

A Data Governance Framework for Ethnography v. 1.0

Alberto Corsín Jiménez
Spanish National Research Council
November 2018

Document licensed under Creative Commons CC BY 4.0

I. Aim

This document is conceived as a resource for regulators, ethics boards, funders and scholars seeking to understand how ethnographic research complies with new regulatory frameworks on data protection (GDPR) and open science.

The document performs four tasks:

1. It identifies regulatory provisos within the GDPR that accommodate for the specific requirements of ethnographic research. The document is intended to help regulators, ethics and data officers, and funders recognize why and how ethnography is compliant with data protection mandates.
2. By identifying such provisos, the document is also intended to be used by ethnographers to make their own data governance plans more robust.
3. The document describes a minimum of technical features that an open science framework (e.g. an institutional data repository) would need to include to warrant its compliance with the ethical exigencies of professional anthropological associations. It is intended to instruct regulators, ethics and data officers, and funders about the infrastructural needs for the ethical compliance of ethnographic research.
4. By identifying such technical features, the document is also intended to help ethnographers identify which data infrastructures might meet the requirements and standards of ethical praxis in their professional communities.

II. Regulatory compliance

Ethnographic research has become a staple of social-scientific research over the past fifty years. The primary methodology of ethnography is participant-observation in one or more social groups over a long period of time. In many such contexts, the prior design of research questions and data-extraction protocols defeats the very purpose of inquiry. Anthropologists have a distinguished tradition of engaging with the ethical challenges of ethnographic fieldwork, which have informed professional codes of conduct and ethics.¹ I refer readers to these codes for a proper understanding of the ethical epistemology of

¹ For example, the [*Principles of Professional Responsibility*](#) (2012) of the American Anthropological Association, and the [*Ethical Guidelines for Good Research Practice*](#) (2011) of the Association of Social Anthropologists.

ethnographic fieldwork.² What follows is not a code of ethics but a guide for regulatory compliance.

1. The open-ended nature of ethnography.

Ethnographers cannot anticipate when and how data will yield itself for observation. It is not always possible to pre-format its appearance and recollection, nor indeed its future uses.

Article 6(4); Recital 50 allows for processing data for purposes other than that for which it was originally collected, so long as the processing is compatible and lawful. Scientific research is considered to be a compatible lawful operation.

Furthermore, article 5(1)(b) deems that further processing of data for archiving purposes in the public interest, scientific or historical research purposes is not considered incompatible with the initial purposes.

Article 89 sets out the safeguards and derogations that must be in place to enable such further processing of data.

2. Ethnography is sometimes irreducible to consent.

Ethnographic participation in a social milieu lends itself to dynamics that are hardly controllable by the researcher and for which it is not always possible (indeed, it is often impossible) to obtain consent.

Recital 33 recognizes that it is not always possible to identify the purpose of personal data processing for scientific research purposes at the time of data collection, yet one should always act “in keeping with recognised ethical standards for scientific research”.

To address this challenge Article 6(4) allows for processing data where no consent has been obtained so long as its processing is deemed “compatible” (as per Recital 50)

Article 6(1)(f) also allows for processing data, without consent, for the purposes of the legitimate interests of the controller (the researcher, in this case).

The concept of “legitimate interests” is explained in Recital 47 and includes contexts “where there is a relevant and appropriate relationship between the data subject and the controller.”

In particular, Recital 47 describes a situation which accommodates to a typical ethnographic scene: “the existence of a legitimate interest would need careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place.”

3. Ethnography with vulnerable subjects

Ethnographers often carry out fieldwork in contexts strained by political conflict or violence. In these contexts, should ethnographic data fall into the wrongs hands it could

² For a recent account of the ethical challenges of anthropology in the age of data governance see Peter Pels et al., “Data Management in Anthropology: The next Phase in Ethics Governance?,” *Social Anthropology* 26, no. 3 (August 1, 2018): 391–413, <https://doi.org/10.1111/1469-8676.12526>.

have very serious consequences for some of the vulnerable people with whom ethnographers work.

The GDPR prohibits in principle the processing of “special categories of data”, such as data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, among others.

However, the prohibition is qualified under Article 9(2) for various circumstances, including for scientific purposes (Article 9(2)(j)).

Repurposing or reprocessing sensitive data is allowed under Article 6(4).

There is therefore discretionary room within the GDPR for ethnographers to process sensitive data, or not, as they see fit.

4. Working in/with third countries

The GDPR prohibits in principle the transfer of data to countries outside the EU, unless the latter have adequate levels of protection and regulations in place. This could pose a problem for anthropologists, who have a long tradition of doing fieldwork outside their home countries, not infrequently in the global South.

However, under Article 49(1), a controller may transfer data to a third country when “necessary for the purposes of compelling legitimate interests pursued by the controller which are not overridden by the interests or rights and freedoms of the data subject.”

Recital 113 allows for the definition of “compelling legitimate interests” to include “the legitimate expectations of society for an increase of knowledge”.

5. Data retention and storage limitation

The GDPR states that personal data should be kept for no longer than is necessary for the purposes for which it was processed (Article 5(1)(e)).

However, in accordance with Article 89(1) data may be stored for longer periods (indefinitely, even) if processed for scientific purposes.

III. Data archiving and management

The contribution of anthropology to archival theory and practice is long-standing and undisputed, in natural history museums or ethnological collections, but also in the very legacy of ethnography as an archival field. Anthropology has further played an important role in recent efforts at redefining the postcolonial nature of the archive, for example by thinking through the challenges of repatriation, intellectual property rights or the renegotiation of protocols of access and display in collections.

More recently, anthropologists have been at the vanguard of the development of digital infrastructures for data archiving. In collaboration with indigenous communities, data scientists and software engineers, ethnographers have developed sophisticated digital archives whose designs incorporate technical features modelled after the complex ethical demands of ethnographic fieldwork.

I offer below a survey of some of the conceptual and system requirements that anthropologists have designed to address the challenges of digital data management.

The description is offered also as guidance for policy-makers, data officers and regulators regarding the minimum requirements that institutional repositories should include to warrant the ethical management of ethnographic data.

1. Ownership vs. collaboration

Ethnographic data is rarely owned by researchers. Much ethnographic data is produced collaboratively and remains ‘alive’ beyond the moment of collection.

Digital infrastructures should allow for data to be remediated by such ongoing collaborative and processual relations, for example, by incorporating features for version control, reuse, commentary, annotation or edit, in different degrees, forms and formats by different parties.

[*Digital Himalaya*](#) is a digital humanities project and collection for multimedia anthropology on the Himalayan region which has developed various strategies for remediation.³

[*Ciudad Escuela*](#) is a free software digital infrastructure for urban anthropology and a meta-archive of urban educational projects, where participants can contribute content, design educational itineraries, as well as modulate anew the site’s navigational features into systems of shared apprenticeships.⁴

2. Access and rights

It is not uncommon for ethnographers to work with indigenous communities whose systems of knowledge and cultural materials do not conform, and cannot be translated into, Euro-American regimes of intellectual property rights, and whose management requires specific protocols of access, use, and legibility.

Digital platforms should therefore allow for the registry and application of different legal licenses and labels to ethnographic materials.

[*Mukurtu*](#) is a digital heritage project specifically designed to enable indigenous communities share knowledge, stories, and cultural materials using their own protocols. Moreover, the project has developed a set of Traditional Knowledge Labels that can be used to qualify protocols of access and use to material already in the public domain or when a third party holds copyright.⁵

The [*Plateu People’s Web Portal*](#) is a collaboratively curated and reciprocally managed archive of Plateu cultural materials, which has enabled Plateu people to add their expert

³ Mark Turin, “Born Archival: The Ebb and Flow of Digital Documents from the Field,” *History and Anthropology* 22, no. 4 (December 1, 2011): 445–60, <https://doi.org/10.1080/02757206.2011.626776>.

⁴ Alberto Corsín Jiménez and Adolfo Estalella, “Ethnography: A Prototype,” *Ethnos* 82, no. 5 (October 20, 2017): 846–66, <https://doi.org/10.1080/00141844.2015.1133688>.

⁵ Kimberly A. Christen, “Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness,” *International Journal of Communication* 6, no. 0 (November 30, 2012): 24.

voices to public collections as well as incorporate traditional cultural protocols for the viewing, circulation, and reproduction of materials.⁶

3. Privacy and permissions

Digital infrastructures should allow for defining and specifying different roles and levels of access within a system, for example: researcher, administrator, contributor, guest, collaborator, etc. It should also be possible to control the availability or appearance of different layers or bundles of content, for example by making certain types of content open, restricted, private, editable, shareable, etc.

The fluxogram below exemplifies how to translate privacy and anonymity into a system of permissions:

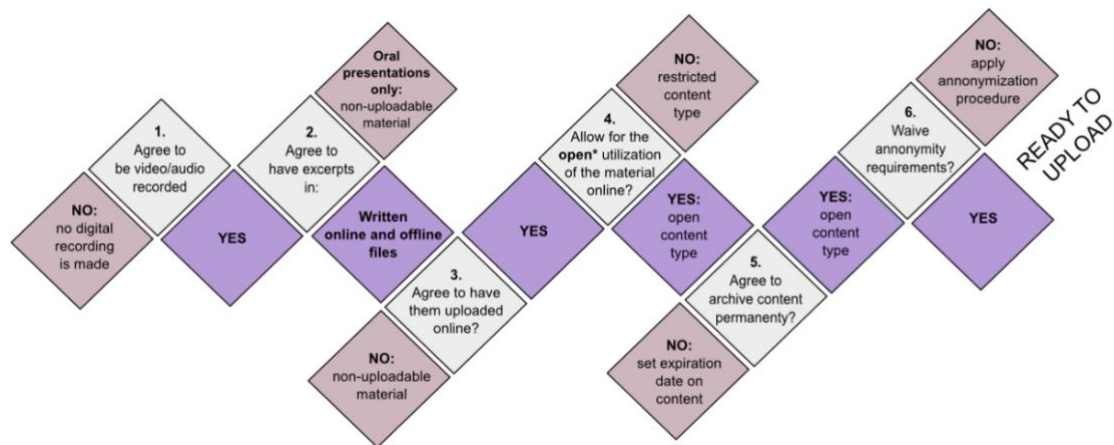


Figure 1. Consent questions transposed into data management permissions.⁷

4. Sustainability and experimentation

Ethnographic data is sometimes bestowed to researchers in recognition of their own role as custodians or stewards of a community’s cultural heritage. The long term preservation of this material is therefore of paramount importance, and researchers should only deposit it in a digital archive if there are safeguards and guarantees in regard of the infrastructure’s long-term socio-technical sustainability.⁸

The socio-technical sustainability of a digital project is of importance not just for its preservation, legacy, or contribution to public interest, but for its capacity to cultivate and promote scholarly experimentation also. A robust, sustainable platform allows scholars and collaborators the freedom to experiment with new forms of curating, publishing and networking content.

⁶ Kimberly Christen, “Opening Archives: Respectful Repatriation,” *The American Archivist* 74, no. 1 (April 1, 2011): 185–210, <https://doi.org/10.17723/aarc.74.1.4233nv6nv6428521>.

⁷ From Luis Felipe Rosado Murillo, “What Does ‘Open Data’ Mean for Ethnographic Research?,” *American Anthropologist* 120, no. 3 (September 1, 2018): 580, <https://doi.org/10.1111/aman.13088>.

⁸ See the Socio-Technical Sustainability Roadmap, <https://sites.haa.pitt.edu/sustainabilityroadmap/>

Experimentation is facilitated if infrastructures are designed using free/libre or open source software, and contents are freely licensed with Creative Commons licenses (subject to the caveats noted above).

The *[Platform for Experimental, Collaborative Ethnography](#)* is an open-source digital platform that supports multi-sited, cross-scale ethnographic and historical research. The platform is designed to enable scholars working within the empirical humanities to design and develop infrastructures for collaborative hermeneutics in their fields.⁹ The platform itself was designed as an experimental re-functioning of a previous collaborative archival project, *[The Asthma Files](#)*, whose source code was repurposed to enable a plurality of collaborative analytics to inhabit and cross-fertilize the same architecture.¹⁰

Data Governance Framework for Ethnography
v. 1.0
November 2018

⁹ Kim Fortun et al., “Hosting the Platform for Experimental, Collaborative Ethnography” (forthcoming), http://worldpece.org/sites/default/files/artifacts/media/pdf/kfortun_pece_for_hau_3.4.2018.pdf.

¹⁰ Kim Fortun et al., “Experimental Ethnography Online,” *Cultural Studies* 28, no. 4 (July 4, 2014): 632–42, <https://doi.org/10.1080/09502386.2014.888923>.