

OPEN SCIENCE DALLA A ALLA Z

6 - EOSC, FAIR, DMP



OpenAIRE



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Photos are mine, available for reuse on Flickr, <https://www.flickr.com/photos/eg65/albums/>



Elena Giglia
25 giugno 2020
elena.giglia@unito.it
[@egiglia](https://twitter.com/egiglia)

Impareremo in questo modulo

1. Cosa significa FAIR nella ricerca quotidiana

2. come scrivere un DMP

MESSAGGI CHIAVE

- FAIR è il futuro
(se no si resta tagliati fuori)
- È più facile di quanto sembri, basta provarci
- DMP online è pronto all'uso, basta fare pratica

Dati FAIR



FAIR train – GoFAIR video

FARM DATA TRAIN



PERSONAL HEALTH TRAIN



...FAIR SIGNIFICA [anche e soprattutto per le macchine]

FINDABLE

- IDENTIFICATIVI
- METADATI

INTEROPERABLE

- STANDARDS
- ONTOLOGIE



ACCESSIBLE

- DOVE SONO CONSERVATI E A QUALI CONDIZIONI DI ACCESSO
 - **NON SIGNIFICA «OPEN»**
 - FORMATI APERTI

REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE
- LEGGIBILI DALLE MACCHINE

FAIR software

Towards FAIR principles for research software

Data Science, June 12 2020

Issue title: FAIR Data, Systems and Analysis

Guest editors: Paul Groth and Michel Dumontier

Article type: Position Paper

Authors: Lamprecht, Anna-Lena^{a,*} | Garcia, Leyla^b | Kuzak, Mateusz^{c,d} | Martinez, Carlos^e | Arcila, Ricardo^f | Martin Del Pico, Eva^g | Dominguez Del Angel, Victoria^h | van de Sandt, Stephanieⁱ | Ison, Joni^j | Martinez, Paula Andrea^k | McQuilton, Peter^l | Valencia, Alfonso^{m,n} | Harrow, Jennifer^o | Psomopoulos, Fotis^p | Gelpi, Josep LL^{q,r} | Chue Hong, Neil^{s,t} | Goble, Carole^u | Capella-Gutierrez, Salvador^{v,w}

2. Software is not data

Technically, software is a special kind of data. In computing, digital data (ultimately sequences of ones and zeros) are used to represent all information, including factual data as well as computer instructions. In the more abstract context of FAIR, software and data are regarded as different kinds of digital research objects next to each other. As such, they share particular characteristics that allow them to be shared, such as the possibility of having a Digital Object Identifier (DOI) and a Creative Commons license. However, as elaborated by Katz

Quality aspects concerning the form of software can be considered as covered by FAIR, in particular by the interoperability and reusability principles. It is important to realise that unlike data, software is not static and can only be (re)used if it is sustainable and evolves along with the continuous development of the entire software ecosystem. The quality of its codebase is decisive for a software's ability to evolve sustainably. This characteristic is often also referred to as maintainability, and

Summary of the proposed FAIR principles for research software and how they relate to the FAIR Guiding Principles for data. It is indicated whether a given FAIR data principle has been simply rephrased to adjust it to software, extended to cover a broader scope, reinterpreted to match the different context, discarded as it does not apply, or newly proposed as it only applies for research software

	FAIR for data	FAIR for software	Operation
F1	(Meta)data are assigned a globally unique and persistent identifier.	Software and its associated metadata have a global, unique and persistent identifier for each released version.	Rephrased
F2	Data are described with rich metadata.	Software is described with rich metadata.	Rephrased
F3	Metadata clearly and explicitly include the identifier of the data it describes.	Metadata clearly and explicitly include identifiers for all the versions of the software it describes.	Rephrased and extended
F4	(Meta)data are registered or indexed in a searchable	Software and its associated metadata are included in a searchable software	Rephrased

I2	(Meta)data use vocabularies that follow FAIR principles.	–	Reinterpreted, extended and split
I2S.1	–	Software and its associated metadata are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I2S.2	–	Software use and produce data in types and formats that are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I3	(Meta)data include qualified references to other (meta)data.	–	Discarded
I4S	–	Software dependencies are documented and mechanisms to access them exist.	Newly proposed



Nov. 20, 2018

Final Report and Action Plan
from the European
Commission Expert Group
on FAIR Data

TURNING FAIR INTO



Define

Concepts for FAIR implementation

Rec. 1: Define FAIR for implementation

Rec. 2: Implement a Model for FAIR Digital Objects

Rec. 3: Develop components of a FAIR ecosystem

Rec. 16: Apply FAIR broadly

Rec. 17: Align and harmonise FAIR and Open data policy

Implement

FAIR culture

Rec. 4: Develop Interoperability frameworks

Rec. 5: Ensure data management via DMPs

Rec. 6: Recognise & reward FAIR data & stewardship

Rec. 18: Cost data management

Rec. 19: Select and prioritise FAIR digital objects

Rec. 20: Deposit in Trusted Digital Repositories

Rec. 21: Incentivise reuse of FAIR outputs

FAIR ecosystem

Rec. 7: Support semantic technologies

Rec. 8: Facilitate automated processing

Rec. 9: Certify FAIR services

Rec. 22: Use information held in DMPs

Rec. 23: Develop components to meet research needs

Rec. 24: Incentivise research infrastructures to support FAIR data

Skills for FAIR

Rec. 10: Professionalise data science & stewardship roles

Rec. 11: Implement curriculum frameworks and training

Above line = priority recommendations

Below line = supporting recommendations

Embed and sustain

Incentives and metrics for FAIR data and services

Rec. 12: Develop metrics for FAIR Digital Objects

Rec. 13: Develop metrics to certify FAIR services

Rec. 25: Implement and monitor metrics

Rec. 26: Support data citation and next generation metrics

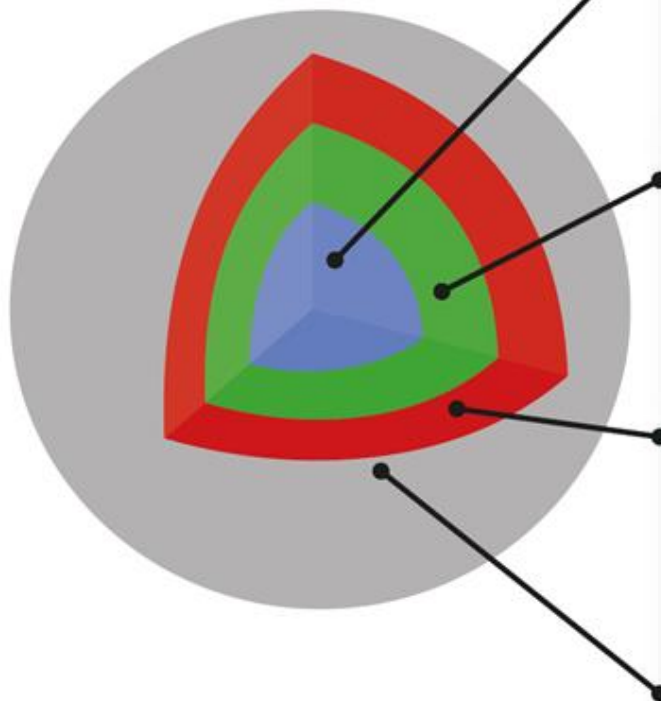
Investment in FAIR

Rec. 14: Provide strategic and coordinated funding

Rec. 15: Provide sustainable funding

Rec. 27: Open EOSC to all providers but ensure services are FAIR

FAIR data: the ideal FAIR object



DIGITAL OBJECT

Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and support citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to process and analyse the data.

METADATA

Contextual documentation

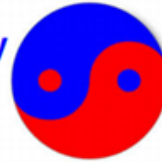
In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

FAIR: technology VS domain

FAIR Principles

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Technology



Domain-relevant content

Findable:

F1 (meta)data are assigned a globally unique and persistent identifier;

F2 data are described with rich metadata;

F3 metadata clearly and explicitly include the identifier of the data it describes;

F4 (meta)data are registered or indexed in a searchable resource;

Interoperable:

I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2 (meta)data use vocabularies that follow FAIR principles;

I3 (meta)data include qualified references to other (meta)data;

Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2 the protocol allows for an authentication and authorization procedure, where necessary;

A2 metadata are accessible, even when the data are no longer available;

Reusable:

R1 meta(data) are richly described with a plurality of accurate and relevant attributes;

R1.1 (meta)data are released with a clear and accessible data usage license;

R1.2 (meta)data are associated with detailed provenance;

R1.3 (meta)data meet domain-relevant community standards;

FAIR è graduale



EOSC FAIR
Executive Board Working Group

Feb. 2020

Interim recommendations on FAIR Metrics for EOSC
February 2020
Draft for consultation

Françoise Genova, Observatoire Astronomique de Strasbourg, ORCID [0000-0002-6318-5028](https://orcid.org/0000-0002-6318-5028)
Jan Magnus Aronsen, University of Oslo, Norway, [0000-0003-2593-1744](https://orcid.org/0000-0003-2593-1744)
Oya Beyan-Fraunhofer FIT, Germany, [0000-0001-7611-3501](https://orcid.org/0000-0001-7611-3501)
Natalie Harrower, Digital Repository of Ireland, [0000-0002-7487-4881](https://orcid.org/0000-0002-7487-4881)
András Hóli, Hungarian Academy of Sciences, [0000-0002-6873-3426](https://orcid.org/0000-0002-6873-3426)
Rob W.W. Hoof, Dutch Techcentre for Life Sciences, [0000-0001-6925-9439](https://orcid.org/0000-0001-6925-9439)
Leslie McIntosh, Research Data Alliance, [0000-0002-3507-7468](https://orcid.org/0000-0002-3507-7468)
Pedro Principe, University of Minho, [0000-0002-8588-4196](https://orcid.org/0000-0002-8588-4196)
Kostas Repanas, European Commission, [0000-0002-7848-2834](https://orcid.org/0000-0002-7848-2834)
Ana Slavec, InnoRenew CoE, [0000-0002-0171-2148](https://orcid.org/0000-0002-0171-2148)
Sarah Jones, Digital Curation Centre, [0000-0002-5094-7126](https://orcid.org/0000-0002-5094-7126)

Edited by Françoise Genova & Sarah Jones, Co-Chairs of the EOSC FAIR Working Group

Inclusiveness: consider FAIR as a process

If FAIR is not seen as a continuum, we risk losing communities who are not well advanced in sharing their data in a FAIR way, as well as advanced communities for whom the effort to attain optional indicators doesn't outweigh the effort required. In addition to avoiding "mandatory" criteria, using multi-step maturity scales to measure the FAIRness level of a resource, instead of a yes/no evaluation for each criterion, would provide an inclusive system, and a way to set up goals and measure progress. A bar representation of the levels of compliance of all the criteria would provide a visualisation of the level of FAIRness of a given resource.

... sfumature di FAIR

Findable

Does the dataset have any identifiers assigned?

No identifier

Is the dataset identifier included in all metadata records/files describing the data?

No

How is the data described with metadata?

The data is not described

What type of repository or registry is the metadata record in?

The data is not described in any repository

Accessible

How accessible is the data?

No access to data or metadata

Is the data available online without requiring specialised protocols or tools once access has been approved?

No access to data

Will the metadata record be available even if the data is no longer available?

Unsure



The screenshot shows the ANDS Training page. At the top, there are search bars for 'Research Data' and 'ANDS Site'. Below these are navigation tabs: 'About us', 'News and Events', 'Partners and Communities', 'Working with data', 'Online Services', and 'Guides and resources'. The 'Working with data' tab is selected, leading to a page titled 'FAIR data training'. On the left, a sidebar lists 'The FAIR data principles', 'FAIR webinar series (Aug/Sep 2017)', 'FAIR data training' (selected), 'Findable', 'Accessible', 'Interoperable', 'Reusable', and 'FAIR data training narkane'. The main content area includes social media links and a list of resources for FAIR data training, such as a basic checklist, a more comprehensive breakdown, a self-assessment tool, and case studies.

MA SOGGETTIVO...



home

news

events

programs

about

FAIR self-assessment tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

<https://www.ands-nectar-rds.org.au/fair-tool>

FAIR maturity evaluator

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

patience ! If you notice any unexpected failures in the tests, please report them to mark.wilkinson@upm.es



Import MI Tests

Import Maturity Indicators Tests as YAML, smartAPI interface annotation

Get started



Create collections

Assemble Maturity Indicators Tests into community centered collections

Get started



Evaluate resources

Evaluate resources FAIRness against Collections of Maturity Indicator Tests

Get started

Evaluating FAIR maturity through a scalable, automated, community-governed framework

Mark D. Wilkinson , Michel Dumontier, Susanna-Assunta Sansone , Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Mercè Crosas & Erik Schultes 

Scientific Data 6, Article number: 174 (2019) | Download Citation 

13 Altmetric | Metrics 

Sept. 20, 2019

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.



FAIR METRICS GEN2 - IDENTIFIER PERSISTENCE

FAIR evaluation service

Philosophy of FAIR testing

Status: Failure

Principle tested: F1

Description: Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema). For URLs that don't follow a schema in FAIRSharing we test known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).

Created on: Feb 18, 2019 by [Mark D Wilkinson](#) (updated on Feb 20, 2019).

Test results

INFO: The metadata GUID appears to be a URL. Testing known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).

FAILURE: The metadata GUID does not conform with any known permanent-URL system.

- OGGETTIVO
- LEGGIBILE DALLE MACCHINE –
COME I DATI FAIR

FAIR Data management wizard

Data Stewardship Wizard

Data integration

7

Data interpretation

3

Information and insight

14

Is there any pre-existing data?

Are there any data sets available?

 Data Stewardship for Open Science

☐ No

☒ Yes 

Will reference data be created?

Will any of the data that you create be shared with others?

 Data Stewardship for Open Science

☐ No

☒ Yes 

Will you be storing samples?

ELIXIR
Data Stewardship Wizard

Smart Data Management Plans for FAIR Open Science

For serious researchers and data stewards

 Data Stewardship Wizard

[Questionnaire Demo](#)

[Log In](#)

[Sign Up](#)

Will you store licenses with the data?

What's up?

Always consider the use of your data beyond the original purpose. One of the issues with re-using other people's data is that they cannot be assumed to be reusable from an ethical or legal standpoint without explicit permission. Assuming that unlicensed data are 'free to use for whatever purpose' is intrinsically wrong, and in the case of for instance Pharmaceutical industry can lead to court cases later on. Therefore, whenever you publish a data set or any other kind of information or digital object, it is important to define a license for reuse. For software many licences exist, and for data, increasingly, standard licenses are available or under development. Please note that a given license is also a defined concept and therefore deserves a Persistent Identifier and a URI pointing to where the license can be studied (machine readable licenses are also under development in some areas). This means that in the metadata, the license under which the data or the workflow can be reused is 'just another PID in the right place'. Users can then specify in their search or workflow container that 'only data with the following licenses should be included'.

For instance, if you include some data in your analysis that cannot be used for commercial purposes, that decision may render your entire results not usable for commercial purposes (at least in the view of some lawyers). This means that not licensing your data at all, even if you don't care who uses them and for what purpose is very counterproductive and will severely undermine the actual reuse of your data by others and in particular by industry. It will also lower the attribution-rate (usually part of the license conditions) and thus the citation and the impact score of your data.

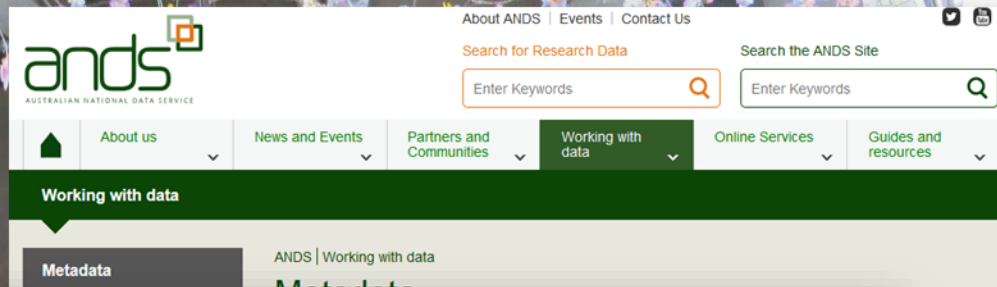
Do

- Always carefully choose a license to be attached to your data upon publication.
- Include and clearly mark the licenses PID as a concept + attributes in the metadata.
- Store and 'expose' the license as part of the metadata in Open Access environments where search engines can easily find the license, even of the data they describe are not (yet) FAIR or even highly restricted in access. The 'fact' that a data set with a specific license is 'out there' is a first step toward effective reuse of your data or information source.
- Make sure, especially when you restrict use of your data, that you are able to enforce the license you choose. Licenses that are not enforceable make no sense. (please note that the enforcement is usually not done by an individual research group but at institutional or repository level)

Don't

- Ever publish data without a license attached or choose a license lightly, without considerations of anticipated reuse of your data.
- Choose a license that is not transitive (i.e. can not be transferred with subsets of the data), but make sure its transitivity does not unduly restrict the reuse of your data.
- Choose an unnecessary complicated license with many clauses and wherever possible one that is already widely adopted in the research community for either software

F = Findable. Metadata



Types of metadata	Goal	Example
Descriptive metadata	<i>The minimal metadata, required to find a digital object.</i> <i>If there are additional contextual metadata, a user will have a better idea on how to use the data</i>	<i>Author, title, abstract, date</i> <i>Contextual metadata are for example location, time, data collection method (tools)</i>
Structural metadata	<i>These link the individual objects of a unity</i>	<i>Links to related digital objects, (e.g. the article written based on the linked research data)</i>
Technical metadata	<i>Information on the technical aspects of the data set</i>	<i>Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard</i>
Administrative metadata	<i>Metadata focusing on user rights and management of digital objects</i>	<i>License, possible reasons for an embargo, waivers</i>

F = Findable. Metadata standards

Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

Arts and Humanities [Edit](#)

- [Archaeology](#) [Edit](#)
- [Creative art and design](#) [Edit](#)
- [Heritage Studies](#) [Edit](#)
- [Historical and Philosophical Studies](#) [Edit](#)
- [History by Area](#) [Edit](#)
- [History](#) [Edit](#)
- [Law](#) [Edit](#)
- [Music](#) [Edit](#)

Engineering [Edit](#)

- [Architecture](#) [Edit](#)
- [Building Conservation](#) [Edit](#)

Life Sciences [Edit](#)

- [Agricultural Economics](#) [Edit](#)
- [Agricultural Science](#) [Edit](#)
- [Animal pathology](#) [Edit](#)
- [Animal physiology](#) [Edit](#)
- [Biochemistry](#) [Edit](#)
- [Biodiversity](#) [Edit](#)
- [Bioengineering](#) [Edit](#)
- [Biogeography](#) [Edit](#)
- [Bioinformatics](#) [Edit](#)

Physical Sciences & Mathematics [Edit](#)

- [Astronomy](#) [Edit](#)
- [Astrophysics](#) [Edit](#)
- [Cartography](#) [Edit](#)
- [Chemistry](#) [Edit](#)
- [Climatology](#) [Edit](#)
- [Crystallography](#) [Edit](#)
- [Environmental Science](#) [Edit](#)
- [Geology](#) [Edit](#)
- [Geoscience](#) [Edit](#)
- [Glaciology](#) [Edit](#)
- [Hydrogeology](#) [Edit](#)
- [Hydrography](#) [Edit](#)
- [Hydrology](#) [Edit](#)
- [Marine Science](#) [Edit](#)
- [Maritime Geography](#) [Edit](#)
- [Materials Science](#) [Edit](#)
- [Meteorology](#) [Edit](#)
- [Minerology](#) [Edit](#)
- [Nuclear and Particle Physics](#) [Edit](#)
- [Oceanography](#) [Edit](#)
- [Palaeontology](#) [Edit](#)
- [Physics](#) [Edit](#)
- [Planetary science](#) [Edit](#)
- [Remote Sensing](#) [Edit](#)
- [Soil Science](#) [Edit](#)
- [Solar physics](#) [Edit](#)

Social and Behavioral Sciences [Edit](#)

- [Anthropology](#) [Edit](#)
- [Demography](#) [Edit](#)
- [Economics](#) [Edit](#)
- [Geography](#) [Edit](#)
- [Health Policy](#) [Edit](#)
- [Human and Social Geography](#) [Edit](#)
- [Planning \(Urban, Rural and Regional\)](#) [Edit](#)
- [Politics](#) [Edit](#)
- [Sociology](#) [Edit](#)

General Research Data [Edit](#)

- [Multi-disciplinary](#) [Edit](#)

F = findable. Metadata tools

What CEDAR does

<https://metadatacenter.org/>

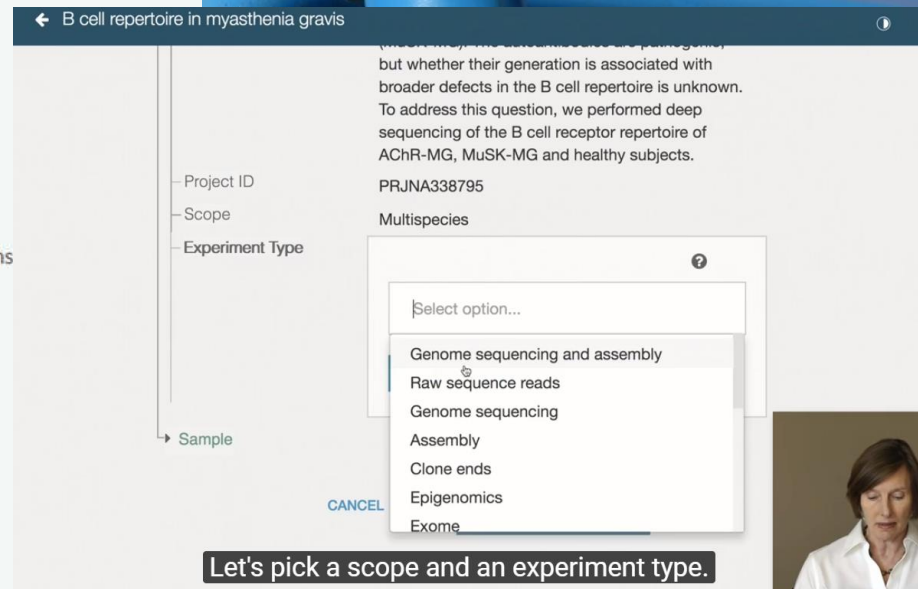
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools will create a metadata record to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools help users collect, store, and download the information that users have provided.

What can CEDAR do for me already?

As of its production release, in February 2017, CEDAR addresses these scenarios:

- create user-friendly, shareable forms for collecting metadata, with features like
 - nested and repeatable elements and fields
 - reusable elements
 - control over tool tips, field titles, and field descriptions
- share your forms and metadata
 - provide a link to your metadata editors, so they can enter metadata responses based on your forms
 - share your forms and other content with individuals or a group
 - create and manage groups to make permissions simpler
- associate your questions (fields) and possible answers (values) with controlled terms
 - select any term or collection of terms from the NCBO BioPortal semantic repository
 - combine different terms from different controlled vocabularies into a single set of options
 - create your own terms, or term lists ('value sets') that can be re-used
- view responses meeting your (simple) search criteria, in several forms
 - CEDAR Metadata Editor's metadata view
 - an in-line JSON-LD format, used by CEDAR for all its metadata instances
 - download of JSON-LD files via the [CEDAR REST API](#), for offline integration with your workflow
- use the Workbench Desktop interface to manage your content
 - use My Workspace to see your items, or Shared with Me to see other items you can access
 - select an item and control-click or use the 3-dot menu in the upper right to share it, copy it, delete it, or get info on it
- enable intelligent metadata suggestions in your template by using a field's Suggestions tab
 - CEDAR keeps track of metadata entered for that field
 - users will see a drop down list of the most popular metadata entries, and can select from them
- remotely access CEDAR content and capabilities using the [CEDAR REST API](#)

With these capabilities, you can capture simple or rich metadata for your project, build a repository of project metadata, or design particular needs. Advanced users can even submit metadata entries through CEDAR's REST API.



Let's pick a scope and an experiment type.

F = Findable. Persistent identifiers

The screenshot displays the DataCite website interface. At the top, the DataCite logo is on the left, and navigation links for 'About us', 'Services', 'Resources', 'Community', and 'Become a member' are on the right. A large banner reads 'WELCOME TO DATACITE' with the tagline 'the leading global provider of DOIs for research data.' Below this, a section for ORCID is featured, including a search bar, navigation tabs for 'FOR RESEARCHERS', 'FOR ORGANIZATIONS', 'ABOUT', 'HELP', and 'SIGN IN', and a registration process titled 'DISTINGUISH YOURSELF IN THREE EASY STEPS'. The steps are: 1. REGISTER (Get your unique ORCID identifier), 2. ADD YOUR INFO (Enhance your ORCID record), and 3. CITE YOUR RESEARCH (Cite your research source). A 'LATEST NEWS' section is also visible. At the bottom, there are three service boxes: 'Find our registry to find datasets, software, images, and other research material.' (with a binary code graphic), 'Find an appropriate repository to access and deposit research data with re3data.org', and 'Generate your references automatically with our easy-to-use citation formatting tool.' (with a citation formatting icon). The URL <https://www.datacite.org/> is displayed at the bottom right.

ORCID
Connecting Research and Researchers

FOR RESEARCHERS **FOR ORGANIZATIONS** **ABOUT** **HELP** **SIGN IN**

SIGN IN REGISTER FOR AN ORCID ID LEARN MORE

6,055,250 ORCID IDs and counting. See more

We need your feedback! Please tell us about your understanding and perceptions of ORCID and your experience of using your ID by completing our [community survey](#). Thank you!

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more](#)

1 REGISTER Get your unique ORCID identifier [Register now!](#)
Registration takes 30 seconds.

2 ADD YOUR INFO Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn). <https://orcid.org/>

3 CITE YOUR RESEARCH Cite your research source with confidence, and receive proper credit when your work is re-used.

LATEST NEWS
Tue, 26 Feb 2019
Construyendo una Infraestructura para Apoyar a los Investigadores - Una entrada a...

re3data.org
Find an appropriate repository to access and deposit research data with re3data.org

Generate your references automatically with our easy-to-use citation formatting tool.
<https://www.datacite.org/>

ORCID – ID. Tutto intorno a me

CREATE E MANTENETE UN
PROFILO ORCID. ATTIVA TUTTI I
MECCANISMI!

ORCID INTEGRATIONS

based on: <https://orcid.org/organizations/integrators/integration-chart>

AAAS AMS

AGU APS

MLA acm

SfN OSA

ROYAL SOCIETY OF CHEMISTRY AAC

AIP OXFORD UNIVERSITY PRESS ELSEVIER

IOP Publishing PNAS SPRINGER NATURE

SAGE Taylor & Francis WILEY

F1000 Research Hindawi PLOS

Funders

wellcome trust NIH osti.gov NHS NIHR

Societies

Generating / sharing data

astrophysics data system

figshare SSRN

DRYAD Europe PMC

scienceOPEN KUDOS

publons Peerage of Science PRE Peer Review Evaluation

PLUM ANALYTICS Altmetric Impactstory

SYMPLECTIC Elements Clarivate Analytics

Crossref über research alyc

ProQuest OCLC airiti

Publishers

Outreach Assessment Discovery

A = Accessible

ACCESSIBLE≠OPEN
«ACCESSO» PUÒ ANCHE ESSERE
RISERVATO O SOTTO EMBARGO

- **Open access**

Data that can be accessed by any user whether they are registered or not.
Data in this category should not contain personal information unless consent is given (see '[Informed consent](#)').

- **Access for registered users (safeguarded)**

Data that is accessible only to users who have registered with the archive.
This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

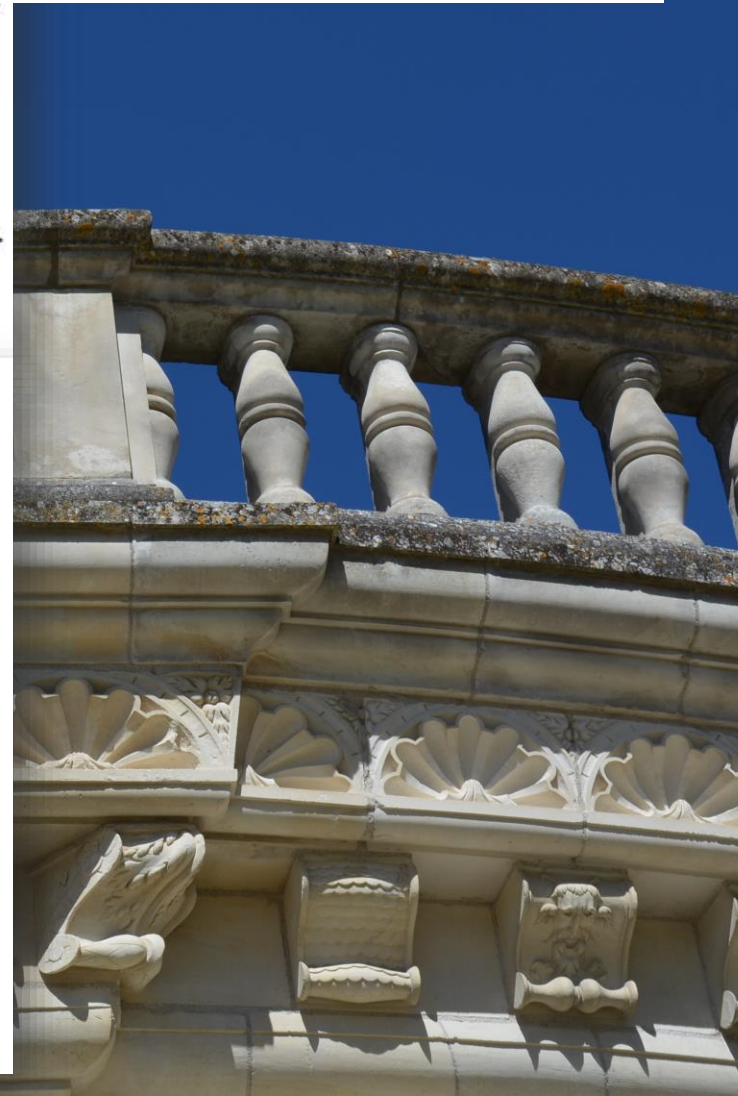
- **Restricted access**

Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.

Restricted access requires the long-term commitment of the researcher or person responsible for the data to handle the upcoming permission requests.

- **Embargo**

Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after a certain period of time.



HARVARD
Dataverse

We've changed hosting! Contact us if you see any issues.

Metrics 3,074,518 Downloads

Share, archive, and get credit for your data. Find and cite data across all research fields

Search this dataverse... Find Advanced Search

☒ Datasets (2,537) 1 to 10 of 78,308 Results

☒ Datasets (75,771)

☐ Files (372,751)

Dataverse Category
Research Project (774)
Researcher (722)
Organization or Institution (238)

Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multi-component, community-based intervention in Zimbabwe
Feb 27, 2018 - Replication Studies Dataverse
Yu, Fang; Hein, Nicholas; Bagenda, Danstan, 2018, "Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multi-component, community-based intervention in Zimbabwe", doi:10.7910/DATaverse.V1

DRYAD About For researchers For organizations Contact us Log in Sign up

Publishers: Simplify data submission. Strengthen links between articles and data. For free.

Integrate your journal with Dryad today »

Submit data now
How and why?

Search for data
Enter keyword, author, title, DOI, etc. Go
Advanced search

Latest from @datadryad

Tweets by @datadryad

Dryad @datadryad
Excited to announce submission integration with JAMIA Open, a new gold OA title from @AMIAinformatics! academic.oup.com/jamiaopen #OpenData

Browse for data
Recently published Popular

Recently published data
Barnett T, Smith G, Young RJ (2017) Data from: The tonic immobility test: do wild and domesticated mantella frogs (Mantella aurantiaca) have the same response? PLOS ONE 12(10):1-10. doi:10.5061/dryad.v172f

zenodo Search Upload Communities

All versions Found 359191 results.

Access Right
☐ Open (338962)
☐ Closed (19395)
☐ Restricted (328)
☐ Embargoed (306)

File Type
☐ Png (128118)
☐ Pdf (117520)
☐ Jpg (49381)
☐ Zip (30738)

February 27, 2018 (v1.0.1) Software Open Access
airr-community/airr-standards: Early revision of AIRR definitions
Ahmad Syed, Christian Busse, Uri Laserson, Scott Christley, Jason Vander
An early revision of the AIRR definitions with corresponding reference lists
Uploaded on February 27, 2018

February 27, 2018 (v0.9.7) Software Open Access
gperrea/fqtrim: fqtrim release v0.9.7
Geo Perreae
filtering and trimming next generation sequencing reads
Uploaded on February 27, 2018

Open Science Framework
A scholarly commons to connect the entire research cycle

figshare search on figshare Browse Upload

store, share, discover research

get more citations for all of the outputs of your academic research over 5000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

General depositories for research data

The following depositories are of interest to researchers in all domains:

- Zenodo (not-for-profit, hosted by CERN): <https://zenodo.org>:
- Dryad (not-for-profit membership organisation): <http://www.datadryad.org>
- Figshare (free service provided by private company): <https://figshare.com>
- Open Science Framework (not-for-profit, developed and maintained by the Center for Open Science¹): <https://osf.io>
- Harvard Dataverse (not-for-profit, hosted by the Institute for Quantitative Social Studies IQSS at Harvard University): <https://dataverse.harvard.edu>

A = Accessible. Data lab [in progress]



Save your data during research

Data lab

Many research projects involve collaboration with several researchers, and a large amount of research data is generated. To manage such quantities of data and to allow the researchers to share them with each other, 4TU.Centre for Research Data offers the possibility of establishing a 'data lab'.

A data lab is a secure online environment (with or without screening) for storing, processing and sharing dynamic research data, software, visualisations and other items with fellow researchers.

DataverseNL



Members of the three technical universities (Eindhoven, Twente and Delft) can draw upon DataverseNL through 4TU.ResearchData. The objective of this open source application is to make research data accessible to others. You can store data in the DataverseNL environment and grant multiple individuals controlled access to them.

Once you have requested the project environment or dataverse from 4TU.ResearchData, you are free to arrange this environment as you wish.

Why DataverseNL?

- Organization of data files in dataverses and datasets
- Addition of metadata and documentation
- Version management
- Management of access rights
- Easy collaboration with fellow researchers or project partners, even beyond your own university or research institute
- Centralized professional storage and backup
- To fulfil funding agency requirements for data management and sharing during your research.

See our quick guide for getting started or contact us when you need additional support.

Costs

Members of the three technical universities (Eindhoven, Twente and Delft) can store up to 100 GB of data free of charge. Storage of more than 100 GB of data costs € 150 per additional 50 GB per year.

A = Accessible. Data repositories

Checklist: is it the right repository for your data?

The checklist that follows addresses the five key questions posed in this guide:

1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

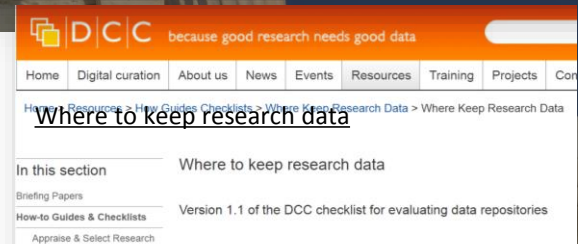
3. Will the data be safe in legal terms?

For this criterion we first consider the basic legal terms and conditions to check. Here only one capability level is given, as a repository either will or will not comply. Then we consider licensing, disclosure risk and access control, when repository may offer different levels of capability you can match to your needs. Note that Re3data provides relevant details for the repositories it lists, under 'terms'.

Legal terms and conditions
Level 1
Personal data or data which may identify individuals when linked to other data should not be stored outside the European Economic Area, unless in a legal jurisdiction that ensures personal data is adequately protected ☐
By agreeing to the terms and conditions the depositor will not be breaching other Data Protection principles, or the terms of any confidentiality agreement with data subjects or owners (e.g. consent form, consortium agreement) ☐

Findable, accessible and interoperable

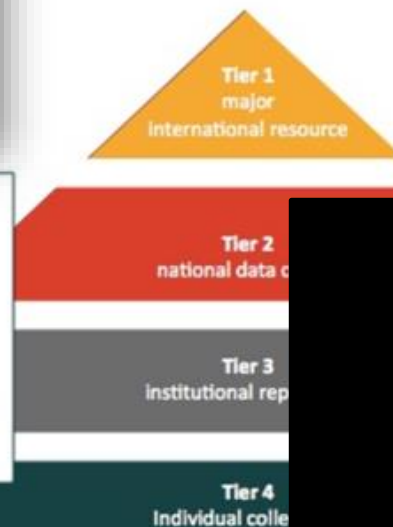
Level 1	Level 2	Level 3
Metadata publishing: Data collections are catalogued in a repository according to funder expectations so that they are discoverable by title, creator, and date of deposition ☐	Repository publishes other pertinent information as metadata fields to enhance cross-disciplinary discovery ☐	Metadata is catalogued to enhance reuse according to sector-leading standards, or to fulfil domain-specific purposes ☐
Stable identifiers: Enables a DOI or other open standard identifier to be assigned to a landing page for each ingested dataset/ collection ☐	Supports assignment of related persistent IDs per dataset/ collection ☐	Supports assignment of multiple persistent IDs at different levels of granularity within dataset/ collection ☐
Discovery metadata: Provides Datacite mandatory metadata and exposes it according to open	Provides metadata elements to enable broader discovery (e.g. geo-spatial) to reflect best practice	Exposes discovery metadata as Linked Open Data to optimise automatic discovery ☐



Value of data increases up the tiers: from individual to community to social value.

Each higher tier brings greater responsibility and demands for access.

As infrastructure increases so must the attention given to standards, sustainability and provenance.



«PRIORITÀ»

1. DISCIPLINARY
2. ZENODO
3. INSTITUTIONAL

(A MENO CHE SIA DATAVERSE, CHE VIENE USATO DURANTE IL PROGETTO)

A = Accessible. Cercate un archivio?



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search

re3data.org

Filter

Subjects ▾
Content Types ▾
Countries ▾
AID systems ▾
API ▾
Certificates ▾
Data access ▾
Data access restrictions ▾
Database access ▾
Database access restrictions ▾
Database licenses ▾
Data licenses ▾
Data upload ▾
Data upload restrictions ▾
Enhanced publication ▾
Institution responsibility type ▾
Institution type ▾
Keywords ▾
Metadata standards ▾
PID systems ▾
Provider types ▾
Quality management ▾
Repository languages ▾
Software ▾
Syndications ▾
Repository types ▾
Versioning ▾

genetic

← PREVIOUS 1 2 3 4 Next →

Found 87 result(s)

GESDB

Genetic Epidemiology Simulation Database

Subject(s) Epidemiology

Content type(s) Raw data

Country Taiwan, P.R.

GESDB is a platform for sharing simulation data and discussion manuscripts. The forum provides a platform for Q&A for the sim

Drosophila Genetic Reference Panel 2

DGRP2

Subject(s) General Genomics

Content type(s) Plain text

Country United States

The Drosophila Genetic Reference Panel (DGRP) is a population of complex traits, and a community resource for whole genome ass

Online Mendelian Inheritance in Man

OMIM

Subject(s) Human Genetics



2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management

By offering detailed information on more than 2,000 research data repositories, re3data has become the most comprehensive source of reference for research data infrastructures globally. Through the development and advocacy of a framework for discipline...

[Read more](#)

Three new DOI Fabrica features to simplify account management

Last month we launched DOI Fabrica, the modernized version of the DataCite Metadata Store (MDS) web frontend. It is the one place for DataCite providers and their clients to create, find, connect and track every single DOI from their organization...

[Read more](#)

One step closer towards instant DOI search results

Art Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to created it, did not exist. And the answer to what this illustrates – this is simply a distorted...

[Read more](#)

<https://www.re3data.org/>

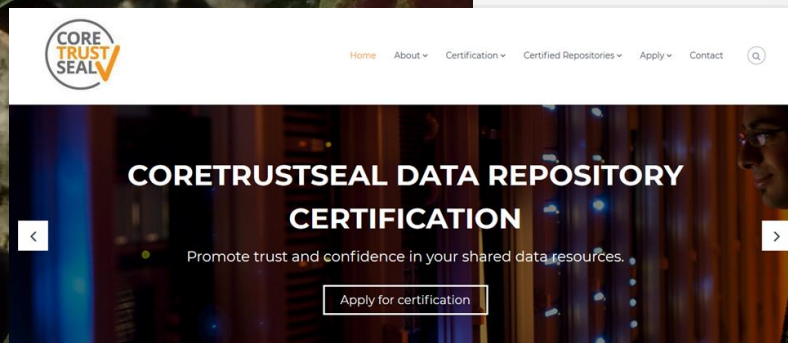
[trusted repositories]



CRITERIA FOR THE SELECTION OF TRUSTWORTHY REPOSITORIES

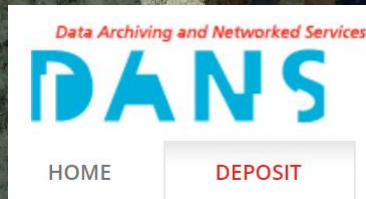
Trustworthy repositories should meet the following minimum criteria:

- ☐ **1. Provision of Persistent and Unique Identifiers (PIDs)**
 - a. Allow data discovery and identification
 - b. Enable searching, citing, and retrieval of data
 - c. Provide support for data versioning
- ☐ **2. Metadata**
 - a. Enable finding of data
 - b. Enable referencing to related relevant information, such as other data and publications
 - c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
 - d. Use metadata standards that are broadly accepted (by the scientific community)
 - e. Ensure that metadata are machine-retrievable
- ☐ **3. Data access and usage licences**
 - a. Enable access to data under well-specified conditions
 - b. Ensure data authenticity and integrity
 - c. Enable retrieval of data
 - d. Provide information about licensing and permissions (in ideally machine-readable form)
 - e. Ensure confidentiality and respect rights of data subjects and creators
- ☐ **4. Preservation**
 - a. Ensure persistence of metadata and data
 - b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)



European Commission Workshop
Alternative Open Access Publishing Models: Exploring New Territories in
Communication
Brussels, 12 October 2015

A = Accessible. Formati



Type	Preferred format(s)	Non-preferred format(s)
Text documents	PDF/A (.pdf)	- ODT (.odt) - MS Word (.doc, .docx) - RTF (.rtf) - PDF (.pdf)
Plain text	Unicode text (.txt)	Non-Unicode text (.txt)
Markup language	- XML (.xml) - HTML (.html) - Related files: .css, .xslt, .js, .es	SGML (.sgml)
Spreadsheets	- ODS (.ods) - CSV (.csv)	- MS Excel (.xls, .xlsx) - PDF/A (.pdf) - OOXML (.docx, .docm)
Databases	- SQL (.sql) - SIARD (.siard) - DB tables (.csv)	- MS Access (.mdb, .accdb) (v. 2000 or later) - dBase (.dbf) - HDF5 (.hdf5, .he5, .h5)
Statistical data	- SPSS Portable (.por) - SPSS (.sav) - STATA (.dta) - DDI (.xml) - data (.csv) + setup (.txt)	- SAS (.7dat; .sd2; .tpt) - R (* under examination)
Raster images	- JPEG (.jpg, .jpeg) - TIFF (.tif, .tiff) - PNG (.png) - JPEG 2000 (.jp2)	DICOM (.dcm) (by mutual agreement)

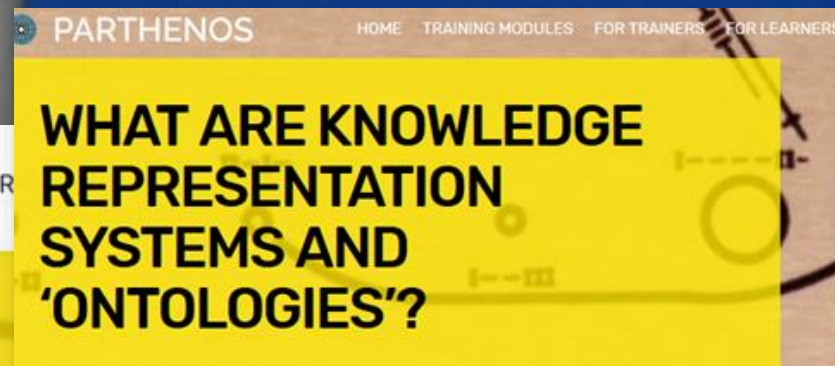
I = Interoperable. Standards



Even perfect metadata may not allow data to become interoperable if a different standard is used. A "standard" refers to a system that structures what types of information are captured in a collection. In our .mp3 library system, a standard is expressed in the header categories such as 'name,' 'time,' 'artist,' and 'album' are listed, with every entry having this filled in. Standards are used to ensure that metadata is as useful as possible for organising a collection, ensuring that common questions (how many songs are there on the album "Big B") can be easily and accurately answered.

How Many Standards Are There and Who Decides Which One To Use?

Different standards have arisen in different kinds of cultural heritage institution: the most common standards in museums are different from those in archives, and those common in libraries are different again.



In addition to metadata and standardised metadata schemas, research infrastructures can also use other forms of "knowledge representation system" to enhance the researcher's experience of the interoperable data they present. When we talk about 'Knowledge Representation Systems' in research infrastructures, we usually mean a specific category of hierarchical systems of terms known more commonly as an 'ontology'. Before the digital age, philosophers referred to an ontology as "the study of the kinds of things that exist". Ontologies are similar to taxonomies, another knowledge organisation framework you probably remember from early lessons in biology.



What is Metadata?

What are Standards?

What Are Knowledge Representation Systems and 'Ontologies'?

Sustainability

Methods and Tools

Networks

I = Interoperable. Standards

Standardization Survival Kit

A collection of research use case scenarios illustrating best practices in Digital Humanities and Heritage research



Browse scenarios



Add a new scenario



About the SSK

Increase efficiency, interoperability and sustainability by using standards

Incorporating standards in all the steps of your research process will make it last longer, easier to update, improve and share. Standards are non legally binding documents produced by an organisation ensuring :

<http://ssk.huma-num.fr/#/>

I = Interoperable

To speed up discovery and uncover new insights, research data should be easily combined with other datasets by humans as well as computer systems.

INTEROPERABLE

10 Establish well documented machine-actionable APIs

Well documented and machine-actionable APIs - a set of subroutine definitions, protocols, and tools for building application software - allow for automatic indexing, retrieval and combining of (meta)data from different data repositories.



Document APIs well and make it possible to deliver the schema of the (meta)data model. Consider showing examples of how to successfully mine data from different endpoints and combine them into new data sets usable for new research.

11 Use open well-defined vocabularies

The description of metadata elements should follow community guidelines that use open, well defined and well known vocabularies. Such vocabularies describe the exact meaning of the concepts and qualities that the data represent.



Use vocabularies relevant to your field, and enrich and structure your research output accordingly from the start of your research project.



Give examples of vocabularies the research community may use, based on research domain specifics.

12 Document metadata models

Clearly documenting metadata models helps developers to compare and make mappings between metadata.



Publish the metadata models in use in your research infrastructure. Document technical specifications and define classes (groups of things that have common properties) and properties (elements that express the attributes of a metadata section as well as the relationships between different parts of the metadata). For metadata mapping purposes, list the mandatory and recommended properties.

13 Prescribe and use interoperable data standards

Using a data standard backed up by a strong community, increases the possibility to share, reuse and combine data collections.



Check with the repository where you want to deposit your data what data standards they use. Structure your data collection in this format from the start of your research project.



Clearly specify which data standard your institution uses, pool a community around them and maintain them especially with a perspective on interoperability. Good examples are CMDI (language studies) and the SIKB0102 Standard (archaeology).

14 Establish processes to enhance data quality

To boost (meta)data quality and, therefore, interoperability, establish (automatic) processes that clean up, derive and enrich (meta)data.



Establish procedures to minimise the risk of mistakes in collecting data. E.g. choose a date from a calendar instead of filling it in by hand.



Invest in tools to help clean up (meta)data and to convert data into standardised and interoperable data formats. Combine efforts to develop workflows and software solutions for such automatic processes, e.g. by using machine learning tools.

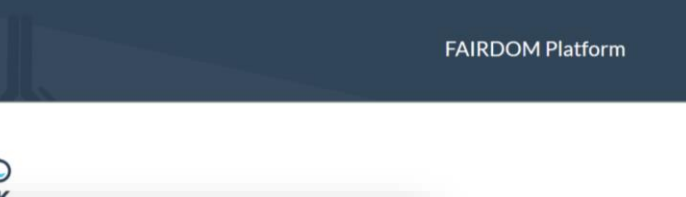
15 Prescribe and use future-proof file formats

All data files held in a data repository should be in an open, international, standardised file format to ensure long-term interoperability in terms of usability, accessibility and sustainability.

GUIDELINES
to FAIRify data
management
and make data
reusable

PARTHENOS

2019



FAIRDOM

About

FAIRDOM Platform

SFFK

...s. Rightfield is used by a 'Template Creator' to
 ...sed by Scientists to collect and annotate their
 ...ed. Rightfield embedded templates are used

...fic research datasets, models or simulations, processes and
 ...ation about the people and organisations involved. The
 ... based on the [ISA-Tools](#) format. When paired with our
 ...ction through to publication. Norwegian users benefit from
 ... simplifies upload and download of files.

...ata sharing within groups and consortia. In addition,

<https://fair-dom.org/platform/>

fic research datasets, models or simulations, processes and
sation about the people and organisations involved. The
) based on the *ISA-Tools* format. When paired with our
ction through to publication. Norwegian users benefit from
P) simplifies upload and download of files.

<https://fair-dom.org/platform/rightfield/>

Metadata	Values (examples)	Notes
Asset Title		The name of the data file
Uploader		The person submitting the asset to SEEK
Uploader SEEK ID		If you add your own SEEK ID, this will help us link this asset with your profile
Project	Project	The project that the asset belongs to
ASSAY		
Assay SEEK ID		If referring to an existing Assay, you can link to it via the Assay SEEK ID.
Assay Title		The title of an existing assay
Assay_type	ExperimentalAssayType	The assay_type describes the type of experiment you are performing
Technology_type	^	Describes the type of instruments and/or equipment used for the experiment
Description	amplification by-product_formation catabolic_response cell_growth_optimisation cell_size Comparative_genomic_hybridization comparative_genomics continuousEnzymatic	A brief, human readable description.
Experimentalist		The names of the people who carried out the experiments. These can either be SEEK members or external scientists
Date		The start date for the experiment if different from the upload date
SOP		Links to SOPs and protocols used to carry out the experiment. If they are already in SEEK, you can refer to them by their SEEK ID
Publication (optional)		If this data appears in a publication, you can link it directly, or via the assay or study. If it is already registered in SEEK, you can use the PubMed ID or DOI as a reference.
Experimental_conditions		
Item	ExperimentalConditions	The name of the experimental condition you are fixing in your experiment (e.g. temperature, concentration, pH etc). If there is more than 1, please list them in columns across the spreadsheet
Compound (if concentration)		The compound name is only required if the item is concentration.
Unit		The SI units of the experimental conditions measurements.
Start_value (optional)		This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
End_value (optional)		This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
Comments		Additional information that would be useful for people reading this data file

R = Reusable. Documentazione

DOCUMENTAZIONE (README FILE) PER
- EVITARE USO SCORRETTO/CATTIVE
INTERPRETAZIONI DEI VOSTRI DATI
- MANTENERE INTEGRITÀ



Project-level documentation



Project-level documentation explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc. In the accordion the questions which your project-level documentation should answer are stated in more detail:

detail:

- ⊕ 1. For what purpose was data created
- ⊕ 2. What does the dataset contain
- ⊕ 3. How was data collected
- ⊕ 4. Who collected the data and when
- ⊕ 5. How was the data processed
- ⊕ 6. What possible manipulations were done to the data
- ⊕ 7. What were the quality assurance procedures
- ⊕ 8. How can data be accessed

Data-level documentation

Data-level or object-level documentation provides information at the level of individual objects such as pictures or interview transcripts or variables in a database. You can embed data-level information in data files. For example, in interviews, it is best to write down the contextual and descriptive information about each interview at the beginning of each file. And for quantitative data variable and value names can be embedded within the data file itself.



⊖ Quantitative data

Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.



Data-level documentation for quantitative data

For quantitative data document the following:

- **Information about the data file**
Data type, file type and format, size, data processing scripts.
- **Information about the variables in the file**
The names, labels and descriptions of variables, their values, a description of derived variables if available, frequency, distribution, etc. The most critical

R = Reusable. Documentazione

≡ ✓ protocols.io

Make your science more reproducible
protocols.io is the #1 open access repository for science methods

Editing: Fixation of yeast cells for RNA-FISH

DESCRIPTION
GUIDELINES & WARNINGS
MATERIALS
STEPS

INOCULATION AND GROWTH

1. Around 10am, start a cell culture in a 50ml tube...
2. Grow for 8-10 hours in a shaker at 30 °C.
3. Measure OD in the evening and dilute into 250ml.

4. Transfer to 50ml falcon tubes.
5. Add 5ml of Formaldehyde, invert a few times, set...

4 Fixation
Transfer to 50ml falcon tubes.

new step paste from text paste from buffer

5 NEW SECTION (optional)

Fixation

Add 5ml of Formaldehyde, invert a few times, set...

Duration / Timer

Expected Result

External Link

Timer

hr 45 sec timer label

Notes

Optional: Transfer to gentle rocking over

recommended per Anne Dodson, Marc T

Reply

Delete

INSERT BLANK DELETE

RStudio

<https://www.rstudio.com/>

RStudio

Open source and enterprise-ready professional software for R

What is an Open Notebook?

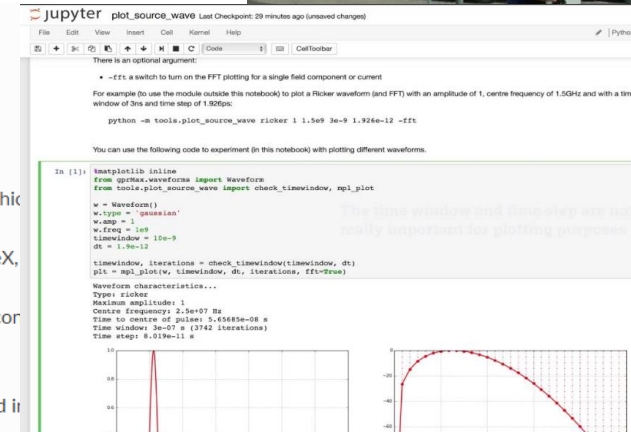
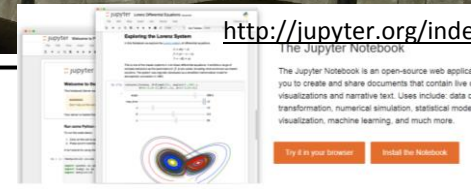
Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.

Notebook web application

The notebook web application enables users to:

- Edit code in the browser, with automatic syntax highlighting, and tab completion/introspection.
- Run code from the browser, with the results of computations attached to the code which generated them.
- See the results of computations with rich media representations, such as HTML, LaTeX, SVG, PDF, etc.
- Create and use interactive JavaScript widgets, which bind interactive user interface controls and visualizations to reactive kernel side computations.
- Author narrative text using the Markdown markup language.
- Include mathematical equations using LaTeX syntax in Markdown, which are rendered in the browser by MathJax.



...why not?

- protocols.io per i protocolli
- Open Lab Notebook tiene traccia di tutto il lavoro (difficile la prima volta, poi...)

R= Reusable. License

Copyright: protects the STRUCTURE, selection or arrangement of their contents" (Art. 3) NOT THE DATA

Sui generis database right: protects the «substantial effort» in OBTAINING data [NOT «CREATING»]... the right owner often is the institution

Database=a collection of independent works, data or other materials arranged in a systematic or methodical way (Art.1)

RICORDA: NESSUN
COPYRIGHT SUI DATI
(NON CREATIVI)

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 11 March 1996

on the legal protection of databases

COUNCIL OF THE EUROPEAN UNION,
in Community, and in particular Article 57 (2), 66 and 100a thereof,

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?

semplici dati e
informazioni

nessuna tutela

database
non creativo

solo diritto
sui generis

database
creativo

diritto sui generis
+ diritto d'autore

livello diritto
d'autore

livello diritto
sui generis

[DMP]

[webinar]



OpenAIRE 2019

SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

More Information about the 2019 webinar series.

data management plan | OA to research data | open science

RDA
RESEARCH DATA ALLIANCE
EUROPE

IT
NATIONAL
NODE

OpenAIRE

Aspetti legali nella gestione dei dati della ricerca

Thomas Margoni
University of Glasgow - CREATE
OpenAIRE project

Support

RESOURCES
Open Science Primers
Guides
Factsheets
Use cases

HELPDESK
FAQs
Ask a Question

TRAINING
Webinars
Workshops
Community of Practice

- POSSONO ESSERCI ALTRE FORME DI PROTEZIONE DEI DATI (ES. CONTRATTI)
- PER DATI CHE RICADONO SOTTO GDPR VA SEMPRE ESPLICITATA LA BASE LEGALE SULLA QUALE SI CONDUCE LA RICERCA

2020

OpenAIRE Legal Policy Webinars

Supporting researchers on the reuse of data: legal aspects to consider

29th April and May 4th, at 2 PM CEST

[non suoniamo tutti la stessa musica]

Obstacles to the trans-European archiving and sharing of research data

Making research data as openly available as possible is a widely recognised goal. For researchers working on an interdisciplinary project involving several countries, it can be difficult to fully comprehend in which ways open access to research data can be legally obtained. European national laws still diverge.

- **Diversity in copyright owner**

If protection applies, the right holder's consent is required for sharing the data. However, the designation of the copyright owner is also different in different jurisdictions. Although in many cases the maker of the work will be considered to be the author and therefore the right holder, only Dutch and UK law designate the employer as the right holder if the work was made in the course of employment.

[CESSDA guide](#)

A report from [Knowledge Exchange](#) (Knowledge Exchange, 2011) concludes that it will remain difficult to predict when particular files of research data are protected because of:

- **Diversity in copyright protection**

Even though most research data will fail to meet the criteria for copyright protection because they are not likely to be considered as "works" (they mainly concern facts), the lack of harmonisation of the criteria for copyright protection in Europe is tricky. E.g., whereas Germany, Denmark and the Netherlands have a relatively similar (higher) originality standard, the UK has a very low standard (skill, judgment and labour) making it

⊕ Czech Republic

⊕ Finland

⊕ Sweden

⊕ Switzerland

⊕ UK

[regole chiare dall'inizio]

Don't even
think of park-
ing here! :)



Safe to be open 2016

Safe to be open

Study on the protection
of research data and
recommendations
for access and usage

Edited by Lucie Guibault
and Andreas Wiebe

... FISSATE REGOLE CHIARE DA SUBITO

- CHI HA I DIRITTI SUI DATI (SE ESISTE)
- CHI DEVE PROVVEDERE A CONSERVARE
 - CHI PUÒ SFRUTTARE

R = Reusable. License

 | D | C | C because good research needs good data

How to License Research Data

This guide will help you decide how to apply a licence to your research data, and which licence would be most suitable. It should provide you with an awareness of why licensing data is important, the impact licences have on future research, and the potential pitfalls to avoid. It concentrates on the UK context, though some aspects apply internationally; it does not, however, provide legal advice. The guide should interest both the principal investigators and researchers responsible for the data, and those who provide access to them through a data centre, repository or archive.

<http://www.dcc.ac.uk/resources/how-guides/license-research-data>

Creative Commons



FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE V0.1

This information guide contains questions and responses to common concerns surrounding open science and the implications of licensing data under Creative Commons licences. It is intended to aid researchers, teachers, librarians, administrators and many others using and encountering Creative Commons licences in their work.

<https://doi.org/10.5281/zenodo.840651>

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the [CC0 Public Domain Dedication](#), which is first and foremost a waiver, but [can act as a licence](#) when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is not protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the [CC0 Public Domain Dedication](#), which is first and foremost a waiver, but [can act as a licence](#) when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is **not** protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

But I would like attribution when others use my dataset. In that case, shouldn't I use a CC BY licence?

We recommend that you avoid using a CC BY licence. Here's why:

While attribution is a genuine, recognisable concern, not only might using a CC BY licence be legally unenforceable when no underlying copyright or SGDR protects the work, but it may also communicate the wrong message to the world. A better solution is to use CC0 and [simply ask for credit](#) (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease. Such requests are consistent with scholarly norms for citing source materials.

Legally speaking, datasets that are **not** subject to copyright or related rights (and are thus in the public domain) cannot be the object of a copyright licence. Despite this, agreements based in contract law may be enforceable. Creative Commons licences, however, are copyright licences. Therefore, where the conditions for a copyright or related right are not triggered, copyright licences, such as the CC BY licence, [are unenforceable](#).

In some cases, however, rights may exist (like the *sui generis* database right previously mentioned), and permission for others to use your dataset may be legally required. These rights are meant to protect the maker's investment, rather than originality. As such, database rights do not include the moral right of attribution. So by using a CC BY licence, you signal to users that you restrict access to your dataset beyond the protections provided by the law. We are not saying that this cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

Commons Open

USARE UNA CC0

- NON SIGNIFICA DIVENTARE ACCADEMICAMENTE MALEDUCATI
- LA FONTE VA CITATA SEMPRE
- USATE LA CC0 E ASSOCIATE UNA FORMULAZIONE DELLA CITAZIONE CHE RICHIEDETE (DA COIPARE/INCOLLARE)

cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

I'm uncomfortable with others using my research for commercial purposes. Should I use a non-commercial licence for my dataset?

We recommend you avoid using a non-commercial licence. Here's why:

For legal purposes, drawing a line between what is and is not 'commercial' can be tricky; it's not as black and white as you might think. For example, if you release a dataset under a non-commercial licence, it would clearly prohibit an organisation

I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

We recommend you avoid using a 'No Derivatives' licence. Here's why:

Similar to how a non-commercial licence might restrict meaningful reuse of your dataset, a ND licence can have the same effect: it may prevent someone from recombining and reusing your data for new research. For data to be truly Open Access, it must permit these important types of reuse.

It sounds like you're really pushing for the use of CC0 for open science datasets.

Exactly. Data is only open if anyone is free to use, reuse, and distribute it. This means it must be made available for both commercial and non-commercial purposes under non-discriminatory conditions that allow for it to be modified.

When data is made available for all reuse, others can create new knowledge from combining it. This leads to the enrichment of open datasets and further dissemination of knowledge. Accordingly, CC0 is ideal for open science as it both protects and promotes the unrestricted circulation of data.

And remember, it's bad science not to cite the source of data you use. To help others cite your data [include a citation](#) that users can copy and paste to give you credit for your hard work.

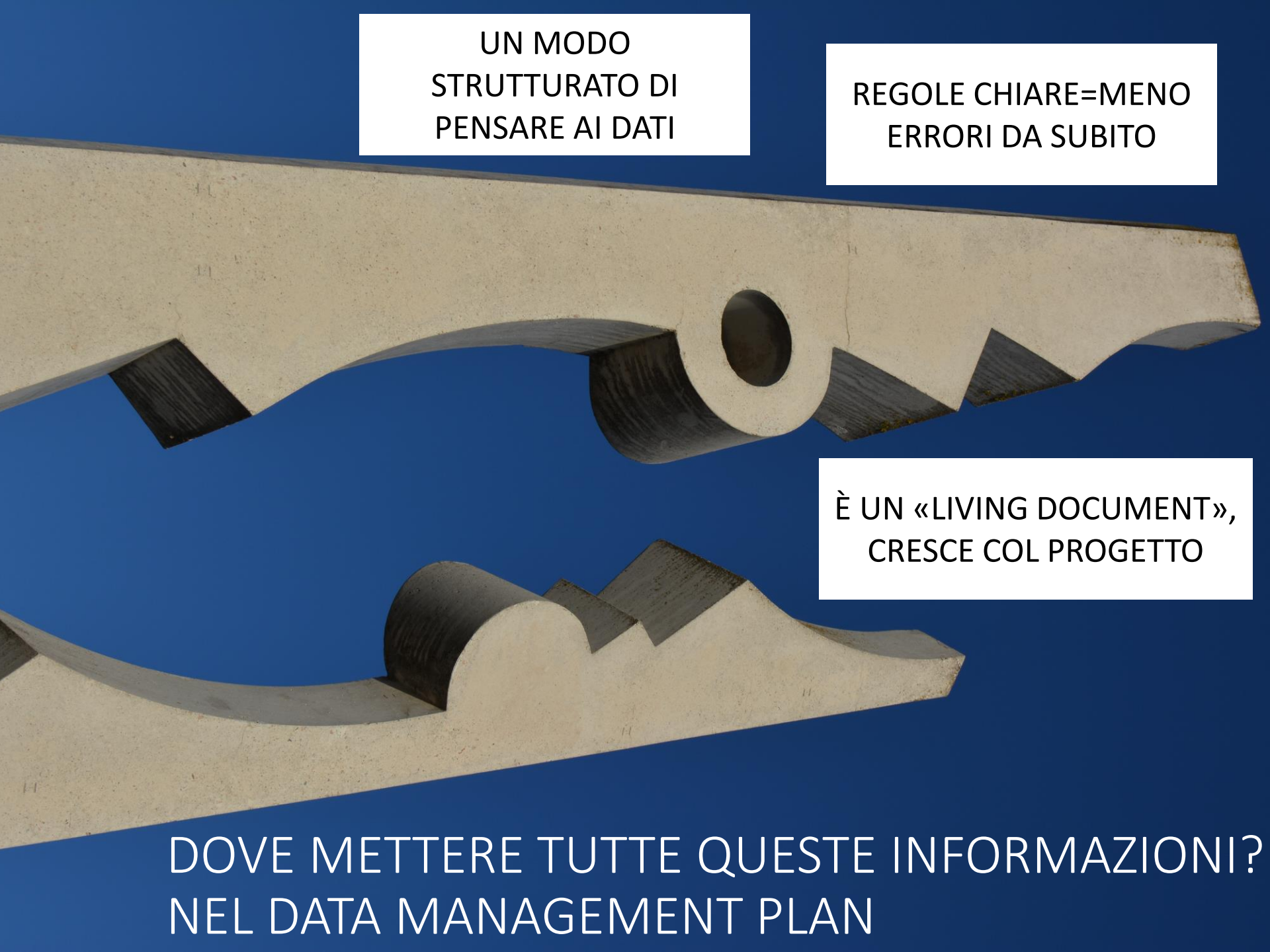


Ci sono solo TRE modi
per rilasciare open data
senza creare inutili
complicazioni:
CC Zero, CC Zero e
CC Zero



...e se proprio non potete
farne a meno, al massimo
usate una CC BY 4.0





UN MODO
STRUTTURATO DI
PENSARE AI DATI

REGOLE CHIARE=MENO
ERRORI DA SUBITO

È UN «LIVING DOCUMENT»,
CRESCe COL PROGETTO

DOVE METTERE TUTTE QUESTE INFORMAZIONI?
NEL DATA MANAGEMENT PLAN

Trucchi e suggerimenti

**SINTETICO E
SPECIFICO**

Top tip - keep it short and specific!

This very short extract from a presentation by Peter Dukes, Medical Research Council, is really useful advice on writing a DMP from the funding body perspective. The advice applies to all disciplines. The quality of the video isn't great, definitely is!



FOSTER toolkit

NON COPIATE

OGNI DATASET È
UNICO, OGNI
INFRASTRUTTURA È
UNICA, OGNI RICERCA
HA LA SUA
IMPOSTAZIONE

CIÒ CHE DICHIARATE NEL
DMP POI VA FATTO
VERAMENTE... QUINDI

- A) NON FATE GLI
SPLENDIDI
 - B) NON IMPEGNATEVI A
FARE COSE CHE
SAPETE IMPOSSIBILI
- Es. DATI PSEUDONIMIZED, non
ANONIMIZED

ESSERE GENERICI NON SERVE
A NULLA

[we expect a huge size of
data; data will be available]

- USATE TABELLE,
ELENCHI PUNTATI
- SIATE SCHEMATICI E
NON DILUNGATEVI

Vantaggi di un DMP

CESSDA Guide



⊖ Benefit 3. Clarifies needed budget

Data management is not free. You do not want to find yourself running out of funding before the end of the project because you have ignored or underestimated the cost of structured, detailed, and safe data management. Therefore, an important aspect of a DMP is its use in calculating how much money will be required for managing your research data during your research project.

A DMP can be useful in the process of applying for funding. Grant applications should not only include time and resources for collecting, analysing, and publishing on data in their budget, time and resources for careful documentation as well as server space, backup solutions, and documentation software need to be included as well. A DMP is also useful once funding is granted to plan and manage your expenses. Many research funders require a DMP as part of the application and decision-making process. The arguments for making data available are several, the most popular being that the data produced by public funds should be used to the greatest extent possible and available to the public. Unless there are legal, ethical or commercial barriers, data should also be openly available so that research results can be verified, replicated and reused.

Examples of Data Management cost assessments are given by the [University of Utrecht](#) (n.d.) and the Dutch Landelijk Coördinatiepunt Research Data Management ([LCRDM](#), n.d.) inspired by the '[Data management costing tool](#)' by UK Data Service, 2013.

È FONDAMENTALE PER
STIMARE I COSTI DI GESTIONE
- STIMATE LA DATA
STEWARDSHIP (IN-KIND?)
-POSSIBILI COSTI DI STORAGE
COSTI SONO RIMBORSABILI IN
H2020 (6.2.D.3 AMGA)

Costs – data curation



Costs related to the implementation of the Open Research Data pilot (e.g. costs for providing open access, related research data management costs, data curation and data storage costs) may be eligible (see [Article 6.2.D.3](#)).

- ELIGIBLE COSTS:
- DATA CURATION
 - DATA STORAGE
 - DATA MANAGEMENT

DMP Core Requirements

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS



When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

- ☐ **1. Data description and collection or re-use of existing data**
 - a. How will new data be collected or produced and/or how will existing data be re-used?
 - b. What data (for example the kinds, formats, and volumes) will be collected or produced?
- ☐ **2. Documentation and data quality**
 - a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
 - b. What data quality control measures will be used?
- ☐ **3. Storage and backup during the research process**
 - a. How will data and metadata be stored and backed up during the research process?
 - b. How will data security and protection of sensitive data be taken care of during the research?
- ☐ **4. Legal and ethical requirements, codes of conduct**
 - a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
 - b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
 - c. How will possible ethical issues be taken into account, and codes of conduct followed?

SEZIONI MINIME IN UN DMP

- ☐ **5. Data sharing and long-term preservation**
 - a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
 - b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
 - c. What methods or software tools will be needed to access and use the data?
 - d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?
- ☐ **6. Data management responsibilities and resources**
 - a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
 - b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS



DMP Core Requirements

Translating the Core Requirements into a DMP template

The following example of a data management plan template is based on the core requirements for DMPs.⁶ These core requirements should be considered as a minimum standard, leaving the flexibility to formulate additional guidelines according to the needs of specific domains or to national or local legislation.

The template presented below refers to the 15 questions covering six core requirements for good data management. Additional guidance and explanations are provided to help researchers fill out such a template and to assure that all relevant aspects of research data management are covered. The below table is an example of how the core requirements can be transformed into a DMP template. It will be up to the individual organisations and disciplines to develop templates that fit their needs.

GENERAL INFORMATION

- Administrative information**
- Provide information such as name of applicant, project number, funding programme, version of DMP.

1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA

- 1 a**
- How will new data be collected or produced and/or how will existing data be re-used?**
- Explain which methodologies or software will be used if new data are collected or produced.
 - State any constraints on re-use of existing data if there are any.
 - Explain how data provenance will be documented.
 - Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

UTILI COME
SPUNTO PER LE
PRIME VOLTE

2 DOCUMENTATION AND DATA QUALITY

2 a

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data.
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
- Use community metadata standards where these are in place.
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.
- Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.

2 b

What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.



DMP questions



Overview

Title of the project

Date of this plan

Description of the project

- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data

- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers

- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)

- What are their contact details and their roles in the project?

Funder (if applicable)

- If funding is granted, what is the reference number of the funding granted?

Data producer

- Which organisation has the administrative responsibility for the data?

Project data contact

- Who can be contacted about the project after it has finished?

Data owner(s)

- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles

- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs

- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses for (preparing the data for) archiving covered?

UTILI COME
SPUNTO PER LE
PRIME VOLTE

Organising and documenting your data

Data collection

- How will the data be collected?
- Is specific software or hardware or staff required?
- Who will be responsible for the data collection?
- During which period will the data be collected?
- Where will the data be collected?

Data organisation

- How will you organise your data?
- Will the data be organised in simple files or more complex databases?
- How will the data quality during the project be ensured?
- If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

Data type and size

- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- After the project: What is the total amount of data collected (in MB/GB)?

File format

- In what format will your data be?
- Does the format change from the original to the processed/final data?
- Will your (final) data be available in an open format?

Folder structure and names

- How will you structure and name your folders?

File structure and names

- How will you structure and name your files?

Documentation

- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

Metadata

- What metadata will be provided with the collected/ generated/ reused data?
- How will metadata for each object be created?
- Is there any program that can be used to document the data?
- Can metadata be added directly into the files or will the metadata be produced in another program or document?

Metadata standard (if applicable)

- What metadata standard(s) will you use?

ORGANISE &
DOCUMENT

DMP questions

Processing your data

Versioning

- What is your strategy concerning versioning your data files (and scripts) during the project?
- Will you create and/or follow a convention for versioning your data?
- Who will be responsible for securing that a "Masterfile" will be maintained, documented and versioned according to the project guidelines?
- How can different versions of a data file be distinguished?

Interoperability

- Will you make use of established software and hardware? If not, how does the software and hardware you use relate to other research?

If applicable:

- Will you make use of established terminologies/ontologies (i.e. structured controlled vocabularies) in the project? If not, how do your terminologies relate to established ones?
- Which coding is used (if any)? Will you build on established coding schemes? If not, how does your coding relate to other research?

Storing your data

Storage

- How and where will the data be stored during the project?
- For how long will the data be stored?

Backup

- How, where and at what intervals will the data be backed-up?
- How will data be recovered in the case of a data loss incident?

Security

- How will sensitive data be protected? (if applicable)
- How will data access be managed?

Protecting your data

Ethical review (if applicable)

- Does your project require approval by a local ethics committee?

Informed consent (if applicable)

- Do you require informed consent for your project?
- If so, how will permission be obtained?
- How are consent files organised and stored?

(sensitive) Personal data /confidential information (if applicable)

- How will access to (sensitive) personal data during the project be controlled?
- How will collaborators be granted access to the data in a secure way?
- If the research project is going to have data that includes confidential information or information that requires informed consent, is there a requirement to notify a privacy officer?
- Is there any confidential information within the material that requires special treatment and/or limits the access to it during/after the project?
- How will the material be protected during/after the project?
- How will permissions and restrictions be enforced?

Intellectual property rights (IPR)/Copyrights

- Are there IPR or copyright issues to consider?
- Will permission be needed to collect/reuse the data?
- Will these rights be transferred to another organisation for data distribution and archiving?

Agreements (if applicable)

- What are the agreements with other stakeholders?

Restrictions (if applicable)

- Are there any other restrictions that need to be considered?

UTILI COME
SPUNTO PER LE
PRIME VOLTE

Basic Information.

- State the purpose of the data collection/generation.
- Explain the relation to the objectives of the project
- Consider what data will be collected or created as part of the study (RAW data).
- Consider what data will be produced by processing the RAW data (Secondary, processed data).
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- Specify the types and formats you plan to use for the data generated/collected (raw, processed, published).
- Consider what data will be published as the result of your study (Published data).

Volume and Life Cycle of the Data.

If you are using FAIRDOM, we will look after data that will be retained and potentially exchanged by your projects. It will help with local storage for temporarily-held local data prior to processing.

For RAW data, please consider the following:

- How much RAW data you think will be produced (Estimates, per month, year, full project duration)?
- Will all of the RAW data be kept for the duration of the study or will the RAW data be deleted once it is processed?
- For large scale RAW data (images, sequence) have you planned the local storage capacity necessary for processing?
- Do you require help to organise a suitable local management system for RAW data?
- Do you have policies that govern the management and usage of RAW data?
- How long will RAW data be kept? Will there be a long-term archive?

For Secondary and Published data, please consider the following:

- What data processing is foreseen in the project?
- How much processed data will be produced, and stored (can you make estimates per month, year, full project)?
- How much of this data will be published? (Estimates per month, year, full project)?
- Does your institution, or the project funders, have policies governing the access and usage of processed data?

Additional for personally sensitive data (e.g medical data)

- When looking at the data flow through the project, define what data is:
 - aggregated (typically safe to share, if names cannot be recovered)
 - anonymized (name cannot be recovered from the data)
 - pseudonymized (name can be recovered by some)
 - non-anonymized (name linked to data)
- Determine which organisational boundaries have to be traversed by which data.
- Make sure with your "local" data protection officer and ethics commission that the data can be shared with your partners along the flow described with the anonymisation levels as described. Why local? Some laws change across surprising boundaries. E.g. in Germany Universities and other public organisations are subject to another data protection law than enterprises. Why seek advice? In some cases you may be required to be able to recover the name-data-relation, e.g. to enable study participants to "leave" a study.

Data Management Checklist

<https://fair-dom.org/knowledgehub/data-management-checklist/>

UTILI COME
SPUNTO PER LE
PRIME VOLTE

PERSONALIZZABILE

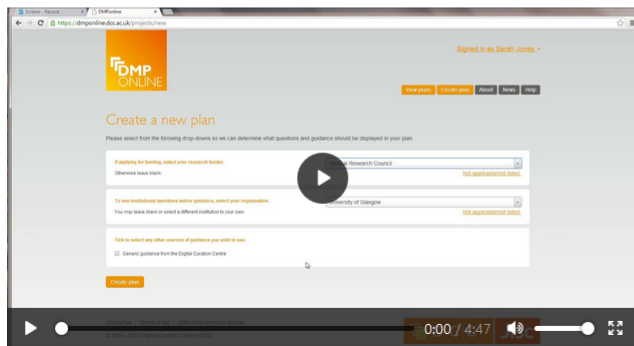
DMP online

[Home](#)[About](#)[Future plans](#)[Help](#)[Change language](#)

Welcome.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

Screenshot on how to use DMPonline



<https://dmponline.dcc.ac.uk/>

Sign in

Veteran tapes

[Project Details](#)[Plan overview](#)[Write Plan](#)[Share](#)[Download](#)

expand all | collapse all

13/13 answered

Data Collection (2 / 2)

What data will you collect or create?

B *I*

The "Veteran tape " project will collect and generate different types of datasets:

Type of data	Volume	Format	Storage format
Video recordings	600 x 1Gb	.mkv	.mkv
Transcriptions	600 x 1500Kb	MS Word	.txt
Structured interview text	1 x 500Kb	MS word	.txt

For the video recordings the selected format is .mkv; the same .mkv format will be used for the long-term preservation .

Transcriptions will be written in MS Word and then stored as .txt files.

We checked the format compatibility against EASY File format
<https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats>

As the total volume of data is greater than 50Gb, DANS requires a fee for the storage. We are currently in touch with EASY to determine the costs of archiving.

Save

[Guidance](#)[Comments \(1\)](#)[DCC](#)[DCC guidance](#)

Guidance

Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

☒ mock project for testing, practice, or educational purposes

* Select the primary research organisation

- or - ☐ No research organisation associated with this plan or my research organisation is not listed

* Select the primary funding organisation

☒ European Commission (Horizon 2020)☐ European Research Council (ERC)

- or - ☐ No funder associated with this plan or my funder is not listed

PREDISPOSTO PER
LE 3 VERSIONI

ALLA FINE,
SCARICA IL PDF

CONDIVISIBILE
PER
SCRIVERLO
INSIEME

IDENTIFICATIVI

GUIDA

Prova

Project Details

Plan overview

Initial DMP

Detailed DMP

Final review DM

Share

Download

* Project title

☒ mock project for testing, practice, or educational purposes

Funder

Grant number

Project abstract

Briefly summarise your research project to help others understand the purposes for which the data are being collected or created.

[See the full list](#)

Save

Guidance

write your plan, DMPonline can show from a variety of organisations.

5 organisations to see their

Curation Centre

from additional organisations

ID

Principal Investigator

Name

ORCID iD

Email

Phone

Data Contact Person

☒ Same as Principal Investigator

Save

Project Details

Plan overview

Initial DMP

Detailed DMP

Final review DMP

Share

Download

Set plan visibility

Public or organisational visibility is intended for finished plans. You must answer at least 50% of the questions to enable these options. Note: test plans are set to private visibility by default.

- ☐ Private: visible to me, specified collaborators and administrators at my organisation
- ☐ Organisation: anyone at my organisation can view
- ☐ Public: anyone can view

PRIVATO O PUBBLICO?

Manage collaborators

Invite specific people to read, edit, or administer your plan. Invitees will receive an email notification that they have access to this plan.

Email address	Permissions
elena.giglia@unito.it	Owner

Invite collaborators

* Email

* Permissions

- ☐ Co-owner
- ☐ Editor
- ☐ Read only

SI POSSONO INVITARE
COLLABORATORI CON LIVELLI DI
PERMESSO DIVERSI

Horizon 2020 DMP

This plan is based on the "Horizon 2020 DMP" template provided by European Commission (Horizon 2020).

The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot).

Projects participating in the pilot must submit a first version of the DMP (as a deliverable) within the first 6 months of the project. The DMP needs to be updated over the course of the project whenever significant changes arise.

Further details are provided in the [Guidelines on FAIR Data Management in Horizon 2020](#) (v.3, 26 July 2016).

Template version 1, published on 16 May 2019

Initial DMP (6 sections, 9 questions) +

Detailed DMP (9 sections, 31 questions) +

Final review DMP (9 sections, 31 questions) +

Initial DMP (6 sections, 9 questions)

Instructions

Once a project has had its funding approved and has started, you must submit a first version of your DMP (as a deliverable) within the first 6 months of the project. The Commission provides a DMP template, the use of which is recommended but voluntary. That template has been replicated by the UK Digital Curation Centre for use in DMPOnline.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a *living document* in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates.

Write plan

1. Data summary

- Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

2. FAIR data

- 2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you use Digital Object Identifiers?

- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

3. Allocation of resources

- Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project

expand all | collapse all

0/9 answered

1. Data summary (0 / 1)

2. FAIR data (0 / 4)

3. Allocation of resources (0 / 1)

4. Data security (0 / 1)

5. Ethical aspects (0 / 1)

6. Other (0 / 1)

...E NON «A
HUGE AMOUNT
OF DATA»

COSTI

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).
- Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?
- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

expand all | collapse all

0/9 answered

1. Data summary (0 / 1)

GUIDA RAPIDA

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

= README
FILE

B I [bulleted list icon] [numbered list icon] [link icon] [table icon]

SCRIVETE
DIRETTAMENTE
(USATE TABELLE E
PUNTI ELENCO)

Save

IN OGNI PASSAGGIO POTETE
SALVARE E CONTINUARE DOPO

ALTRE GUIDE
SPECIFICHE

USATE FORMATI
STANDARD (CON
ELENCO)

Guidance

DCC

Data volume

Data format

Data description

Data format

TABELLA PER DATI DI
FORMATO DIVERSO

- Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tiff).
- Explain why you have chosen certain formats. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community.
- Using standardised, interchangeable or open formats ensures the long-term usability of data; these are recommended for sharing and archiving.
- See UK Data Service guidance on [recommended formats](#) or DataONE Best Practices for [file formats](#).

2. FAIR data (0 / 4)

In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Guidance

Comments

EC

DCC

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

APRE LA
DIRECTORY DEI
METADATA
STANDARDS

B I    

Save

2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

B I    

Save

RICORDA GENTILEMENTE CHE
TUTTI I DATI DEVONO ESSERE
APERTI (SE POSSIBILE)

Guidance

Comments

EC

DCC

Participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather, the ORD pilot follows the principle "**as open as possible, as closed as necessary**" and focuses on encouraging sound data management as an essential part of research best practice.

The Commission recognises that there are good reasons to keep some or even all research data generated in a project closed. Where data need to be shared under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

APRE REGISTRY OF
REPOSITORIES

The [Registry of Research Data Repositories](#) provides a useful listing of repositories that you can search to find a place of deposit.

2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

B I [List Icon] [List Icon] [Link Icon] [Table Icon]

Save

Guidance

Comments

EC

Interoperability means allowing exchange and re-use between institutions, organisations, countries (i.e. adhering to standards for as much as possible compliant with (open) software applications, a particular facilitating re-combining different datasets from different

Guidance

Comments

Add comments to share with collaborators

B I [List Icon] [List Icon] [Link Icon] [Table Icon]

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

B I [List Icon] [List Icon] [Link Icon] [Table Icon]

Save

Guidance

Comments

EC

DCC

The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Reasons for embargoes may include time to publish or seek patents. If an embargo is sought, specify why and for how long,

29.3 Open access to research data

OPTION 1a for actions participating in the open Research Data Pilot: Regarding the digital research data generated in the action ('data'), the beneficiaries must:

- deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
 - the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*

RICORDATE LORO CHE HANNO FIRMATO UN GRANT AGREEMENT...

3. Allocation of resources (0 / 1)

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

Guidance

Comments

EC

DCC

Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Costs are eligible for reimbursement during the duration of the project under the conditions defined in the H2020 Grant Agreement, in particular [Article 6](#) and [Article 6.2.D.3](#), but also other articles relevant for the cost category chosen.

RICORDARE SEMPRE
CHE EVENTUALI
COSTI SONO
RIMBORSABILI

Save

4. Data security (0 / 1)

Address data recovery as well as secure storage and transfer of sensitive data

Guidance

Comments

EC

DCC

Also consider whether the data is safely stored in certified repositories for long term preservation and curation.

...QUI SAREBBE UTILE POLICY
DI ATENEO; SAREBBE SEZIONE
PRECOMPILATA

Save

5. Ethical aspects (0 / 1)

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Guidance

Comments

EC

DCC

Consider whether there are any ethical or legal issues than can have an impact on data sharing. For example, is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

BASE LEGALE PER DATI
PERSONALI; CONSENSI
INFORMATI VANNO IN
QUESTA SEZIONE

Save

6. Other (0 / 1)

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Guidance

Comments

DCC

Related policies

- Consider whether there are any existing procedures that you can base your approach on. If your group/department has local guidelines that you work to, point to them here.
- List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security.

QUI SAREBBE
URGENTE UNA
POLICY DI ATENEO

Save

Project Details

Plan overview

Initial DMP

Detailed DMP

Final review DMP

Share

Download

Download settings

Select phase to download Initial DMP ▼

Optional Plan Components

- ☐ project details coversheet
- ☒ question text and section headings
- ☒ unanswered questions

Format

pdf ▼

PDF formatting

Font

Face

Arial, Helvetica, Sans-Serif ▼

Size (pt)

10 ▼

Top

25 ▼

Bottom

20 ▼

Left

12 ▼

Right

12 ▼

Margin (mm)

Download Plan

AL TERMINE SI SCARICA NEL
FORMATO PREFERITO
(PDF, CSV, HTML, TXT) CON UN
MINIMO DI FORMATTAZIONE

2. Data Management Plan – general definition

Data Management Plans (DMPs) are a *key element* of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt submit a DMP on a voluntary basis.

IN HORIZON 2020
DMP ENTRO M6
COME DELIVERABLE

VA FATTO SEMPRE
PERCHÉ RIGUARDA LA
CORRETTA GESTIONE
DEI DATI, NON
APERTURA/CHIUSURA

4. Research data management plans during the project life cycle

Once a project has had its funding approved and has started, you **must submit a first version of your DMP** (as a deliverable) within the first 6 months of the project. The Commission provides a DMP template in annex, the use of which is recommended but voluntary.

The DMP needs to be **updated** over the course of the project whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP should be updated as a minimum in time with the periodic evaluation/assessment of the project. If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest. Furthermore, the consortium can define a timetable for review in the DMP itself.

Periodic reporting

For general information on periodic reporting please check the following sections of the online manual

- How to fill in [reporting tables for publications, deliverables](#)
- [Process for continuous reporting](#) in the grant management system.



H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

Guide 2016

Version 3.0
26 July 2016

Open data – H2020

How?

Open access to digital research data involves 3 steps:

Procedure for open access (research data):

- Step 1 —** Deposit the digital research data, preferably in a research data repository.
- Step 2 —** Provide **open access** by taking measures to enable users to access, mine, exploit, reproduce and disseminate the data free of charge (*e.g. for databases: by attaching an appropriate creative commons licence (CC-BY or CC0 tool) to the data; if the access/use is not subject to any rights: by indicating that no licence is needed*).

Open access must not be given immediately; for data needed to validate the results presented in scientific publications, as soon as possible; for other data, beneficiaries are free to specify embargo periods for their data in the data management plan (as appropriate in their scientific area).

- Step 3 —** Provide **information**, via the repository, about **tools and instruments** for validating the results.

Where possible, the beneficiaries should provide those tools and instruments (*e.g. specialised software or software code, algorithms, analysis protocols, etc.*).



- 1) DEPOSITARE IN UN DATA REPOSITORY
- 2) RENDERE I DATI OPEN ACCESS CON LA LICENZA PIÙ APERTA POSSIBILE
- 3) POSSIBILE EMBARGO
- 4) FORNIRE ANCHE TUTTE LE INFORMAZIONI UTILI A VALIDARE I DATI

A wooden sign made of several vertical planks is placed on a brick patio. The sign features a motivational quote in bold, black, sans-serif capital letters. The quote is arranged in six lines. The sign is positioned in front of a wooden bench made of thick, weathered planks. The background consists of a brick wall and a brick patio. The ground in the foreground is a dark, textured surface.

**"IF YOU ARE NOT
DOING WHAT
YOU LOVE,
YOU ARE
WASTING
YOUR TIME."**

... ora tocca a voi!