Plan Overview

A Data Management Plan created using DMPonline

Title: Assessment of institutional quality checks to improve the Findability and Accessibility content of life sciences manuscripts

Creator: Fiona Booth

Principal Investigator: Fiona Booth

Contributor: Nicholas Beazley-Long

Affiliation: University of Bristol

Funder: Economic and Social Research Council (ESRC)

Template: ESRC Template

ORCID iD: 0009-0008-1617-9822

Project abstract:

This study aims to assess if whether independent checks performed manually by academic institutions have a significant impact in increasing the first two elements of the FAIR Guiding Principles, Findability and Accessibility. We will compare this prospective intervention in a randomised, controlled manner to providing a checklist and training material alone to authors. We will also examine the cost-benefit ratio of the intervention should the check show positive impact.

In addition to the randomised controlled trial of the intervention of independent checks we will collect qualitative data on factors which influence their decisions and the practical implications of openly sharing their research data.

ID: 155441

Start date: 06-01-2025

End date: 05-01-2027

Last modified: 12-07-2024

Grant number / URL: APP46065

Assessment of institutional quality checks to improve the Findability and Accessibility content of life sciences manuscripts

Assessment of existing data

Provide an explanation of the existing data sources that will be used by the research project, with references

Literature searches have shown than there is no available data to demonstrate the potential of independently performed checks for FAIR guiding principles in improving the content of research manuscripts. The study involves 3 groups, 1 of which will use existing data:

Group 0: research published by University of Bristol and University of Northumbria Health and Life Science researchers will act as a control arm for Findability and Accessibility checks

Groups 1 & 2: these groups require the generation of novel research data to assess the impact of the intervention of independently performed Findability and Accessibility checks

Provide an analysis of the gaps identified between the currently available and required data for the research

There are no datasets generated which match the interventions of this study for Groups 1 and 2.

Information on new data

Provide information on the data that will be produced or accessed by the research project

The following data will be produced:

1. Data on Compliance with Findability and Accessibility:
1.1 Scores will be recorded for each category:
Group 0: Published manuscripts will be scored once against the defined Findability and Accessibility criteria.
Group 1: Pre-submission manuscripts submitted following review of training data and checklist will be scored against the defined Findability and Accessibility criteria.
Group 2: (A). Baseline Score: pre-submission manuscripts submitted following review of training data and checklist will be scored against the defined Findability and Accessibility criteria.

(B) Post FA check score: pre-submission manuscripts submitted following provision of the results of the Findability and Accessibility check to the researcher will be scored against the Findability and Accessibility criteria again.

1.2 Duration of each assessment: the time taken to assess each manuscript for Findability and Accessibility will be recorded: date of check and duration of check in minutes.

2. Qualitative Assessment
All participants will be asked to provide feedback on their experience of completing the training and, for those in Group 1, the experience of receiving specific feedback on Findability and Accessibility.

Questionnaire will include both closed and open-ended questions.

Quality assurance of data

Describe the procedures for quality assurance that will be carried out on the data collected at the time of data collection, data entry, digitisation and data checking.

Scoring guidance: the Project Co-Lead and the Research Co-Lead will review a variety of research datasets and create a scoring guidance document which provides specific instructions for the range of datasets reviewed. This will assist the Research Associate in scoring manuscripts in a consistent manner.

To minimise errors in the scoring of manuscripts by the Research Co-Lead, 5% of manuscripts in each randomised group will be checked by the Project Co-Lead as well as the Research Co-Lead. The results of each check will then be compared to ensure consistency and reference made to the scoring guidance to resolve any conflicting scores.

Data collected on the date and duration of checks: date range validation checks and max/min duration validation checks will be perform to help identify erroneous data.

Backup and security of data

Describe the data security and backup procedures you will adopt to ensure the data and metadata are securely stored during the lifetime of the project.

During the research study the data will be stored on SharePoint provided by University of Bristol: this is backed-up according the University's IT Security Policy and access will be restricted to the Project Lead, Research Co-Lead and Project Co-Lead.

Following completion of the project the data will be moved to the Research Data Storage Repository (RDSF) which is backed up to tape.

Management and curation of data

Outline your plans for preparing, organising and documenting data.

1. Research Dataset

1.1 Findability and Accessibility Scores:

Scores will be recorded for each category in a MS Excel, one file and one worksheet per pre-submission manuscript (Groups 1 and 2) and one file per published paper (Group 0.)

1.2 Qualitative Assessment

The questionnaire will be administered via Microsoft Forms or a similar tool which creates a record of the survey responses held in the University's secure Microsoft ecosystem. All responses will be

downloaded and stored into one MS Excel file.

2. Metadata

2.1 Informed consent record. A MS Excel record of the participant ID number, date and time that the participant provided informed consent to enter the study, and, if a participant chooses to withdraw their consent, the date and time that the request was made and consent withdrawn. Research discipline will also be recorded here.
2.2 Date and time of receipt of all pre-submission manuscripts for both Group 1 and Group 2 logged in a MS Excel file
2.3 Date and time that the results of the FA check were provided to Group 2 researchers
2.4 Attribution metadata accompanying each manuscript score i.e. author, when, plus justification/evidence for scores plus comments
2.5 Feedback generated following each score - txt file with attribution metadata
2.6 Date and time of Qualitative Assessment Survey responses

Difficulties in data sharing and measures to overcome these

Identify any potential obstacles to sharing your data, explain which and the possible measures you can apply to overcome these.

Datasets can be fully anonymised and contextual information which may identify a participant redacted. This is described in the next section.

Consent, anonymisation and strategies to enable further re-use of data

Make explicit mention of the planned procedures to handle consent for data sharing for data obtained from human participants, and/or how to anonymise data, to make sure that data can be made available and accessible for future scientific research.

Consent for data-sharing (in anonymised form) will be included in the informed consent form which participants are asked to review and complete as a condition of entering the study.

The following measures will be taken to produce a dataset which is fully anonymised and can therefore be shared openly:

1. Upon study entry participants will be assigned an ID number which will be recorded along with their name and email address in a password-protected file in the secure SharePoint site. These details will be used to contact participants during the study. Once data analysis is complete this file will be destroyed.
2. The Qualitative Assessment Questionnaire will be reviewed to check for any contextual information which may identify a participant. This data will be redacted from the version of this file which is prepared for data-sharing. The full set of responses will be held in the RDSF and not made publicly available.

Copyright and intellectual property ownership

State who will own the copyright and IPR of any new data that you will generate.

Data will be shared under a CC by 4.0 licence. The IPR will belong to the University of Bristol.

Responsibilities

Outline responsibilities for data management within research teams at all partner institutions

All data will be handled and stored by the University of Bristol:
PL: Project Lead
PcL: Project Co-Lead
RcL: Research Co-Lead

1. Assign permissions to SharePoint file and periodically monitor (PL & CL)
2. Create and maintain data management plan (RcL)
3. Create secure SharePoint site for data storage (PL & CL)
4. Create a file hierarchy which is described in an overarching README file (RcL)
5. Create standard file formats based on a recognised tabular data standard and README files for metadata (RcL)
6. File data in SharePoint on an ongoing basis (RcL)
7. File analysis datasets (RcL)
8. Prepare anonymised datasets (RcL)
9. Perform periodic checks of file contents (CL)
10. Transfer all data to RDSF at study end (PL & CL)

Preparation of data for sharing and archiving

Are the plans for preparing and documenting data for sharing and archiving with the UK Data Service appropriate?

Datasets will be provided in open formats (e.g. xlsx) for archiving with the UK Data Service.

Is there evidence that data will be well documented during research to provide highquality contextual information and/or structured metadata for secondary users?

This data is described in this section: Outline your plans for preparing, organising and documenting data.