

## **VERBATIM: Qualitative Data Archiving and Secondary Analysis in a French Company**

*Dominique Le Roux & Jean Vidal*

**Key words:**

qualitative data  
analysis,  
qualitative data  
archival resource,  
CAQDAS,  
secondary  
analysis

**Abstract:** Archiving qualitative data analysis and secondary analysis, first developed in the USA and in the Northern European countries, is not a current practice in France. To our knowledge, we are the first to have developed it in a French context. Within the Research and Development Department of the French electricity company, Electricité de France, a social science research group has the task of undertaking various qualitative surveys in order to better understand customers' requirements and the problems encountered by the company's employees. The aim of *VERBATIM* is to keep trace not only of these studies but also of the interviews carried out in this context in order to be able to re-use them in different ways. We are presenting here the results of an experiment undertaken two years ago, the way it was carried out, the problems posed, mainly ethical and methodological, and also future development plans.

### **Table of Contents**

- [1. Introduction](#)
- [2. Context for the Development of \*VERBATIM\*](#)
  - [2.1 Social science research in an electricity company](#)
  - [2.2 Model of the qualitative data archiving database](#)
- [3. Collection and Organization of the Data](#)
  - [3.1 Problems encountered](#)
  - [3.2 Presentation of the database](#)
- [4. Research and Data Treatment Tools](#)
  - [4.1 Available searches](#)
  - [4.2 A dedicated tool for qualitative data analysis](#)
- [5. Re-use of Qualitative Data](#)
  - [5.1 First results](#)
  - [5.2 Advantages and difficulties of secondary analysis](#)
- [6. Future Developments Anticipated](#)
- [References](#)
- [Authors](#)
- [Citation](#)

## **1. Introduction**

Quantitative data have for a long time, in France and elsewhere, been capitalized on in term of archiving, re-use, and development of secondary analysis methodologies. On the other hand, this same kind of operations applied to qualitative data, while well represented in the world of Anglo-Saxon sociology, remains rare in France. Different kinds of constraint explain the reticence of the French to launch into the archiving of qualitative data, such as: more restrictive

laws regarding the protection of individual data; half-hearted contacts with Anglo-Saxon sociology; and especially methods of using computers for aiding qualitative analysis which make the analysis of large amounts of data and cooperative work easier. Moreover, though sociologists in universities have already been thinking about it, they have not taken action. Working in a company and urged by the necessity of an industrial context, we had to try out the experience of archiving data. [1]

We thought it would be useful to share here our experience, spanning 2 years, of archiving and re-using qualitative data, carried out in an company environment. In concrete terms, this experience brought about firstly the development of a database called *VERBATIM*, and secondly the establishment—in the course of developing it, a methodology for secondary analysis which relies on qualitative data from the capitalization base. This paper presents the elaboration conditions of *VERBATIM*, the benefits and the limitations of the experience, and, finally, the prospects for evolution. [2]

## **2. Context for the Development of *VERBATIM***

### **2.1 Social science research in an electricity company**

For about twenty years now, the GRETS group, which is part of the Research and Development Department of Electricité de France, has brought together sociologists, statisticians, semioticians and linguists who are working on the development of social sciences for the needs of the company and are undertaking studies on behalf of the operational services within it. The subjects tackled cover many different fields, such as: the internal sociology of organizations (contributing thus another angle on looking at the way the company functions internally), understanding customers needs, understanding partners who, in collaboration with the employees provide electricity services. [3]

The methods used combine the quantitative and qualitative aspects, the most recent of which we shall develop specifically in this paper. Each qualitative survey gives rise to the collection of semi-structured qualitative interviews which follow an interview guide and are carried out among diverse population groups such as customers or partners in the installation of electrical equipment. The analysis of these interviews (which are recorded and then mostly transcribed) has allowed a large number of studies to be carried out since the department was founded. But after completion of a study, until a recent date, only the results had been archived and the complete absence of archiving or availability of qualitative data had prevented any possibility of re-use of these data. At the end of 1997 the possibility of following the example of a center such as [Qualidata](#) at the University of Essex (CORTI & THOMPSON 1998) was presented as a challenge. [4]

## **2.2 Model of the qualitative data archiving database**

The database *VERBATIM* (developed on Lotus Notes) has enabled us to collect:

1. the identifying description of the study;
2. the methodology used and the interview guide;
3. the interviews carried out in the course of the qualitative studies;
4. the results obtained;
5. the list of past requests for information from the database and of past results of secondary analyses;
6. essential articles and references to the Web pages concerned regarding archiving methodology for qualitative data, qualitative analysis and the use of the CAQDAS (Computer Assisted Qualitative Data Analysis; see e.g. JENNY 1997). We needed to create a bank of past qualitative studies and an information resource that could be shared between and re-used by the researchers in the department. [5]

## **3. Collection and Organization of the Data**

### **3.1 Problems encountered**

The first part of the *VERBATIM* project concerns the collection and organization of the data. In the first instance, it was a matter of recovering from researchers and the archives the maximum number of ancient qualitative interviews (from 1994 to 1998) in an electronic form, equipped with the necessary context information for re-use. After that, procedures had to be set in order to collect the recently done interviews in a standardized way. This is what we are presently working on, through reviewing the practices of sociologists in the department. This task, which is set within a quality assurance plan by the department, must end up with the documents being presented in a standardized form, by using a guide for transcribing interviews on cassettes in order to make their re-use easier. [6]

This phase of the project, is, to all intents, quite a simple one. Nevertheless, it posed numerous problems of varying types. First, as when any new technology in information and communication is introduced, sociologists needed to be associated with the development of the project, their wishes, fears, warnings listened to and safeguards put in place to avoid any blunders. This is why, for developing the database, we selected the principle of rapid modeling which allows continual exchanges between the users and the designers. It was also necessary to preserve the dynamics of the project, to prevent it from sinking into the forgotten depths of aborted developments. [7]

The main barriers to the development of the experiment are as following:

1. Ethical problems involving the necessity of ensuring that anonymity is totally maintained in its most extreme form (i.e. not only must the qualitative data be anonymous but it must also be impossible to recognize an interviewee from the sociological data characterizing him. This problem is particularly crucial where internal inquiries are concerned or when the professionals involved would be easily identifiable);
2. The sociologists' fear that this may give rise to some re-use of the contents of the database poorly controlled in respect of the methodology;
3. The importance of the contract made at the start of the interview between the sociologist and the interviewee;
4. The feeling certain researchers have that they are the exclusive owners of the data they themselves have collected (which is not true in the framework of a company). [8]

The very sensitive ethical and legal problem to be sorted out, especially in the context of the absence of any French example concerning re-use of qualitative data, was examined by various researchers and legal advisors. A charter for the use of the data archived in the *VERBATIM* database acts as a safeguard for the re-use of interviews and is subject to checks by the head of department. To ensure the security of the archived documents, access to the database has been made secure and provides different levels of access to the data depending on the status of the user (a manager, supplier of data, departmental researcher) and on the degree of confidentiality of the data (obviously data collected from employees about the management of the company are more sensitive than information about the attitude of users to air-conditioning). This ethical charter is at present undergoing revision in order to be able to respond to new cases for use. [9]

Another problem encountered, but one nearly resolved, is that of the transcription of qualitative data. The absence of any policy until recently on archiving qualitative data had meant that there was a great heterogeneity in transcription methods for interviews, especially in regard to the status of the transcription (selective note taking, complete transcription), the representation of the socio-demographical data (sometimes very poorly described) etc. [10]

### **3.2 Presentation of the database**

The size of the database is at present 130 Mb. It comprises 65 studies from 1994 to 2000. It is continually increasing in size and is structured on the model shown by illustration 1:

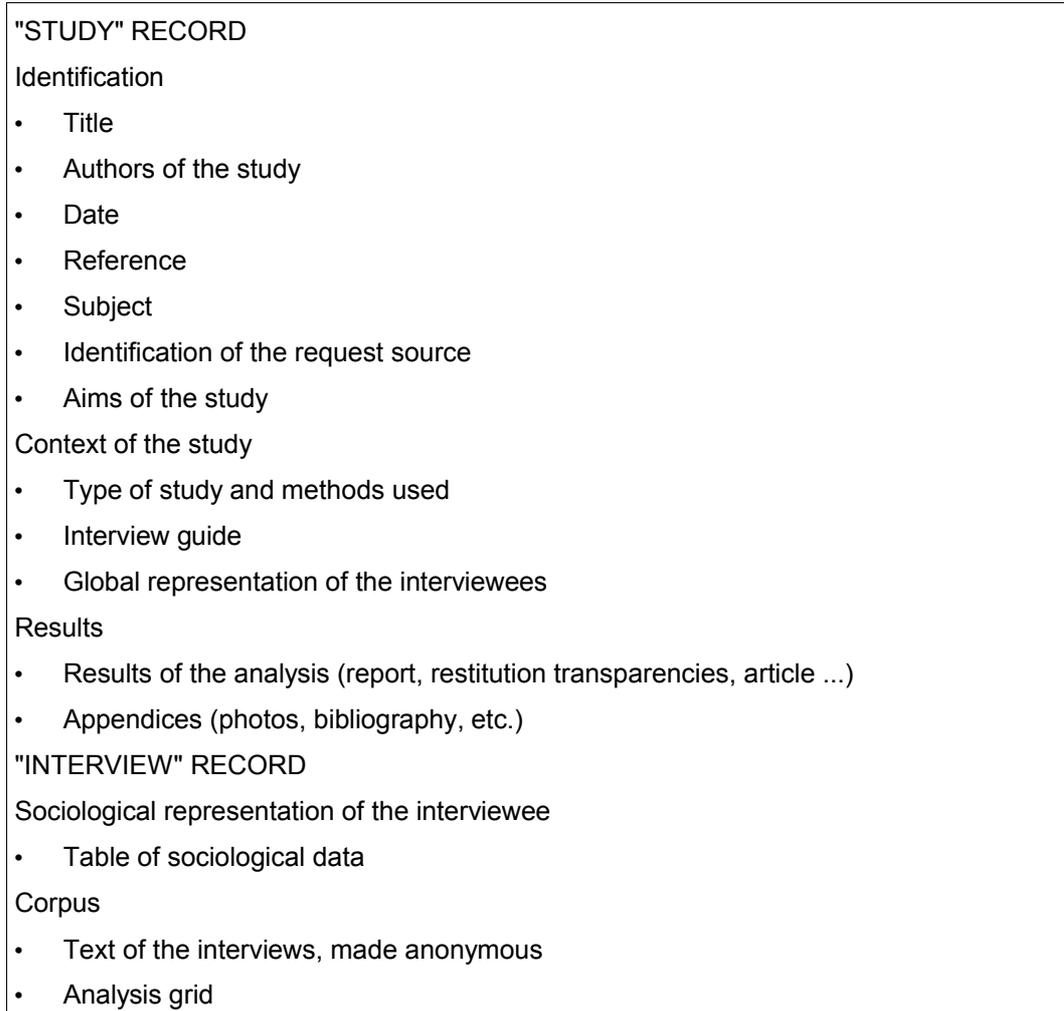


Illustration 1: Architecture of Verbatim [11]

There is one record per study which gives all the information considered necessary for proper re-use of the data. Attached to each "study" record are several "interview" records which contain both the socio-demographical characteristics of the person interviewed and the text of the interview transcribed and made anonymous. These "interview" records can be exported directly to a tool for computer-aided qualitative data systems such as Atlas-ti, that we chose for carrying out our analyses. [12]

There are four socio-demographical data presentation models:

1. user (for feedback from experiences of the applications of electricity);
2. EDF employee (for a better understanding of their problems);
3. resident (for problems concerning conflicts about high-tension line construction, for example);
4. external actor (for people outside the company who play an important role). [13]

## 4. Research and Data Treatment Tools

### 4.1 Available searches

Lotus Notes, apart from the advantage it has of being used within the whole company, has shown itself to be a convenient tool enabling transverse access across the data (LE ROUX & VIDAL, 2000). It is thus possible to define views depending on the fields defined in the database

1. search by type of study or methodology;
2. search by author's name;
3. search by subject (from a descriptive list *electrical vehicle, heating, council housing ...*). [14]

It is also possible to move around in the database searching for specific information, for example, to list all the interview guides produced, for methodology research purposes, to build up a collection of texts from the interviews archived on each record. [15]

Finally, integral text searches can be carried out using the indexing tool provided by Lotus Notes. This tool is applied to the entire database including items forming appendices in Word format. On the other hand, in the case of the latter, the words indexed do not appear highlighted which does make research within the text more difficult. Lotus Notes is always evolving and it is hoped that this problem will soon be overcome. [16]

As regards to tools for the treatment of data, there are two major types of software which can be used in a complementary fashion:

1. software helping to sweep and to analyze the collection of texts. Certain are already used at GRETS such as Alceste, Tropes (BRUGIDOU & LE QUEAU 1999), Discursus (BRUGIDOU et al. 2000); their task is to deal with large volumes in order to extract trend indicators;
2. software assisting the qualitative analysis of the interviews such as CAQDAS. Their objective, compared with the previous types, is to enable a more detailed analysis, including both coupling of sociological data and the added value of sociological interpretation. Very close to sociological practice, these software are aimed at facilitating the handling of fragments of text and the creation of links with subjects of the analysis plan in order to test hypotheses (BUSTON 1997). They also make it possible to go through qualitative data in a team from which a creation process of richer meaning might be expected because knowledge and individual experiences are mutually enriching and create added value and individual cognitive slants are toned down (FOUET 1997). [17]

## 4.2 A dedicated tool for qualitative data analysis

[Atlas.ti](#), developed by Thomas MUHR (Scientific software, Berlin), has been selected by the team to assist with qualitative analysis. A presentation guide and basic manual in French has been written to meet the needs of sociologists. It has been tested in full on several studies, chiefly on the analysis of feedback messages coming from employees involved in resolving the crisis that followed the major storms of December 1999. We hope that sociologists will continue to use the software in the future. [18]

## 5. Re-use of Qualitative Data

### 5.1 First results

Secondary analysis enables a mine of data to be interrogated transversely: these data have been gathered as a function of research objectives and work hypotheses other than those undertaken in a secondary way (DALE 1993). [19]

For example, one of the studies currently being carried out aims to obtain better knowledge of the different partners of EDF employees concerned with the delivery of electricity (installers, architects, promoters, wholesalers etc.). Up until now, no study has focused directly on this subject. On the other hand, a test search in the *VERBATIM* database permitted numerous interviews to be found which had been carried out in the context of various collections of feedback and which contained information very relevant to this subject, such as judgments made by customers on the different partners acting with the company, opinions given by these partners themselves about their own mission or those of other participants, diagnoses made by EDF employees concerning malfunctions or successful activities undertaken in partnership. [20]

Another study dealt with how the customers who have air-conditioning system understand the health problem: the fears they had before acquiring the system (air conditioning has the firm reputation of encouraging colds), whether these fears have been overcome after using their appliance, whether they have other health problems that might be caused by air conditioning. The *VERBATIM* database was asked these questions and more and interrogated several studies the subjects of which had nothing specifically to do with health. In some cases the interview guide asked the question, in other cases the subject had been touched on spontaneously by the people being interviewed. The results are always interesting. Certainly they do not replace the richness provided by a new study but they allow an idea to be obtained about the question and the hypothesis to be laid for a possible new survey. [21]

In all cases interesting results can be obtained for our work in return for always respecting certain basic methodological rules. Overall they are matters of good sense and methodological stringency which are equally essential in the framework of a "primary" analysis. [22]

## 5.2 Advantages and difficulties of secondary analysis

Secondary analysis of qualitative data, as we have said, does appear to be current practice, but not in the context of French sociology. As with all practices, secondary analysis has its advantages and also its limitations and one must be aware of this to be able to make valid use of the data. [23]

Among the advantages are (DALE 1993):

1. the immediate availability of data that are ready for use, enabling either results to be obtained without having to implement the slow and costly process of a survey, or conditions for a new survey, if one is found to be necessary, to be better defined (the last part does not make sense);
2. the possibility of carrying out past studies to a fruitful conclusion;
3. the possibility of applying questions to a wider population than that which a single qualitative survey can cover;
4. the possibility on the contrary of creating sub-populations of specific groups. [24]

Among the disadvantages of secondary analysis, which are due essentially to the way the data are collected are the following:

1. the value of a file of data depends on the quality of the data collected: the researcher is not in any way taking part in either drawing up the questionnaire or in gathering the interviews;
2. one has to be content with the data available and adapt one's question to the limitations imposed by the former (DALE 1993); given the distance of the analyst from the phase when the data were constructed, one of the obvious dangers is of losing from sight the context in which the original study was undertaken (the age of the data, the economic and social context of the period, the type of questioning etc.). [25]

The theoretical problems posed by secondary analysis are very real and have furthermore been the subject of several studies (DALE, ARBER, GAMBLE & PROCTER 1988), (HAKIM 1982). Though it is always easy to draw interpretations from a text, it is on the other hand more complex when it comes to controlling this hermeneutics in order to make the results legitimate. In order not to lead to false conclusions, a certain number of safeguards must be set up, among which the following have been suggested:

1. a good knowledge in advance of the area of study and of the practices of qualitative analysis (including therefore the necessity of asking oneself the question, "Who uses the database?" before envisaging any sharing of the data with anyone else;
2. when possible, consultation by the analyst in charge of a secondary analysis with the sociologist responsible for the original study;

3. validation of the hypotheses by coupling them with quantitative data;
4. sharing of working data (interview grid, non-modified interviews, sorting grid etc.) which ensures an indirect validation by confrontation. [26]

At the present time we are working to lay down the criteria necessary to ensure good methodological practice of re-use of these secondary analyses. They will be presented in the next version of the charter for use of *VERBATIM*. [27]

The secondary analysis practice in the GRETS department can be represented in the following way (Illustration 2):

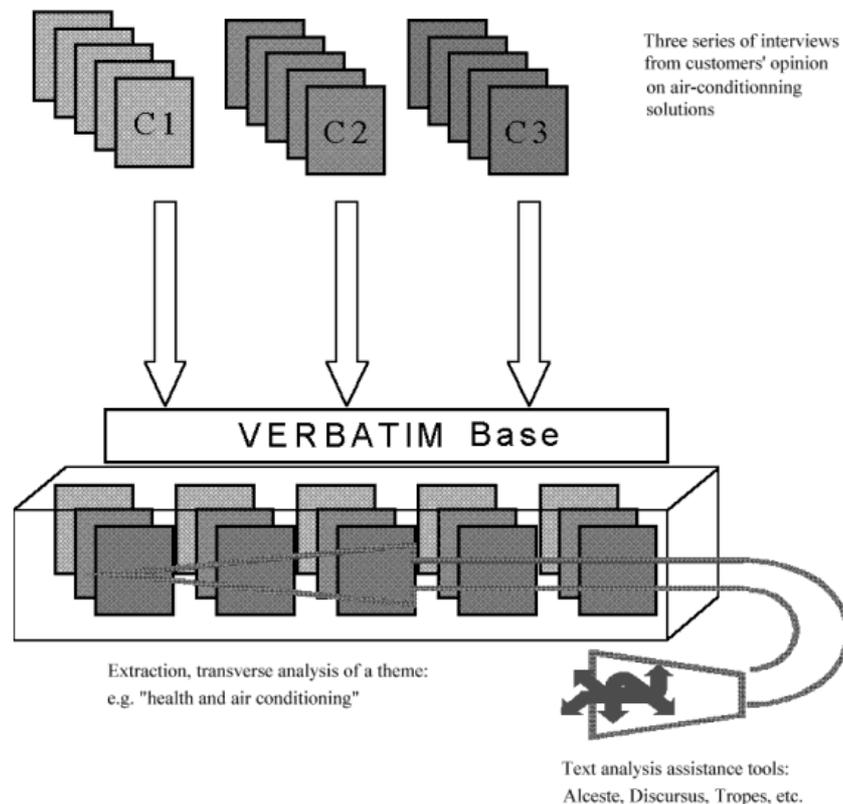


Illustration 2: Corpus extraction and transverse analysis [28]

## 6. Future Developments Anticipated

Now that the architecture of *VERBATIM* has been defined and the system has been evaluated, the two points to be looked into more closely are basically:

- Firstly, as we have already seen, the increasing complexity of the charter for use of the data. This charter has to be adapted for collaboration between several entities in the company, which makes the problem of the management of confidential material considerably more complex on both the technical and administrative levels;
- secondly, the establishment of a methodology for secondary analysis adaptable to our needs. A secondary analysis on the subject of comfort will

enable the methods used to carry out the work to be described. To this effect, collaboration with the University of Grenoble is planned. [29]

In addition, we intend to explore different types of extraction tools such as as Nomino, developed by Centre ATO (Université du Québec à Montréal, UQAM), in order to obtain a more discriminating search in the database. Finally, we expect a great deal from this conference, moreover, in order to have a worthwhile exchange about these practices for archiving and re-using qualitative data with other researchers. [30]

## References

- Brugidou, M. & Le Queau, P. (1999). Les "rafales", une méthode pour identifier les différents épisodes du récit : contribution au traitement et interprétation des entretiens non-directifs de recherche. *Bulletin de méthodologie sociologique*, 64, 49-82.
- Brugidou, M.; Escoffier, C.; Folch, H.; Lahlou, S.; Le Roux, D.; Morin-Andreani, P. & Piat, G. (2000). Les facteurs de choix et d'utilisation de logiciels d'analyse de données textuelles, JADT 2000. *Journées internationales d'analyse statistiques des données textuelles*, 373-380.
- Buston, K. (1997). NUD\*IST in action: its use and its usefulness in a study of chronic illness in young people. *Sociological Research Online*, 2(3), <http://www.socresonline.org.uk>.
- Corti, L. & Thompson, P. (1998). Are you sitting on your qualitative data? Qualidata's mission. *International Journal of Social Research Methodology*, 1(1), 85-89.
- Dale, A. (1993). Le rôle de l'analyse secondaire dans la recherche en sciences sociales. *Sociétés contemporaines*, 14-15, 7-21.
- Dale, A.; Arber, S.; Gamble, J. & Procter, M. (1988). *Doing secondary analysis* (Contemporary Social Research Series). London: Unwin Hyman.
- Fouet, J.M. (1997). *Connaissances et savoir-faire en entreprise. Intégration et capitalisation*. Paris: Hermès.
- Hakim, C. (1982). *Secondary analysis of social research*. London: George Allen & Unwin.
- Jenny, J. (1997). Méthodes et pratiques formalisées d'analyse du contenu et de discours dans la recherche sociologique française contemporaine: état des lieux et essai de classification. *Bulletin de Méthodologie Sociologique (BMS)*, 54, 64-112.
- Le Roux, D. & Vidal, J. (2000). Verbatim, une expérience de capitalisation d'entretiens qualitatifs. *Bulletin de Méthodologie Sociologique (BMS)*, 65, 58-67.

## Authors

**Dominique Le ROUX**: Dr in Linguistics and information science, Research Engineer in the Social science research group in the Research and Development Division of Electricité de France. Manager of the *VERBATIM* project (Qualitative data archiving and secondary analysis)

Contact:

Dominique Le Roux  
EDF / PI / DRD / CLEO / GRETS  
1 avenue du General de Gaulle  
92141 CLAMART Cedex  
France

E-mail: [dominique.le-roux@edf.fr](mailto:dominique.le-roux@edf.fr)

**Jean VIDAL**: Senior consultant, Research Engineer in social sciences, Marketing Division of Electricité de France

Contact:

Jean Vidal  
EDF / PC / DMK / DE  
3 rue de Messine  
75017 PARIS Cedex  
France

E-mail: [jean.vidal@edf.fr](mailto:jean.vidal@edf.fr)

FQS 1(3), Art. 15, Dominique Le Roux & Jean Vidal:  
VERBATIM: Qualitative Data Archiving and Secondary Analysis in a French Company

## Citation

Le Roux, Dominique & Vidal, Jean (2000). *VERBATIM: Qualitative Data Archiving and Secondary Analysis in a French Company* [30 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3), Art. 15, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0003150>.