OPEN SCIENCE DALLA A ALLA Z
6-DATI FAIR E DATA MANAGEMENT PLAN

Elena Giglia

Elena Giglia
Perugia, 27 febbraio 2020
elena.giglia@unito.it
@egiglia

Impareremo in questo modulo

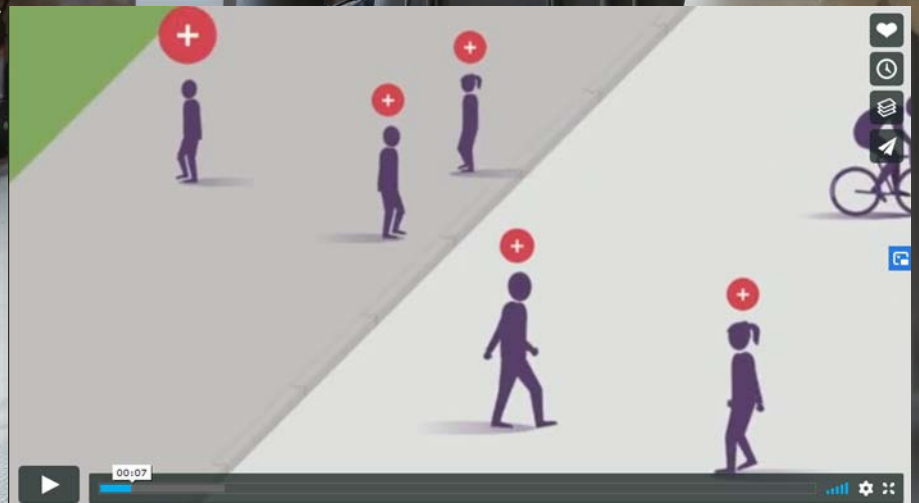
1. Cosa significa FAIR nella ricerca quotidiana
2. come scrivere un DMP (Data Management Plan)

MESSAGGI CHIAVE

- FAIR è il futuro (se no si resta tagliati fuori
- È più facile di quanto sembri, basta provarci

FAIR train – GoFAIR video

FARM DATA TRAIN



PERSONAL HEALTH TRAIN

…FAIR SIGNIFICA
[anche e soprattutto per le macchine]

FINDABLE

- IDENTIFICATIVI
- METADATI

ACCESSIBLE

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

INTEROPERABLE

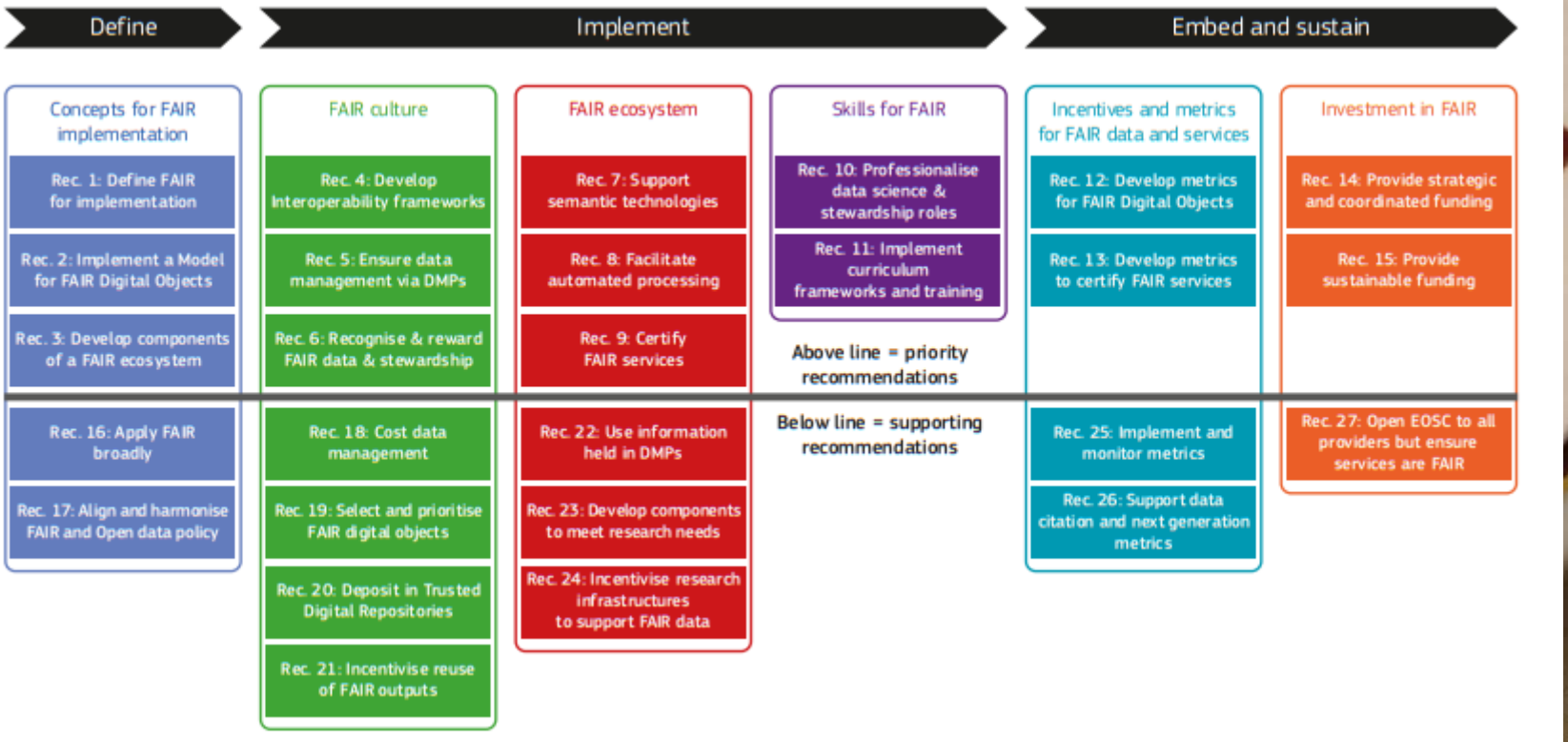
- STANDARDS
- ONTOLOGIE

REUSABLE

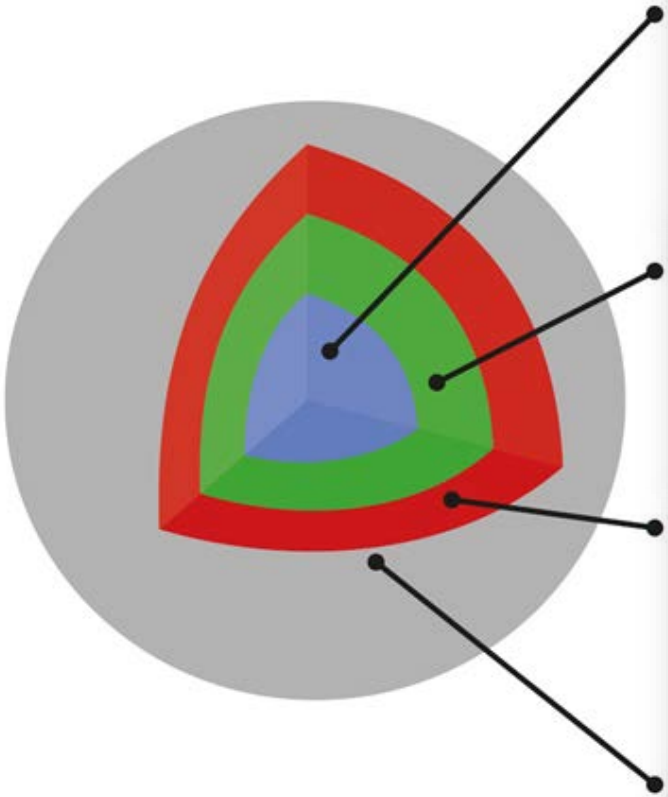
- LICENZE D'USO
- DOCUMENTAZIONE
- LEGGIBILI DALLE MACCHINE

Nov. 20, 2018


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


Define
Implement
Embed and sustain
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
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
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 use 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.

ECD, 2015

FAIR: technology VS domain

Technology

FAIR Principles

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

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;

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;

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;

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;

https://www.go-fair.org/wp-content/uploads/2018/11/26102018_Country_meeting_GFISCO_staff_presentation.pdf

… 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

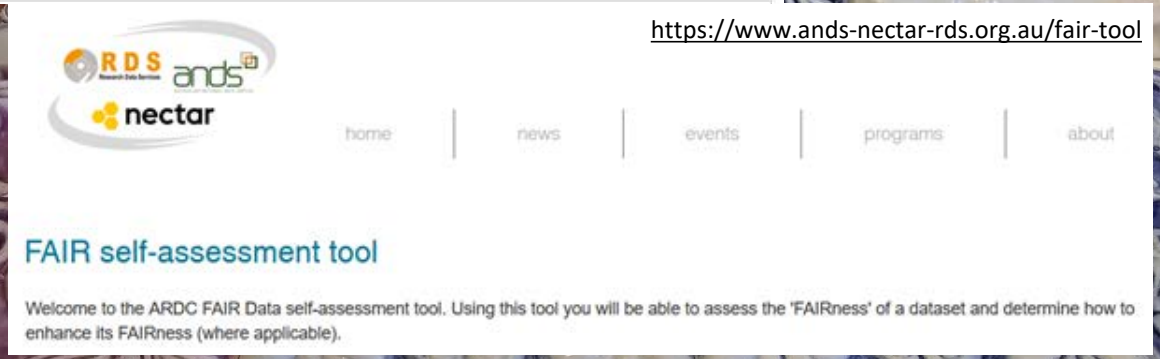
Accessible

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

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



FAIR maturity evaluator

- OBJECTIVE
- MACHINE-READABLE … AS FAIR DATA ARE

Top 10 FA

Top 10 FAIR Data & Software Things

February 1, 2019

Sprinters:

Reid Otsuji, Stephanie Labou, Ryan Johnson, Guilherme Castelao, Bia Villas Boas, Anna-

Geoscience

Sprinters:

John Brown, Janice Chan, Niamh Quigley (Curtin Univer

Audience:

Researchers

Things

Findable

Thing 1: Data sharing and discovery

Thing 6: Vocabularies for data description

Thing 7: Identifiers and linked data

Thing 10: Spatial data

Accessible

Thing 2: Long-lived data: curation & preservation

Thing 3: Data citation for access & attribution

markup a dataset can be found here. The dataset is part of a project to reconstruct the domestic market for colonial goods in the Dutch Republic.

ACTIVITIES:
1. Try to find one or two terms that are relevant to your research using the resources that are mentioned above. You can aso use Swoogle to search for vocabularies related to your research. 2. Search for a term related to your research in the CIDOC Conceptual Reference Model (CRM) concept search. Were you able to find it? Tip 1: Search for "person" to get an idea of how the thesaurus works. Tip 2: All the terms used can be found in the last release of the model: http://www.cidoc-crm.org/get-last-official-release.

Thing 7: FAIR data modelling

The fourth and the fifth star in Berner Lee's model can be awarded when the data are stored in a format in which the topics their properties and their characteristics are identified using URIs whenever possible. More concretely, it implies that you record your data using the Resource Description Framework (RDF) format. RDF, simply put, is a technology which enables you to publish the contents of a database via the web. It is based on a simple data model which assumes that all statements about resources can be reduced to a basic form,

FAIR Data management wizard

ELIXIR
Data Stewardship Wizard

Smart Data Management Plans for FAIR Open Science

For serious researchers and data stewards

Data Stewardship Wizard

Data integration 7

Data interpretation 3

Information and insight 14

Is there any pre-existing data?

Are there any data sets av

Data Stewardship for C

No

Yes

Will reference data be cr

Will any of the data that y
others)?

Data Stewardship for C

No

Yes

Will you be storing sample

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.

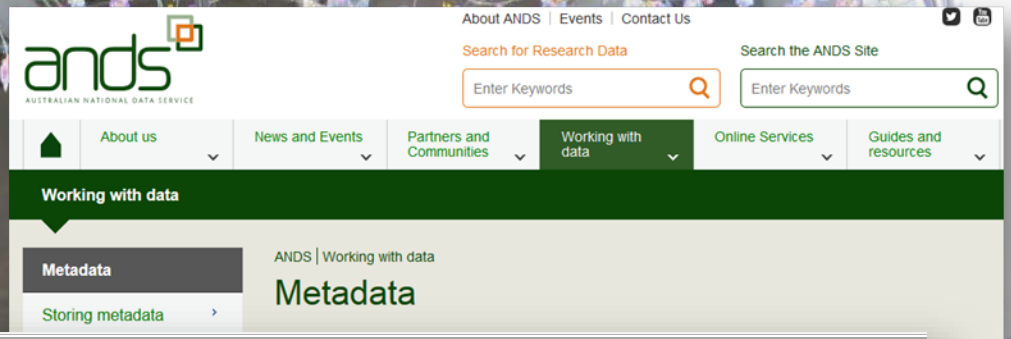
Do

- Always carefully choose a license to be attached to your data upon publication.
- Include and clearly mark the licences 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	*The minimal metdata, required to find a digital object.* *If there are additional contextual metadata, a user will have a better idea on how to use the data*	*Author, title, abstract, date* *Contextuele metadata are for example location, time, data collection method (tools)*
Structural metadata	*These link the individual objects of a unity*	*Links to related digital objects, (e.g. the article written based on the linked research data)*
Technical metadata	*Information on the technical aspects of the data set*	*Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard*
Administrative metadata	*Metadata focusing on user rights and management of digital objects*	*License, possible reasons for an embargo, waivers*

data

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

http://rd-alliance.github.io/metadata-directory/standards/

F = findable. Metadata tools

What CEDAR does

https://metadatacenter.org/

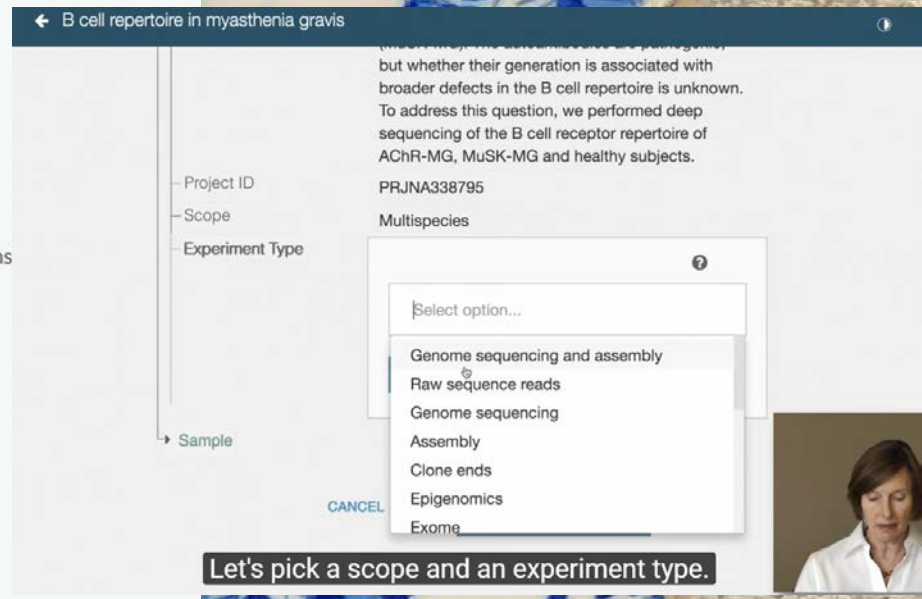
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools wil
metadata record is created to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools hel
to users, and download the information that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

What can CEDAR do for me already?

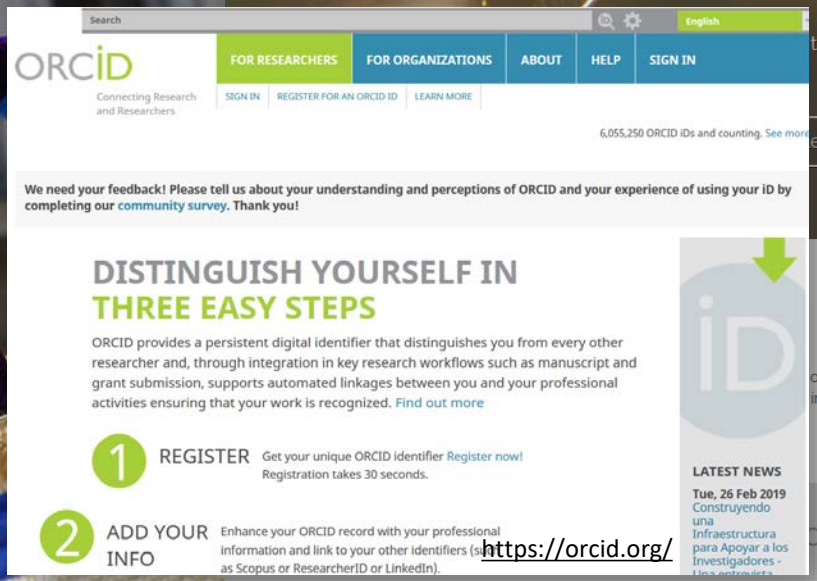
As of its production release, in February 2017, CEDAR addresses these scenarios:

- 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 critieria, 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.



F = Findable. Persistent identifiers



https://orcid.org/

https://www.datacite.org/

- ASSIGN A DOI
DIGITAL OBJECT IDENTIFIER
- USE ORCID

ORCID – ID. Tutto intorno a me

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

F = Findable. Persistent identifiers

Here are some identifier schemes:

- ARK (Archival Resource Key) – a URL with extra features allowing you to ask for descriptive and archival metadata and to recognize certain kinds of relationships between identifiers. ARKs are used by memory organizations such as libraries, archives, and museums. They are resolved at "http://www.nt2.net". Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- DOI (Digital Object Identifier) – an identifier that becomes actionable when embedded in a URL. DOIs are very popular in academic journal publishing. They are resolved at "http://dx.doi.org". Resolution depends on HTTP redirection and the Handle identifier protocol, and can be managed through an API or a user interface.
- Handle – an identifier that becomes actionable when embedded in a URL. Handles are resolved at "http://www.handle.net/". Resolution depends on HTTP redirection and the Handle protocol, and can be managed through an API or a user interface.
- InChI (IUPAC International Chemical Identifier) – a non-actionable identifier for chemical substances that can be used in printed and electronic data sources, thus enabling easier linking of diverse data compilations.
- LSID (Life Sciences Identifier) – a kind of URN that identifies a biologically significant resources, including species names, concepts, occurrences, and genes or proteins, or data objects that encode information about them. Like other URNs, it becomes actionable when embedded in a URL.
- NCBI (National Center for Biotechnology Information) ACCESSION
- a non-actionable number in use by NCBI.
- PURL (Persistent Uniform Resource Locator) – a URL that is always redirected through a hostname (often purl.org). Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- **URL (Uniform Resource Locator)** – the typical "address" of web content. It is a kind of URI (Uniform Resource Identifier) that begins with "http://" and consists of a string of characters used to identify or name a resource on the Internet. Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using the HTTP protocol. Well-managed URL redirection can make URLs as persistent as any identifier. Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- **URN (Uniform Resource Name)** – an identifier that becomes actionable when embedded in a URL. Resolution depends on HTTP redirection and the DDDS protocol, and can be managed through an API or a user interface. A browser plug-in can save you from typing a hostname in front of it.

https://dmptool.org/general_guidance#persistent-identifiers

… [SE NON POTETE USARE UN DOI]

FORCE11
The Future of Research Communications and e-Scholarship

ABOUT | COMMUNITY | GROUPS | RESOURCES | NEWS + BLOGS | EVENTS | PUBLICATIONS

FORCE11 » Community News » Introducing a new standard for the citation of research data

INTRODUCING A NEW STANDARD FOR THE CITATION OF RESEARCH DATA

Posted by Jennifer McLennan | May 8, 2018 | Sign In or Join Now! to post comments

Rules, registry and recommendations

Compact Identifiers. A "compact identifier" is a string constructed by concatenating a namespace prefix, a separating colon, and a locally unique identifier (LUI), e.g. pdb:2gc4.

Provider Specification. To specify a specific provider, where multiple providers exist, prepend the provider code and a "/" to the compact identifier, e.g. rcsb/pdb:2gc4.

Provider Default. Where multiple providers exist, and the provider is not specified in the compact identifier, the resolver will determine where to resolve the request based on its own rules, e.g., taking into account uptime availability, regional preference, or other criteria.

Redirect Rule. A URL template associated with the provider code is maintained in the namespace registry, defining how to forward compact identifiers to any specific provider (see 4.2.3 below).

SCIENTIFIC DATA

Altmetric: 20 May 8, 2018 More detail »

Article | OPEN

Uniform resolution of compact identifiers for biomedical data

Sarala M. Wimalaratne, Nick Juty, John Kunze, Greg Janée, Julie A. McMurry, Niall Beard, Rafael Jimenez, Jeffrey S. Grethe, Henning Hermjakob, Maryann E. Martone & Tim Clark

tember 2017
uary 2018
y 2018

A reasonable solution to the identifiers problem is to assign Digital Object Identifiers (DOIs) to identify datasets. DOIs are already widely used in the publishing world as persistent identifiers for scholarly publications. They have been adopted by generalist data repositories such as Dryad, FigShare, Zenodo and Dataverse, as well as by domain data repositories outside of biomedicine. Handles[14], which underlie the DOI system, may also be used directly. The DataCite consortium provides a robust central means for assigning DOIs to data.

However, DOIs are not commonly used for biomedical data, which is partitioned across over 600 autonomous repositories that are independently funded. Instead, in biomedicine there has been a

A = Accessible

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

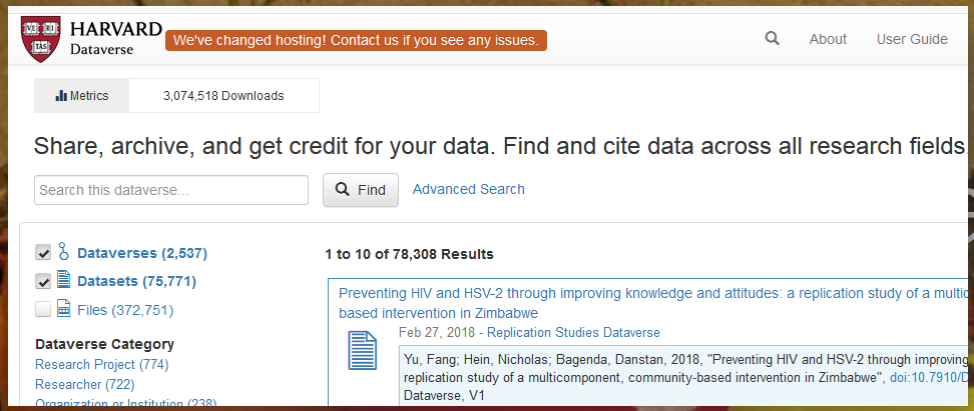
- **Open access**
  Data that can be accessed by any user whether they are registered or not. Data in this category shouldn't contain personal information (unless consent is given (see 'Informed consent').

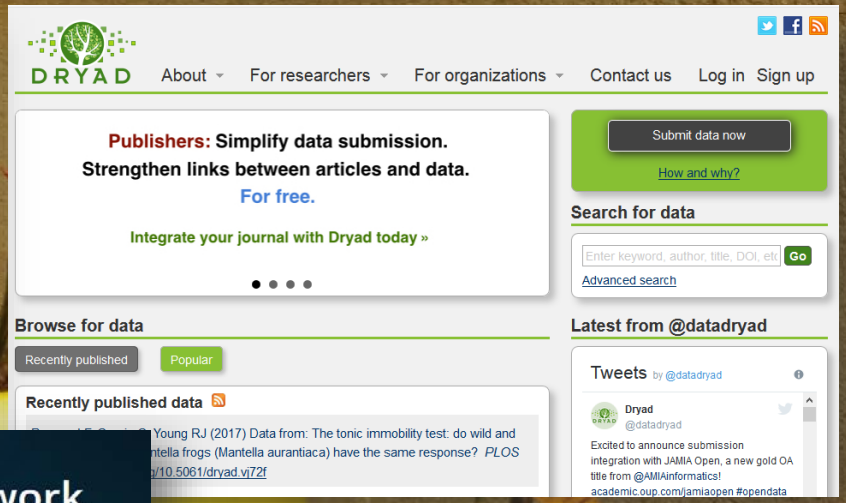
- **Access for registered users (safeguarded)**
  Data that is accessible only to users who have registered with the archive. This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

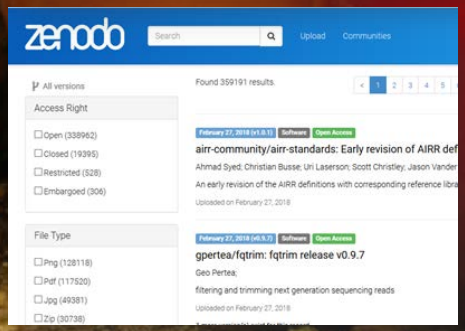
- **Restricted access**
  Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.
  Restricted access requires long-term commitment of the researcher or person responsible for the data to handle the upcoming the permission requests.

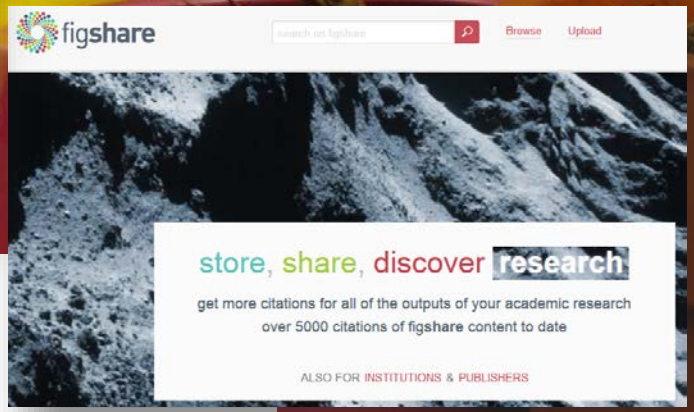
- **Embargo**
  Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after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

CESSDA Guide







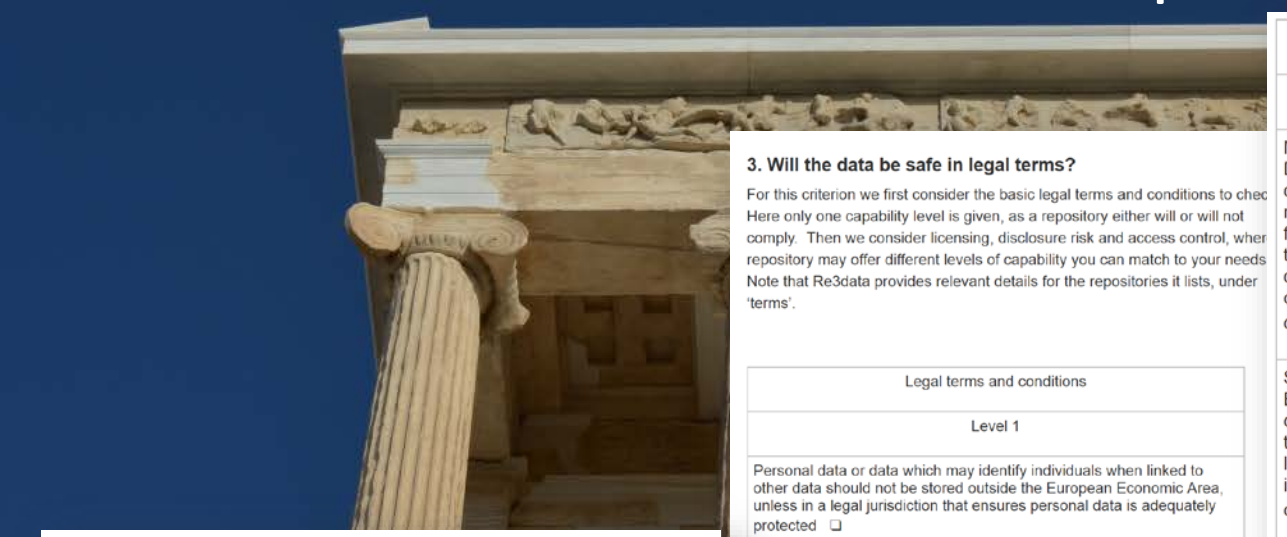


General depositories for research data

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 Science[1]): 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 repositories



3. Will the data be safe in legal terms?

For this criterion we first consider the basic legal terms and conditions to chec
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, wher
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'.

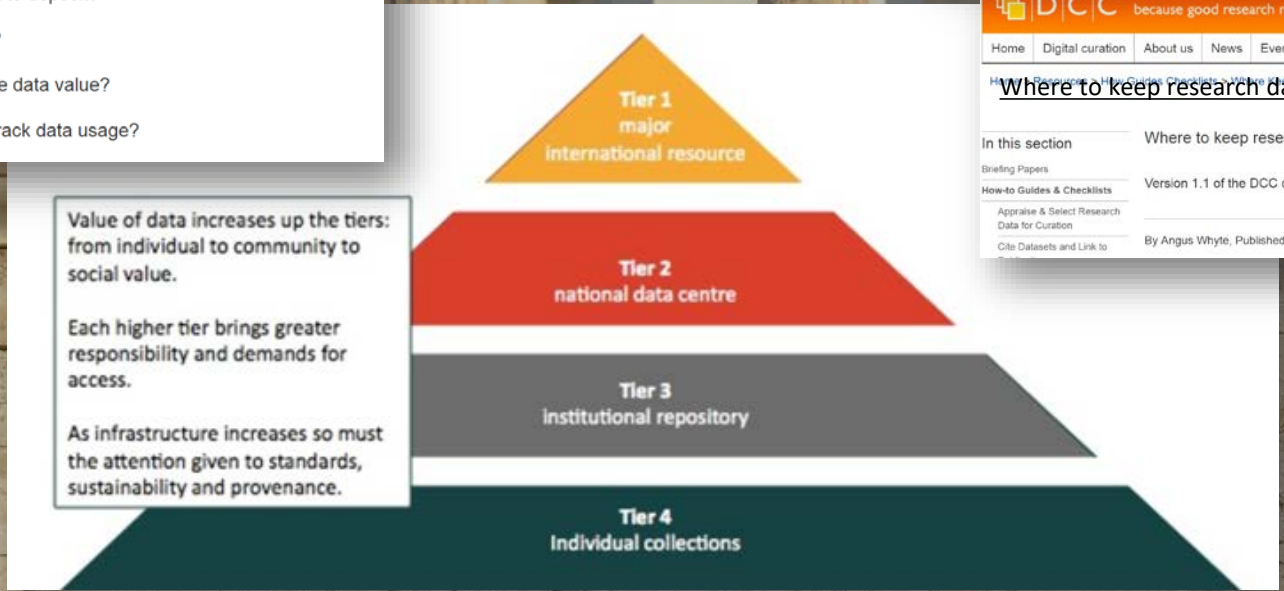
Legal terms and conditions
Level 1
Personal data or data which may identify individuals when linked to other data should not be stored outside the European Economic Area, unless in a legal jurisdiction that ensures personal data is adequately protected ❑
By agreeing to the terms and conditions the depositor 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	Provides metadata elements to enable broader discovery (e.g. geo-spatial) to	Exposes discovery metadata as Linked Open Data to optimise automatic

Checklist: is it the right repository for your data?

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

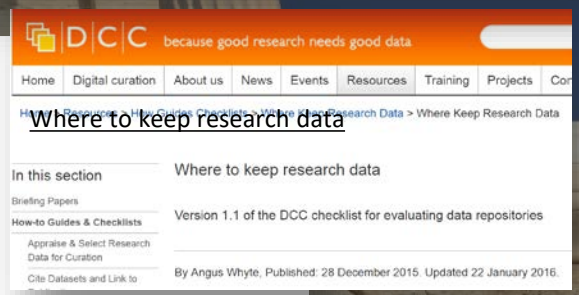
1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?



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.



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)
Content type(s)
Country
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)
Content type(s)
Country
The Drosophila Genetic Reference Panel (DGRP) is a populatio
complex traits, and a community resource for whole genome as

Online Mendelian Inheritance in Man
OMIM
Subject(s)

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 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

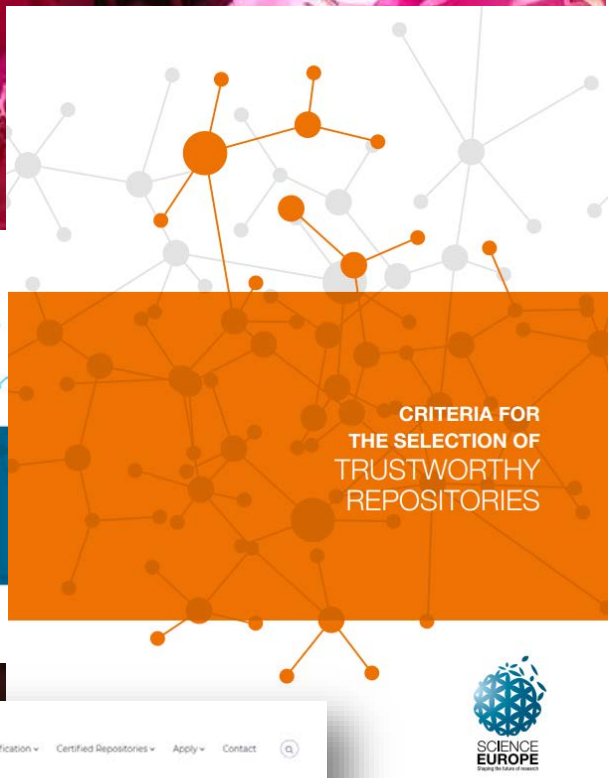
Art Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to created it, did not exist. And the answer to what this illustrates – this is simply a distorted...

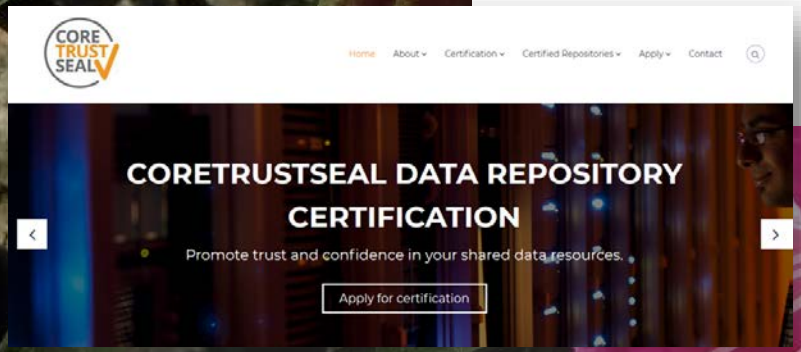
Read more

https://www.re3data.org/

[trusted repositories]

Dec. 2018





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)

A = Accessible. Data journals

Data Journals

Hier entsteht eine Liste von Data Journals, die vorwiegend Data Papers

- Atomic Data and Nuclear Data Tables (Elsevier)
- Biodiversity Data Journal (Pensoft Publishers)
- Biomedical Data Journal (Procon Ltd.)
- BMC Research Notes (Biomed Central)
- Chemical Data Collections (Elsevier)
- Data (MDPI)
- Data in Brief (Elsevier)
- Dataset Papers in Science (Hindawi Publishing Corporation)
- Earth System Science Data - ESSD (Copernicus Publications)
- Ecological Archives (Ecological Society of America - ESA)
- European Data Watch (European Data Watch)
- F1000Research (F1000 Research)
- Genomics Data (Elsevier)
- Geoscience Data Journal (Wiley)
- GigaScience (BioMed Central)
- Internet Archaeology (Internet Archaeology)
- Journal of Open Psychology Data (JOPD) (Ubiquity Press)
- Journal of Chemical & Engineering Data (ACS Publications)
- Journal of Physical and Chemical Data (AIP Publishing)
- Nuclear Data Sheets (Elsevier)
- Open Archaeology Data (Ubiquity Press)
- Open Data Journal for Agricultural Research (diverse)
- Open Health Data (Ubiquity Press)
- Open Journal of Bioresources (Ubiquity Press)
- Open Network Biology (BioMed Central)
- Research Data Journal for the Humanities and Social Sciences (Brill)
- Scientific Data (Nature Publishing Group)

Dataset Description

Object Name

- *walkers* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for records made by individual walkers during stage-one fieldwalking.
- *counts* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for pot-sherds countedduring stage-one fieldwalking.
- *pottery* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main pottery database, assembled various artefact specialists.
- *petrography* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for those sherds sampled for thin section petrography.
- *lithics* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main lithics database.
- *other* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all non-ceramic and non-lithic finds.
- *structs* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all standing remains, except for terraces.
- *coast* – a vector polygon dataset (.shp and associated files) with the shape of Antkythera's coastline.
- *geology* –a vector polygon dataset (.shp and associated files) with the main bedrock units on Antkythera.
- *tracts* – a vector polygon dataset (.shp and associated files) with the main stage-one survey units.
- *grids* – a vector polygon dataset (.shp and associated files) with the main stage-two survey units.
- *terraces* – vector line dataset (.shp and associated files) with all observable agricultural terraces (i.e. the location

- *other* – primarily Andrew Bevan (UCL), with further assistance from James Conolly (Trent)
- *geology* – a combination of fieldwork by Ruth Siddall (UCL) and remote sensing by Andrew Bevan (UCL)

Repository Location
UK Archaeology Data Service Collection 1115 (doi: 10.5284/1012484)

Publication Date
05/02/2012

Language
English (a Greek language summary of the project methods and results can be found at www.ucl.ac.uk/asp/ or www.tuarc.trentu.ca/asp/).

License
Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. http://dx.doi.org/10.5284/1000271 http://dx.doi.org/10.5284/1006 dx.doi.org/10.528 to the fact that th cal. The ASP data locations, dates a ally in the databas structures and ter

Data journals

Panayiota Polydoratou

Alexander Technological Educational Institute of Thessaloniki

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

A = Accessible. Formati

Data Archiving and Networked Services

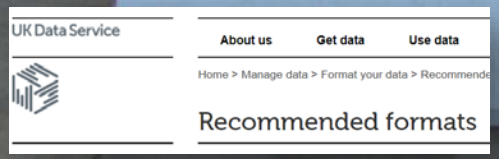
DANS

HOME DEPOSIT

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)

https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats

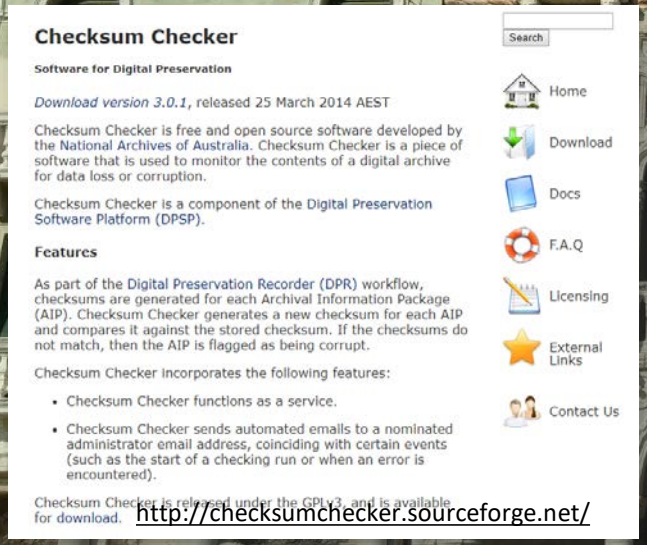
Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages
Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0



https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx

A = Accessible. Conservazione

LUNGO O BREVE TERMINE?



Checksum Checker

Software for Digital Preservation

Download version 3.0.1, released 25 March 2014 AEST

Checksum Checker is free and open source software developed by the National Archives of Australia. Checksum Checker is a piece of software that is used to monitor the contents of a digital archive for data loss or corruption.

Checksum Checker is a component of the Digital Preservation Software Platform (DPSP).

Features

As part of the Digital Preservation Recorder (DPR) workflow, checksums are generated for each Archival Information Package (AIP). Checksum Checker generates a new checksum for each AIP and compares it against the stored checksum. If the checksums do not match, then the AIP is flagged as being corrupt.

Checksum Checker incorporates the following features:

- Checksum Checker functions as a service.
- Checksum Checker sends automated emails to a nominated administrator email address, coinciding with certain events (such as the start of a checking run or when an error is encountered).

Checksum Checker is released under the GPLv3, and is available for download.

Search

Home

Download

Docs

F.A.Q

Licensing

External Links

Contact Us

http://checksumchecker.sourceforge.net/

Storage Solutions	Advantages	Disadvantages	Suitable for
Personal Computer & Laptop	*Always available* *Portable*	*Drive may fail* *Laptop may be stolen*	*Temporary storage*
Networked drives File servers managed by your university, research group or facilities like a NAS-server	*Regularly backed up* *Stored securely in a single place*	*Costs*	*Master copy of your data* *(if enough storage space is provided ..)*
External storage devices USB flash drive, DVD/CD, external hard drive	*Low cost* *Portability*	*Easily damaged or lost*	*Temporary storage*
Cloud services	*Automatic synchronization between folders and files* *Easy to access and use*	*It's not sure whether data security is taken care of* *You don't have direct influence on how often backups take place and by whom*	*Data sharing*

1 2 3 4 5 **6**

Organize and document research data. Make digital versions of paper data documentation in a PDF/A format (suitable for long-term storage).

A = Accessible. Conservazione

Portable devices | Cloud storage | Local storage | Networked driv

Laptops, tablets, external hard-drives, flash drives and Compact Discs

Advantages	Disadvantages/Risks	Pre (ser dat
• Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field. • Low-cost solution.	• Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk. • Not robust for long-term storage or master copies of your data and files. • Possible quality control issues due to version confusion.	Use enc pas

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
• Automatic backups. • Often automatic version control.	• Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens. • Insufficient control over where the data is stored and how often it is backed up. • Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes. • Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.	• Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see 'Security' for more information on encryption; also see 'Protecting data').

Recommendations

- Do: use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.
- Do: Read the terms of service. Especially focus on rights to use content given to the service provider.
- Do: Opt for European, national, or institutional cloud services which store data in Europe if possible.
  - B2drop (EUdat, n.d.) is an example of a European cloud storage solution.
  - SWITCHdrive (SWITCH, 2017) is a Swiss solution.
  - DataverseNL (Data Archiving and Networked Services, 2017) is an example of a service for Dutch researchers that allows the storage and sharing of data both during and after the research period.
- Don't: make this your only storage and backup solution.
- Don't: use for unencrypted (sensitive) personal data.

CESSDA Guide

Differenti bisogni, differenti strumenti. Durante l'esperimento, dovete poter condividere con il team

Cosa conservare?

DCC because good research needs good data

Home | Digital curation | About us | News | Events | Resources | Training | Projects

Home > Resources > How Guides > How Develop Rdm Services

How to develop

In this section

How to Develop RDM Services - a guide for HEIs

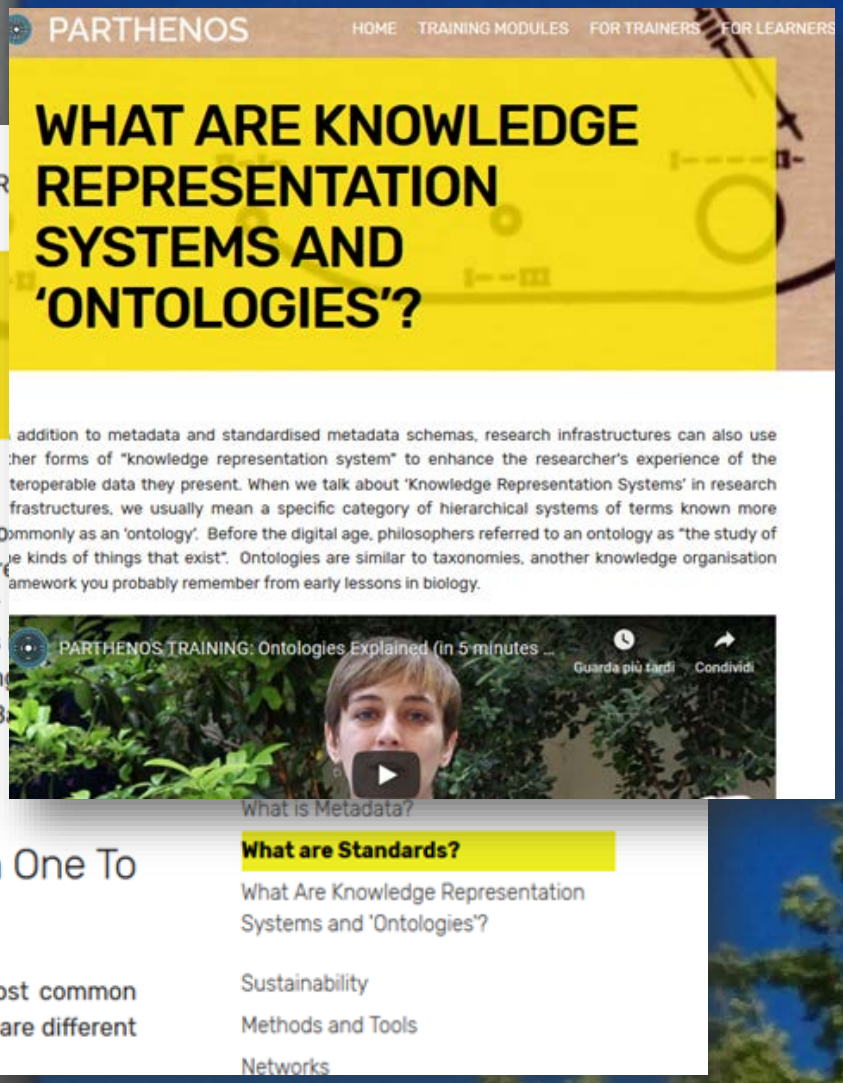
Establishing criteria for selection decisions

You should establish criteria to guide selection decisions. The DCC's How to Select and Appraise Research Data for Curation[56] proposes seven criteria as outlined below:

1. **Relevance to mission**: the resource content fits any priorities stated in the institution's mission, or funding body policy including any legal requirement to retain the data beyond its immediate use.
2. **Scientific or historical value**: is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
3. **Uniqueness**: the extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
4. **Potential for redistribution**: the reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
5. **Non-replicability**: it would not be feasible to replicate the data/resource or doing so would not be financially viable.
6. **Economic case**: costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
7. **Full documentation**: the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource's provenance and the context of its creation

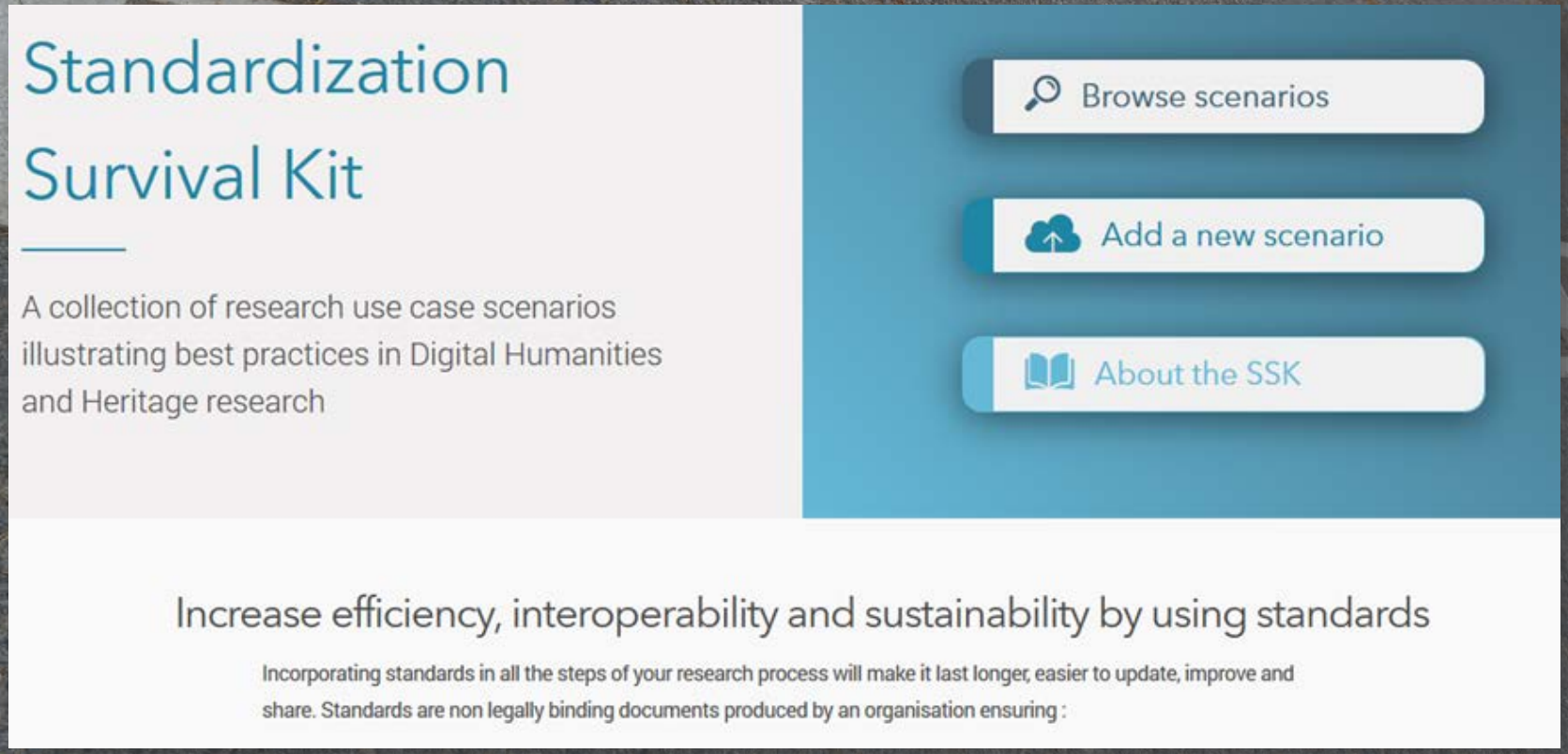
- Rilevanti per la missione dell'ente
- Valore storico
- Unicità
- Potenziale di riuso
- Non replicabili
- Costo/benefici
- Documentazione completa

I = Interoperable. Standards



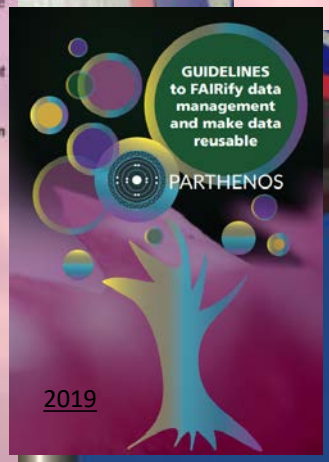
http://training.parthenos-project.eu/sample-page/intro-to-ri/interoperability/what-are-standards/

I = Interoperable. Standards

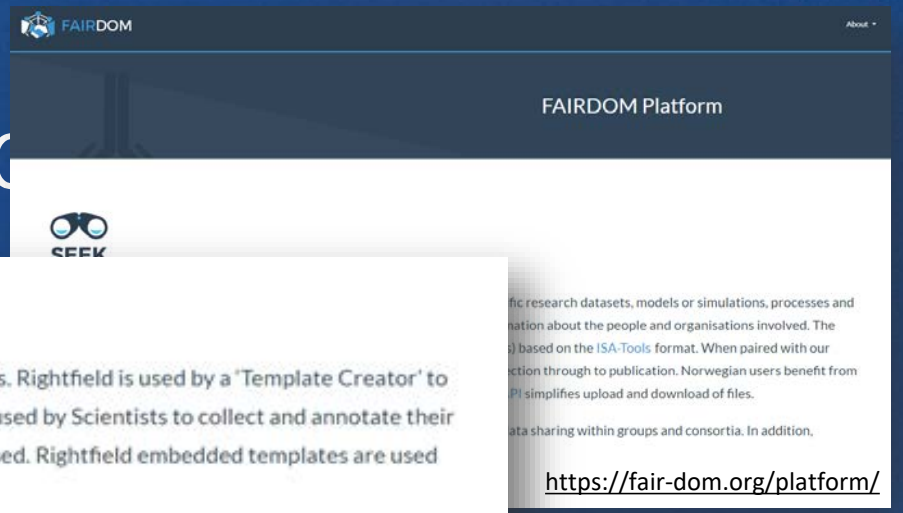


http://ssk.huma-num.fr/#/

I = Interoperable



I= Interoperable – ac



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

RightField

Rightfield is an open-source tool for adding ontology term selection to Excel spreadsheets. Rightfield is used by a 'Template Creator' to create semantically aware Excel spreadsheet templates. The Excel templates are then reused by Scientists to collect and annotate their data; without any need to understand, or even be aware of, Rightfield or the ontologies used. Rightfield embedded templates are used within the Samples framework of the SEEK.

MORE INFORMATION

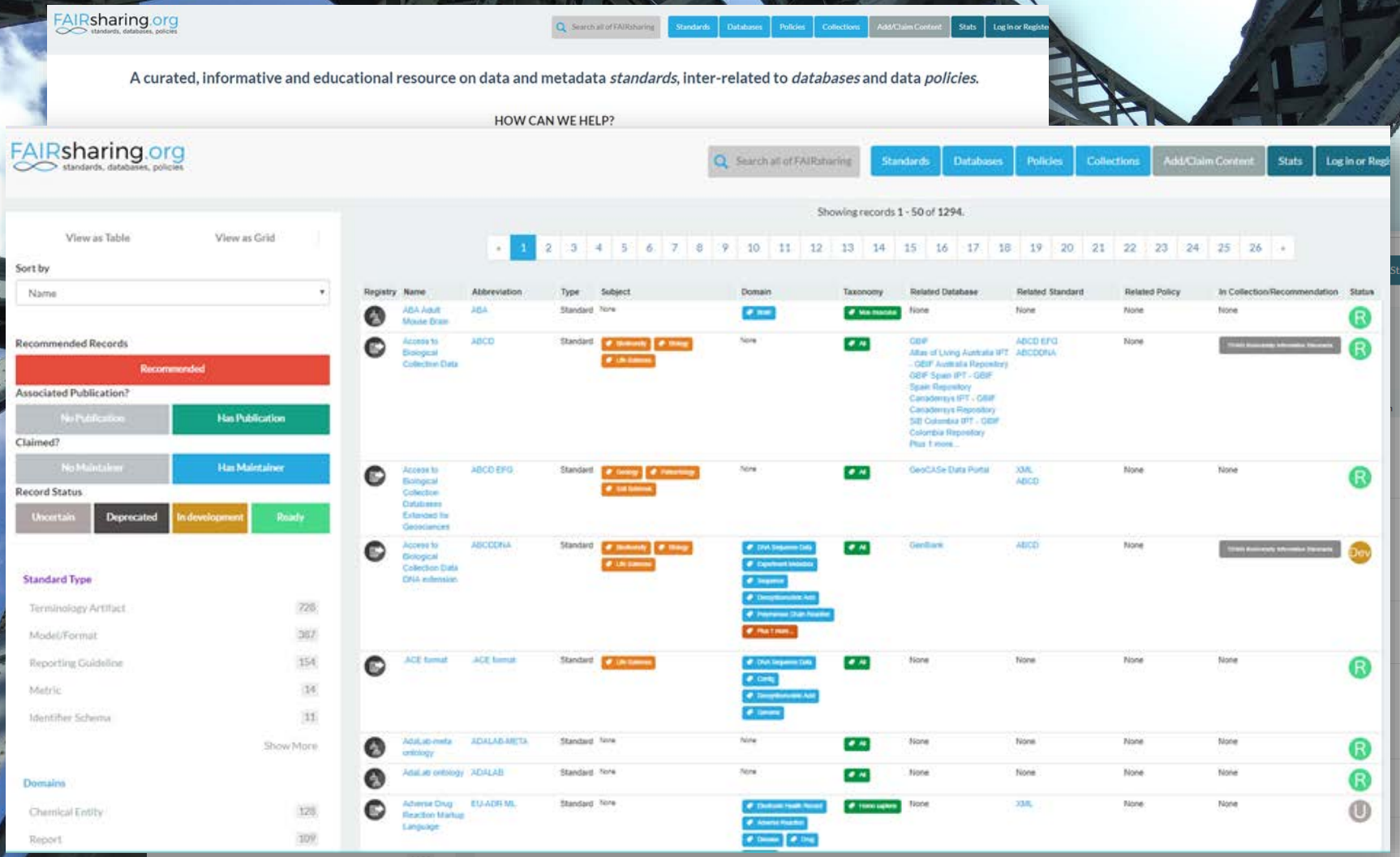
https://fair-dom.org/platform/rightfield/

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 exisitng Assay, you can link to it via the Assay SEEK ID.
Assay Title			The title of an exisitng 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			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	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

amplification
by-product_formation
catabolic_response
cell_growth_optimisation
cell_size
Comparative_genomic_hybridization
comparative_genomics
continuousEnzymatic

Metadata | Organism_Sample | Instrument_(optional) | Data_ | Result_

[FAIRsharing. Per essere interoperabili]



https://fairsharing.org

R = Reusable. Documentazione

cessda TRAINING

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:

- ⊕ 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

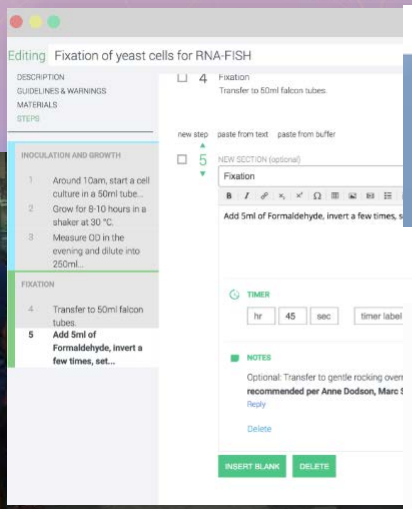
https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata

R = Reusable. Documentazione

protocols.io

Make your science more reproducible

protocols.io is the #1 open access repository for science methods

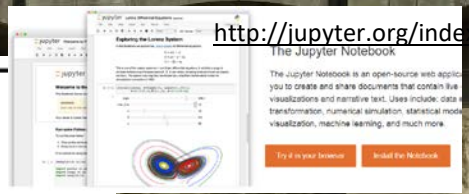


RStudio

https://www.rstudio.com/

RStudio

Open source and enterprise-ready professional software for R



http://jupyter.org/index.h

What is an Open Notebook?

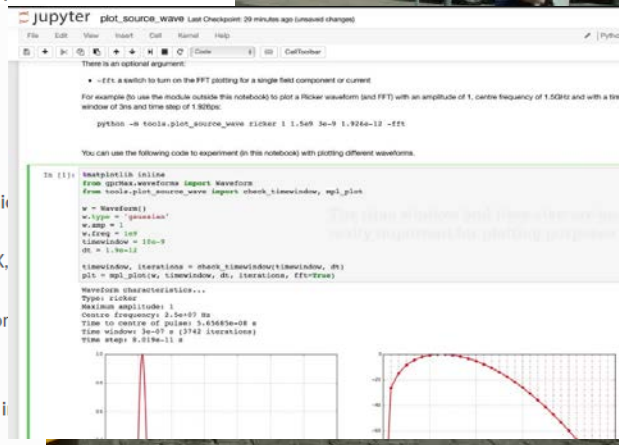
Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

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, indentation, and tab completion/introspection.
- **Run code from the browser**, with the results of computations attached to the code whi 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 cor 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 i browser by MathJax.



…why not?

- protocols.io to deposit your methods
- Open Lab Notebook to track anything you do [time consuming the first time, then…]

R= Reusable. Licenze

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

KEEP CALM AND

RICORDA: NESSUN COPYRIGHT SUI DATI (NON CREATIVI)

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

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 11 March 1996

on the legal protection of databases

NCIL OF THE EUROPEAN UNION,

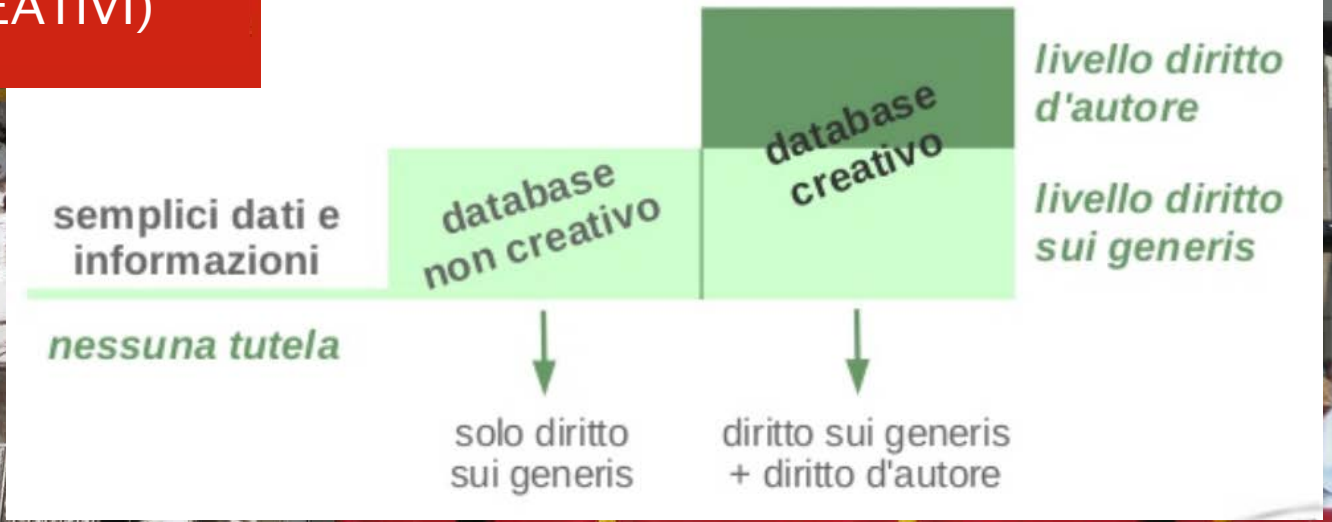
n Community, and in particular Article 57 (2), 66 and 100a thereof,

(1),

Simone Aliprandi

2014

QUALI DIRITTI SUI DATI?



[webinar]


OpenAIRE
SERVICES
SUPPORT
OPEN SCIENCE IN EUROPE
ABOUT
Italian Open Science Webinar Series 2019
CONTROL
REGULATIONS
LAW
POLICY
COST
STANDARDS
SECURITY
COMPLIANCE
RISKS
RULES
STRATEGY
OpenAIRE
RDA
RESEARCH DATA ALLIANCE EUROPE
IT NATIONAL NODE
Italian Open Science Webinar Series 2019


OpenAIRE
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

[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.

A report from Knowledge Exchange (Knowledege 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

- **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 th right holder, only Dutch and UK law designate the employer as the right holder if the worl was made in the course of employment.

CESSDA guide

⊕ Czech Republic

⊕ Finland

⊕ Switzerland

⊕ UK

[regole chiare dall'inizio]

Safe to be open 2016

… FISSATE REGOLE CHIARE DA SUBITO

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

R = Reusable. Licenze

DCC 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

R = Reusable. Licenze

DCC 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.

Creative Commons at a glance

Good for

- very simple, factual datasets
- data to be used automatically

Watch out for

- versions: use v. 4 or later
- attribution stacking
- the NC condition: only use with dual licensing
- the SA condition as it reduces interoperability
- the ND condition as it severely restricts reuse

ODC-By at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking

ODC-ODbL at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking
- the copyleft condition as it reduces interoperability
- the DRM clause as it may put off some reusers

Public domain at a glance

Good for

- most databases and datasets
- data to be used by anyone or any tool
- data to be used for any purpose

Watch out for

- lack of control over how database is reused
- lack of protection against unfair competition

Creative Com

creative commons UK

FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE V.0.1

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

PUBLIC DOMAIN MARK 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.

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

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN

CC0	Public Domain
'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

CC0	Public Domain
'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

nmons e Op

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.

USE A CC0

- THEN ASK FOR CREDIT
- PROVIDE A CITATION TO C&P
- BEAR IN MIND IT'S BAD SCIENCE NOT TO CITE THE SOURCE
- CC0 DOES NOT MEAN ACADEMIC UNPOLITENESS

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.

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.

...in sintesi...


12
3
PATRIOTS
12
9
Green Bay Packers
San Francisco 49ers
New Orleans Saints


OHIO STATE
2
AUBURN
21
TEXAS A&M
5
TEXAS
25
Ohio State Buckeyes
Auburn Tigers
Texas A&M
Texas Longhorns

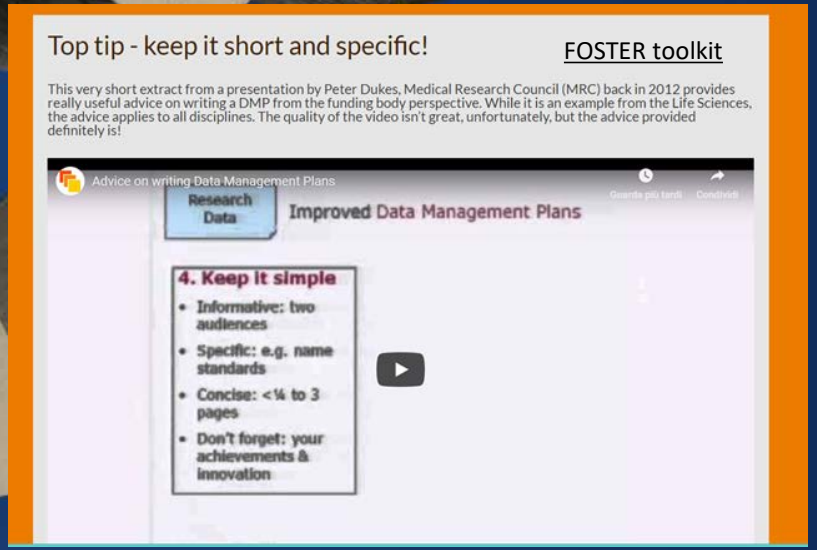

SPURS
BULLS
SIANTS


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
S. Aliprandi

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

Vantaggi di un DMP

CESSDA Guide

Benefit 3. Clarifies needed budget

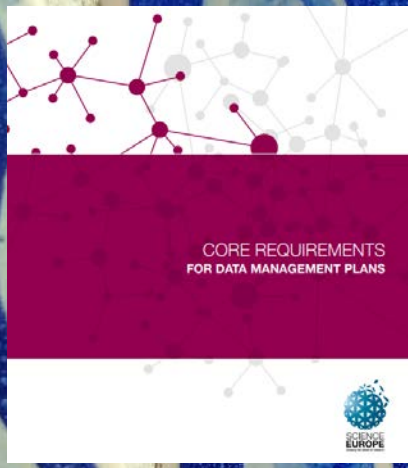
Data management is not free. You do not want to find yourself running out of funding before the end of the project because you have ignored or underestimated the cost of structured, detailed, and safe data management. Therefore, an important aspect of a DMP is its use in calculating how much money will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

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?

Dec. 2018

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)?

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	
1a **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.

2 DOCUMENTATION AND DATA QUALITY	
2a **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.
2b **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

PLAN

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?

ORGANISE & DOCUMENT

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?

https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf

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?

cessda
Consortium of European Social Sciences Data Archives

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?

DMP question

Archiving and publishing your data

Archiving
- How and where will the data be stored after the project's completion?
- Will you archive your data in a trusted data repository?
- Will your data receive a persistent identifier?

Data formats
- What formats will you provide your data in for archiving (and sharing)?
- Will specific software be required to process your data? Can this software be deposited with the data?

Access (if applicable)
- Will your data be available (Open Access)?
- Will all data or only parts of it be published?
- What licenses do you need for your data?
- How should your data be cited when reused?
- Will there be an embargo period for (all or some of) the data?
- Are there other agreements or restrictions (see above) that need to be considered?
- Are there any legal/ethical restrictions that prevents the publication of all the material?
- Will these restrictions mean that action must be taken before the material can be made available?
- Is there a risk of delayed publication/making data available (all or parts of)? If so what might be needed to do to avoid this?

DISCOVER

Discovering data

Identification of needs
- Do you plan to use existing data for your research?
- What is the purpose for which you need the data?
- What do you want to learn from the data?
- What type of data do you need?

Search for data
- Do you know where the data may be located?
- How do you plan to search for the data?

Evaluation of data quality
- What is the minimal required quality of the data (in terms of origin, contents, scope, size, methods, etc.)?
- How do you plan to evaluate data quality (evaluation of metadata, tests, analysis, comparisons)?

Gaining access to data
- What are the (expected) terms and conditions for data access and use?
- What is the (expected) process for gaining access to the data?
- What is the (expected) time-span of the process for gaining access to the data?
- What are the (expected) costs for data access and use?

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).

FAIRDOM About Platform

Data Management Checklist

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

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.

FAIRDOM About Platform

Data Management Checklist

Making data findable (documentation and metadata management)

- What documentation and metadata will accompany the data (assist its discoverability)? (Details on methodology, definitions, procedures, SOPs, vocabularies, units, dependencies, etc)
- What information is needed for the data to be read and interpreted in the future?
- What naming conventions will be used?
- How will you approach versioning your data?
- How will you capture / create this documentation and metadata?
- How do you ensure the completeness of the captured data?

Making Data Accessible

- Specify which data will be made openly available taking into consideration
  - What ethics and legal compliance issues do you have if any? Do you need consent for data preservation and sharing? Do you have to protect certain data? Is any data sensitive?
  - Do you think you might have Intellectual Property Rights issues? Have you considered ownership of the data, licensing, restrictions on use?
  - Do you think you will need to embargo any data?
- How will you make the data available? (consider the platforms you will use: databases, repositories, etc)
- What methods or software tools are needed to access the data? You should list where the software can be obtained. You should also document how to use the software to access the data. The documentation should be as complete as possible, including examples. If you distribute your system, include the access software and its documentation as part of any distribution.
- If there are any restrictions on accessibility, how will you provide access?

Making Data Interoperable

- What standards (metadata vocabularies, formats, checklists) or methodologies will you use?
- How do you address data and model quality? What validation steps do you foresee?
- Will you use standardised vocabulary for all data types to allow inter-disciplinary interoperability?
- Where you can not used standardised vocabulary for all types of data, can you map to more commonly used ontologies?

Making data Re-usable

- How will you licence your data to permit the widest re-use possible?
- When will the data be made available for re-use? Does this include an embargo period? (if yes, please detail why)
- Which data will be available for re-use during/after the project? For data that is not re-usable, please detail why
- What are your data quality assurance processes?
- How long do you expect your data to remain re-usable?

PERSONALIZZABILE

DMP online

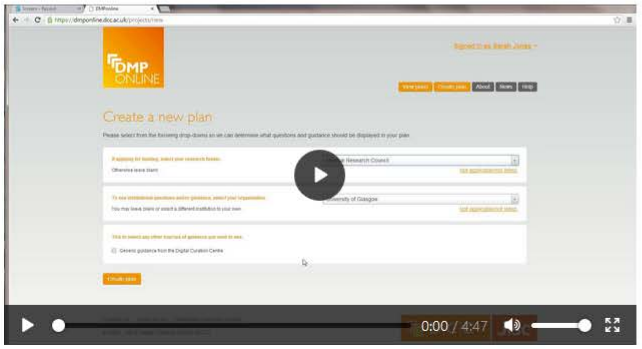
DMP ONLINE

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).

Screencast on how to use DMPonline



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?

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.

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

… servono esempi [NON modelli]?…

https://dmptool.org/public_plans

DMPTool
Build your Data Management Plan

Public DMPs

Public DMPs are plans created using the DMPTool service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Search

Project Title	Template	Organizatio
CAPERS - Cloud Applications and Platforms Enhancing Research Scalability	NSF-CISE: Computer and Information Science and Engineering	Georgia Tec
sample	NSF-SBE: Social, Behavioral, Economic Sciences	Binghamton University
Generic NSF DMP	NSF-GEN: Generic	Binghamton University
Sample NEH Plan	NEH-ODH: Office of Digital Humanities	Binghamton University
sample DMP Plan for workshop	NSF-CHE: Chemistry Division	Binghamton University
Sample NIH-GEN dmp	NIH-GEN: Generic	Binghamton University
As bibliotecas universitárias diante das mudanças contemporâneas	IMLS: Datasets	São Paulo S University (U
Gestão de acervos audiovisuais em repositórios	Digital Curation Centre	Non Partner Institution
Scientific Journals publishing	Digital Curation Centre	Non Partner Institution
The contribution of water retention, nutrient loading and microbial community to mosquito breeding and West Nile virus transmission in Spokane County	U.S. Geological Survey (USGS)	Non Partner Institution

View all

Horizon 2020

AtlantOS Data Management Plan
A 2016 DMP covering the concept of FAIR data for an Altalntic Ocean observation systems project

Centre of Excellence in Simulation of Weather and Climate in Europe (2016)
DMP provided by the German Climate Centre (DKRZ)

FREME Data Management Plan
Plan developed by the Open framework of e-services for multilingual and semantic enrichment of digital content project

Helix Nebula open science cloud
A High Energy Physics example covering the Large Hadron Collider experiments

Open Sea Operating Experience to Reduce Wave Energy Costs
First version of a DMP by the opera project

Tweether
An engineering based micro-electronics example

AutoPost
An example DMP from an industry-driven innovation action that will deliver ICT-based solutions to enhance established post-production workflows.

Evaluation rubric for assessing H2020 plans
Assessment framework developed according to the original H2020 template

http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

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.

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Guide 2016

Version 3.0
26 July 2016

IN HORIZON 2020
DMP ENTRO M6
COME DELIVERABLE

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.

Open access to data

GRANT AGREEMENT
ART. 29.3
pag. 248

https://goo.gl/sryNTg

3. Open access to research data (Extended Open Research Data Pilot)

What?

Beneficiaries of actions that participate in the Open Research Data Pilot must give **open, free-of-charge access** to the end-user to **digital research data** generated during the action (⚠ new in Horizon 2020).

⚠ As of the Work Programme 2017, the Open Research Data pilot has been extended to all thematic areas of Horizon 2020 (except ERC PoC actions, SME instrument Ph1 actions, ERA-NET Cofund actions that do not produce data, EJP Cofund actions, and prizes).

Participation is therefore now in principle **the default**. However, actions may **opt out** at any stage — both before signing the GA and afterwards (through an amendment; *see Article 55*) —, if:

- participation is incompatible with the obligation to protect results *(see Article 27)*
- participation is incompatible with the security obligations *(see Article 37)*
- participation is incompatible with rules on protection of personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate/collect any research data or
- there are other legitimate reasons not to take part.

- PILOT PROJECT EXTENDED TO ALL DISCIPLINES
- OPT OUT CLAUSES

PRINCIPLE: **«AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»**

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

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf#page=238

Open data – H2020

https://goo.gl/sryNTg

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

... ora tocca a voi!