



Digital anthropology as an alternative to “big data”  
analysis for decision support. The case  
of health care

Author(s): Alberto Cottica, Marco  
Manca, Ivan Cukerić

March 2023

SHS/MOST/DA/2023/PI/5

Referred to this call: <https://liivcenter.org/wp-content/uploads/2022/07/Call-for-Evidence-ext.pdf>

## Title

Digital anthropology as an alternative to “big data” analysis for decision support. The case of health care.

## Abstract

In health care, the European digital transition mandates the creation of a shared, federated data space. With each interaction with the health care system, patients and caregivers leave a trail of data, whose analysis informs the design and provision of services. While the technology to do this is new, the idea is coherent with the 20th century trend of governance as a superimposition on reality of an administrative order, evoked in James Scott’s “Seeing Like a State” [1]. As a result, care service design optimizes not for individual patients, but for statistical abstractions (“the average diabetic”). This flies in the face of the inherent complexity of human physiology and of the relationships of care, and introduces distortions and epistemological errors. The latter generalize beyond health care, to fields such as environmental or inequality-reducing policies.

We argue that digital ethnographic methods could mitigate the problem by offering a methodological approach capable of embracing diversity. Recent advances allow for analysis of human conversations at scale as data. We illustrate the point with data from a study on community provision of health and social care, and propose an argument for the epistemological superiority of digital anthropology over big data analysis.

## Background Literature

In COM(2022) 197/2 the Directorate-General for Health and Food Safety of the European Commission proposed a regulation to set up a European Health Data Space. The initiative is rooted in a study assessing EU Member States’ rules on health data in light of the GDPR [2], which proposed that uneven implementation and interpretation of the GDPR by Member States creates considerable legal uncertainties, resulting in barriers to secondary use of electronic health data that hinder natural persons access to innovative treatments and prevents policy-makers from effectively reacting to a health crisis, due to barriers impeding access for researchers, innovators, regulators and policy makers to necessary electronic health data. Despite other benefits being mentioned for natural persons, mostly focusing around access to data, the proposal is at large geared towards creating a EU wide data space from where IP generating initiatives and policy makers could tap high quality data to draw maps of adjacent possibilities. Approaches that focus on “high quality data” rest upon an objectivistic rhetoric but fail to deliver in reality, because data are known to be not representational, but relational. In other words, they carry the world view and biases of those who have the power to create them [3]. Additionally, they encounter a fundamental limitation that has only recently been acknowledged: the challenge of ergodicity (*the property of a system whose time averages are equal to the ensemble averages*) breaking in human datasets [4] and the very delusion of objectivity [4], that is radical enough to touch on the act of collecting/generating data [6].

## Digital ethnography as a path to an inclusive governance of the Health Data Space

Against this backdrop, we propose that digital ethnography might be explored as a way to make the governance of the proposed Health Data Space more inclusive and effective. Ethnography evolved to describe cultures; it is well-suited to the task of surfacing hidden assumptions and biases. The digital element serves two purposes. The first one is to enable analysis of relatively large amounts of qualitative data (*corpora* in the language of ethnography). The second one is to improve the accountability of the research process itself. Specifically, we build on prior research by our group in the field of digital ethnography [7] and propose that digital ethnographic methods might and should be adopted within a health data space to realize an OODA (*Observe–Orient–Decide–Act*, in military discipline, the description of how a human navigates an evolving landscape) [8] or PDSA (*Plan-Do-Study-Act*, the OODA translation to management theory) [9] system wide cycle.

The reasoning behind this proposal is as follows. Start by imagining a public forum where patients and other users of the health care system discuss their interaction with the system itself, giving each other advice and support, or simply venting frustration, or expressing happiness and gratitude. We argue that **people on such a forum constitute an *issue public***. This concept harkens back to the political philosopher John Dewey. He believed that, in a democracy, competent publics emerge in response to issues by which their members are affected, and where expertise is contested and established frames for problem-solving break down. Members of issue publics then set out on a quest to educate each other about what this issue really is, and how it should be framed [10]. This makes participating in such a forum into a sort of public consultation, and the analysis of such participation into an inquiry into the results of that consultation.

Furthermore, **the interactions occurring in such a forum constitute a corpus, which can be studied with a digital ethnographic or netnographic approach** [11]. The discipline has evolved quali-quantitative mixed methods to take advantage of the digital nature of this form of qualitative data in order to be relatively robust to the challenges of analyzing corpora with thousands of informants and tens of thousands of contributions (like forum posts). We ourselves have been working on a method based on the representation of ethnographic data as networks of co-occurrences between codes [7].

**The analysis of the corpus described above has a greater degree of accountability than standard ethnographic analysis.** This is because the data are structured and in digital form (which means they come with metadata describing how, when and by whom they were collected); and because their analysis is partially algorithmic (for example, inducing a co-occurrence graph and calculating that graph's statistics are easily verifiable steps). Data science – especially, but not only, when data represent aspects of human behavior – is undergoing a replicability crisis. Apparently innocent choices in data cleanup or model specification can lead different researchers to divergent results, even when controlling for known biases like confirmation bias or the researcher's skill levels [12]–[14]. This situation breeds distrust in scientific findings; accountability of research can go some way towards restoring trust.

All this, of course, does not make the ethnography we propose replicable. But then, **ethnographic research is not expected to be replicable, and because of this it has developed a culture of epistemological humility**; this includes the refusal to claim objectivity; the transparency around the researchers' biases and assumptions, encoded in positionality statements; and the tradition of involving informants in the analysis of the data they provided, that harkens back to Jean Rouch's shared ethnography [15]. We argue that this goes some way towards remediating the opacity of big data study.

To recapitulate, we propose an ethnographic study, conducted by means of quali-quantitative methods on a corpus that consists of the interactions on an online forum where the Deweyan issue public consisting of patients and practitioners discuss the provision of health care services. **We claim it can provide trustable data analysis by better accountability than traditional ethnography, and more epistemological humility than big data analysis.**

An obstacle to our program is that ethnography does not have a tradition of open science practices. Publication of data is rare; publication of FAIR data even more so [16]. Data analysis remains a black box in most ethnographic studies. Despite ethnography's participatory tradition, we are not aware of any study that meets the standards of what Haklay calls "extreme citizen science" or "participatory science". These are studies where non-professional scientists participate in data analysis with professional researchers, using a common platform of data, software for analysis, and debate [17].

All this means that **ethnographic research could be made even more accountable and trustable by leveraging the digital nature of the mixed methods we advocate to embrace open and citizen science practices**. A "citizen ethnography" of the Health Data Space could contribute to leveraging the collective intelligence of patients and practitioners in a way that we all can trust, because (1) the data are available and open, (2) the progression from data through analysis to conclusions can be traced using data and simple software for analysis, and (3) health care's issue publics can challenge the conclusions of professional researchers, applying their own categories (via ethnographic coding) to the same data to draw their own conclusions.

## Discussion

The 19th century's Industrial and scientific acceleration emerged as the result of a number of converging elements whose root cause can be traced to the establishment of new practices of sharing artifacts (like publications and patents), and organizing inquiry and knowledge transfer around those artifacts [18]. We suggest that the introduction of digital ethnographic practices would produce a similar cascade and evolution in healthcare management and digitalization by enabling the emergence of competing narrative, not just around the effectiveness and efficiency of health care, but about its underpinning values, strategic goals, and key concepts (for example, what it means to be "healthy" depends on who you ask, and should be object of debate). The roadmap we propose enables health care leaders to make decisions and run evaluations by organizing conversations and sensemaking processes around artifacts (datasets, analytics, visualizations, causal models, counterfactuals, ...) which (1) encode the world views of patients and (2) by virtue of being open (in the sense of being FAIR data, and of the techniques for data analysis being relatively simple and intuitive instead of highly

mathematical), enable and encourage a plurality of agents to use them. Patient groups, for example, would then be empowered not only to bring their experiences and reflections (by individual contributions to the forum), but also to provide their own accounts of what those experiences and reflections actually mean. In the language of citizen science, they would be not only data collectors, but participants in research design and data analysis.

In addition to its advantages in openness and contestability, digital ethnography is free from the ergodicity fallacy that affects much statistical analysis. It focuses on intersubjective meaning rather than supposedly (but not really) objective general models; thus, it carries the voices of many individual humans that traverse the health care system, as opposed to representing “the average patient with pathology X”.

## Conclusion

Digital ethnography applied to digital healthcare offers tools to resolve counterfactual reasoning and negotiations by embracing the biases and hypotheses behind their emergence. In so doing, it allows a more inclusive, accountable, epistemically diverse discourse on health care. It is reasonable to expect that such discourse fosters strategies which, in turn, are themselves more inclusive and accountable.

## References

- [1] J. C. Scott, *Seeing like a state: how certain schemes to improve the human condition have failed*. New Haven: Yale University Press, 1998.
- [2] Consumers, Health, Agriculture and Food Executive Agency., *Assessment of the EU Member States' rules on health data in the light of GDPR*. LU: Publications Office, 2021. Accessed: Sep. 06, 2022. [Online]. Available: <https://data.europa.eu/doi/10.2818/546193>
- [3] S. Leonelli and A. Beaulieu, *A critical introduction to data and society*, 1st ed. Thousand Oaks: Sage Publications Ltd, 2021.
- [4] A. J. Fisher, J. D. Medaglia, and B. F. Jeronimus, “Lack of group-to-individual generalizability is a threat to human subjects research,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, no. 27, Jul. 2018, doi: 10.1073/pnas.1711978115.
- [5] A. Birhane and D. J. T. Sumpter, “The games we play: critical complexity improves machine learning.” arXiv, May 18, 2022. Accessed: Sep. 06, 2022. [Online]. Available: <http://arxiv.org/abs/2205.08922>
- [6] D. Balazka and D. Rodighiero, “Big Data and the Little Big Bang: An Epistemological (R)evolution,” *Front. Big Data*, vol. 3, p. 31, Sep. 2020, doi: 10.3389/fdata.2020.00031.
- [7] A. Cottica, A. Hassoun, M. Manca, J. Vallet, and G. Melançon, “Semantic Social Networks: A Mixed Methods Approach to Digital Ethnography,” *Field Methods*, vol. 32, no. 3, pp. 274–290, Aug. 2020, doi: 10.1177/1525822X20908236.
- [8] J. Boyd, *Destruction and creation*. US Army Comand and General Staff College, 1976.
- [9] R. Aguayo and W. E. Deming, *Dr. Deming: the American who taught the Japanese about quality*, 1. Fireside ed., [6. Dr.]. New York, NY: Simon & Schuster, 1991.
- [10] N. Marres, “Issues spark a public into being. A key but often forgotten point of the Lippmann-Dewey debate,” in *Making things public: atmospheres of democracy*, MIT Press, 2005.
- [11] L. Costello, M.-L. McDermott, and R. Wallace, “Netnography: Range of Practices, Misperceptions, and Missed Opportunities,” *International Journal of Qualitative Methods*, vol. 16, no. 1, p. 160940691770064, Dec. 2017, doi: 10.1177/1609406917700647.

- [12] A. Decuyper, A. Browet, V. Traag, V. D. Blondel, and J.-C. Delvenne, "Clean up or mess up: the effect of sampling biases on measurements of degree distributions in mobile phone datasets." arXiv, Sep. 29, 2016. Accessed: Sep. 06, 2022. [Online]. Available: <http://arxiv.org/abs/1609.09413>
- [13] R. Silberzahn *et al.*, "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results," *Advances in Methods and Practices in Psychological Science*, vol. 1, no. 3, pp. 337–356, Sep. 2018, doi: 10.1177/2515245917747646.
- [14] N. Breznau *et al.*, "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty," MetaArXiv, preprint, Mar. 2021. doi: 10.31222/osf.io/cd5j9.
- [15] A. M. Jørgensen, "Filmmaking as Ethnographic Dialogues: Rouch's Family of 'Scoundrels' in Niger," *Visual Anthropology*, vol. 20, no. 1, pp. 57–73, Jan. 2007, doi: 10.1080/08949460601061701.
- [16] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016, doi: 10.1038/sdata.2016.18.
- [17] M. Haklay, "Citizen science and volunteered geographic information: Overview and typology of participation.," *Crowdsourcing geographic knowledge*, pp. 105--122, 2013.
- [18] U. Wengenroth, "Science, technology, and industry in the 19th century," *Munich Center for the History of Science and Technology Working Paper*. [http://www.mzwtg.mwn.tum.de/fileadmin/w00bmt/www/Arbeitspapiere/Wengenroth\\_scitech-ind-19c.pdf](http://www.mzwtg.mwn.tum.de/fileadmin/w00bmt/www/Arbeitspapiere/Wengenroth_scitech-ind-19c.pdf), 2000.