RESEARCH DATA GUIDE

This Library guide provides information on research data management, including data discovery; terms and conditions of access and use; data security; dataset metadata and documentation; data repositing, preservation and open data sharing.

This third edition contains expanded details about data management plans (DMPs), data management in the EU Horizon 2020 research funding programme, and options for open data sharing.

The focus is on digital research data outputs - primarily numerical statistical data. Guidelines can also be applied to other forms of research data: qualitative, textual, audio and visual - the generation of which is anticipated to increase over time.

The guide also addresses data sharing, focusing on whether, when, how, where and under what terms, research data outputs can be openly shared. Please contact the Library to arrange a research data appointment.

Thomas Bourke
Economics Information Specialist
econlibrary@eui.eu
July, 2015

An online version of this document is available at:
http://www.eui.eu/Research/Library/ResearchDataServices/Guide.aspx
1. Data discovery and access
2. Terms and conditions of use
3. Support, software and infrastructure
4. Research data management
5. Data management in EU Horizon 2020
6. Metadata: describing data outputs
7. Data preservation, repositing and open data
8. EUI Library research data services
9. International resources and guidelines

1. DATA DISCOVERY AND ACCESS

This section describes access to EUI Library licensed data resources, access to data hosted by other institutions and access to open data resources.

1(a) The EUI Library Data Portal

The EUI Library maintains a Data Portal providing access to both (i) licensed data resources and (ii) open data resources of relevance to EUI research themes.

There are three sub-directories: one for macroeconomic data; one for micro-socioeconomic data and one for Europe-related data. The Data Portal is primarily dedicated to quantitative data resources, however the number of qualitative resources is steadily increasing.

**Macroeconomic** databases provide statistics on national, regional and global economic and political-economic developments from major international institutions and data publishers, including Thomson Reuters, Eurostat, the ECB, OECD, IMF, World Bank and WIIW.

**Micro-socioeconomic** databases provide individual, family and household-level data observations. Major providers include GESIS, Bureau van Dijk, ICPSR, UKDS, Eurostat and DIW. Access to micro data provided to EUI members under Library license requires additional registration. Special terms and conditions relating to micro-socioeconomic data access and use are described in Section 2(b) below.

**European, EU and Eurozone** data resources provide macroeconomic, financial, monetary and micro-socioeconomic data for research on pan-European topics, EU states and sub-state regions, and Europe in the global context.

All resources in the EUI Data Portal have a dedicated resource page providing:

- Data description
- Time period and release/wave information
- Support links (online manuals, software transfer routines, user networks)
- Details of access and terms of use
Data provided to EUI members under Library license is accessible via internet protocol (IP) or - for micro-socioeconomic data - restricted server. The Datastream financial database is run directly from the EUI desktop ‘programmes’ menu. Data from the Inter-university Consortium for Political and Social Research (ICPSR) archive is downloaded by Library staff on request.

1(b) Access to data at other institutions

EUI scholars who require access to data held at another facility should apply for access early in their research projects, as applications can be lengthy. Some micro-socioeconomic geo-coded data can only be accessed at the issuer's secure facility. Contact the Library for assistance with access applications. Scholars who require access to unpublished datasets (eg. underlying data associated with a journal article) should contact the Library before writing to data owners/authors. In certain instances, it may be possible for EUI scholars to obtain access via Library consortia. See Section 2(c) on data user undertakings below.

1(c) Access to open data resources

Open data resources are available on the internet. Major data repositories are indexed by subject, data type and host location in the re3data.org\(^1\) research data registry and the data repositories section of the Open Access Directory.\(^2\) Section 7 of this guide provides information on how research data outputs can be shared as open data.

2. TERMS AND CONDITIONS OF USE

This section gives an overview of terms and conditions of data access and use; special security measures for micro-socioeconomic data; individual data user undertakings; and terms and conditions of use for open data.

2(a) License agreements

The use of data provided by the EUI Library is subject to license agreements, copyright terms and data protection provisions. Data users are individually responsible for compliance with such terms. Violation of terms and conditions puts at risk other EUI members’ future access to data resources. All users are requested to scrupulously abide by the terms and conditions of access to, and use of, data. Open data, freely accessible via internet, are normally subject to terms and conditions of use and restrictions on republishing; see Section 2(d) below.

Under the EUI Library’s license agreements for data resources: Users may not distribute or allow any other party to have access to data which is provided under license; users may not modify or create a derivative work of the licensed materials without the permission of the licensor; users may not remove, obscure or modify any copyright or

---

1 http://www.re3data.org/
2 http://oad.simmons.edu/oadwiki/Data_repositories
other proprietary notices included in licensed materials; users may not use licensed materials for commercial purposes; users may not retain or distribute substantial portions of a database, and must comply with any post-project data destruction undertakings in the license. Additional guidelines are the Library’s Terms and Conditions page. The University of Michigan maintains a database of data user agreements.

2(b) Micro data access and use

Special terms and conditions apply to access and use of micro-socioeconomic data. These reflect the sensitive nature of data observations about individuals, families and households. Such terms and conditions apply to both (i) micro data hosted by the Library for EUI members and (ii) micro data hosted by third-parties and provided directly to EUI members under individual license. Specific terms and conditions are accessible via the Micro-socioeconomic Data Directory at 'full details'.

Micro data users must preserve the confidentiality of observations pertaining to dataset subjects: individuals, families and households. Users must not attempt to identify any individual or other data subject in the dataset. Micro data users must also comply with any post-project data destruction undertakings in the license.

EUI members must use the Library micro data access form when applying for access to micro-socio economic data hosted by the Library by (i) Selecting one dataset from the dropdown menu (a separate form is required for each dataset) (ii) Signing the EUI paper form 'Terms and Conditions of Use of Micro Data Provided by the EUI Library' (iii) Signing the data provider's terms and conditions paper form. Forms can be signed at the Economics Information Office (Badia Library, 085). For access to some datasets, providers require parallel application procedures. In some cases (eg. Eurostat) it is necessary to present a project proposal when applying for access to micro data.

There is no off-campus access, VPN access, or laptop access to micro-socioeconomic data hosted by the EUI Library. Non-EUI members and short-term visitors do not have access to EUI-hosted micro data, and should contact their institution's data librarian, statistics manager, or data publishers directly for information about access at their location.

2(c) Data User Undertakings: individual agreements with data providers

Some data issuers require that access contracts be established directly with the user(s) of the data. EUI members who require access to data provided by third parties are required to sign a data user agreement with the provider/licensor. Some data issuers also request that an EUI administrator co-sign individual user agreements. In such cases the Library will require the data user(s) to sign an internal undertaking that they will abide by the terms and conditions of use and access of the data.

---

3 [http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx](http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx)
4 [http://www.psc.isr.umich.edu/dis/data/dc/grid.html](http://www.psc.isr.umich.edu/dis/data/dc/grid.html)
2(d) Open Data: terms and conditions of use

Open data, freely accessible via internet, are normally subject to terms and conditions of use, re-distribution and/or re-platforming - even in the absence of an explicit user contract. By accessing and using open data, individual users are responsible for complying with terms of use. In cases where open data are made available without explicit licence, users should contact the data owner, or the EUI Library, before re-hosting or re-distributing large sections of an open dataset. See also Section 7(b) below.

3. SUPPORT, SOFTWARE AND INFRASTRUCTURE

This section outlines EUI Library and ICT Service support for data users.

3(a) Library support

Data support is available at the Library, Badia Fiesolana, office 085, on weekday mornings and Tuesday and Thursday afternoons. Data support in the Economics Department (Villa San Paolo) is provided on Monday, Wednesday and Friday afternoons from 15:00 to 18:30. Training sessions take place throughout the year.

The Library maintains the Data Portal and the Micro Data Restricted Server, and provides assistance with access to, and use of, open data by advising users on research data management and open data options.

Data software manuals (print versions) are available in the Badia Library at shelfmarks 001 to 005. A directory of online research data software manuals, with full-text links, is available on the Library web site. The Library also provides a comprehensive collection of monographic works on data methodology, ontology and statistical theory. Books and manuals in any language can be suggested for acquisition by the Library.

Every Friday during term, the Library issues an e-Bulletin with data news, including details of new data waves, data errata, platform upgrades and tips on how to access Library databases and open data. EUI members can sign up via econilibrary@eui.eu.

3(b) ICT support

Software, infrastructure and connectivity support is provided by the EUI ICT Service. Research software programmes available at the EUI are listed on the ICT web site. Technical support is provided by the site offices of the ICT Service. Advice on the use of statistical software is provided by the EUI research software tutors. The EUI provides access to all major data software including: Stata, Gauss, MATLAB, OxMetrics, Stat/Transfer, WinEdt, WinRATS, R for Windows and other programmes.

4. RESEARCH DATA MANAGEMENT

This section provides an introduction to research data management, the security of data during research projects and data management plans.

4(a) Research data management

Research data management during the project cycle includes the control of data inputs, the elaboration, collation and protection of data, and the creation of data outputs. RDM covers the description of data and tools; the storage of data during analysis; the provision of clear and accurate metadata; the preservation of data; and - where possible - making research data outputs available to other researchers. Research data management is carried out by individual researchers and research teams throughout the duration of a research project.

4(b) Security of data during research projects

During the project cycle, it is important to keep data safe and secure. Researchers are encouraged to use a desktop computer for data elaboration, and to make regular backups on the EUI network server or a safely-secured external memory device. In accordance with contractual agreements, micro data should only be elaborated on a desktop computer in a secure location. Data CD-roms and associated documentation, should be kept in locked storage when not in use.

4(c) Data management plans

Data management plans (DMPs) - now required by most science funding agencies - are short documents outlining how data are handled, stored, documented, preserved, and - where possible - made available for sharing. DMPs include information on:

- The generation of data and the sources of data
- How data is elaborated
- How data is documented; which metadata scheme is used
- Where data will be stored during the project
- How data is preserved
- Which researcher (in a team) decides about future sharing of data outputs

It is important to keep an accurate record of dataset changes, variables, characteristics, software versioning and - in the case of survey data - pre-agreed terms of disclosure, so that decisions about data sharing can be made at the end of a project. 'Metadata' refers to the descriptors or 'tags' that identify a dataset (see Section 6 below).

Data management plans can be used as the basis for deciding whether, when, how, where and under what terms, research data outputs can be openly shared. Open data refers to the trend among international organisations, scholars and government agencies to share data outputs.
EUI members who are required to submit a data management plan - either as part of a funding proposal or early in a research project – should contact the Library for assistance. Tools such as DMPonline\(^\text{10}\) – the Digital Curation Centre's data management planning tool - can be used to construct a tailored data management plan. To use this tool, enter an email address, name of organisation and create a password. EUI users of DMPonline should select 'other organisation' from the drop-down menu. First-time users are taken to the ‘edit profile’ section of the platform. When using the tool to create a data management plan, it is possible to select a funder template (eg. European Commission, Horizon 2020) which generates an appropriate matrix. The author(s) complete each free-text section of the plan - responding to the prompts. It is possible to share editing rights with project collaborators by entering their emails and assigning status as ‘co-owner’, ‘editor’ or ‘read only.’ For assistance using DMPonline, contact the EUI Library.

The name of the Principal Investigator (P.I.) should be clearly stated in data management plans. For research teams, this should be the contact person for all decisions regarding whether, when, how, where and under what terms, research data outputs emanating from the research project might be openly shared. If the Principal Investigator is not the same person as the project Data Manager, this should be clearly stated. In international collaborative projects, the name of the person who has authority with regard to decisions on the sharing of data outputs should be clearly indicated. Unless otherwise indicated, the Principal Investigator undertakes this role. Many data repositories require a 'data ownership statement' from the data depositor.

5. DATA MANAGEMENT IN EU HORIZON 2020

Horizon 2020 is the EU Research and Innovation funding programme for the period from 2014 to 2020. EUI members preparing applications for Horizon 2020 funding are required to submit general information on data management as part of their proposal. This is evaluated by the European Commission under the criterion 'impact.' Social science and economic research projects are included in the 'Societal Challenges' cluster - which includes projects in the 'Europe in a changing world - inclusive, innovative and reflective societies' category. The EUI Library can assist project leaders with the data sections of funding applications as well as in-project data management plans.

Project managers should address the following questions - provided by the European Commission - when completing the general information section of H-2020 applications:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.

\(^\text{10}\) http://www.dcc.ac.uk/dmponline
In the context of Horizon 2020, the European Commission has launched a research data pilot. Project managers must provide a data management plan (DMP) within six months of the official start of the research project. The European Commission mandates two further versions of the DMP; one at the mid-point and one at the completion of the funding period. The EC's DMP template is on p.5 of the document ‘Guidelines on Data Management in Horizon 2020’. The DMPonline tool (see 4(c) above) can also be used for H-2020 data management plans - by selecting the H-2020 template in the Funder section.

H-2020 grant beneficiaries must "deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate - free of charge for any user - the following: (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; (ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan." (European Commission Guidelines, p.3.) The EC suggests attaching a license; eg. CC-BY. The EC does not specify or recommend specific repositories. Further details are provided in these official documents:

- Guidelines on Data Management in Horizon 2020
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
- Fact Sheet: Open access to publications and data in Horizon 2020

6. METADATA: DESCRIBING DATA OUTPUTS

Metadata are data about data, presented in a systematic scheme. It is important to keep an updated record describing data capture, use and elaboration throughout the project cycle. There are many different schemata for describing datasets. An introduction to metadata standards for social science and humanities data (eg. Dublin Core, DDI) is maintained by the Digital Curation Centre.

The headings below can be used as a guide when creating metadata. Different repositories require different combinations of metadata fields. Not all of the fields suggested below will be required by data funders or repository managers. Metadata can be used as a 'checklist' to determine whether, when, how, where and under what terms,
research data outputs can be shared as open data. Metadata should be updated as research projects evolve. Some research data outputs may also require multi-lingual metadata and data tags.

**TITLE OF DATASET**
The title should succinctly convey the nature and scope of the dataset. Dossiers and files should be assigned consistent tags.

**NAME(S) OF DATASET CREATOR(S)**
The name, or names, of the researchers, and technical collaborators, who created/elaborated the dataset should be indicated - preferably with institutional contacts. The name of the Principal Investigator (where applicable) should be provided. If the project is undertaken in the context of a consortium, the name of the Data Manager should be provided (if different from the P.I.). Where researcher ID numbers are available, these should be provided (eg. ORCID).

**DESCRIPTION OF DATA**
A meaningful 'data abstract', indicating the kind of data, the scope of the research context and a note on data methodology should be provided. The abstract should be updated as the research project evolves. This information will be useful if a technical paper about the data is submitted to an open data journal.\(^{19}\)

**SOURCE OF DATA**
The source(s) of the data should be indicated. If the dataset has been generated during the research project, this should be stated with details of data collection (eg. survey parameters). If the dataset is derived from a pre-existing database, the source(s) should be clearly cited; eg. institution, publisher or data author. If there are multiple streams of input - all should be cited.

**DATE OF CREATION OF THE DATASET**
Give the date of the completion of the dataset. If the dataset is a new wave of a pre-existing series, this should be clearly indicated with version designation.

**DATE-RANGE COVERAGE OF DATASET**
Indicate the start- and end-dates of data observations contained in the dataset.

**SPATIAL COVERAGE OF DATASET**
Indicate the geographical scope of the dataset (national, regional, global &c.) as applicable.

**FORMAT**
The software format and version should be given (eg. Stata 14, csv, &c.). Software version changes and transfers should be noted.

**LANGUAGE**

\(^{19}\) [https://www.fosteropenscience.eu/foster-taxonomy/open-data-journals](https://www.fosteropenscience.eu/foster-taxonomy/open-data-journals)
The language of data variables, notes and codebooks should be indicated. Multi-lingual documentation, tags, questionnaires and variable descriptions may be necessary (eg. international surveys).

LOCATION OF DATA
During the research cycle, researchers should carefully store and protect their data. After the research project, the location of the data should be given (eg. subject or institutional repository) - even if the data is not openly available.

DIGITAL OBJECT IDENTIFIER
Reposited datasets should be assigned a digital object identifier (DOI). Most data repositories automatically assign DOIs, which is useful for linking to and citing datasets.

ACCESS STATUS AND EMBARGO
After the project, the status of access to the data should be indicated. The status 'Open Data' should be assigned to datasets that are publicly accessible on the internet. If the data is subject to embargo, the end-date of the embargo period should be indicated. If the data can only be accessed at a secured facility this should be indicated. If the data can only be accessed by requesting permission from the Principal Investigator, contact information should also be provided here. (See also Section 7 below.)

LICENSE
Terms and conditions of access and use of data by other persons should be stated. It is advisable to provide a license (eg. CC-BY).

FUNDING STATEMENT
The name of funding bodies, and research grant numbers, should be provided where applicable.

RELATED PUBLICATIONS
Bibliographical details of publications based on the dataset should be listed, with links to abstracts and, where possible, full texts.

DATASET CITATION
A short ‘ready-to-use’ citation reference for the dataset should be provided - incorporating the core descriptive elements.

7. DATA PRESERVATION, REPOSITING AND OPEN DATA

During the course of research projects, scholars should safely preserve their data and associated documentation, code, software and tools. Researchers are encouraged to use a desktop computer for data elaboration, and to make regular backups on the EUI network server or a safely-secured external memory device.

20 https://creativecommons.org/licenses/by/3.0/
7(a) Preservation and repositing

The EUI Library can assist with the preservation of research datasets. Where possible, datasets should be preserved in a data repository. Major data repositories are indexed in the re3data.org registry\(^{21}\) and the data repositories section of the Open Access Directory.\(^{22}\) The Library can assist researchers to deposit datasets in Zenodo, GESIS and other repositories. The Library is also testing the DSpace software as a possible local solution for the repositing of EUI-generated datasets. Reposited data should have at least one human-readable documentation file (eg. txt manual). Note that most repositories will conduct a 'data quality review' of file structure, metadata, format and documentation.\(^{23}\) The UKDS provides guidelines on preparing data for deposit.\(^{24}\)

7(b) Open data

There is a growing trend among scholars, government agencies and international organisations to share data and associated documentation, code, software and tools.

By carefully noting the metadata elements suggested in Section 6 above, scholars will have a ready checklist for determining whether, when, how, where and under what terms, research data outputs can be shared as open data. Datasets that are made available as open data should be the product of original research. Data outputs should be either (i) original data generated during the research project or (ii) data that is the product of significant, value-added elaboration of pre-existing data.

Not all research data outputs can be openly shared. The two most significant considerations when determining whether a research dataset can be made available on an open data basis are:

- **Data protection**: It is not permitted to openly publish datasets in which persons, households or families are identified. In such instances, it may be possible (pending terms) to generate an anonymised version of the dataset for sharing.

- **Database copyright**: It is normally not possible to openly publish significant portions of data that are sourced from pre-existing data resources provided under contractual license. The Library can help scholars determine whether, and how, derivative datasets can be made available as open data.

7(c) Levels of access

Research data outputs can be shared under a variety of access models. Datasets can be made openly available for all users via the internet; can be subject to registration prior to access and use; can be subject to user contracts (sometimes requiring a project plan); can be embargoed for a defined period (or indefinitely); can be restricted to on-site

\(^{21}\) [http://www.re3data.org/](http://www.re3data.org/)

\(^{22}\) [http://oad.simmons.edu/oadwiki/Data_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)


\(^{24}\) [http://ukdataservice.ac.uk/deposit-data/preparing-data](http://ukdataservice.ac.uk/deposit-data/preparing-data)
access and use; or can be subject to an access request directly to the data owner(s) for case-by-case access authorisation. Data can also be reposed solely for preservation purposes (dark archive). Access status can change over time.

Where data are provided on an open basis, it is normally not possible to re-publish freely. Open data are normally subject to licensed terms and conditions of access. Even in cases where there is no explicit license for the re-use of open data resources, scholars should contact data owners/licensors before re-publishing significant portions of an open dataset. Contact the EUI Library for assistance.

8. EUI LIBRARY RESEARCH DATA SERVICES

Research data management assistance is provided by the EUI Library in the following areas:

- Terms and conditions of data access and use
- Data user undertakings for access to third-party resources
- Metadata and ontology for research datasets
- Data documentation
- Data security during research projects
- Data repositing and preservation
- Data sharing and guidelines for open data

Write to econlibrary@eui.eu for data support. Further information is on the Research Data Services section of the Library web site.25

9. INTERNATIONAL RESOURCES AND GUIDELINES

Further information on research data management and open data is available from these sources:

- re3data.org registry of data repositories
- EU Horizon-2020 Open Research Data Pilot
- EC Report on Open Research Data
- Zenodo data repository (CERN/OpenAire)
- OpenAire network and portal
- GESIS repository - Leibniz Institute for the Social Sciences
- Metadata for Social Science and Humanities data (DCC)
- European Union Open Data Portal
- EUI Open Data Registry
- Digital Curation Centre guidelines on data sharing
- Göttingen/OpenAIRE Study on the Protection of Research Data
- Open Economics Principles - Open Knowledge Foundation
- MANTRA data advice for scholars

25 http://www.eui.eu/Projects/EUIdatarepository/Home.aspx
- UK Data Service guidelines on data management
- NSF guidelines for social science and economic data plans
- American Economic Association guidelines on data availability
- DLA Piper – IP rights in data handbook
- U.S. Government open data policy
- GitHub software collaboration
- Research Data Curation Bibliography
- Data User Agreements Directory
- Open Data Handbook - Open Knowledge Foundation
- The Hague Declaration on Knowledge Discovery in the Digital Age
- JISC Directions for Research Data Management in UK Universities

An online version of this document is available at:

http://www.eui.eu/Research/Library/ResearchDataServices/Guide.aspx

econlibrary@eui.eu

July 2015