Reimagining Public Values in Algorithmic Futures

A Data Management Plan created using DMPonline.be

Creator: Maiju Tanninen

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / **URL:** 3H220639

ID: 198416

Start date: 01-11-2022

End date: 31-10-2025

Project abstract:

This study analyses the role of public values, such as solidarity, autonomy and trust, within algorithmic systems and automated decision-making (ADM). Recent scholarly work has argued that developments around algorithmic systems have led to negative social impacts that are undermining democratic processes and strengthening inequalities. The policy responses to algorithmic systems typically invoke values such as fairness, accountability and transparency. These values, however, are often only understood in narrow, technical senses. Moreover, it is not self-evident how such values can be programmed into algorithmic systems. This project mobilizes a sociological perspective on value and examines how values are deployed in situated practices rather than in the abstract, suggesting ways forward in thinking about algorithmic futures.
The study observes the people behind algorithms: what do people do when they build, promote and evaluate algorithmic systems? Specifically, we analyse how values are at stake in the developing of algorithmic systems in two distinct fields: private (re)insurance industry and in silico medicine. Although both fields are global industries, we focus mainly on the practices happening in the European regulatory and market space.
New data types, algorithmic technologies and computational methods are believed to transform the practices of both the insurance industry and medical research. (Re)insures have started experimenting with data-driven technologies and investing in them heavily, hoping that the new technologies will transform risk prediction, prevention and pricing (Tanninen 2022). In in silico medicine, the vision is that computational techniques could revolutionize the diagnosis and treatment of diseases by modelling human biological processes in a virtual environment. The algorithmic operations happening in these different fields are, thus, driven by the belief that increased automation will create optimized, personalized and more efficient operations. However, the operations might have (unintended) negative effects; they might, for instance, reinforce existing inequalities by using biased data (in silico medicine) or by excluding people deemed ‘high risk’ from coverage (insurance).
Drawing from these exemplary cases, the study examines emerging technological practices that have promises to change the operational logics of their fields completely. We are especially interested in how values and valuations are at stake in these experimental operations and examine the co-constitution of data-driven operations and central values: What is deemed valuable in these practices? What gets lost in the process? Furthermore, we analyse the frictions and balancing acts between the ‘automation’ values (efficiency, personalization, optimization) and the more traditional public values of the fields: solidarity and trust in insurance; care and public good in medical research. Finally, by following this practice-oriented approach, we explore ways to think and talk about algorithmic systems beyond the optimization rhetoric, analysing how algorithmic technologies could help strengthen the more traditional public values, such as equality and solidarity.

Last modified: 06-04-2023

Reimagining Public Values in Algorithmic Futures

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
In silico focus group discussions	Transcriptions of focus group discussion with medical professionals and people working in in silico medicine.	N	D	T	.doc	<100GB	
In silico interviews	Interviews with people working with in silico medicine or other involved actors, such as policymakers. Original soundfiles (will be destroyed when no longer needed) and transcriptions.	N	D	S T	WAV .doc	<100GB	
Insurance interviews	Interviews with people who have a relevant connection to the (re)insurance industry or tech industry, for instance, insurance professionals, tech professionals, policymakers. Original soundfiles (will be destroyed when no longer needed) and transcriptions.	N	D	S T	WAV .doc	<100GB	
Observation data	Fieldnotes from observing in silico and insurance professionals in professional settings, such as business conferences and seminars.	N	D	T	.doc	<100GB	
Secondary data	Interviews with Finnish insurance professionals and focus group discussions with Finnish policyholders.	E	D	T	.doc	<100GB	
Qualitative analysis via Atlas.ti	Atlas.ti project for coding and analysing interviews and focus group data.	N	D	T	.atlproj	<100GB	
Insurance document data	Freely available reports, articles and publications by (re)insurance comapanies.	N	D	T	pdf jpg	<100GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Pseudonymized Finnish language transcriptions of 16 interviews with insurance professionals.
Pseudonymized Finnish language transcriptions of 11 focus group discussions with insurance customers.
These data were collected by researcher Maiju Tanninen in a previous research project.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

SMEC G-2023-6260, approval is not finished yet.

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

G-2023-6260
The following data sets will include forms of personal data:
In silico focus groups
In silico interviews
Insurance interviews
Secondary data

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

The use of the secondary data is restricted by research collaboration agreement between Maiju Tanninen (MT) and the participating insurance companies to only MT.
The use of the in silico focus group data is restricted to the researchers of the 'In Silico World' project (project number 3M210137) where researcher Elisa Elhadj works besides this project.

The in silico and insurance interviews are still under negotiations. Given that the research concentrates on the professional sphere and on the practices of developing new digital products, it is likely that in order to get an access to the field, we need to make agreements that restrict the dissemination of original data only to the researchers.
The insurance document data is already freely available online. We will make the specific document set used in this research accessible.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

For the interview and focus group transcriptions, the documentation (a Word document) will include description of the setting (location, and modality, e.g. in person or online), the informed consent process, the subjects discussed in the interview/focus group and the instructions given to interviewers/moderators. It will also include the topic lists of the interviews/focus groups and the process through which direct identifiers have been removed from the data.
For the observation notes, the documentation will include when and where the data was collected, the main themes of the observation data and how consent was achieved/the ethical justification for making observations (e.g. public event).
For the document data, the documentation will include details of the company, date of publication, date of collection and brief description of the contents.
For the Atlas.ti project, the documentation will include the codebook (names and meanings of the codes and their relation to one another (code tree).

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

A metadata file will be created using MS Excel. The metadata from Atlas.ti qualitative analysis and coding will be incorporated. The following metadata will be registered:
Title (name given to the resource)
Subject (the topic of the resource)
Description (an account of the resource)
Creator (the actor primarily responsible for making the resource)
Respondent id code (anonymized respondent id, the key will be stored in another secure location available only to the researcher)
Date (date of creation and last modification)
Type (the genre of the resource)
Format (the file format)
Identifier (An unambigous reference to the resource in a given context, code of the data file that can be linked to another data)
Source (A related resource from which the resource is derived, e.g. sound file to transcription file)
Language
Rights (information about the rights held in and over the resource)

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)

The data will be stored on the KU Leuven OneDrive, that will be protected by multifactor identification.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

The data stored in KU Leuven Onedrive will be automatically backed up in the cloud.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The KU Leuven OneDrive uses multifactor identification procedure that effectively secures the data.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The data storage will not generate extra costs.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

The pseudonymized transcriptions and document data will be stored according to the KU Leuven RDM policy.

The original interview/focus group audio files will be destroyed after transcriptions are produces and they are no longer needed for analysis.

Where will these data be archived (stored and curated for the long-term)?

- Other (specify below)

The data that will be stored in a secured KU Leuven OneDrive folder. Access to the data will depend on data agreements with the research participants.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

There will be no additional costs. OneDrive is part of the Microsoft 365 Education A3 plan. The cost of the Microsoft 365 Education A3 licenses within the EES agreement is financed centrally for all KU Leuven students and the majority of active KU Leuven staff.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Other (specify below)

The availability of data depends on the data type.
Depending on the data agreements between the participating actors/companies, the focus group data and the interview data can have a closed access. This is likely the trade off with being able to enter to the fields of in silico medicine and (re)insurance.
Data sets consisting of publicly available document data will be made available to reuse.

If access is restricted, please specify who will be able to access the data and under what conditions.

The restrictions will be negotiated with the research participants (individuals and companies involved in the research). In the most strict case, the pseudonymized data types with restricted access will be accessible to the KU Leuven and partnering researchers involved in the project.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- Yes, other

The interview data on both in silico medicine and (re)insurance industry might include confidential information about the development of new products/business knowledge. This, depending on the data agreements formed with the participating actors, might restrict the sharing of (some of) the data.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

From the parts that data can be made available, it will be shared in KU Leuven RDR

When will the data be made available?

- Other (specify below)

From the parts that data can be made available, it will be made available at the end of the project.

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

The publicly available document data can have a creative commons license.
Depending on the level of accessibly that the data agreements with the research participants grant to the focus group/interview data, the data will have a Data Transfer Agreement.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

What are the expected costs for data sharing? How will these costs be covered?

There are no expected costs for data sharing.

Responsibilities

Who will manage data documentation and metadata during the research project?

Researchers Maiju Tanninen and Elisa Elhadj.

Who will manage data storage and backup during the research project?

Researchers Maiju Tanninen and Elisa Elhadj.

Who will manage data preservation and sharing?

Researchers Maiju Tanninen and Elisa Elhadj.

Who will update and implement this DMP?

Researchers Maiju Tanninen and Elisa Elhadj.