Research Data Guide
EUI Library
6th edition
March 2018
INTRODUCTION

1. Data discovery and the EUI Library Data Portal
2. Data protection, database copyright and ethical use
3. Data management plans (DMPs)
4. Managing data during the research project cycle
5. The EUI ResData repository: preserving and sharing data
6. Research Data in EU Horizon 2020
7. Open Data
8. Qualitative data in the humanities and social sciences
9. Infrastructure, software and support
10. International research data guidelines

INTRODUCTION

The structure of this Guide broadly reflects the ‘research data cycle’, encompassing data discovery, data management planning, data use, data protection, data preservation and open data. The Library has launched a new repository, called ResData, to provide an infrastructure for the preservation and sharing of EUI research data outputs. ResData is fully compliant with the data management requirements of the EU Horizon 2020 programme.

Although the trend towards open data is gaining momentum, scholars also require access to copyrighted resources which are only accessible under more restrictive terms. For this reason, the Guide begins with data discovery and the EUI Library Data Portal – presenting the concepts of data protection, database copyright and ethical use. The focus of the Guide is on quantitative and qualitative digital data in the social sciences and humanities.

1. DATA DISCOVERY AND THE EUI LIBRARY DATA PORTAL

This section describes access to 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\(^1\) which (i) provides access to licensed data resources for EUI members and (ii) introduces open data resources of relevance to EUI research themes. All resources in the Data Portal have an individual guide with:

- Data description and scope
- Time-period coverage and release/wave information
- Support links (online manuals, software transfer routines, user networks)
- Terms and conditions of access and use.

---

There are three sub-directories. Macro databases provide statistics on global, regional and national economic and political-economic developments from international organisations and major data publishers, including Databanks, ECB, Eurostat, Global Financial Data, IMF, OECD, Standard & Poor’s, Thomson Reuters, World Bank, UN and WIW. Micro-socioeconomic databases provide individual, family, household and company observations. Major providers include DIW, Eurostat, GESIS, ICPSR, Moody’s Analytics/BvD and UKDS. Access to micro data on the EUI Library restricted data server requires additional registration. European, EU and Euro area resources provide data for research on pan-European topics, EU states, European sub-state regions and Europe in global context.

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. Modes of access are described in the Data Portal resource guide for each resource.

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 project, because applications can be lengthy. Some qualitative and micro geo-coded data can only be accessed at issuers’ secure facilities. Contact the Library for assistance with access applications. 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 scholars, government agencies and international organisations to share datasets, codebooks and software via the internet. Open research datasets can be located via the re3data registry of research data repositories, which is indexed by discipline, sub-discipline, data type and host location. The Open Access Directory and internet search engines also help locate open data. Data management plans (Section 3 below) can be used to determine whether, when, how, where and under what terms, research data outputs can be shared as open data. The research data outputs of EUI members can be reposited in the ResData repository for preservation and sharing (Section 5).

---

2 https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/MacroDataSet
3 https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/MicroDataSet
4 https://www.eui.eu/Research/Library/RequestForms/Register
5 https://www.eui.eu/Research/Library/ResearchGuides/Economics/EuropeanEUandEurozoneStatistics
7 econlibrary@eui.eu
8 In this document, the term ‘scholar’ includes researchers, fellows, professors and all other scientific collaborators on research projects.
9 http://www.re3data.org/
10 http://oad.simmons.edu/oadwiki/Data_repositories
2. DATA PROTECTION, DATABASE COPYRIGHT AND ETHICAL USE

This section provides an overview of data protection and ethical use requirements for micro-socioeconomic and qualitative data; database copyright; and individual data user undertakings. Further information is provided in the EUI Guide to Good Data Protection Practice in Research.11

2(a) Data protection

Special terms and conditions apply to access and use of micro-socioeconomic and qualitative data. This reflects 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 hosted by the Library are given in the ‘full details’ section of the individual resource guides in the Library’s Micro Data Directory.12

Persons, families and households cannot be identifiable in any dataset. When using micro datasets, users must not attempt to identify any individual, family or household.

The storage, transfer and use of personal data is subject to data protection rules.13 The processing of particularly sensitive data (referring, for example, to ethnicity, health, political or sexual orientation, religion and biometrics) requires even stricter safeguards.14 Those rules are currently under revision in light of the new European Union General Data Protection Regulation which comes into effect on 25 May 2018. As written in the EUI Guide on Good Data Protection Practice: “As a general rule, the processing of such personal data is prohibited. However, Article 7 of the EUI’s Data Protection Policy does allow it to be lawfully processed in specific circumstances. The most common in research is upon the data subject’s explicit consent.”15

When collecting (eg. via interviews or forms), storing, using or transferring personal data, scholars are responsible for obtaining the specific informed consent of subjects (Section 4(e) below) and for preserving the confidentiality of data observations pertaining to human subjects, families and households. Consent must be intelligible: it should refer clearly and precisely to the scope and the consequences of the data processing, and the retention period. Blanket consent without specifying the exact purpose of the processing is not acceptable.

EUI members must use the online micro data form16 when applying for access to micro-socioeconomic data hosted on the Library’s restricted server. A separate form is required for each dataset. Applicants must also sign (i) the EUI ‘Terms and Conditions of Use of

16 http://www.eui.eu/Research/Library/RequestForms/Register.aspx
Micro Data’ paper form, 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). Some providers (eg. Eurostat) require the submission of a detailed project proposal before access is granted.

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. Further information is provided in the EUI Guide to Good Data Protection Practice in Research\(^\text{17}\) and the OECD’s guidelines on Research Ethics and New Forms of Data for Social and Economic Research.\(^\text{18}\)

2(b) License agreements and database copyright

Access to, and use of, databases provided by the EUI Library are subject to contractual license agreements and copyright terms. Full details are on the Library’s Terms and Conditions’ web page.\(^\text{19}\) Data users are individually responsible for compliance with terms and conditions of access of use. Violation of license terms 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 provided under license to the EUI community.

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 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.\(^\text{20}\)

2(c) Agreements between individual users and data providers

Some data issuers require that access contracts be established directly with end-users. If a data issuer also requests the counter-signature of an EUI administrator (‘guarantor’), contact the Library. Upon completion of the access application, the Library will request that the EUI data user sign an internal undertaking to abide by the terms and conditions of access and use.\(^\text{21}\)

3. DATA MANAGEMENT PLANS (DMPs)

Data management plans are short documents, normally required by science funding agencies. Research data management is carried out by scholars throughout the duration of a research project. Data management plans should address:

---


\(^{19}\) [http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx](http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx)

\(^{20}\) [http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx](http://www.eui.eu/Research/Library/ElectronicResources/TermsAndConditions.aspx)

\(^{21}\) econlibrary@eui.eu
- How data is generated and/or sourced
- How data is used, elaborated and organised
- How data, and data subjects, are protected
- How data, tools and ancillary elements are described and documented
- How data is stored and secured, and how long it will be retained
- How data authorship and credit are assigned
- How data is preserved
- How, whether, and under what terms, research data outputs can be shared.

During the research project it is important to keep an accurate record of changes to dataset design, folders, files, variables and versioning. Scholars undertaking experiments should ensure that pre-agreed terms of disclosure are documented and preserved. Codebooks and other documentation should describe the data, the methodology and the metadata – also known as ‘data about data’ (Section 5(c) below).

Data management plans (DMPs) 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 scholars, government agencies and international organisations to share data outputs and ancillary materials via the internet. The FAIR data principles (findable, accessible, interoperable and re-usable) should be considered during the preparation, and revision, of data management plans. Further information is in the European Commission guide to the FAIR data principles.\(^{22}\)

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.\(^{23}\) Tools such as DMPonline\(^{24}\) – the Digital Curation Centre’s data management planning tool – can be used to prepare a structured data management plan, complying with (eg.) the EU Horizon 2020 mandate. 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. The Principal Investigator (P.I.) should be identified in the data management plan. For research teams, the P.I. can assign co-author rights to the data management plan, by entering email addresses and assigning status as ‘co-owner’, ‘editor’ or ‘read only.’

The Principal Investigator should be the contact person for decisions regarding whether, when, how, where and under what terms, research data outputs might be openly shared. If the Principal Investigator is not the same person as the project Data Manager, this should be stated. In international collaborative projects, the name of the person who has final authority regarding decisions on the sharing of data outputs should be indicated. Unless otherwise stated, the Principal Investigator undertakes this role.

Science funders normally require a revised DMP at the mid-point of the research project and a definitive DMP as a project deliverable. It is important to update plans throughout

\(^{23}\) econlibrary@eui.eu
\(^{24}\) [http://www.dcc.ac.uk/dmponline](http://www.dcc.ac.uk/dmponline)
the research project by incorporating information about changes in the scope of the project, new data generation and use, and changes in the composition of the research team or consortium.

4. MANAGING DATA DURING THE RESEARCH PROJECT CYCLE

Data should be carefully managed throughout the duration of the research project. Particular attention should be given to data input; quality control; dataset design; folder structure; file structure; file naming; variable naming; software format; documentation and codebooks; security and backup during the project; data protection: in particular, obtaining and documenting informed consent of data subjects25; anonymisation26; copyright; preservation; archiving; data citation and – where possible – sharing data outputs as open data. Three guides are particularly useful for scholars in the social sciences and humanities: Expert Tour Guide on Data Management by the Consortium of European Social Science Data Archives (2017)27, Research Data Lifecycle by the UK Data Service (2017)28 and Guide to Social Science Data Preparation and Archiving by the Inter-university Consortium for Political and Social Research (2012).29

4(a) Data input and quality control

The generation and/or sourcing of data should be informed by high-quality control standards. Datasets which are generated from empirical work during a research project should adhere to the ethical best-practices of the discipline or sub-discipline. In cases where research datasets are partially based on pre-existing resources, scholars should ensure that the underlying sources are of high quality, and that there are no infringements of copyright. Contact the Library for assistance. The ingestion of observations into a research dataset should be undertaken with close attention to accuracy, completeness and consistency. Data codes and variables should be assigned in a logical, accurate and sustainable manner.

4(b) Folders, files, variables, format and versioning

The design of the research dataset should be carefully considered at the outset of the project. Dataset design varies by conventions of the discipline and sub-discipline, medium, types of variables, units of analysis, methodology, relationship between data elements, and whether or not the dataset is part of a series. Clear and consistent metadata for folders, files, variables and versioning, helps facilitate future data retrieval, reuse and replicability.

27 www.cessda.eu/DMGuide
28 https://www.ukdataservice.ac.uk/manage-data/lifecycle The UKDS is developing a tool for quality control of social science datasets, QAMYData (Quality Assessment for My Data), due for release in late 2018. Further information is at: https://www.ukdataservice.ac.uk/about-us/our-rd/qamydata The Horizon 2020 project OpenUP Hub is developing an Open Peer Review for Research Data https://www.openuphub.eu/community/blog/item/openup-pilot-2-open-peer-review-for-research-data
29 https://www.icpsr.umich.edu/files/deposit/dataprep.pdf
The folder structure of the research dataset should be considered at the beginning of the project (eg. hierarchical or horizontal). Qualitative datasets containing text, interviews, photographs &c. may require individual files for each element. File names should be standardised, eg: date, descriptor, version. Variables, such as age, country and sex should be clearly tagged, avoiding special characters and spaces. Temporary identifiers should be removed from the schema. Different file versions should be systematically named, using a standardised date system (Y-M-D) or version numbering.

The software format of the dataset should facilitate flexible use of the data. Scholars may consider a different format for final reposit, taking into consideration accessibility via open source (if the dataset can be shared as open data). For reposit in EUI ResData, large datasets should be submitted in in archive format (eg. .zip). It is not possible to create multiple ResData entries for individual elements of a dataset. Subsequent waves of a dataset can be assigned new entries in ResData (Section 5 below).

4(c) Documentation and codebooks

Clear and accurate documentation should be provided about the purpose and context of the research project, and about the research data output. Good documentation makes datasets findable, accessible, interoperable and re-usable (FAIR data principles). Documentation should include a detailed description of the dataset, providing information about folders, files, variables, versioning, and – where applicable – information about problematic values, missing observations and weightings. Codebooks, questionnaires and data dictionaries should be included. A concise note on methodology, or methodologies, should be given. Information should be provided on how and when the data was generated and/or collected; how the data was collated and elaborated during the research project; and how the dataset might be used for future research. Documentation in the EUI ResData repository is included in a ‘Docs’ dossier in PDF/A format, or plain text, for long-term preservation.

4(d) Security and backup during the research project

During the research project it is important to keep data secure at all times. Scholars 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 to facilitate recovery. 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, after completing the required Library registration protocols. Preliminary findings and associated documentation should be kept in locked storage when not in use. The EUI ICT Service provides a strong password policy to prevent the loss, exposure, or corruption of sensitive information.

---

30 https://www.ukdataservice.ac.uk/manage-data/format/organising
32 https://www.icpsr.umich.edu/files/deposit/dataprep.pdf
33 https://www.eui.eu/Research/Library/RequestForms/Register
34 https://www.eui.eu/ServicesAndAdmin/ComputingService/PolicyDocuments/StrongPasswordPolicy
4(e) Data protection, informed consent of data subjects and anonymisation

The data protection requirements for data use, described in Section 2(a) above, also apply to data generation during the research project cycle. OECD guidelines recommend: “The default position should be that personal data is not collected, processed or shared without informed consent.” The mode of consent obtained from subjects depends on the nature of the research project, the kind of data collected, and how and when the data will be used. Where possible, it is recommended to obtain written consent, using a template appropriate to the discipline or sub-discipline in which the research is being conducted.

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, rounding, sampling &c.); and derivative data items and banding.

Copyright considerations for data access and use, described in Section 2(b) above, also apply to in-project data management. It is not possible to publish a dataset containing significant portions of data sourced from pre-existing copyrighted databases governed by contractual license. The Library provides advice on data protection and copyright.

5. THE EUI RESDATA REPOSITORY: PRESERVING AND SHARING DATA

The Library launched the EUI ResData repository in October 2017. ResData is a solution for the preservation and sharing of EUI research data outputs. This section provides information on preparing research data outputs for long-term preservation and – where possible – sharing with other users. Section 5(c), further below, explains how to create metadata – data about data – which are essential for repot, retrieval and reuse.

5(a) Preparing data for repot in EUI ResData

Scholars presenting datasets for inclusion in the ResData repository should pay particular attention to data quality control and dataset structure. It is not possible to repot unstructured data elements and observations in ResData. Large datasets should be submitted in archive format (eg. .zip). Data should be submitted in original file format version. It is not possible to create multiple ResData entries for sub-sections of a dataset. Sub-sections should be accommodated in the folder structure. Documentation should include a concise overview of the research project and methodology. Codebooks and other relevant documentation should be provided in sustainable PDF/A format, or plain text, for preservation in a ‘Docs’ dossier.

37 econlibrary@eui.eu
5(b) Completing the online EUI ResData submission form

EUI members submitting a dataset for inclusion in the EUI ResData repository should first complete the Library’s online metadata form. Library staff will use information from the form to generate dataset-level metadata in the Dublin Core schema. An appointment for data transfer will subsequently be made by Library staff.

By submitting the online form, EUI members acknowledge that the dataset for reposit is the result of original data generation; 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 at least one EUI member of a research team. The name of the Principal Investigator, researcher(s), and – where applicable – technical collaborator(s), must be provided. EUI email contacts must be given. 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).

By completing the ResData submission form, EUI dataset creators certify that their work complies with the Code of Ethics in Academic Research of the European University Institute. The source(s) of the data must be indicated. If the dataset is the output of original data generation, details must be provided. If the dataset is derived from pre-existing sources, those sources must be clearly indicated (eg. data creator, institutional source and/or publisher).

EUI members submitting to ResData should state whether or not the dataset can be shared as open data. The Library can offer guidance on data protection and database copyright. Submitters should indicate if datasets presented for reposit are to be subject to embargo. In some cases it may be possible to create a limited public version of a larger restricted dataset.

Persons, families and households cannot be identifiable in any dataset. The collection and use of observations relating to ethnicity, health, orientation, religion, biometrics &c. are subject to data protection laws. 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 (Section 4(e) above). Creators of research data outputs which have been elaborated from pre-existing copyrighted sources may need to obtain the permission of rights’ owners before open data sharing. It is not possible to publish a dataset containing substantial portions of data sourced from pre-existing databases governed by contractual license. The EUI Library can provide advice, and can also assist scholars to reposit datasets in discipline-specific data repositories and the multi-disciplinary Zenodo repository. Major data repositories are indexed in the international re3data registry and the data repositories’ section of the Open Access Directory.

39 https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow
40 http://dublincore.org/metadata-basics/
42 http://zenodo.org/about
43 http://www.re3data.org/
44 http://oad.simmons.edu/oadwiki/Data_repositories
5(c) Metadata

Metadata are data about data, presented in a systematic schema. Accurate metadata are necessary for the organisation, use, repositing, sharing and machine-discovery of datasets. Throughout the research cycle, it is important to keep a detailed and updated record of data capture, input, use and elaboration. An introduction to metadata standards for social science and humanities’ data has been published by the Digital Curation Centre.45

Metadata elements 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 consistent during the research project. Some research data outputs may require multi-lingual metadata.

EUI members submitting a dataset for inclusion in the EUI ResData repository should first complete the Library’s online submission form.46 These are the principal metadata fields:

NAME(S) OF DATASET CREATOR(S)
The name, or names, of the scholars 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 research team. Guidelines on authorship and credit for research outputs are provided by CASRAI.47 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, eg. ORCID, these should be provided.

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 abstract, describing the data, the purpose of the research project and the methodology, must be provided.

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

TYPE OF DATA
The type of data must be indicated: eg: statistical; textual; observational; computational; experimental or simulational.

YEAR OF COMPLETION OF DATASET

45 http://www.dcc.ac.uk/resources/subject-areas/social-science-humanities
46 https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow
47 http://casrai.org/credit
The date of completion of the dataset must be provided. If part of a data series/wave, this should be indicated.

**DATE-RANGE COVERAGE OF DATASET**
The start- and end-dates of dataset coverage must be provided.

**GEOGRAPHICAL COVERAGE OF DATASET**
Where applicable, the geographical scope of the dataset (national, regional, global &c.) should be indicated.

**FORMAT OF DATA**
The software format and version must be given (eg. STATA 15, .csv, Excel, .txt, &c.).

**CODEBOOK / SUPPORTING DOCUMENTATION**
Codebooks and supporting documentation should be provided in PDF/A or simple text format.

**ACCESS STATUS**
The status of access to the data must be indicated. The status ‘open data’ should be assigned to datasets that are to be made publicly available via the internet. If the data is subject to embargo, the expiry date of the embargo should be indicated.

**LICENSE**
It is advisable to provide a license (eg. CC-BY International).

**DIGITAL IDENTIFIER**
The Library will assign a unique object identifier to the dataset, for the purpose of locating, linking and citation.

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

**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**
In cases where it is intended to generate future iterations of the dataset, details should be provided.

**DATASET CITATION**
A short ‘ready-to-use’ data citation should be provided, incorporating core descriptive elements: family name; first name; year; dataset title; ‘EUI ResData’. For example: ‘Hutcheson, Derek; Jeffers, Kristen (2013): Integration Indicators (CITINT). EUI ResData.’

---

48 [https://creativecommons.org/licenses/](https://creativecommons.org/licenses/)
6. RESEARCH DATA IN EU HORIZON 2020

EU project managers who are preparing EU Horizon 2020 applications are required to submit preliminary information about data management in the proposed project. That section of the application is evaluated by the European Commission under the criterion 'impact.' When completing the general information section of the Horizon 2020 application, EUI project applicants should address the following questions: (i) What types of data will the project generate/collect? (ii) What standards will be used? (iii) How will the data be exploited and/or shared for verification and re-use? (iv) If data cannot be made available, explain why (v) How will the data be curated and preserved?49

If the project proposal is approved, a data management plan (DMP) must be provided within six months of the start of the contract period. The European Commission mandates two further versions of the DMP; one at mid-point and one as a final project deliverable. The EC provides further details in the Guidelines on Data Management in Horizon 2020.50 The DMPonline tool51 (Section 3 above) can be used for Horizon 2020 data management plans, by selecting the Horizon 2020 template in the Funder section. The EUI ResData repository is a Horizon 2020 compliant solution for the preservation and sharing of research project outputs.52

The European Commission’s policy is for research data to be “as open as possible, as closed as necessary”.53 A list of opt-outs is provided by the EC.54 The Horizon 2020 research data pilot requires grant beneficiaries to deposit research data outputs in a research data repository. The EC does not recommend specific repositories. The guidelines state that “as far as possible, projects must then take measures to enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data. One straightforward and effective way of doing this is to attach Creative Commons Licences (CC BY or CC0) to the data deposited.”55

7. OPEN DATA

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

50 econlibrary@eui.eu
52 http://www.dcc.ac.uk/dmponline
53 https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow
55 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf p.8: "Projects can opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions - if: participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited; participation is incompatible with the need for confidentiality in connection with security issues; participation is incompatible with rules on protecting personal data; participation would mean that the project's main aim might not be achieved; the project will not generate / collect any research data or; there are other legitimate reasons (you can enter these in a free-text box at the proposal stage)."
Open data resources are available via the internet. Major data repositories are indexed in the international re3data registry\textsuperscript{57} and the data repositories section of the Open Access Directory.\textsuperscript{58}

By carefully noting the metadata elements explained in Section 5(c) 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. Outputs should be either (i) the result of original data generation, or (ii) the output of significant, value-added, elaboration of pre-existing sources.

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 and documenting the informed consent of subjects for the collection and processing of personal data. Dataset creators are responsible for the anonymisation of data observations (Section 4(e) above). The EUI Guide to Good Data Protection Practice in Research provides further information on data protection.\textsuperscript{59}

- **Database copyright**: Research data outputs which are elaborated from pre-existing copyrighted sources may require permission to reposit and share. It is not possible to publish a dataset containing substantial portions of data sourced from pre-existing databases governed by contractual license. The Library can provide advice on copyright and data protection.\textsuperscript{60}

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. In some cases it may be possible to create a limited public version of a larger, restricted dataset. Data can also be reposited solely for preservation purposes (dark archive).

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 database.

8. **QUALITATIVE DATA IN THE HUMANITIES AND SOCIAL SCIENCES**

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 surveys. The definition of 'data' varies across academic disciplines and sub-disciplines – especially where there is a mix of qualitative and quantitative methods. It is important that dataset design be located in the culture of the discipline or sub-discipline in which the research is undertaken.

8a) 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 scholars 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 and/or rights' holders in advance of digitisation.

The generation of qualitative data is also subject to strong ethical considerations and standards. 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 or sub-discipline. Scholars are responsible for obtaining the informed consent of subjects for the collection and processing of personal data. The mode of consent obtained from subjects depends on the nature of the research project, the kind of data collected, and how and when the data will be used. Where possible, it is recommended to obtain written consent, using a template appropriate to the discipline or sub-discipline in which the research is undertaken. Dataset creators are responsible for the anonymisation of sensitive data observations (Section 4(e) above).

Due to the sensitive personal nature of many qualitative data observations, scholars should pay particular attention to ethical standards when handling data relating to ethnicity, health, orientation, religion, biometrics &c. Persons, families and households cannot be identifiable in any dataset. The linking of variables on ethnicity &c. to individuals, families or households, is governed by data protection legislation and academic best-practice.

8b) 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. 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. Many of the tools used for the analysis of quantitative data (eg. Gauss, Julia, MATLAB, Python, R, Stata) can also be used for qualitative data analysis, where aggregate statistical observations are

required.

The analysis of restricted personal data may require scholars to work in a ‘safe-room’ environment. If this is stipulated by a data provider or project funder, contact the EUI Library. 63

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 additional considerations. Research data management encompasses the control of data inputs, the handling and use of data, the protection of data, the creation of data outputs and the preservation of data. Research data management is carried out by scholars throughout the duration of a research project (Section 3 above). During data analysis, 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 or prepared for digitisation.

Data management plans (DMPs) are normally required by science funding agencies. Due to the heterogeneous, multi-media and complex nature of qualitative data in the humanities and social sciences, it is particularly important for scholars to keep a record of data sources, including notebooks, questionnaires, codebooks and multilingual thesauri. Supporting documentation serves as the basis for accurate metadata, and facilitates future retrieval and reuse. In the case of non-repeatable, time-sensitive, socio-political research, data management plans may require a detailed explanation of the qualitative research methods used.

8(d) Metadata for qualitative data

Metadata are data about data, presented in a systematic schema. Accurate metadata are necessary for the organisation, use, repositing, sharing and machine-discovery of datasets (Section 5(c) above). The metadata fields used for quantitative data outputs can also be used for qualitative data outputs. However there are additional considerations. As well as the name(s) of scholars 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 reposed in the EUI ResData repository or in a subject repository, or a multi-disciplinary repository. By carefully noting the metadata elements explained in Section 5(c), 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. Qualitative datasets that are made

63 econlibrary@eui.eu
available as open data should be the product of original research and should adhere to best-practice regarding data protection and database copyright (Section 2 above).

9. INFRASTRUCTURE, SOFTWARE AND SUPPORT

Infrastructure, software and connectivity are provided by the EUI ICT Service. Research software programmes are listed on the ICT web site.\(^6^4\) Technical support is provided at the site offices of the ICT Service. The EUI Library maintains the Data Portal, the Micro Data Restricted Server and the ResData repository. Advice on the use of statistical software is provided by EUI research software tutors.\(^6^5\) The Data and Methods Clinic organises workshops and data consultations for EUI members.\(^6^6\)

The EUI ICT Service provides research data software, including Fortran, Gauss, MATLAB, OxMetrics, Python, R, Stata, Stat/Transfer, WinEdt, WinRATS, and supports the high-performance computing cluster at the EUI. A guide to the High Performance Cluster (HPC) is available by writing to HPC.support@eui.eu (EUI members only).

The EUI Library maintains the Data Portal, the Micro Data Restricted Server and the ResData repository. The Library provides assistance with data discovery, data access and use of licensed digital databases, and open data. The Library also helps EUI members prepare data management plans. The workflow for the submission, preparation and metadata enrichment of EUI data outputs is also managed by the Library.

A directory of online research data software manuals, with full-text links, is available on the Library web site.\(^6^7\) Data software manuals are available in the Badia Library and the Economics Departmental Library (shelfmarks 001 to 006). The Library holds a substantial collection of works on statistical science and data methodology (shelfmarks 500 to 519). Books and manuals in any language may be suggested for acquisition by the Library.\(^6^8\)

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 is provided on Monday, Wednesday and Friday afternoons from 14:45 to 18:30 (Departmental Library, Villa La Fonte).

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 Blog\(^6^9\) and Twitter.\(^7^0\)

\(^6^4\) http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/SoftwareResearch.aspx
\(^6^6\) https://www.eui.eu/ServicesAndAdmin/AcademicService/Academic-and-Professional-Development/Data-Clinic/Workshops
\(^6^7\) http://www.eui.eu/Research/Library/ResearchGuides/Economics/Software.aspx
\(^6^8\) Write to econlibrary@eui.eu
\(^6^9\) https://blogs.eui.eu/library/category/economics
\(^7^0\) https://twitter.com/econlibrary
10. INTERNATIONAL RESEARCH DATA GUIDELINES

Information on data use and management, data protection, data preservation and open data is available from these sources:

- American Economic Association guidelines on data availability
- Big Data Europe - Big Data Aggregator Platform
- Consortium of European Social Science Data Archives (CESSDA)
- 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 ResData repository
- EUI ResData submission
- 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
- Guide to Social Science Data Preparation and Archiving (ICPSR)
- 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
- 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

6th edition

March 2018