

Bibliographic Control of Research Datasets: reflections from the EUI Library*

Thomas Bourke^(a)

a) European University Institute

Contact: Thomas Bourke, thomas.bourke@eui.eu

Received: 7 April 2021; **Accepted:** 3 June 2021; **First Published:** 15 January 2022

ABSTRACT

The exponential growth in the generation and use of research data has important consequences for scientific culture and library mandates. This paper explores how the bibliographic control function in one academic library has been expanded to embrace research data in the social sciences and humanities. Library bibliographic control (BC) of research datasets has emerged at the same time as library research data management (RDM). These two functions are driven by digital change; the rise of the open science and open data movements; library management of institutional repositories; and the increasing recognition that data sharing serves the advancement of science, the economy and society. Both the research data management function and the bibliographic control function can be enhanced by librarians' awareness of scholarly projects throughout the research data lifecycle (input, elaboration and output) – and not only when research datasets are submitted for deposit. These library roles require knowledge of data sources and provenance; research project context; database copyright; data protection; data documentation and the FAIR Guiding Principles, to make data findable, accessible, interoperable and reusable. This case study suggests that by creating synergies between the research data management function (during research projects) and the formal bibliographic control function (at the end of research projects) – librarians can make an enhanced contribution to good scientific practice and responsible research.

KEYWORDS

Research data | Datasets | Research data management | Bibliographic control.

* With special thanks, for comments and contributions, to Carlotta Alpigiano (EUI Acquisitions and Library Budget, Co-ordinator), Tommaso Giordano (Former Director, EUI Library), Simone Sacchi (EUI Open Science Librarian), Monica Steletti (EUI Special Collections Librarian), Lotta Svantesson (EUI Repository Manager) and Pep Torn (EUI Library Director). Thomas Bourke is EUI Library information specialist for economics.

© 2022, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. JLIS.it is a journal of the SAGAS Department, University of Florence, Italy, published by EUM, Edizioni Università di Macerata, Italy, and FUP, Firenze University Press, Italy.



1. Introduction

Research libraries have a long history of collecting, managing and providing access to data resources – in particular, statistical data series in support of the social sciences. While there are important differences between disciplines and sub-disciplines, most research libraries had a limited role in the management of their institutions' research data *outputs* until the 21st Century. Today, the collation, bibliographic control, preservation and dissemination of research data are important library functions, due to increasing awareness of datasets as 'first-class' outputs of research.

This case study treats the management and bibliographic control of research dataset outputs in the social sciences and humanities at the Library of the European University Institute (EUI).

While research data management (RDM) is primarily carried out by scholars during research projects, librarians have steadily increased their collaboration and training to fill the skills and capabilities' gap in this area. RDM is undertaken throughout the research data lifecycle, embracing the control of data inputs, the elaboration of data, the protection of data and the creation of research data outputs and documentation. The main reasons for librarians' involvement in research data management include: the exponential growth in the availability and production of digital data; the rise of the open science and open data movements; the establishment of research repositories – frequently managed by libraries; and the increasing recognition that the sharing of research data serves the advancement of science. All of the above are in contexts which require the transfer of knowledge and expertise of library staff – through consultation with, and training of, researchers. While librarians' research data management takes place *during* research projects; bibliographic control normally takes place *after* (or towards the end of) research projects.

Both research data management (Section 3 below) and the bibliographic control of datasets (Section 4 below) require librarians to strengthen their liaison with researchers in order to enhance familiarity with research project design, data generation and use, and research data outputs. Research data management also requires librarians to support the generation of data management plans (DMPs) which are increasingly required by science funders.¹ Support for data management planning raises librarians' awareness of the nature and scope of research projects before final data outputs are presented for deposit. While campus libraries have important roles regarding research data management and bibliographic control, it is acknowledged that – in some institutions – there are lead roles for data centres, ICT services and/or research administration offices.

2. Data, digital change and scientific culture

The generation, collection and use of vast quantities of data – and the retro-digitisation of non-digital collections and content – places research libraries at the vanguard of recent transformation.² In

¹ The European context is described by Filip Kruse and Jesper Boserup Thestrup (Kruse and Thestrup 2018).

² The evolution of library research data roles is analysed by Robin Rice and John Southall (Rice and Southall 2016); by Lynda Kellam and Kristi Thompson, et al. (Kellam and Thompson 2016); and by Rossana Morriello (Morriello 2020).

addition to the volume of data, the tools for the elaboration of data have become more sophisticated. In the social sciences and the humanities, these developments have had an impact on scientific culture – facilitating more empirical and applied research; experimental research; evidence-based policy research, and data-driven methodologies.

The definition of ‘data’ varies across academic disciplines and sub-disciplines and the scope of the term itself has been debated for several decades.³ Data types in the social sciences and humanities include: numerical data, minable text, survey data, experimental data, interview transcripts, archival material, field notes, images, and audio and video recordings. The long history of library expertise in the management of multi-media collections constitutes a solid basis for library curation of research data outputs.⁴

Although definitions vary by discipline, it is useful for librarians to distinguish between collected or acquired ‘databases’ (eg. databases of monographic, journal or statistical content) and individual ‘datasets’ (eg. data outputs from research projects hosted at their institutions). This paper does not treat the traditional library database acquisition and management function (which includes the acquisition, classification, cataloguing and access control of subscription resources; eg. financial market data).⁵ The data treated in this case study are *outputs* generated by university scholars, which are managed, bibliographically controlled, curated, classified and repositied by library staff for the purpose of preservation and – where possible – sharing with other researchers.

The open data movement – an extension of the open access movement – refers to a growing trend whereby government agencies, international organisations and researchers share data outputs, documentation, codebooks and software via the internet. Here it is necessary to distinguish between ‘public data’ and ‘research data.’ Most governments and international organisations provide some level of access to ‘open public data.’ In the research community ‘open research data’ refers to outputs from scholarly research projects which are openly available, usually via institutional repositories.

3. Research data management: library roles

The impact of technological change on scientific culture has necessitated the expansion of library data support roles. The traditional function of acquiring access to subscription databases has been joined by two newer library roles: (i) library support for data-intensive research and research data management *during the research data lifecycle* and (ii) the bibliographic control, collation, reposit and preservation of research data outputs *at the end of the research project*. Research data management during research projects is carried out by both scholars and librarians, complementing their

³ For a theoretical treatment, see the entry “Data” in the Encyclopaedia of Knowledge Organization: <https://www.isko.org/cyclo/data> (International Society for Knowledge Organization, n.d.).

⁴ See Joudrey 2015, chap. 5; and Pradhan 2018.

⁵ At the EUI, these resources are presented in the Library Data Portal: <https://www.eui.eu/Research/Library/Research-Guides/Economics/Statistics/DataPortal> and classified at: <https://www.eui.eu/Documents/Research/Library/Research-Guides/Economics/Statistics/MacroMicroLocations.xls> (Accessed 7 April 2021).

respective expertise.⁶ Bibliographic control, reposit and preservation of research data outputs are carried out by librarians. This is especially true when it comes to datasets in disciplines where the culture of managing (and sharing) research data is not yet fully developed, or where established subject-oriented data repositories (e.g. GenBank, HEPData) do not exist.

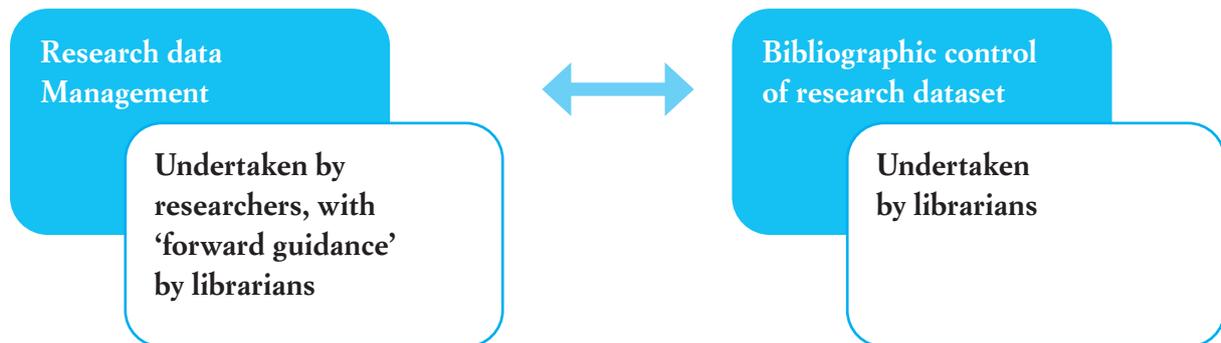


Fig. 1. RDM and BC roles: scholars and librarians

An important component of research data management is the generation of data management plans (DMPs).⁷ A request from a principal investigator for DMP support is often the first point of contact between a librarian and a new research project. Data management plans provide information on how data is generated and/or sourced; how data is organised, used and elaborated; how data – and data subjects – are protected; how data and tools are described and documented; how data is stored and secured during the research project; how data authorship and credit are assigned; how data will be preserved and whether research data outputs can be shared.⁸ The involvement of librarians in data management planning constitutes a solid foundation for the eventual bibliographic control of dataset outputs.

Contemporary research data management is underpinned by the FAIR Guiding Principles, to make data *findable*, *accessible*, *interoperable* and *reusable* – frequently used by librarians to promote awareness of good research data management practices.⁹ Both the research data management function and the bibliographic control function help advance the FAIR Guiding Principles. Library research data management 'forward guidance' is provided via individual user support, library-web documentation and group training. Librarians provide advice that data outputs must be carefully structured, because the 'objects' (outputs) for reposit will be datasets (not unstructured data observations) and that researchers should carefully consider the design of datasets early in their research projects. Dataset structure varies by discipline and sub-discipline, medium, types

⁶ The Consortium of European Social Science Data Archives (CESSDA) maintains an annually-updated Data Management Expert Guide. <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide> (Consortium of European Social Science Data Archives, n.d.).

⁷ Online template tools such as DMPonline <https://dmponline.dcc.ac.uk/>, maintained by the UK Digital Curation Centre (Digital Curation Centre, n.d.), and Argos <https://argos.openaire.eu/splash/>, maintained by OpenAIRE (OpenAIRE, n.d.), can be used to generate structured data management plans.

⁸ <https://www.eui.eu/Research/Library/ResearchDataServices/Guide> (European University Institute Library 2021).

⁹ See Wilkinson, Dumontier, Aalbersberg et al. 2016. Barend Mons provides a practical overview (Mons 2018).

of variables, units of analysis, relationships between data elements, and whether or not the dataset is part of a series. Librarians also explain the importance of clear and consistent naming of folders, files, variables, versioning and documentation, and how good practice helps facilitate findability, accessibility, interoperability and reusability.

Supporting documentation should be updated throughout research projects because, when datasets are presented for deposit, documentation – such as codebooks and questionnaires – must also be submitted. Comprehensive documentation – describing dataset structure, folders, files, variables, versioning and (where applicable) information about problematic values, missing observations and weightings – makes research data findable, accessible, interoperable and re-usable (FAIR). Librarians – who are familiar with a wide variety of data documentation across disciplines and sub-disciplines – can offer feedback on data documentation and help edit dataset abstracts at the time of deposit.

Although all research institutions have data protection officers (DPOs), the library is frequently the first point of contact for scholars who have questions about database copyright and data protection. Librarians' long-standing experience with copyright and terms and conditions of access and use, has been extended to database copyright – which is important when library-licensed databases are used by researchers to generate new research data outputs. In many social science research projects, data outputs are the product of 'mixing' pre-existing data resources (frequently acquired and made available by the campus library) with new project-generated data (eg. surveys and experiments). For librarians, who are 'custodians' of subscription databases, it is important to inform database users of terms and conditions of access and use before research datasets based on licensed resources are openly shared.

During the research data lifecycle – and in collaboration with the DPO – librarians also inform scholars of their data protection obligations regarding the collection, use and security of data observations relating to persons, families and households. Librarians can advise on anonymisation and pseudonymisation techniques – which are particularly relevant for micro-level socio-economic data.

At the library of the European University Institute it is observed that many of the features of research data management (RDM) during the research project overlap with – and help prepare for – formal bibliographic control (BC) at the end – or near the end – of research projects.

4. Bibliographic control, infrastructure and workflow

Due to technological change, the exponential increase in digital content, and the momentum of the open access movement – universities began to establish institutional research repositories at the turn of the 21st Century.¹⁰ Initially these infrastructures only indexed full-text documents and bibliographic records of publications. Gradually research dataset outputs and multi-media have been added, due to an ongoing research culture change towards open science and the increasing requirements of funding agencies. Research scholars also have the option to deposit

¹⁰ Data on the growth of repositories (2000-2020) is available from the Registry of Open Access Repositories: https://en.wikipedia.org/wiki/Registry_of_Open_Access_Repositories. Accessed 7 April 2021.

their data outputs in subject/domain repositories and ‘catch-all’ multi-disciplinary repositories, such as Zenodo.¹¹

The EUI Library launched the Cadmus institutional repository, based on the DSpace infrastructure, in 2003.¹² The beta version of the EUI ResData repository was launched in 2016 and was merged with Cadmus in 2019. University librarians are increasingly aware of data-driven research projects because the campus library is the primary source for subscription databases; researchers usually require access to data software manuals provided by the library; data management plans are supported and reviewed by librarians and researchers frequently approach the library for advice on database copyright and data protection during the research data lifecycle.

However, it is not possible for librarians to be aware of every data-intensive project on campus, as there is no mandate for such information to be shared. Sometimes, librarians will only become aware of research data outputs when a principal investigator, or research team, approaches the library for advice on the preservation, reposit and open sharing of research dataset outputs – often due to funding agency requirements, such as the Horizon 2020 Framework Programme Open Research Data Pilot. Figure 2 provides an overview of the roles of researchers, librarians and ICT staff at the EUI during the research data lifecycle.

	ACTIVITY	RESEARCHERS	LIBRARY	ICT Service
Data input	Data discovery	Researchers discover data via library collections; the internet; and non-digital resources	Maintenance of data portal, indices and OPAC records	/
	Data generation	Researchers generate data (eg. surveys, experiments)	/	/
	Terms of access and use; database copyright and data protection	User compliance	Library promotes awareness of terms and conditions of access and use	ICT service and library provide access protocols
	Data management plans (DMPs)	Researchers write data management plans	Library provides training on DMP template tools and helps edit DMPs	ICT service provides standard description of infrastructure and security



¹¹ <https://zenodo.org/>. Accessed 7 April 2021.

¹² <https://cadmus.eui.eu/>. Accessed 7 April 2021.

	ACTIVITY	RESEARCHERS	LIBRARY	ICT Service
Data elaboration / In-project data management	Dataset structure: folders, files, variables, observations	Researcher activity	Library advisory role	/
	Data anonymisation	Researcher activity	Library advisory role	
	Standardisation of file names, versioning, in-project metadata	Researcher activity	Library advisory role	/
	Documentation, codebooks and associated software/ routines	Researcher activity	Library advisory role	ICT advisory role
	In-project security and backup	Researcher activity	Library advisory role	ICT infrastructure and encryption software
Data output	Submitting research datasets	Researchers submit details of data outputs via online form	Library reviews submission	/
	Bibliographic control and metadata	/	Library checks structure and sources of dataset; converts submission information into metadata	/
	Repositing and infrastructure	/	Library reposit datasets in the institutional repository	Support for institutional repository infrastructure

Fig. 2. Research data lifecycle roles of researchers, librarians and ICT staff

4.1 Data submission to the institutional research repository

Over the past two decades there has been a growing awareness of the scientific value of making research data more openly available. While many academic disciplines have a long history of sharing underlying data within epistemic communities – the open science movement advocates wider access to research data as a public good, of benefit to scientific endeavour. Researchers in the social sciences and humanities are increasingly aware that the academic community is awarding recognition to research datasets as outputs in their own right. Reposited datasets can promote awareness of related publications, or in themselves become part of promotion and tenure procedures. In some cases, researchers become aware of these issues (open data; open science) late in the research project – for example, when an academic colleague or a funding agency requests information about underlying data. Researchers at the European University Institute who submit datasets for reposit are required to complete the library’s online data submission form.¹³ It is important to distinguish between three

¹³ <https://www.eui.eu/Research/Library/ResearchDataServices/EUIResDataWorkflow>. Accessed 7 April 2021.

types of data description activities. Firstly, researchers generate essential descriptors for their data (names of folders, files, tabs, variables &c.) during the research project. These descriptors do not always constitute formal ‘metadata’ in the sense of bibliographic control. Secondly, the observations entered by researchers in the EUI library’s online data submission form constitute ‘raw’ information about a dataset, and are never ingested directly into the repository without review. Thirdly, librarians generate bibliographic-standard metadata for the research repository; to make research datasets findable, accessible, interoperable and reusable.

This case study suggests that the in-project research data management (RDM) function complements the end-of-project bibliographic control (BC) function. The creation of synergies between the two library functions helps contribute to overall scientific quality control.

4.2 Initial review

The EUI library’s online data submission form captures information which is used for verification, provenance and bibliographic control. At the EUI, the name and institutional email address of the principal investigator (or delegated submitter) is required for verification that the submitter is a member of the institution. Only works generated by EUI members – or research teams with at least one EUI member – can be included in the institutional repository. Alumni and former professors can submit datasets if the substantive part of the research was conducted while a full member of the university.

The EUI library became a member of ORCID, the Open Researcher and Contributor ID service, in November 2017. ORCID is a solution for authority control of authors’ name variations across the EUI’s Central Person Registry (CPR); the research repository Cadmus, and the ORCID registry. Both publications and datasets are associated with authors’ ORCID IDs, providing increased visibility for researchers and the institution in the digital environment. EUI authors’ names in the Cadmus repository are linked to the ORCID record – pushing publication and dataset metadata to their ORCID profiles.¹⁴

- When completing the dataset submission form, the names of all creators of the dataset must be listed – including technical collaborators if they have significantly contributed to the creation of the dataset.
- The title of the dataset submitted should not be identical to the title of a project or a publication. Librarians frequently offer suggestions regarding title clarity and, in many cases, titles are modified.
- The online submission form captures both the year of completion of the dataset (which may, or may not, be the current year); the date-range of data coverage – which is of great importance for any time-series data; and (where applicable) the geographical coverage of the dataset.
- Submitters are required to provide a description of the dataset – which is a first draft of the abstract displayed prominently in the repository entry.
- One of the most important submission form fields is the ‘Source(s) of data’. If the dataset is the output of original data collection and elaboration, details should be provided. If the

¹⁴ The complete workflow is explained by Lotta Svantesson and Monica Steletti, in their presentation at the Open Repositories conference (Svantesson and Steletti 2019). The EUI ORCID connect page is at: <https://cadmus.eui.eu/ORCID/>. Accessed 7 April 2021.

dataset is derived from pre-existing sources, those sources should be clearly indicated (data creator, institutional source, publisher).

- The online submission form requires a preliminary statement of whether the data can be made available for open sharing immediately, or is to be repositied under embargo. Librarians help determine this status in consultation with researchers.
- Submitters are required to provide the file format of data files. If the data is in a proprietary format, librarians can recommend (where possible) options for open format versions. This information is always translated into the related media type.¹⁵
- The number of data files within the dataset is an important field which allows librarians to discuss the relationship between the repository entry and the constituent elements. In some cases, it is necessary to create two entries (works) for a dataset which the submitter might be submitting as a single work. In other cases, multiple data submissions from the same project might be consolidated into one entry with multiple sub-sets.
- The online form also requests information regarding projected future waves of the dataset being submitted. This can require that a data sub-set which is intended to have future iterations, might need a separate entry in the repository.
- The online form also gathers information about supporting documentation, codebooks and (where applicable) software routines to enable the use of the data by others.
- The library advises on the appropriate reuse licence for open research data; eg: Creative Commons Attribution (CC-BY) or Public Domain (CC0).
- Submitters are asked to include references to related publications. This information can also be added when publications become available.

4.3 Provenance

The establishment of research repositories in universities and other institutions requires librarians to have a strong role regarding provenance. While research documents (working papers, theses, articles, chapters, monographs &c.) are normally subject to editorial review either inside the university or by external peer reviewers and publishers – the situation regarding research data outputs is more complex.

Very few universities have formal faculty-level ‘editorial’ review procedures for dataset outputs. The research data lifecycle is predominantly undertaken by researchers – with support from library and ICT professionals. At the end of research projects, the library becomes involved in issues of provenance, originality, data protection and database copyright. Here it can be seen that there is an overlap between the research data management (RDM) function and the bibliographic control (BC) function.

Although there are multiple ways in which librarians can undertake verification and provenance, it is impossible for librarians and information specialists to have detailed knowledge of data in every discipline and sub-discipline. It is also impossible for librarians to guarantee that every element and observation in a dataset is correct.

¹⁵ Formerly known as MIME Type: https://en.wikipedia.org/wiki/Media_type. Accessed 7 April 2021.

At the EUI, librarians build trust with researchers during the data lifecycle as part of the research data management function – informing scholars that;

By submitting this [online submission] form, EUI members acknowledge that the dataset for deposit is the output of original data collection and elaboration; or is the output of significant, value-added, elaboration of pre-existing sources; and conforms with the EUI *Guide to Good Data Protection Practice in Research*.¹⁶

Librarians build trust with researchers through outreach and training; assistance with data management plans; provision of in-project services during the research data lifecycle and advice about database copyright, data protection, research ethics, scholarly reputation and scientific impact. When data is presented to the library for deposit, the ‘Source(s) of data’ field in the online submission form reveals whether the dataset output is partially based on pre-existing, library-licensed resources. At this point, it may be necessary for library staff to liaise with data suppliers to control for potential license issues regarding the open sharing of derivative datasets via the university repository.

4.4 Metadata generation

The generation of metadata about research datasets renders research datasets findable, accessible, interoperable and reusable (FAIR) and helps librarians decide whether research data outputs can be shared as open data. The research data management activities undertaken by librarians during research projects constitute a solid foundation for library bibliographic control and metadata generation. At the EUI, librarians use the raw information from the online dataset submission form to generate repository metadata using:

- The Dublin Core schema
- Library of Congress subject headings
- Dewey Decimal 23 classification
- A modified UN/Eurostat classification originally developed for the paper-format statistics collection¹⁷ and,
- An internal data series identifier.

When setting up the EUI Research Data Collection structure in the Cadmus institutional repository, the EUI’s institutional setup was reflected in the sequential, internal ID (dc.identifier.other) – eg: EUI_ResData_00032_HEC. The numeric value is a running sequence, with alpha-suffixes for:

- Economics: ECO
- History and civilisation: HEC
- Law: LAW
- Social and political sciences: SPS and
- The inter-disciplinary Robert Schuman Centre for Advanced Studies: RSC.

Here follows an example of the metadata record for a dataset deposited in 2020.

¹⁶ <https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf> (European University Institute 2019).

¹⁷ <https://www.eui.eu/Research/Library/ResearchGuides/Economics/StatisticsClassification>. Accessed 7 April 2021.

Informal politics of codecision dataset	
dc.contributor.author	BRESSANELLI, Edoardo
dc.contributor.author	HERITIER, Adrienne
dc.contributor.author	KOOP, Christel
dc.contributor.author	REH, Christine
dc.coverage.spatial	European Union
dc.coverage.temporal	1999-2009
dc.date.accessioned	2020-09-09
dc.date.available	2020-09-09
dc.date.created	2014
dc.date.issued	2020
dc.identifier.other	EUI_ResData_00028_RSC
dc.identifier.uri	https://hdl.handle.net/1814/68095
dc.description	1 data file; 1 documentation file
dc.description.abstract	This dataset, created as part of the research project on ‘The Informal Politics of Codecision’ - funded by the Research Council of the European University Institute (EUI) and the Economic and Social Research Council (ESRC; Grant RES-000-22-3661) - is constituted by all 797 legislative files concluded under codecision between 1999 and 2009. It presents a new variable, ‘early agreement’, indicating whether legislation has been agreed informally, in trilogues, by the Council of Ministers and the European Parliament. It also includes variables with characteristics of the legislative file (legal nature, policy area, complexity, media salience, policy type, duration) and of the legislative negotiators (priorities of the Council Presidency, ideological distance between the Parliament’s rapporteur and the national minister, the Presidency’s workload).
dc.format	Excel file
dc.format.mimetype	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
dc.language.iso	en
dc.publisher	European University Institute, RSCAS
dc.relation.ispartofseries	EUI Research Data
dc.relation.ispartofseries	2020
dc.relation.ispartofseries	Robert Schuman Centre for Advanced Studies
dc.rights	info:eu-repo/semantics/openAccess
dc.rights.uri	http://creativecommons.org/licenses/by/4.0/
dc.subject	Legislative bodies
dc.subject.classification	FS-CA
dc.subject.ddc	328.4077
dc.subject.lcsh	Legislative bodies - European Union countries



dc.title	Informal politics of codecision dataset
dc.type	Dataset
eui.subscribe.skip	TRUE
dc.rights.license	Creative Commons Attribution 4.0 International
dc.description.version	The dataset documentation is available in: BRESSANELLI, Edoardo, HERITIER, Adrienne, KOOP, Christel, REH, Christine, The informal politics of codecision : introducing a new data set on early agreements in the European Union, EUI RSCAS, 2014/64, EUDO - European Union Democracy Observatory -- Retrieved from Cadmus, European University Institute Research Repository, at: http://hdl.handle.net/1814/31612

Fig. 3. Example of metadata record, full view, from the EUI Cadmus repository Research Data Collection

This case study suggests that the in-project research data management (RDM) function complements the end-of-project bibliographic control (BC) function. The generation of metadata about research datasets helps to make research datasets findable, accessible, interoperable and reusable (FAIR). For example, the unique and persistent identifier helps researchers to find the dataset; the retrievability of the metadata via the repository protocol helps make the dataset accessible; the Dublin Core schema allows for broad sharing and interoperability, and the license information facilitates reusability.

4.5 Transfer and uploading of datasets

When EUI librarians have prepared the metadata record for the dataset, an appointment is made for the transfer of data, documentation and (where applicable) codebooks. At this stage, there may be further discussions about structure, format, provenance, copyright and data protection. Once the dataset is received and approved, the metadata file, the dataset and the documentation are uploaded in the Research Data Collection of the EUI Cadmus repository. A digital object identifier is generated and a data citation can be exported, eg:

BRESSANELLI, Edoardo, HERITIER, Adrienne, KOOP, Christel, REH, Christine, *Informal politics of codecision dataset*, EUI Research Data, 2020, Robert Schuman Centre for Advanced Studies. Retrieved from Cadmus, European University Institute Research Repository, at: <https://hdl.handle.net/1814/68095>

The repository metadata schema allows further discovery of the resource, for example via library discovery tools and online aggregator services.¹⁸ Accurate bibliographic control will also facilitate forthcoming machine discoverability of datasets and artificial intelligence applications.

¹⁸ The EUI's Cadmus repository is interoperable with, and harvested by, CORE, Google Scholar, OpenAIRE, RePEC and Worldcat.

5. Conclusion

Contemporary research data management is underpinned by the FAIR Guiding Principles, to make data findable, accessible, interoperable and reusable. Both the research data management (RDM) function and the bibliographic control (BC) function can be combined in the service of these principles.

Based on the experience of EUI library staff – this paper suggests that research data management during research projects and bibliographic control at the end of research projects are complementary elements of an emerging ‘continuum’ of library support for modern scientific culture – contributing to overall scientific quality control.

References

- Consortium of European Social Science Data Archives. n.d. "Data Management Expert Guide." Accessed 7 April 2021. <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>.
- Digital Curation Centre. n.d. "DMPonline data management tool." Accessed 7 April 2021. <https://dmponline.dcc.ac.uk/>.
- European University Institute. 2019. *Guide to Good Data Protection Practice in Research*. Accessed 7 April 2021. <https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf>.
- European University Institute Library. 2021. *Research Data Guide*. Accessed 7 April 2021. <https://www.eui.eu/Research/Library/ResearchDataServices/Guide>.
- International Society for Knowledge Organization. n.d. "Data". In *Encyclopaedia of Knowledge Organization*. Accessed 7 April 2021. <https://www.isko.org/cyclo/data>.
- Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. 2015. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited.
- Kellam, Lynda, and Kristi Thompson, eds. 2016. "Databrarianship: the Academic Data Librarian" In *Theory and Practice*. Chicago, IL: Association of College and Research Libraries.
- Kruse, Filip, and Jesper Boserup Thestrup, eds. 2018. *Research Data Management: a European Perspective*. Berlin: De Gruyter Saur.
- Mons, Barend. 2018. *Data Stewardship for Open Science: implementing FAIR Principles*. Boca Raton, FL: CRC Press.
- Morriello, Rossana. 2020. "Birth and Development of Data Librarianship." *JLIS.it* 11 (3): 1-15. <http://dx.doi.org/10.4403/jlis.it-12653>.
- OpenAIRE. n.d. "Argos data management tool." Accessed 7 April 2021. <https://argos.openaire.eu/splash/>.
- Pradhan, Sanghamitra. 2018. *Cataloguing of Non-Print Resources: a Practical Manual*. New Delhi: Ess Publications.
- Rice, Robin, and John Southall. 2016. *The Data Librarian's Handbook*. London: Facet Publishing.
- Svantesson, Lotta, and Monica Steletti. 2019. "DSpace ORCID integration: name authority control solution at the European University Institute." Presented at the The 14th International Conference on Open Repositories (OR2019), Hamburg, Germany. <https://doi.org/10.5281/ZENODO.3553926>.
- University of Southampton. "Registry of Open Access Repositories." Accessed 7 April 2021. <http://roar.eprints.org/>.
- Wilkinson, Mark D., Michel Dumontier, Barend Mons et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.