Data Hazards Chapter

Susana Roman Garcia, Ceilidh Welsh, Natalie Zelenka, need to add Irma

2023-11-13

Table of contents

Landing Page

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

This is the hackmd we worked on the May Book dash: https://hackmd.io/ZoJpaBsnTBuAfwGWnMCwGg?view

Part I

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

2 How to use the Data Hazards project

(er-datahazardshowtouse)=

There are four steps to using the Data Hazard labels: 1. **{ref}Learning<er-datahazardshowtouse-labels>**: familiarising yourself with the Data Hazard labels. 2. **{ref}Applying<er-datahazardshowtouse-apply>**: deciding which Hazard labels are relevant to your project. 3. **{ref}Reflecting<er-datahazardshowtouse-prec** on what to do differently and what mitigations to make. 4. **{ref}Display<er-datahazardshowtouse-display>**: displaying the labels alongside your work can help you to communicate that you've thought about these broad ethical issues and how you'd like others to use your work.

This spells LARD , which makes it pretty easy to remember! Rest-assured, this project is still suitable for vegans .

In addition, whether you have used the labels yet or not, you can also **{ref}contribute<er-datahazardshowtou** to the project. If you think any labels are missing or could be improved, this project is looking to evolve with collaboration.

(er-datahazardshowtouse-labels)= ## **1. Learning** about the Data Hazard labels The first step in using the Data Hazards materials is learning about the Data Hazards labels: familiarising yourself with them so that you can later {ref}apply<er-datahazardshowtouse-apply> them to a project.

Data Hazards labels are supposed to represent as broad a selection of ethical risks associated with data-centric work as possible. This includes, but is not limited to the risks considered by {ref}ethics committees<er-committees>, which often focus on risks to research participants that could lead to legal repercussions for the research organisation, such as consent and privacy. It also includes issues like algorithmic bias or danger of misuse that might result from downstream outputs of research, rather than the research process itself.

Learning about the Data Hazard labels is usually part of a Data Hazards workshop, but you can also do it independently by: 1. Reading the labels on the website - click on each of them for more information! 2. Practicing applying them to a project. 3. Talking about them with other people. 4. Printing out label cards and playing games with them.

(er-datahazardshowtouse-apply)= ## **2. Applying** the Data Hazard labels *Applying* the Data Hazard labels means deciding which labels are relevant to a project.

2.0.1 What projects can I apply the labels to?

We recommend that you apply the labels to your own work unless you've been invited to give feedback on another piece of work.

The labels have been applied to many different types of projects that use data or data-intensive methods. Some examples of projects that have used Data Hazards include: - A predictive model that aims to use machine learning to predict human traits from their genetic mutations. - A digital humanities project that uses web-scraping and natural language processing to analyse text of political speeches. - An NHS data-linkage project to create a new database for researchers to use. - A modelling project that aims to improve whole-cell models of a bacteria.

2.0.2 When should I apply Data Hazards in my work?

Data Hazards labels can be *applied* at any stage in the research life-cycle. Ideally, the best time to *apply* the labels is close to the beginning of the research project, at the same time as you might consider your {ref}`university ethics process<>`, or {ref}`pre-registering your analysis<>` (you can also pre-register your Data Hazards analysis!). However, much of our research builds on itself, so researchers have also found workshops useful to reflect and plan what they can do better in follow-on projects.

Read {ref}`the PhD research life-cycle case study<>` for an example of how applying, reflecting, and displaying labels can fit into a research life-cycle.

2.0.3 How: Workshop

Applying labels through a collaborative and interdisciplinary discussion can give project owners perspectives they might not otherwise consider.

2.0.4 How: Self reflection

(er-datahazardshowtouse-precautions)= ## **3. Reflecting** on Safety Precautions

(er-datahazardshowtouse-display)= ## **4. Displaying** the labels

(er-datahazardshowtouse-contribute)= ## **Contributing** to Data Hazards materials

3 DERR Symposium and Data Hazards Workshop Case Study

(sectioninitials-filename)=

(sectioninitials-filename-overview)= ## Overview

The Data Ethics and Reproducible Research (DERR) Symposium was a one-day event held at the Alan Turing Institute 10th March 2023. This symposium aimed to showcase implementations of reproducibility and ethics for research. The symposium structure included a variety of speakers, an interactive workshop and networking oppurtunities.

insert picture of the event here

A central part of this symposium was the {ref}`Data Hazards<er-datahazards>` Workshop to provide a collaborative training opportunity for attendees. The workshop encourages attendees to explore, discuss and reflect on the ethical implications and wider societal impact of specific data-intensive projects. It was an opportunity for attendees to appreciate that ethics is complex, situational and important to discuss in our own contexts. It aims to shine a light on an event that promotes data ethics through the content of the event, and also its planning and delivery.

This case study is an account of our first-hand experiences organising and hosting an event. It may be useful to you if you would like to host or organise your own accessible workshop in data ethics and reproducbility (whether a Data Hazards workshop or not).

For more detailed information and checklists, please see the Turing Way page for {ref}`Organising a Conference<cl-organising-conferences>`.

(sectioninitials-filename-goals)= ### DERR Symposium Goals

Central Goal: to provide an open, inclusive and accessible space for attendees to learn from one another and discuss first-hand experiences applying ethics and reproducibility to their work. This aims to discuss the successes and challenges we might face, and importantly how to consider ethics as more than a tick-box exercise in a research project.

Goals for Event Organisers: To develop our skills in event management and organisation including: - hosting event platforms. - attendee registration. - questionaries and polls. -

8

requesting catering. - learn about accessibility considerations and requirements. - how to: equality impact assessments.

Moreover, to improve our collaborative, team working and networking skills with both volunteers and colleagues to provide a curated and thoughtful event.

Goals for Event Attendees: - identify areas of research projects where wider ethical implications can be considered in different research contexts. - learn through a hands-on workshop how to embed ethics and reproducibility into current work or research. - to discover what other attendees are working on in these areas.

(sectioninitials-filename-organising)= ## Organising the Event Here we discuss the logistics for organising an event, from setting-up and promoting the event on an external platform to collaborating with volunteers and speakers to help successfully run the event on the day.

(sectioninitials-filename-organising-agenda)= ### Create an Agenda: The very first step was to put together some initial ideas for the agenda. We worked together using google docs. Please find our finalised agenda here. What did this look like for us? - brainstorming themed talks, - deciding on keynote speakers - practicalities of running the Data Hazards Workshop. - identified target audience for the symposium, - when the event would start and end, including rest breaks*, - catering options - decided to run the event as hybrid (online and in-person).

*To make the event more accessible in terms of able-bodied capacity, it was important for us to **make sure there were enough 'bio breaks'** to allow people to stretch their bodies, use the toilet and breathe a little.

(sectioninitials-filename-organising-advertising)= ### Advertise the Event

Choosing a date and gauging interest: We reached out to our community *before beginning to organise* the symposium to gauge interest and spread the word. This allowed us to know what other events were going on in our organisation and research field so that we could avoid clashes or busy times of the year. It also ensures you have engaged participants who have already indicated their availability.

Eventbrite: After finding an available date for our interested participants, we chose to advertise our event on Eventbrite. Because we were running a hybrid event, we created two event pages (one in-person and one online-only). Eventbrite worked well for us because: - it could be accessed by everyone, not just those in our organisation. - we could use it to gather participant information (clearly explaining why we were asking for the sensitive information and how the data would be handled), including accessibility information participants wanted to declare and dietary requirements.

(sectioninitials-filename-organising-communication)= ### Communication We emailed the registered participants at several time-points in the run-up to the event. This aimed to: - encourage those who could no longer attend to cancel their reserved slot so that someone else could attend in their place. - provide a small nudge and reminder that the event was happening. - provide participants a Code of Conduct in advance. - provide a zoom link

for online participants and directions to the location of the symposium for those attending in-person.

This is what a template email could look like:

> Dear Participants, Thank you for registering for the *(In person) Data Hazards, Ethics and Reproducibility One-Day Symposium*, on the *10th March 2023*. We hope you are as excited as we are about this symposium!
>
> **Checking in: can you still make it?** We ask you to please be considerate and kindly let us know in advance if you can no longer make it to the event. This will help us reduce waste when ordering catering as well as help with organizing room numbers. Due to the high demand of this event, we have people in a waiting list, so please do cancel your ticket in advance if you cannot make it any more, in order to allow someone else to join instead. Please cancel directly through Eventbrite or by emailing the organisers.
>
> **Agenda**: Please find the tentative agenda here. It is mostly set up by now, please keep your eyes open for final details in the coming days. We are looking forward to showcasing these wonderful people and talks.
>
> **Code of Conduct**: We ask all attendees to familiarize themselves with the Code of Conduct for the event (add your own link here). Please do have a read in your own time in order to allow for an open and welcoming environment among all participants.
>
> **How to reach the Alan Turing Institute, Enigma room**: Once you arrive to the main Alan Turing Reception area, there will be someone there to greet you and guide you towards the Enigma room. Instructions on how to get to the Turing Institute can be found here.

(sectioninitials-filename-organising-collaboration)= ### Collaboration: The DERR symposium had two primary organisers, however would not have been possible without the help of different collaborators. Here we discuss the steps we took to reach out to volunteers, symposium collaborators and speakers to help the event run smoothly.

Internal Teams We had several collaborators from within our organisation including: - the **Skills Team** who assisted us with: - completeing equity, diversity and inclusion assesment and data proctection forms. - understanding the financial requirements of the event, from paying speaker travel expenses to setting-up a purchase order for an event caterer. - providing grassroots funding to cover event expenses. - the **Facilities Team** who assisted with: - delivery of the catering on the day. - room bookings to host the event. - microphone and audio-visual set-up for the room to host the hybrid event.

Volunteers Volunteers were an essential part of the day, and our event would not have been successful without their help! For clear communication with volunteers, we: - **provided an availability form** for volunteers to fill in, indicating specific sessions or hours they could

help, what they would like to do, and their own accessibility requirements. - **hosted an on-boarding call** to talk through the agenda and highlight the parts of the day where we would require help from volunteers. - **created a dedicated volunteer's Slack channel**, so that volunteers could ask questions relating to the event, and we could easily do any problem shooting during the event.

Find our volunteer form here.

Speakers As part of the symposium, we had to reach out to speakers for the event:

- we reached out to directly to key-note speakers whose research area or field considered data ethics and reproducibility.
- we invited speakers from both inside and outside our organisation to include a variety of experiences via a speaker registration form.
- we asked for abstracts and titles, providing potential speakers with and estimated length of talk (including time for questions) and session they would present in.

Find our speaker form here

(sectioninitials-filename-organising-catering)= ### Catering For our event we choose a local, entirely vegan caterer, to align with the core values of the event. To ensure successful ordering and delivery of the catering on the day we: - estimated the number of attendees to provide catering - including volunteers, speakers and organisers. - recieved a quote from the caterers in advance and ensure we had the budget to cover the quote. - collected dietary requirements to ensure that all attendees were catered for, and ensured the catering company recieved these requirements in advance. - arranged a delivery time and location. - ensured payment to catering company as per company and organisation policies.

Tip: make sure terms and conditions are read and timing requirements from different parties involved are clear from the beginning. It might be that the caterer you are ordering from requires a purchase order to be completed a certain time in order to complete the order action.

(sectioninitials-filename-organising-accessibility)= ### Accessibility

Code of Conduct We put together a code of conduct for the event that reflected our values and the values of the community we aimed to create during the symposium. It also summarised the expectations of participants joining the symposium and laid out reporting guidlines and contact points for if the code of conduct was breached during the event.

Please find our Code of Conduct here.

(sectioninitials-filename-hosting)= ## Hosting the Event Some tips for the day of the event. ### Starting the Event - Arrive at least 1h hour prior to start time to the in-person venue. To ensure the zoom link and audio-visual equipment (speakers, screens and microphones) are all working. - Open the online call early to allow for: - volunteers to settle in and open any

rooms necessary. - participants to join in advance and set-up break out rooms for networking events and workshops later in the day.

3.0.1 Hybrid Event Tips

We choose to host the Symposium in a hybrid format as this would allow attendance for those interested in the topic, and speakers, who would be unable to attend in-person. - Have session chairs, which will introduce speakers and make sure people keep to agenda timings. - Set a non-disruptive 1min warning. For example, have a 1min sign which you can rise, or type in the chat, when the time approaches. - Keep questions to the end of the talks to avoid disruptions and talks becoming longer.

3.0.2 Accessibility

There is already a chapter on Organising Conferences which includes an in-depth dive into requirements for accessibility. As well as on Organising Hybrid meetings. Here, we highlight some points to ensure that the hybrid-format and the event were accessible: - collected accessibility requirements at the time of advertising, to help with planning. - used Otter.ai, a speech to text transcription application using artificial intelligence and machine learning. It shows captions for live speakers, and generates written transcriptions of the speeches. - ensured all participants in-person were using microphone when speaking. - provided regular accessibility breaks. - reminded all participants which sections of the day would be recorded. - provided a space for participants to include their name and pronouns using name labels for in-person and Zoom names online. - provided a code of conduct, and contact points for if the code of conduct was breached. - used the raised-hand function in Zoom for direct questions, but included a session chair to monitor for questions in the chat.

(sectioninitials-filename-hosting-workshop)= ## Hosting a Workshop: Data Hazards ### Communication ### Collaboration **Project Owners** *insert the project owner case study here ### Accesibility*

(sectioninitials-filename-postevent)= ## Post Event ### Feedback Forms To cover all avenues for feedback after the event we: - provided a QR code and link to the feedback form in the final session of the day and encouraged participants to complete this form. - emailed participants after the event to thank them for their attendance and share the feedback form again. - collected slides and recordings from speakers. - created an public Git repository where all content is openly avaiable for attendees and those interested in organising and hosting a similar event.

Find our feedback form here.

Feedback is really invaluable to ensure that future organisers can learn from previous events. In our case, this was our first symposium, so feedback from participants was really important! Here we provide some feedback we were given after the event.

Positive Feedback: - attendees that came to the symposium with no experience of the data hazards enjoyed the interactive nature of the workshop, saying that: *'the structure of the course made it very easy to follow, and the case study element was very interactive with lots of opportunities to ask questions.'* - attendees also enjoyed the wide applications of data ethics and reproducibility in the research space - attendees enjoyed networking over lunch, and our vegan catering got lots of compliments!

Constructive Feedback: - some attendees struggled with people joining and leaving parts of the online event, but recognised this was out of the organisers control. A consideration fur future events would be to create initiatives to encourage people to stay online for the whole day. - attendees felt that an additional industry perspective would have been beneficial.

3.0.3 Git Repository

To embody the topics of reproducibility and open science, we have shared all of our content, from forms to recordings on a public Git Repository. The aim of this is to help anyone who wants to run a workshop or symposium in the future consider all the different elements of event organisation and hosting.

Find our full Git Repository, including recordings of speakers, code of conduct, and forms here.

(sectioninitials-filename-challenges)= ## Challenges As this was the first event (on this scale!) that we had organised, there were several challenges that popped up throughout the process.

The primary challenge was understanding the number of hours required to organise an event. For our one day event, it took over 50 hours of organisational time, from form writing, code of conduct, creating the event page, advertisement, organising catering and so on. Particularly if this is the first time you are organising an event in a new context, either as an organisation or on that research topic, there won't be clear expectations or guidelines to follow. This means different tasks can take longer than expected to complete. So we would really recommend that you have a large organising commitee of passionate people that can also recognise the boundaries and constraints of the event type and work within them. This also taught us that organising an event will **always** take longer than you expect, so plan accordingly!

(sectioninitials-filename-reflections)= ## Reflections

Upon reflections after the event, we reflected that it is important to understand your organisations guidelines and frameworks for hosting events, what support they can offer you, and reach out to ask for help where you need it. The importance of collaboration in successful event organisation cannot be understated here. This is an important value in the Turing Way, with a

{ref}`Guide for Collaboration<cl-collaboration>`providing insights in how to collaborate remotely, running open source projects and organising events.

We also recognise the importance of {ref}`self reflection<er-self-reflection>` throughout the process of organising an event, and how our lived experiences give us a particular lens on how we view the world. So ensuring you collaborate with others, take on different points of view that will ensure a responsible, inclusive, fair and ultimately ethical event!

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.or
g/10.1093/comjnl/27.2.97.