

## Participant Information Sheet

**Project title:** Data Hazards: using worst case scenarios to make data ethics relevant to abstract technologies.

We would like to invite you to take part in our research project about how we can use worst case scenarios to help people to think more about the consequences of their research in its early stages. It is based around a workshop where participants will consider some real data science projects.

Before you decide whether or not to participate, we would like you to understand why the research is being conducted and what it would involve for you. Please ask us any questions if anything is unclear.

### **What is the purpose of the project?**

Data science products that go on to cause harm to individuals, groups or the environment are the result of years of research and development. We think it should be easier for people working in the early stages of new technologies to consider what the future harms of their work could be, so that we don't leave these issues until they are being deployed. This project is about developing a set of 'Data Hazards' (a bit like the COSHH chemical hazard symbols) that acknowledge the harm that data projects could cause, and clearly communicate that these new technologies need to be 'handled with care'. This has two purposes: the first is to make data scientists reflect more regularly on the potential implications of their work, before it is applied to a product, and also to share the responsibility of considering potential harms across all researchers who 'pick up' this technology along its cycle of development.

### **Why have I been invited to participate?**

There are two possible roles to participate in the workshop: audience members and data science project owners.

**Audience members:** You have been invited to participate as someone who has an interest in data science and its application in the real world. We would like a wide variety of inputs, so participants do not need to be data scientists themselves. They could be social scientists, philosophers, or just people who like data.

**Data project owners:** You have been invited to take part as someone who is working on a data science project that could be presented in the workshop. The project should not yet be being developed as a product, but should involve using data.

### **Do I have to take part?**

It is entirely up to you to decide whether you want to participate in the workshop. We are very happy to answer any questions you might have over email, or a video call. If you agree to take part, we will ask you to sign a consent form. You are free to withdraw at any time, without giving a reason.

Once the workshop has taken place we ask that you let us know within 1 month if you would like to withdraw, since after this time we will not be able to de-anonymise all the data we have collected. We will not be able to remove your contribution to the discussion, but we can delete your survey responses, and not use any quotes from you.

**What will happen to me if I take part and what will I have to do?**

If you take part we will ask you participate in an online workshop that will last up to 2 hours. The purpose of the workshop is to explore which 'Hazard Labels' would best apply to a series of 'live' data science projects that will be presented during the workshop.

*If you would like to be a data science project owner:*

Before the workshop we will ask you to do a short survey about your project, and which hazard labels you think best fit it from a list of options.

Then, at the workshop we will ask you to present a 5-7 minute overview of your project to the 'audience members' (up to 30 people), who will all be participants in this research or researchers. We will ask you then to meet with a small group of the participants in a breakout room (a researcher will be there too to facilitate). You will take questions from the 'audience' members about your project, and then we will ask the audience to discuss the ethical implications of your project without any input from you. After this you can discuss with the audience your thoughts on their conversation, and how the conversation impacts on your decisions about the Hazard Labels for your project.

*If you would like to be an audience member:*

We will ask you to attend the workshop where you will hear from data science project owners about their projects, and we will ask you to then take part in the breakout room discussions around particular projects and your thoughts on which of the hazard labels best fit. We will encourage you to speculate around 'worst case scenarios' for the projects, and make your own assessments of which Hazard Labels you would apply.

All participants will be asked to do a survey at the end of the workshop about their experience, and changes they would make the labels or process of applying them.

**What are the possible disadvantages and risks involved in taking part in the project?**

As a data science project owner, we hope the experience of participating will be useful to you, but it is possible that you might find it uncomfortable to hear about negatives of your research. We do not anticipate any disadvantages to audience members.

**What are the possible benefits of taking part?**

Our hope is that this reflective process, and the Data Hazards, will help data scientists to consider ethical implications of their work earlier, and more regularly. This could have great benefits for wider society, and contribute to us all using data in more responsible ways.

**Will my participation in this project be kept confidential?**

We will record the workshop and this recording will be saved directly to University of Bristol digital storage. This recording is only for the use of the research team and will be deleted at the end of the study. No full transcripts of the workshop will be shared or stored long term.

We would like to use verbal quotes from participants as part of the reporting of the evaluation, and this is the reason for recording the workshop. Any quotes we used would not be attributed to individuals.

The data from surveys filled out by participants will be shared openly on an online repository ([The Open Science Framework](#)), without names, directly identifying information, or free-text fields. We will not ask anything sensitive in the survey – it will be about your opinion of the Data Hazards framework, and your experience of the workshop, and which broad subject background you are from.

While you will not be identified by name or project title, it may be possible for observers to infer who survey results belonged to for project owners. When describing the projects

presented by project owners we will use broad descriptions of the field (e.g. “Machine Translation”, “Computational Genomics”), rather than specific descriptions of the exact research area and use case. If you are one of a small number of people in this field his could be used to de-anonymise your survey responses.

Surveys will be collected using University approved survey software.

**What will happen to the results of the research project?**

We intend to write the evaluation as a research article and publish it in an academic journal. We will also share outcomes through blog posts. Project owners will have the chance to approve these outputs before publication, to ensure that they are happy with the level of project anonymisation. We will share these outputs through the JGI and Data Ethics Club mailing list.

**Who is organising and funding the research?**

Natalie Thurlby (Data Scientist, Jean Golding Institute, University of Bristol) and Nina Di Cara (GW4 BioMed PhD Student, University of Bristol) are organising this research. It is supported by the Jean Golding Institute and has received funding from the School of Computer Science, Electrical and Electronic Engineering and Engineering Mathematics at the University of Bristol.

**Who has reviewed the study?**

This study has been reviewed by the Engineering Faculty Research Ethics Committee.

**Further information and contact details**

Natalie Thurlby: [natalie.thurlby@bristol.ac.uk](mailto:natalie.thurlby@bristol.ac.uk)

Nina Di Cara: [nina.dicara@bristol.ac.uk](mailto:nina.dicara@bristol.ac.uk)

If you have any concerns related to this study please direct them to the Faculty of Engineering Research Ethics Committee, via Liam McKervery, Research Governance and Ethics Officer (Tel: 0117 331 7472 email: [Liam.McKervery@bristol.ac.uk](mailto:Liam.McKervery@bristol.ac.uk))