

universität  
wien

Universitätsbibliothek



# Setting up a digital archive for ethnographic data: challenges, strategies, experiences

Wolfgang Kraus

Department of Social and Cultural Anthropology, University of Vienna  
and Ethnographic Data Archive, Vienna University Library

International Conference *Frontiers in African Digital Research*

Africa Multiple Cluster of Excellence, Section “Digital Solutions”

February 1, 2020, University of Bayreuth

# Opening and sharing data?

## The call for Open Data: definitions and assumptions

“Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings...”

<https://epsrc.ukri.org/about/standards/researchdata/scope/>

“Research Data [are defined as] Data that are descriptive of the research object, or are the object itself.”

<https://wiki.bath.ac.uk/display/ERIMterminology/ERIM%20Terminology%20V4>

## Assumption 1 (epistemology)

Data...

- reflect the real world independently of their specific research context;
- are therefore **unproblematic** to re-use;
- may, and should, serve “to reproduce and verify the results”\* of research.

\* <https://www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/>

## **Assumption 2 (economy)**

Once public money is invested to fund research...

- the data collected become assets that cannot be owned by the researcher;
- the public (including other researchers) has a right to access and re-use them.

**These sets of assumptions imply ideals of...**

- Objectivity;
- Reproducibility/replicability;
- Cost efficiency;
- Efficiency of knowledge generation.

... and a sharp discontinuity between everyday knowledge/experience and research-based knowledge.

# What are ethnographic data?

## Characteristics of ethnographic research

- Research is typically done over extended periods of time in close collaboration and exchange with research subjects.
- Research deals with personal lifeworlds, implying issues of trust and responsibility.
- Ethnography relies on an open and flexible combination of methodological tools;
- Data are multimodal and technically diverse;
- Various kinds of data are interpreted in relation to each other and to the overall research context and experience.

## Basic assumptions of ethnographic research

- Data are not found but co-constructed in dialog between the researcher and the research subjects.
- Therefore, to some extent 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.
- Being products of relation and dialog, ethnographic data are neither objective nor subjective.
- There is a gradual difference but no discontinuity between ethnographic knowledge and everyday knowledge and experience.
- Ethical considerations take precedence before considerations of efficiency.

## **Additional assumptions of our archival activities** (partially controversial)

Ethnographic data have an **intrinsic interest** beyond the **primary research context**...

- ... because they are rich in a way that is generally **not fully exploited** in the **original analysis**;
- ... because, being situated in time and space, they are **historical by nature**.
- There are good reasons to make ethnographic data **accessible and re-usable** but it would be **mistaken** to treat them as **independent** of their **original research context**.
- Since ethnographic knowledge is **embedded** in **complex corporeal experience**, specific data sets **represent the research process only incompletely**.

## **Ethnographic Data Archive (eda)**

Scientific leader: Wolfgang Kraus (Dept. of Social and Cultural Anthropology, DSCA)

Coordination: Birgit Kramreiter (Social and Cultural Anthropology Library)

Archive management: Igor Eberhard (Vienna University Library/DSCA)

Archive assistance: Jasmin Hilbert

Permanent storage in PHAIDRA, “the repository for the permanent secure storage of digital assets at the University of Vienna” (<https://phaidra.univie.ac.at>)

Pilot project (2017–19), now part of normal operations

web > <https://eda.univie.ac.at>

email > [eda.ksa@univie.ac.at](mailto:eda.ksa@univie.ac.at)

## **Starting point**

Generational change at DSCA; primary focus on historical data in analog form.

## **Main activities**

- 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;
- Medium to long term: development of a **comprehensive research data management strategy** for social and cultural anthropology and related fields.

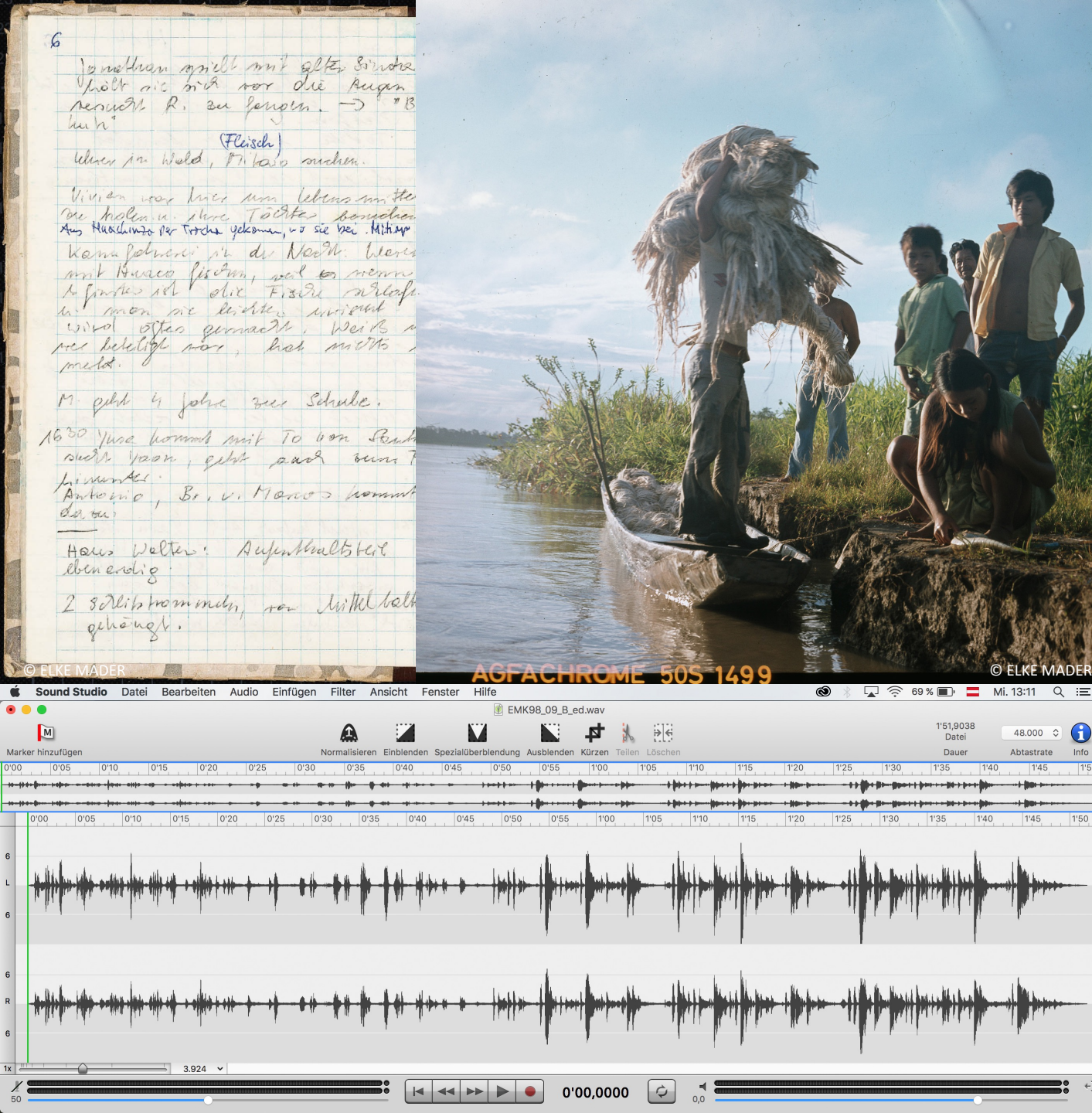
## **Main strategic considerations**

- Cooperation with researchers rather than administration of legacies;
- Comprehensive contextualisation and interlinking of data objects in research settings and researchers' biographies;
- Respect for the legitimate interest of research subjects and source communities in the data;
- Sustainability through optimised workflows, file formats, standardised procedures and metadata, quality control

## Challenges

Contradictory demands: keeping data meaningful for re-use vs. protecting research subjects

- How to adequately represent the wider research context of data objects, their interlinked nature, and the researcher's positionality?
- How to protect research subjects' privacy, interests and security?
- How to pseudonymise/anonymise without loss of context, making data next to useless?
- How to make data accessible and meaningful for their source communities?
- In the case of critical data: how to identify who is entitled to represent these communities?



## Kinds of data and data carriers

- Text (on paper and digital): e.g., notebooks, field diaries; transcripts
- Images (on film, paper and digital): e.g., photos, documents, maps
- Audio recordings (analog and digital): e.g., interviews, recitations, rituals and events, music
- Film and video: not yet part of our activities

For archived data objects see:

<https://phaidra.univie.ac.at/search#?q=eda>

## eda How To... Digitalisierung von Kompaktkassetten

Zur Formatierung: Text in *halbfett kursiv* bezieht sich auf *Hardware*, Text in halbfett auf *Software*.  
Text in *rot* bezieht sich auf *Hardware-Einstellungen* bzw. *Software-Eingaben*.

Beteiligte Geräte: Kassettendeck *Tascam 122 MKIII* >> *Babyface USB Audio Interface* >>  
*Mac Pro* Workstation; Kopfhörer (am *Babyface* angeschlossen)

1. *Mac Pro* starten (Benutzer *ksamultimedia*). Wenn *Babyface* über USB verbunden ist, dann starten automatisch die Programme *RME Totalmix* und *Fireface USB Settings*. In *Fireface USB Settings* überprüfen: *Sample Rate* = **96000 Hz**; *Clock Source* = **Internal**. In *RME Totalmix* überprüfen: *Gain* muss auf **0,00** stehen.



Sieht das *RME Totalmix*-Fenster anders aus?

Dann im Menü unter *File > Load Workspace...* **Discard all changes?** bestätigen, dann die Datei *Documents/Cassetten\_digitalisieren.tmxs* öffnen.

2. Programm *Sound Studio* starten (über das Dock am linken Rand des Bildschirms).
3. *Tascam 122 MKIII* einschalten. Einstellungen am Gerät: *Pitch Cont* = **off**; *Output* = **6** (muss nur bei sehr laut ausgesteuerten Kassetten eventuell angepasst werden, s. Punkt 9). *Dolby NR*: s. unten, Punkt 5.
4. An der Kassette mit nichtmagnetischem Werkzeug (Schraubenzieher) Tabs abbrechen (verhindert das versehentliche Neubspielen bzw. Löschen); lose Bandwickel spannen (durch Drehen mit Finger oder Stift); Kassette einlegen, einmal vollständig vor- und zurückspulen.
5. Gibt es auf der Kassette eine Angabe über die Verwendung von *Dolby NR* (Noise Reduction) bei der Aufnahme? Falls ja, am *Tascam* mit dem Schalter *Dolby NR* entsprechend (**OFF**, **B** oder **C**) einstellen. Falls nicht: Kassette wiedergeben, über Kopfhörer Höhenwiedergabe kontrollieren: besser mit oder ohne *Dolby*? (Anm.: *Dolby C* ist eher selten anzutreffen.) Relevant ist dabei weniger die Höhenwiedergabe als ob hochfrequente Geräusche und bei Sprachaufnahmen Sibilanten (d.h. „Zischlaute“ wie s und sch) natürlich und nicht überbetont klingen.

## eda Guidelines

### Step by step How To guides for digitising processes:

- Digitising compact cassettes
- Using the flatbed scanner for best quality

### In preparation:

- Digitising film with the film scanner (Hasselblad X1)
- Digitising film with the high resolution DSLR (Pentax K-1)
- Copywork with the high resolution DSLR

### eda File Name Convention for consistent naming of files (file name = object call number)

Firefox File Edit View History Bookmarks Tools Window Help

Phaidra


https://app01.oc.univie.ac.at/#/detail/o:995050

Most Visited Google Advanced Google Images Universität UBW Suche uaccess urfind Institut Webmail

universität wien phaidra.univie.ac.at

Search Submit My objects Bookmarks Groups Templates Logout

Search / Detailpage o:995050



**Identifiers**

PID <https://phaidra.univie.ac.at/o:995050>

Other

hdl:11353/10.995050

<https://phaidra.univie.ac.at/o:995050>

**Object data**

Owner

Object Type Picture

Created 11.07.2019 12:15:31

**Metadata**

[Show metadata](#)

[Dublin Core](#)

[Data Cite](#)

**Edit**

[Edit metadata](#)

[Manage object](#)

[Upload container member](#)

**Links**

[View](#)

[Download](#)

**Title (eng)** Woman churning butter, Ibudign, Imdgas, Morocco

**Photographer** Wolfgang Kraus

**Researcher** Wolfgang Kraus

**Description (eng)** Photograph taken November 11, 1983 during ethnographic fieldwork in the central High Atlas mountains in Morocco.

**Remark (deu)** eda – Ethnographisches Datenarchiv/UB Wien | eda – Ethnographic Data Archive/Vienna University Library

**Remark (eng)** Related files:

**Digitization note (eng)** Digitized by Wolfgang Kraus: negative copied with Pentax K-1 camera (Grey gamma 2.2; 16bpc)

**Language** No linguistic content | Not applicable [xxx]

**Keywords (ber)** Imazighen Imazign Ayt Hdiddu Imdgas

**Keywords (eng)** Berber

**Project (eng)** Fieldwork for doctoral thesis (Dissertation), 1983 and 1985

**Filename** WK01\_BN1983204\_18.tif

**Resource type (eng)** image

**Persistent identifier** <https://phaidra.univie.ac.at/o:995050>

**License** © All rights reserved

**Rights statement (deu)** Wolfgang Kraus

**Subject**

**Title (eng)** Woman churning butter, Ibudign, Imdgas, Morocco

**Photographer** Wolfgang Kraus

**Researcher** Wolfgang Kraus

**Reproduction note (eng)** Original

**Date created** 1983-11-11

**Provenance (deu)** Wolfgang Kraus

**Depicted/Represented place (deu)** Imdgas valley, central High Atlas

**Depicted/Represented place (eng)** Ibudign, Imdgas

**Call number** WK01\_BN1983204\_18

**Material (eng)** B&W negative (35mm)

**Technique (eng)** black-and-white photography

**Technique (eng)** Pentax MX, HP5 film/Microphen 1+1 developer (EI 1600)

Members (0)

## eda Metadata standard and templates

eda-specific metadata set and submit form, project-related metadata templates

Distinction of several object categories in metadata:

- born digital-objects,
- digital copy,
- first-order analog object,
- second-order analog object

Challenges: standardisation of metadata (localisation, names of collectivities, controlled vocabularies...)

## Other strategic considerations for the Phaidra repository

- Conceptualisation of the notion of *complex object* (now termed **container object**), consisting of several data files that represent the same analog object (e.g., a compact cassette)
- Research and testing of **best-practice file formats** (e.g., introduction of .flac format for losslessly compressed audio files)
- Development of the category of *context data* (information on the wider research context with an **authorship** and the possibility to **restrict access**)

## Archiving and sharing ethnographic data

As historically situated representations of a world in flux, ethnographic data do have a **value beyond the primary research context**.

However, the dialogic character of ethnography and the access to personal life-worlds raises important **issues of confidentiality, privacy and reciprocity**. Hence, ...

- ... the categories of **ownership of, control of, access to and licence to use data** need to be distinguished and clarified.
- ... for ethical and legal reasons, in many cases **fully open data is not an option**. This makes access management highly complex.
- ... in the future development of Phaidra, several **graded levels of access to data** must be defined and implemented.

## The bottom line

Peter Pels suggests there are several reasons “to consider social science data as indigenous or global heritage”<sup>\*</sup>—a perspective that makes preservation of data an ethical obligation.

However, long-term digital preservation of data objects—even if openly sharable—is merely a first step.

Making ethnographic data findable, accessible, interoperable and re-usable for all interested actors involves a lot more. How this can be accomplished on a large scale is not yet clear.

<sup>\*</sup> Data management in anthropology: the next phase in ethics governance? *Social Anthropology/Anthropologie Sociale* 26 (3), 2018, p. 3

Thank you for your attention!

