
Plan Overview

A Data Management Plan created using DMPonline

Title: GRAIL (Getting Responsible about AI and machine Learning in research funding and evaluation) - Workshop series

Creator: Denis Newman-Griffis

Principal Investigator: Denis Newman-Griffis

Data Manager: Denis Newman-Griffis

Project Administrator: Denis Newman-Griffis

Affiliation: The University of Sheffield

Template: The University of Sheffield Research DMP

ORCID iD: 0000-0002-0473-4226

Project abstract:

This is a knowledge exchange project. This is part of the larger GRAIL project ongoing with the Research on Research Institute (RoRI) and partner funding organisations; the scope of this application is for the KE workshop series to be run from Summer 2023 - end of project in April 2025.

The aim of the project is to develop a community of learning around the use of artificial intelligence (AI) and machine learning (ML) techniques and technologies within research funding and evaluation. As with many other public bodies, research funders are interested in exploring the use of AI/ML techniques to more effectively use the data and expertise they have built up in funding and evaluating research programmes to improve the delivery of their strategic aims. However, there are no established best practices on how to integrate AI inquiry and techniques into existing organisational processes, or on how to ensure that AI use is aligned with the public-serving mission of funding organisations.

ID: 126412

Start date: 19-06-2023

End date: 01-04-2025

Last modified: 19-06-2023

Grant number / URL: [Cost centre pending]

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

GRAIL (Getting Responsible about AI and machine Learning in research funding and evaluation) - Workshop series

Defining your data

- What digital data (and physical data if applicable) will you collect or create during the project?
- How will the data be collected or created, and how will you assure the quality of your data collection and processing?
- Approximately how much digital data will be generated (in GB, MB, etc), and what formats will they be in (e.g. .doc, .txt, .jpeg)?
- Are you using pre-existing datasets? Give details if possible, including conditions of use.

Data will be collected in the course of a Knowledge Exchange workshop series. We will run between 12-15 2-hour long virtual workshops over the course of the project period. Each will focus on a different aspect/use case of AI/ML use within research funding and evaluation, to serve as a springboard for discussion among participants about broader principles and practices of AI use.

Data will be created in the following form(s):

1. Written notes including quotes and key discussion points from workshop discussions.
2. Audio/video recording of workshop sessions. Workshops will NOT be recorded in the first instance, however participants may request that a session be recorded with the consent of all participants. No recordings are planned at this point but may be collected over the course of the project.

Looking after data during your research

- Where will you store digital data during the project to ensure it is secure and backed up regularly? (E.g. [University research storage](#), or University Google drive)
- How will you name and organise your data files?
- If you collect or create physical data, where will you store these securely?
- Will you use extra security precautions for any of your digital or physical data? (E.g. for sensitive and/or personal data)
- What metadata/documentation will you create for your data? (E.g. a README file including methodology and file structure; descriptive metadata to enable discovery in a data repository)

Workshop notes may be recorded with identifying initials of speakers.

These notes will be stored on a secure University of Sheffield Google Drive folder with access limited to the PI and RA (TBN) only.

After conclusion of each workshop session, anonymised versions of discussion notes will be produced by the research team for storage on the Research on Research Institute Google Drive, with distribution of copies to workshop participants as part of the co-productive workshop process.

Storing data after your research

- Which data supporting your research conclusions will be stored on a long-term basis after the end of the project?
- Where will the data be stored after the project (e.g. University of Sheffield repository [ORDA](#), or a subject-specific repository) and for how long (e.g. standard TUoS retention period of minimum 10 years after the project)?
- Will your chosen long-term data storage incur any financial costs?

Anonymised versions of discussion notes will be stored on the Research on Research Institute Google Drive; based on content of notes the research team will discuss further public sharing of anonymised notes via ORDA and/or RoRI FigShare with the

project partners & workshop participants.

Project outputs, including a working paper summarising the workshop process and main discussion points, will be posted both on RoRI FigShare and ORDA for long-term archival storage.

No long-term storage options are anticipated to incur financial costs.

Sharing data after your research

- How will you make data available outside of the research group after the project? (E.g. openly available through a repository, or on request through your department)
- Will you make all of your data available, or are there reasons you can't do this? (E.g. personal data, commercial or legal restrictions, very large datasets)
- If you can't share all of your data, how might you make as much of it available as possible? (E.g. anonymisation, participant consent, sharing analysed data only)
- How will you make your data as widely accessible as possible? (E.g. include a data availability statement in publications; ensure published data has a DOI)
- Will there be any delay before making data available? If so, give the reasons for this.

Personal data (consisting solely of workshop notes with initialised attribution of quotes and/or discussion points) will NOT be shared externally and will be destroyed 3 years after the final publication/dissemination of project outputs.

As this is a co-production project, we will work with the workshop participants and project partners to establish the best approach to distribute anonymised versions of workshop notes. At minimum the output working paper/report from the project, which will summarise key takeaways from workshop discussions, will be made available publicly via RoRI FigShare and ORDA, with DOI associated.

Putting your plan into practice

- Who will be responsible for data management in the project? (There may be more than one person)
- Do you require any extra resources to put your data management plan into practice? Will this incur any financial costs?

PI Newman-Griffis will be responsible for data management.

No additional resources or financial costs are anticipated.