

RESEARCH DATA GUIDE

EUI LIBRARY

11th edition 2023



Research Data Guide

11th edition, 2023

1. **Data discovery and the EUI Library Data Portal**
 - 1.a The EUI Library Data Portal
 - 1.b Registration for access to Library-hosted micro-socioeconomic data
 - 1.c Access to Eurostat micro-socioeconomic data
 - 1.d Terms and conditions of use and database copyright
 - 1.e Annual call for EUI early stage researcher funding applications
 - 1.f Access to data at other institutions and open data resources
2. **Data generation, data protection, data processing and ethical use**
 - 2.a Data protection
 - 2.b EUI Ethics Committee review: processing and use of sensitive data
 - 2.c Information Security Risk Assessment
 - 2.d Anonymisation and pseudonymisation
 - 2.e Library Secure Data Room and security protocols
 - 2.f Third-party data access
3. **Data management plans**
 - 3.a Generating a data management plan (DMP)
 - 3.b EUI support for principal investigators and project managers
 - 3.c Data management planning tools
4. **Research data management in Horizon Europe, ERC and other EU programmes**
5. **Managing data during the research project cycle**
 - 5.a Data input and quality control
 - 5.b Secure storage and network backup during the research project
 - 5.c Structure, folders, files, variables, format and versioning
 - 5.d Documentation and codebooks
 - 5.e Deletion of sensitive data at the end of the research project
6. **Repositing and preserving data in the EUI research repository Cadmus**
 - 6.a Preparing data for reposit in Cadmus
 - 6.b Completing the dataset submission form
 - 6.c Metadata
 - 6.d Linking datasets and publications
7. **Open Data, FAIR Principles and Open Science**
 - 7.a Data repositories and Library support
 - 7.b Data metadata and access rights
 - 7.c FAIR data principles: Findable, Accessible, Interoperable and Reusable
 - 7.d Pre-registration research plans
8. **Qualitative and unstructured data in the humanities and social sciences**
 - 8.a Protection and ethical use of qualitative data
 - 8.b Support, software and infrastructure for qualitative data
 - 8.c Research data management and data management plans for qualitative data
 - 8.d Metadata for qualitative data
 - 8.e Qualitative data archiving, preservation and access
9. **EUI infrastructure, software, protocols and support**
10. **International research data guidelines**

The hyperlinked version of this Guide is available at:
<https://www.eui.eu/Research/Library/ResearchDataServices>

This 11th edition of the EUI Library Research Data Guide provides information and best practices about data discovery, data access, data generation, data processing, data protection, data management planning, data preservation and data sharing. Information in this guide helps support EUI data users throughout their research projects and is intended to raise awareness of how good research data management contributes to the overall quality of scientific research.

1. DATA DISCOVERY AND THE EUI LIBRARY DATA PORTAL

This section describes access to EUI licensed data resources; access to restricted micro-socioeconomic data; terms and conditions of access and use; access to data hosted by other institutions and resources for open data discovery.

1.a The EUI Library Data Portal

The Library maintains a Data Portal¹ providing access to licensed resources for EUI members and introducing selected open data resources of relevance to EUI research themes. All 78 resources indexed in the Data Portal have an online guide with (i) data description and scope (ii) time-period coverage and release/wave information (iii) supporting resources (online manuals, software transfer routines, user networks) and (iv) terms and conditions of access and use. Modes of access for each database are described in the individual resource guides linked from the Data Portal. There are five sub-directories:

- Macroeconomic, Financial and Historical Data Resources for national, regional and global economic, political and historical research²
- Micro-socioeconomic Data Resources for household, family, individual and company-level research³
- European, EU and Euro Area Data Resources providing statistics for research on pan-European topics, the European Union, European countries and European sub-state regions⁴
- Data Directory for Research on COVID-19⁵
- Data Topics Index for matching research themes with data resources.⁶

Databases indexed in the EUI Library Data Portal have diverse modes of access which are described on each resource page. The most common modes of access are:

- Proxied internet HTML interface (mostly for macroeconomic, political economic and historical time series)
- Library restricted data server (mostly for micro-socioeconomic resources)
- Desktop application: eg. Datastream via Programmes, Compustat via Stata.

¹ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal>

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/ResearchGuides/Economics/EuropeanEUandEurozoneStatistics>

⁵ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Crisis> See also the pilot COVID-19 Social Sciences and Humanities Data Portal: https://opac.eui.eu/client/en_GB/covid19-ssh-data/search

⁶ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/Topics>

Macroeconomic data providers include Databanks, ECB, Eurostat, Global Financial Data, IMF, OECD, Refinitiv-LSEG, Standard & Poor's, Statista GmbH, World Bank, UN and WIIW.⁷

1.b Registration for access to Library-hosted micro-socioeconomic data

EUI members can register for access to micro data hosted by the EUI Library as follows:

- Complete the micro-socioeconomic data registration form (selecting the required dataset from the drop-down menu)⁸
- For some providers, parallel registration with data issuers is required
- Registered users are given access to the Library's restricted data server
- Specific instructions are provided under 'How to access' on the individual micro data resource pages linked from the Data Portal⁹
- Short-term EUI visitors do not have access to EUI-hosted micro data.

1.c Access to Eurostat micro-socioeconomic data

The EUI is a Eurostat-approved research data facility. Access to Eurostat collated data (EU Adult Education Survey, EU Labour Force Survey, EU Statistics on Income and Living Conditions and European Community Household Panel) requires individual end-user contracts with Eurostat:

- Complete the Library's Micro Data Registration Form, selecting the Eurostat dataset required from the dropdown menu¹⁰ (EU-AES, EU-LFS, EU-SILC, ECHP)
- EUI scholars must create an EU Login to apply for Eurostat micro data access¹¹
- Special note: Eurostat informs that for EUI members, it is necessary to use a non-institutional (non-EUI) personal email account for the creation of the EU Login. (It is not necessary for all members of a research team to create an EU login - only the person submitting the application)
- Before completing the application, consult the relevant Eurostat micro data resource page for details of variables, reference years and population¹²
- Applications for data access are processed via the online Microdata Access Workflow Tool¹³
- When completing the application, EUI researchers should enter the name of their thesis supervisor as Principal Investigator (P.I.)
- For institutional Contact Person indicate 'Dr. Thomas Bourke'
- Write to econlibrary@eui.eu requesting (i) the Research Entity ID and (ii) the standard description of the EUI's secure data infrastructure and protocols
- The names and email contacts of all scholars on the project must be included in the form (researchers, fellows, professors)

⁷ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/MacroDataSet>

⁸ <https://www.eui.eu/Research/Library/RequestForms/Register-micro-data>

⁹ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal>

¹⁰ <https://www.eui.eu/Research/Library/RequestForms/Register-micro-data>

¹¹ https://ec.europa.eu/eurostat/documents/203647/771732/EU_Login_Tutorial/

¹² <https://ec.europa.eu/eurostat/web/microdata>

¹³ <https://webgate.ec.europa.eu/multisite/microdata/>

- The applicant uses the Microdata Access Workflow Tool¹⁴ to initiate the application procedure. Full instructions are online¹⁵
- After receipt of the application, the Eurostat Technical Unit may provide feedback and/or requests for further information about the project
- The submitter should monitor their EU Login account for updates regarding the application. (Library staff do not have access to individual EU Login accounts)
- If the application is approved by Eurostat, the applicant should print the research project proposal (RPP) and confidentiality declaration(s)
- The Principal Investigator initials all pages of the RPP, and signs and dates the final page
- The application is sent to the institutional Contact Person for signature: econlibrary@eui.eu
- Separate confidentiality declarations must be signed by (i) the Principal Investigator (ii) the Applicant(s) and (iii) the Contact Person
- The applicant scans the signed documents and uploads to the EU Login account
- When Eurostat approves access to the data, send this to econlibrary@eui.eu
- The Library provides access to the restricted data server.

No person(s) other than those named in the Eurostat contract can access the data. Breach of contract conditions may lead to the withdrawal of the EUI's research entity recognition, and the loss of access to Eurostat micro data by other EUI scholars. There is no off-campus access to Eurostat micro data. Short-term visitors to the EUI do not have access to Eurostat micro data. At the end of the project contract period, Eurostat will contact the P.I. requesting a signed confirmation of destruction of local copies/sub-sets of the data (Section 5.e below). Publications based on the data should be reported via the Eurostat online form.¹⁶

1.d Terms and conditions of use and database copyright

Access to, and use of, databases provided by the EUI Library are subject to contractual license agreements and database copyright terms. Full details are on the Library's Terms and Conditions' web page.¹⁷ Data users are individually responsible for compliance with terms and conditions of access and use. Violation of license terms puts at risk other EUI members' future access to data resources and may expose the university to financial penalties. EUI users may not distribute, or allow any other party to have access to data which is provided under license. Users may not use licensed materials for commercial purposes. In some cases, the sharing of micro-socioeconomic data with journal referees is not permitted (check individual data contracts for details).

1.e Annual call for EUI early stage researcher funding applications

The EUI Dean of Research issues an annual call for early stage researcher (ESR) funding support. EUI researchers and fellows can apply for financial support to cover some, or all, of the costs incurred in data collection and data generation (surveys, interviews, fieldwork, experiments &c.). ESRs should check the funding opportunities

¹⁴ <https://webgate.ec.europa.eu/multisite/microdata/>

¹⁵ Instructions for using the Microdata Access Workflow Tool are at:
https://ec.europa.eu/eurostat/documents/203647/771732/EU_Login_Tutorial/

¹⁶ <https://ec.europa.eu/eusurvey/runner/Results-from-microdata>

¹⁷ <https://www.eui.eu/Research/Library/Collections/TermsAndConditions>

intranet page under 'Internal (EUI) funding opportunities'¹⁸ for the text of the call, application form, FAQs and updated deadline.

- Before the deadline of the ESR call, applicants must take, and pass, the online CITI¹⁹ training module 'Essentials of Grant Proposal Development' and (for empirical projects dealing with human-subject non-secondary data) 'GDPR for Research and Higher Ed'.
- Some projects may require EUI Ethics Committee approval (Section 2.b below). ESR call applicants may request Ethics Committee approval after the funding decision, but before the commencement of data collection and processing. Approval cannot be granted retroactively.
- ESR applicants who are applying for data funding should first check the Library Data Portal for availability. If the required data is not available, write to DPR@eui.eu asking whether the Library can contribute to the cost of data acquisition.
- The reply of the Library Director should be included in the ESR funding application.
- Academic units rank the proposals and a Selection Committee decides on financial support.

The ESR funding application form and full details are provided to all EUI researchers and fellows by email at the launch of the call.

1.f Access to data at other institutions and open data resources

EUI members who require access to restricted data at another institution (data centre, statistics agency, university, archive) should apply for access early in their research projects because application periods can be lengthy. Some sociological, qualitative and micro geo-coded data can only be accessed at issuers' secure on-site facilities. 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. Scholars, government agencies and international organisations increasingly share data, codebooks and software via the internet. Open research datasets can be located via Google Dataset Search²⁰ and the re3data²¹ registry of research data repositories which is indexed by discipline, sub-discipline, data type and host institution.

2. DATA GENERATION, DATA PROTECTION, DATA PROCESSING AND ETHICAL USE

This section provides information on data protection requirements for scholars who generate and process sensitive data; the EUI Ethics Committee review procedure; anonymisation and pseudonymisation techniques and the Library Secure Data Room.

¹⁸ <https://www.eui.eu/en/private/funding-opportunities>

¹⁹ <https://www.eui.eu/en/private/citi-program>

²⁰ <https://datasetsearch.research.google.com/>

²¹ <http://www.re3data.org/>

2.a Data protection

Data Protection at the EUI is governed by President's Decision No.10/2019²² which was introduced following the adoption of the EU General Data Protection Regulation (GDPR).²³ The EUI 'Guide to Good Data Protection Practice in Research' provides detailed information about data protection at the EUI.²⁴ Special terms and conditions apply to access and use of personal data, including micro-socioeconomic and qualitative data. This reflects the sensitive nature of observations about human subjects, families and households. Data protection provisions apply to (i) personal data generated by scholars for research purposes (ii) licensed micro-socioeconomic data provided to EUI members by the Library and (iii) data provided directly to EUI members under third-party contract.

EUI principal investigators are responsible for compliance with the laws and regulations governing data protection and database copyright. Scholars are subject to personal data protection regulations when generating, processing, collecting and storing personal data.²⁵ When using pre-existing datasets, scholars must not attempt to identify any person, family or household. When generating new data (eg. surveys) scholars must adhere to data protection legislation and best practices. The processing of sensitive data relating to health, racial/ethnic origin, religious beliefs, political opinions, genetic/biometric observations, sexual orientation, memberships &c. requires strict safeguards.²⁶ The EUI 'Guide to Good Data Protection Practice in Research' states: "As a rule, the processing of sensitive data is prohibited. However, Article 8 of the EUI's Data Protection Policy provides for specific circumstances which allow for the processing of sensitive data. The most common in research is upon the data subject's explicit consent."²⁷ Persons, families and households cannot be identifiable in any subsequently published research dataset. It must not be possible to identify human subjects by combining observations with external sources or other means. Anonymisation and pseudonymisation are described in Section 2.d below.

When collecting, processing, using and storing personal data; scholars are responsible for obtaining and documenting the informed consent of subjects. Any and every collecting, processing, storing and future sharing of data related to human subjects requires consent. These aspects must be foreseen and explicitly stated in the consent form. Consent must be intelligible; it should refer clearly and precisely to the scope and the consequences of the data processing, as well as the retention period. EUI scholars can use the sample consent form.²⁸ Where written consent is not possible, audio or video recording may be used. Human subjects must be informed of their rights, and scholars should consider risks to data subjects before embarking on personal data

²² EUI PD 10/2019 <https://www.eui.eu/Documents/AboutEUI/Organization/DataProtection/PresDecision10-2019-DataProtection.pdf>

²³ Regulation (EU) 2016/679, GDPR: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

²⁴ <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

²⁵ See the EUI President's Decision on Privacy:

<https://www.eui.eu/Documents/AboutEUI/Organization/PresidentDecisionOnPrivacy.pdf>

²⁶ EUI Guide to Good Data Protection Practice in Research, p.7:

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

²⁷ EUI Guide to Good Data Protection Practice in Research, p.9:

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

²⁸ <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/sampleInformedconsentform.pdf>

gathering via interviews, questionnaires or other means.

2.b EUI Ethics Committee review: processing and use of sensitive data

The collection and processing of personal data for research purposes may require EUI scholars to apply for Ethics Committee review before data collection and processing can begin. Principal investigators should consult the online Initial Checklist on the Ethics Committee homepage.²⁹ EUI members can initiate the review via the link 'Procedure to request an Ethics Review' (intranet login required). The Ethics Committee review must be completed before data collection begins - approval cannot be granted retroactively. The review can take up to six weeks.

EUI members have access to the online Collaborative Institutional Training Initiative (CITI) platform, providing training on research ethics including: Human subjects research; Responsible conduct of research; Observational research protocols; General Data Protection Regulation (GDPR) for research, and Essentials of grant proposal development. EUI members can create a CITI account (username and password) via the EUI intranet (login required)³⁰ selecting 'European University Institute' from the drop-down menu.

EUI members requiring access to the survey tool Qualtrics³¹ should write to the HelpDesk³² to request an individual account. Users should read the EUI Declaration on the Processing of Data via Surveys³³ before collecting personal information.

2.c Information Security Risk Assessment

EUI scholars and principal investigators who require ICT Service approval for information security should complete the Information Risk Management Questionnaire³⁴ and send it to the Data Security Officer (DSO) at: data_security_officer@eui.eu Information Security Guidelines regarding compliance with data protection requirements, and the checklist of security controls, are provided on the EUI intranet.³⁵ The information will be used by the DSO to assess the project risk level and to provide advice on the adequate safeguards to implement.

2.d Anonymisation and pseudonymisation

Throughout the research data cycle, scholars are responsible for preserving the confidentiality of human subjects, families and households. Sensitive data must be anonymised so that persons, families and households cannot be directly or indirectly identifiable. Anonymisation techniques include; the removal of direct identifiers from datasets (names, workplaces, telephone numbers &c.); the aggregation, banding or

²⁹ <https://www.eui.eu/en/public/about/committees/ethics-committee>

³⁰ <https://www.eui.eu/en/private/citi-program> [EUI intranet login required]

³¹ <https://www.eui.eu/ServicesAndAdmin/ComputingService/PolicyDocuments/DataProtectionforSurveys>

³² <https://www.eui.eu/helpdesk>

³³ <https://www.eui.eu/ServicesAndAdmin/ComputingService/PolicyDocuments/DataProtectionforSurveys>

³⁴ <https://www.eui.eu/Documents/ServicesAdmin/ComputingService/Intranet/EUI-Information-Risk-Management-Questionnaire.xlsx> [EUI intranet login required]

³⁵ <https://www.eui.eu/Documents/ServicesAdmin/ComputingService/Intranet/EUI-Information-Security-Guidelines.pdf> [EUI intranet login required]

reduction of the precision of variables (eg: year of birth rather than year/month/date of birth); the reporting of general levels of observation (eg. replacing specific geo-codes with general locations); the generalisation of categories (eg. sector of employment rather than precise employment function); and the removal of the upper or lower ranges of observations.³⁶

Pseudonymisation can be used when it is necessary to retain personal data for future processing (eg. a follow-up survey). Pseudonymisation entails the substitution of personal information with a unique identifier not connected to the person's real identity, using unique, randomly-generated, numeric or alphanumeric identifiers. The data and the key file must be kept separate and secure at all times. Pseudonymised data must be treated as personal data in accordance with GDPR.³⁷ The use of cloud computing services must not reduce the level of data protection. Data can be encrypted with a tool such as TrueCrypt.³⁸

2.e Library Secure Data Room and security protocols

The EUI Secure Data Room is located on the upper floor of the Library, Badia Fiesolana (BF-173). The Room is only accessible to scholars processing sensitive data, or using data from providers whose contracts explicitly require that data be used in a locked room. The room is not for 'general' data access and use. New EUI data users who require access to the Secure Data Room should send a copy of their data access contracts to econlibrary@eui.eu. Authorised users can book a computer (Monday to Saturday) via the Library's online form.³⁹ New users can select from Workspaces #1 and #2. (Workspaces #3, #4, #5 and #6 are reserved for designated datasets.) To access the Secure Data Room, scholars should enter the Library via the Badia loggia, ascend to the first floor and show their EUI card at the Library Loan Desk where the key will be issued. The key must be returned to the Loan Desk at the end of the working session.

2.f Third-party data access

Many third-party data providers require individual end-user contracts. In cases where the signature of an 'authorised representative' is required, scholars should write to the office of the Legal Advisor in the Central Coordination Unit.⁴⁰ The Library provides support for data access applications, contract workflow and data management. Send the following information to econlibrary@eui.eu:

- Name of data issuer/publisher
- Title of dataset, with internet URL (if available)
- Names of EUI and external research team members, with email contacts
- What special conditions are required for access to the data (eg. remote protocol)?
- What special conditions are required for use of the data?
- Does the data issuer require a contract counter-signature from the EUI DPO?

³⁶ See also the Finnish Social Science Data Archive anonymisation guidelines: <https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/>

³⁷ Regulation (EU) 2016/679, GDPR: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

³⁸ <https://sourceforge.net/projects/truecrypt/>

³⁹ <https://www.eui.eu/Research/Library/ResearchDataServices/SensitiveData>

⁴⁰ <https://www.eui.eu/en/public/about/central-coordination-unit>

- Does the data issuer require that the data be used in the EUI's Secure Data Room?
- Does the data issuer require that the data be used on a stand-alone, off-line computer?
- For projects with non-EUI collaborators: What controller/processing operations will be undertaken by research partners?

Data provided by third parties can be elaborated on a provider's infrastructure or (where permitted) can be transferred to the EUI. When sensitive personal data is transferred to the EUI, users have to complete the EUI 'Notification of Processing Operations' form, available from the office of the EUI Legal Advisor.⁴¹ In cases of joint research collaborations between EUI members and scholars at other institutions (eg. in the context of research consortia) it is important to distinguish between the 'data controller' function and the 'data processor' function - and to indicate the division of labour between scholars - especially where research team members are based at different institutions. The General Data Protection Regulation distinguishes between 'data controller' (eg. principal investigator or head of department) and 'data processor' (eg. PhD researcher). The GDPR defines 'controller' as a "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data" (Article 4(7)). The GDPR defines 'processor' as a "natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller" (Article 4(8)).⁴² If required by the data issuer, the EUI ICT Data Security Officer can provide a description of ICT security protocols and infrastructure.⁴³

3. DATA MANAGEMENT PLANS

Data management plans (DMPs) are documents normally required by research funders at the beginning, mid-point and end of research projects. Early stage researchers should learn how to create DMPs, required when applying for future project funding.

3.a Generating a data management plan (DMP)

The principal investigator, or project manager, or delegated member of the research team, is responsible for writing and delivering the data management plan. The following supporting materials should be consulted during the preparation of EUI DMPs:

- EUI 'Guide to Good Data Protection Practice in Research'⁴⁴
- EUI Guidelines on Ethics and Integrity in Academic Research⁴⁵
- EUI Library 'Research Data Guide' [this document, also online with hyperlinks].⁴⁶

Scholars preparing a data management plan should carefully follow the instructions provided by the relevant science programme office (eg. Horizon Europe, European Research Council &c.). Where a DMP template is provided, this must be used; for

⁴¹ <https://www.eui.eu/en/public/about/central-coordination-unit>

⁴² Regulation (EU) 2016/679, GDPR: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

⁴³ Data_Security_Officer@eui.eu

⁴⁴ <https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf>

⁴⁵ <https://www.eui.eu/ServicesAndAdmin/DeanOfStudies/Ethics-and-Integrity-in-Academic-Research>

⁴⁶ <https://www.eui.eu/Research/Library/ResearchDataServices>

example: the Horizon Europe DMP template⁴⁷ in the Horizon Europe Programme⁴⁸ or the ERC DMP template⁴⁹ in the ERC Work Programme.⁵⁰

The sections below provide advice and prompts which can be used when preparing a data management plan. Every data management plan is a unique document, grounded in the individual research project. It is the responsibility of the P.I. (or delegated team member) to write and deliver a comprehensive, accurate and original data management plan. The Library provides feedback and guidance. Data management plans describe:

- How data is collected and generated
- How data is used, elaborated and organised
- How data, and data subjects, are protected
- How data, code and ancillary elements are described and documented
- How data is stored and secured during the project, and how long it will be retained
- How dataset authorship and credit are assigned
- How data is preserved
- How, whether and under what terms, research data outputs can be shared.

The FAIR data principles - to make data findable, accessible, interoperable and reusable⁵¹ - must be considered during the preparation and revision of data management plans. Detailed information about these four principles are in the grid below.

For international collaborative projects, the name of the person who has final authority regarding credit for dataset creation, and the sharing of data outputs, should be indicated. Unless otherwise stated, the P.I. undertakes this role. It is important to update DMPs throughout the research project, incorporating new information about data collection, data generation, methodologies and relevant changes to the composition of the research team or consortium.

3.b EUI support for principal investigators and project managers

EUI principal investigators and Project Managers can use the general prompts in the grid below when compiling a data management plan. Where the science funder provides a DMP template, this should be used. The data management plan must include detailed, project-specific information before submission to the science funder. During preparation of the DMP, EUI members may need to liaise with Library staff, the Data Protection Officer, the Data Security Officer and/or the Ethics Committee (if approval is required). The following elements should be provided in a data management plan:

⁴⁷ <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>

⁴⁸ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf (pp.43-50). Further information is contained in these Horizon Europe documents: Model Grant Agreement (MGA) article 17; Work Programme General Annexes; Proposal template and; Annotated Grant Agreement (AGA), article 17.

⁴⁹ <https://horizoneurope.apre.it/wp-content/uploads/sites/3/2022/01/ERC-Data-Management-Plan.pdf>

⁵⁰ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2022/wp_horizon-erc-2022_en.pdf

⁵¹ <https://www.go-fair.org/fair-principles/>

Overview of research project
<ul style="list-style-type: none"> • Provide an overview of the research project, indicating discipline/sub-discipline, scope, methodology and primary research questions • List the types of data being collected, generated, processed and used.
Anticipated data outputs and utility
<ul style="list-style-type: none"> • Describe expected data outputs • Distinguish (where applicable) between (i) in-project data which will not be shared and (ii) post-project research dataset outputs (public or embargoed) • Describe expected public utility; eg. for researchers, policymakers, media organisations and the general public • Describe possible inter-disciplinary use • Describe data quality assurance processes.
Resources required for data collection, generation, processing and use
<ul style="list-style-type: none"> • Provide a detailed description of data collection, generation, processing and use • Describe anticipated technical and research assistance requirements • Give an overview of the data management responsibilities of research project leaders, team members and (where applicable) institutional partners.
Data security, infrastructure and protocols
<ul style="list-style-type: none"> • EUI members should write to the Data Security Officer⁵² to request the standard description of data security, infrastructure and protocols. This internal document describes layered physical and network security, authentication protocols and other elements for inclusion in data management plans.
Ethical considerations
<ul style="list-style-type: none"> • Describe data protection measures in place to scrupulously guarantee that persons, families and households are not identifiable in any dataset outputs • EUI Ethics Committee review may be required before the collection and processing of personal data. Ethics Committee approval cannot be granted retroactively; EUI members should initiate the request for review before collecting data⁵³ • Consultation with the EUI Data Protection Officer and/or Data Security Officer may be required • Maintain a record of pre-existing data resources used in the project, and comply with copyright provisions • Contact the Library for assistance.

⁵² Data_Security_Officer@eui.eu

⁵³ <https://www.eui.eu/en/public/about/committees/ethics-committee>

FAIR Data Principles to make data
(i) Findable (ii) Accessible (iii) Interoperable and (iv) Reusable

Making data Findable

- Identify the repository where data outputs will be preserved and, where possible, made openly available; eg. the EUI research repository, Cadmus⁵⁴
- Describe how folders, files, variables and versions will be consistently named to aid discovery
- Datasets should be assigned accurate and consistent metadata to aid findability and machine-retrievability (Section 6.c)
- Datasets should have persistent, unique digital object identifiers (eg. handles generated by the EUI research repository, Cadmus).

Making data Accessible

- Distinguish between in-project data which will remain closed (or will be destroyed) after the project; and data outputs which will be preserved and (where possible) openly shared after the project
- For restricted data, give reasons for restricted status (eg. personal data protection and/or database copyright compliance)
- For restricted data the terms and conditions for access should be explicitly stated (even if these might be revised as the research progresses)
- For sharable data, the status 'open' (accessible immediately) or 'embargoed' (accessible at a later date) should be assigned
- Describe data protection and anonymisation measures in detail
- Describe the software packages/tools required to use the data
- Metadata records are indexed in the EUI research repository, Cadmus, which provides machine-readable metadata for indexing by search engines and application programming interfaces (eg. Google Scholar, OpenAIRE and CORE).

Making data Interoperable

- Provide details of any standard controlled vocabularies (ontologies, thesauri and taxonomies) from the relevant discipline (social sciences, humanities, ethnography &c.)
- Variable schemas should be readable by standard software packages to facilitate data interoperability
- Standard metadata and naming conventions should be used
- The EUI research repository, Cadmus, uses web standards (the underlying DSpace software adopts standards for embedded micro-data and the OAI-PMH interoperability protocol) providing both human- and machine-readable interfaces to search, discover and access repositied data.

Making data Reusable

- Open data generated by the research project should be made available under an open licence that clearly states reuse conditions - either the Creative Commons Attribution International license: CC-BY International⁵⁵ or the Creative Commons Public Domain Universal license: CC0⁵⁶

⁵⁴ <https://cadmus.eui.eu/>

⁵⁵ <https://creativecommons.org/licenses/by/4.0/>

⁵⁶ <https://creativecommons.org/publicdomain/zero/1.0/>

- Supporting documentation and codebooks (PDF/A documents and 'readme.txt' files) should be deposited with the data to make the observations comprehensible and reusable by other scholars and stakeholders
- Provide information about any tools or instruments necessary to reuse or verify the data (software, algorithms, routines, models &c.)
- Note any data embargo period, where applicable.

EUI research data guidelines and standards

- EUI members can cite these institutional guidelines in their data management plans:
 - EUI 'Code of Ethics in Academic Research' 2021⁵⁷
 - EUI Library 'Research Data Guide' (11th ed) 2023.⁵⁸

3.c Data management planning tools

Online tools can be used to prepare structured data management plans - complying with EU Horizon Europe, European Research Council, and other science funder DMP requirements.

- **DMPonline**⁵⁹ (maintained by the Digital Curation Centre) can be accessed via 'Create Account' at the upper right of the DMPonline homepage. For organisation enter 'other'. Select the science funder's template. The principal investigator (P.I.) should be identified in the data management plan. For research teams, the P.I. can assign DMP sharing rights by entering colleagues' email addresses and assigning the status of 'co-owner', 'editor' or 'read only.' Enter project details (title, abstract, &c.) and click on 'Initial DMP' in the top menu. Complete the sub-fields to generate the data management plan. A Word document can be generated. If the science funder provides a DMP template, the text generated in DMPonline can be transferred into the funder's template.
- **The Research Data Management Organiser** (RDMO)⁶⁰ is an online tool available in German and English to support research data management. RDMO is maintained by the Deutsche Forschungsgemeinschaft (DFG).
- **Argos**⁶¹ is an online DMP creation tool maintained by OpenAIRE. Create an account on the Argos platform, click 'Launch Wizard' and follow the step-by-step instructions.
- **The DMP Evaluation Rubric**⁶² developed by Science Europe, provides core criteria which can be used by principal investigators and project managers to evaluate data management plans.
- **FAIRsharing** is a data support service of the Data Readiness Group and the Bodleian Library, University of Oxford, providing resources for data and metadata standards and research data management.⁶³

⁵⁷ <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/CodeofEthicsinAcademicResearch.pdf>

⁵⁸ <https://www.eui.eu/Research/Library/ResearchDataServices/Guide>

⁵⁹ <http://www.dcc.ac.uk/dmponline>

⁶⁰ <https://rdmo.readthedocs.io/>

⁶¹ <https://argos.openaire.eu/splash/>

⁶² https://scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf pp.32-51

⁶³ <https://fairsharing.org/>

- **The F-UJI Automated Fair Data Assessment Tool**⁶⁴ maintained by the Fostering FAIR Data Practices In Europe project, helps scholars comply with the FAIR Guiding Principles for scientific data management that ensure data are - and remain - findable, accessible, interoperable and reusable (FAIR).
- **FAIR-Aware**⁶⁵ maintained by the Data Archiving and Networked Services Institute is an online checklist tool to help researchers build compliance with the FAIR data principles.
- **The DMP Catalogue**⁶⁶ maintained by LIBER, provides examples of research data management plans.

4. RESEARCH DATA MANAGEMENT IN HORIZON EUROPE, ERC AND OTHER EU PROGRAMMES

EUI principal investigators and project managers who are preparing EU Horizon Europe and European Research Council (ERC) grant proposals are required to submit preliminary information about data management provisions. The Horizon Europe Programme Guide⁶⁷ and the ERC Work Programme⁶⁸ provide guidance for project proposers.

Data management sections of project proposals are evaluated by the European Commission under the criterion 'impact', which includes scientific and societal benefit. When completing the general information section of EU funding applications, scholars should address: (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 outputs cannot be made available, explain why (v) How will the data be curated and preserved? The EUI Library assists principal investigators and project managers with the data management sections of funding proposals and the subsequent data management plans (Section 3 above).⁶⁹

Open Science is defined by the European Commission as “an approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology... Open science is a policy priority for the European Commission and the standard method of working under its research and innovation funding programmes as it improves the quality, efficiency and responsiveness of research. When researchers share knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.”⁷⁰

In the Horizon Europe programme, open science - which includes open access to research data - is evaluated under the 'excellence' criterion. The European Commission's recommendation for sharing research data outputs is: “As open as

⁶⁴ <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

⁶⁵ <https://fairaware.dans.knaw.nl/>

⁶⁶ <https://libereurope.eu/working-group/research-data-management/plans/>

⁶⁷ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf (pp.43-50). Further information is contained in these Horizon Europe documents: Model Grant Agreement (MGA) article 17; Work Programme General Annexes; Proposal template and; Annotated Grant Agreement (AGA), article 17.

⁶⁸ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2022/wp_horizon-erc-2022_en.pdf

⁶⁹ resdata@eui.eu

⁷⁰ https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

possible; as closed as necessary”.⁷¹ The Horizon Europe programme requires grant beneficiaries to reposit open research data outputs in a research repository. The EUI research repository, Cadmus, can be used for the reposit of EUI Horizon Europe and ERC project data outputs. The EC 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 way of doing this is to attach Creative Commons Licences (CC BY or CC0) to the data reposit.”⁷² Project results can also be shared via the Horizon Results Platform.⁷³

The European Open Science Cloud portal was launched in 2018.⁷⁴ The EOSC portal enables the discovery of tools and resources for data analysis, and information about storage, computation, training and security. The European Commission provides two supporting guides: Prompting EOSC in Practice⁷⁵ which explains the governance of the service and Turning FAIR into Reality⁷⁶ which explains how to make research data findable, accessible, interoperable and reusable. The European University Institute endorsed the European Open Science Cloud Declaration in 2017. The Social Sciences & Humanities Open Marketplace was launched in 2022 to provide “tools, services, training materials, datasets, publications and workflows” for researchers.⁷⁷ Institutional partners are coordinated by the Consortium of European Social Science Data Archives (CESSDA).⁷⁸

5. 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, data protection, data security and backup during the project. At an early stage, scholars should define dataset structure, file structure and variable naming. Documentation and codebooks should be updated throughout the research project. The following guides are useful for scholars working in the social sciences and humanities:

- CESSDA Data Management Expert Guide⁷⁹
- JISC Research Data Management Toolkit⁸⁰
- Finnish Social Science Data Archive Guide⁸¹
- UKDS Prepare and Manage Data Guide.⁸²

⁷¹ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

⁷² http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
p.10

⁷³ <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/horizon-results-platform>

⁷⁴ <https://eosc-portal.eu/>

⁷⁵ <https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1>

⁷⁶ <https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>

⁷⁷ <https://marketplace.sshopencloud.eu/> This is the successor to the Social Sciences and Humanities Open Cloud (SSHOC) initiative, 2019-2022.

⁷⁸ <https://www.cessda.eu/>

⁷⁹ <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

⁸⁰ <https://rdmtoolkit.jisc.ac.uk/>

⁸¹ <https://www.fsd.tuni.fi/en/services/data-management-guidelines/processing-quantitative-data-files/#syntax>

⁸² <https://www.ukdataservice.ac.uk/manage-data/>

5.a Data input and quality control

The collection and generation of data should be informed by high standards of quality control. Datasets generated from fieldwork must be accurate, complete and consistent - and must adhere to data protection legislation and the ethical best-practices of the discipline or sub-discipline. During the research project, scholars should keep an accurate record of data inputs and data generation. Scholars undertaking surveys and experiments should ensure that anonymisation techniques are documented. 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.

Scholars must comply with the data protection requirements described in Section 2 above. Sensitive data must be anonymised so that no persons, families or households can be directly or indirectly identified in any research data outputs. In those cases where it may be necessary to retain personal data for future research use, scholars may use pseudonymisation techniques (Section 2.d above). Scholars must comply with any end-of-project data destruction requirements indicated by providers (Section 5.e below).

5.b Secure storage and network backup during the research project

During research projects it is important to keep data secure at all times. The EUI ICT Service maintains a strong password policy to prevent the loss, exposure, or corruption of sensitive information.⁸³ The ICT Service advises that the personal G:\ network drive “should always be used to store work-related data.”⁸⁴ The drive is backed up by the ICT Service. Full details about EUI drive quotas are provided by the ICT Service.⁸⁵ (EUI members also have a personal C:\ drive for storing non-work related data (photos &c.) which is not backed up.) The EUI protects computing devices by applying patches, and computers are protected by anti-malware solutions which are updated regularly to prevent malware infections.

5.c Structure, folders, files, variables, format and versioning

The design of research datasets should be carefully considered at the outset of the project. Dataset design varies by discipline, types of data and variables, medium of data, 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 make research data findable, accessible, interoperable and re-usable (FAIR principles).⁸⁶

The folder structure of research datasets (eg. hierarchical / horizontal) should be considered early in the project. Qualitative datasets containing text, interviews, images &c. may require individual files for every element. File names should be standardised and consistent, eg: date, descriptor, version. Variables, such as age, country and sex should be clearly tagged, avoiding special characters and spaces. Temporary identifiers

⁸³ <https://www.eui.eu/ServicesAndAdmin/ComputingService/PolicyDocuments/StrongPasswordPolicy>

⁸⁴ <https://www.eui.eu/ServicesAndAdmin/ComputingService/OfficePCs/StoringData>

⁸⁵ <https://www.eui.eu/ServicesAndAdmin/ComputingService/Storage/DiskQuota>

⁸⁶ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

should be removed from the schema. Files should be systematically named using a standardised and consistent versioning system, eg. YYYY-MM-DD. Changes to the structure should be documented. Versioning systems can be used to track changes to data and code, eg. the GIT (global information tracker) version control resource.⁸⁷

The software format of the dataset should facilitate flexible use of the data. Scholars using one format during a research project, may consider a different format for preservation - taking into consideration open source accessibility. To avoid future file obsolescence, it is best to avoid proprietary formats, where possible. For example .rtf instead of .doc/.docx; .tif instead of .jpg; .flac instead of .mp3 &c. Details of how to submit dataset outputs to the EUI research repository, Cadmus, are in Section 6 below.

5.d Documentation and codebooks

Clear and accurate documentation should be provided about the purpose, context and methodology of the research project and the data outputs. Good documentation helps make datasets findable, accessible, interoperable and re-usable (FAIR principles). Documentation should include a detailed description of the dataset, providing information about structure, folders, files, variables and versioning. Information about definitions, variable names, problematic values, missing observations and weightings should be included. A concise note on methodology should be provided, along with information on how and when the data was collected or generated, and elaboration techniques (sub-setting, combining &c.). At the end of the research project, codebooks, questionnaires and data dictionaries should be reposit with the data.

5.e Deletion of sensitive data at the end of the research project

At the end of a research project, it may be necessary to permanently delete data used in the elaboration phase. Many micro-socioeconomic data providers (eg. Eurostat) require deletion of work-files and non-aggregate outputs. EUI members can use software such as Disk Wipe⁸⁸ which performs byte-by-byte deletion of data, ensuring that data cannot be retrieved. If the provider of the underlying data requires a compliance statement, the name of the software used, the location of the device, the name of the drive and the time of deletion should be recorded. Contact the EUI ICT Service⁸⁹ for assistance with deletion software, and the Library for further information about compliance.⁹⁰

6. REPOSITING AND PRESERVING DATA IN THE EUI RESEARCH REPOSITORY CADMUS

EUI research data can be preserved in the research repository, Cadmus, which is an infrastructure for EUI research outputs, a platform for sharing research data and a solution for funder-mandated research data management compliance.⁹¹ Scholars are encouraged to reposit and - where possible - to share data in compliance with the European Commission's policy on Open Science to spread knowledge "as soon as it is

⁸⁷ https://www.w3schools.com/git/git_intro.asp?remote=github and EUI Library introduction: <https://www.eui.eu/events?id=535720>

⁸⁸ <https://www.diskwipe.org>

⁸⁹ <https://www.eui.eu/ServicesAndAdmin/ComputingService/PolicyDocuments/StorageDisposal>

⁹⁰ resdata@eui.eu

⁹¹ The EUI workflow is explained in Bourke, T. (2022). Bibliographic Control of Research Datasets: reflections from the EUI Library. *JLIS.It*, 13(1), 321–334. <https://www.jlis.it/index.php/jlis/article/view/441/434>

available using digital and collaborative technology.”⁹² Special provisions for the preservation and sharing of qualitative data (audio, transcripts &c.) are described in Section 8 below.

6.a Preparing data for deposit in Cadmus

Datasets presented for deposit in the EUI research repository, Cadmus, must be compliant with data protection regulations pertaining to persons, families and households. Research datasets which are partially based on pre-existing sources must not violate database copyright provisions. Library staff can advise on compliance. The European Commission published a report on Open Science and Intellectual Property Rights in 2022.⁹³ EUI members submitting to the Cadmus repository should create a DATA folder and a DOCS folder. The DOCS folder should contain precise documentation about the methodology, context and discipline of the research project. Where applicable, codebooks should also be presented for deposit. Documentation should be in PDF/a format for preservation. Data should be submitted in the original file format. Large datasets should be submitted in a compressed archival format (eg. .zip, .tgz). Subsets should be accommodated within the DATA folder - not submitted as multiple repository entries. If new iterations of a dataset are planned, scholars can consult with Library staff to determine whether these should be added to the original Cadmus entry (with updated date range), or deposited as a new Cadmus entry. The uploading of datasets and documentation is done by Library staff after the submission of the online submission form. The archiving of more restricted qualitative multi-media outputs (audio, transcripts &c.) is described in Section 8 below.

6.b Completing the dataset submission form

EUI researchers, fellows and professors, from any discipline, can submit research data outputs for deposit in the Cadmus repository. Datasets presented for inclusion must be the output of research by either (i) a current EUI member or (ii) an EUI research team or (iii) a team of scholars with at least one EUI member. The first step is to complete the online submission form.⁹⁴ Library staff will use the submitted information to generate a formal metadata description in the Dublin Core schema which will be used to generate the description page for the dataset in Cadmus. This process helps render the data findable, accessible, interoperable and reusable (FAIR Principles). Library staff will then make an appointment for data transfer. The name of the principal investigator, researcher(s), and - where appropriate - technical collaborator(s), must be provided. The email contacts of all collaborators 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).

By submitting the online data submission form, the P.I. acknowledges that the dataset presented for inclusion in the Cadmus repository is the output of original data generation and/or collection; or the output of significant, value-added elaboration of pre-existing sources. 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, these sources must

⁹² https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

⁹³ https://ec.europa.eu/info/publications/open-science-and-intellectual-property-rights_en

⁹⁴ <https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow>

be listed (publisher, rights owner and/or institutional source). Creators of research data outputs which have been elaborated from pre-existing copyrighted sources may need to obtain the permission of rights owners before research data outputs can be openly shared. Write to the Library for assistance. It is not possible to publish a dataset containing substantial portions of data sourced from pre-existing databases governed by contractual license. Scholars can also reposit datasets in the discipline-specific data repositories indexed in the international re3data registry⁹⁵ and by Google Dataset Search.⁹⁶

EUI members submitting data to the Cadmus repository should state whether or not the dataset can be shared as open data or will be subject to embargo. Library staff can assist. In some cases it may be possible to create a limited, or aggregated, public version of the dataset for open sharing. Embargo status can change over time.

6.c Metadata

Metadata are 'data about data' presented in a formal 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, generation and elaboration. An introduction to metadata standards for social science and humanities' data is maintained by the Digital Curation Centre.⁹⁷ Metadata should be consistent throughout the research project. Some research data outputs may require multi-lingual metadata. EUI Cadmus metadata are generated by Library staff, using information from the online submission form.⁹⁸ These are the principal fields:

NAME(S) OF DATASET CREATOR(S)

The name, or names, of the scholars and (where applicable) 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. 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 IDs are available (eg. ORCID) these should be provided.

EMAIL CONTACT(S)

The email contacts of the dataset creator(s) must be provided. Datasets presented for inclusion in the EUI Cadmus repository must be the output of research by a current EUI member, or an EUI research team, or a team of researchers with at least one EUI member.

TITLE OF DATASET

The title should precisely convey the nature and scope of the dataset. The title should not be identical to the project title, or a related publication title.

⁹⁵ <http://www.re3data.org/>

⁹⁶ <https://datasetsearch.research.google.com/>

⁹⁷ <http://www.dcc.ac.uk/resources/subject-areas/social-science-humanities>

⁹⁸ <https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow>

DESCRIPTION OF DATA

A meaningful abstract, describing the data, the purpose and scope of the research project, and the methodology, must be provided. The Library can help edit data abstracts.

SOURCE(S) OF DATA

The source(s) of the data must be provided. 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 pre-existing resources, these must be listed.

TYPE OF DATA

The type of data must be indicated: eg: statistical; textual; computational; experimental; simulational, observational &c.

YEAR OF COMPLETION OF DATASET

The date of completion of the dataset must be provided. If the dataset is part of a 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

Where applicable, the geographical scope of the dataset (number of countries, sub-states, regions &c.) should be indicated.

FORMAT OF DATA

The file format and version must be given (.csv, Excel, .flac, .rtf, Stata 17, .tif .txt &c.).

CODEBOOK / SUPPORTING DOCUMENTATION

Codebooks and supporting documentation should be provided in PDF/A format.

ACCESS STATUS

The status of access to the data must be indicated. The status 'open data' should be assigned to datasets which can be shared on reposit. If the data is subject to embargo, the expiry date of the embargo should be indicated.

LICENSE

Dataset submitters should choose a license option from the drop-down menu: CC-BY (international)⁹⁹ or CC0.¹⁰⁰

UNIQUE OBJECT IDENTIFIER

Library staff will assign a unique object identifier (handle) to the dataset for the purpose of discovery, linking and citation.

LANGUAGE

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

⁹⁹ <https://creativecommons.org/licenses/by/4.0/>

¹⁰⁰ <https://creativecommons.org/share-your-work/public-domain/cc0/>

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. (The Library can also cross-index publications and datasets after deposit.)

PROJECTED FUTURE WAVES OF DATASET

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

DATASET CITATION

The Cadmus repository generates standardised citations for datasets, eg: CHIOCCHETTI, Paolo, Datasets on legal integration and differentiation in the EU (1958-2020), EUI Research Data, 2023, Robert Schuman Centre for Advanced Studies: <https://hdl.handle.net/1814/75313>

When submitting details of research publications to the EUI Cadmus repository, EUI scholars should cite datasets which support the research findings.¹⁰¹ Further information about data citation is provided by DataCite.¹⁰²

6.d Linking datasets and publications

Datasets deposited in Cadmus can be linked to related publications. EUI members should provide details of publications when submitting related datasets. Subsequent publications can also be cross-linked in Cadmus.

7. OPEN DATA FAIR PRINCIPLES AND OPEN SCIENCE

7.a Data repositories and Library support

Scholars, government agencies and international organisations increasingly share datasets, codebooks and software via the internet. Major data repositories are indexed in the international re3data registry¹⁰³ and Google Dataset Search.¹⁰⁴ The EUI Library provides comprehensive support for open access publishing and open science.¹⁰⁵ Since 2021 the EUI has an internal network of data liaisons, composed of one academic and one administrative member for each academic unit. The Dean of Research Office and the Legal Officer are observers. Data liaisons meet for discussion on best practices related to research data once or twice per year.

7.b Data metadata and access rights

Accurate dataset metadata help scholars determine whether, when, how, where and under what terms, research data outputs can be shared as open data. Not all research

¹⁰¹ <https://cadmus.eui.eu/>

¹⁰² DataCite metadata schema: <https://schema.datacite.org/>

¹⁰³ <http://www.re3data.org/>

¹⁰⁴ <https://datasetsearch.research.google.com/>

¹⁰⁵ <https://www.eui.eu/Research/Library/PublishingAndOpenScience>

data outputs can be openly shared at the conclusion of a research project. The two most important considerations when determining whether a research dataset can be made openly available are: the protection of data pertaining to human subjects (Section 2.a) and database copyright (Section 1.d). Library staff can help EUI members determine whether and when a dataset can be openly shared. Data access status may change over time. Data can be made openly available on deposit; data can be embargoed for a defined period; data can be restricted to on-site ('safe room') access; data can be subject to pre-access registration terms (eg. access permission from data owner); or data can be subject to user contract (sometimes requiring a project proposal). In cases where underlying data observations cannot be shared, it may be possible to create a limited or aggregated public version. Data can also be repositied solely for preservation purposes ('dark archive').

7.c FAIR data principles: Findable, Accessible, Interoperable and Reusable

The FAIR data principles - to make data findable, accessible, interoperable and reusable - provide a framework for good data management practice.¹⁰⁶ Section 3.b above provides information on how to use the FAIR principles when preparing a data management plan. To make data **findable**, datasets should be assigned accurate and consistent metadata; the location of the data should be indicated; the dataset should be clearly structured; and the dataset should have a unique and persistent identifier. Datasets should be **accessible** ("as open as possible, as closed as necessary"). The access status eg. 'open' (accessible immediately) or 'embargoed' (accessible at a later date) should be indicated. Data anonymisation and protection measures should be described. Software and tools required to use the data should be indicated. To make data **interoperable**, standard metadata and naming systems should be used; controlled vocabularies or discipline-specific taxonomies should be explained; variables and other elements should be readable by standard software. To make data **reusable**, an open license should be assigned; either the Creative Commons Attribution International license: CC-BY International or the Creative Commons Public Domain Universal license: CC0; Supporting documentation and codebooks should be repositied with the data to make the observations reusable by other scholars, and information about tools or instruments necessary to reuse the data should be provided.

7.d Pre-registration research plans

Scholars may pre-register research plans before substantive work on the project commences.¹⁰⁷ The Berkeley Initiative for Transparency in the Social Sciences provides resources for research data openness and reproducibility.¹⁰⁸ The GitHub hosting service (a subsidiary of Microsoft Corp.) provides resources for software development, version control and code hosting and sharing.¹⁰⁹

¹⁰⁶ <https://www.go-fair.org/fair-principles/>

¹⁰⁷ <https://www.cos.io/initiatives/prereg>

¹⁰⁸ <https://www.bitss.org/>

¹⁰⁹ <https://github.com/features>

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

This section treats qualitative data generation, use, preservation and access in the humanities and social sciences. Examples of qualitative data include; audio and audio-visual interviews, transcripts, images, minable text, survey diaries, field notes and multimedia materials. The definition of 'data' is broad, as it varies across academic disciplines (eg. history, ethnography, anthropology, sociology &c.). Data collection methodology, dataset structure and data documentation should conform to the best practices of the discipline in which the research is undertaken.¹¹⁰

8.a Protection and ethical use of qualitative data

Scholars generating, accessing, processing and sharing qualitative data should pay particular attention to the data protection and ethical use provisions described in the EUI 'Guide to Good Data Protection Practice in Research'.¹¹¹ Scholars are subject to data protection legislation when collecting, generating, processing, using and storing personal data. The processing of sensitive data relating to health, racial/ethnic origin, religious beliefs, political opinions, genetic/biometric observations, sexual orientation, memberships &c. may require EUI Ethics Committee clearance before data collection and processing begins.¹¹² Ethics Committee approval cannot be sought retroactively. The Ethics Committee review can take up to six weeks. Principal investigators should consult the online Initial Checklist and the supporting documentation on the EUI information page for Ethics and Integrity in Academic Research.¹¹³ When collecting, processing, using and storing personal data, scholars are responsible for obtaining and documenting the informed consent of subjects. The EUI sample consent form can be used.¹¹⁴ Throughout the research data lifecycle, scholars are responsible for preserving the confidentiality of data observations pertaining to human subjects, families and households. Anonymisation and pseudonymisation techniques are described at Section 2.d above. Software such as the open source QualiAnon anonymisation tool can be used for the anonymisation and redaction of interview transcripts.¹¹⁵

Scholars who intend to generate images, videos or photocopies of materials and objects in museums and archives, should seek the rights owners' permission in advance. Scholars who wish to reproduce pre-existing images or multimedia of such materials (eg. photographs in catalogues raisonn e) should seek copyright clearance before reproducing.

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.¹¹⁶ Many of the tools used for the elaboration and analysis of quantitative data (eg. Excel, Python, R, Stata) can be used to generate

¹¹⁰ Learned societies, academies and international research associations provide guidelines by discipline.

¹¹¹ <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

¹¹² <https://www.eui.eu/en/public/about/committees/ethics-committee>

¹¹³ <https://www.eui.eu/en/public/about/committees/ethics-committee>

¹¹⁴ <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/sampleInformedconsentform.pdf>

¹¹⁵ <https://www.qualiservice.org/en/current/news-events-details/anonymization-tool-qualianon-now-usable.html>

¹¹⁶ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/Index.aspx>

aggregate statistical observations from qualitative data. The analysis of restricted personal data may require scholars to work in a 'safe-room' environment. If this is required by a data provider, contact the Library.¹¹⁷

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 (Section 5 above) - there are some additional considerations. Scholars completing data management plans should describe how unique qualitative data materials (eg. primary sources) will be handled and secured. 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 complete and accurate record of all data sources, and to preserve 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 should describe the qualitative research methods used.

8.d Metadata for qualitative data

The metadata schema used for quantitative data outputs (6.c above) can also be used for qualitative data outputs, with some additional considerations. As well as the name(s) of scholars and technical collaborators who generate the research dataset, it may be necessary to acknowledge the authors/creators of pre-existing subsidiary resources. Multi-lingual documentation should be provided where relevant. The format and version of software used to elaborate the data should also be indicated. The Oral History Association (OHA) maintains an online Manual of Best Practices providing guidance on metadata and description.¹¹⁸ The Consortium of European Social Science Data Archives (CESSDA) provides guidelines for data management in the social sciences.¹¹⁹

8.e Qualitative data archiving, preservation and access

From 2023, the EUI Library and the Historical Archives of the European Union provide support for the archiving and preservation of qualitative data, multi-media, transcripts and ancillary research outputs. Audio-visual and other qualitative data, transcripts and ancillary materials can be securely archived at the HAEU as follows.

- Fill in the online submission form¹²⁰
- Upon receipt of the form, EUI Library staff (resdata@eui.eu) will contact the submitter, with a link to a safe online folder (SharePoint) to transfer the data. The SharePoint folder can only be accessed by Library and HAEU staff directly involved with the process of data repositing
- The submitter will also receive more detailed information on how to structure the folder and how to provide metadata.

¹¹⁷ <https://www.eui.eu/Research/Library/ResearchDataServices/SensitiveData>

¹¹⁸ <https://www.oralhistory.org/archives-principles-and-best-practices-complete-manual/>

¹¹⁹ <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/Data-in-the-social-sciences>

¹²⁰ <https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow>

- The submitter will be contacted again by resdata@eui.eu when the repositing of the dataset has been completed. The data will be permanently deleted from the temporary folder on SharePoint.

Qualitative data presented for preservation must be accompanied by documentation describing the resource and, where applicable, indices or codebooks describing the material.

9. EUI INFRASTRUCTURE, SOFTWARE, PROTOCOLS AND SUPPORT

The EUI ICT Service provides infrastructure, software, hardware and connectivity support. Research software programmes are listed on the ICT web site.¹²¹ Technical support is provided by the site offices of the ICT Service.

The EUI ICT Service provides and supports Fortran, Gauss, MATLAB, OxMetrics, Python, R, Stata 17, Stat/Transfer, WinEdt, WinRATS - and the High Performance Cluster.¹²² Advice on the use of statistical software is provided by the ICT Service and software tutors.¹²³ For assistance with functionalities of statistical databases such as Amadeus, Compustat, Datastream, Global Financial Data, S&P Capital IQ Pro and other Library-licensed resources, contact the Library.

The Library's subscription to SAGE Campus provides self-paced interactive learning on Python and R.¹²⁴ The companion SAGE Research Methods provides access to over 1,000 eBooks on all aspects of methodology in the social sciences - including searchable, full-text access to the Quantitative Applications in the Social Sciences series and the Qualitative Research Methods series. The Data Management Training Clearinghouse provides online learning resources for Python, R and other software.¹²⁵

A directory of online research data software manuals, with links to full-texts, is available on the Library web site.¹²⁶ Data software manuals are available at the Badia Library and the Economics Departmental Library (Villa La Fonte) at shelfmarks 001 to 006. The Library holds a substantial collection of works on statistical theory, data science, applied mathematics, probability, calculus, data mining, modelling, optimisation, regression analysis, experimental design, survey methods, time series, artificial intelligence and machine learning (shelfmarks 500-519). Books, eBooks and manuals in any language may be suggested for acquisition by the Library.¹²⁷

¹²¹ <http://www.eui.eu/ServicesAndAdmin/ComputingService/Software/SoftwareResearch.aspx>

¹²² <https://www.eui.eu/ServicesAndAdmin/ComputingService/ParallelComputing> A guide to the HPC is available by writing to HPC.support@eui.eu (EUI members only)

¹²³ <https://www.eui.eu/ServicesAndAdmin/ComputingService/Software/ResearchSoftwareTutors>

¹²⁴ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/SAGEmethods>

¹²⁵ <https://dmtclearinghouse.esipfed.org/home>

¹²⁶ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Software>

¹²⁷ econlibrary@eui.eu

10. INTERNATIONAL RESEARCH DATA GUIDELINES

This guide draws on the international sources listed below, which are strongly recommended to scholars working with research data.

- [Argos online DMP creation tool \(OpenAIRE\)](#)
- [CESSDA Data Management Expert Guide](#)
- [DMPonline data management planner](#)
- [DMP Evaluation Rubric \(Science Europe\)](#)
- [EUDAT collaborative data infrastructure](#)
- [EUI Library Data Portal](#)
- [EUI Research Data Collection in Cadmus](#)
- [EUI Research Data submission form](#)
- [European Data Portal](#)
- [European Open Science Cloud \(EOSC\)](#)
- [Finnish Social Science Data Archive anonymisation guidelines](#)
- [F-UJI automated FAIR data assessment tool](#)
- [GESIS repository - Leibniz Institute for the Social Sciences](#)
- [Good Data Protection Practice in Research \(EUI\)](#)
- [Google Dataset Search](#)
- [Guide to Social Science Data Preparation and Archiving \(ICPSR\)](#)
- [Horizon Europe Programme Guide](#)
- [How and Why You Should Manage Your Research Data \(JISC\)](#)
- [LIBER data management plan catalogue](#)
- [Metadata for Social Science & Humanities \(Digital Curation Centre\)](#)
- [OECD Guidelines on Research Ethics & New Forms of Data](#)
- [Open Data Handbook - Open Knowledge Foundation](#)
- [OpenAire network](#)
- [Practical Guide to Sustainable Research Data \(Science Europe\)](#)
- [Prepare and Manage Data \(UKDS\)](#)
- [re3data registry of data repositories](#)
- [Research Data Alliance](#)
- [Research Data Management Toolkit \(JISC\)](#)
- [The Hague Declaration on Knowledge Discovery in the Digital Age](#)
- [Where to Keep Research Data Checklist \(DCC\)](#)
- [Zenodo data repository \(CERN\)](#)

The hyperlinked version of this Guide is available at:

<https://www.eui.eu/Research/Library/ResearchDataServices>

11th edition, 2023