

Natalia Rozalia Avlona (2023): In: Proceedings of the 21st European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies – Doctoral Colloquium, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.48340/ecscw2023-dc08

Copyright 2023 held by Authors, DOI: 10.48340/ecscw2023_dc08

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, contact the Authors.

Natalia Rozalia Avlona

Department of Computer Science, University of Copenhagen, Marie Curie Fellow, DCODE Network

naav@di.ku.dk

PhD Supervisor: Prof. Irina Shklovski

Department of Computer Science, University of Copenhagen, ias@di.ku.dk

PhD short description

This PhD project aims to explore how economic value is generated by data in data-driven technologies and the role of contextual and informal social relations in these valuation processes. Conducting ethnography in existing socio-technical networks ‘in the wild’, this PhD situates its focus on the healthcare sector. This project aims is to explore how the existing data practices, from the micro level of the data- practitioners to the mezzo level of the data stewards, construct value propositions for the ways data value is traded and regulated by the policymakers. Ethnographic fieldwork for this PhD research has already been conducted in two northern European companies in the Health Tech Industry. Further research is planned to be conducted in the public and private healthcare sectors in UK, Denmark and Greece.

Bellow, please do find the articulation of the PhD research in three studies. Study 1 and Study 2 have already been conducted and are about to be published. Study 3 is partly ethnographically conducted.

Study 1

The Logics of Data Quality: Challenges in Creating High Quality Data for Algorithmic Systems in Healthcare

Data quality is an important consideration in the development and deployment of algorithmic systems in general but especially in the healthcare domain. However, the process of achieving appropriate data quality is a source of active debate. CSCW researchers have debated challenges of data work, data annotation and data quality, proposing a range of different considerations, approaches, and points of contention. In this paper we present findings from an interview study of data creation approaches in two health technology organizations. Taking the traditional

dimensions of data quality - accuracy, completeness, consistency, and timeliness - we investigate how these dimensions are contested and achieved. We find that each of these dimensions encapsulates a logic of practice and contestation to arrive at something that can be termed a quality dataset. For each logic we analyse its imagined ideal features and the implications of failing to achieve these, and then consider the contestations and paradoxes emerging from the intended practices to achieve it. As the goal of the process was the creation of training data, we highlight the implications and the dependencies amongst the processes of different logics as our participants describe the often frustrating, laborious, contested process of creating what one participant called “a well-orchestrated ground truth” with all its limitations

Study 2

Design of Ground Truth Schemas for Medical AI - Who Decides What and How to Label On Medical Data?

**In collaboration with Hubert Zajac,*

In this study we answer the research question "What factors influence the design of ground truth schemas for medical AI datasets and how". In this paper, we articulate the factors that crucially mediated the design process of the Ground Truth Schema towards the design of responsible AI, in our three studies. We ground our contributions in ethnographic findings from three projects that use AI in the medical domain: (I) screening chest x-rays, (II) AI-powered diagnostic tools for lung and pancreatic diseases (III) AI-powered platform for automatic patients-to-clinical trials matchmaking. We explore the negotiations, tensions, and compromises made by medical professionals, data scientists, and designers on their quest to create suitable medical AI datasets in highly constrained environments. The data includes approximately 50 hours of observations, 37 interviews, with medical professionals, data scientists and ML engineers, designers, as well as observation notes, email communication, reports, and artefacts. We followed the Grounded Theory recommendations to construct the five factors influencing the design of Ground Truth Schemas [Charmaz 2014].

Through this research, we first uncover the factors that affect the design decisions which further shape the medical AI datasets, even before their creation. We identify these factors, as internal to the labelling process and external ones.

Study 3¹

From data care to data stewardship in the Healthcare Sector: Care-ful enactments and deviations

Over the last years, the increasing digitisation of healthcare services has transformed the provision of patient-centric care into data-centric healthcare (Gotz et al 2016; Zahid et al, 2021). Particularly, the employment of AI-powered systems for health service management and diagnostic decision-making in the public and private healthcare sector has mutated patient-centrism, as a human-centred care practice and value, to a data-centric one (Sunarti et al, 2021). Within this context, scholars have critically addressed issues emerging through this increasing datafication and automation of healthcare provision. One strand of critical research has shown how automation has asymmetrically affected the often invisible and devalued data work in the public health sector (Bossen, Pine et al 2019; Moller et al 2020). Scholars engaged with the notion of care from a Feminist STS perspective have problematised how these data-centric practices in caregiving signalled “technosolutionism” in the ways the care for health data is enacted (Kaziunas et al, 2017; Murphy, 2015). In this complex territory of the healthcare datafication and AI-fication, data stewardship, as this process of “taking care” of the appropriate and legally compliant provision, use and reuse of medical data, seems to acquire a renewed importance (Bukowski et al, 2019). Nevertheless, limited research has been conducted by Critical Data Scholars in order to conceptualise the role of data stewardship in this context.

In this paper, we investigate how data-centric care is enacted from the standpoints of data practitioners who are engaged in two different processes of data handling: **a.** The experts who are engaged with the “data creation” for the design of AI-powered systems for patients and clinical trials match-making, and AI-powered diagnostic tools, **b.** The data stewards who are responsible for the legally compliant provision, use and reuse of the medical data. To do so, we analyse the findings of multi-sited ethnographic research in two Northern European Health Tech Companies developing AI-powered systems and three hospitals in the UK and Greece. The objective of this paper is twofold. **Firstly**, we articulate the enactments of data-centric care by the experts in the particular context of our studies by looking at how the values of each domain of expertise inform the data care practices. **Secondly**, we show how data stewardship, as a matter of processual “care for compliance” is differentiated from the situated data care enactments and what conceptual implications this bears for the critical data studies.

¹ This study has been partly conducted. The fieldwork in UK and Greece is planned to be conducted during Spring and Summer 2023.

Research Questions

RQ1: How is data produced in specific contexts and for specific purposes?

RQ2: Which are the processes that transform “quality data” into “valuable” data?

Research Objectives

1. Unpack the nitty-gritty processes of “data creation”/“data production” at the micro-level of domain expertise.

2. Articulate the dependencies and contestations of these practices.

3. Translate the implications of these practices to the ways that data value is perceived and traded at the mezzo level and macro level of the Industry and Policymakers.

Methodology

This PhD research project employs multi-sited ethnographic research as its core methods of choice. Ethnography has already been conducted in two Health Tech companies in order to gain a better understanding of how practitioners in the health tech industry think and practice health data. In total 26 experts were interviewed, with an average of 65 minutes in length, whilst we conducted follow-up interviews with two of the experts. Further ethnographic research is planned to be conducted in four public and private hospitals in UK, Denmark and Greece.

Short Bio

Natalia-Rozalia Avlona is a lawyer, researcher and Marie Curie PhD Fellow (DCODE) the Computer Science Department of the University of Copenhagen. Her research focuses on unworking data as a concept and practice in the Data-Driven Healthcare Sector. She studied law at the School of Law of the National and Kapodistrian University of Athens (2006), obtained her master’s degree in Human Rights Law (LLM) from King’s College London (2007), and followed courses in the department of Geography at Royal Holloway, University of London and in the department of Curating Contemporary Art at the Royal College of Art.

Her expertise is on the intersection of open and emerging technologies with law and society, through a feminist framework. Furthermore, her wider interests have led her into working on the legal and ethical consequences of the emerging technologies.

Natalia has international experience working in several Organizations and European Research Programmes in UK, Belgium, Greece and Cyprus. Among those are the Future

Emerging Technologies Department (DG Connect, EU Commission), the Aristotle University of Thessaloniki, the Organisation of Industrial Property in Greece, the Royal College of Art in London, Abandon Normal Devices in Manchester, the University of Nicosia in Cyprus, the Heinrich Boell Stiftung in Greece, the General Secretariat for Gender Equality and the GUnet (Greek Universities Network) in Athens. Before joining the University of Copenhagen as Marie Curie Fellow, she was a Research Fellow at ELIAMEP (Hellenic Foundation for European & Foreign Policy) in Athens, where she led the AMIF ATHENA Research programme, and has worked as a Researcher on the Horizon2020 TARGET programme (Taking a Reflexive Approach to Gender Equality for Institutional Transformation) since 2019.

She is currently member of the Management Committee of the CA21118 Cost Action Platform Work Inclusion Living Lab (P-WILL) .She has also been smember of the Management Committee of the CA16121-From Sharing to Caring: Examining Socio-Technical Aspects of the Collaborative Economy (2017-2021). She is also the co-founder of the techno-feminist hacking network Restorative Infrastructures.

Besides her research path, Natalia has a strong involvement and commitment to feminist ethics as an activist in the field of commons, and Social and Solidarity Economy. She has run a series of workshops on Wikipedia for the Galleries, Libraries, Archives and Museums (GLAM) sector, co-organized feminist workshops on FOSS at hackerspace.gr (Django Girls Athens), Techno-feminist Festivals (/EtcAthens) and (un)conferences on the commons, and has given seminars on Gender & Open Technologies, Gender & Social and Solidarity Economy and co-organized workshops on Everyday Feminisms and Care for Degrowth Strategies. During the covid-19 pandemic, she has co-initiated the Emergency Making Aid, a local, bottom-up initiative of makers, architects, researchers that aimed at the making and donation of 3d -printed protective equipment to the medical personnel of the hospitals.

Statement for Feedback

It would be extremely useful to receive feedback on the following issues:

1. How to conceptualize these different studies within a stronger theoretical framework.
2. If the rest of the ethnographic fieldwork makes sense to be solely conducted in the public health care sector, or it should be comparatively done in the private versus the public hospitals.
3. How this research can be more organically situated in the CSCW community and scholarship.