Schweizerische Ethnologische Gesellschaft
Société Suisse d'Ethnologie
Associazione Svizzera di Antropologia
Swiss Anthropological Association

Open Science and Data Management in Anthropological Research. Position Paper of the Swiss Anthropological Association (SAA)[*]

Adopted at the SAA Annual Meeting,

held on-line on November 4, 2021

---

[*] Numerous colleagues have contributed directly or indirectly to this Position Paper. I would like to thank the members of the first DMP working group launched by the SAA's Scientific Commission: Anja Sieber Egger, Judith Hangartner, Peter Larsen, Barbara Waldis and Olaf Zenker. Sabine Strasser took over this work with me in 2018, assisted by Martine Stoffel and Irina Wenk, and by Dr. Luc Gauthier, Research Counsel at the University of Neuchâtel; my warm thanks for their crucial input. The Board and interested SAA members have provided invaluable criticism and suggestions, specifically Laurent Amiotte-Suchet, Sylvain Besançon, David Bozzini, Filipe Calvao, Julia Eckert, Judith Hangartner (again!), Lena Kaufmann, Olivia Killias, Anne Lavanchy, Monika Salzbrunn, Valerio Simoni, Eva Soom Ammann, Wiebke Wiesigel and members of the SAA's Ethical and Deontological Think Tank. Finally, my appreciation goes to the data management specialists at the Swiss national science foundation (SNSF), swissuniversities, the Swiss Academy of the Humanities and Social Sciences (SAGW-ASSH), and the Swiss Center of Excellence in the Social Sciences (FORS), specifically Pablo Diaz and Alexandra Stam, all of whom answered our questions and helped us through the learning process with great patience.

Introduction

Context

Over the past decade, anthropologists, like other scientists, have been confronted with formal demands to conduct their research based on principles of "open science"[1]. Broadly speaking, the movement for open science aims at maximizing public access to research results, and increasing transparency, collaboration and sharing in the research process. These principles reflect different objectives: to enhance replicability and validity of data and results; to avoid wasteful duplication of projects and infrastructure; to promote democratic access to and influence over research; and to encourage the dissemination of knowledge to scholars, institutions and countries that cannot afford access to commercially based scientific publication and data-sharing platforms.

The rapidity with which this new paradigm has been adopted by governments, universities and funding agencies alike is a phenomenon worth anthropological investigation in its own right. Clearly, the push for open science is closely associated with social transformations induced by digitalization, so-called "Big Data" and the omnipresence of the Internet. While the open science movement is generally portrayed as a form of resistance to the privatization of research data, it is more properly understood as multidimensional, promoted by a wide variety of actors for a variety of reasons[2]. Swiss regulatory and funding agencies have been quick to subscribe to its principles and, over the past fifteen years, have worked to integrate them into the institutional reality that Swiss anthropologists must understand and work within.

The open science paradigm in Switzerland is broadly based on the 2003 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, jointly signed by the SNSF, swissuniversities and the Swiss Academies in January of 2006. This declaration contains the main arguments for increased openness of scientific knowledge and provides a legal basis for open access and open data instruments in the age of the Internet. In conjunction with this move towards openness, however, came a countervailing concern with privacy, or closure. The key text in this regard is the European Union's General Data Protection Regulation (GDPR) which came into effect in May of 2018, and which strictly regulates the accessibility, circulation and use of personal data, particularly in commercial settings.

To translate these broad principles into national policy, in late 2018 the Swiss Federal Council adopted its "Digital Switzerland" strategy[3]. This strategy addresses the challenges caused by the digital turn and proposes a "coherent legal basis regarding the legal rights relating to data, access to data and data handling". The document also locates national initiatives to harmonize data accessibility in the context of the ongoing revision of the Federal Act on Data Protection (FADP). Meanwhile, the three institutions party to the Berlin Declaration have worked to transposed these principles into their internal regulations.

As Switzerland's national funding institution for science, the SNSF has the legal obligation to set forth a framework and prerequisites for funding in conformity with the Berlin Declaration. These principles are defined in Article 47 of its Funding Regulation[4]. Article 47 requires research institutions and researchers to provide both greater public access to research results (through its "Open Access to Publications" policy) and greater access to the data on which these results are based (through its "Open Research Data" policy). The latter further requires researchers to submit a data management plan (DMP) in which they spell out how and to what extent they will make data available to the public,

---

[1] The tendency to capitalize this term, often referred to simply as "Open Science", strikes us as reifying and just plain odd. We use the terms "movement for open science" or "open science paradigm" instead, without capital letters.

[2] See the useful introduction in "Open Science et recherches en sciences sociales: entre injunction institutionnelle et spécificités disciplinaires. Eléments de cadrage et de réflexion", a report compiled by Laurent Amiotte-Suchet (2019) for the Haute Ecole de Santé – Vaud (contact the author). See also an interesting discussion in the Wikipedia entry "Open Science".

[3] Relevant documents are available at : https://www.bakom.admin.ch/bakom/en/homepage/digital-switzerland-and-internet/strategie-digitale-schweiz/digitale-schweiz.html.

[4] http://www.snf.ch/en/funding/documents-downloads/Pages/regulations-funding-regulations.aspx.

in compliance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) for data management. In addition, the SNSF manages two data service centres – SWISSUbase (formerly known as FORSbase)[5] and DaSCH[6] – designed for data storage, sharing and archiving in conformity with the FAIR principles. Their storage, maintenance and archiving services are available (cost-free!) to Swiss researchers in the social sciences and humanities, respectively.

In 2019, the Swiss Academy of Humanities and Social Sciences (SAGW-ASSH), responsible for a substantial part of SAA's funding, adopted its own "Open Science Policy" in conformity with the Swiss legal framework. The SAGW-ASSH's open science policy has had a direct effect on the SAA in that it led to our professional journal, *Tsantsa,* transitioning to full open access in 2019.

Finally, in October of 2019, the State Secretariat for Education, Research and Innovation (SeFRI) mandated swissuniversities, the Swiss consortium of institutions of higher learning, to prepare a national strategy, or "Open Science Action Plan", for the period 2021-2024. This action plan, which will focus on coordinating investments in data management infrastructure across universities and cantons, is currently under development, and is scheduled for validation, after consultation with relevant stakeholders, in November of 2021.

It is important to note that all three institutions make a clear distinction between "open" and "fair" data. The key difference is that "open data" is:

> "available to everyone to access, use, and share, without licenses, copyright or patents. It is expected that open data at most should be subject to attribution/share-alike licenses. [FAIR data, by contrast,] uses the term 'accessible' to mean accessible by appropriate people, at an appropriate time, in an appropriate way. This means that data can be FAIR when it is private, when it is accessible by a defined group of people, or when it is accessible by everyone (open data). The accessibility of FAIR data depends on the purpose for which they were collected, where they currently are in their lifecycle, and their projected end-use."[7]

This distinction takes into account, among other elements, the obligations imposed by countervailing legal and ethical principles of personal privacy and data protection. Thus, while the SNSF clearly requires the researchers it funds to comply with the FAIR principles, it does not require "open data" in the strict sense. That being said, the continued use of the term "open" rather than "fair" in key documents, as well as the awkward application of the FAIR principles to the realities of qualitative data management, has led to questions, confusion and concern within the anthropological community. The purpose of this Position Paper is to address these concerns.

Origins and objectives of this Paper

This Paper originated in an informal working group convened under the auspices of the SAA's Scientific Commission in 2017. In 2018, catalyzed by the SNSF's new data management requirement, but also by lively debates within the discipline in Europe[8], the SAA governing board took up the issue. At its September 2018 meeting, the Board mandated two independent researchers to produce an initial analysis of the state of the field, which was discussed at an extended Board meeting in June of 2019[9].

---

[5] https://forsbase.unil.ch/.

[6] https://dasch.swiss/.

[7] See https://ask-open-science.org/1116/what-the-difference-between-fair-data-and-open-data-there-any.

[8] See, notably, the "Leiden Statement" in Pels P. *et al.* (2018), "Data Management in Anthropology: the Next Phase in Ethics Governance?", *Social Anthropology/ Anthropologie sociale* 26/3: 391-413; Corsin Jiménez, A. (2018). "Data Governance Framework for Ethnography 1.0". Madrid: CSIC; the European Association of Social Anthropologists (EASA)'s "Statement on Data Governance in Ethnographic Projects" (n.d.); and the Société International d'Ethnologie et de Folklore (SIEF)'s "Statement on Data Management in Ethnology and Folklore" (2019). These papers touch on issues that are not discussed in the present Position Paper, in particular open and important questions surrounding intellectual property rights.

[9] Ellen Hertz and Sabine Strasser hereby warmly thank Martine Stoffel and Irina Wenk for their very helpful contributions.

Through this discussion, Sabine Strasser (then vice-president) and Ellen Hertz (president) identified a series of open questions, which they decided to submit to the three institutions mentioned above: swissuniversities, the SNSF, and the SAGW-ASSH. In parallel, in April of 2019, Prof. David Bozzini (UniFR) convened a conference/workshop with specialist Kim Fortun who has designed and runs the Platform for Experimental and Collaborative Ethnography (PECE), an open-source digital platform for anthropological and historical research based at the University of California (Irvine).

The results of these exchanges were transmitted to SAA members at the SAA General Assembly held in Geneva in November of 2019, in a document entitled "Data Management Framework for Anthropological Research. Discussion Paper of the Swiss Anthropological Association". Further research and discussions were carried out by Hertz and an *ad hoc* group of interested SAA members (informally named the "Data Management Working Group" or "DMWG") on the following occasions: at the SAA's fall Board meeting in September 2020; at the SAA General Assembly in November of 2020; at an *ad hoc* meeting of the DMWG in April of 2021; at a special meeting convened with experts at FORS in May of 2021; and at the SAA's fall Board meeting in September of 2021. At this last meeting, it was decided to produce two separate papers: the present Position Paper, addressed to our membership, and a second, much shorter "Executive Summary", addressed to relevant funding and regulatory agencies.

The present Paper, along with the Executive Summary, were presented to the SAA General Assembly at its annual meeting on November 4, 2021 and accepted unanimously (less abstentions). These two papers aim to help the Swiss anthropological research community navigate its way through this complex new field without too much anxiety, and perhaps even find it interesting. They provide certain concrete suggestions as to how our members can meet their legal and regulatory obligations under this new regime, and how Swiss funding and regulatory agencies can better respond to the needs of this community.

Key questions for the anthropological research community

The SAA sees the push towards open science as a welcome occasion to rethink and clarify disciplinary practices on a wide variety of issues, ranging from collaborative research to informed consent, from data protection to procedures for sharing our results with the people with whom we work. None of these issues is entirely new, as anthropologists have for decades contributed actively to fundamental thinking on the politics of research and the restitution of research results[10]. Nonetheless, the institutionalization of data management policies and requirements, as well as increased concerns over data security, have served as catalysts for experimentation with and critical reflection on how anthropologists produce knowledge in a digital age, and how they can share this knowledge with wider publics and preserve it for future generations.

Despite this general openness to the principles of open science and data management, anthropologists are also acutely aware that new regulatory and legal requirements are often based on research paradigms that originate in other disciplines (the natural sciences in particular) and that can be inappropriate or counterproductive when applied to ethnographic research. While our discussions with relevant experts and authorities in this field have been very reassuring to date, in particular with the highly informed and cooperative experts at FORS, as social scientists we know all too well how regulatory drift can transforms norms into obligations, creating subtle hierarchies amongst knowledge practices, and leading to standardization, bureaucratization and the creation of mutual distrust between researchers and the institutions that regulate them. It is with these concerns in mind that the SAA has decided to provide guidance to our members and to make our needs and perspectives available to policymakers and the public.

---

[10] For foundational texts, see Hymes, D., ed. (1972). *Reinventing Anthropology.* New York: Random House and Brettell, C. B., ed. (1993). *When They Read What We Write: The Politics of Ethnography.* Westport, CT: Bergin & Garvey.

There are two main thrusts to new regulatory activity in the area of data management: the push for increasingly open data on the one hand, and legal requirements for increased attention to data security and privacy rights of research subjects on the other. All experts we have read and talked with acknowledge that these two policy goals can enter into tension if not properly understood. The most general formula for resolving these tensions is summed up in the phrase that was quoted to us by swissuniversities: "as open as possible, as closed as necessary".

This policy signifies that anthropologists should make their data "fair and open" (through storage, sharing or archiving on data management platforms) ***only to the extent that*** *these procedures conform with legal requirements concerning short- and long-term data security, respect for personal privacy and protection of the interests of the people with whom they work.* SAA members and anthropologists practicing in Switzerland can be reassured by this general conclusion: regulatory and funding institutions will not oblige researchers to publicize their data if there is any risk that these data can cause harm. To the contrary, careful data management protocols can and should work to reinforce ethical standards and increase deontological vigilance in anthropological practice.

That being said, the legal, technical and regulatory requirements for adequate data management are relatively new and complex. The following discussion is organized around the two distinct goals of "closure" and "openness", in that order.

"As closed as necessary": implications of the "do no harm" principle in anthropology

As SAA members well know, the cornerstone of anthropological research ethics is and has always been long-term concern for the people with whom we work[11]. There is broad consensus within the anthropological community that our discipline should be a force for good, our primary "good" being the production of knowledge about the life ways of all human groups in society and about their interactions with each other and with their natural, institutional and built environments. It flows from this principle that we must avoid harm to the fullest extent possible: to our research subjects, to other researchers and to the institutions that fund our work.

The American Anthropological Association distinguishes broad categories of harm[12], and insists that anthropologists must consider all forms of potential, indirect or long-term risk that their presence, inquiries or publications might present. The commissioned paper for the Ethics Unit B6, DG Research and Innovation of the European Commission (Iphofen 2015) breaks this down further into "forms of distress" and "material harm" (pp. 23-25). Serious risks include involuntarily alerting authorities to practices, networks or organizations classified as illegal or creating conflict within the networks anthropologists work with, for example through inadvertent sharing of confidential information. In addition, certain categories of research (for example medical anthropology) or certain categories of research subjects (for example with minors or people deemed legally incompetent) may be subject to targeted regulation in relation to specific risks[13]. However, all professional associations agree that the forms of harm attending research can be surprisingly varied and cannot be entirely foreseen before the research begins. Anthropological research ethics, like anthropological knowledge production itself, are processual, not procedural (see SAA EDTT 2011, 2018), and require long-term engagement and on-going, contextualized judgement.

With the digitalization revolution, certain categories of risk have become particularly salient, as electronic data are notoriously accessible and manipulable. These include the risk that anthropological data fall into the wrong hands or are inadvertently revealed to the wrong audiences. It is important for our members to understand the extent to which this area is regulated today. This implies that

---

[11] Key texts of reference include the two long thought pieces produced by the SAA Ethical Deontological Think Tank (EDTT), the American Anthropological Association's Principles of Professional Responsibility, and the European Commission's document entitled "Research Ethics in Ethnography/Anthropology".

[12] See https://ethics.americananthro.org/ethics-statement-1-do-no-harm/.

[13] Medical anthropologists and researchers working with vulnerable populations may have specific regulatory frameworks to comply with, such as the Federal Act on Research involving Human Beings.

anthropologists may be subject to legal liability above and beyond the moral and deontological responsibilities they have under current codes of professional ethics. Both the GDPR (European Union) and, in Switzerland, the FADP require that researchers make thorough, good faith efforts to attend to issues of data security and protection of privacy. In this section, we will examine this issue by discussing four interrelated questions, all of which require increased attention from the anthropological community:

1. How can researchers best secure their data against third-party misuse?
2. How can researchers best "anonymize" their data?
3. What are the requirements for and implications of obtaining the full and informed consent of research subjects?
4. How can researchers be protected from legal obligations to reveal information about their research subjects?

(1) Data security and legal liability: how better to secure our data

Anthropologists have traditionally celebrated the virtues of the pen and paper, and recorded the greater part of their interactions and observations in field notebooks. Today, while the notebook remains a useful tool for recording data *in situ*, our notes are generally re-transcribed in digital form and stored on our computers, memory sticks or the cloud. We also record interviews in digitalized formats and take digitalized pictures and videos. Furthermore, much of this data is exchanged over the Internet, often in non-crypted format or via non-secured platforms. These practices raise real problems of data protection for the people we work with and create risks of legal liability for researchers themselves.

The new data protection frameworks put in place by the European Union and the Swiss Confederation apply first and foremost to businesses and seek to regulate the commercialization of data. However, they also apply to data collected by governments and local authorities. Because our research is publicly funded, our data falls within the purview of these new rules and regulations, although they may receive special consideration as the knowledge they produce is conceived of as a public good.

The SNSF's new DMP requirement reflects the fact that, strictly speaking, researchers are legally responsible for safeguarding their data to the best of their abilities. While universities own the data produced by their researchers, they do not own the intellectual property rights to exploit it, nor are they legally responsible for the technical, methodological and deontological choices researchers make. While researchers may not be held responsible for all security breaches that could occur, nor for all of their consequences, they must make good faith efforts to think through the potential risks of their research protocols and to mitigate these risks as thoroughly as possible. Filling out a Data Management Plan (DMP) is an excellent way to go through this exercise at the very initial stages of research.

The apportionment of rights and responsibilities in this area is complex, and not something anthropologists have generally been trained to deal with. Questions that researchers should reflect on when beginning research include:

- *What are the most secure servers for data storage?*
  Generally, the organizations we have consulted recommend storage on university servers (with accompanying data encryption software) during the research period, and discourage even short-term storage on commercial cloud-based servers, personal computers, memory sticks or even in paper format under lock and key. Universities may be fully or partially responsible for security breaches if they have not fully informed their researchers of necessary procedures or if they have not properly secured their servers, but most universities have now put in place the necessary procedures and infrastructures. It is up to the anthropological community to learn how to use them, and to insist on infrastructural changes at their host institutions if these facilities do not provide adequate protection to our research subjects.

- *How can university servers be used in a secure manner?*
  Even if university servers are secure, storage on university servers does not guarantee the security of data as it is being transferred to these servers unless strict VPN protocols are followed. And of course, data security breaches are easily committed when data is temporarily stored on personal computers or recording devices. Furthermore, access to university servers may be time-consuming or even simply impossible from field sites without stable Internet connections. Thus, while university servers are the best solution for some purposes, they do not entirely solve the problem. Researchers should become familiar with and regularly use encryption software on their personal computers.

- *How can researchers with temporary contracts with their universities guarantee access to the data they produce and that of their research groups?*
  Storage on university servers also raises questions of access to data for researchers who may lose their formal affiliation to their research institutions at the end of their contracts. This has particularly important implications for young researchers with temporary positions. We recommend that researchers take up this question with their host institutions. This may also be a reason for keeping a separate, encrypted version of all research material on a server that is accessible for researchers no longer formally affiliated with a research institution.

In summary, the experts we talked with strongly recommend that anthropological researchers use encryption software to secure their data in all of the locations where it may be stored. This represents a major shift in disciplinary practice for many anthropologists, and one that requires immediate attention. *Thus, our first recommendation for the anthropological research community is to participate in specialized trainings on data security and protection whenever possible. (Paired with this is a recommendation for our funding and host institutions that these training modules be freely available – through the SNSF, through FORS and through university research offices and graduate training programs such as the transversal skills modules offered by CUSO* (Conférence universitaire de la Suisse occidentale). *We also encourage the experts at FORS to produce a research paper specifically dedicated to encryption)*[14].

As a follow up to this recommendation, researchers, and in particular independent researchers and researchers in the field, should signal to us any difficulties they are having securing their data so that we can correctly assess the full range of issues raised by data security.

(2) The problem of "anonymization"

In current anthropological practice, one of the principal tools used to protect research subjects is "anonymization", for example by assigning code names or numbers to interviewees or interviews and keeping the code separate from the data itself. However, as anthropologists have long pointed out, full anonymization is neither possible nor desirable, as anthropological data is highly context-bound, and therefore can often be traced to specific speakers. Furthermore, the meaning of our data depends on the particular characteristics of the person speaking or the situations of interaction, and these cannot therefore be eliminated to "clean up" the data.

The contextual, situated nature of anthropological data is well recognized by specialists active in the area of social science data management in Switzerland. These experts readily admit that total anonymization should not be the goal and encourage anthropologists rather to think in broad terms about what good-faith efforts they can make to disguise identities (pseudonyms, altering non-relevant details, creating composites, etc.). Because total anonymization would dilute or distort the meaning of empirical data, most anthropologists will not and should not pin their hopes on this means to ensure the protection of the people with whom we work.

---

[14] To begin informing themselves on matters of encryption, surveillance and privacy, members may consult these resources, among others: the Association of Internet Researchers' "Internet Research: Ethical Guidelines 3.0" (2019); "Ethics and data protection" (Iphofen 2015); The Electronic Frontier Foundation (EFF); the EFF "Surveillance Self-Defense" kit; and EFF guide to crossing U.S. borders with electronic devices. For an example of open-source disk encryption software, see VeraCrypt.

As guidance for our members, we recommend the highly instructive study entitled "Data anonymization: legal, ethical, and strategic considerations", drafted by Alexandra Stam at FORS[15]. It recommends that anonymization be considered as only one pillar of an overall strategy for data protection, along with other measures such as informed consent, research design and data access control. Similar suggestions and guidance can be found at university library websites on data management, which we urge our members to consult systematically.

(3) The status of informed consent under Swiss law

As mentioned above, the obtention of "informed consent" is not only a central tenet of anthropological ethics, it is also part of the process of managing data security and legal liability. If the people we work with understand and agree to participate in anthropological research, it is argued, they can also control the kinds of data they wish to convey to researchers, thereby protecting themselves. As with anonymization, however, these formalist approaches to "doing no harm" do not do justice to the complex questions underlying anthropological ethics.

Currently, Swiss law does not require researchers to obtain explicit, written informed consent. Oral and implicit forms of informed consent also satisfy legal requirements, as long as they are obtained in good faith. However, current revisions of the Swiss Data Protection Act are moving towards a requirement of *explicit* consent, though not necessarily in written form. The EU GDRP tends to be less lenient; it currently requires that researchers show their research subjects explicit informed consent forms at the beginning of the research process. However, these forms need not be signed until the end of the process. This complex legal situation is well summed up in a recent paper published by FORS entitled "The informed consent as legal and ethical basis of research data production".

As a professional association, we will remain attentive to any up-coming changes in this area. Anthropological researchers have long argued that consent must be understood not as a formalized moment in time, but as part of a long-term relationship of mutual trust established between researchers and the people with whom they work (see EDTT 2018 "Searching for Ethics"). Anthropology is generally a "hypothesis-generating" and not a "hypothesis-testing" science, based on inductive reasoning and a continual back-and-forth between data collection and problem formulation. Anthropologists may thus find it difficult to present research subjects with a complete description of their research design at the beginning of the research process. Indeed, in some cases formalized written consent obtained at the beginning of this process may in fact violate ethical and deontological norms, as genuine guarantees of consent can only be obtained over time, and with a full understanding of how research questions have evolved, how results will be disseminated and in which contexts. Anthropological guidelines tend to insist, therefore, on the notion of *dynamic and processual consent*. The EDTT has complied a series of detailed examples of ethical problem-solving by Swiss researchers. In the future, it would be useful to synthesize and systematize these case studies, so as to help Swiss regulatory agencies understand the nature of the ethical issues at stake[16].

(4) Relations with law-enforcement authorities

Related to questions of data protection is the further (and largely unaddressed) question of whether and how anthropologists can prevent their data from being requisitioned by law-enforcement officials. Anthropologists, like other social scientists, have a long tradition of studying illegal or "informal" behavior. The information we collect can interest intelligence, judicial, security and police forces in various national and international contexts. Border crossing is a particularly delicate moment in this regard. Our members should understand that, unlike journalists, anthropologists do not benefit from a clear-cut legal framework that protects their data against law enforcement requests for access.

---

[15] Many of the short papers produced by FORS can be helpful to anthropologists. See generally FORS Guides to Data Management, and "How to draft a DMP from the perspective of the social sciences, using the SNSF template".

[16] The FORS experts in qualitative research methods would specifically welcome case studies that illustrate how questions, and therefore also consent procedures, can evolve over time in anthropological research.

The German Association for Social and Cultural Anthropology (DGSKA), in cooperation with other disciplines, is currently initiating a discussion on potential legislative proposals for a right to refuse to provide evidence for social scientists. This is clearly an area where the SAA must become active in the future. However, according to our discussions with experts at FORS, the current international trend is not in our favor. Quite to the contrary, it seems that the question of researcher- subject privilege in the giving of evidence has been discussed and adjudicated in different national contexts, and no exceptions for social science research has thus far been granted.

We encourage the experts at FORS to draft a factsheet on these questions, which no doubt interest other disciplines than anthropology. In the meantime, our members should know that the best protection for our research subjects in largely technical in nature: solid and stable encryption technologies, data storage in secured locations, and good faith, proactive efforts at pseudonymization.

"As open as possible": implications of the FAIR principles for anthropological data

The SAA is convinced that, *under certain conditions,* anthropologists have every interest in opening up and systematizing access to our research data, not simply for scientific purposes (restudy, collective long-term fieldwork, validity-testing, historical documentation) but also because anthropological analyses nourish social debate and provide insights that are critical to democratic societies' ability to govern themselves. Furthermore, we could hope that the more the public comes to understand how anthropologists come to their conclusions, the more legitimate anthropological contributions to these public debates will be. However, while anthropologists have long encouraged collaborative research, co-authorship and systematic restitution of research results to the populations with whom we work, the discipline does not have an established practice of making our data available to researchers or populations outside specific research networks, nor have anthropologists developed the technological tools or platforms necessary to pool their efforts in this direction. The new regulatory framework provides an opportunity for the discipline to innovate in this area.

That being said, the conditions under which increased public access to anthropological data could become a regular aspect of established research practice are far from realized. The fine work performed by the EDTT can serve as a reference on these issues. Two detailed papers explore the characteristics of much anthropological research that make data sharing and archiving difficult: the situated, informal and highly personal production of data through intersubjective interactions; the embodied and experiential nature of anthropological knowledge, and perhaps most importantly, the fact that anthropologists frequently work with populations in situations of political or social vulnerability, or in contexts where the state is not a reliably ally for defending individual or group rights.

As mentioned above, our exchanges with experts involved in data management policy at the national level – particularly at the FORS Centre of Excellence in the Social Sciences – have convinced us that these specificities of qualitative, inductive research methodologies are well understood by Swiss regulatory bodies. Rather than review these questions here, we direct our readers to the EDTT papers, and concentrate on the more technical questions raised by data management requirements and the application of the FAIR principles. Below, we discuss four issues that require increased attention from the anthropological community and that can form the basis for requests that the SAA addresses to the institutions governing open science policy in Switzerland. These are:

1. What counts as anthropological data and which kinds of data should be made available to the public under which conditions?
2. What are the FAIR principles and how do they apply to anthropological data?
3. Who is responsible for data maintenance over time?
4. What resources (training, data management assistance) can be made available to researchers to ease the transition to an appropriate "open data" paradigm for anthropology?

(1) What counts as anthropological data and what should be made "open"?

In the initial discussions over data management and DMPs organized by the SAA, we spent a considerable amount of time discussing what counts as "data" within this framework. In particular, many felt that, for different reasons, fieldnotes and personal journals should not be classified as "data" at all. Fieldnotes often contain a mixture of personal reflections by the anthropologist and more objectified descriptions of places, people or interactions. Separating the anthropologist's subjective experience of the field from the objectivized data that s/he produces is both epistemologically undesirable and practically impossible. Rather, the consensus within the discipline is that anthropologists work with "situated knowledge", knowledge that is created not from nowhere but from the point of view (personal, social, cognitive, cultural, etc.) of a researcher in interaction with the world. Situating knowledge production through reflexivity is one of anthropology's primary methods for striving towards objectivity, as it allows readers better to understand how the researcher arrived at the conclusions s/he did. It is also one of the original contributions of our discipline to the social sciences.

These initial discussions within the SAA were certainly worth having and helped bring out a surprising degree of consensus about our epistemological and professional self-understandings. They led us to formulate the following position about fieldnotes and personal journals. While the material collected therein should be qualified as data from an epistemological point of view (it is indispensable for drawing conclusions and formulating our research results), it is not conducive to "open data" treatment. Fieldnotes and personal journals are often incomprehensible (and sometimes literally illegible) to anyone other than the person who wrote them. They are also virtually impossible to anonymize, creating real security risks for the anthropologist and for his or her research subjects. *Thus, as a general matter, we encourage anthropologists to maximize data protection measures for this type of data, and not to (feel obliged to) make them available to other researchers or the public through data repositories or other data-sharing mechanisms.*

In contrast to fieldnotes, other types of data – recorded or transcribed interviews, drawings, photos, focus group discussions, analytical memos, secondary literature, collections of articles from the press and/or audio-visual media, and material objects – all with their accompanying contextualizing metadata, might well be useful for and made available to colleagues and to the public, *but once again only under certain conditions.* Clearly other scientists and the public could be interested in consulting this material for a variety of purposes: comparative or historical analysis, information on specific contexts or events, or simply well-meaning curiosity about anthropological knowledge practices. The main questions concern how to guarantee that this material is correctly contextualized, and how to provide this necessary contextualization while also guaranteeing protection for the people with whom we work.

(2) Application of the FAIR principles to anthropological data

The "FAIR Guiding Principles for scientific data management and stewardship" were spelled out in an influential review paper in 2016 and were rapidly integrated into the SNSF's data management framework. The acronym "FAIR" represents the principal concepts mobilized by this framework: data must be findable (essentially through the use of digitalized metadata), accessible (through long-term secured and publicly financed data-sharing platforms or "repositories"), interoperable (through metadata that allow for exchange and interpretation) and reusable (through documentation and metadata that clarify research protocols).

As this brief description suggests, the FAIR principles were developed for the natural sciences, and it is not obvious on first reading how they apply to social science data, particularly that of a qualitative nature. In direct opposition to anthropological insistence on the need for situated interpretation of situated realities, the FAIR principles "emphasize machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data" (sited here on the GO FAIR Initiative website). It is important to stress how off-putting this language is for anthropological researchers. However, the discussions we have had with experts at FORS suggest that it is not meant to apply to the social sciences in these terms. *Essentially, FAIR data for qualitative and ethnographic research is data that is accompanied by keywords, facilitating access and sharing, and that has been correctly contextualized through metadata that specifies the circumstances under which they were collected.*

In practical terms, the FORS data repository provides the possibility for tailoring access to individual researchers' needs, in conformity with this general FAIR philosophy. Data can be made accessible under certain conditions, after a certain amount of time, or only as metadata. In the latter case, interested members of the research community or of the public could be directed to contact researchers themselves for full access, once again under conditions that researchers are free to set. *Thus, the purely technical questions of how and where to store data for open access seem relatively simple: anthropologists should address their requests to experts at FORS, who are available to provide advice before, during and after the research process*[17]. Specialists in data management at universities (we exchanged with Basel and Bern) are also willing to offer workshops for organizations such as the SAA in order to identify problems and find solutions for data sharing and archiving.

As to the new regulatory requirements under the SNSF's data management protocol, *experts we consulted assured us that anthropologists can be frank and straight-forward about the reasons motivating their decisions to make or not to make their data available to other researchers or to the general public, and under which conditions.* Reasons for not providing open access can include the need to protect the researcher's own privacy, security or personal integrity; the requirement that s/he protect the integrity of research subjects; or the fact that certain kinds of anthropological data are simply not useful for other researchers other than for biographical or historical purposes. *What is required of researchers is a brief but well-reasoned assessment of which kinds of data they wish to make available to others, and which kinds of data they wish to keep closed, or open only under certain conditions. All of these choices should be justified in the SNF DMP submission, in sufficient detail to show that the research has given the issue genuine thought and is not just checking the boxes.*

This is all good news and provides initial guidance to researchers wondering how to go about data management on a practical level. However, the problems of time, expense and training remain unresolved, and we have very little concrete experience to go on. Our hope is that discussions with the experts at FORS can help provide easily assimilated and appropriate protocols for researchers facing new challenges in this area. To clarify these questions and reach a common understanding of precisely what the FAIR paradigm means for anthropology, *we believe it would be necessary to establish an* ad hoc *working group for qualitative research in conjunction with the specialists at FORS and specialized university divisions.* The kinds of substantive questions this working group could answer are: what can usefully count as a "data set" for anthropological data and what kind of PID (persistent identifier) would make sense for our discipline? What does it mean to "describe" a data set, and to make this information "accessible"? What kinds of information count as "metadata", should it be standardized and how? How could these standardized metadata be stabilized over time and across scientific communities? What language(s) would metadata be recorded in? Adapting the FAIR framework to anthropological data would also raise significant practical and financial issues (see below).

(3) Data "minimization" and "destruction"

Other issues present in the literature also indicate how thoroughly data management principles have been formatted by activities and domains that have little to do with anthropological research. The first involves the notion of "data minimization", recommended in a key European Commission paper on "Ethics and Data Protection". With "data minimization", researchers are encouraged to collect "only the data that [they] need to meet [their] research objectives" (p. 10). Clearly this recommendation has been formulated with other audiences in mind, notably commercial enterprises doing market research by culling Big Data on the Internet.

---

[17] See the information available on the FORS website.

For anthropologists trying to understand and appropriate the positive elements in the open science movement, statements such as these understandably raise hackles. Inductive research involves a back-and-forth between data collection and hypothesis generation that makes it impossible to collect targeted data of this sort. Indeed, in so doing, anthropology would lose much of the added value it brings to the social sciences, for it would oblige itself only to research issues for which is has formulated precise questions or testable hypotheses in advance. *Once again, our discussions with experts at FORS have cleared up much of our initial resistance on this issue, as we have concluded that the evolutive, processual nature of anthropological research is well understood and accepted by funding and regulatory agencies. From these discussions, it has become clear that exploratory and open-ended research protocols are acceptable, as long as these epistemological and methodological choices are clearly explained in the data management plan*. Nonetheless, we would like to emphasize how poorly the language of "data minimization" sits with the anthropological research tradition, *and we urge the SNSF to consider creating a document specifically addressed to the qualitative social sciences in order to clarify potential misunderstandings.*

The second issue revolves around expectations for the length of time that data should be maintained in repositories. The SNSF protocol for DMPs currently suggests that personal data should be destroyed or only kept for a short period at the end of the project. This may make sense for experimental data in the natural sciences that has no intrinsic interest in and of itself once it has been integrated into or compared to other data sets or meta-analyses. However, the notion that anthropological data ought to be destroyed after analysis goes against much of what makes anthropological data useful and potentially sharable in the first place: the historical perspective they give researchers on the evolution of events, life trajectories, and research sites.

Once again, discussions with experts at FORS have helped clarify how the language of "data minimization" and "destruction" could be applied to our discipline. Personal data that is no longer necessary for analysis and writing-up could be destroyed to minimize risks to our research subjects. However, non-sensitive data could be stored and made accessible for as long as it is useful, under the conditions that our research subjects are aware and have actively agreed that data to which they have contributed will be archived. It may be that given the different embargos that researchers place on access, much of the data that anthropologists wish to make available will only become public years after the project has ended. This would require that researchers make choices about what they wish to make accessible far in advance and under circumstances that will most certainly differ from those prevailing at the time the embargo is lifted. Furthermore, long-term data storage raises practical issues, such as how to maintain contact with researchers who may well change institutions or move out of the professional altogether. *In all of these matters, anthropological researchers will need institutional support to help them think through new issues and translate them into appropriate disciplinary protocols. These are precisely the sorts of questions that could be discussed in the ad hoc working group on anthropological open data*[18].

(4) Training, time and resources

Clearly, the move to make anthropological data publicly available opens up new opportunities for anthropologists. These include increased collaboration between researchers, increased sharing of hard-won interviews, descriptions and mappings, and increased possibilities for restituting our analyses to the people we work with and to our various publics. It also enables long-term, cumulative approaches to evolving issues and field sites. However, as should be clear by now, moving disciplinary practice in this direction is not a small task, nor is it entirely evident that anthropologists will have the motivation, skills, time and funding necessary to adopt these new practices in a meaningful way[19].

---

[18] For a preliminary look at the literature on this subject, see Diaz and others in a recent special issue of the *Bulletin of Sociological Methodology* (2021).

[19] For a thorough and amusingly honest examination of the reality of data sharing and archiving practices in the qualitative social sciences, see Amiotte-Suchet, L. *op cit.,* note 2.

What is glaringly obvious is that none of this can happen without a significant investment of resources on the part of researchers, universities and funding agencies. Anthropologists, and qualitative social scientists more generally, would need to have the institutional space and support to imagine new kinds of research design, digital infrastructures and operational logics. We are aware that the SNSF offers the possibility of budgeting up to CHF 10'000 in funding per project for data management. This is a very helpful first step that allows researchers to undertake anonymization and standardized metadata annotations for the data they wish to make available. *However, the risk is real that without a concerted, cross-disciplinary and even international effort to create standardized vocabulary and common protocols for sharing and archiving qualitative research, these data will simply languish in data repositories, and will be neither findable, accessible nor reusable.*

We do not wish this to be taken as a rejection of the open data enterprise. We do, however, want to avoid that the push for open data become a kind of formalized "jumping through the hoops" that does not lead to increased access in the real world. These questions, and many more, could be studied in detail in the proposed working group on FAIR qualitative data management proposed above.

Conclusion

The open science paradigm represents a fascinating development, a quasi-revolution for the natural sciences and an intriguing set of challenges to the human and social sciences. The SAA wishes to "join the bandwagon" and continue to reflect on the principles and practices that the move toward open science promotes. We see many interesting opportunities for transforming some of the more individualistic, inward-oriented habits of our discipline. Indeed, a number of anthropologists are already developing innovative models for shared research platforms, such work being done on the prestigious Smithsonian Institution Archives or the Platform for Experimental and Collaborative Ethnography (mentioned above). However, we also see potential traps and difficulties for a discipline that is already underfunded and understaffed in comparison to the natural sciences.

From the point of view of the SAA, and somewhat paradoxically, the single clearest benefit of this new paradigm lies not in its encouragement towards openness but in its reminders about closure. With the digital data revolution, anthropologists must thoroughly and urgently rethink their relation to data security, for the protection of our research subjects and also for our own protection. Over the next few years, the SAA will prioritize closure over openness, through training and infrastructure development, while contributing in parallel to on-going discussions on open data.

Ellen Hertz
President of the Swiss Anthropological Association
Neuchâtel, November 4, 2021