choose specific cases for the sample according to theoretically and empirically relevant aspects (see KELLE & KLUGE 1999, p.46). Furthermore, the combination of methods make it possible to control for the intermediate results produced by one method in light of the other. At the end, the final research results achieved by the use of both methods can be tested accordingly to the following criteria:

1. Do they conform and therefore validate each other mutually?
2. Are they complementary when illuminating different aspects of the research task?
3. Do they contradict each other? In this case, one has to examine if either one method has produced unreliable results or if the theoretical models do not cover the evidence (see KELLE, KLUGE & SOMMER 1998). [21]

In order to archive and transmit these mixed datasets, it is an advantage if both data types are available at all time and can be managed in only one database system:
• It is evident at once which data records belong together. An additional database system is not needed for the quantitative variables because the qualitative and quantitative data can be handled in one single system.
• For the appropriate analysis of the qualitative interviews, it is very helpful if the quantitative information on the individual cases is available at any time. In its standardised form, this information can be called up instantly.
• Further, new variables can be generated at any time in order to register the results of the qualitative data analysis in a compressed form.
• The variables can also be used to make so-called selective retrievals in order to compare different subgroups or to pursue other questions. [22]

3.3 The database system "QBiQ"

In order to meet these requirements, colleagues of our project programmed "QBiQ" (pronounced "cubic"): A database system with which both the interview texts, codes, codings and memos and the standardised data sets can be managed. This is why it is called "QBiQ"—because of the combination of qualitative and quantitative data. [23]

Because of its programming language "QBiQ" can be used under different operating systems and it can be passed on to other researchers as freeware. With "QBiQ" the verbal data can not only be managed easily but they can also be edited, because the most important basic functions of the common text analysis systems are integrated in the program including coding the data material, carrying out different kinds of text retrievals, searching specific terms, writing and administration of memos etc. Thus, with "QBiQ" you can also prepare the data for your own studies. [24]

Currently, we are developing additional interfaces for "QBiQ" for other text analysis systems according to the standards of statistical software packages or other database systems. It is possible to get rapid access to the interview texts and to relevant codes, text codings or memos, independently of the program with which the data providing project has worked. Up to now we have an interface for winMAX. In order to import data from other text analysis systems we have to use an external program. [25]

4. Data Protection and Ethical Aspects

Because of the far-reaching possibilities of modern data processing and the sensitivity of qualitative interviews, requirements of data protection and ethical aspects must be considered very well during the archiving process in order to protect the right of informal self-determination. Therefore, the data material must be anonymised very carefully and further precautions should be taken (see also KLUGE & OPITZ 1999). [26]
4.1 Anonymisation

Normally, personal information like the name or the postal address is already anonymised during the transcription of the interviews in order to prevent an identification of the interviewed persons. This procedure corresponds to regulations which demand an anonymisation as soon as possible according to the research purpose (see §40/3 Bundesdatenschutzgesetz (BDSG), the German Data Protection Law). Qualitative interviews often contain many personal details and extensive descriptions of life events. Therefore, information or individual passages sometimes must be deleted and/or be modified to prevent any identification. If any cases can not be anonymised in accordance to these principles, they cannot be filed or passed. [27]

Transmitting research data material to other scientists is only possible, if the interviewed persons can not be identified any more or only with a disproportionally high expenditure of time, money and manpower (see §3/7 BDSG). If such degree of anonymisation can not be guaranteed, the data transmitting should only be allowed with the agreement of the interviewees. [28]

4.2 Consent of interviewees

For ethical reasons qualitative interviews generally should only be filed with the consent of the interviewees (see also HOPF 1991, 2000). It is necessary to inform the interviewees of the whereabouts of their interviews and in order not to reduce them to a "purely information object" (see BVerfGE 65, p.48; METSCHKE 1994, p.16; HAMM & MÖLLER 1999, p.7). Such a proceeding also corresponds to the guidelines of the Ethical Codex of the German Society of Sociology (Deutsche Gesellschaft für Soziologie, DGS) and of the Association of German Sociologists (Berufsverband Deutscher Soziologen, BDS) that explicitly requires:

- that the individual rights of persons included into sociological investigations should be respected as well as their right of free decision in participating in a research project (ETHIC CODEX 1998, p.81) and
- that participation in sociological investigations generally should be voluntary and carried out on the basis of complete information about the aims and methods of the research project (ETHIC CODEX 1998, p.81). [29]

4.3 Use for research purposes only

It is obvious that the data may be passed on only for investigation purposes. The data protection law in Germany allows no exception to this rule6. [30]

---

6 “Personal data collected or stored for scientific research may only be used or processed for purposes of the scientific investigation.” (§40/1 BDSG)
4.4 Further regulations

Data users should sign a contract including the following regulations:

- The interviews may only be used for investigation purposes (§40/1 & 2 BDSG) and only for the written agreed purpose, that means only for a concrete research project.
- Data material may not be passed to third persons and must also be stored in a way that external persons gain no access (see also §5 BDSG7).
- Any attempt for de-anonymisation of the data is forbidden (§40/3; see also ETHIC CODEX 19988).
- Personal details are not allowed to become quoted or published (§40/4 BDSG).
- The end of research must be declared and the borrowed material must either be destroyed or given back to the archive or to the data providing project.
- These regulations apply to all members of the research team who have access to the data (see ETHIC CODEX 1998, p.82). [31]

4.5 Conditions for access and transmitting of data

Although the data can be stored by the (original) researchers, qualitative data material in the long run should be deposited in a professional archive. Here it can be stored more safely and the access to the data material as well as its transmitting can be controlled much better. Generally, transmitting of data should be very restrictive and it has to be decided from case to case who may or may not have access to the data. [32]

5. Summary

The computer-assisted archiving of qualitative data material not only reduces the great amount of work and storage requirements, but also accelerates and facilitates the transmitting of this material considerably. It also provides an easy access for academic institutions and small research studies. Particularly, if the data providing project used a text analysis system and additional information like codes, codings and memos is also available. [33]

The computer-assisted archiving of qualitative data does not only facilitate the access to already collected qualitative material but it also increases the transparency of the research process. This facilitates the validating of the research findings. Thus, qualitative research methods can be established as a permanent part of empirical social research. [34]

7 §5 BDSG regulates the data secrecy: “The persons who are engaged in the process of data analyse are prohibited to use personal data unauthorised (data secrecy). (…)”

8 “The anonymity of the examined persons has to be preserved.” (ETHIC CODEX 1998, I B 5, p.81)
However, the protection of the interviewees have to be taken into account. To preserve the right of informal self-determination during the archiving and transmitting of biographic interviews, the interviewees must agree to the use of their data. The (future) data users have to submit to the same safety rules as the researchers of the data providing project. [35]

The risk that other researchers would attempt to identify the interviewees must be regarded as very small, even if a misuse of the data can never be completely excluded. Normally, researchers have no interest in re-anonymisation, as they are not interested in individuals but in general results. [36]

References


Hopf, Christel (2000). Forschungsethik und qualitative Sozialforschung. In Uwe Flick, Ernst v. Kardoff & Ines Steinke (Eds.), Qualitative Sozialforschung (pp.589-600). Reinbek: Rowohlt.


Computer-aided Archiving of Qualitative Data with the Database System "QBiQ"


Authors

Susann KLUGE
Dr. phil., Dipl.-Sozialwiss.; born in 1963; study of Sociology, Psychology and Law in Bremen, Münster and Passau; since 1992 at the Special Collaborative Centre 186 "Status Passages and Risks in the Life Course" of the University of Bremen.

Main research fields: methodology and methods of qualitative social research, empirically grounded construction of types and typologies in qualitative social research, integration of qualitative and quantitative methods, computer-aided methods of qualitative social research, archiving and anonymisation of qualitative data, women and gender studies.

Diane OPITZ
Dipl.-Sozialwiss.; born in 1972; study of Social Sciences in Potsdam and Berlin; since 1998 at the Special Collaborative Centre 186 "Status Passages and Risks in the Life Course" of the University of Bremen.

Activity fields: archiving and anonymisation of qualitative data, development of a meta-databank, integration of the databank in the WWW.

Contact:
Dr. Susann Kluge
Sonderforschungsbereich 186
"Statuspassagen und Risikolagen im Lebensverlauf" der Universität Bremen / Bereich Methodenentwicklung und EDV
Wiener Str. / FVG-West
Postfach 330 440
D - 28344 Bremen
Phone: +49 / (0)421 / 218 - 4166, - 4169
Fax: +49 / (0)421 / 218 – 415
E-mail: skluge@sfb186.uni-bremen.de
URL: http://www.sfb186.uni-bremen.de

Contact:
Dipl. Soz.wiss. Diane Opitz
Sonderforschungsbereich 186
"Statuspassagen und Risikolagen im Lebensverlauf" der Universität Bremen / Bereich Methodenentwicklung und EDV
Wiener Str. / FVG-West
Postfach 330 440
D - 28344 Bremen
Phone: +49 / (0)421 / 218 - 4166, - 4169
Fax: +49 / (0)421 / 218 – 415
E-mail: dopitz@sfb186.uni-bremen.de
URL: http://www.sfb186.uni-bremen.de

© 2000 FQS http://www.qualitative-research.net/fqs/
Citation


Revised 7/2008