

ETHICAL 'COLLECTIONS AS DATA' STEWARDSHIP IN THE AGE OF ARTIFICIAL INTELLIGENCE

Participant Information Leaflet (PIL)

This leaflet provides information about the research study **Ethical 'Collections as Data' Stewardship in the Age of Artificial Intelligence**¹ conducted by Giulia Osti, d-real/SFI-funded PhD candidate at the University College Dublin, School of Information and Communication Studies, under the supervision of Assoc. Prof. Amber Cushing (University College Dublin) and Prof. Suzanne Little (Dublin City University).

Before you decide whether to participate in this study, I would like to provide you with some detailed information to help you make a choice. Please take the time to read the following information as it will help you understand the process you may be involved in. Thank you very much for your time and consideration.

What is this research about?

The Collections as Data (CaD)² initiative attempts to provide an ethically informed framework and resources to guide the creation of collections in the form of data to meet the needs of real communities, empowering them and the stewards who take care of these collections. CaD initiative is a relatively recent phenomenon that has started influencing the understanding of heritage data in multiple countries and disciplines.

This study aims to investigate how CaD principles are interpreted and implemented by professionals in various institutional and disciplinary contexts. Additionally, it seeks to explore the underlying ideas and values associated with CaD stewardship, ultimately approaching the ethical challenges that arise at the intersection of CaD and Artificial Intelligence (AI) systems.

Why am I doing this research?

You have been invited to participate in this research as we seek individuals who:

- Are familiar with the CaD initiative.
- Support or actively implement CaD principles in their professional practices.

¹ This work is conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

² <https://collectionsasdata.github.io/>

- Work in fields such as library and information science, archival studies, digital humanities, etc. particularly (but not exclusively) with digital heritage collections.
- Are interested in discussing the ethical challenges posed by artificial intelligence to CaD.

How will my data be used?

The findings from the data analysis will inform my PhD research and its subsequent refinement, facilitating the generation of relevant academic outputs (e.g. presentations, papers, workshops) in collaboration with my supervisors.

What will happen if I decide to take part in this research study?

If you choose to participate, you will be asked to take part in an online interview through Zoom. After filling in the interview recruitment form we will schedule an interview slot according to your availability. During the interview, you will be asked to answer 8-15 questions and perform 2 card-sorting tasks, which I estimate will take you 45 to 60 minutes. Following the interview, you will be sent a copy of the edited (de-identified) version of the interview transcript.

How will my privacy be protected?

Your privacy is of the utmost importance to us. To safeguard it, all participant information will be handled with strict confidentiality. Your personal information and original audio recordings will never be shared with any third parties and will be stored securely on the UCD Google Drive cloud system with restricted access to me and my main supervisor.

Before any publication, all interview data (including the ones collected through the interview recruitment form) will undergo de-identification, and participants will be referred to by using unique identifiers (e.g., PI0xx), following the dedicated **de-identification protocol**³. For data collection purposes, a re-identification key will be created, to be destroyed at the end of the study, alongside the audio recordings and raw transcripts. The de-identified versions of interview transcripts and participants' demographics will only be disseminated through publication; a copy will be archived in the UCD Google Drive for up to 5 years.

³https://github.com/semanticnoodles/phd-research/blob/main/study-1/de-identification_protocol.pdf

What are the benefits of taking part in this research study?

While there may not be immediate or direct benefits, your insights will significantly contribute to future research beyond the scope of my PhD. Additionally, you will have the chance to be selected to receive a National Book Token e-gift card worth €25 or equivalent. These tokens can be used online or in bookshops across Ireland and the UK, and many participating bookshops offer worldwide shipping.

What are the risks of taking part in this research study?

There are no foreseen risks from participating in this research. This research has been identified as a low-risk level study and has received clearance from the UCD Office of Research Ethics (reference no. HS-LR-24-223-Osti-Cushing).

Can I change my mind at any stage and withdraw from the study?

You can withdraw from the study at any time without giving a reason or incurring any penalty. If you decide to withdraw from the study before it is completed, your data will be destroyed immediately.

How will I find out what happens with this project?

You will have the chance to decide whether or not to be notified when the outcomes of this study will be published when filling in the interview recruitment form.