
Plan Overview

A Data Management Plan created using DMPonline

Title: Engineering biology: AI, intermediaries, and responsible innovation

Creator: Claire Holland

Affiliation: University of Manchester

Funder: Biotechnology and Biological Sciences Research Council (BBSRC)

Template: EPSRC Data Management Plan Customised By: University of Manchester

ORCID ID: 0000-0003-2488-5985

Project abstract:

The project explores the framing and operationalization of responsible innovation in the context of AI-enabled engineering biology. A key project focus is on innovation intermediaries involved in aspects of research, applied development and translation of AI-enabled engineering biology. We probe the development of innovation intermediaries from (1) linear bridging organisations of research-to-business translation to (2) hubs that catalyse innovation ecosystems for economic benefit; and suggest evolutionary needs and opportunities for intermediaries to (3) embed greater attention to societal and sustainability objectives and concerns as well as economic goals. We also explore how researchers, practitioners, and other stakeholders understand responsible innovation and embed attention to societal and sustainability implications of AI-enabled engineering biology. We identify and examine framings, practices, and challenges in engaging with responsible innovation, sustainable development, governance, and societal control and alignment through targeted case studies and interviews with a sample of UK-based innovation intermediaries, researchers, practitioners, and other stakeholders in engineering biology. Insights for intermediary development, research management, and policy are explored.

ID: 104435

Start date: 25-07-2022

End date: 31-01-2024

Last modified: 15-08-2022

Grant number / URL: BB/W013770/1; BB/W014483/1

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

Engineering biology: AI, intermediaries, and responsible innovation

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- None of the above

2. Is The University of Manchester collaborating with other institutions on this project?

- Yes - Part of a collaboration and owning or handling data

3. What data will you use in this project (please select all that apply)?

- Re-use existing data (please list below)
- Acquire new data

Existing data that will be used includes public domain data about research publications and patents, information included in published documents, and information posted on publicly-available web sites. Mixed sources of data will be deployed, with analyses drawing on primary field interviews, and secondary sources (available public domain literature, reports, and government and business statistics).

4. Where will the data be stored and backed-up during the project lifetime?

- University of Manchester Research Data Storage Service (Isilon)

5. If you will be using Research Data Storage, how much storage will you require?

- 1 - 8 TB

6. Are you going to be receiving data from, or sharing data with an external third party?

- No

7. How long do you intend to keep your data for after the end of your project (in years)?

- 5 - 10 years

Guidance for questions 8 to 13

Highly restricted information defined in the [Information security classification, ownership and secure information handling SOP](#) is information that requires enhanced security as unauthorised disclosure could cause significant harm

to individuals or to the University and its ambitions in respect of its purpose, vision and values. This could be: information that is subject to export controls; valuable intellectual property; security sensitive material or research in key industrial fields at particular risk of being targeted by foreign states. See more [examples of highly restricted information](#).

Personal information, also known as personal data, relates to identifiable living individuals. Personal data is classed as special category personal data if it includes any of the following types of information about an identifiable living individual: racial or ethnic origin; political opinions; religious or similar philosophical beliefs; trade union membership; genetic data; biometric data; health data; sexual life; sexual orientation.

Please note that in line with [data protection law](#) (the UK General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de-identified) as soon as practically possible. You must obtain the appropriate [ethical approval](#) in order to use identifiable personal data.

8. What type of information will you be processing (please select all that apply)?

- Pseudonymised personal data
- Personal information, including signed consent forms
- Audio and/or video recordings

9. How do you plan to store, protect and ensure confidentiality of any highly restricted data or personal data (please select all that apply)?

- Where needed, follow University of Manchester guidelines for disposing of personal data
- Anonymise data
- Store data in encrypted files, folders, computers or devices
- Store data on University of Manchester approved and securely backed up servers or computers

Original contact details, and interview recordings and transcripts will be stored in encrypted folders and/or password protected form only. The data we are collecting primarily relates to business, sustainability, AI, and policy. Participation in interviews and workshops will be voluntary and compliant with consent and ethical approvals. We will offer all participants anonymity.

Participant information (including contact details) will be stored on University of Manchester data servers in password protected and encrypted folders that are kept separate from password protected and encrypted folders containing transcripts of interviews and focus groups.

Recorded interviews will be transferred from the recording device onto an encrypted device at the earliest opportunity.

During data processing data will be anonymised with any identifying information and recordings kept separate from the data files in encrypted and password protected form.

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- No

Personal information including contact details of interviewees, workshop participants, and consent forms will not be kept beyond the end of the project.

11. Will the participants' information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- No

12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?

- Not applicable

13. Are you planning to use the personal information for future purposes such as research?

- No

14. Will this project use innovative technologies to collect or process data?

- No

15. Who will act as the data custodian for this study, and so be responsible for the information involved?

Claire Holland

16. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2022-07-21

Data Collection

What data will you collect or create?

This is a mixed-method study. The project stages outlined in this proposal will deliver different sets of data: (i) publication and public company data (ii) interview data, (iii) case data and (iv) personal data (of interviewees, workshop participants, etc.), which will be anonymised. Initial data collection time is planned to be completed in 6 months. Additional data (through further interviews and cases) may occur in 2023. We estimate the cumulative duration of interview audio recordings at about 15 to 25 hours (12 to 20 interviews @ up to 1.25 hours each) = a storage volume of 9.8 to 16.3 GB (@0.65 GB/hour). Up to another 10GB will be required for storage of associated documentation and secondary data.

(i) Initial data collection will be through landscape and bibliometric analysis of existing data from scientific publications, publicly available company resources, and relevant literature (Microsoft Academic, Scopus). The data from existing publications are subscription-restricted but access is widely available at higher education institutions and have been used in numerous empirical applications including by the PI and Co-I. Microsoft Academic also maintains records of pre-print publications, allowing us to ensure the information we have is as up to date as possible.

(ii) Primary research will be conducted through interviews with at least 12 researchers involved in the integration of AI and Engineering Biology, found through searches of publicly available resources such as EPSRC and UKRI funding databases. Interviews will be conducted using standard audio recording equipment or online recording tools where in person interviews are not feasible. Audio files will be provided directly to a transcriber for transcription. The original audio files and the transcribed files will be stored only on encrypted devices. We shall explain the interview/case study/assessment intentions to participants as well as the intentions of the research plan and other required information, to ensure informed consent.

(iii) Case study data include company level information as well as information regarding collaborators, supply chains, feedstock, technology, and policies and processes predominately collected from publicly available datasources such as company and Government websites, publications and funding databases.

(iv) Personal data: Participants will complete informed consent forms and be provided with participant information sheets. Any personally identifying data will be protected by anonymisation, encryption, and secure storage using the UoM RDM service. For all data files we will provide any scripts used for data preparation or cleaning and maintain detailed documentation. For all files there will be annotations in an accompanying file, documenting the process of data collection. The raw data files will be kept intact but all data preparation routines (conducted with SQL) will have a version control and be backed up. This will allow us to reproduce all analysis steps at a later point in time.

How will the data be collected or created?

Secondary/existing data collection will be through landscape and bibliometric analysis of existing data from scientific publications, biointermediary online resources, and relevant literature (Microsoft Academic, Scopus) and public company and Governmental consultation documents/profiles/resources.

Primary data will be collected through audio recordings of semi-structured interviews. Recordings will be saved on the recording device in encrypted form and transferred in encrypted form to the UoM server for transcription. This transcription will take place as soon as possible after the interview, in most cases on the same or following day. After transcription, the recordings will be deleted and the transcription saved in encrypted form using the UoM IT department RDM service. Any later changes (such as answers

removed or amended from transcripts) will be noted and dated in the transcript .doc file. Further primary data collection will be through case studies. Consent will be sought from participants and encryption of sensitive/ publicly unavailable information will be available.

Documentation and Metadata

What documentation and metadata will accompany the data?

We will collect informed consent forms from all interview and case study participants. These will be kept separate from the data collected and anonymised but can later be connected if required by ethics review. We will also maintain a research diary listing dates of interviews, changes to original data, research context, methods, quality control and so on. This will also be pseudonymised and stored on the UoM RDM services. Other data will be found in relevant literature and public information, some of which can be reused. Data types: raw data, processed data, documents and scripts, and tabular data with minimal metadata. Widely used formats: MS Word (.doc / .docx), MS Excel (.xls / .xlsx), JPEG (.jpeg, .jpg, .jp2), PDF (.pdf), etc.

Ethics and Legal Compliance

How will you manage any ethical issues?

We put forward our proposed work to the University of Manchester AMBS ethics review chair. We were informed (by email) that no ethical approval was required for it in its current form. However, if things change throughout the development of the project we will communicate with them and request further guidance to ensure compliance with ethical standards. Participants will complete informed consent forms and be provided with participant information. Archiving of transcribed interviews will be discussed with participants and their consent sought, and their approval or refusal recorded properly.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The existing public domain data (such as publications, patents) collected by this project does not have copyright or IPR. We will include details of fair use for the qualitative data in the informed consent form, to be presented to respondents prior to each interview. All terms of use in the University's IP Policy will be complied with. The rights to the newly generated data will be owned by the University of Manchester and the investigators.

Storage and Backup

How will the data be stored and backed up during the research?

The project research data will be stored and backed up on the University of Manchester Research Data Storage Service (Isilon). During data collection, data will be kept by the investigators and PDRAs on secure university servers in password protected or encrypted form.

For long term storage, this data will be stored and backed up on Isilon. Original contact details, interview and survey data is sensitive and will therefore be stored in encrypted folders and/or password protected form only. Recorded interviews will be transferred from an encrypted recording device onto an encrypted device at the earliest opportunity. After completing the file transfer for safe research data storage, all interview records prepared on the laptop will be permanently deleted. The transfer will be done electronically. All research data will be recorded electronically.

During the project, copies of the original data will be maintained on the secure University of Manchester servers and the university has a backup routine in place to ensure no data is lost. After the end of the project all data will be maintained by the investigators as long as required by regulations.

How will you manage access and security?

The desktop with access to the UoM server is password protected. Data is stored on the UoM server. After transcribing, we will

transfer the transcriptions in encrypted form to the UoM IT department RDM service and delete the recording from the UoM server. Access to data stored with the RDM service will also be password protected. All data (transcripts, as well as existing quantitative data) will be stored long-term on the RDM service. Data will be encrypted at every stage.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

We will retain transcripts and analysis of interviews, plus analysis of the publicly available resources and policy data, through to 5-10 years after project completion.

What is the long-term preservation plan for the dataset?

Data such as interview recordings, transcripts and analysis will initially be saved on the University of Manchester server. After transcribing, recordings will be deleted and remaining data will be saved using the Research Data Management service provided by the University of Manchester IT Department. We will store these transcripts using the RDM service for 5-10 years after project completion.

Data Sharing

How will you share the data?

Individualised identifiable data will not be shared with any third party. Anonymised and aggregated data (where no individuals are identifiable) may be shared (in a paper or as an appendix) for purposes of publication or in response to requests from others, in accordance with open sharing of data good practices. For open access sharing of anonymised and aggregated data (including methodologies, thematic analysis codes, and other materials), we will use University's institutional data repository (Figshare).

Are any restrictions on data sharing required?

Data will only be shared after the results are published. The license will be provided by CCBY. No personal information or identifiable information will be released publicly.

Responsibilities and Resources

Who will be responsible for data management?

Responsibility for data management, quality control and data archiving of the UK data will be shared by the investigator, PDRAs, and PhD students involved. The designated data manager is Dr Claire Holland.

What resources will you require to deliver your plan?

Access to the Research Data Management service provided by the University of Manchester.