Long-term Archiving of Ethnographic Research Data: The *Ethnographic Data Archive* at the Vienna University Library

Wolfgang Kraus

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

UNICA Scholarly Communication Webinar "Research data management"

11 October 2022

The Ethnographic Data Archive: a grassroots initiative with a disciplinary/ methodological focus

Small but no island: archival infrastructure provided by the PHAIDRA repository

Cooperation between the Department of Social and Cultural Anthropology and the Vienna University Library

Open Science, Open Data

Promises of better science:
more transparent, more reliable, more reproducible, more replicable, more efficient, more accountable, more relevant

vs.

Scepticism in anthropology and related disciplines
A positivist model of research not appropriate for qualitative and ethnographic research

Nevertheless good reasons for archiving and sharing ethnographic 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

Assumptions 1 (epistemology)

Data…

- document the real world **independently of their specific research context**;

- are therefore largely **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.

Basic assumptions of ethnographic research

- Data are not found but co-constructed in dialogue between the researcher and the research subjects.

- Therefore, 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 dialogue, 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.

Rationale of our archival activities

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**.

- … because the **research subjects** have **their own legitimate interest** in the data and should be given access to them.

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

Starting point

Generational change at DSCA; primary focus on historical data in analog form; Cooperation with researchers rather than **administration of legacies**

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.

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 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?

Partial solutions

- Sustainability through **optimised workflows, file formats, standardised procedures and metadata, quality control**

- Comprehensive **contextualisation** in research settings and **interlinking of data objects** from the same research context

- Conceptualisation of the notion of **container object**, consisting of several data files that represent the same analog object

- Development of the category of *context data:* information with an **authorship** and the possibility to **restrict access**
  E.g., extended information on specific objects;
  information on research context;
  information on researchers and their biographies

- For the future: collaborative contextualisation with research subjects

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 in dialogue with research subjects

- ... 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.

Thank you for your attention!