

# La diversità dei dati della ricerca: una strategia comune in un contesto multidisciplinare

Sara Coppini



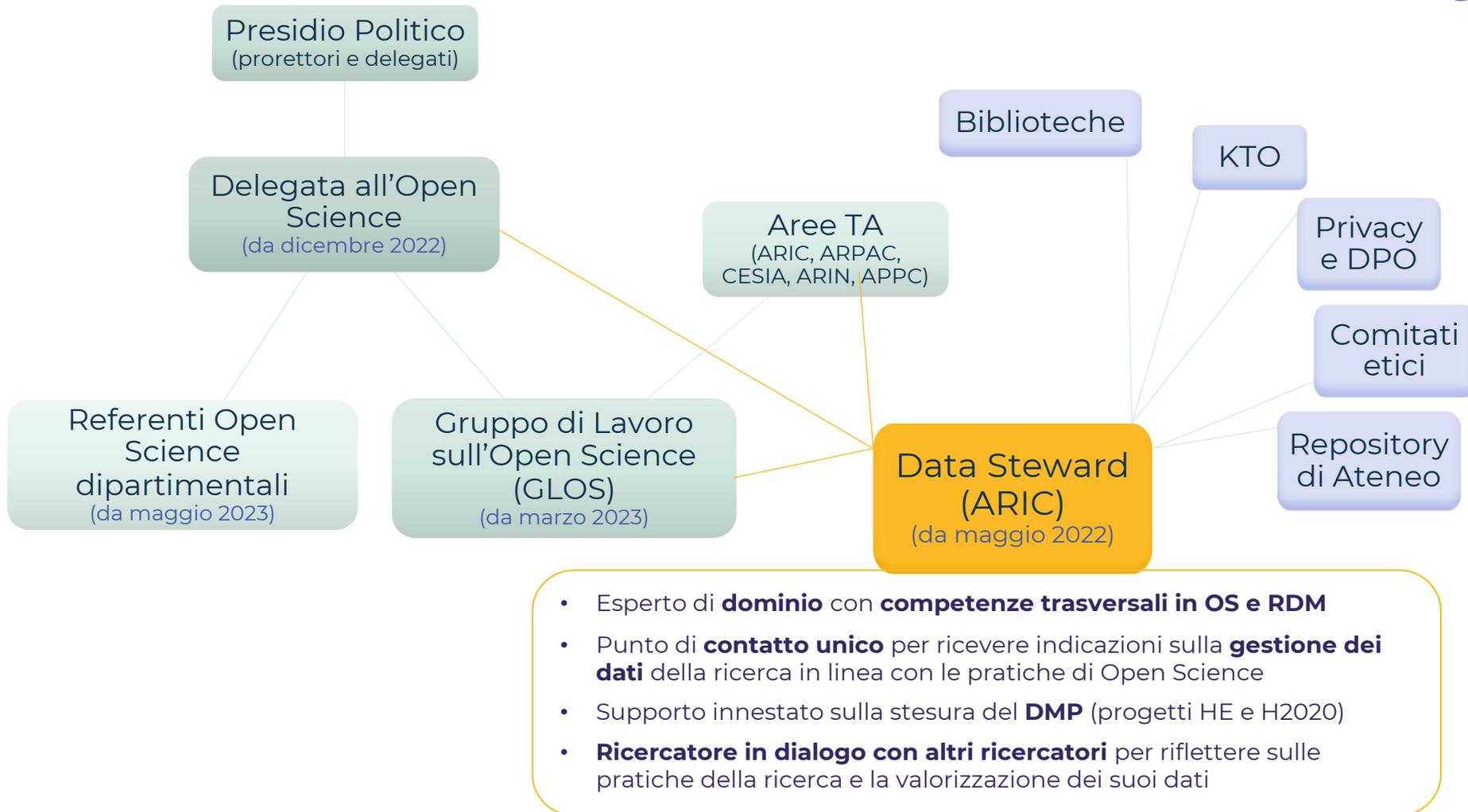
ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

ConfGARR23  
**SAPERI INTERCONNESSI**



# Contesto istituzionale

## Unibo per l'Open Science



# Contesto istituzionale ed europeo

## Unibo per l'Open Science in linea con EOSC



Riconoscimento della **centralità del dato** e della sua valorizzazione da parte di EOSC e dell'UE come ente finanziatore – specialmente in **ambito disciplinare** con la **condivisione di buone pratiche e metodologie** in linea con i principi FAIR e dell'Open Science per massimizzare il valore e l'impatto della ricerca scientifica: *“field-specific data stewards (...) are essential for successful implementation of FAIR and take-up by researchers”*.

Il servizio di **data stewardship in UniBO** è centralizzato presso ARIC ma impostato per **specificità disciplinare nella gestione del dato**: 4 data steward per ogni macro-area disciplinare.

La **mappatura dei dati per ambito disciplinare** si pone come strumento di riflessione sulle pratiche per una gestione efficace in linea con l'Open Science



### Why does research data matter?

Research data are a primary output of research.

### Six Recommendations for Implementation of FAIR Practice

Shaping Europe's digital future

#### A European Strategy for data

The strategy for data focuses on putting people first in developing technology, and defending and promoting European values and rights in the digital world.

Data is an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general.

The **European strategy for data** aims at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty.

European Commission. Directorate General for Research and Innovation. & EOSC Executive Board. (2022). *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)*. Publications office of the European Union. <https://data.europa.eu/doi/10.2777/935288>

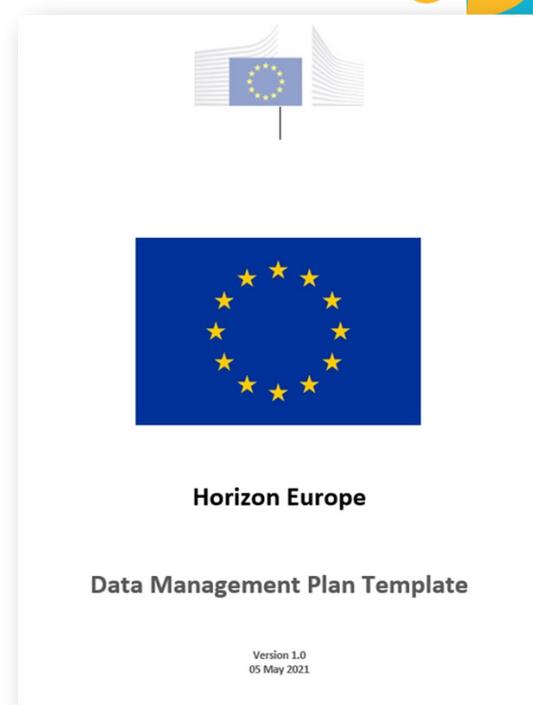
<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>, <https://www.scienceeurope.org/our-priorities/research-data/>  
European Commission, Directorate-General for Research and Innovation, *Six Recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group* – , Publications Office, 2020, <https://data.europa.eu/doi/10.2777/986252>

# La mappatura

**Riflessione teorico-metodologica sui dati di ricerca** prodotti, gestiti o riutilizzati nel contesto dei progetti competitivi europei a cui l'ateneo partecipa.

La ricerca si basa sulle descrizioni dei dati all'interno del **Data Management Plan (DMP)** dei **progetti Horizon Europe e Horizon 2020** in cui UniBO è coinvolta o come **coordinatore** o come **partner responsabile del DMP**.

data_type	data_content	format
corpora	textual corpora of selected sources	xml
text	physics and mathematical points for equations	json, dat
tabular	genetic data on model organisms	csv
text	census of different neural network architectures	json
3D model	3D reconstructions (e.g. neural)	obj
tabular	maps of finger movements - combination of mapping algorithms	csv
image	visual results obtained from deformable linear objects (DLOs) graspi	tiff
text	settings for prototypes: component descriptions, kinematics of the c	docx
sound	recording of interviews and field notes for multi-sited ethnographic	wav
geospatial data	GIS mapping data from satellite	gis
software	AI-on-Demand Platform software	py



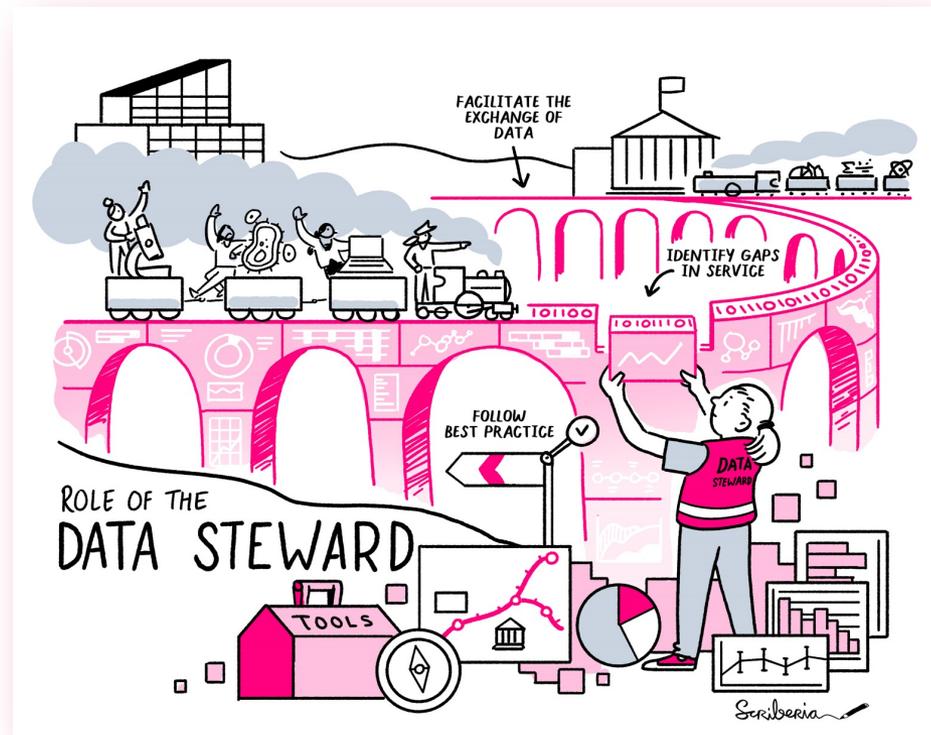
Analisi dei dati descritti dei DMP su diversi livelli:

- **tipologia di dato a livello formale** (es. immagine, testo, video)
- **contenuto** (es. scansione di un documento del secolo XIII)
- **formati** in cui i dati vengono depositati (es. JPG, SVG)

# La mappatura: obiettivi

L'obiettivo principale è di individuare **esigenze e problemi nella gestione dei dati**:

- tra le peculiarità disciplinari e i trend comuni – analisi con **focus disciplinare** in linea con le aree della data stewardship all'UniBO
- per **migliorare il servizio di data stewardship** e definire le strategie per supportare operativamente la ricerca affinché questa sia aperta, trasparente ed affidabile
- per lo sviluppo di **risorse, policy** e di **infrastrutture e modelli** (es. schema di metadatazione transdisciplinare) a supporto dei ricercatori

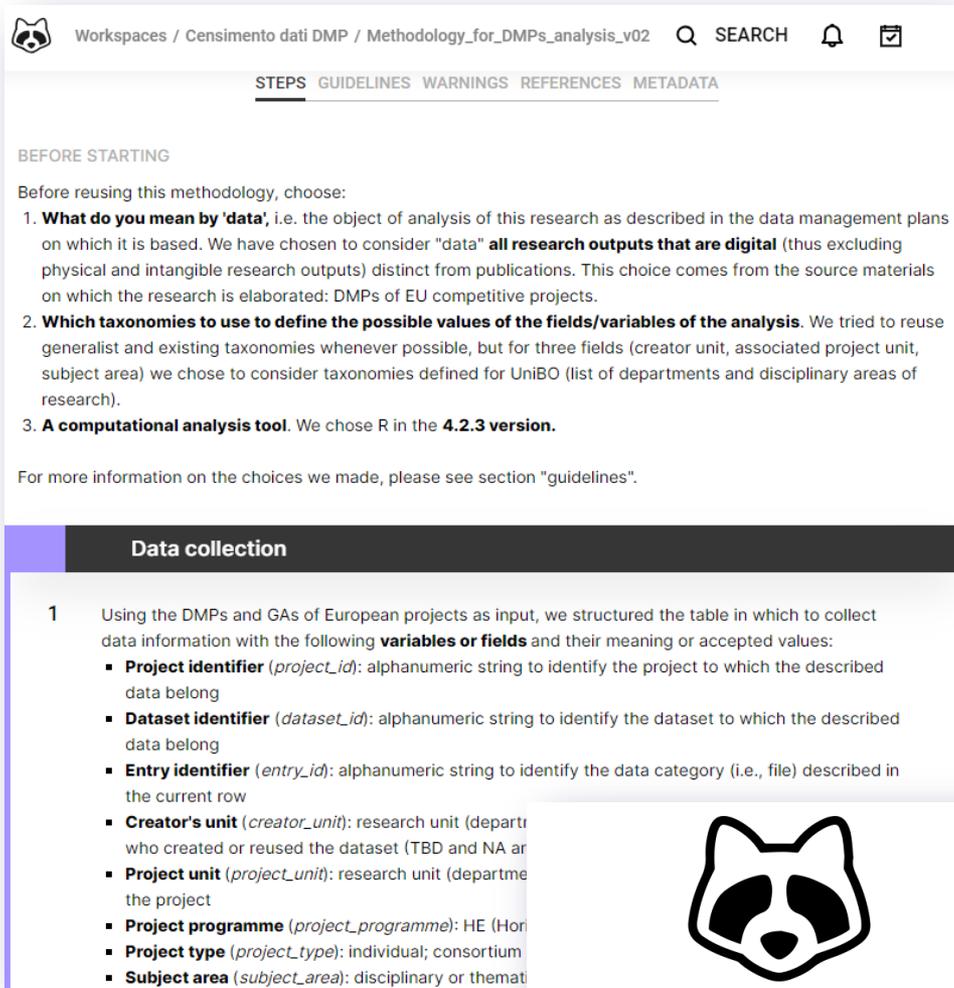


The Turing Way Community, & Scriberia. (2023). Illustrations from The Turing Way: Shared under CC-BY 4.0 for reuse. Zenodo. <https://doi.org/10.5281/zenodo.7587336>

# La mappatura: metodologia

La definizione, strutturazione e validazione di una metodologia di lavoro è il primo passo per lo sviluppo della ricerca.

- In linea con i principi dell'Open Science: metodologia **trasparente, aperta e riproducibile**.
- Sviluppo di una metodologia computazionale attraverso **strumenti Open e FAIR: protocols.io**.
- I **risultati** della ricerca (protocollo, dataset di test, codice, dataset finale di input e dataset finale di output) saranno **depositati in formato aperto** non proprietario con **licenza CC BY**.



Workspaces / Censimento dati DMP / Methodology\_for\_DMPs\_analysis\_v02

STEPS GUIDELINES WARNINGS REFERENCES METADATA

BEFORE STARTING

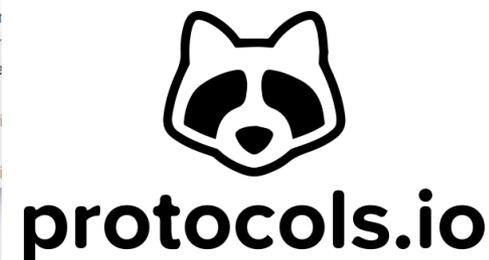
Before reusing this methodology, choose:

1. **What do you mean by 'data'**, i.e. the object of analysis of this research as described in the data management plans on which it is based. We have chosen to consider "data" **all research outputs that are digital** (thus excluding physical and intangible research outputs) distinct from publications. This choice comes from the source materials on which the research is elaborated: DMPs of EU competitive projects.
2. **Which taxonomies to use to define the possible values of the fields/variables of the analysis**. We tried to reuse generalist and existing taxonomies whenever possible, but for three fields (creator unit, associated project unit, subject area) we chose to consider taxonomies defined for UniBO (list of departments and disciplinary areas of research).
3. **A computational analysis tool**. We chose R in the **4.2.3 version**.

For more information on the choices we made, please see section "guidelines".

**Data collection**

- 1 Using the DMPs and GAs of European projects as input, we structured the table in which to collect data information with the following **variables or fields** and their meaning or accepted values:
  - **Project identifier** (*project\_id*): alphanumeric string to identify the project to which the described data belong
  - **Dataset identifier** (*dataset\_id*): alphanumeric string to identify the dataset to which the described data belong
  - **Entry identifier** (*entry\_id*): alphanumeric string to identify the data category (i.e., file) described in the current row
  - **Creator's unit** (*creator\_unit*): research unit (department) who created or reused the dataset (TBD and NA are not allowed)
  - **Project unit** (*project\_unit*): research unit (department) of the project
  - **Project programme** (*project\_programme*): HE (Higher Education)
  - **Project type** (*project\_type*): individual; consortium
  - **Subject area** (*subject\_area*): disciplinary or thematic





# La mappatura: step e timeline

Lettura dei DMP e GA dei progetti seguiti ad aprile 2023

Identificazione delle informazioni che si vogliono mappare, i.e., identificazione delle variabili per l'analisi dati

Creazione e strutturazione di un protocollo su protocols.io: inserimento di linee guida e step per la raccolta dati, analisi dati e pubblicazione dei risultati

Creazione di un dataset fittizio per sviluppo e test del codice per l'analisi dei dati

Test della metodologia secondo il codice e il dataset fittizio (con eventuali correzioni)

Applicazione della metodologia a tutti i DMP consegnati da maggio 2022

Analisi dei risultati e deposito degli output della ricerca: dataset iniziale, codice, dataset di analisi

Studiare i dati di ricerca nel contesto dei progetti EU vuole essere l'occasione per evidenziare le esigenze specifiche del ricercatore, promuovere un adeguato supporto alla gestione del dato secondo principi FAIR e sviluppare risorse e infrastrutture.



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

**Contatti:**

**[aric.datasteward@unibo.it](mailto:aric.datasteward@unibo.it)**

ConfGARR23

**SAPERI INTERCONNESSI**