



HAL
open science

ArchEthno - a new tool for sharing research materials and a new method for archiving your own research

Florence Weber, Carlo Maria Zwölf, Arnaud Trouche, Agnès Tricoche, José
Sastre

► To cite this version:

Florence Weber, Carlo Maria Zwölf, Arnaud Trouche, Agnès Tricoche, José Sastre. ArchEthno - a new tool for sharing research materials and a new method for archiving your own research. *Journal of Data Mining and Digital Humanities*, 2024, Atelier Digit Hum, 10.46298/jdmdh.11617.hal – 04166298v3

HAL Id: hal-04166298

<https://hal.science/hal-04166298v3>

Submitted on 16 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

ArchEthno – a new tool for sharing research materials and a new method for archiving your own research

Florence Weber^{1*}, Carlo Maria Zwölf², Arnaud Trouche³, Agnès Tricoche⁴, Jose Sastre⁵

1 Université Paris Sciences et Lettres, Centre Maurice Halbwachs, France

2 Université Paris Sciences et Lettres, Laboratoire d'Etudes du Rayonnement et de la Matière en
Astrophysique et Atmosphères, France

3 Artenum, Paris. Science & Groupware, France

4 Université Paris Sciences et Lettres, Archéologie & Philologie d'Orient et d'Occident, France

5 Université Paris Sciences et Lettres, France

*Corresponding author: Florence Weber florence.weber@ens.psl.eu

Abstract

The archiving of ethnographic material is generally considered a blind spot in ethnographic working methods which place more importance on actual investigations and analysis than on how archives are constructed. A team of computer scientists and ethnographers has built an initial tool for sharing ethnographic materials, based on an SQL relational data model that suited the first survey processed but proved difficult to transpose to other surveys. The team developed a new tool based on dynamic vocabularies of concepts which breaks down archiving into three stages. Firstly ethnographers can select and contextualise their survey materials; secondly they structure them in a database according to the research question discovered during their survey; finally, they share this data with other researchers subject to the opinion of an ethics committee whose members are competent in ethnography.

keywords

ethnographic method; survey materials; database; ethics committee; epistemology; data sharing; multidisciplinary research

INTRODUCTION

From the 2000s onwards, it became extremely difficult for sociologists and anthropologists to defend their publications which were increasingly being questioned by respondents to their surveys or other third parties. Cases with varyingly high profiles destabilised certain researchers and even the discipline as a whole. Therefore why should researchers bother with the delicate matter of sharing unpublished material? This then is the context which has led to the open data policy promoted in European-level research policies being perceived by the ethnographers' community as a pointless and even harmful injunction imposed on them from above in the name

of transparency in science [*Genèses*, 2022]. Ethnographers are artisans who provide a specific type of material namely their field notes based on ethnographic encounters [Pina Cabral, 2011]. These are distinct from the data constructed by large public or private organisations and also from archives the preservation of which depends on political decisions. The validity of ethnographers' findings is now subject to peer review which is not always competent to judge the ethnographic approach and also to multiple forms of review by the people concerned by the research, whether as interviewees or because they feel authorised to speak on behalf of those who were.

These discouraging observations risked calling into question the actual exercise of the ethnographer's profession. This situation led to a team of ethnographers, sociologists, anthropologists and photographers working with data base and computer engineers to develop a method to protect respondents and ethnographers against the circulation of responses and practices which ethnographers have access to because a relationship of trust has been established [Béliard & Eideliman, 2008]. The intention was also to protect them from suspicions and tactlessness within the collective ethnography teams themselves [Laferté, 2016] and in the public space [Avanza, 2016]. This work took as its starting point a multidisciplinary collective survey whose ethnographic materials were archived in print format. The method for achieving this form of protection is based on an initial sideways step. It does not address individual ethnographers and instead relies on partial pooling of ethnographic materials self-regulated by professional ethnographic researchers and academics in compliance with ethnographic deontology and scientific ethics.

Producing these materials is expensive in terms of time and emotional cost as is their analysis. It is similarly costly to format them for re-use. The fundamental idea underpinning our team's work is that such materials are a useful resource for other research, particularly multidisciplinary research. We work on the complex task of making these materials available to researchers from other disciplines and in this way are breaking with the dichotomy established over the last forty years in the worlds of science and art in the name of the materials' heritage value. This dichotomy posits the idea that the value (and even the price) of certain materials is immense if their producer is famous and zero if the producer is unknown. It was imposed on the scientific community from outside and has led to the destruction of a great deal of material. Within the scientific community itself, it has created a certain form of impunity as regards the theft of research material when creators do not have the means to publish their results quickly. Our method proposes a profound modification in the way materials are disseminated with their value being linked to their quality rather than to the status of their producer. This quality is gauged by the work invested in making them available and by the scientific interest of their re-use. This form of transformation enables us to re-open the question of the cumulative nature of knowledge produced by the sciences studying societies.

I IS ARCHIVING ETHNOGRAPHIC MATERIALS IN 2020 IN A DEADLOCKED SITUATION?

Who are ethnographic materials archived for and why and how is this done? Over the last twenty years or so, these questions have been the subject of much discussion in the international scientific community working in the two disciplines that have forged the ethnographic tradition since the nineteenth century - social and cultural anthropology (which is commonly referred to as ethnology in France) and the so-called qualitative sociology.

1.1 A context of conflict

1.1.1 *The 2000s – a turning point*

Many post-publication conflicts between researchers and survey respondents marked the beginning of the 21st century, revealing the structural weakness of ethnographers as academics and social scientists studying the contemporary world in different socio-historical contexts and situations. To give just a few examples, in 2006 the Quebec anthropologist Natacha Gagné was challenged on the international scientific scene by other anthropologists claiming that a non-Maori had no right to speak about Maori cultures [Gagné, 2008]. A few years later, her university did not defend her when a former survey respondent made an accusation against her. In France, the Thérème affair [Weber, 2008] highlighted the different issues facing an interview respondent who was also a writer and a sociologist of literature working on an analysis of the field of 1970s French literature. The former's self-image and reputation as an author were at stake while the latter was more concerned by her contribution to scientific knowledge of the historical risks linked to the construction of value in the literary field. There was a mix-up involving two judgement criteria – firstly did the sociologist damage the writer's self-image and reputation? And secondly does the writer's complaint invalidate the increase in knowledge her work provided? This situation could undoubtedly have been avoided if the sociologist had asked the writer to read her article before publication or if a judge had been able to help them come to a negotiated compromise.

An initial summary report of the situation in France [Laurens and Neyrat, 2010] showed that controversies tended to be much more violent when they had a political dimension or highlighted the structural weakness of ethnographers in unambiguous power relationships, for example relations between a boss and employees carrying out an incognito investigation. The Alice Goffman affair in 2015 combined all these ingredients – a woman, a brilliant anthropologist and heir to Erving Goffman's reputation, was dragged through the mud. She was accused of being complicit in abuses committed by her interviewees while she was observing them and then actually accused of lying [Avanza, 2016]. The mix of morality, politics, judicialisation and criminality became a white-hot issue.

1.1.2 *Two anthropological traditions that know nothing of each other*

Do these conflicts testify to a shift in the balance of power between observers and the observed or even peer judgement being worryingly replaced by public judgement? To judge this, we need to take into account the existence of two distinct anthropological traditions. Anthropology's editorial, encyclopaedic and museographic traditions have long promoted a genuine policy of publishing and conserving heterogeneous materials, texts, sound and visual recordings that are often linked to objects. This tradition dates from the earliest ethnographic expeditions and has successfully alimented some splendid museographic experiments around the world¹ as well as grandiose encyclopaedic hopes². In France, several national institutions such as the *Muséum d'Histoire Naturelle*, the *Musée de l'Homme*, the *Musée du Quai Branly* and the *Musée des Arts et Traditions Populaires* which has been succeeded by the MUCEM³ are at the forefront in showcasing these ancient collections. Their collections tend to be of more interest to social science historians than to contemporary ethnographers. An exception would be cases in which they engage in dialogue with movements that aim to revive indigenous arts and customs [Le

¹ Spencer and Gillen. *A journey through Aboriginal Australia*. 2009: http9 (access 15.12.2023)

² *Human Relations Area Files*. Yale University (New Haven). 2023: http6 (access 15.12.2023)

³ Their collections feature on CALAMES, the Online Catalogue of Archives and Manuscripts in French University and Research libraries, see for example the archives of the *Laboratoire du Musée de l'Homme*. 2007: http4 (access 15.12.2023)

Gonidec, 2020]. Another tradition stemming from Malinowski's first survey published in 1922 merged the two roles of investigator and researcher in a single ethnographer-author role [Stocking, 1983]. This tradition spread in France in a form of opposition to the museum tradition [Segalen, 2019]. For a long time it was spared the radical criticism of the scientific pretensions of social and cultural anthropology that occurred in the United States and Great Britain.

1.1.3 *Incertitude about the status of field diaries*

An ethnographer's most important prime material is his or her field diary but this tool has never had a unified status. The publication of Malinowski's '*A Diary in the Strict Sense of the Term*' [Malinowski, 1967] written between 1914 and 1920 created shockwaves. The ethnographer revealed the hidden sides of a revered and radiant work, *Argonauts of the Western Pacific*, the ethnographers' bible to this day [Malinowski, 1922]. His diary was not initially intended for publication and the author revealed therein his doubly difficult relationship with the members of the colonial society he was forced to mix with and with the natives of the colonised societies he was observing and studying by vocation. Anthropologists were often careful to destroy their diaries – at least partially – to protect their allies in the field and themselves in conflict situations. This was notably the case in colonial contexts like for example Evans-Pritchard [Grootaers, 2001] or criminal contexts for example Tarrus [2017] or Alice Goffman [Yanow, 2021]. The distinction between a personal diary and an investigative research diary [Weber, 1991] coincided with the ethnographic method spreading beyond anthropology [Beaud & Weber 1993, Parent & Sabourin, 2016] to reach the field of design [Lécho Hirt et al., 2015].

1.2 **The promotion of tools that are unsuited to the ethnographers' inductive approach**

1.2.1 *The lexicon of qualitative issues*

From 1994 to 2020, ethnographers were encouraged to deposit their materials to facilitate their re-use through several major initiatives aimed at depositing and sharing ethnographic materials. The most important initiatives of this kind in Europe were Qualidata in Great Britain⁴ and Bequali in France⁵. Taking stock of these initiatives shows they have all come up against the following difficulties:

- the reluctance of researchers to deposit what they consider to be their own professional property has led university authorities to attempt to make this legally compulsory (difficult to implement except in exceptional cases) and to offer financial incentives which is now effective but still viewed as yet another administrative constraint [Duchesne & Noûs, 2019];
- the target audience's low level of interest in re-using such materials has led the promoters of platforms of this kind to target captive audiences like students despite the risk of discouraging future ethnography researchers from using the tool;
- there have been continuing communication difficulties between researchers and archivists [Wasamba, 2013] despite some successes, for example in the field of the political history of social science research [Wolikow, 2003];
- a terminological disagreement has arisen between ethnographers who refer to materials and academic authorities who refer to data with the latter influenced by the powerful models of statistical surveys and administrative data.

⁴ Created in 1994 and absorbed into the Data UK Archive Service in 2012: <http12> (access 15.12.2023)

⁵ Created based on an experimental approach in 2010 and led by Sciences Po Paris: <http3> (access 15.12.2023)

In the social sciences, ethnography corresponds to the inductive phase of scientific research: researchers do not know precisely what they want to prove before the beginning of their fieldwork. Sometimes they have questions and choose places, events and people they think relevant to find answers. Sometimes the scientific question emerges from the analysis of the first ethnographic encounters. In no case the ethnographer is able to follow biomedical research's protocolised methods or hypothetico-deductive methods from economics [Olivier de Sardan, 2008]. Additionally, ethnography is based on a doubly private form of participant observation in which ethnographers and their interviewees commit to a greater extent than to purely professional obligations. This means ethnography now needs to redefine the circle in which its observations are shared. This is the idea foreseen by Geertz in 1988 when he asked the somewhat anxious question: "Who do we need to convince now? Africanists or Africans? And of what?" [Geertz, 1988]. Today, this question is posed in a more precise fashion. Who can we open up our intermediate results and confidential materials to and why?

1.2.2 The lexicon of data. Do open data and the General Data Protection Regulation (GDPR) constitute a double bind?

Additional difficulty has arisen because of the European-level General Data Protection Regulation (which prohibits the retention and dissemination of personal data) combined with regulations stipulating that the head of a research project is responsible for applying the law. Universities and research organisations are beginning to measure the likely impact of all this on research based on the ethnographic method. Some universities have appointed a GDPR referent to work on protecting students, particularly those working on PhDs, from potential misuses of these laws by certain employers or respondents who may aim to retain a monopoly on legitimate discourse about the population they represent. With a few exceptions, university data platforms mainly disseminate tools and methods for using quantitative data which have only recently become available. As is well known, the two models for the relationship of researchers with empirical material are statistics (reinforced by the Big Data boom) and therapeutic trials. Teachers and students in the social sciences still sometimes find themselves referred to ethics and scientific integrity committees that do not have the competence to pronounce on issues involving the ethnographic method. These committees follow the model of Institutional Review Boards in the English-speaking world which have long been badly adapted to ethnographic research [Desclaux, 2008]. This situation has led to researchers developing avoidance tactics or taking radical stances against injunctions to open up research data [Genèses, 2022].

1.2.3 Science archives – heritage or accumulation?

The French science archives sphere [Charmasson, 2006] has acquired high-quality instruments as the history of science and technology has developed, driven by major heritage institutions like the *Collège de France*, the *Bibliothèque Nationale*, *Ecole des Chartes*, prestigious libraries, national anthropology museums and the National Archives. However, the study of some of the most important archives in the history of fieldwork [Müller and Weber, 2003] has shown that it is difficult to create collections that bring together archives from among various institutions over the course of turbulent political and social histories. How should the value of research materials be assessed? According to the value of their authors? Of their subjects? Of the new questions such materials can help to explore? In a world overwhelmed by existing digital data that is however eminently fragile due to the rapid obsolescence of their writing formats, how can we create a new science of data that takes into account the skills of producers, historians, digital specialists and future users in all their unpredictable variety?

Despite misunderstandings between the different professions working on this new data science, ethnography could in fact prove an excellent example of how to bring together robust multidisciplinary teams before materials and data are made available and also before enabling other new multidisciplinary teams to appropriate these materials and data.

1.3 The development of new practices for surveys and for analysing materials

In the meantime, today's young ethnographers are working with no metaphorical safety net on developing investigation and analysis methods aimed at saving time (sometimes illusorily) and accumulating material with the potential to give them a head start in the race to publish their work. Certain of these researchers seem caught up in a specific form of quantophobia which involves obtaining as many documents 'of interest' (recorded interviews, photographs, videos, thousands of pages of field diaries, etc.) as possible despite the risk of losing all control of the thought process about the issue at hand. Others find themselves blocked by anxiety deriving from their own uncertainty about the rules they are supposed to follow in the field and in the construction of their research object. Some use their smartphones to record 'on the spot' interviews and are then faced with the current tendency to discredit 'undercover' survey practices or with technical problems linked to exporting files designed to trap their users within a commercial system. Novice ethnographers can also find themselves faced with data security and privacy issues of varying levels of justification and thus entangled in arcane GDPR bureaucracy.

Other practices are developing for editing rough field notes and classifying materials. Ethnographers are as subject as anyone to the temptations of mainstream consumer cognitive tools that promise to turn amateurs and enthusiasts into professionals for the right price. They can however find something positive for their work in this. To manage and process their proliferating research materials, they can use professional tools that are more or less adapted to their requirements such as the classification of photographs on tropy⁶. Ethnographers are at the crossroads of several disciplinary traditions in the cultural professions - linguistics, with its tools for the analysis of language material; the history of literature with its tools for the analysis of texts and the history of art with its tools for the analysis of images. They can also use office tools for automatic transcription like Sonal⁷, translation and report writing or alternatively ask themselves questions about all such tools. The method ArchEthno proposes is located between technophilic euphoria and technophobic anxiety and aims to collectively take back control of the profession of ethnographer. Like the profession of archaeologist, this profession is historically attached to writing, images and schematisation rather than computing, as [Fabry, 2021] has demonstrated.

1.4 Towards ethnographic lucidity?

The effects of these new practices on the methodology of field surveys and their results have yet to be examined or debated. The practice of open questioning enabled ethnographers to explain their academic status, their projects and their purpose to respondents and guaranteed the confidentiality of the information respondents being asked to provide. This had established itself as one of the norms of ethnography in a context of inter-acquaintance [Beaud, Weber, 2010]. It is clear that this practice was overtaken by the widespread practice of covert questioning in the 2010s combined with the sometimes overt aim of 'deceiving' to be able to 'denounce' a given issue. Proponents of this aggressive approach seem unaware of its effects and it has also been countered by forms of anxiety about confidentiality issues. We may

⁶ Presentations of tropy: <http10>; <http11> (access 15.12.2023)

⁷ A presentation of Sonal: <http8> (access 15.12.2023)

reasonably hope a new era is dawning with an ethnography that is aware of its effects. This could be based both on ethnographies in extreme conditions [Shukan, 2016] and on a renewed alliance between photography and the social sciences [Dantou et al., 2020].

II THE GENESIS AND CHARACTERISTICS OF THE ARCHETHNO METHOD

With this hope for a new era of ethnography in mind, since 2013 we have been working on the development of an initial tool for sharing structured materials within a multidisciplinary team and then on enhancing it so that it can be adapted to several different studies [Dantou, 2014, Blum, 2017, Blum, Goudet and Weber, 2022]. We have found a solution that can be adapted to materials shared in multidisciplinary teams and those produced and deposited by individual researchers. The production, transformation and uses of this tool have directly resulted in the ArchEthno method. The tool consists of a software suite with three elements that enable researchers to construct a dictionary, enter their materials and metadata locally and allow authorised persons to consult their work on the internet. These are accompanied by instructions for use and a guide to archiving ethnographic materials. The method requires the involvement of an ethics committee for researchers working in institutions or, failing that, professional legal advice.

2.1 Epistemological convictions

2.1.1 *Researchers alone should decide whether to publish or share scientific material*

A dual conviction influenced the method's development. It became clear to us at a very early stage of the project that researchers had to control the entire research cycle and also that they needed to be helped to identify the issues linked to sharing and publishing the hitherto concealed aspects of their research – from the production and collection of materials to their transformation into reusable data.

This conviction can be summed up by the following four points:

- researchers alone are able to document their own materials (each material has an identity document: who produced it, when, where and in what circumstances). We believe it is counter-productive to entrust the 'investigation of the investigation' to anyone other than researchers except in the case of historical research based on archival collections;
- researchers alone are entitled to structure their materials according to the research question being explored which changes as the survey and initial analyses progress;
- researchers alone are capable of implementing a multi-disciplinary approach based on common goals;
- researchers need to discuss the elements they feel are worth sharing with other researchers and need to be supported by an ethics committee or a legal advisor who will consider and discuss the consequences of sharing with them.

In short, the question of publishing research materials is a fully scientific issue rather than of a technical or legal nature. The technical and legal obstacles need to be overcome so researchers can then rely on the evaluation system of thesis and recruitment juries and the scientific publishing system with its reviewers and others in publishing professions so they can propose the 'material' component that is complementary to the published article or book.

2.1.2 *Constructing ArchEthno 2017 for the Medips-Alzheimer 2003 survey*

Disseminating research material requires multidisciplinary skills ranging from IT techniques (creating and structuring relational databases or websites, managing security or encrypting

confidential information) to regulatory (for example the new National Plan for Open Science) and legal aspects. Isolated researchers or those working in small teams can find it very difficult to bring together the skills required to ensure materials are effectively disseminated.

A team of researchers carried out the 2004 Medips-Alzheimer exploratory survey [Joël and Gramain, 2005, Medips 2006] which involved a large-scale collaboration between economists from LEGOS (Paris Dauphine) and anthropologists and sociologists from the Centre Maurice Halbwachs (ENS). The project received several public (the Ministry of Research's ACI Blanche Young Researchers initiative, 2000) and private (Fondation Médéric Alzheimer, 2003-2005) grants. When the print files corresponding to the ethnographic analysis of the 91 'cases' that made up the survey were deposited in the ENS library in 2011 new funding from the National Solidarity Fund for Autonomy (CNSA) helped extend the survey to other populations and this combined with the TransferS Labex's support made it possible to archive the files digitally.

Agnès Tricoche, an archaeologist and data engineer joined the team and worked with us between 2014 and 2017 on the creation of our first shared ethnographic data entry tool (ArchEthno2017) based on a relational database. The underlying data is from research carried out by LEGOS at Paris-Dauphine and the 'Surveys, Fieldwork and Theory' team at the Centre Maurice Halbwachs into the vital family and professional care provided for fully dependent people. This 'Medips-Family' research was carried out in several phases. The most standardised phase involved a survey carried out in 2004 by a dozen or so interviewers of 91 people with Alzheimer's-type disorders contacted through 5 institutions in the Paris region.

For each case involving a person receiving care, the survey material was made up of a set of questionnaires completed by the people contacted, the interviewer's field diary and a summary sheet. A number of additional documents on the institutions were involved in the survey. The survey was based on an original combination of the ethnographer's field diary and structural econometrics. It resulted from long-term research carried out by a health economist and a kinship anthropologist. A simplified version was reproduced by the National Institute for Statistics and Economic Studies (INSEE) in the context of its *Handicap Santé* surveys⁸. The original survey has also been the subject of several sociology theses based on its ethnographic complements.

2.2 ArchEthno2020: a paradigm shift in IT

The ArchEthno2017 prototype constructed around a relational database enabled researchers to think in depth about the conditions for standardising and rationalising ethnographic field data. It also introduced the issues of systematising the documentation of survey conditions and the long-term archiving of digital or digitised material.

The database's structuring was based on a set of metadata that was hard to transpose to other surveys. Any changes would have required help from IT experts to modify the data model, adapt the old data to new formats (with a risk of corruption during this work) and review the data entry and retrieval interfaces. This all meant the model was not viable in the long term.

2.2.1 Extending the usages and simplification of the model

A paradigm shift was therefore required and a new tool (ArchEthno2020) was developed between 2017 and 2020 to overhaul the initial prototype. This involved a shift from a rigid SQL structure to approaches based on the semantic web where metadata is structured in modular and scalable concept dictionaries which also provide finely tuned confidentiality management for all the data and metadata. The structuring of metadata and the conditions for its reusability are

⁸ For a presentation of the 2008 *Handicap-Santé* surveys, please see <http5> (access 15.12.2023)

thus based on custom-built dictionaries rather than on pre-established fixed SQL schemas (Figure 1).

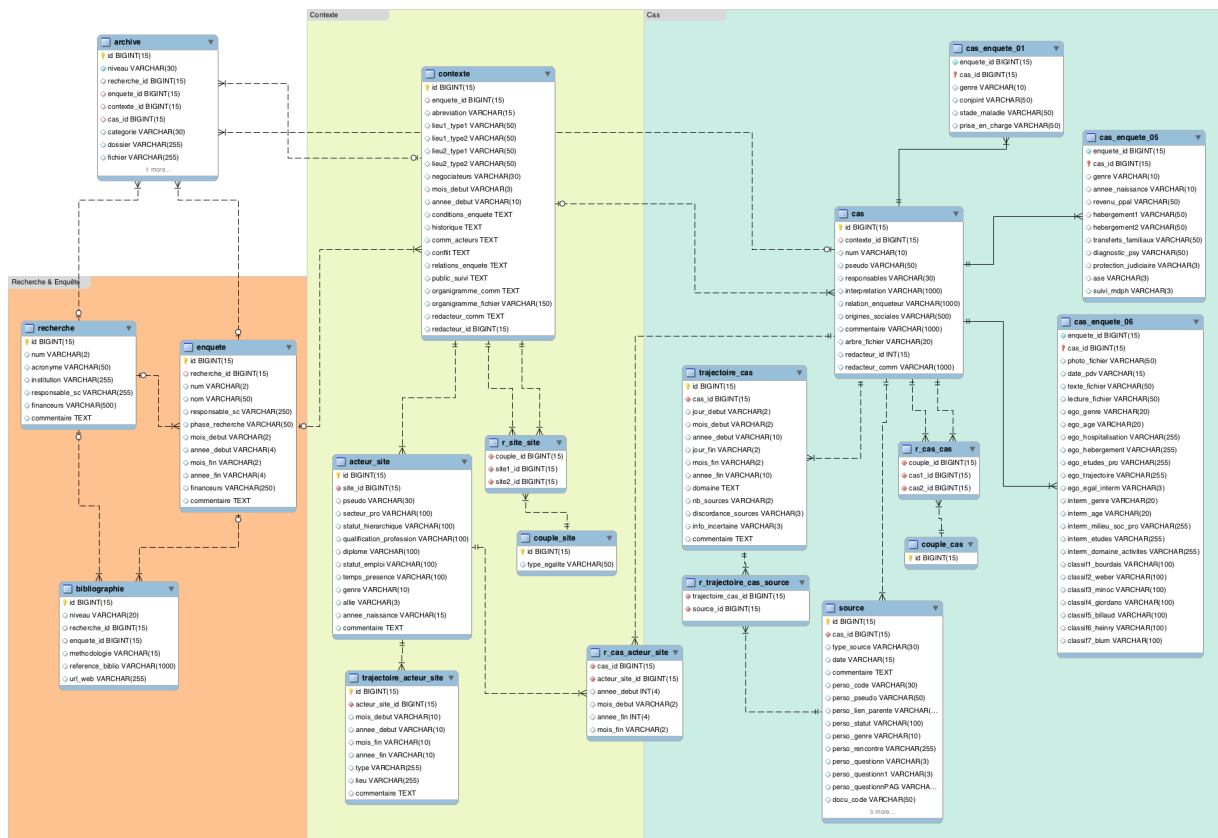


Figure 1a

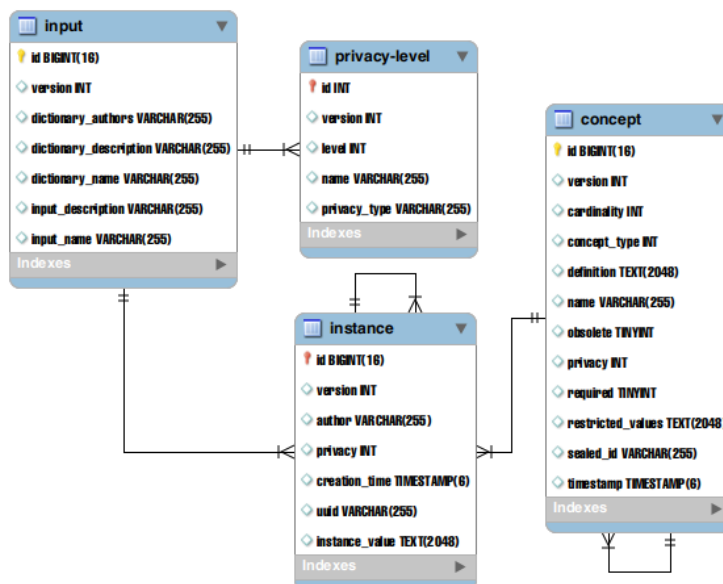


Figure 1b

Figure 1. ArchEthno2017's SQL relational data model (1a) compared with the ArchEthno2020 model based on dynamic concept vocabularies (1b). The simplification carried out between 2017 and 2020 means the structure of the database now only contains references to concepts whose description and hierarchisation are the subject of the dictionary.

These concept dictionaries are stored in ad hoc files and act as configuration files for a set of generic software components. They are only written once and definitively. Any extensions or modifications that may be required will be achieved by modifying the concept dictionary but not the software stack.

The generic software stack is made up of 4 components:

- a dictionary editor used to create or modify a dictionary;
- a data entry client - an application for entering and storing data on a personal computer;
- a centralised database which enables users to merge data entered via the data entry client;
- a consultation website that enables users to view the data in the centralised database using a web browser.

These four components make up a coherent whole in which:

- the software's fine level of granularity means confidentiality can be managed and enables researchers to think about their decisions as and when they need to rather than all at once in advance while following a protocol;
- a data construction framework enables compliance with scientific research presentation and documentation standards while respecting the inventiveness of researchers carrying out their analyses (in a Structuring module);
- a guide to confidentiality decisions helps raise researchers' awareness of the practical implications of their confidentiality decisions before these are submitted to an ethics committee in combination with the user guide which has been simplified as much as possible;
- a data entry guide linked to each dictionary available in free access provides guidance for researchers in structuring their material and for those who want to learn about the conditions data are produced under before requesting access to confidential materials (a search engine can be used to check the existence and interest of confidential material).

Even researchers lacking IT skills or the support of IT specialists can use ArchEthno2020 to develop the tool to adapt it to the requirements of their research.

The complete software solution was developed as part of an industrial support project funded by the Université Paris Sciences & Lettres (PSL). It has the following strengths and original features:

- it can manage the polymorphous nature of data and metadata. This requires ongoing development through concept dictionaries that enable data and their production contexts to be characterised in a unique and persistent way;
- it integrates the crucial need for confidentiality in the humanities and social sciences whereby certain data and metadata can only be made available to authorised users. The levels of visibility and authorisations are managed by dictionaries and can be modified when the data is entered. The tool can manage an infinite number of levels of confidentiality;
- it offers innovative solutions combining the ability to openly cite data with the need for confidentiality and respect for privacy.

The concept dictionary approach facilitates interdisciplinarity and enables the migration of data between disciplines. Also, from a technical standpoint, the data entry and retrieval interfaces are automatically generated from the dictionaries.

2.2.2 Using mutualisation as a guide to moving back and forth between material and hypotheses

The tool possesses flexibility and the capacity to be used as an instrument for a structuring approach during the dictionary creation phase. These factors open up prospects for wider use than the dissemination of research material. These could include helping researchers to structure field notes and formulate sociological reasoning or to pool materials, questions and hypotheses during a survey's collective phases. In both cases, the tools offer users the certainty of being able to return to the ethical issues involved in disseminating the material with the help of trusted third parties once the analysis is complete. Thus researchers have the peace of mind required for the sociological imagination.

We have proposed basic structuring in three modules that we consider suited to the ethnographic approach. The '*Matériaux/Materials*' module can be used to describe each brick of the field diary as the survey progresses. The '*Recherche/Research*' module recalls the approach's institutional framework. The '*Structuration/Structuring*' module can be used to construct an analysis of the material with reference to the research question as it is being stabilised and to formalise hypotheses that have been tested, invalidated or retained and possibly expanded or made clearer.

The prospect of sharing materials, in the framework of a teaching relationship or a research team, made us aware of the tool's richness. Before sharing, description is required. Mutualising materials helps build sharing communities, truly brings a team into existence and enables collective work to be brought to a close through deciding together the subsequent re-opening opportunities that will be possible. Sharing materials enables researchers to progress in structuring them. This approach is particularly well-suited to multi-disciplinary research which it provides with a practical framework focused on the most important points. What is the research question? How is work divided between team members according to their specialisation? How can any misunderstandings about each discipline's specific approach be avoided? Finally, when it comes to scientific publication, editing materials opens up new ways of sharing so who will be entitled to question decisions taken as regards structuring the research and analysing the materials? The ArchEthno tool offers the ethnographic practice of translation between several social worlds (those of survey respondents, scientific publications and possibly of decision-makers or journalists) a life-size test of what needs to be translated, simplified and finally highlighted or stressed.

2.2.3 Solitary and collective usages of ArchEthno

When a researcher is carrying out a survey alone, ArchEthno can help him/her to clean up field notes without undue 'fetishising' of their textual, narrative or rhetorical dimension. It also helps researchers as the survey progresses to integrate materials that are crucial to the analysis of how the survey was carried out like exchanges of emails, text messages or visual notes. These all make it possible to reconstruct the institutional and material conditions for survey appointments for example. Typed research diaries retain their temporal structure and their character as non-modifiable material. They are also used to help formulate a research question as the analysis progresses along with theoretical hypotheses based on constant 'back-and-forth' references to scientific concepts and analysis of the materials. The tool then helps ethnographers to put into practice the sometimes mysterious suggested rules on how to develop a "description in concepts" [Passeron, 1991] of an observed event, interaction or situation.

During the collective survey training courses that introduce students to the ethnographic approach, the tool can help make researchers more aware of the two stages of work in this research field. The first is actually carrying out a survey which involves the need for patience, disappointments and meeting people which requires a researcher to reformulate a question that 'speaks' to the respondents as much as to sociologists. The second is an analysis that uses the

materiality and content of the collected textual, auditory and visual notes to reflect on the nature of the investigative relationship and the interactions' dynamics before formulating hypotheses that will need to be tested by using sociological materials (description of infrastructures and institutions, even those that are initially invisible to the investigator). ArchEthno can thus become a tool for centralising the various techniques, for making a break with the interviewer's ethnocentrism and which helps sharpen an ethnographer's eye and ear. These include replaying recorded interviews, focusing on the spatio-temporal dimensions of the way interaction occurs and progresses and finally questioning the meaning attributed by others to their own behaviour.

2.3 ArchEthno 2020 and the FAIR principles

The ArchEthno2020 tool was constructed based on the principles of FAIR data (Findable, Accessible, Interoperable, Reusable). The data, metadata and the structures used to model the data are all FAIR-ised. The concept dictionaries and their construction procedures comply with the criteria defined in the FAIR Semantic recommendation [http11]. The tool's design was discussed at the 'Digital practices in History and Ethnography' session at the 11th plenary meeting of the Research Data Alliance in spring 2017 [http7]. Its innovative solutions for openly citing data while fully respecting confidentiality and privacy requirements were considered of interest for applying the GDPR to the HSS with no negative consequences for researchers' working conditions. The tool was developed in dialogue with the main French open data stakeholders – the INSEE, the Secure Data Access Centre (CASD), the National Institute for Demographic Studies (INED) and the *Bulletin de Méthodologie Statistique* – and also in relation to the social and cultural anthropology of contemporary worlds in Europe (Portugal, Spain, Netherlands, the European Anthropological Association) and the United States (Princeton, Harvard for medical anthropology, the American Anthropological Association). All the FAIR principles have been followed and implemented.

- The data is 'findable' because the tool offers a search portal for users to consult existing archives and their metadata. We also intend for this service to be indexed, offered and hosted by the Huma-Num TGIR (very large research infrastructure) as a generic tool for use by different disciplinary communities.
- The tool's confidentiality management function ensures a balance between accessibility and reusability. All data have a unique persistent identifier that is public and resolvable even if they are confidential. Once the PID has been resolved the information available depends on the user's level of accreditation. A PID is a public reference to data present in the information system but only authorised users will be able to view the private content of the public reference. In this way, we are complying with the European principle of "as open as possible, as closed as necessary"⁹ by guaranteeing the traceability and reproducibility of scientific work and making it public even if 'closed' data are involved.
- The technological choices our solution is based on mean that interoperability is dependent on the definition of concept dictionaries which set the epistemological limits for each piece of data (production context, potential field of validity, underlying assumptions, modelling) beyond which the piece of datum could become a source of error. As stated above, the process of definition of the dictionary complies with the FAIR Semantic recommendation [http7, http13].

⁹ Commission Recommendation (EU) 2018/790 dated April 25th 2018 on access to and the preservation of scientific information

A community of developers and users (data producers and site visitors) has been constructed to make this a long-term development effort. The classification into different levels of confidential information for each survey is submitted to an ethics committee at least 75% of whose members actually practice the ethnographic method. When a team of ethnographic producers linked to a laboratory receiving public funding – including students, PhD students, post-doctoral fellows, independent service providers, contract or tenured engineers or teachers – opens part of its metadata and data it needs to consult the ethics committee for its laboratory. Ethnographic producers who are not in a publically-funded laboratory will be informed that they can consult a legal specialist before opening their research data.

The ArchEthno method is not just the software solution involved and obviously still needs to find its place in the socio-technical system being constructed around the challenges linked to opening up research material. It will need to fit in with all the platforms dedicated to anthropology currently being set up on most continents. Indeed [Murillo, 2018] has proposed the idea of an Open Anthro Source that could link these together in ways that are still to be invented. The method will also have to become a link in the chain running from the 'primary' production of material and data by past and present researchers right up to the various 'post-production' stages whether these involve real or virtual dissemination or editorial development.

2.4 Publishing the dictionary and the data. Depositing and reusing material. Examples and future prospects

First and foremost ArchEthno is a tool for research. As such, its success depends on academic recognition of the work required to construct the database in a form similar to software copyright, film credits, printed credits or book *colophons*. This work includes software development, dictionary design (each dictionary put online is credited to its authors who are duly cited by the producers who deposit their materials and by database visitors) and the formatting and editing of material and data by the ethnographer and his/her editors (each collection put online is credited to its producers and duly cited as a scientific publication). The division of editorial work will become routine as people gradually publish the first editions of materials and data from ethnographic surveys. As with digital publishing, a great deal of attention will be paid to dating the various phases of the work and to the risks of free labour being exploited and added value being captured by the *front office* [Godechot, 2006].

The first databases made available on ArchEthno2020 will be as follows:

- the first Medips Alzheimer database constructed with ArchEthno2017 will be put online on ArchEthno2020 in 2024 along with its dictionary¹⁰. Its confidential material will be deposited to be accessible to authorized readers;
- the dictionary [Goudet, Vieujean and Weber 2022] used to design [Blum, Goudet, Weber 2022] will be put online in 2024 along with non-confidential material and data. Confidential data will be gradually deposited as they are prepared.
- Special mention needs to be made of databases designed to promote and disseminate students' work carried out during field surveys that combine the ethnographic method with the tradition of documentaries. The ethnographic and documentary traditions do not work with the same responses to crucial anonymity and confidentiality issues [Béliard and Eideliman, 2008, Dantou and Weber, 2015]. Also, until currently they do not share the same financial stakes linked to the trade in works [Weber, 2011] with a few exceptions. Furthermore, they do not have the same ways of involving respondents in the process of publishing the results of the survey.

¹⁰ For the ArchEthno dictionaries and databases, <http2> (access 15.12.2023)

Thanks to these first experiments, we will be able to propose criteria and processes to assess the value of research material: number of scientific publications based on this material; precisions on the conditions of its production, including analysis of the fieldwork relation and restitutions workshops; pedagogical issues; interest of the material for multidisciplinary teams. In order to avoid arbitrary selections we will participate to collegial discussions with dean's offices, archivists, scholars and students.

Other forms of usage are currently being examined.

- In art history and fields involving artistic creation, the ArchEthno tools could be used to document works in anthropology museums' (*Musée du Quai Branly - Jacques Chirac*, regional ethnology museums, MUCEM, BNF) collections of objects and photographs more effectively. This is particularly the case when the dissemination of certain information could be potentially dangerous for the respondents or their families. This is seen as more of an issue than the 'restitution' of such works and opening them to these people. From this point of view, contemporary ethnographers are the most aware of issues linked to the unwelcome circulation of confidential information [Tarrus, 2017].
- In history and archaeology, the ArchEthno tools could be used to construct multidisciplinary databases covering survey contexts, events, stratigraphic units and graphic and photographic documents including confidential geolocalised metadata via a dictionary constructed by several people. The tool can of course be used to store and revisit primary documentation from archaeological digs and also to publish heritage archives produced by an archaeology laboratory over several decades.

More broadly, ArchEthno could have three other forms of usage:

- methodological usage through the dissemination of the concept dictionary design method to professional ethnographers. The team is considering the possibility of adapting the tool to metadata in other languages and alphabets;
- multi-disciplinary usage through sharing materials equipped with metadata within teams of researchers in which ethnographers work alongside statisticians (sociologists, economists, geographers, demographers, epidemiologists) or environmental science researchers (ecologists, geologists, hydrologists, etc.);
- a technological usage by transferring the technical solution to other disciplines like archaeology or astrophysics.

Conclusion

The Archethno method responds to several types of difficulties encountered by researchers in different situations and in different disciplinary traditions and not just in the field of ethnography. Such difficulties are linked to epistemological, ethical or scientific integrity issues.

- From an epistemological standpoint, the method respects researchers' inventiveness as they can publish their own dictionaries when they decide to do so. It also makes it possible to combine 'objective' (documentation of a context) and 'subjective' material (where the investigative relationship is made explicitly clear) without opposing these or reducing their specific features.
- From an ethical standpoint, the method allows confidentiality decisions to be made explicitly clear as all materials submitted are initially confidential by default. Also confidentiality restrictions are lifted as late as possible when any confidentiality issues

have been fully analysed and the procedures for lifting confidentiality restrictions have been designed. Anonymisation techniques cannot be standardised as each research project has its own individual priority issues to manage.

- From a scientific integrity standpoint, the method authorises the citation of data whose content remains confidential and also provides a publication platform that *de facto* protects ethnographers' work from plagiarism.

Finally, the ArchEthno method makes it possible to clarify and deal with the tensions raised by the delimitation of 'sharing collectives' in contemporary societies that would do anything to defend what they see as their individual borders. We live in a society of cultural hyper-consumption that authorises or values the transparency of practices but also develops sophisticated secrecy techniques. Putting data sharing issues forward for analysis as issues linked to power [Butnaru, 2023] could serve as an effective *dispositif* or device for 'cooling' conflicts. In a world that is currently suffering from no longer knowing how to think about borders [Debray, 2010], the method could possess an inherent strategy for 'slowing down' cognition to maintain our capacity for analysis when we are faced with the endless deluge of unwanted information that characterises the contemporary public space [Fenoglio and Fleury, 2022].

Acknowledgements

We would like to thank all those who, at one time or another, have contributed to our reflections: Pauline Blum, Julien Clément, Jean-Robert Dantou, Julien Forest, Agnès Gramain, Jean-Marc Goudet, Pascal Mulet, Luis Felipe Rosado Murillo, Sabrina Pastorelli, Amandine Reist, Léa Saint Raymond, Karine Sitcharn et Olivia Vieujean.

References

- Avanza M. Le 'cas Alice Goffman' et les risques d'un backlash méthodologique. *Revue française de Science politique*. 2016;66(6):995-1001.
- Beaud S. and Weber F. *Guide de l'enquête de terrain*. La Découverte (Paris), 2010.
- Béliard A. and Eideliman J.-S. Au-delà de la déontologie. Anonymat et confidentialité dans le travail ethnographique. In Bensa A. and Fassin D. *Les politiques de l'enquête*. La Découverte (Paris), 2008:123-141.
- Blum P. *Un quotidien ébranlé : des jeunes patients de la psychiatrie et leur famille, dans la France contemporaine*. Thèse de sociologie, ENS-PSL (Paris), 2017.
- Blum P., Goudet J.-M. and Weber F. *Troubles psychiques en milieu scolaire. Que fait l'école ?* Éditions Rue d'Ulm (Paris), 2022.
- Butnaru D. Being loyal to fieldwork: on building the 'contract of silence'. *Journal of Data Mining and Digital Humanities*. Atelier Digit_Hum, 2023. <https://doi.org/10.46298/jdmdh.11394>
- Charmasson T. Archives scientifiques ou archives des sciences : des sources pour l'histoire. *La revue pour l'histoire du CNRS*, 2006;14. <https://doi.org/10.4000/histoire-cnrs.1790>
- Dantou J.-R. *Objets sous contrainte*. Agence Vu' (Paris), 2014. [http](http://).
- Dantou J.-R. et al. *Pour une alliance entre photographie et sciences sociales*. Auto-édition, 2020.
- Dantou J.-R. and Weber F. *The Walls Don't Speak*. Kehrer Verlag (Heidelberg), 2015.
- Debray R. *Éloge des frontières*. Gallimard (Paris), 2010.
- Desclaux A. Les lieux du 'véritable travail éthique' en anthropologie de la santé : terrain, comités, espaces de réflexion ? *ethnographiques.org*. 2008. <http://www.ethnographiques.org/2008/Desclaux.html>
- Duchesne S. and Noûs C. Apories de la mise en banque : retour d'expérience sur la réutilisation d'enquêtes qualitatives. *Tracés. Revue de Sciences humaines*. 2019;19:89-100.
- Fabry L. *L'anthropologie comme science rigoureuse : une lecture épistémologique de l'œuvre de Claude Lévi-Strauss*. Thèse de philosophie (Paris, ENS-PSL), 2021.
- Fenoglio A. and Fleury C. *Ce qui ne peut être volé. Charte du Verstoehlen*. Gallimard (Paris), 2022.
- Gagné N. Le savoir comme enjeu de pouvoir : l'ethnologue critiquée par les autochtones In Bensa A. and Fassin D. *Les politiques de l'enquête*. La Découverte (Paris), 2008:277-298.
- Geertz C. *Works and Live. The Anthropologist as Author*. Stanford University Press (Stanford), 1986.
- Genèses. Le procès des données. 2022;129(4).
- Godechot, O. Hold-up en finance. Les conditions de possibilité des bonus élevés dans l'industrie financière. *Revue française*

- de sociologie, 2006;47(2):341-371.
- Goudet J.-M., Vieujean O. and Weber F. Dictionnaire Troubles psychiques 2022. http2.
- Grootaers J. -L. De l'exploitation des archives de terrain. Une textualisation en chaîne. *Gradhiva : revue d'histoire et d'archives de l'anthropologie*. 2001;30-31:73-80.
- http1 <https://agencevu.com/serie/objets-sous-contrainte-2014/>
- http2 <https://www.archethno.org/>
- http3 <https://bequali.fr/fr/>
- http4 <http://www.calames.abes.fr/pub/#details?id=FileId-2182>
- http5 <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/02-les-enquetes-handicap-sante>
- http6 <https://hraf.yale.edu/>
- http7 <https://rd-alliance.org/ig-digital-practices-history-and-ethnography-rda-11th-plenary-meeting>
- http8 <https://sonal.hypotheses.org/>
- http9 <http://spencerandgillen.net/>
- http10 <https://tropy.org/>
- http11 <https://distam.hypotheses.org/236>
- http12 <https://ukdataservice.ac.uk/help/other-data-providers/other-data-providers-qualitative-data/>
- http13 <https://doi.org/10.5281/zenodo.4314320>
- Joël M.-E. and Gramain A. *L'aidant familial aux patients atteints de démence de type Alzheimer : agent économique producteur de services. Rapport final*. Fondation Mederic-Alzheimer, 2005.
- Laurens S. and Neyrat F. *Enquêter : de quel droit ? Menaces sur l'enquête en sciences sociales*. Éditions du Croquant (Vulaines-sur-Seine), 2010.
- Laferté G. Retours d'expériences. Plaidoyer pour l'ethnographie collective. *ethnographiques.org*. 2016;32.
<https://www.ethnographiques.org/2016/Laferte>
- Lécho Hirt L., Nova N., Kilchör F. and Fasel S. Design et ethnographie. *Techniques & Culture*. 2015;(64):64-77.
- Le Gonidec M.-B. Les fonds sonores de l'ancien musée national des Arts et Traditions populaires : collections ou archives ? *Bulletin de l'Association française des archives orales sonores et audiovisuelles*. 2020;46(1):16-28.
- Malinowski B. *Argonauts of the Western Pacific, An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. Routledge and Kegan (London), 1922 [*Argonautes du Pacifique occidental*. Gallimard (Paris), 1989].
- Malinowski B. *A Diary in the Strict Sense of the Term*. Routledge and Kegan (London), 1967. [*Journal d'ethnologue*. Le Seuil (Paris), 1985].
- Medips. Comprendre la mobilisation familiale autour des personnes âgées. *Connaissance de l'emploi*. 2006;33.
- Müller B. and Weber F. Réseaux de correspondants et missions folkloriques. Le travail d'enquête, en France, vers 1930. *Gradhiva : revue d'histoire et d'archives de l'anthropologie*. 2003(33):43-56.
- Murillo Rosado L.-F. Multimodal Anthropologies. What Does 'Open Data' Mean for Ethnographic Research? *American Anthropologist*. 2018;120(3):577-600.
- Olivier de Sardan J.-P. *La rigueur du qualitatif. Les contraintes empiriques de l'interprétation socio-anthropologique*. Academia-Bruylant (Louvain-La-Neuve), 2008.
- Passeron J.-C. *Le Raisonnement sociologique. L'Espace non-popperien du raisonnement naturel*. Nathan (Paris), 1991.
- Pina-Cabral J. Ethnography as tradition in Africa. *Etnográfica*. 2011;15(2):379-394.
- Segalen M. Le Musée national des arts et traditions populaires, 1936-2005. Récit d'un brillant fiasco. Deuxième partie : Chronique d'une mort annoncée (1980-2005). *Bérose - Encyclopédie internationale des histoires de l'anthropologie*. 2019.
<https://www.berose.fr/article1684.html>
- Shukan I. *Génération Maïdan. Aux origines de la résistance ukrainienne*. Éditions de l'Aube (La Tour-d'Aigues), 2022.
- Stocking G. W. *Observers observed: essays on ethnographic fieldwork*. University of Wisconsin Press (Madison), 1983.
- Tarrius A. Fluidité de l'identité du chercheur en situation d'immersion le long des territoires de circulations migratoires. Extraits de carnets d'enquête : Prizren, Kosovo, 2004. La Palme, Aude, 2014. *Migrations Société*. 2017;167(1):41-52.
- Vieujean O. Une mère et son fils dans un quartier prioritaire. In Blum P., Goudet J.-M. and Weber F. *Troubles psychiques en milieu scolaire. Que fait l'école ?* Editions Rue d'Ulm (Paris), 2022.
- Wasamba P. Going beyond data collection in ethnography: Options for bridging the gap between researchers and archivists. *International Journal of African Renaissance Studies - Multi- Inter- and Transdisciplinarity*. 2013;7(2):4-17.
- Weber F. L'enquête, la recherche et l'intime ou : pourquoi censurer son journal de terrain ? *Espaces Temps*, 1991;47-48:71-81.
- Weber F. Publier des cas ethnographiques : analyse sociologique, réputation et image de soi des enquêtés. *Genèses*. 2008;70(1):140-150.
- Weber F. La déontologie ethnographique à l'épreuve du documentaire. *Revue de synthèse*. 2011;132-6(3):325-349.
- Wolikow S. L'historien face aux archives du Komintern. In Sébastien Laurent (dir.), *Archives 'secrètes', secrets d'archives ? Historiens et archivistes face aux archives sensibles*. CNRS Editions (Paris), 2003:177-188.
- Yanow D. Ethnography on trial: introduction to the dialogue. *Politics, Groups, and Identities*. 2021;9(4):826-834.