Plan Overview

A Data Management Plan created using DMPonline

Title: Comparing the accessibility, metadata content, and interoperability of export formats from website searches of six academic search systems commonly used in biomedical evidence syntheses

Creator: Emma Wilson

Principal Investigator: Malcolm Macleod

Affiliation: University of Edinburgh

Template: UoE Default DMP template for Research Staff

ORCID iD: 0000-0001-9187-9839

Project abstract:

When downloading search results from online academic search systems for evidence synthesis projects or bibliometric analyses, access to bibliographic metadata is important for retrieval of further data and streamlining of projects. For instance, digital object identifiers (DOI) can be used to access the record online and retrieve further metadata from additional sources such as OpenAlex (RRID:SCR_022787) or CrossRef (RRID:SCR_003217). Database accession numbers can be helpful to ensure that data is not duplicated in projects, such as when conducting a systematic review update. Additionally, International Standard Serial Numbers (ISSN) can be used to identify journal names that may appear slightly differently across databases.

Bibliographic databases can be assessed by online platforms or academic search systems (i.e., PubMed is used to access MEDLINE) which in turn allow various export format options. These academic search systems may place restrictions on exports, such as the number of records that can be exported in a file, and the data fields exported may differ across different export formats (e.g., the abstract text may be exported in one export type but not another). Additionally, some file formats are more supported than others for importing into reference manager software such as Zotero (e.g., RIS, BIB) or programming software such as R (e.g., CSV).

This project aims to provide a comprehensive overview of the accessibility (e.g., download restrictions), data content (e.g., which data fields can be exported), and interoperability (e.g., which file formats can be imported easily into specific software and how data from different bibliographic databases can be combined) of different export types accessed from from six academic search systems commonly used in biomedical evidence syntheses.

ID: 180483

Start date: 07-07-2025

End date: 31-10-2025

Last modified: 04-07-2025

Copyright information:

Comparing the accessibility, metadata content, and interoperability of export formats from website searches of six academic search systems commonly used in biomedical evidence syntheses

Administrative Information

1) School or Institute

- CMVM - Edinburgh Neuroscience

2) Project start date

2025-07-07

3) Project end date

2025-10-31

4) Project funder or sponsor.

Simons Initiative for the Developing Brain (SIDB)

Data Collection

5) Data Collection

Academic search systems
Exports from the following academic search systems / database will be assessed:

- PubMed (RRID:SCR_004846; https://pubmed.ncbi.nlm.nih.gov/)
- EuropePMC (RRID:SCR_005901; https://europepmc.org/)
- OVID: Embase (RRID:SCR_001650)
- OVID: PsycINFO (RRID:SCR_014799)
- Scopus (RRID:SCR_022559)
- Web of Science: Core Collection (RRID:SCR_022706)

Web of Science Core Collection, Scopus, Embase, and PsycINFO web platforms will be accessed via

University of Edinburgh secure logins.

These search systems and databases were chosen as they are commonly used in healthcare and life sciences evidence synthesis projects, which is relevant to our previous and current work.

What data will be collected

- Data related to bibliographic database accessibility, metadata content, and interoperability
- Search result export files from bibliographic databases

How will the data be collected

- Accessibility, content, and interoperability data will be collected through manual testing of web interfaces
- Search result export files will be retrieved via using the "export" function on bibliographic database web platforms

What format will data be collected in

- All data will be in text format
- All data generated through the project will be collected in csv format
- Data from search results will be in a variety of text formats (txt, csv, bib, and ris)

What volume of data will be collected

- Only a small sample (e.g. 100 records) of data will be exported in each format

Documentation & Metadata

6) Documentation & Metadata

What documentation and metadata will be kept

- Data on accessibility, content, and interoperability will be labelled with the name of the database the information pertains to
- Search exports will be named in the format: "[name of database]_export_[export type]_[date as YYMMDD].[file extension]"
- A data dictionary and README explaining all data collected will be kept up to date

Ethics & Legal Compliance

7) Ethics & Legal Compliance

Some bibliographic data included in exports (e.g., abstract text) will be subject to copyright or other licenses. Before collecting any data, I will familiarise myself with each databases' licensing policies to ensure ethical and legal compliance. I will store data to comply with all copyright restrictions and only share data where I have the right to do so.

Storage and Back-Up

8) Where will your data be stored and backed-up during the project?

Data will be stored and backed-up on the UoE DataStore platform for the duration of the project.

Selection and Preservation

9) Where will the data be stored long-term?

Edinburgh DataShare will be used for long term data storage and sharing.

10) Which data will be retained long-term?

Accessibility, content, and interoperability data will be shared. As I will be unable to share search export files due to publisher or data provider licensing, I will share metadata on what fields were included in each export.

Data Sharing

11) Will the data produced from your project be made open?

- Yes: go to 12

12) How will you maximize data discoverability & access?

Data will be shared in Edinburgh DataShare along with comprehensive and clear metadata and contributor ORCID IDs. The DataShare DOI will be linked to in Data Availability Statements made in open access publications and/or preprints from the project.

Responsibilities & Resources

14) Who will be responsible for the research data management of this project?

Emma Wilson will be responsible for data collection, documentation, storage, and sharing.

15) Will you require any training or resources to properly manage your research data throughout this project?

Emma will take refresher training in DataStore and DataShare and writing high quality data documentation as required.