Designing Interactive Data Stories to Uncover and Navigate Uncertainty

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: Audiovisual Images Sound Numerical Textual Model SOftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Bibliographic references	Bibliographic references managed in Zotero	N	D	T	.pdf	<100GB	
Literature review	Notes and images	N	D	I/T	.md .jpg .png	<100GB	
card set	Set of printed cards designed by researcher that show visual design strategies	N	D/P	I/T	.pdf	<1GB	About 10-30 cards (playing card size)
Informed consent forms	Signed ICF (paper and scanned)	N	D/P	T	.pdf	<1GB	1 ICF (2 sheets)/ participant
Personal data of participants	Personal data of participants collected via LimeSurvey	N	D	T	.csv	<1GB	
Design exploration	Digital and physical drawings (which are scanned) to explore design possibilities	N	D/P	I	.png .jpg	<100GB	20 sheets of A3 paper and a sketchbook
Datasets for data visualisations	Datasets selected by the researcher to create the data visualisations	E	D	N	.csv	<100GB	
Data visualisation for news item (several iterations)	A design for a data visualisation, including earlier iterations	N	D	I/T	.png .svg	<100GB	
News data story (several iterations)	A design for a data story, which is a digital news article containing multiple visualisations (including earlier iterations)	N	D	I/T/SO	.png .svg .html .css .js	<100GB	
Interactive data visualisation for news (several iterations)	A design for an interactive data visualisation, including earlier iterations	N	D	I/T/SO	.png .svg .html .css .js	<100GB	
Co-creation recordings (3)	Audio files of the co-creation sessions	N	D	S	.wav	<100GB	
Co-creation transcripts	Pseudonymized transcripts of audiofiles of the co-creation sessions	N	D	T	.odt	<1GB	

Co-creation drawings	Drawings made by participants during co-creation sessions	N	P	I			approximately 6 A2 sheets of paper
Co-creation notes	Notes taken during co-creation sessions	N	D	T	.md	<1GB	
Co-creation reports	Written reports that combine notes, drawings end insights from the transcribed recordings	N	D	T/I	.pdf	<1GB	
Interview recordings (36)	Audio files of the interviews	N	D	S	.wav	<100GB	
Interview transcripts	Pseudonymized transcripts of audiofiles of the interviews	N	D	T	.odt	<1GB	
Interview notes	Notes taken during Interviews	N	D	T	.md	<1GB	
Qualitative data analysis	Qualitative data analysis documents	N	D	T	.docx	<1GB	
Log Data	Log Data of user interactions with digital prototype (via Adobe Analytics)	N	D	T	.csv .pdf	<1GB	
Questionnaires	Digital questionnaires after user interaction with prototype (via LimeSurvey)	N	D	T	.csv	<1GB	
quantitative data analysis	quantitative data analysis documents	N	D	T/N	.docx	<1GB	
Design guidelines document	A living, online document that functions as a guidebook	N	D	T	.pdf	<1GB	

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:

Global Carbon Budget (2023) – with major processing by Our World in Data. "Fossil fuels – GCB" [dataset]. Global Carbon Project, "Global Carbon Budget" [original data]. Andrew, R. M., & Peters, G. P. (2023). *The Global Carbon Project's fossil CO2 emissions dataset* (Version 2023v36) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10177738

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)

Ethical issues have been carefully considered as human participants will be involved in co-creation sessions, semi-structured interviews, and quantitative evaluations. Informed consent will be obtained from all participants, and pseudonymization measures will be applied to protect participants' identities and personal data. As audio recordings may contain identifiable information they will be transcribed by the researcher and deleted as soon as possible. The study has received ethical approval from SMEC (PRET approval number: G-2024-8513-R2(MIN)) and complies with the ethical guidelines set forth by AVG and SMEC.

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)

The personal data collected in this project includes identifiers and demographic data: names, phone numbers, and email addresses for co-creation sessions and interviews, as well as IP addresses for the quantitative evaluation. All participants will have their age and gender recorded, and additional information will be gathered on media attitudes and data visualization literacy for general public participants.
The data types vary by evaluation method:

- **Co-creation sessions**: Participants' profession will be recorded. Audio recordings of up to three hours will be made, alongside observational notes and participant drawings.
- **Semi-structured interviews**: Audio recordings of interviews lasting up to 20 minutes will capture topics such as participant understanding of the data visualization, perceptions of uncertainty and reliability, opinions on author credibility, and the

impact of visual language on these perceptions.
- **Quantitative evaluation**: Participants will complete a survey with a 'feeling thermometer,' questions on perceived uncertainty and data reliability, and author credibility. Interaction data will also be collected to assess user engagement with the visualisations.

Data collection complies with GDPR regulations, ensuring informed consent, pseudonymization, and secure data storage.
The project has received ethical approval and is registered under number G-2024-8513-R2(MIN)

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.

- No

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.

- Yes

The datasets used to create the visualizations are published under a CC BY 4.0 license. We will ensure proper attribution to original data sources.

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).

Documentation of data will be included in the form of README.txt files to ensures accurate interpretation, and facilitates future data reuse. Data will describe features such as when it was created, file type, the context in which it was gathered (e.g. transcript of audio captured during a co-creation session with group a), a description of the contents of each dataset.
These README.txt files will be placed inside each folder alongside the data and organised on KU Leuven's OneDrive, ensuring easy sharing among researchers at KU Leuven. Bellow the top-level folder, which is the main project folder, there will be a folder for each workpackage. The work packages will have a subfolder for each type of data (e.g., audio files, transcripts, visualisations, images, survey responses).
Each file name will follow this structure:

1. Case study number (1, 2, 3, or none)
2. Data Description (e.g.Interview)
3. Date in format: YYYYMMDD

The folder structure and file naming conventions will be described in a README.txt file in the main project folder.

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 specific metadata standard will not be used. README.txt files will explain the folder structure and file naming scheme, ensuring the data is findable and understandable.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)

How will the data be backed up?

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

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?

Data will be securely stored in OneDrive, with full access restricted to the PhD researcher. Documents can be selectively shared with supervisors via OneDrive. If documents need to be shared outside the research project (e.g., within the research group), this will be done with restricted access—shared via email with view-only permissions.
Physical documents will be securely stored in a locked drawer accessible only to the PhD researcher.

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

No costs are expected for data storage and backup during the research project, as we will exclusively use the available free storage of KU Leuven OneDrive, a service provided to all staff of KU Leuven.

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...).

- Certain data cannot be kept for 10 years (explain below)

Most data will be retained for 10 years to support reproducibility and potential follow-up research. However, personal data that is not essential for long-term analysis will be deleted when it is no longer necessary: IP addresses collected during online surveys, used to ensure each participant is unique, will be deleted immediately after verification. Contact details for participants will be deleted once they are no longer needed for ongoing communication or study invitations. Additionally, audio recordings will be deleted after transcription to prevent potential identification of participants.

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

- KU Leuven RDR

Physical materials, such as drawings, will be digitized for long-term storage. The only items retained in physical form will be the signed informed consent forms, which will be securely stored in a locked container.

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

No costs are expected for data preservation during the retention period, as all digital data will be stored on KU Leuven's infrastructure, which is provided at no additional cost.

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.

- Yes, as open data
- Yes, as restricted data (upon approval, or institutional access only)
- No (closed access)

Open data

Data will be open unless access is restricted or closed where necessary because of ethical issues regarding privacy.

Restricted Data

Pseudonymized transcripts, questionnaire responses, data analysis documents, log data and researcher notes from co-creation sessions and interviews will be restricted and only made available upon approval.

Closed access

Sensitive data, including informed consent forms, non-pseudonymized personal data, and audio recordings of sessions and interviews, will remain securely stored and inaccessible to external parties to ensure privacy and compliance with ethical guidelines.

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

The collected data will be accessible to the phd researcher Gijs Ipers, the main project supervisor (prof. Dr. Sandy Claes) and the co-promotor (prof. Dr. Michaël Opgenhaffen).

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, privacy aspects

Personal data of the research participants (co-creation participants, survey respondents and interviewees) will not be made public for privacy reasons. Only pseumdonimised data will be available.
Audio recordings will be transcribed and deleted to preserve the privacy of participants.

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)
- Other (specify below)

Design guidelines documents and the code for the visualisations will (in addition to the RDR) be uploaded to GitHub.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- MIT licence (code)

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.

- Yes, a PID will be added upon deposit in a data repository

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

There are no expected costs as both the KU Leuven RDR as GitHub are provided for free.

Responsibilities

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

During the project, the doctoral student (Gijs Ipers) will be responsible for documenting data and generating metadata to improve data readability. They will be closely supported by the main promotor of the project (prof. Dr. Sandy Claes) and the co-promotor (prof. Dr. Michaël Opgenhaffen).

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

During the project, the doctoral student (Gijs Ipers) will be responsible for managing the data. They will be supervised in this task by the main promotor of the project (prof. Dr. Sandy Claes) and the co-promotor (prof. Dr. Michaël Opgenhaffen).

Who will manage data preservation and sharing?

The main promotor of the project (prof. Dr. Sandy Claes) will be designated responsible person for longterm management, preservation and facilitation of data generated within this research project.

Who will update and implement this DMP?

This is the collective responsibility of all three involved researchers: Gijs Ipers, Sandy Claes, and Michaël Opgenhaffen.