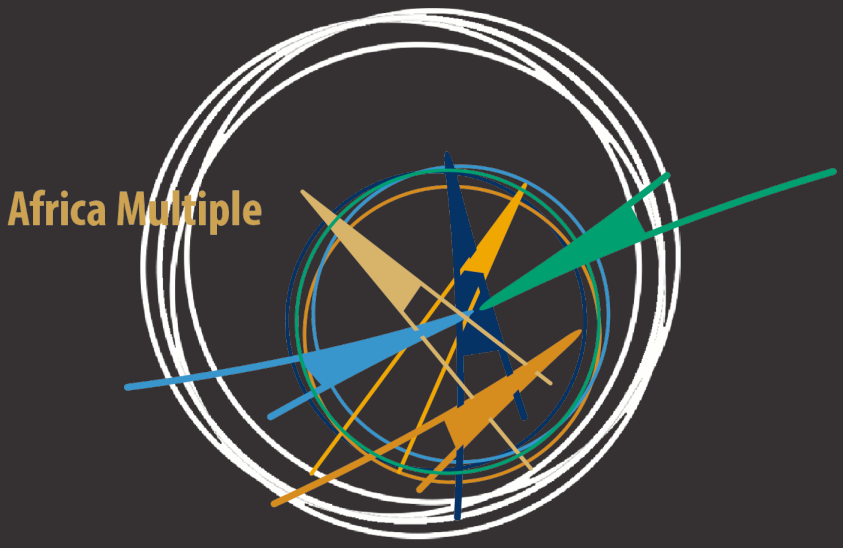
9

University of Bayreuth
African Studies Online



Frontiers in African Digital Research

Anja R. Dreiser and Cyrus Samimi, 2022

UNIVERSITÄT BAYREUTH

INSTITUTE OF AFRICAN STUDIES

africamultiple CLUSTER OF EXCELLENCE

9

University of Bayreuth African Studies Online

Frontiers in African Digital Research

Conference Proceedings

Anja R. Dreiser, Cyrus Samimi, 2022

CC BY 4.0

Institute of African Studies (IAS)

The *University of Bayreuth African Studies Online* are published by the Institute of African Studies (IAS) in Bayreuth, Germany.

The IAS promotes African Studies at the University of Bayreuth by supporting scholars from a wide range of academic disciplines from almost all faculties. It facilitates cooperation between researchers and institutions engaged in Africa-related projects, as well as teaching, both on campus and around the world. The IAS consists of three central bodies: the Iwalewahaus, the Africa Multiple Cluster of Excellence, and the Africa Research Center (*Forschungszentrum Afrika*), which is currently under construction.

University of Bayreuth African Studies Online presents results of Africa-related research at the University of Bayreuth. Publications include articles, edited collections, and single authored monographs – including but not limited to PhD-theses. Submitted papers are subject to external double-blind review. Contributions may be submitted to the editor-in-chief: Dr. Lena Kroeker (mailto:Lena.Kroeker@uni-bayreuth.de)

The *University of Bayreuth African Studies Online* is registered under ISSN 2628-1791 on the EPub document server at the university library:

2005 - 2018

https://epub.uni-bayreuth.de/view/series/Bayreuth_African_Studies_Online.html

2018 - ongoing

https://epub.uni-bayreuth.de/view/series/University_of_Bayreuth_African_Studies_Online.html

INSTITUTE OF AFRICAN STUDIES

Institute of African Studies

Director: Prof. Dr. Eva Spies
Vice-Director: Prof. Dr. Thoko Kaime

University of Bayreuth
Wölfelstr. 2
D-95440 Bayreuth

Phone: +49 (0)921 554511
Fax: +49 (0)921 554502

www.ias.uni-bayreuth.de
IAS@uni-bayreuth.de

Table of Content

Institute of African Studies (IAS) **iii**
Africa Multiple Cluster of Excellence **v**
About the Authors **vi**
Table of Content **xiii**

Frontiers in African Digital Research

Frontiers in African Digital Research **15**
Anja R. Dreiser, Cyrus Samimi

Section 1: Four Perspectives on Digital Transformation **19**
The Double Enclosure of Co-Produced Knowledge **21**
Stefan Ouma
Digitalisation of the Academe in Africa **39**
'Tunde Ope-Davies
Digital Colonialism and Infrastructure-as-Debt **65**
Michael Kwet
Mediatization of ethnographic Research Practice in the digital Age **79**
Mirjam de Bruijn

Section2: The Legal Framework of and Ethical Considerations in Data Processing **97**
GDPR and Research Data in the European Union **99**
Oliver Watteler

Problematic Practices and New (Digital) Avenues in Copyright and Intellectual Property Rights in African Music **129**

Markus Coester

Section 3: Capturing Semantics for Knowledge Management Systems **147**

Modelling in the Humanities **149**

Øyvind Eide

Towards an Integrated and Contextualized Research Data Management at the Cluster of Excellence *Africa Multiple* **155**

Jonas Huisl, Myriel Fichtner, Philipp Eisenhuth, Oliver Baumann, Mirco Schoenfeld, Stefan Jablonski, Cyrus Samimi

Linguistic Diversity and Bias in Online Dictionaries **173**

Gábor Bella, Khuyagbaatar Batsuren, Temuulen Khishigsuren, Fausto Giunchiglia

Connecting Knowledge **187**

Matthias Fuchs, Stefan Wolff

Section 4: Digital Repositories in African Heritage and African Studies **191**

Digital Return as a Question of Ethics **193**

Jody Butterworth

Setting Up a Digital Archive for Ethnographic Data **213**

Wolfgang Kraus

Portals and Platforms **233**

Mireille Djenno

Setting Up a Digital Archive for Ethnographic Data

Challenges, Strategies, Experiences

Wolfgang Kraus (University of Vienna)[112]

1 Introduction

Can material from ethnographic research and related qualitative methodologies be made useful and meaningful beyond its original research context? Many ethnographers are deeply sceptical about the possibility of others reusing their data, and they have good arguments to support their position. A major one among them is that submitting to the increasing demand for open data might have a damaging impact on relations between researchers and research subjects because it risks to undermine the relation of trust that is fundamental to most ethnographic research. Giving away data means giving up control about what is done with them, while many researchers feel they can never hand over to others the responsibility for the ways in which their data are used (cf. Imeri 2017; 2019: 49 f.; Pels et al. 2018).

These arguments and others against the open data optimism prevalent in many other fields are well founded and must be taken very seriously. Conversely, there are equally good reasons for archiving ethnographic data in order to preserve and make them available for various kinds and scenarios of reuse – provided that it is done right.

In what follows I will report on what might be described as a grassroots and researcher-centred initiative that I have been leading at the University of Vienna, the establishment of an

[112] Contact: Wolfgang.Kraus@univie.ac.at

ethnographic data archive since 2017. I will discuss the main motivations, strategic considerations and experiences made in a project that is still in its early stages, even though it is permanently established by now. Most of my argument is framed in terms of the specificity of ethnographic research; however, a large part of the challenges we faced and the strategies we developed is relevant for many kinds of qualitative data in the social sciences and humanities.

I have been considering questions of data archiving at various points in my career as a researcher, not least because of my interest in ethnographic photography and audio documentation, which tend to produce *data objects* in a more obvious sense than other forms of ethnographic communication. I also had the good fortune of cooperating with and being supported by an audio archive right from the start of my own field research – indeed, the oldest institution of its kind world-wide, the Vienna-based Phonogrammarchiv of the Austrian Academy of Sciences – where I was able to archive a selection from my field recordings made in Morocco in the mid-1990s (see Kraus 2007).

The specific idea of setting up an ethnographic data archive at the Vienna Department of Social and Cultural Anthropology first came up in 2013–14, while I was head of department, in conversations with my deputy and friend, Gertraud Seiser. We were aware that the department was heading towards a major generational transition, with several colleagues bound to retire over the next years, and discussed the possibility of creating a local archival infrastructure to preserve the ethnographic data they had collected during their research careers, involving them in the process while they were still available. In addition to such considerations, our ideas about why and how this should be done were based on our own long-term experiences of conducting and teaching ethnographic research (e.g., Seiser 2012; Seiser and Schweitzer 2010; Kraus 1998, 2004).

For both of us, the increasing call for data management and open access to research data was not yet relevant when we began to reflect on meaningful ways of preserving ethnographic material for secondary reuse. However, starting in the natural sciences, this has become a major concern in all scientific fields over the last decade (see Allianz 2010). The debate on researchers' responsibilities in producing and handling their data, under such headings as Open Research Data, FAIR data (FORCE11 n.d.) or data management, now provides an important context for all considerations of data preservation and archiving (see, e.g., Imeri 2017, 2019; Mosconi et al. 2019; Pels et al. 2018).

2 Managed and Open Research Data

Therefore, I will begin by looking at this wider context of demands to manage, share and open research data. It is evident that these demands are predicated on an understanding of what research data are, itself based on certain assumptions about the scientific process of producing knowledge. Let us see how appropriate these assumptions are for ethnographic research.

The notion of research data is often treated as self-evident or only descriptively specified in listings of what may constitute research data[113]. Trying to turn this into a definition, as is

[113] "Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden.

sometimes done[114], risks becoming a circular argument: research data are defined by their role in the research process, while research itself can be defined through the systematic use of data.

A definition in the strict sense of the term must in some way also specify how data relate to the object of research. A much-quoted formulation originating from the United States Office of Management and Budget reads, "Research data is defined as the *recorded factual material* commonly accepted in the scientific community as necessary to validate research findings…" (OMB 2006, my emphasis). A slightly expanded variation has been picked up by several UK institutions (see, e.g., EPSRC n.d., itself quoted by other institutions). Another definition states, "Research Data [are defined as] Data that are *descriptive* of the research object, or are the object itself" (University of Bath 2011, my emphasis).

The understanding of data entrenched in such definitions seems to me to be based on two sets of assumptions. In epistemological terms, it assumes that research data document aspects of the real world in a factual or, at least, descriptive sense and in a way that is largely independent of their specific research context. Although due allowance is often made for disciplinary specificities and differences, the basic model is often that of the natural sciences. If data reflect the real world independently of their research context, then they are unproblematic to reuse in a different context. Another important assumption is that access to research data serves "to reproduce and verify the results" of research, as the FWF, Austria's main funding institution for basic research, states in the context of its Open Access policy (FWF n.d.).

A second, related set of assumptions goes like this: once public money is invested to fund research, the results (publications) should be made openly accessible. "Taxpayer-funded research" is an important buzzword here, and one that raises interesting questions concerning the role of national boundaries with regard to access to knowledge (see, e.g., the US Alliance for Taxpayer Access[115]; for a British example, see the Foreword in Concordat 2016: 2). By the same logic, the data collected during research become assets that cannot be owned by the researcher. Rather, they must be shared and the public – including other researchers – has a right to access and re-use them.

Even though I can partially agree with some of the ideas involved in the second set, while remaining sceptical of their neoliberal thrust, I want to underline that taken together these assumptions imply three things that sit rather uncomfortably with the practice and self-understanding of ethnographic research:

- Ideals of objectivity and replicability of research that I consider epistemologically mistaken;

Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden" (DFG 2015).

[114] "Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen" (Allianz n.d.).

[115] https://www.taxpayeraccess.org/. Last accessed 1 June 2022.

- Principles of cost efficiency as well as efficiency in generating knowledge;
- A sharp discontinuity between everyday knowledge/experience and research-based knowledge.

3 Characteristics of Ethnography

Most readers are probably aware of the basic principles underlying ethnographic research. Nevertheless, for the sake of my argument and in contrast to the assumptions outlined above, let me quickly sum up the main characteristics and assumptions of ethnography. Since the methodological approach of ethnography emerged in the field of social and cultural anthropology, I use *ethnographic* and *anthropological* as overlapping and almost synonymous terms, notwithstanding the fact that ethnographic methods are also practiced in many other fields in the social sciences and humanities.

Research is typically done over extended periods of time in close collaboration and exchange with research subjects. It is based on communicative relations with those being researched and sees them as active participants rather than passive objects of observation. Much ethnographic research deals with personal lifeworlds, a fact that implies important issues of trust and responsibility.

Ethnography is not a method but rather a methodology that relies on an open and flexible combination of tools. Hence, ethnographers tend to produce varied and multiple forms and formats of data. Whether in analogue or in digital formats, ethnographic data tend to be technically diverse, and different kinds of data are interpreted in relation to each other and to the overall research context and experience. In recent years and with the establishment of new digital tools, media and ways of disseminating and sharing knowledge based on ethnographic research, this characteristic has come to be discussed as "multimodal anthropology" (e.g., Collins and Durington 2018).

In typical suggestive but vague fashion, Clifford Geertz wrote of the fact "that what we call our data are really our own constructions of other people's constructions of what they and their compatriots are up to" (Geertz 1973: 9). Even if the understanding of the task of anthropology, and the category of data in ethnography, may have expanded a lot since Geertz (1973) proposed his notion of "thick description", most anthropologists today would subscribe to a vision of ethnographic research as based on a relation of dialogue between researchers and the people they work with. From such an understanding, several fundamental assumptions of ethnographic research follow:

- Data are not simply "found" but co-constructed in a process of dialogue between the researcher and the research subjects. They do not merely document facts "out there" but are representations that contain the voices, or perspectives, of both sides involved: the researchers and the researched.
- Therefore, ethnographic data cannot simply belong to the researchers (and even less to their institutions). They also belong to the research subjects and their communities, who have their own interest in the data.

- There are no "raw", uninterpreted data in ethnography. The dialogic process of "making" data is by necessity a process of interpretation.
- Being products of relation and dialogue, ethnographic data are neither objective nor subjective. Both of these notions presume a clear distinction between the observer and the observed object, a distinction that is neither meaningful nor possible in ethnographic research. Therefore, ethnographic data are never simply "descriptive of the research object" (University of Bath 2011), but first of all of the ethnographic relation between researchers and research subjects.
- Both sides involved have their social and cultural context which they bring into the ethnographic encounter. Therefore, there is a gradual difference but no discontinuity between ethnographic knowledge and everyday knowledge and experience.
- Taking into account the historical context in which the ethnographic methodology emerged and, even more importantly, the responsibility that comes with its practice, ethical considerations clearly must take precedence over considerations of efficiency.

On a side note, I should make it clear that I am not claiming that there should be a separate epistemology for anthropological and ethnographic research in contrast to, say, the natural sciences. Rather, I am convinced that the epistemological basis of the assumptions outlined above holds for all forms of science, but that is not a point to be elaborated here.

A second clarification concerns the established distinction between "raw" and "processed" or interpreted data, a distinction that makes no sense with regard to ethnographic material[116]. In what appears to be an opposite view, the anthropologist Peter Pels states, "Anthropologists should… insist on making an epistemological distinction between 'raw' and 'processed data'" (Pels et al. 2018: 394), but it turns out that he means something different by these terms. Simplifying his complex argument, the main point is that ethnographic data typically contain deeply personal information, thus, "Extensive processing of raw materials (beyond mere anonymisation) becomes inevitable if others are to reuse them" (Pels et al. 2018: 394). Yet, that risks to render them useless for further research. I agree; however, "processing" here is not used in the predominant sense of making data usable for analysis; rather, it refers to the task of preparing them for reuse by others. Nevertheless, I do not agree with Pels' argument – if I am reading him correctly –that sharing data in principle cannot be reconciled with the research subject's "rightful claims to knowledge shared with researchers" (Pels et al. 2018: 394). It can – provided that research subjects are explicitly conceptualised as forming part of those for whose access and reuse data should be prepared and archived, and that necessary precautions are taken. These topics are discussed at length in later sections.

4 Towards an Ethnographic Data Archive

Most anthropologists will probably tend to agree with my characterisation of principles of ethnographic research. However, the project of setting up an ethnographic data archive and the

[116] The distinction is also problematic with other kinds of data. From a statistician's point of view, Barrowman (2018) argues convincingly that quantitative data are never raw nor context-free.

archival strategies we devised in the course of our activities relied on additional assumptions which may hold more potential for controversy.

Our basic premise is that, generally speaking, ethnographic data have an intrinsic interest beyond the primary research context, for two main reasons. First, they are complex and rich in ways which are hardly ever fully exploited in the original analysis. This is often a matter not only of complexity but also of sheer quantity.[117]

Second, being situated in time and space, ethnographic data are historical by nature. With the passage of time and as a consequence of transformation and change, they may become interesting and relevant in unforeseeable ways.

Thus, there are good reasons to preserve ethnographic data in order to make them accessible and reusable beyond the original research context – reasons that are entirely unrelated to the rationale of the Open Research Data discourse. However, it would be mistaken to treat them as independent of their original research context. Ethnographic knowledge is embedded in social relations and in complex corporeal experience. Specific data sets can represent the research process only incompletely. As stated above, ethnographic data tend to be diverse, and different kinds of data must be interpreted in relation to each other and to the overall research context. A guiding principle for our archival activities is therefore that the link between specific data and their research context must be retained as far as possible.

I proposed the idea of a pilot project for setting up digital ethnographic data archive in early 2015 via the Faculty of Social Sciences and again a year later. It took the University of Vienna almost a year to evaluate how this idea fit into the overall digitisation and emerging data management strategy before the decision was taken to finance a two-years pilot project that would be attached to the Vienna University Library. The data were to be archived in PHAIDRA, the "repository for the permanent secure storage of digital assets at the University of Vienna"[118]; our cooperation with the PHAIDRA team has proved to be a highly productive and most pleasant experience.[119]

The main people involved were Igor Eberhard, who was hired for a half-time position as archive manager, Birgit Kramreither, the head of the Social and Cultural Anthropology Library, as coordinator, and myself as scientific leader. One of our more obvious insights during the pilot phase was that work capacity was notoriously short. When several months before its end we pondered how to make the best use of the remaining funds, we decided to employ a student assistant, Jasmin Hilbert, to support us.

[117] In my own case, a significant proportion of the material from my second field research phase in central Morocco between 1995 and 2005 remains unused. I doubt that this is a singular experience.

[118] See the PHAIDRA website, https://phaidra.univie.ac.at, last accessed 1. June 2022.

[119] Several people have substantially contributed to establishing and developing the *Ethnographic Data Archive*; most noteworthy among them are Maria Seissl, Susanne Blumesberger, Raman Ganguly, Rastislav Hudak and Claudia Feigl. I and the *eda* team are deeply indebted to them for their ongoing support.

After the pilot project ended in February 2019, the Vienna University Library decided to continue the activities as part of its regular operations. Eberhard's employment was extended and has now been made permanent. Hilbert also continues working with us but on a terminable basis. With the permanent establishment of the archive, we changed its name from the earlier *Projekt Ethnographische Datenarchivierung* to *Ethnographisches Datenarchiv*, or *Ethnographic Data Archive* in English; the acronym *eda* remains the same.[120]

5 Objectives and Strategic Considerations

As mentioned above, the point of departure for *eda* was the generational change at the Department of Social and Cultural Anthropology. In the case of colleagues bound to retire, as in Seiser's and mine, the primary focus was on the digital preservation of historical data from earlier researches, mostly in analogue form. The insight that it was also necessary to support ongoing research and provide data management expertise followed logically from our experience of working with the historical material. It is important to note that our initiative came out of our and our colleagues' research practice. Thus, it is not representative of the "'top down' policy push" that Mosconi et al. identify as a characteristic of Open Science, but rather of what they refer to as the "collegial desire to share data" (Mosconi et al. 2019: 756).

The *eda* team set its main objectives as follows:

- Defining archival and metadata strategies and standards adapted to the specificities of ethnographic research.
- Testing and defining best practice digitisation workflows.
- Networking and exchange with other data management and archival initiatives in related fields.
- Identifying the ethical and legal issues involved and proposing solutions.

In the medium to long term, we aim at developing a comprehensive research data management strategy for social and cultural anthropology and related fields.

Among the guiding strategic considerations, one is that cooperation with researchers makes more sense than the administration of legacies. This reflects experiences we made working with materials existing in the departmental archive, with often insufficient metadata and context information. More importantly, this conviction is based on our holistic understanding of ethnographic data as representing an interactive research process rather than separate aspects of an independent reality. This means for us that all data must be linked to the research setting and the researchers' biographies or careers and should be made accessible in a way that enables the user to take this context into account. Only the researchers themselves can accomplish this task of contextualising and interlinking the data objects in a comprehensive manner.

[120] The *eda* website (https://eda.univie.ac.at/) gives a short overview of the team, activities and cooperations; the *eda* team can be contacted at: eda.ksa@univie.ac.at.

Another guiding principle is the respect and support for the legitimate interest of research subjects and source communities in the data, and their protection from harm. Here again, only the researchers can make informed decisions about the interests and risks involved with specific data sets. As Pels notes, it is their "ethical duty to control how research materials 'go public'" (Pels et al. 2018: 395).

On a more pragmatic level, we aim at sustainability through optimised workflows, appropriate file formats, standardised procedures and metadata, and ongoing quality control. Devising best practice digitisation workflows requires a balancing of contradictory demands. The amount of work and cost involved and the required storage space should be kept low, while the technical quality of a digital copy should be such that it can be expected to be taken as an adequate representation of the analogue original even several decades from now.

Note that this also involves a judgement about the relevant aspects of an object. Imagine the case of a faded photograph. Are we mainly interested in its current appearance – that is, in the effects of history on it – or in the original information that can be restored by proper illumination and by digital editing, or both? In each case, the optimal digital copy or copies will be different. Once again, the researchers can help to make such decisions.

A further non-technical aspect is an ideal of autonomy in the digitising processes. We try to avoid having to hand over objects to third parties for digitising purposes, and have so far been successful in doing so. Finally, our digitising workflows also rely on considerations of the relation between analogue objects as potential carriers of knowledge and their digital representations, and the transformation from one state to the other. The question of the faded image raised previously shows how a conception of what constitutes the data object in relation to the research context must guide the digital representation. Another example is a Compact Cassette, which can contain several field recordings. Is the single recording the object we are interested in, or is it the cassette as an entity representing a specific moment or time span in the field? We have opted for the second as our predominant perspective. Finally, in line with the considerations above, we leave the selection of what to archive to the researchers, while offering our advice when being asked for it.

6 Challenges

Archiving ethnographic data for reuse clearly is a case of contradictory demands. On the one hand, we wish to make the data accessible while preserving their richness and complexity and the connections between them in order to keep them meaningful. On the other hand, precisely these aspects may expose the research subjects or their communities, but it is our duty to respect their interests and keep them safe from harm as best we can. This may force us to restrict or postpone access to the data and/or to process data so as to avoid identification. Finally, researchers themselves may also have an interest in protecting their privacy.[121]

[121] These and other related issues have extensively been discussed in the last few years (see, e.g., Cliggett 2016; Eberhard and Kraus 2018; Imeri 2017, 2018, 2019; Lederman 2016; Pels et al. 2018; Sterzer and Kretzer 2019; Zeitlyn 2012).

Following from this dilemma of principles, there are many practical problems that must be addressed during the process of archiving data. There are no abstract general solutions for these problems, nor is there a single recipe for balancing the underlying contradictions and tensions. Viable compromises must be found for each single case. Therefore, rather than proposing dubious answers, I phrase some of these issues – the list is not exhaustive – as necessary and pressing questions of ongoing relevance:

- How should we adequately represent the wider research context of data objects, their interlinked nature, and the researcher's positionality?
- How can we best protect the research subjects' privacy, interests and security?
- How should we deal with data from historical research where research subjects were not, and can no longer be, asked to make informed decisions about the material?
- How can we pseudonymise without risking to make the data next to useless due to loss of context? (Anonymising in the strict sense is hardly an option with ethnographic material.) The question is further compounded in the case of visual material.
- How do we make sure that the value we attach to personal non-identifiability does not harm interests that research subjects in a given field may value higher, such as social embeddedness, visibility and recognition?[122]
- What could be the possible consequences for research subjects if we decide not to pseudonymise them?
- How can we best make data accessible and meaningful for the research subjects and source communities?
- At least in the case of data that pose a potential threat to personal interests or security, it is also necessary to break up the simplifying notion of "source communities". How can we identify who is entitled to represent communities in claiming access to data?

7 Data, Digitising and File Formats

The forms and formats of data we typically deal with include:

- Text (on paper and digital, the latter often in obsolete file formats): e.g., notebooks, field diaries, transcripts.
- Images (on film, paper and digital): e.g., photos, documents, drawings, maps.
- Audio recordings (analogue and digital on various kinds of data carriers): e.g., interviews, narrations, recitations, rituals, music. With some digital material, again, the problem of file formats comes up, as with some files from closed system dictation devices in highly

[122] See Zeitlyn 2012: 470 f.

compressed proprietary formats. Today, most high-quality digital recorders produce files that can be archived without any conversion, depending on the settings.

Film and video material have not yet been part of our activities. This is due to the specific complexities of digital video such as the multiplicity of formats and the need for data compression; unlike image and audio, storing video uncompressed and in highest resolution is not a practical option. Furthermore, we do not have a lot of personal experience in handling video data. This does not however mean that we exclude video as a matter of principle.

Text or images on paper and similar support materials are scanned or photographically copied. Depending on the size of the original, a photographic copy with a high-resolution digital camera and a high-quality macro lens can provide better real resolution and image quality than most flatbed scanners and always allows better control of lighting characteristics. Digital copies are stored either as image or PDF files. The former option is preferred if it might make sense to edit the image during use, for instance to improve the readability of manuscripts; the latter provides easier access, especially with multi-page documents.

For digitising photos on film (negatives or diapositives) we employ a Hasselblad Flextight X1 scanner, one of the highest-quality film scanners on the market today. Alternatively, and more often, we copy film material with a high-resolution digital camera (Pentax K-1) and extremely high-resolution industrial (line-scan) lenses.[123] Extensive testing proved that with a carefully optimised workflow this is a much faster and in most respects technically superior solution, at least for 35mm film.[124] Image files are archived as TIFF files, usually in Adobe RGB (1998) or Gray Gamma 2.2 space with a resolution of 16 bits per channel (or rarely as lossy JPEG 2000 files with some images reproducing text information).[125]

We use an RME Babyface USB interface as analogue-to-digital converter to transfer analogue audio material – typically on Compact Cassette tape – to LPCM encoded audio with a sampling rate of 96 kHz and 24 bits resolution. A challenge during playback is to determine whether the source material has been recorded using Dolby noise reduction, a fact that is often undocumented. Audio

[123] With its Pixel Shift Resolution system, the 36-megapixel Pentax K-1 DSLR provides full colour information for each pixel, unlike competing full-frame cameras with the same nominal but less real resolution. This is an essential feature to achieve best image quality and avoid the typical imaging artifacts of conventional Bayer sensor cameras. It has a very visible advantage for copywork too where it helps to avoid aliasing issues, such as moiré.

[124] The Flextight X1 has a maximum nominal resolution of 6300 ppi while the Pentax K-1 images provide about 5000 ppi for 35mm film. However, testing showed that this does not translate into a visible advantage of the X1 scans. With medium format film the X1 is able to provide visibly better resolution: 3200 ppi versus about 2200 ppi with the Pentax K-1 for 6x6 negatives or slides.

[125] 16-bit resolution is mandatory for all image editing steps. It is also preferrable for archival files because there may be a need for further editing during reuse. Note that the lossless LZW compression that may be used with TIFF files does *not* reduce files sizes with 16-bit images.

files are archived as WAVE or as losslessly compressed FLAC (with the advantage of smaller file sizes).[126]

Born-digital materials – those that originate in digital form – are best archived in their original format, but this may be problematic in a long-term perspective in the case of proprietary and/or obsolete formats. In these cases, we try to convert them to suitable, preferably open archival formats and archive these as well. Moreover, they may be stored on obsolete data carriers and require special hardware to convert them to archivable files. So far, we have been able to gain experiences with obsolete text files which we managed to read using the open-source LibreOffice software. The preferred archival format for text files is PDF/A. We also processed a large number of digital audio recordings on MiniDisc and from a proprietary and closed dictation system, both with their own specific difficulties. These recordings are stored in their original sampling rate with 16 bits resolution.

We have not yet started archiving born-digital photographs. Here again, image files are best stored in their original format. The lossy JPEG format commonly used by digital cameras is acceptable for digital field photographs but should be avoided for digitisation purposes. Whenever raw file formats – which provide better image quality and allow users to profit from future developments in raw converting technology – are used, it makes sense for archival purposes to convert proprietary raw formats to the open DNG format using the Adobe DNG Converter or other equivalent software. However, it is wise to archive processed TIFF files too. There is disagreement as to whether DNG is a suitable format for long-term archiving,[127] but it should be understood that raw files hold information which is lost when converting them to conventional image formats. Regardless of how this may be judged from the archival perspective, it must be kept in mind that converting raw files into images is not just a technical process but involves of necessity a creative interpretation. With all relevant digital image formats the options of embedding metadata, e.g. the IPTC and EXIF standards, can be extensively used in order to provide technical, descriptive and context information.

For some of the routine processes, we have written detailed step by step guidelines based on our equipment and preferred workflows. For others, this remains to be done. We have also developed a file naming convention and guideline.[128] These guidelines are meant to ensure consistency in processes and compliance with best practice workflows.

8 Metadata

In the process of establishing *eda* we gave a lot of thought to metadata-related questions, for several reasons. Seemingly innocent descriptive information – such as names attached to collectivities – could transport and cement problematic perspectives or ethnocentric

[126] See https://xiph.org/flac/index.html. Last accessed 1 June 2022. The downside is that FLAC files are somewhat less accessible than WAVE, but the open-source VLC media player, for instance, can be used for playback.

[127] See, e.g., DFG 2016: 19 for a negative view.

[128] File names also serve as call numbers for digital objects, but are separate from the object numbers and handles assigned by the repository.

assumptions. Attention to the research context was a guiding principle for our approach to metadata. Another important question was how to facilitate access to data objects for research subjects and source communities. The PHAIDRA team proved extremely flexible and supportive in helping us to develop an *eda*-specific metadata scheme and submit form.[129]

For metadata the PHAIDRA repository relies on a linked-data approach combining several established metadata standards, including Dublin Core, BIBFRAME, SKOS and others. We suggested several additional metadata fields that we considered useful, which were then translated into existing categories.[130]

While access to data objects in PHAIDRA can be restricted or blocked – an indispensable feature for *eda* for obvious reasons –, a basic given of metadata in PHAIDRA is that they are always public and open. They cannot be restricted, and there can be no authorship and no copyright (or *Urheberrecht* in the German and Austrian sense, which is based on a different reasoning) in them.

Our interest in making the research context transparent and tangible turned out to be in potential conflict with this policy. What about personal information that helped to contextualise the data but was not fit for public access? What about background information about historical materials that relied on a research effort and personal interpretation and opinion, in other words, an authorship that should be credited? Our solution is the notion of *context data* that we devised as a pragmatic workaround for these issues. Context data are data objects – generally in text format (PDF/A) – providing information on other objects that is not suited for metadata because it is too complex and extensive, needs being protected, or requires authorial responsibility and copyright. As separate data objects, they can have access restrictions; they are referenced in the metadata of the related objects. However, in the present form, they are not searchable. A significant category of context data is information on research projects and researchers' biographies and research careers, as in the biographical interviews that Igor Eberhard conducted in 2017 with Elke Mader, then professor at the Vienna Department of Social and Cultural Anthropology.[131]

Another consideration was that, given the potential complexity of data objects, we wanted to clearly distinguish between several object categories in metadata, some of which might coincide in the case of a given object. We referred to these categories as:

- born digital-object,
- digital copy,
- first-order analogue object, and
- second-order analogue object.

[129] Rastislav Hudak from the Vienna University Computer Center's Software Design & Development team deserves special credit for his invaluable support in this context.

[130] For an example of an object with fairly complete metadata, see https://phaidra.univie.ac.at/o:1069269.

[131] See https://phaidra.univie.ac.at/o:1146526. Sadly, Elke Mader passed away in August 2021.

While the first two are obvious, the others require an expanded definition. Imagine the case of a musical recording. A straight WAV file from a digital recorder is a born digital-object. A recording on a Compact Cassette, when digitised, is the digital copy of the audio content of a first order analogue object, the cassette. When the recording, beyond the immediate musical event, is considered a representation of a specific instrument that is being played, the instrument is a second-order analogue object. Each one of these categories, as far as they exist in a given object, requires its own metadata. An analogue object in the second sense may also exist in the case of a born-digital object. The *eda* metadata scheme lets us follow this approach, even if an appropriate terminology – not necessarily using our working terms, which might be improved – has not yet been implemented because it is difficult to translate into existing metadata standards.

The submit form that the PHAIDRA team designed for *eda* is relatively transparent and self-explanatory. This will allow us to provide researchers with personalised or project-specific accounts in order to upload their own data and metadata (something we have not yet tried). This option is further facilitated by the possibility to create project-specific metadata templates.

A major metadata-related issue that remains is the lack of appropriate controlled vocabularies. The need for such vocabularies is now widely perceived as pressing, and activities are taking place in various contexts, e.g., in the project GND for Cultural Data (GND4C) (DNB 2019). Our aim is to stay abreast of these developments and, if possible, to participate in them in the future. Some other problems with the standardisation of metadata, concerning for instance the local context of data objects, are also not yet satisfactorily resolved. For these and other reasons, the whole area of index terms is still very underdeveloped in *eda*.

9 Archived Material

We started working with material from a few selected persons who were prepared to participate in the process of finding our way through the intricate practicalities of archiving ethnographic data. The first were Elke Mader with text, audio and photographic material from her two research phases in South America (1979–81 in Peruvian Amazonia and 1990–99 in Ecuador), and Gertraud Seiser with interviews from a field school with MA anthropology students in rural Austria which she taught in the context of the KASS project in 2005.[132] In addition to this material, we worked with images and documents from the ethnographic collection and archive of the Department of Social and Cultural Anthropology. Finally, I brought in field notes, photographs and audio recordings from my research in Morocco (1983–85 and 1995–2005). Simultaneously working with concrete data objects and developing our strategies, workflows and metadata approach proved very productive because different items tended to raise specific issues and forced us to come up with workable solutions.

In addition to the data that have already been archived, we prepared and processed a large number of objects which have not yet been ingested in PHAIDRA for various reasons – lack of work capacity, incomplete metadata, open questions concerning access and so on. By January 28,

[132] http://web.eth.mpg.de/kass/, last accessed 1 June 2022; https://phaidra.univie.ac.at/o:1071491; Seiser and Schweitzer 2010.

2021, *eda* has archived 561 objects. More than half of them are photographs,[133] several are text documents (text- or image-based PDF/A files)[134] and single audio files[135]. A few are collections which can be used to bring together related objects for easier access.[136] The rest are container objects.

The container object is a data object that consists of of several files representing the same analogue object. We conceived this notion based on the idea that the original analogue object is an entity capable of better elucidating the research context than the detached information it might hold. Consider, again, the example of a Compact Cassette containing several recordings made over a time span. Rather than archiving separate tracks, we opted for the format of the container object allowing us to retain the original connection between the recordings. In addition, a Compact Cassette often comes with notes written on it, referring to its content, or on the cardboard insert inside the box. Most of the container objects archived so far represent Compact Cassettes and consist of audio files and photos of the two cassette faces and inserts.[137] Other container objects represent photographs from the ethnographic collection of the Department with extensive captions on the verso.[138]

In addition to these objects, a large number of items from the archive and ethnographic collection of the Department has been digitised and ingested using the *eda* metadata scheme and submit form, but with the Ethnographic Collection as owner rather than *eda*. Most of these are documents such as correspondence and educational slides.

Other than the persons named above, several colleagues have voiced their interest or agreed to work with *eda* in order to archive material from their researches, but we have not yet started cooperating. In principle we are also open to working with researchers from other institutions than the University of Vienna. We have made contacts with some researchers whose material seems to us to promise synergies with data sets we already have but, again, this has not yet resulted in concrete cooperations.

10 Conclusion and Outlook

The main assumption underlying *eda*'s activities is that ethnographic data constitute historically situated representations of a world in flux. As such they do have a value beyond the primary research context and should in principle be preserved and shared. However, the dialogic character

[133] E.g., https://phaidra.univie.ac.at/o:1096509.

[134] E.g., https://phaidra.univie.ac.at/o:947378.

[135] E.g., https://phaidra.univie.ac.at/o:1048725. This object and some other audio files are short excerpts from longer field recordings containing oral traditions in *Tamazight* Berber language that I published and analysed (Kraus 2019: 79). Being cross-referenced with the publications, they exemplify one way of using the repository to connect data and publications.

[136] E.g., https://phaidra.univie.ac.at/o:1146527.

[137] E.g., https://phaidra.univie.ac.at/o:957555.

[138] E.g., https://phaidra.univie.ac.at/o:936768.

of ethnography and the access ethnographers gain to personal life-worlds raise important issues of confidentiality, privacy and reciprocity.

As discussed above, there is a basic contradiction between the aim of making data accessible to others outside the primary ethnographic relation, and the ethical obligations that follow from that relation. This contradiction cannot be resolved on the level of principles. Nevertheless, despite the difficulties and challenges discussed above, it makes sense to find pragmatic but ethically sound compromises that allow us to preserve and share – with various possible forms of restrictions – ethnographic data without harming our research subjects. After all, the situation is not so unlike the challenges that come along in the process of ethnographic writing and publishing – challenges for which we routinely find pragmatic single-case solutions.

However, these inherent tensions require a number of additional measures that *eda* will have to deal with in the future, and before archiving more sensitive data than we have had so far. They have to do with its implementation within the PHAIDRA repository but also with matters of principle such as the relation between research subjects and the data they helped to generate. The categories of ownership of, control of, access to and licence to use data need to be distinguished and clarified. Since for ethical and legal reasons in many cases opening data is not an option, flexible models of access management are required, but complex and difficult to implement. If we want to avoid the necessity of defensively precluding access to all sensitive material – an option that would defeat the short to mid-term purpose of the archive – then in the future development of PHAIDRA several graded levels of access to data must be defined and implemented.[139] This will be an important step, even if making decisions about appropriate levels then will add another layer of complexity to the archiving process.

On the surface, these are ethical and legal issues that arise after the fact of data production, unless they have been clarified through informed consent, which should be the case with more recent research. Nevertheless, I hope to have made it clear that, given the collaborative nature and other characteristics of ethnography, it is impossible to draw a strict line between research and subsequent forms of organising knowledge. It is both consistent and productive to extend the understanding of ethnography as a co-production of knowledge to the archiving process. We have not yet made serious steps in that direction, but consider it necessary not only in making decisions about access to data but also with regard to producing metadata and context data. In recent years, various approaches to organising knowledge have emerged that build on communities and collaboration and could provide models of how to proceed.[140]

If I read him correctly, Peter Pels suggests there is more than one "reason to consider social science data as indigenous or global heritage" (Pels et al. 2018: 393) – a perspective that makes both the preservation of ethnographic data and the inclusion of research subjects and their communities in the process an ethical obligation. Only such an inclusive approach will be able to

[139] For an elaborate model of how this can be done, see Sterzer et al. 2018.

[140] For an example of a community-centred approach concentrating on scholarly communities, see Herrmann and Kurzawe 2020; for a museum-based approach of collaboration with source communities, see Scholz 2017a, 2017b.

safeguard data that are findable, accessible, interoperable and re-usable (FORCE11 n.d.) in the long term for all interested parties.

11 References

Allianz der deutschen Wissenschaftsorganisationen. n.d. Forschungsdaten. https://www.allianzinitiative.de/archiv/forschungsdaten/. (last accessed 31 May 2022).

Allianz der deutschen Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten. https://doi.org/10.2312/ALLIANZOA.019. (last accessed 31 May 2022).

Borrowman, Nick. 2018. Why Data Is Never Raw. *The New Atlantis* (56): 129–135. https://www.thenewatlantis.com/publications/why-data-is-never-raw. (last accessed 31 May 2022).

Cliggett, Lisa. 2016. Preservation, Sharing, and Technological Challenges of Longitudinal Research in the Digital Age. In *eFieldnotes: The Makings of Anthropology in the Digital World*, ed. Roger Sanjek and Susan W. Tratner, 231–250. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812292213-014.

Collins, Samuel Gerald, and Matthew Slover Durington. 2018. Multimodality. In *The international encyclopedia of anthropology*, ed. Hilary Callan, 1–7. Hoboken, NJ, Chichester, West Sussex: Wiley Blackwell. https://doi.org/10.1002/9781118924396.wbiea2304.

Concordat. 2016. Concordat on Open Research Data. https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf. (last accessed 31 May 2022).

DFG. 2015. Leitlinien zum Umgang mit Forschungsdaten. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf. (last accessed 31 May 2022).

DFG. 2016. DFG Practical Guidelines on Digitisation: DFG form 12.151 – 12/16. https://www.dfg.de/formulare/12_151/12_151_en.pdf. (last accessed 31 May 2022).

DNB. 2019. GND für Kulturdaten (GND4C). https://www.dnb.de/DE/Professionell/ProjekteKooperationen/Projekte/GND4C/gnd4c.html. (last accessed 31 May 2022).

Eberhard, Igor, and Wolfgang Kraus. 2018. Der Elefant im Raum: Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 41–52. https://doi.org/10.31263/voebm.v71i1.2018.

Engineering and Physical Sciences Research Council. n.d. EPSRC policy framework on research data: Scope and benefits. https://epsrc.ukri.org/about/standards/researchdata/scope/. (last accessed 31 May 2022).

FORCE11. n.d. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing Version b1.0. https://www.force11.org/fairprinciples. (last accessed 31 May 2022).

FWF. n.d. Open Access to Research Data. https://www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/. (last accessed 31 May 2022).

Geertz, Clifford. 1973. Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures*, ed. Clifford Geertz, 3–30. New York, NY: Basic Books.

Herrmann, and Felix Kurzawe. 2020. Discuss Data: Community-zentrierter Ansatz für das Forschungsdatenmanagement in den Geistes- und Sozialwissenschaften (2): 56–64. https://doi.org/10.17192/BFDM.2020.2.8277.

Imeri, Sabine. 2017. Open Data? Zum Umgang mit Forschungsdaten in den ethnologischen Fächern. In *E-Science-Tage 2017*, ed. Jonas Kratzke and Vincent Heuveline, 167–178. University Library Heidelberg: heiBOOKS, 167–178. http://doi.org/10.11588/heibooks.285.377.

Imeri, Sabine. 2018. Archivierung und Verantwortung: Zum Stand der Debatte über den Umgang mit Forschungsdaten in den ethnologischen Fächern. In *Archivierung und Zugang zu Qualitativen Daten*, ed. Betina Hollstein and Jörg Strübing, 69–79. Berlin: Rat für Sozial- und Wirtschaftsdaten. https://doi.org/10.17620/02671.35.

Imeri, Sabine. 2019. Open Data' in den ethnologischen Fächern: Möglichkeiten und Grenzen eines Konzepts. In *Forschungsdesign 4.0: Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive*, ed. Jens Klingner and Merve Lühr, 45–59. Dresden: Institut für Sächsische Geschichte und Volkskunde. https://doi.org/10.25366/2019.07.

Kraus, Wolfgang. 1998. Contestable Identities: Tribal Structures in the Moroccan High Atlas. *Journal of the Royal Anthropological Institute* 4 (1): 1–22.

Kraus, Wolfgang. 2004. *Islamische Stammesgesellschaften: Tribale Identitäten im Vorderen Orient in sozialanthropologischer Perspektive.* Wien/Köln/Weimar: Böhlau. https://doi.org/10.26530/oapen_574825.

Kraus, Wolfgang. 2007. Als Anthropologe in Zentralmarokko: Methodologische Reminiszenzen. In *Um-Feld-Forschung: Erfahrungen - Erlebnisse - Ergebnisse*, ed. Julia Ahamer and Gerda Lechleitner, 165–179. Wien: Verl. der Österreichischen Akad. der Wiss. https://austriaca.at/0xc1aa5576%200x00152aec.pdf.

Kraus, Wolfgang. 2020. Overlooking the past from the High Atlas mountains: Ayt Ḥdiddu oral traditions. *Études et Documents Berbères* 41 (1): 71–86. https://doi.org/10.3917/edb.041.0071.

Lederman, Rena. 2016. Archiving Fieldnotes?: Placing "Anthropological Records" among plural Digital Worlds. In *eFieldnotes: The Makings of Anthropology in the Digital World*, ed. Roger

Sanjek and Susan W. Tratner, 251–271. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812292213-015.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three Gaps in Opening Science. *Computer Supported Cooperative Work (CSCW)* 28 (3-4): 749–789. https://doi.org/10.1007/s10606-019-09354-z.

OMB. 2006. OMB Circular A-110. Uniform Administrative Requirements for Grants and Other Agreements with Institutions of Higher Education, Hospitals and Other Non-Profit Organizations (11/19/1993) (further amended 09/30/1999). *Office of Management and Budget Circulars and Guidance, Washington D.C.: The White House* 39–68. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A110/2cfr215-0.pdf. (last accessed 25 May 2021).

Pels, Peter, Igor Boog, J. Henrike Florusbosch, Zane Kripe, Tessa Minter, Metje Postma, Margaret Sleeboom-Faulkner, Bob Simpson, Hansjörg Dilger, Michael Schönhuth, Anita von Poser, Rosa Cordillera A. Castillo, Rena Lederman, and Heather Richards-Rissetto. 2018. Data management in anthropology: The next phase in ethics governance? *Social Anthropology* 26 (3): 391–413. https://doi.org/10.1111/1469-8676.12526.

Scholz, Andrea. 2017a. Sharing Knowledge as a Step toward an Epistemological Pluralization of the Museum. *Museum Worlds: Advances in Research* 5 (1): 133–148. https://doi.org/10.3167/armw.2017.050112.

Scholz, Andrea. 2017b. „Wissen teilen" als postkoloniale Museumspraxis: Ein Kooperationsprojekt zwischen der Universidad Nacional Experimental Indígena del Tauca (Venezuela) und dem Ethnologischen Museum Berlin. *Sociologus* 67 (1): 59–81. https://www.jstor.org/stable/45174713.

Seiser, Gertraud. 2012. Kooperationen in der Landwirtschaft kontextualisiert: Ethnografische Fallstudie von der Mühlviertler Alm. *Österreichische Zeitschrift für Soziologie* 37 (4): 385–402. https://doi.org/10.1007/s11614-012-0063-1.

Seiser, Gertraud, and Peter P. Schweitzer. 2010. Houses, families, and the making of community: The Austrian rural case. In *The view from below: Nineteen localities*, ed. Patrick Heady and Peter Schweitzer, 109–123. Frankfurt, New York: Campus Verlag.

Sterzer, Wjatscheslaw, and Susanne Kretzer. 2019. Archivierungsstrategien anpassen: Herausforderungen und Lösungen für die Archivierung und Sekundärnutzung von ethnologischen Forschungsdaten. *Bibliothek Forschung und Praxis* 43 (1): 110–117. https://doi.org/10.1515/bfp-2019-2011.

Sterzer, Wjatscheslaw, Sabine Imeri, and Matthias Harbeck. 2018. Zugriff auf ethnologische Forschungsdaten: Anforderungen und Lösungen. Poster for the 107. Deutscher Bibliothekartag in Berlin 2018. https://nbn-resolving.org/urn:nbn:de:0290-opus4-157193.

University of Bath. 2011. ERIM Terminology V4: Terminology for Research Data Management. https://wiki.bath.ac.uk/display/ERIMterminology/ERIM%20Terminology%20V4. (last accessed 31 May 2022).

Zeitlyn, David. 2012. Anthropology in and of the Archives: Possible Futures and Contingent Pasts. Archives as Anthropological Surrogates. *Annual Review of Anthropology* 41 (1): 461–480. https://doi.org/10.1146/annurev-anthro-092611-145721.