Part of

Mapping research data at the University of Bologna: Data Management Plan

Description

Mapping Research Data at the University of Bologna: Dataset

Description

This description concerns the table containing all information estracted from the analysed DMPs (i.e., types of research data, their formats, accessibility, the type of project, funding), related to projects coordinated by the University of Bologna.

Each row of the table corresponds to a data entry, i.e. a minimum unit (i.e., a file) that makes up a dataset described in a DMP. A dataset may have several data entries.

This table serves as input for R code written to analyse the data and explore some research questions.

1. What types of data are produced and managed by researchers at the University of Bologna?

2. What could be trends of problems and patterns useful to improve the Data Stewardship service?

3. Is there interdisciplinarity in data production at the University of Bologna?

 DOI: - 16/12/2024

The outputs of the analysis are graphs of various types, from histograms to sankey diagrams.

Researchers

Bianca Gualandi (orcid:0000-0001-8202-8493), Sara Coppini (orcid:0000-0002-6279-3830), GIULIA CALDONI (orcid:0000-0001-6105-721X), MARIO MARINO (orcid:0000-0003-1499-8425)

Funder

Grant

Description

1 Your research output

1.1 Contact details and responsibilities

1.1.1 Contact person(s)

Contact Name	Contact ID	E-mail address	Affiliation
Coppini, Sara	0000-0002-6279-	sara.coppini2@unib	ALMA MATER

	3830	o.it	STUDIORUM - Università di Bologna

1.1.3 Research output creator(s) and contributor(s

Contributor Name	Contributor ID	E-mail address	Affiliation	Role
Caldoni, Giulia	0000-0001-6105-721X	giulia.caldoni2@unibo.it	ALMA MATER STUDIORUM - Università di Bologna	Creator
Coppini, Sara	0000-0002-6279-3830	sara.coppini2@unibo.it	ALMA MATER STUDIORUM - Università di Bologna	Creator
Gualandi, Bianca	0000-0001-8202-8493	bianca.gualandi4@unibo.it	ALMA MATER STUDIORUM - Università di Bologna	Creator
Marino, Mario	0000-0003-1499-8425	mario.marino6@unibo.it	ALMA MATER STUDIORUM - Università di Bologna	Creator

1.1.4 Institution(s) involved

ALMA MATER STUDIORUM - Università di Bologna

1.2 Resources for data management

1.2.1 What resources will be dedicated to data management and ensuring that data will be FAIR?

To achive a FAIR data management, around 2-5% of creators' time is needed.

1.3 Identifying the research output

1.3.2 Which type of research output are you describing?

Dataset

2 Summary

2.1 Identification of the dataset

2.1.1 Are you re-using an already existing dataset?

No

2.1.3 Creating a new dataset

Due to its very innovative and specific nature, it has been necessary to generate new data for this research activity, right from the sources. In fact, since we had to collect data exclusively concerning the University of Bologna, and coming from DMPs compiled with our support as data stewards, it was impossible for us to reuse existing data.

2.2 Dataset characteristics

2.2.1 Dataset name

Mapping Research Data at the University of Bologna: Dataset

2.2.2 Data types

a. Processed

b. Tabular

c. Structured

d. Quantitative

e. Qualitative

2.2.3 Data formats

a. CSV

b. XLSX

2.2.4 Size of data

kilobytes (KB)

Comment:
Around 75 KB

2.2.5 How is the data generated/collected

Firstly, we have analysed a limited number of DMPs and of Grant Agreements (documents regulating the administrative and financial aspects of EU-funded projects and describing in detail the planned research activities) of projects we had supported in order to identify the type of information to collect, i.e. our variables of interest. Then, we collaboratively entered the information for each variable of interest into the table, based on what was effectively written in the DMPs.

2.2.6 Why is the data generated/collected

The objective of data collection in this dataset has been to organise the information of interest for our analysis in a structured format, to be then machine-readable and analysed by a computer system and computational tool (i.e., in our case, R).

The dataset could also be useful outside our specific research activity, as the variables are rather generic and thus could serve as a model for other similar research in other institutions or research centres.

2.2.7 Dataset description

This is a table containing all information estracted from the analysed DMPs, related to projects coordinated by the University of Bologna. For example: types of research data, their formats, accessibility, the type of project, funding, etc.

Each row of the table corresponds to a data entry, i.e. a minimum unit (i.e., a file) that makes up a dataset described in a DMP. A dataset may have several data entries.

3 Ethics and legal aspects

3.1 Personal data

3.1.1 Are you handling personal data?

Yes

3.1.2 Solutions adopted for handling personal data

We never collected data that could directly identify people and we anonymized all project information (i.e. acronyms, institutional Departments) that could indirectly lead to their identification.

3.2 Legal aspects

3.2.1 Are there any legal issues associated to your ouput? (e.g. IPR or valorization)

No

3.3 Other ethical aspects

3.3.1 Are there other ethical issues associated to your ouput?

No

4 Making research FAIR

4.1 Long-term preservation

4.1.1 Selecting what to preserve

We anonymized all project information (i.e. acronyms, institutional Departments), which could indirectly lead to the identification of people, from the data to be preserved long term.

We have no problems with the size of the file or its technical specifications for filing, as it is under 100 KB and in an extremely accessible and interoperable format.

4.1.2 Chosen repository/ies

Name	Link	Is your repository OpenAIRE compliant?	Is your repository catalogued in re3data?	What type of repository is it?
Zenodo	https://zenodo.org/	Yes	Yes	Generalist

4.2 Findability

4.2.1 Persistent identifier(s)

Type of PID (e.g., DOI, Handle)	PID (e.g., doi.org/10...)
DOI	10.5281/zenodo.14234555

4.2.2 Metadata

Zenodo register DOIs with DataCite, and allows to export the metadata according to many different standards.

4.2.3 Keywords

research data management, open science, data management plan

4.3 Accessibility

4.3.1 Is the research output openly accessible?

Yes

4.3.3 Accessing the research output

The dataset is accessible and openly available via the repository: https://doi.org/10.5281/zenodo.14234555

4.3.4 Is metadata openly accessible?

Yes

4.4 Interoperability

4.4.1 Methodologies

The methodology for data generation and processing is described in detail on Protocols.io at https://dx.doi.org/10.17504/protocols.io.n2bvj87jpgk5/v2

4.4.2 Vocabularies, taxonomies and other standards

We have defined new taxonomies when necessary, i.e., when we have not found any that adhere to our type of investigation in terms of purpose and method (e.g., "reasons of inaccessibility"). Existing taxonomies, either generalist (e.g., DataCite, MIUR settori scientifico disciplinari) or institutional (e.g., UniBO taxonomy for the 5 subject areas of academic research) have been reused when appropriate. We will expand these initial taxonomies or (occasionally) make changes to them if new typologies of data or other aspects not previously considered will emerge during the analysis.

For the field "data type" we reused the taxonomy proposed by DataCite, specifically we reused some of the controlled values for the element 10.a resourceTypeGeneral (http://purl.org/dc/terms/). We reused those in line with the definition of data chosen in this work, so we selected: Audiovisual, ComputationalNotebook, Image, InteractiveResource, Report, Software, Sound, Standard, Text, Workflow, Other, Model, Tabular. The latter has been renamed thus by us, whereas in the original scheme it would be 'dataset'. We made this renaming choice because we found the term "dataset" confusing, since we already use it in the sense of "set or collection of data" (as it is understood in the DMPs that are the subject of our analysis) and since the datacite definition for "dataset" corresponds to the concept of "tabular data" or "structured data" (cf. "Data encoded in a defined structure", https://datacite-metadata-

schema.readthedocs.io/en/4.5_draft/appendices/appendix_1/resourceTypeGeneral.html#dataset).

For some fields, the possible values are those of the UniBO taxonomies.

- "creator's unit" and "project unit": we used the departments in UniBO ("nd" when data is new but creator's name is yet to be defined, and "ext" for "external" when data is reused and thus created by a person external to unibo, since we are not interested in tracking that information if UniBO is not involved in the data reuse or generation)
- "subject area" for which we considered the 5 areas of research as defined by UniBo: Arts, Humanities, and Cultural Heritage (shortened: Humanities); Science; Economics and Management (shortened: Economics); Engineering; Medicine.

4.5 Reusability

4.5.1 Licensing

CC 0, "No Rights Reserved", https://creativecommons.org/share-your-work/public-domain/cc0/

4.5.2 Documentation

We provide a README file to document the software, which is deposited on the repository alongside the research output. We also provide a detailed protocol (https://dx.doi.org/10.17504/protocols.io.n2bvj87jpgk5/v2) as a further type of documentation.

5 Quality and security

5.1 Quality of the research output

5.1.1 Measures implemented to ensure quality

To check the consistency and quality of data, we carried out the following processes: representation with controlled vocabularies, encoding of null values to avoid incomplete data, peer review of data table to avoid inaccurate data.

5.2 Security measures

5.2.1 Storage

Cloud Storage Services

Comment:

MS OneDrive/Sharepoint was chosen as a cloud storage service, since the University of Bologna has signed a contract with Microsoft. In addition, the platform provides to all necessary security strategies, such as back-up and two-factor authentication.

5.2.2 Backup

Semi-automatic backups are provided by the cloud-storage service used and the different file versions are also periodically saved locally (local hard drive/computer) by at least one of the team members.