



RESEARCH DATA GUIDE

EUI LIBRARY

- MAY 2017
- 5.0

1. Data discovery and EUI Library Data Portal
2. Data protection, database copyright and ethical use
3. Support, software and infrastructure
4. Research data management (RDM) and data management plans (DMPs)
5. Data management in EU Horizon 2020
6. The EUI ResData repository: metadata and preservation
7. Open Data
8. Qualitative data in humanities and social sciences
9. Library research data services
10. International resources and guidelines

1. DATA DISCOVERY AND EUI LIBRARY DATA PORTAL

This Guide provides information about both quantitative and qualitative data in digital formats. This section describes access to EUI Library licensed data resources, access to restricted data hosted by other institutions, and access to open data resources.

1(a) The EUI Library Data Portal

The Library maintains an indexed [Data Portal](#) providing access to both (i) licensed data resources for EUI members and (ii) open data resources relevant to EUI research themes. There are three sub-directories: macroeconomic data; micro-socioeconomic data and Europe-related data.

[Macroeconomic](#) databases provide statistics on national, regional and global economic and political-economic developments from international institutions and data publishers, including Thomson Reuters, ECB, Standard & Poor's; Eurostat, OECD, IMF and World Bank.

[Micro-socioeconomic](#) databases provide individual, family, household-level and company data observations. Major providers include GESIS, ICPSR, Bureau van Dijk, UKDS, Eurostat and DIW. Access to micro data on the EUI Library restricted data server requires additional registration.

[European, EU and Euro Area](#) data resources provide macroeconomic, micro-socioeconomic and financial data for research on pan-European topics, EU states, European sub-state regions, the Euro area and Europe in global context.

All resources indexed in the EUI Data Portal have an online resource guide 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 are accessible via internet protocol (IP) or – for micro-socioeconomic data – the Library's restricted data server. The

Datastream (Thomson Reuters) 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 upon request.

1(b) Access to data at other institutions

EUI members who require access to restricted data 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.¹ EUI members who require access to unpublished datasets (eg. underlying data associated with a publication) should contact the Library before writing to data creators/owners. In some instances, it may be possible for EUI members to obtain access via Library consortia.

1(c) Access to open data resources

Open data refers to the trend among international organisations, researchers and government agencies to share data outputs via the internet. Research data management and data management plans (Section 4 below) are the basis for determining whether, when, how, where and under what terms, research data outputs can be shared as open data. The re3data² research data registry and the data repositories section of the Open Access Directory³ provide lists of international research data repositories by discipline, data type and host location. Further information about open data is in Section 7 below.

2. DATA PROTECTION, DATABASE COPYRIGHT AND ETHICAL USE

This section provides an overview of terms and conditions of data access and use; special data protection requirements for micro-socioeconomic data; individual data user undertakings; and terms and conditions of use for open data.

2(a) License agreements and database copyright

Access to, and use of, data provided by the EUI Library is subject to license agreements, copyright terms and data protection provisions. Full details are on the Library's Terms and Conditions web page.⁴ Data users are individually responsible for compliance with terms and access of use. Violation of terms and conditions puts at risk other EUI members' future access to data resources. All EUI users must scrupulously abide by the terms and conditions of access to, and use of, data. Open data – freely accessible via the internet – are also subject to terms and conditions of use, and restrictions on re-publishing (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

¹ econlibrary@eui.eu

² <http://www.re3data.org/>

³ http://oad.simmons.edu/oadwiki/Data_repositories

⁴ <http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx>

permission of the licensor; users may not remove, obscure or modify any copyright or 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.

2(b) Data protection: micro-socioeconomic data access and use

Special terms and conditions apply to access and use of micro-socioeconomic data, reflecting the sensitive nature of data observations about human subjects, families and households. Such terms and conditions apply to both (i) micro-socioeconomic data hosted by the Library for EUI members and (ii) micro-socioeconomic data hosted by third-parties provided directly to EUI members under individual license. Terms and conditions for each micro dataset are given in the 'full details' section of the resource guides in the Library micro data directory.⁵ The EUI guide to Good Data Protection Practice in Research provides further information on data protection.⁶

Micro data users must preserve the confidentiality of observations pertaining to human subjects, families and households. Users must not attempt to identify any individual, family or household in a dataset. Anonymisation is explained in Section 4(c) below.

EUI members must use the online Library micro data access form⁷ when applying for access to micro-socioeconomic data hosted on the Library restricted server. A separate form is required for each dataset. Applicants must also sign (i) the EUI Library paper form 'Terms and Conditions of Use of Micro Data' and (ii) the data provider's terms and conditions paper form. Both forms can be signed at the Economics Information Office (Badia Library, 085) or at the Economics Departmental Library (Villa La Fonte, 035). 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 issuers directly for information about access at their home institutions.

Further information is provided in the OECD's guidelines on Research Ethics and New Forms of Data for Social and Economic Research (2016).⁸

2(c) Individual agreements with data providers

Some data issuers require that access contracts be established directly with data end-users. In such cases EUI members are required to sign a data user agreement with a

⁵ <http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/MicroDataSet.aspx>

⁶ <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

⁷ <http://www.eui.eu/Research/Library/RequestForms/Register.aspx>

⁸ http://www.oecd-ilibrary.org/science-and-technology/research-ethics-and-new-forms-of-data-for-social-and-economic-research_5jln7vnpxs32-en

third-party provider/licensor. If the data issuer also requests the counter-signature of an EUI administrator ('guarantor'), the Library will require the data user to sign an internal undertaking that he/she will abide by the terms and conditions of data access and use.

2(d) Open Data: terms and conditions of use

Open data, freely accessible via internet, are normally subject to terms and conditions of use and redistribution – 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 license, users should contact the data owner, or the EUI Library, before rehosting or redistributing sections of the data.

3. SUPPORT, SOFTWARE AND INFRASTRUCTURE

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

3(a) Library support

Data support is provided at the Badia Library (office 085) on weekday mornings and on Tuesday and Thursday afternoons. Data support in the Economics Department (Villa La Fonte) is provided on Monday, Wednesday and Friday afternoons from 14:45 to 18:30.

The EUI Library maintains the Data Portal, the Micro Data Restricted Server and the EUI ResData (beta) repository, and provides assistance for the discovery, access and use of EUI-licensed digital databases and internet-hosted open data. The Library also advises users on research data management (Section 4 below) and open data options (Section 7 below).

A directory of online research data software manuals, with full-text links, is available on the Library web site.⁹ Paper versions of data software manuals are available in the Badia Library and the Economics Departmental Library, at shelfmarks 001 to 005. The Library also holds a comprehensive collection of monographic works on statistical science and data methodology at shelfmarks 500 to 519. Books and manuals in any language may be suggested for acquisition by the Library.¹⁰

Every Friday during term, the Library issues an e-Bulletin with updates on new data releases, information on how to use Library and internet data resources, and developments in statistical science. EUI members can sign up for the weekly e-Bulletin with an @eui.eu account. Send a message with 'subscribe' in the title to econlibrary@eui.eu. Data news is also disseminated via the EUI Library Blog¹¹ and Twitter.¹²

⁹ <http://www.eui.eu/Research/Library/ResearchGuides/Economics/Software.aspx>

¹⁰ Write to econlibrary@eui.eu

¹¹ <https://blogs.eui.eu/library/category/economics>

¹² <https://twitter.com/econlibrary>

3(b) ICT Service 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.¹³ Technical support is provided at the site offices of the ICT Service. Advice on the use of statistical software is provided by the EUI research software tutors.¹⁴ The EUI provides access to major data software including Fortran, Gauss, MATLAB, OxMetrics, Python, R, Stata, Stat/Transfer, WinEdt, WinRATS, and provides support for the high-performance computing cluster at the EUI.

4. RESEARCH DATA MANAGEMENT (RDM) AND DATA MANAGEMENT PLANS (DMPs)

Research data management (RDM) encompasses the control of data inputs, the handling of data, the protection of data, and the creation of data outputs. Research data management is carried out by individual researchers, and members of research teams throughout the duration of a research project.¹⁵ Research data management covers the description of data and tools; the storage of data during analysis; the provision of clear and accurate metadata and supporting documentation; the preservation of data and – where possible – making research data outputs openly available. The main features of research data management are covered in data management plans (DMPs). Data management plans are increasingly required by science funding agencies.

4(a) Creating a data management plan

Data management plans are short documents outlining how data are handled, stored, documented, preserved, and – where possible – made available for sharing. DMPs provide information on:

- The creation of data and the sources of data
- How data is elaborated, collated and organised
- How data and ancillary elements are documented
- Where data is stored during research projects
- How data authorship and credit are assigned
- How data is preserved.

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

Data management plans can be used as the basis for determining whether, when, how, where and under what terms, research data outputs can be openly shared – or shared under more restrictive terms and conditions. Open data refers to the trend among

¹³ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/SoftwareResearch.aspx>

¹⁴ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/ResearchSoftwareTutoring.aspx>

¹⁵ The term ‘researcher’ in this context, includes Professors, Fellows and PhD Candidates.

international organisations, researchers and government agencies to share data outputs via the internet.

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 EUI Library for assistance.¹⁶ Tools such as DMPonline¹⁷ – the Digital Curation Centre’s data management planning tool – can be used to write a structured data management plan. To use DMPonline, enter an email address, name of organisation and create a password. EUI users should select ‘other organisation’ from the drop-down menu. First-time users are taken to the ‘edit profile’ section of the DMPonline platform. When using this resource to create a data management plan, it is possible to select a funder template (eg. EU Horizon 2020) which generates the appropriate matrix. Author(s) of the data management plan complete each free-text section of the plan – responding to the prompts. Editing rights can be shared with project collaborators by entering their emails and assigning status as ‘co-owner’, ‘editor’ or ‘read only.’

The Principal Investigator (P.I.) should be identified in the data management plan. For research teams, this should be the contact person for decisions regarding whether, when, how, where and under what terms, research data outputs generated by 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.

Data management plans should be updated over the course of the research project, incorporating descriptions of new data generation and use, changes in project policy, and changes in the composition of the research team or consortium. Many science funders require a revised DMP at the mid-point and at the end of the project.

4(b) Security of data during research projects

During the research project cycle, it is important to keep data secure. Researchers should use a desktop computer for data elaboration, and make regular backups on the EUI network server, or on a safely-secured external memory device. In accordance with contractual agreements, micro-socioeconomic data at the EUI can only be accessed and elaborated on a desktop computer in a secure location. Preliminary findings and associated documentation should be kept in locked storage when not in use. EUI members are welcome to submit their datasets for preservation in the Library’s ResData repository, launched in May 2017 (see Section 6 below).

4(c) Data protection and anonymisation

Data pertaining to human subjects, households or families are subject to data protection laws. Such data must be handled with particular care. Persons, families and households cannot be identifiable in any dataset. Researchers are responsible for obtaining the

¹⁶ econlibrary@eui.eu

¹⁷ <http://www.dcc.ac.uk/dmponline>

informed consent of subjects for the collection and processing of personal data. OECD guidelines (2016) state: “The default position should be that personal data is not collected, processed or shared without informed consent.”¹⁸ The EUI guide to Good Data Protection Practice in Research provides further information on data protection.¹⁹

Dataset creators are responsible for the anonymisation of sensitive data observations. Anonymisation techniques include: data masking (partial data removal and data quarantining); pseudonymisation; aggregation (cell suppression, inference control, perturbation, rounding, sampling, synthetic data, tabular reporting); and derived data items and banding.²⁰

A new European Union General Data Protection Regulation will come into effect in 2018. Details will be incorporated in the 6th edition of this Guide. The web site of the EU Article 29 Working Party on Data Protection provides further information.²¹

4(d) Documentation and codebooks

In order to render data for future use, it is important that researchers keep a record of data inputs, usage, elaboration and code throughout the project cycle. Documentation includes: notebooks, questionnaires, codebooks, data dictionaries, software syntax, database schema and notes on methodology. The MANTRA data management training site provides step-by-step data management guidelines for researchers.²²

5. DATA MANAGEMENT IN EU HORIZON 2020 PROJECTS

Horizon 2020 is the European Union Research and Innovation funding programme for the period 2014-2020. 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’ category. The EUI Library assists project administrators with the data management sections of funding applications, as well as in-project data management plan updates.

When completing the general information section of H-2020 applications, EUI project managers should address the following questions – provided by the European Commission:

- What types of data will the project generate/collect?
- What standards will be used?

¹⁸ http://www.oecd-ilibrary.org/science-and-technology/research-ethics-and-new-forms-of-data-for-social-and-economic-research_5jln7vnpxs32-en p.7

¹⁹ <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

²⁰ Anonymisation techniques are described in appendix 2 of the Anonymisation Code of Practice, UK Information Commissioner's Office (2012) <https://ico.org.uk/media/1061/anonymisation-code.pdf> The UKDS provides a guide to anonymisation for both quantitative and qualitative data at: <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation>

²¹ http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=50083

²² <http://datalib.edina.ac.uk/mantra/>

- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?²³

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 contract period. 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.'²⁴ The DMPonline²⁵ tool can also be used for H-2020 data management plans – by selecting the H-2020 template in the Funder section (see Section 4(a) above).

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 Guidelines²⁶, p.9.] The EC suggests attaching a license; eg. CC-BY International.²⁷ The EC does not recommend specific repositories. EUI members may submit data outputs to the EUI ResData repository (see next section). Further details about data management in Horizon 2020 are provided in these official documents:

- Guidelines on Data Management in Horizon 2020²⁸
- Guidelines on FAIR Data Management in Horizon 2020²⁹
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020³⁰
- Fact Sheet: Open access to publications and data in Horizon 2020³¹

6. THE EUI RESDATA REPOSITORY: METADATA AND PRESERVATION

In May 2017, the Library launched the [EUI ResData](#) repository (beta) – a solution for the sharing and archiving of EUI research data outputs. EUI ResData is based on the DSpace infrastructure.³² This section provides information on how to reposit research data outputs for long-term preservation and – where possible – to share data with other researchers. Sub-section 6(b) below explains how to create metadata – data about data – which is essential for data repositing, retrieval and reuse.

²³ Guidelines on Data Management in Horizon 2020, December 2013, p.2.

²⁴ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

²⁵ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

²⁶ <http://www.dcc.ac.uk/dmponline>

²⁷ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

²⁸ <https://creativecommons.org/licenses/>

²⁹ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

³⁰ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

³¹ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

³² https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf

<http://www.dspace.org/>

6(a) The EUI ResData repository (beta)

EUI members who wish to submit their datasets for inclusion in EUI ResData (beta) should complete the online form on the Library web site.³³ This first step is to provide metadata – descriptive information about the dataset. Library staff will make an appointment for data transfer upon receipt of the completed form. By submitting the form, EUI members acknowledge that the dataset for reposit is the output of original data collection and elaboration; or is the output of significant, value-added, elaboration of pre-existing sources.

Datasets presented for inclusion in the EUI ResData repository must be the output of research by a current EUI member or research team. The name of the Principal Investigator, researcher(s), and – if applicable – technical collaborator(s) who created the dataset, must be provided. EUI email contacts must 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 Principal Investigator).

Dataset creators must certify that their work complies with the Code of Ethics in Academic Research of the European University Institute.³⁴ EUI members will be required to sign a declaration that they are creators of the dataset being presented for reposit, and that the dataset is the output of original data collection and elaboration; or is the output of significant, value-added, elaboration of pre-existing sources. The source(s) of the data must be indicated. If the dataset is the output of original data collection and elaboration, details must be provided. If the dataset is derived from pre-existing sources, those sources must be clearly indicated (data creator, institutional source, publisher).

EUI members submitting a dataset for inclusion in the EUI ResData repository must state whether the data can be shared as open data or not – taking into consideration data protection and data copyright. Submitters should indicate if datasets presented for reposit are to be subject to embargo.

Persons, families and households cannot be identifiable in any dataset. Depositors are responsible for obtaining the informed consent of subjects for the collection and processing of personal data. Dataset creators are responsible for the anonymisation of data observations (see Section 4(c) above).

Creators of research data outputs which are elaborated from pre-existing copyrighted sources may need to seek permission from data rights' owners before open sharing. It is not possible to publish a dataset containing significant portions of data sourced from pre-existing databases governed by contractual license. The EUI Library can provide advice where necessary.

The Library can also assist researchers to deposit datasets in discipline-specific data repositories and the international Zenodo repository.³⁵ EUI-generated research datasets are indexed in the EUI Research Data Registry.³⁶ Major data repositories are indexed in

³³ <http://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow.aspx>

³⁴ <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/CodeofEthicsinAcademicResearch.pdf>

³⁵ <http://zenodo.org/about>

³⁶ <http://www.eui.eu/Research/Library/ResearchDataServices/EUIresearchdata.aspx>

the international re3data registry³⁷ and the data repositories' section of the Open Access Directory.³⁸

6(b) Metadata

Metadata are data about data, presented in a systematic scheme. Accurate metadata are necessary for the repositing and sharing of datasets. Throughout the research project, it is important to keep a detailed and updated record describing data capture, use and elaboration. An introduction to metadata standards for social science and humanities' data is maintained by the Digital Curation Centre.³⁹

During the course of research projects, researchers should maintain and preserve documentation, code, software and tools used to generate datasets.

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.

The EUI ResData repository uses Dublin Core metadata.⁴⁰ EUI members submitting a dataset for inclusion in the EUI ResData repository should first complete the Library's online metadata form.⁴¹ An appointment for data transfer will subsequently be made by Library staff.

These are the principal metadata fields:

NAME(S) OF DATASET CREATOR(S)

The name, or names, of the researchers and technical collaborators who created the dataset must be provided. The name of the Principal Investigator must be given if the dataset has been created by a team. Guidelines on authorship and credit for research outputs are provided by CASRAI.⁴² If the project is undertaken in the context of a consortium, the name of the Data Manager must be provided (if different from the P.I.). Where researcher ID numbers are available, these should be provided (eg. ORCID).

EMAIL CONTACT(S)

The EUI email of the dataset creator(s) must be provided.

TITLE OF DATASET

The title should succinctly convey the nature and scope of the dataset.

DESCRIPTION OF DATA

A meaningful data abstract, indicating the kind of data, and the research context must be provided. A brief note on methodology should be provided.

³⁷ <http://www.re3data.org/>

³⁸ http://oad.simmons.edu/oadwiki/Data_repositories

³⁹ <http://www.dcc.ac.uk/resources/subject-areas/social-science-humanities>

⁴⁰ <http://dublincore.org/metadata-basics/>

⁴¹ <http://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow.aspx>

⁴² <http://casrai.org/credit>

SOURCE(S) OF DATA

The source(s) of the data must be indicated. If the dataset has been *generated during a research project*, this should be indicated, with details of data collection (eg. survey parameters). If the dataset is *derived from a pre-existing database*, all source(s) must be clearly cited; eg. institution, publisher etc. If there are multiple sources – all must be cited.

TYPE OF DATA

The type of data must be indicated: eg: statistical; observational; computational; experimental; simulational etc.

YEAR OF COMPLETION OF DATASET

The date of completion of the dataset must be provided. If part of a data series, this should be indicated.

DATE-RANGE COVERAGE OF DATASET

The start- and end-dates of dataset coverage must be provided.

GEOGRAPHICAL COVERAGE OF DATASET

If applicable, the geographical scope of the dataset (national, regional, global etc.) should be indicated.

FORMAT OF DATA

The software format and version must be given (eg. Stata 14, .csv, Excel, .txt etc.).

CODEBOOK / SUPPORTING DOCUMENTATION

Codebooks and supporting documentation should be provided.

ACCESS STATUS

The status of access to the data must 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 expiry date of the embargo should be indicated.

LICENSE

Terms and conditions of access and use of research data outputs by other persons should be stated. It is advisable to provide a license (eg. CC-BY International).⁴³

DATASET CITATION

A short 'ready-to-use' citation reference for the dataset should be provided, incorporating core descriptive elements.

DIGITAL IDENTIFIER

The Library will assign a unique object identifier to the dataset, for locating, linking and citation.

LANGUAGE

Where applicable, multi-lingual documentation, tags, questionnaires and variable descriptions should be provided.

⁴³ <https://creativecommons.org/licenses/>

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, if any, should be listed with links to abstracts and, where possible, full-texts.

PROJECTED FUTURE WAVES OF DATASET

If it is intended to generate a future iteration of the dataset, details should be provided.

7. OPEN DATA

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

Open data resources are available via the internet. Major data repositories are indexed in the international re3data registry⁴⁴ and the data repositories section of the Open Access Directory.⁴⁵

By carefully noting the metadata elements suggested in the previous section, researchers 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. Outputs should be either (i) original datasets generated during the research project or (ii) datasets which are 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:** Persons, families and households cannot be identifiable in any dataset. Depositors are responsible for obtaining the informed consent of subjects for the collection and processing of personal data. Dataset creators are responsible for the anonymisation of data observations (see Section 4(c) above). The EUI guide to Good Data Protection Practice in Research gives further information on data protection.⁴⁶
- **Database copyright:** Research data outputs which are elaborated from pre-existing copyrighted sources may require permission before being repositied and shared. It is not possible to publish a dataset containing significant portions of data sourced from pre-existing databases governed by contractual license. The Library will provide advice where necessary.⁴⁷

⁴⁴ <http://www.re3data.org/>

⁴⁵ http://oad.simmons.edu/oadwiki/Data_repositories

⁴⁶ <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

⁴⁷ econlibrary@eui.eu

Research data outputs can be preserved under a variety of access terms and conditions. Access status may change over time. Data can be made openly available for all users via the internet; data can be subject to pre-access registration terms; data can be subject to user contract (sometimes requiring a project proposal); data can be embargoed for a defined period (or indefinitely) and data can be restricted to on-site access and use. Data can also be repositied solely for preservation purposes (dark archive).

8. QUALITATIVE DATA IN HUMANITIES AND SOCIAL SCIENCES: ACCESS AND USE

This section treats access to, and use of, qualitative data in the humanities and social sciences. Examples of qualitative data include; minable text, transcripts of interviews; images; audio and video recordings; survey diaries; archival material; field notes; and free-text answers to surveys. The definition of 'data' varies across academic disciplines – especially where there is a mix of qualitative and quantitative methods. It is important that project data planning be located in the culture of the discipline in which the research is undertaken.⁴⁸ Non-numerical data is subject to most of the same terms and conditions of access, and use, that apply to quantitative data. Research projects can incorporate a mix of both qualitative and quantitative approaches, and in many cases qualitative data can be processed and expressed numerically.

8(a) Access and terms and conditions of use: qualitative data

The handling, use and sharing of qualitative data in the social sciences and humanities is subject to strong ethical considerations and standards. When accessing qualitative data, it is important for researchers to familiarise themselves with the terms and conditions of access, and use, as indicated by the holding institution and/or rights owner. If a digital database is being generated from non-digital materials, it is important to obtain the consent of subjects or rights' holders in advance of inclusion.

Qualitative data can be generated from surveys, free-text responses to interview questions, focus group recordings or experimental simulations. In all cases, subjects should be informed of their rights as established by jurisdictional data protection legislation, and best-practice guidelines from scholarly societies in the relevant discipline(s). Researchers are responsible for obtaining the informed consent of subjects for the collection and processing of personal data.

Due to the personal nature of many qualitative data observations, researchers should pay particular attention to ethical standards when handling such data. Human subjects, families and households cannot be identifiable in any dataset. Researchers are responsible for obtaining the informed consent of subjects for the collection and processing of personal data; and for the anonymisation of data observations. The linking of variables on gender, religion etc. to individuals, families or households, is governed by data protection legislation and academic best-practice.

⁴⁸ See, for example, the EUI SPS methods' directory:

<http://www.eui.eu/DepartmentsAndCentres/PoliticalAndSocialSciences/ResearchAndTeaching/MethodsAdvisers.aspx>

8(b) Support, software and infrastructure for qualitative data

Support for qualitative data use and elaboration is provided by the EUI Library. Software support is provided by the EUI ICT Service.⁴⁹ Many of the tools used for the analysis of quantitative data (eg. Gauss, MATLAB, Python, R, Stata) can also be used for qualitative data analysis – especially if the data is given a numerical expression, or if aggregate statistical observations are drawn. ArcGIS and ATLAS.ti can be used for analysis, mapping and visualisation of qualitative non-numerical data such as audio, graphics, text and video. Coding Analysis Toolkit (CAT) can be used for content and discourse analysis. Tools for data backup (SyncToy), file zipping (7-Zip), data encryption (TrueCrypt) and image adjustment (Resizer) are also available.⁵⁰

The use of some tools for the analysis and presentation of restricted personal data may require researchers to work in a 'safe-room' environment. If this is stipulated by a data provider or project funder, contact the EUI Library for support.

8(c) Research data management and data management plans for qualitative data

Although research data management for qualitative data is similar to research data management for quantitative data – there are some specific additional considerations. Research data management (RDM) encompasses the control of data inputs, the handling of data, the protection of data, and the creation of data outputs (see Section 4, above). Research data management is carried out by individual researchers, and members of research teams throughout the duration of research projects. During data analysis work, qualitative data materials should be carefully handled and secured, either in a locked storage unit or in a locked room. This is particularly important for confidential, unique and archival material being collated and/or ingested for the purposes of creating a dataset.

Data management plans (DMPs) are increasingly required by science funding agencies (see Section 4(a) above). Due to the heterogeneous, multi-media and complex nature of qualitative data in the humanities and social sciences, it is particularly important that researchers keep a record of data sources, and retain notebooks, questionnaires, codebooks and multilingual thesauri used during research projects. Supporting documentation helps towards the creation of accurate metadata for the reposit, preservation, retrieval and reuse of datasets.

In the case of non-repeatable, time-sensitive, socio-political research, data management plans may require a detailed explanation of the qualitative methods used.

8(d) Metadata for qualitative data

Metadata are data about data, presented in a systematic scheme (see Section 6, above). Metadata fields used for quantitative data outputs can also be used for qualitative data outputs. However there are some additional considerations.

⁴⁹ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/Index.aspx>

⁵⁰ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/SoftwareResearch.aspx>

In addition to the name(s) of researchers and technical collaborators who generate a dataset, it may be necessary to include the authors/creators of subsidiary qualitative data. The dates of creation of subsidiary works included in any new qualitative dataset should be clearly indicated. Linguistic, national and regional metadata should be provided where relevant (eg. multi-lingual surveys). The format and version of software used to elaborate the data should also be indicated.

8(e) Data preservation, repositing and open qualitative data

Qualitative digital data outputs in the humanities and social sciences can be repositing in the EUI ResData repository (see Section 6 above) or in a subject repository, or in a multi-disciplinary repository. Access control is particularly important for qualitative data relating to human subjects, families and households. Researchers should adhere to the standards of the repository, the project funder and the discipline's academic best-practice when determining whether and how to share qualitative data outputs. Data can be shared under open data license (eg. CC-BY International⁵¹) or under more restrictive terms and conditions. Researchers should refer to funding contracts, and to guidelines provided by scholarly societies, when determining whether, and under what terms, qualitative research data outputs can be shared.

9. LIBRARY RESEARCH DATA SERVICES

The following services are provided by the EUI Library:

- Data discovery: the EUI Library Data Portal
- Data protection, database copyright and terms and conditions of use
- Access to the EUI Library restricted server for micro-socioeconomic data
- Preservation of datasets in the EUI ResData repository
- Research data management (RDM) and data management plans (DMPs)
- Data management in EU Horizon 2020
- Metadata for research data outputs
- Open data and guidelines for sharing
- Data user undertakings for access to third-party data resources
- Weekly Library data e-Bulletin
- Badia Library helpdesk: BF-085 (right side, entry-level floor), every weekday morning and Tuesday and Thursday afternoons (tel. 2346)
- Economics Department helpdesk: VLF-035 (second floor): Monday, Wednesday and Friday afternoons (tel. 2904)
- Write to Thomas Bourke, econlibrary@eui.eu for research support.

10. INTERNATIONAL RESOURCES AND GUIDELINES

Information on research data management, open data, and data management plans is available from these sources:

- [American Economic Association guidelines on data availability](#)
- [Big Data Europe - Big Data Aggregator Platform](#)

⁵¹ <https://creativecommons.org/licenses/>

- [Curating Research Assets and Data using Lifecycle Education \(CRADLE\)](#)
- [Data Access and Research Transparency \(DART\)](#)
- [Data User Agreements Directory \(University of Michigan\)](#)
- [Digital Curation Centre: How to develop a data management and sharing plan](#)
- [EUDAT collaborative data infrastructure](#)
- [EUI Research Data Registry](#)
- [European Commission Report on Open Research Data](#)
- [European Data Portal](#)
- [European Open Science Cloud](#)
- [GESIS repository - Leibniz Institute for the Social Sciences](#)
- [GitHub software collaboration](#)
- [Good Data Protection Practice in Research](#)
- [Göttingen/OpenAIRE Study on the Protection of Research Data](#)
- [Horizon 2020 Guidelines on Data Management](#)
- [How and Why You Should Manage Your Research Data \(JISC\)](#)
- [IP Rights in Data handbook \(DLA Piper\)](#)
- [Managing and Sharing Data \(UK Data Service\)](#)
- [MANTRA research data management training](#)
- [Metadata for Social Science & Humanities \(Digital Curation Centre\)](#)
- [OECD Guidelines on Research Ethics & New Forms of Data for Social & Economic Research](#)
- [Open Data Handbook - Open Knowledge Foundation](#)
- [Open Economics Principles - Open Knowledge Foundation](#)
- [OpenAire project and network](#)
- [re3data.org registry of data repositories](#)
- [RECODE – Policy RECommendations for Open access to research Data in Europe](#)
- [Research Data Alliance \(Europe\)](#)
- [Research Data Curation Bibliography \(C.W. Bailey Jr.\)](#)
- [State of the art report on open access publishing of research data in the humanities](#)
- [The Hague Declaration on Knowledge Discovery in the Digital Age](#)
- [Where to Keep Research Data Checklist \(DCC\)](#)
- [Zenodo data repository \(CERN/OpenAire\)](#)

An online version of this document is available:

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

econlibrary@eui.eu

Thomas Bourke

May 2017