

Open Science dalla A alla Z

5 – Dati FAIR

Università Bocconi,
25 novembre 2022

Elena Giglia

Università di Torino

elena.giglia@unito.it



@egiglia



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). Photos are mine. Feel free to reuse from Flickr/eg65

Impareremo in questo modulo
1. Cosa significa FAIR BY DESIGN
nella ricerca quotidiana e in
Horizon Europe

MESSAGGI CHIAVE

- FAIR è il futuro
(se no si resta tagliati fuori)
- È più facile di quanto sembri...

...e non dimenticate Horizon Europe

Proposal template Part B: technical description

1.2 Methodology [e.g. 15 pages]

- Research data management and management of other research outputs: Applicants generating/collecting data and/or other research outputs (except for publications) during the project must provide maximum 1 page on how the data/ research outputs will be managed in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable), addressing the following (the description should be specific to your project): [1 page]

ANNEX 5

SPECIFIC RULES

COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (— ARTICLE 17)

Open science: research data management

- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle ‘as open as possible as closed as necessary’, unless providing open access would in particular:

LA GESTIONE DEI DATI FAIR RIENTRA

- FRA LE PRATICHE RACCOMANDATE
- FRA LE PRATICHE OBBLIGATORIE

DOVETE DECLINARLE ENTRAMBE IN MAX UNA PAGINA NELLA SEZIONE EXCELLENCE DELLA PROPOSTA

2. DATI FAIR BY DESIGN



Perché serve FAIR

I DATI SONO DIFFICILI DA
TROVARE

UNA VOLTA TROVATI,
DIFFICILE ACCEDERE

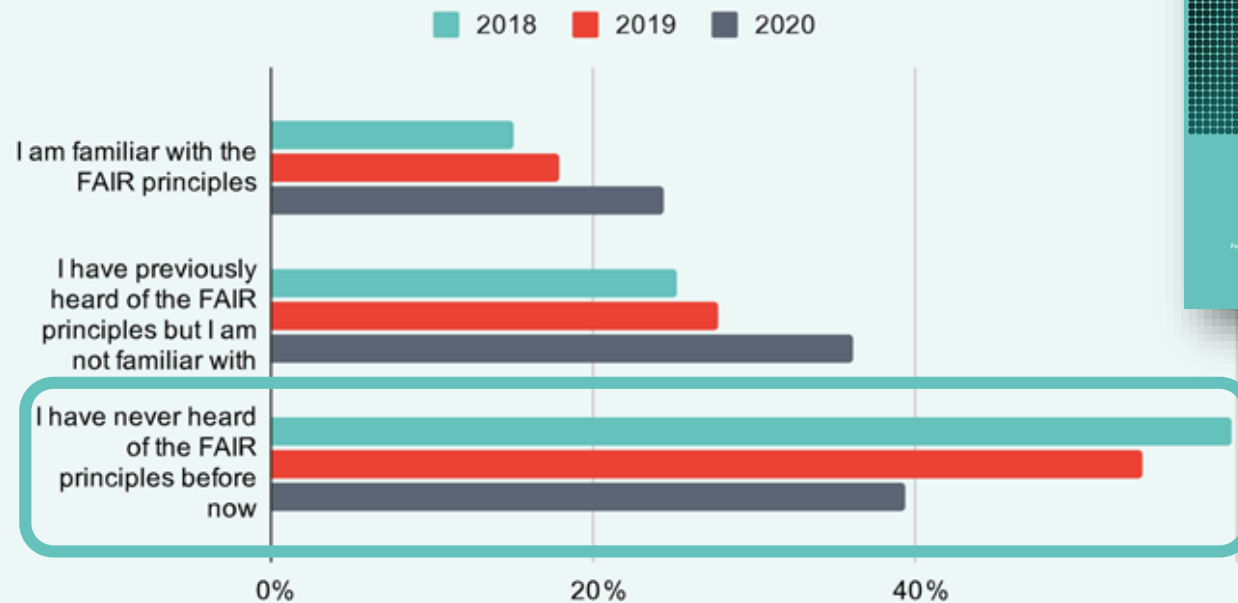
UNA VOLTA AVUTO ACCESSO,
DIFFICILI DA CAPIRE

... E SI PERDE MOLTO TEMPO
A SISTEMARLI O RICREARLI
SE NON TROVATI

PER QUESTO SERVE FAIR

Dati FAIR???

How familiar are you with the FAIR principles in relation to open data?



FAIR train – GoFAIR video

FARM DATA 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

IL TUTTO, LEGGIBILE DALLE MACCHINE

Kissed or missed?



PRINCIPI FAIR SONO
«MACHINE ACTIONABLE»
(DIVERSO DA READABLE)
FAIR = FULLY AI READY



Decision making procedures in data management and data stewardship for Open Science



Data-centric AI

Automated decision making using data.

Data is fundamental for training and deploying AI models.

Data management and/or curation is a crucial step to feed into AI model.

'Machine learning models are only as good as the data they're trained on' -

<https://fairmlbook.org/datasets.html>

(Chapter 8)

Clearbox AI

[Clearbox](#)

We are on a mission to harness powerful AI technologies to improve businesses and society in a trustworthy and human-centered way.

flexible product / Real

clearbox AI

Your

Synthetic Data

provider



Data stewardship challenges & AI ethics



Black box AI - Model inputs and operations remain a mystery. Unknown input data provenance and quality. Automated data retrieval lead to inconsistent results.



AI bias due to generalisation (insufficient representative input data), or unsuitable data collection, processing (cleaning), quality, mislabelling and model design. Synthetic (output) data generated inherits and propagates bias affecting scientific validity.



Data misuse - Using data as input for an AI model that causes harm.



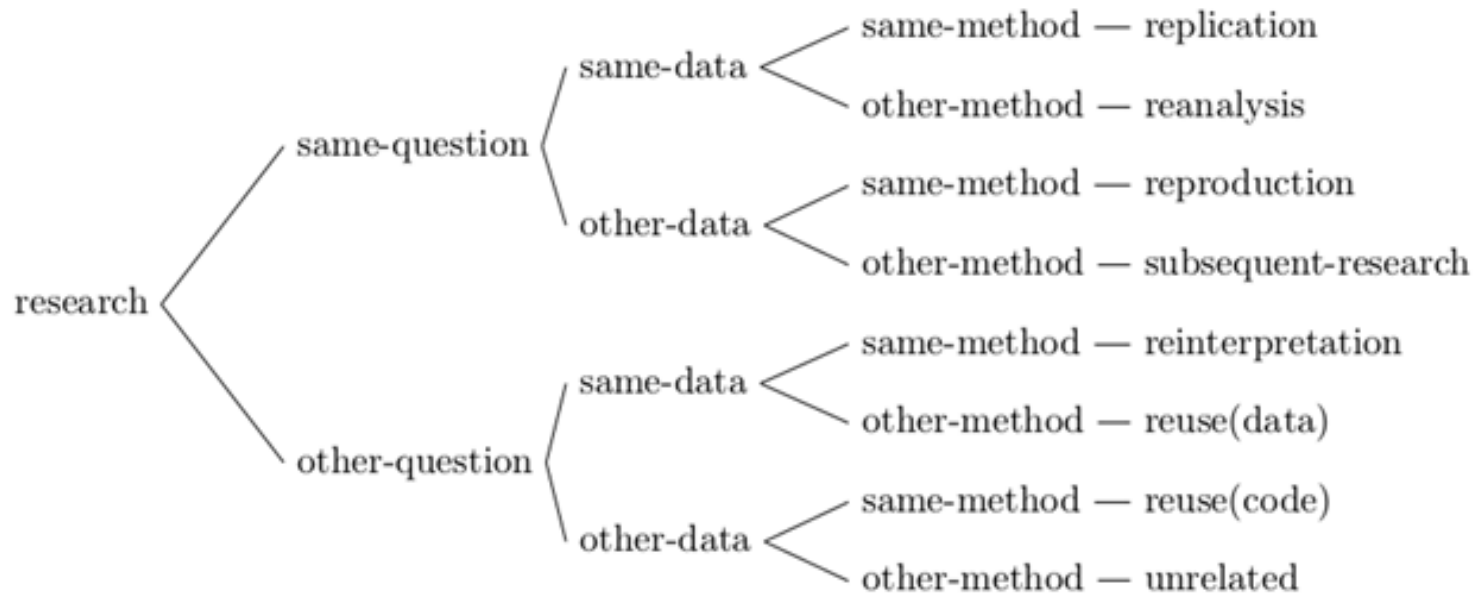
Lack of standards, tools and mechanisms to evaluate data quality and whether datasets are fit for purpose.

INTELLIGENZA ARTIFICIALE

- FUNZIONA SE I DATI SONO BUONI
- PONE QUESTIONI ETICHE

Focus

ASSICURARE IL RIUSO



Van de Sandt et al. 2019

DATA IS A RENEWABLE RESOURCE AS MUCH AS SUN AND WIND.
EVERY 18 MONTHS WE DOUBLE THE AMOUNT OF DATA WE PRODUCE.
INDUSTRIAL AND COMMERCIAL DATA, **85% OF WHICH IS NEVER USED.**
THIS IS NOT SUSTAINABLE. WITHIN THOSE DATA, THERE ARE HIDDEN TREASURES
AND UNTAPPED OPPORTUNITIES FOR BUSINESS AND SOCIETY [Von der Leyen 2020]

...prima di partire per FAIR

Data Intelligence

2020

Issues

Online Early

About ▾

Submit ▾

Volume 2, Issue 1-2

Winter-Spring 2020

January 01 2020

FAIR Principles: Interpretations and Implementation Considerations



< Previous Article

Next Article >

Article Contents

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine D. Kitchin, Tobias Kohn, Jusseli Lehtinen, Barbara Messinger, Peter McQuilton, Natalie Meyers, Annalisa Montuschi, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Barend Mons  , Erik Schultes

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 10–29.

FONDAMENTALE PER
EVITARE ERRORI

- Findability: Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process [14].
- Accessibility: Protocols for retrieving digital resources should be made explicit, for both humans and machines, including well-defined mechanisms to obtain authorization for access to protected data.
- Interoperability: When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity. Similarly, when a digital entity is capable of being processed by an online service, a machine should be capable of automatically detecting this compliance and facilitating the interaction between the data and that tool. This requires that the meaning (semantics) of each participating resource – be they data and/or services service – is clear.
- Reusability: Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource *should* be reused (i.e., is it relevant to the task at-hand?); if a digital resource *can* be reused, and under what conditions (i.e., do I fulfill the conditions of reuse?); and *who to credit* if it is reused.

Scenario

Volume 2, Issue 1-2

Winter-Spring 2020
2020



< Previous Article Next Article >

January 01 2020

The Need of Industry to Go FAIR

Herman van Vlijmen , Albert Mons  , Arne Waalkens , Wo Christine Kirkpatrick , Luiz Olavo Bonino da Silva Santos , Ber Sebastiaan Knijnenburg , Scott Lusher , Rudi Verbeeck , Jean

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 276–284.

https://doi.org/10.1162/dint_a_00050

2. THE VALUE OF FAIR DATA

Research data is one of the most valuable resources we have in the world, as it is the key ingredient to innovation, ultimately leading to societal benefits, like alternative energy options, or treatments of diseases. Every element of data could potentially contain a clue that can lead to an important discovery. However, in industry, much like in academia, research data is rarely leveraged beyond its original intended purpose [2]. This is not only based on deliberate data protection, but also on a lack of findability. That means that making data FAIR in industry, and ensuring interoperability and reusability presents a huge opportunity for industry, but ultimately also for society as a whole.

- RARAMENTE I DATI VENGONO SFRUTTATI AL DI LÀ DEL LORO INTENTO ORIGINALE
- NON SFRUTTATI PERCHÉ NON TROVATI
- MA OGNI DATO HA UN POTENZIALE INNOVATIVO

1. INTRODUCTION AND CONTEXT

2. THE VALUE OF FAIR DATA

3. THE NEED FOR A FAIR PUBLIC PRIVATE PARTNERSHIP (PPP)

4. BENEFITS FOR DATA INTENSIVE INDUSTRY

5. BENEFITS TO FAIR DATA SERVICE PROVIDERS

6. CURRENT LAY OF THE LAND FAIR TOOLING AND SERVICES

7. FAIR DATA AND CERTIFICATION

8. THE PUBLIC PRIVATE PARTNERSHIP AS FAIR TRUSTED PARTY

9. THE FAIR SERVICE PROVIDER CONSORTIUM

Obiettivo

DATI FAIR BY DESIGN

E Giglia, Open Access, ovvero...
Aviano 23 settembre 2015

FAIR principles

To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

TO BE 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.

TO BE 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.

TO BE RE-USABLE:

- R1. meta(data) have 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 their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

Force 11

«ACCESSIBLE»
≠ «OPEN»
= DOVE E A QUALI
CONDIZIONI
I DATI SONO
ACCESSIBILI

FAIR/Open

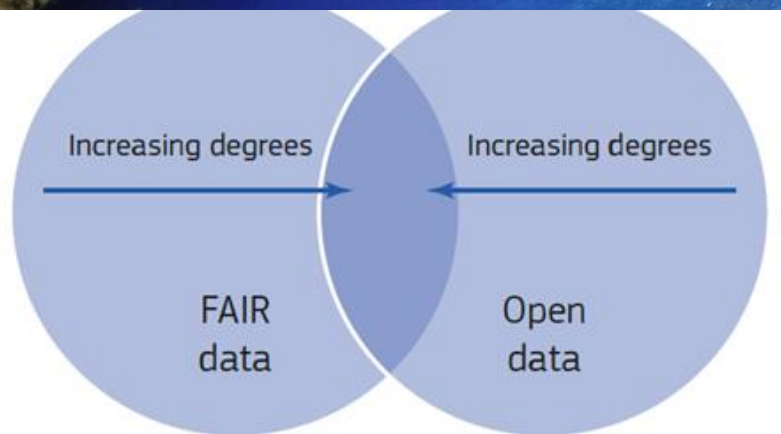


Figure 4. The relationship between FAIR and Open



A TENDERE, I DUE INSIEMI SARANNO SEMPRE PIÙ SOVRAPPOSTI. MA ESISTERANNO SEMPRE DATI PERFETTAMENTE FAIR CHE NON POSSONO ESSERE OPEN

FAIR is not a standard: The FAIR guiding principles are sometimes incorrectly referred to as a 'standard', even though the original publication explicitly states they are not [25]. The guiding principles allow many different approaches to rendering data and services Findable, Accessible, Interoperable, to serve the ultimate goal: the reuse of valuable research objects. Standards are prescriptive, while guidelines are permissive. We suggest that a variety of valuable standards can and should be developed, each of which is guided by the FAIR Principles. FAIR simply describes the qualities or behaviours required of data resources to achieve – possibly incrementally – their optimal discovery and scholarly reuse.

FAIR is not equal to RDF, Linked Data, or the Semantic Web The reference article Scientific Data [25] emphasises the machine-actionability of data and metadata. This implies (in fact, requires) that resources that wish to maximally fulfil the FAIR guidelines must utilise a widely-accepted machine-readable framework for data and knowledge.

FAIR is not just about humans being able to find, access, reformat and finally reuse

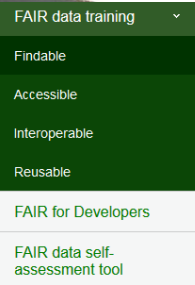
data: The official press release for the authors' position clearly: "The research data publication autonomously, and in accordance with the FAIR Principles. Computers are now an integral part of research. In recent surveys, the time reported for researchers dealing with discovering and reusing data has been pegged at 80% [19]. Were this time spent dealing with FAIR data and services, it would be a great deal better than it is today. The avoidance of time-wasting is a key principle of data stewardship. To serve this potential, data discovery and services should be actionable wherever possible.

FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired by Open Science, explicitly and

FAIR refers to a set of principles, focused on ensuring that research objects are reusable, and actually will be reused, and so become as valuable as is possible. They deliberately do not specify technical requirements, but are a set of guiding principles that provide for a continuum of increasing reusability, via many different implementations. They describe characteristics and aspirations for systems and services to support the creation of valuable research outputs that could then be rigorously evaluated and extensively reused, with appropriate credit, to the benefit of both creator and user.

- PRINCIPI, NON STANDARD [VANNO IMPLEMENTATI]
 - NON SOLO PER UMANI
- DIVERSO DA LINKED DATA, RDF...
 - DIVERSO DA «OPEN»

FAIR in sintes



F1. (meta)data are assigned a globally unique and eternally

There are many resources created by the ARDC on the topic of [metadata](#)

- [Metadata guide](#)
- [Data versioning](#)

The ARDC has information on persistent identifiers on three different levels

- [Persistent identifiers: awareness level](#)
- [Persistent identifiers: working level](#)
- [Persistent identifiers: expert level](#)

It is also a provider of services for minting persistent identifiers of many different types (e.g. the following are some of the data being identified):

- [Digital Object Identifier \(DOI\) System for research data](#)
- [Handle minting Service \(Identify My Data\)](#)
- [International Geo Sample Numbers \(IGSN\)](#)

Complementary to the assignment of persistent identifiers is their proper



I miti su FAIR

Let's talk about FAIR data...

"FAIR data do not have to be openly available. FAIR data can be kept private, but information about the data should be shared"



www.vidensportal.deic.dk/FAIR

FAIR data are: Findable, Accessible, Interoperable and Reusable.
Making your data FAIR means maximizing the project's output, increasing your impact and enhancing your recognition as a researcher.

©2018 by DeIC / Deff / DTU / CBS / AAU / KU / Rigshospitalet / Det Kongelige Bibliotek
Public is made available under a Creative Commons 4.0 license International

Let's talk about FAIR data...

"The FAIR principles for research data are not a standard. They are a set of common guidelines to make any data more valuable"



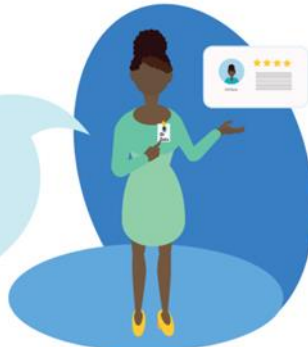
www.vidensportal.deic.dk/FAIR

FAIR data are: Findable, Accessible, Interoperable and Reusable.
Making your data FAIR means maximizing the project's output, increasing your impact and enhancing your recognition as a researcher.

©2018 by DeIC / Deff / DTU / CBS / AAU / KU / Rigshospitalet / Det Kongelige Bibliotek

Let's talk about FAIR data...

"With FAIR data, you can improve the impact of your publications, and make sure that you will get all the credit you deserve"



www.vidensportal.deic.dk/FAIR

FAIR data are: Findable, Accessible, Interoperable and Reusable.
Making your data FAIR means maximizing the project's output, increasing your impact and enhancing your recognition as a researcher.

©2018 by DeIC / Deff / DTU / CBS / AAU / KU / Rigshospitalet / Det Kongelige Bibliotek
Public is made available under a Creative Commons 4.0 license International

EOSC Podcast Special
Making Open Science FAIR For Researchers



Sarah Jones
EOSC Executive Board - FAIR WG Chair



07 Sep 2020

Sept.7 2020

News EOSC Podcast Special: Making Open Science FAIR For Researchers

EOSCsecretariat.eu is launching a series of podcasts, focusing on different topics surrounding EOSC. The first episode featured EOSC FAIR Working Group Chair, **Sarah Jones**, discussing Open Science, FAIR Data, and how and when researchers should be engaged. The conversation is fascinating as she pulls back the curtain on what EOSC is looking to achieve, explaining **what Open Science and FAIR Data mean**. [Read more](#)

CONTINUE READING

FAIR myths

FAIR software

The FAIR4RS Principles are:

F: Software, and its associated metadata, is easy for both humans and machines to find.

F1. Software is assigned a globally unique and persistent identifier.

- F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.
- F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

A: Software, and its metadata, is retrievable via standardized protocols.

A1. Software is retrievable by its 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 software is no longer available.

I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

R1. Software is described with a plurality of accurate and relevant attributes.

- R1.1. Software is given a clear and accessible license.
- R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

R3. Software meets domain-relevant community standards.

Table 1: The FAIR Principles for Research Software

FAIR software

Practical guide to Software Management Plans 2022

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
User documentation	How will your software be documented? Please provide a link to the documentation if available.
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available.
Software licencing and compatibility	What licence will you give your software? How will you check that it respects the licences and dependencies it uses?

Table 1. Core requirements of an SMP and examples of questions at a low level of software management.

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
Repository	How will you make your software publicly available? If you do not plan to make it publicly available you should provide a justification.
User documentation	How will your software be documented for users? Please provide a link to the documentation if available. How will you document your software's contribution guidelines and governance structure?
Software licencing and compatibility	What licence will you give your software? How will you check that it respects the licences and dependencies it uses?
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available.
Citation	How will users of your software be able to cite your software? Please provide a link to your software citation file (CFF) if available.
Developer documentation	How will your software be documented for future developers?
Testing	How will your software be tested? Please provide a link to the (automated) testing results if available.
Software Engineering quality	Do you follow specific software quality standards? If yes, which ones?
Packaging	How will your software be packaged and distributed? Please provide a link to your packaging information (e.g. entry in a public registry, if available).
Maintenance	How do you plan to procure long term maintenance of your software?

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
Repository	How will you make your software publicly available? If you do not plan to make it publicly available, you should provide a justification.
User documentation	How will your software be documented for users? Please provide a link to the documentation if available. How will you document your software's contribution guidelines and governance structure?
Software licencing and compatibility	What type of licence will your software have? How will you check that it respects the licences of libraries and dependencies it uses?
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available. This documentation should include a complete and unambiguous description of dependencies to other software, datasets, and hardware.
Citation	How will users of your software be able to cite your software? Please provide a link to your software citation file (CFF) if available.
Developer documentation	How will your software be documented for future developers?
Testing	How will your software be tested? Please provide

...E SERVONO
SOFTWARE
MANAGEMENT
PLANS!

Dati FAIR e citizen science

STANDARD PER
RICERCHE CON
CITIZEN SCIENCE

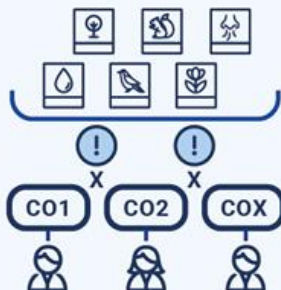
STApplus

Why should you use it?

STApplus aims to standardise citizen science data and make it accessible, interoperable and reusable among different citizen observatories (COs) and services. In particular, STApplus is an extended data model for the SensorThings API*.

01 THERE IS A HUGE AMOUNT OF CITIZEN SCIENCE DATA STORED IN THE CLOUD

The problem is that it is usually extremely difficult to exchange data among different citizen science platforms - known as citizen observatories (COs) - or to reuse their data, as they don't speak a 'common language'.

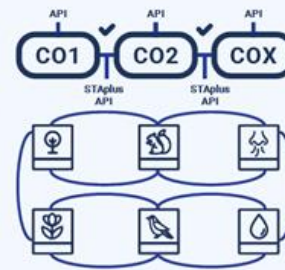


02 STApplus STANDARDISES THE DATA STRUCTURE

To tackle this 'communication problem', STApplus has extended the SensorThings API data structure to apply it to citizen science. In particular, it has added these aspects: ownership of the observations, citizen science project or campaign, link between observations-author-project.



03 COs USING STApplus ARE ABLE TO EXCHANGE AND REUSE THEIR DATA



04 TO USE IT, YOU JUST HAVE TO ADAPT YOUR DATA TO THE STApplus DATA STRUCTURE

Even if you already have an API, you can add STApplus API and have both.

05 A MAP VIEWER: AN EXAMPLE OF USE

STApplus has great potential to craft data requests from the platforms that use it. This way, if you want to create, for example, a map viewer to show certain citizen science data by location, you can do it easily!



STApplus: SCENARIOS IN Cos4Cloud

Natusfera & Pl@ntNet, two of the biodiversity citizen observatories participating in Cos4Cloud, are already using this API.



Nov. 20, 2018

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

TURNING FAIR INTO



Define

Concepts for FAIR implementation

Rec. 1: Define FAIR for implementation

Rec. 2: Implement a Model for FAIR Digital Objects

Rec. 3: Develop components of a FAIR ecosystem

Rec. 16: Apply FAIR broadly

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

Implement

FAIR culture

Rec. 4: Develop Interoperability frameworks

Rec. 5: Ensure data management via DMPs

Rec. 6: Recognise & reward FAIR data & stewardship

Rec. 18: Cost data management

Rec. 19: Select and prioritise FAIR digital objects

Rec. 20: Deposit in Trusted Digital Repositories

Rec. 21: Incentivise reuse of FAIR outputs

FAIR ecosystem

Rec. 7: Support semantic technologies

Rec. 8: Facilitate automated processing

Rec. 9: Certify FAIR services

Rec. 22: Use information held in DMPs

Rec. 23: Develop components to meet research needs

Rec. 24: Incentivise research infrastructures to support FAIR data

Skills for FAIR

Rec. 10: Professionalise data science & stewardship roles

Rec. 11: Implement curriculum frameworks and training

Above line = priority recommendations

Below line = supporting recommendations

Embed and sustain

Incentives and metrics for FAIR data and services

Rec. 12: Develop metrics for FAIR Digital Objects

Rec. 13: Develop metrics to certify FAIR services

Rec. 25: Implement and monitor metrics

Rec. 26: Support data citation and next generation metrics

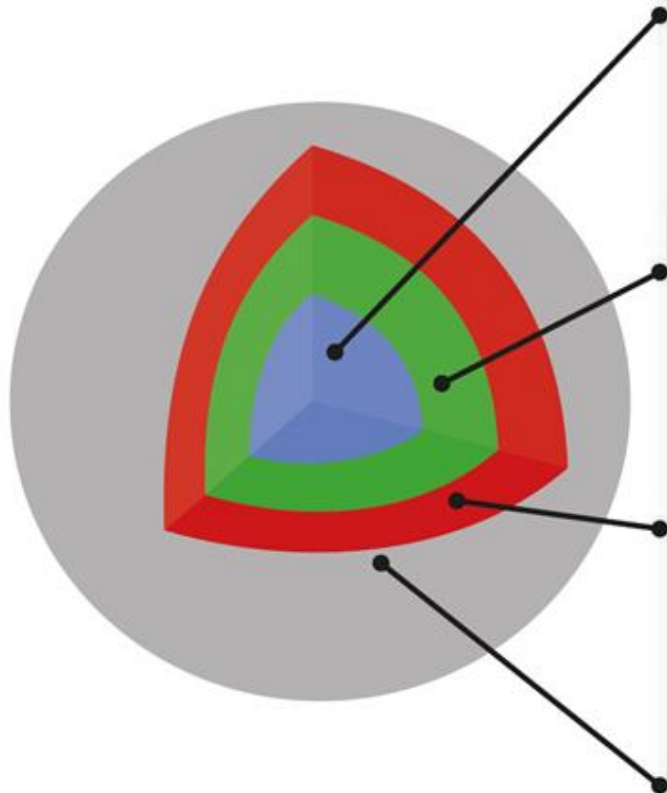
Investment in FAIR

Rec. 14: Provide strategic and coordinated funding

Rec. 15: Provide sustainable funding

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

Oggetto FAIR ideale



DIGITAL OBJECT

Data, code and other research outputs

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

IDENTIFIERS

Persistent and unique (PIDs)

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

STANDARDS & CODE

Open, documented formats

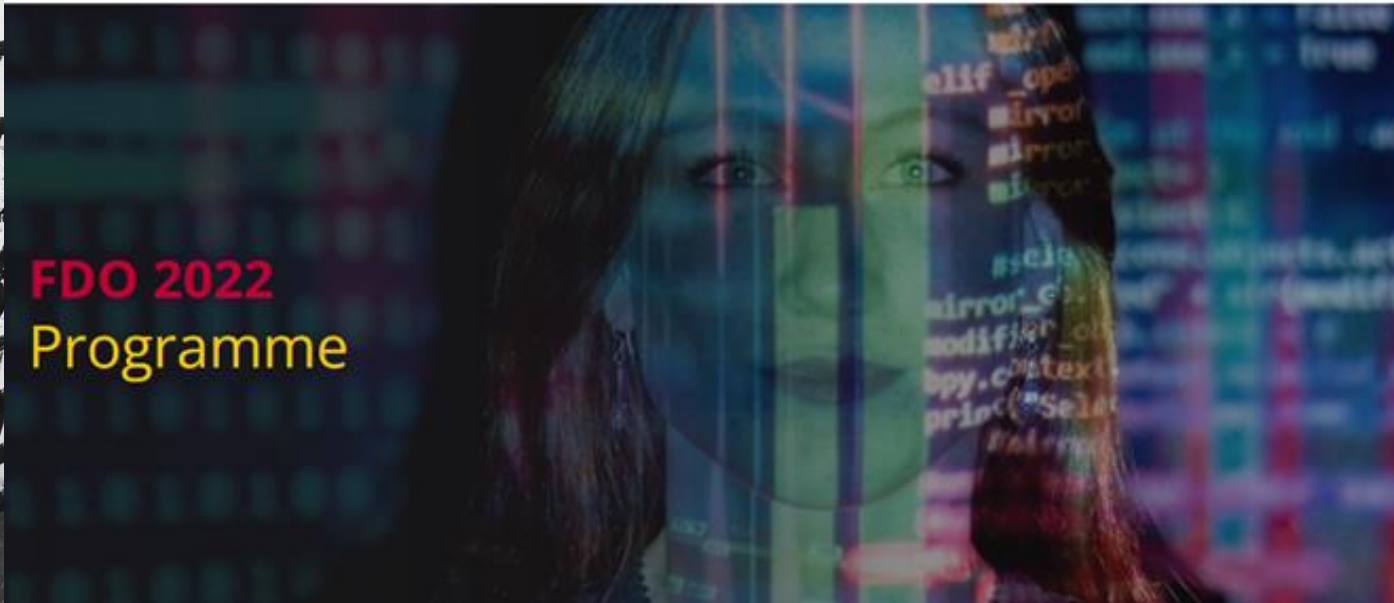
Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to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

METADATA

Contextual documentation

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

1st International Conference on FAIR Digital Objects	<u>FDO 1st 2022</u>	Organisation	Programme	Partners	Logistical Information
---	----------------------------	----------------------------	-------------------------	------------------------	--



FDO 2022
Programme

FAIR: technology VS domain



Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

FAIR RICHIEDE AZIONI
DAI RICERCATORI E DAI
REPOSITORIES
...MA SONO
STRETTAMENTE
INTERCONNESSE

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure
- A2. metadata are accessible, even when the data are no longer available

To be 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

To be 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

The screenshot shows the Zenodo upload interface. At the top, there's a row of icons for different upload types: Publication (selected), Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, Physical object, and Other. Below this, the 'Publication type' dropdown is set to 'Journal article'. The 'Basic information' section includes a 'Digital Object Identifier' field with the example 'e.g. 10.1234/foo-bar', a 'Publication date' field set to '2021-05-26', a 'Title' field, an 'Authors' field with a family name and given names, and a 'Description' field with a rich text editor. There are also links for 'Reserve DOI' and 'Add another author'.

IN ZENODO LA
DESCRIZIONE
DELL'OGGETTO È
GIÀ IMPOSTATA

FAIR for dummies

COSA DEVE FARE IL
RICERCATORE

COSA FA IL
REPOSITORY

Explanation of the [FAIR data principles](#) ²⁰¹⁹

Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Principle	In other words	Researcher's responsibility	Requirements to be fulfilled by the repository	
To be findable: Data and metadata should be easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.	F1. (meta)data are assigned a globally unique and persistent identifier	Each data set is assigned a globally unique and persistent identifier (PID), for example a DOI , ARK , RRID ... These identifiers allow to find, cite and track (meta)data.	Ensure that each data set is assigned a globally unique and persistent identifier. Certain repositories automatically assign identifiers to data sets as a service. If not, researchers must obtain a PID via a PID registration service.	A repository needs to have a predictable way to assign a PID to each component of a dataset (e.g. each file or nanopublication), in order to be able to include these identifiers into the corresponding metadata before the submission.
	F2. data are described with rich metadata (defined by R1 below)	Each data set is thoroughly (see below, in R1) described: these metadata document how the data was generated, under what term (license) and how it can be (re)used, and provide the necessary context for proper interpretation. This information needs to be machine-readable.	Fully document each data set in the metadata, which may include descriptive information about the context, quality and condition, or characteristics of the data. Another researcher in any field, or their computer, should be able to properly understand the nature of your dataset. Be as generous as possible with your metadata (see R1).	Allow researchers to upload metadata for each data set.
	F3. metadata clearly and explicitly include the identifier of the data it describes	The metadata and the data set they describe are separate files. The association between a metadata file and the data set is obvious thanks to the mention of the data set's PID in the metadata.	Make sure that the metadata contains the data set's PID.	Allow researchers to upload metadata for each data set.
	F4. (meta)data are registered or indexed in a searchable resource	Metadata are used to build easily searchable indexes of data sets. These resources will allow to search for existing data sets similarly to searching for a book in a library.	Provide detailed and complete metadata for each data set (see F2).	Request and store part of the metadata in a structured way, for example by providing a form with specific fields to be completed or by providing an XML schema to be used by the researchers. For example the storing of PID's, author names, disciplines, etc. will facilitate the creation of indexes. However, it must remain possible to provide arbitrary metadata in addition.

FAIR – how to



FAIR Data Collective
Making practical FAIR data solutions.

2 followers · World · <https://dk.linkedin.com/company/fair-...>

[Overview](#) [Repositories](#) 20 [Projects](#) [Packages](#) [People](#)

Popular repositories

excel2rdf-template

This resource represents a template Github repository for the workflow automatization described in <https://excel2rdf.readthedocs.io>.

☆ 6 🍴 8

M4M19-subjects-vocabulary

This repository is configured to run sheet2rdf workflow targeted of subject controlled vocabulary.

Python ☆ 2



OSFHOME ▾

[M4M.20 \(FAIRware Psychology\)](#) [Files](#) [Wiki](#) [Analytics](#) [Registrations](#)

<https://osf.io/qcht7/>

50.9MB Public 0

M4M.20 (FAIRware Psychology)

Contributors: Erik Anthony Schultes, Barbara Magagna, Nikola Vasiljevic, Evie Vergauwe, Elena Giglia, Gianvito Laera, Christina Heemskerk, Chiara Scarampi
Date created: 2022-06-01 08:55 PM | Last Updated: 2022-08-20 02:40 PM
Identifier: DOI 10.17605/OSF.IO/QCHT7
Category: Project
License: CC-BY Attribution 4.0 International

Wiki

June 20 2022
Preworkshop homework <https://bit.ly/M4M-20-homework>
Common Notes <http://bit.ly/M4M-20>
User Guide for completing the dataset metadata <https://osf.io/kt7y2/>
[Datasets overview](#)
Video recording <https://bit.ly/M4M-20-recording>
[Read More](#)

Citation

Recent Activity

Erik Anthony Schultes created external identifier(s) doi:10.17605/OSF.IO/QCHT7 on M4M.20 (FAIRware Psychology) 2022-08-20 02:40 PM

Erik Anthony Schultes updated wiki page Home to version 6 of M4M.20 (FAIRware Psychology) 2022-07-08 01:10 AM

FAIR Implementation Profile

FAIR Implementation Profile

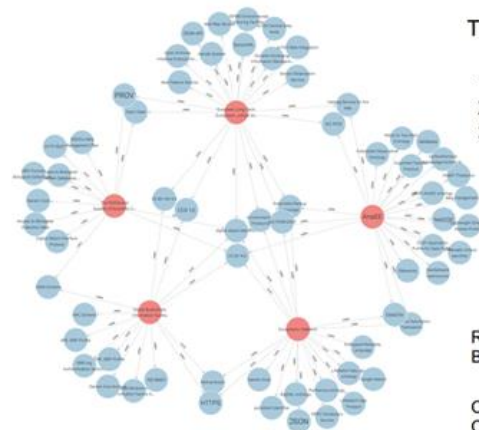
FAIR principle	Question	FAIR enabling resource types
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type
F2	Which metadata schemas do you use for findability?	Metadata schema
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism
F4	In which search engines are your metadata records indexed?	Search engines
F4	In which search engines are your datasets indexed?	Search engines
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique
A2	Which metadata longevity plan do you use?	Metadata longevity
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema
I3	Which models, schema(s) do you use for your datasets?	Data schema
R1.1	Which usage license do you use for your metadata records?	Data usage license
R1.1	Which usage license do you use for your datasets?	Data usage license
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	



Biodiversity-ecosystem sciences

Top 3 shared resources:

1. Digital Object identifier (all)
2. CC-BY-4.0 (4)
3. Extensible Markup Language (3)



Red: communities
Blue: resources

CC-BY Kristina Hettne, Leiden University Libraries
Created with GraphDB

Slides courtesy of Erik Schultes Go FAIR
OSF | HS.3PFF.Oct 2021.pdf

FAIR Implementation profiles

[International Conference on Conceptual Modeling](#)

ER 2020: [Advances in Conceptual Modeling](#) pp 138-147 | [Cite as](#)

2020

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence

Authors

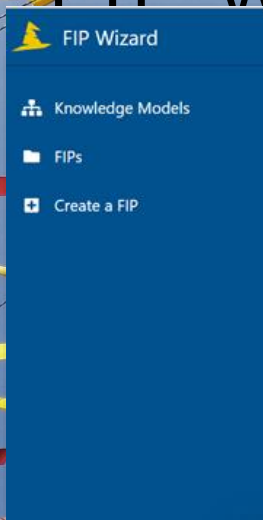
Authors and affiliations

Erik Schultes, Barbara Magagna , Kristina Maria Hettne, Robert Pergl, Marek Suchánek, Tobias Kuhn

Powerful incentives are driving the adoption of FAIR practices among a broad cross-section of stakeholders. This adoption process must factor in numerous considerations regarding the use of both domain-specific and infrastructural resources. These considerations must be made for each of the FAIR Guiding Principles and include supra-domain objectives such as the maximum reuse of existing resources (i.e., minimised reinvention of the wheel) or maximum interoperation with existing FAIR data and services. Despite the complexity of this task, it is likely that the majority of the decisions will be repeated across communities and that communities can expedite their own FAIR adoption process by judiciously reusing the implementation choices already made by others. To leverage these redundancies and accelerate convergence onto widespread reuse of FAIR implementations, we have developed the concept of FAIR Implementation Profile (FIP) that captures the comprehensive set of implementation choices made at the discretion of individual communities of practice. The collection of community-specific FIPs compose an online resource called the FIP Convergence Matrix which can be used to track the evolving landscape of FAIR implementations and inform optimisation around reuse and interoperation. Ready-made and well-tested FIPs created by trusted communities will find widespread reuse among other communities and could vastly accelerate decision making on well-informed implementations of the FAIR Principles within and particularly between domains.

CREARE FAIR
IMPLEMENTATION
PROFILES
RIUSABILI ANCHE
DA ALTRI NELLA
COMUNITÀ
DISCIPLINARE

FIP wizard



FIP wizard



Welcome to the FIP Wizard!



Social Science Survey Research_V1

 Questionnaire  Metrics  Preview  Documents

View 

Current Phase

Before Submitting the Proposal 

Chapters

Background: The FAIR Implementation Profile and I. FAIR Implementation Community 

II. FAIR Implementation Community 

III. Findability 25

IV. Accessibility 12

V. Interoperability 21

VI. Reusability 14

I. Background: The FAIR Implementation Profile and FAIR Implementation Community

The FAIR Implementation Profile (FIP) is a collection of FAIR implementation choices made by a FAIR Implementation Community for each of the FAIR Principles. Community-specific FIPs are themselves captured as FAIR datasets and are made openly available to other communities for reuse. To create a FIP, the data steward of a community needs to fill out this questionnaire where the implementation choices are recorded as resources. The questionnaire is structured as follows: the first section is about the FAIR Implementation Community, which is then followed by a number of questions per FAIR principle. The answer to each of the questions should be a FAIR-Enabling Resource. The questionnaire offers to look up the resource in Nanobench. If the resource cannot be found in any of these applications, there is an option at the end of the questionnaire to register a FAIR-Enabling Resource as a nanopublication in Nanobench. The resource will get a PURL which can then directly be used when further filling out the questionnaire. When the questionnaire is filled in, the FIP is considered to be the implementation of R1.3, which is why there is no separate question addressing this subprinciple.

[FAIR all'opera]

I DATI NON SI
MUOVONO, GLI
ALGORITMI LI TROVANO
I DATI STANNO DOVE
SONO MA POSSONO
SERVIRE L'INTERA
COMUNITÀ

Making data and metadata FAIR ensures that these data are discoverable on the Internet of FAIR Data and Services. Central to this approach is the establishment of FAIR Data Points (FDPs), for COVID-19 relevant digital data objects. Opening up FAIR (meta)data by publishing them on a FDP allows algorithms to search these (meta)data, looking for patterns. The Internet of FAIR Data and Services is a distributed data discovery network; data are NOT moved, but algorithms going over the internet can find the data.

Go FAIR newsletter First data-visiting of data safely held in FAIR Data Points

POSTED ON 1 OCTOBER 2020

After the successful deployment of seven machine-actionable FAIR Data Points in Africa, a test to execute a machine-based querying of FAIR Data Points across continents between Leiden University Medical Center and Kampala International University was successful.

Mirjam van Reisen (IN Coordinator of **IN-Africa** and one of the coordinators of the **Ambassadors IN**) said:

Today we have shown the first data-visiting of data safely held in FAIR Data Points within the hospitals. Data is handled in accordance with the regulatory frameworks that apply in each location. This is a major step forward to ensure that data stays where it belongs but can serve the global health community to find solutions to the pandemic. It is also proof that the Internet of FAIR Data and Services can be realised.

VODAN AFRICA
VODAN Africa&Asia

The VODAN Africa & Asia

Fighting the COVID-19 with FAIR Data

The initiative is funded by the Philips Foundation, Google and the FMO to enable distributed access to the critical data needed from Africa and the rest of the world to fight and contain the COVID-19 pandemic

OAI12 – The Geneva Workshop on Innovations in Scholarly Communication

6-10 September 2021
Virtual workshop

OAI12

Search...

Open Science – its impact and potential as a driver for radical change

OAI12 – The Geneva Workshop on Innovations in Scholarly Communication will be held virtually from 6 – 10 September 2021.

More details about the workshop can be found at the website <https://oai.events/>.

The main themes of this edition are:

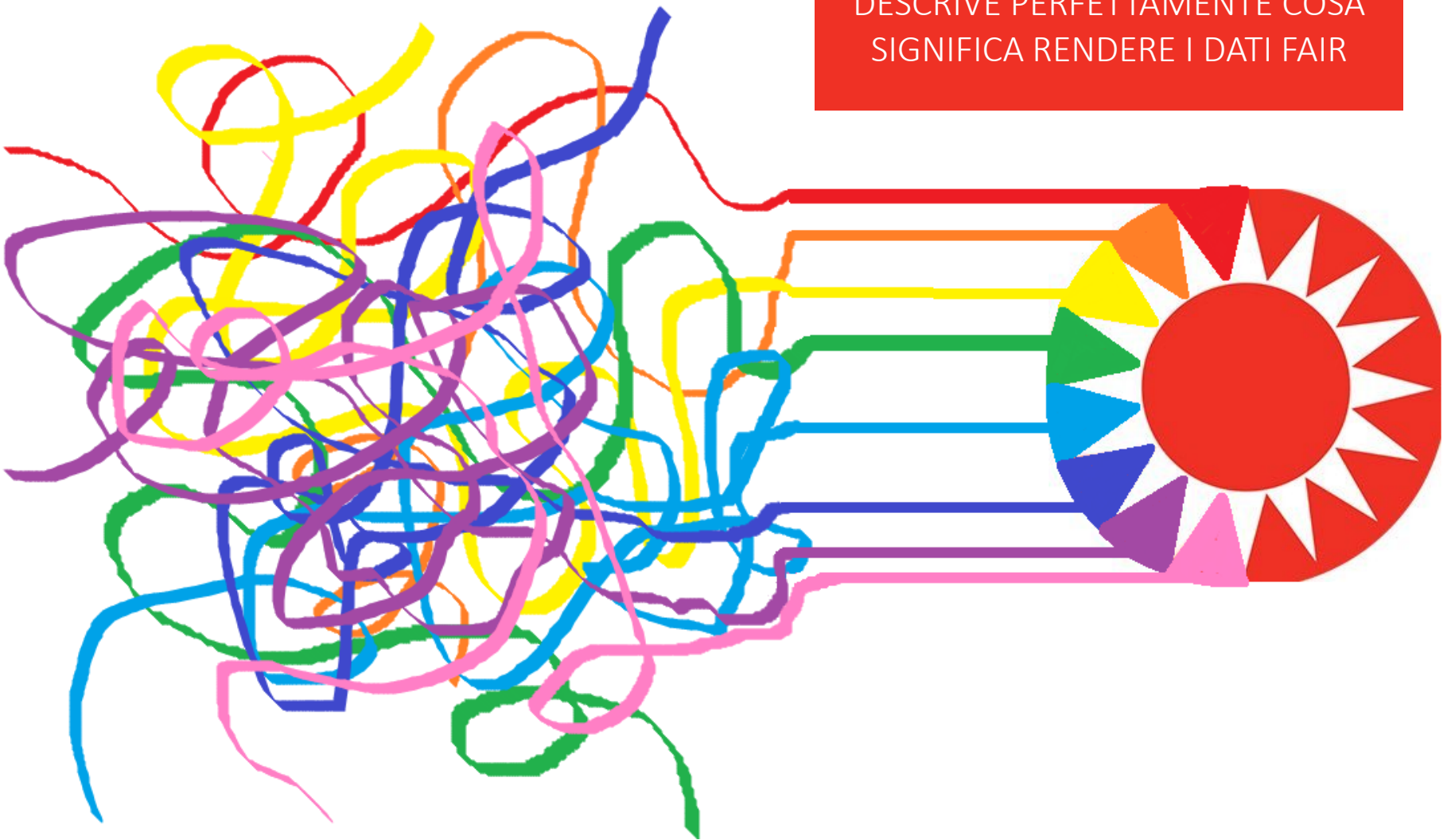
- Scholarly publishing
 - (Call for proposals until 6 April)
- Digital research data
- Reproducibility and research integrity
- Diversity, inclusivity and collaboration
- The future of open science
 - (Call for papers is closed. We thank all)

PRESENTATO A
OAI12



Come fare

QUESTA IMMAGINE, CREATA DA
Noa, Anna, Lilian e Charlotte
DESCRIVE PERFETTAMENTE COSA
SIGNIFICA RENDERE I DATI FAIR



FAIR aware

- DOMANDE
- VERIFICA INTENZIONI
- VERIFICA CONOSCENZA
- FORNISCE INFORMAZIONI



Let's assume you have research data almost ready for uploading to a repository: do you already know how you and the repository can work together to make the data as findable, accessible, interoperable and reusable (FAIR) as possible? By guiding you through the assessment process, the FAIR-Aware tool can help you to better understand the FAIR Principles and how making data FAIR can increase the potential value and impact of your data.

FAIR-Aware is an disciplinary-agnostic online tool developed by the FAIRsFAIR project. Different scientific communities can adapt it to their own use. You should, however, have a target dataset in mind to be able to answer the questions and complete the assessment.

unique persistent and resolvable identifier when deposited with a data repository?

What does this mean?

A **persistent identifier** is a long-lasting reference to a resource. The **data(set)** you deposit in a **data repository** should be assigned a globally unique, persistent and resolvable identifier (PID) so that both humans and machines can find it. Persistent identifiers are maintained and governed so that they remain stable and direct the users to the same relevant object consistently over time. Examples of PIDs include Digital Object Identifier (DOI), Handle, and Archival Resource Key (ARK).

Why is this important?

If your data(set) or metadata does not have a PID, you run the risk of "link rot" (also known as "link death"). When your data(set) or metadata is moved, updated to a new version, or deleted, older hyperlinks will no longer refer to an active page. Without a PID, others will not be able to find or reuse your data(set) or metadata in the long-term.

How to do this?

When you upload your data(set) or metadata to a data repository, the data repository (or other service providers) usually assigns a PID. Repositories ensure that the identifier continues to point to the same data or metadata, according to access terms and conditions you specified.

There are many different types of PIDs, each with their own advantages, disadvantages, and disciplines they are typically used in. Generally speaking, the data repository will have thought about these aspects before deciding which PID type to use. In case you have to choose the PID type yourself, you can visit the Knowledge Hub on the PID Forum for guidance. Some disciplines or organisations also provide tools to help you make this choice, see for example this Persistent Identifier Guide for cultural heritage researchers. Once you have chosen a PID type, you can search for data repositories providing that specific PID in registries such as Re3data or FAIRsharing (see related databases).

Not all data you produce during your research will need a PID. In general, those that underpin published findings or have longer term value are worth assigning a PID. If in doubt about which data should be allocated a PID, speak to your local research data management support team or the data repository.

FINDABLE

1. Are you aware that a data(set) should be assigned a global persistent and resolvable identifier when deposited with a data repository?
2. Are you aware that when you deposit a data(set) in a data repository you will need to provide discovery metadata in order to make your data(set) findable, understandable and reusable to others?
3. Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans?

ACCESSIBLE

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which your data(set) can be reused?
5. Are you aware that metadata should remain available over time if the data(set) is no longer accessible?

...FAIR

AUTO-
VALUTAZIONE SU
QUANTO L'ENTE DI
RICERCA «ABILITA»
I DATI FAIR

DO I-PASS FOR FAIR?

Oct. 2020



Self assessment tool to
measure the FAIR-ness
of an organization

BEGINNER

INTERMEDIATE

ADVANCED

DOES YOUR ORGANIZATION...

- 1** **POLICY**
...have a FAIR research data policy?
- 2** **SERVICES**
...have a DCC which provides services to
allow research(ers) to comply with FAIR?
- 3** **SKILLS**
...acknowledge that FAIR capacity building
requires specific roles and skills?
- 4** **INCENTIVES**
...have incentives for FAIR data?
- 5** **ADOPTION**
...have adoption of FAIR?

...FAIR per enti di ricerca

DO I-PASS FOR FAIR?

Oct. 2020



Policy: Does your organization (institute / university (of applied sciences)) have a FAIR research data policy?

1. Are the FAIR principles explicitly mentioned in the data (or research data) policy of your organization?	The FAIR principles are not explicitly mentioned, but there is a reference to sustainable and long term storage.	The F and A principle are explicitly mentioned.	All FAIR principles are explicitly mentioned (or there is an explicit reference to the overarching Open Science concept).	☐ Beginner ☐ Intermediate ☐ Advanced
2. Is the general research data policy translated into faculty/institutes specific policies?	Less than 20% of the faculties or sub-	20-80% of the faculties or sub-units have a specific policy.	More than 80% of the faculties or sub-units have	☐ Beginner ☐ Intermediate

Services: Does your organization have a (virtual) DCC which provides services (including infrastructure) to allow comply with FAIR?

3. Is the general research data policy supported by dedicated service units (e.g. DCC) ?	5. Which services does your organization provide in order for researchers to comply with the F principles	We provide or refer to a service to deliver a PID for a data set	We provide or refer to a service for PID and adding metadata (including reference to the dataset).	On top of adding PIDs and metadata, we provide or refer to a service to make the data findable through indexed resources.	☐ Beginner ☐ Intermediate ☐ Advanced
	6. Which services does your organization provide in order for researchers to comply with the A principles?	We provide or refer to a service with a standard and open communication protocol as to access the data.	We provide or refer to a service to enable restricted access to data and meta data (authentication protocol)	Our organization as an archiving policy, both on data and meta data.	☐ Beginner ☐ Intermediate ☐ Advanced
	7. Which services does your organization provide in order for researchers to comply with the R principles?	We provide or refer to a service that enriches the dataset metadata, referencing to other datasets and articles (I3).	We provide or refer to a service to attribute (meta)data that use a formal, accessible, shared, and broadly applicable language for knowledge representation.	We provide or refer to a service to attribute (meta)data that use vocabularies that follow FAIR principles in order to facilitate machine readability.	☐ Beginner ☐ Intermediate ☐ Advanced
4. Which means and channels used to communicate the research data policy and services to the researchers	8. Which services your organization provide in order for researchers to comply with the R principles?	We provide advice on licenses and provide or refer a service to add licenses in the metadata	We monitor the reuse of our data, both within the institute as well as outside	We perform quality control on the (meta)data, e.g. adequate read me file as to assure that somebody else can actually reuse the data. We assure that (meta)data meet domain-relevant community standards	☐ Beginner ☐ Intermediate ☐ Advanced ☐ Not yet

Supporto / Come essere FAIR

zenodo

Search Upload Communities

January 11, 2022 **2022** Book Open Access

D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions

Engelhardt, Claudia; Biernacka, Katarzyna; Coffey, Aoife; Cornet, Ronald; Danciu, Alina; Demchenko, Yuri; Downes, Stephen; Erdmann, Christopher; Garbuglia, Federica; Germer, Kerstin; Helbig, Kerstin; Hellström, Margareta; Hettne, Kristina; Hibbert, Dawn; Jetten, Mijke; Karimova, Yulia; Kryger Hansen, Karsten; Kuusniemi, Mari Elisa; Letizia, Viviana; McCutcheon, Valerie; McGillivray, Barbara; Ostrop, Jenny; Petersen, Britta; Petrus, Ana; Reichmann, Stefan; Rettberg, Najla; Reverté, Carmen; Rochlin, Nick; Saenen, Bregt; Schmidt, Birgit; Scholten, Jolien; Shanahan, Hugh; Straube, Armin; Van den Eynden, Veerle; Vandendorpe, Justine; Venkataram, Shanmugasundaram; Wiljes, Cord; Wuttke, Ulrike; Yeomans, Joanne; Zhou, Biru

5 – FAIR lesson plans

6 – Implementing FAIR

6.1 Introduction

6.2 Getting to FAIR institutional policies

6.3 Data management planning

6.4 Data processing and documentation

6.5 Support infrastructure

6.6 Data publication

6.7 Data reuse

FAIR cookbook

Search this book...

FAIR Cookbook

FOREWORD

Introduction

Ethical values of FAIR

Glossary

RECIPES

Findability

Accessibility

Interoperability

FAIR Cookbook

FAIR cookbook

Created by researchers and data managers professionals, the FAIR Cookbook is an online resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable (FAIR).

Turning FAIR into practice

The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. However, the FAIR Principles are aspirational and generic. The FAIR Cookbook guides *researchers* and *data stewards* of the Life Science domain in their FAIRification journey; and also provides *policy makers* and *trainers* with practical examples to recommend in their guidance and use in their educational material.

HOW TO FAIR

How to FAIR

What is FAIR Why FAIR How to FAIR About Quiz

A deep dive into FAIR data

This website will take you on a deep dive into the subject matter of FAIR research data. Over the course of

HOW TO FAIR

What is FAIR Why FAIR How to FAIR About Quiz

18 min read Documentation	12 min read File formats	20 min read Metadata
10 min read Access to data	7 min read Persistent identifiers	5 min read Data licences

FAIR toolkit

Practical Support for FAIR Data

An overview of how the FAIR Toolkit provides practical support for implementation of FAIR data management through numerous use cases from industry and relevant tools, training and change methods.

Practical Support

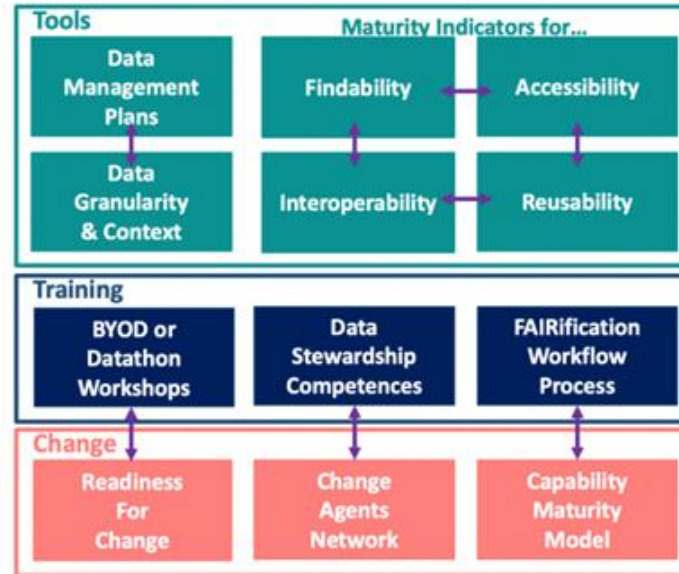
FAIR Toolkit

The FAIR Toolkit is designed to provide support for management of the FAIR data life cycle as illustrated in Figure 1 below. It places emphasis on the practical aspects of FAIR data management through the leverage of existing resources that are most relevant to the needs of Life Science industry.



FAIR TOOLKIT CON
STRUMENTI, CASI
D'USO,
FORMAZIONE

FAIR Toolkit methods



TOOLS

Data Management Plans

A Data Management Plan documents the specific attributes expected for your FAIR objectives.

- Prepare the Data Management Plan as early as possible

[Find out more >](#)

Findability Maturity Indicators

Find out how to apply the FAIR Maturity Indicators to measure the FINDABILITY of the data and metadata.

- Findability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

[Find out more >](#)

Accessibility Maturity Indicators

Learn how to apply the FAIR Maturity Indicators to measure the ACCESSIBILITY of the data and metadata.

- Accessibility of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

[Find out more >](#)

Data Granularity and Context

Consider how the granularity and context of data and associated metadata to help to inform your FAIR objectives.

- Understand the granularity and context of the data as early as possible

[Find out more >](#)

Interoperability Maturity Indicators

Read how to apply the FAIR Maturity Indicators to measure the INTEROPERABILITY of the data and metadata.

- Interoperability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

[Find out more >](#)

Reusability Maturity Indicators

Discover how to apply the FAIR Maturity Indicators to measure the REUSABILITY of the data and metadata.

- Reusability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

[Find out more >](#)

FAIRification

Volume 2, Issue 1-2
Winter-Spring 2020



January 01 2020

A Generic Workflow for the Data FAIRification Process

Annika Jacobsen, Rajaram Kaliyaperumal, Luiz Olavo Bonino da Silva Santos, Barend Mons, Erik Schultes, Marco Roos, Mark Thompson

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 56–65.

https://doi.org/10.1162/dint_a_00028





FAIR è graduale

FAIR = CONTINUUM
«AS FAIR AS POSSIBLE»

Inclusiveness: consider FAIR as a process

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



Interim recommendations on FAIR Metrics for EOSC

February 2020

Draft for consultation

Feb. 2020

... sfumature di FAIR

Findable

Does the dataset have any identifiers assigned?

No identifier

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

No

How is the data described with metadata?

The data is not described

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

The data is not described in any repository

Accessible

How accessible is the data?

No access to data or metadata

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

No access to data

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

Unsure

The screenshot shows the ANDS Training page. The header includes the ANDS logo and navigation links: About us, News and Events, Partners and Communities, Working with data, Online Services, and Guides and resources. The main content area is titled 'Working with data' and features a sidebar with links to 'The FAIR data principles', 'FAIR webinar series (Aug/Sep 2017)', 'FAIR data training' (selected), 'Findable', 'Accessible', 'Interoperable', 'Reusable', and 'FAIR data training narkane'. The main content area is titled 'FAIR data training' and includes a list of resources: 'A basic checklist' (or more comprehensive breakdown), 'Use the FAIR data self-assessment tool' in training or consultation, 'Discussing the components via a process of transforming a dataset to be more FAIR', and 'Case studies of domain specific consideration of the principles'.

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

UTILISSIMO PER PORSI
LE DOMANDE GIUSTE...
MA SOGGETTIVO...

The screenshot shows the FAIR self-assessment tool page. The header includes the RDS, ANDS, and NECTAR logos. The main content area is titled 'FAIR self-assessment tool' and includes a welcome message: 'Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).' The footer includes navigation links: home, news, events, programs, and about.

Per sapere quanto siete FAIR

[FAIRassist.org](https://fairassist.org)

<https://fairassist.org/#/>

Help you discover resources to measure and improve FAIRness.

FAIRassist is the new, under development, educational component of the well established FAIRsharing resource.

Resource ▾	Execution Type	Key Features	Organisation	Target Objects	Reading Material
5 Star Data Rating Tool	Manual - questionnaire	Based on rating systems and maturity models	CSIRO OzName	Datasets	
AutoFAIR	Semi-automated	A portal for automating FAIR assessments for bioinformatics	Department of Computer Science, University of Maastricht		
Data Stewardship Wizard	Predictive; based on a manually filled questionnaire	Helps researchers to design a data stewardship process to achieve the highest reasonable FAIR data.			
F-UJI	Automated	The REST API support a programmatic assessment of objects based on a set of core metrics developed by the FAIR community. The metrics specification is available at https://doi.org/10.26434/chemrxiv-2019-08-01			
FAIR Data Self-Assessment Tool	Manual - questionnaire	Educational and Informational purposes			
FAIR Evaluator	Automated	1. Core universal maturity indicators 2. Compliance tests 3. Evaluation tool			
FAIR enough	Automated				
FAIR-Aware	Manual - questionnaire				
FAIR-Checker	Automated				
FAIRdat	Manual - questionnaire				
FAIRness self-assessment grids	Manual - checklist				
FAIRshake	Manual - questionnaire				

FINDABLE



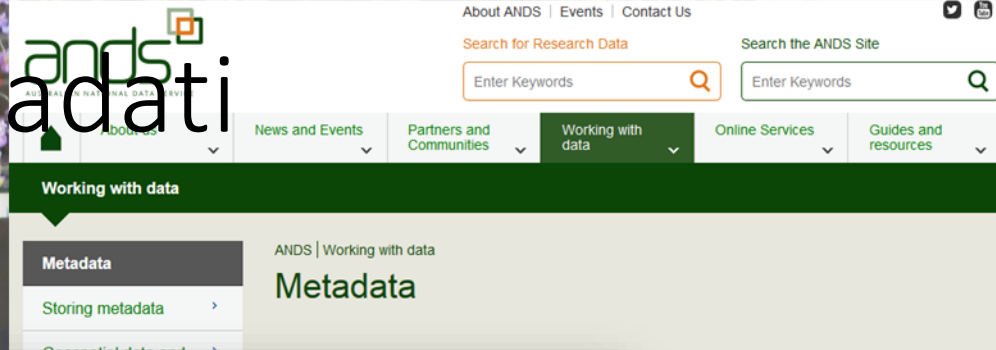
Findable - Metadata



RECOMMENDATIONS

- » A good starting point is to consult the Metadata Standards Directory, a community-maintained directory hosted by the Research Data Alliance: <https://rd-alliance.github.io/metadata-directory/>
- » Metadata works best when terminology is consistent, e.g. naming conventions are followed, spelling is normalised, and so on. Depending on the complexity and size of your metadata, consider using a tool such as Open Refine to 'clean' your metadata.
- » For greater searchability and interoperability, researchers should also consider using controlled vocabularies to identify common terminology when populating metadata fields. For example, the Library of Congress maintains a controlled vocabulary for subject headings: <http://id.loc.gov/authorities/subjects.html>
- » For findability, the metadata should include a clear and explicit reference to the dataset it describes, through the inclusion of a PID in the metadata (see section on Trustworthy Data Repositories and Persistent Identifiers).
- » Make your metadata as rich as possible in order to better contextualise your data and facilitate reuse. Consider more detailed descriptions, and fuller provenance information, as well as using the spectrum of available metadata fields.
- » Metadata should be machine-readable.

F = Findable. Metadata



- **Descriptive metadata:** information required for discovery and assessment of the collection,
 - e.g. title, contributors, subject or keywords, study description, and location and dates of the study.
- **Provenance metadata:** this relates to the origins and processing of the data, and enables interpretation and reuse of the data. It ranges from the human to the highly technical, and usually requires some knowledge of the domain to create.
 - e.g. Where did the data come from? Why was it collected? Who collected it, when and where? What instruments/technologies were used to collect the data, and how were they set up? How has the data been processed?
- **Technical metadata:** fundamental information for a person or a computer application to read the data.
 - e.g. How is the data set up? What formats, and versions of formats, are used? How is the database configured? How does it relate to other data?
- **Rights and access metadata:** information to enable access, and licensing or usage rules.
 - e.g. How can someone access the data? Who is allowed to view or modify the data, or the metadata, and under what conditions? Who has some kind of authority over the data? Are there costs associated with access? Under what licence is the data being made available?
- **Preservation metadata:** this builds on the history from the Provenance, Rights and Technical metadata, and also includes information to allow the data to be managed for long-term accessibility.
 - e.g. Has there been any restructuring or other changes to the files, e.g. due to migration to new file formats? What software has been used to access the data?
- **Citation metadata:** information required for someone to cite the data
 - e.g. Creator(s), Publication Year, Title, Publisher, Identifier.

F Findable – metadati di base



Data management stage	Metadata fields	Standardize public resources
Sample	Latitude, longitude, date/time, temperature, biome/ecosystem, depth and/or elevation of sampling site, etc.	Environmental Ontology (ENVO), Minimum Information about x Sequence (MIXS), International Geo/General Sample Number (IGSN)
Preparation	Laboratory protocol(s): DNA extraction, purification, amplification.	Protocols.io, e-laboratory notebook/management software
Data processing	Software tools for QA/QC, assembly, annotation. Include reference (if published), version, and parameters used.	Community guidelines for describing and citing software [23–25]
Feature	E.g., Annotations of sequence data, such as taxonomy or function	NCBI Taxonomy, Genome Taxonomy Database toolkit (GTDB-tk); Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), etc.
FAIR: Findable (i.e., PID metadata)	Data owner(s), organization, keywords	ORCID, Researcher Organization Registry (ROR); keyword selection [26] to enhance search engine optimization (SEO)
FAIR: Accessible	Usage license, privacy protocols, transfer protocols	Creative Commons, HTTP
FAIR: Interoperable	Type and size of data, file formats, etc.	.csv, .tsv, etc.
FAIR: Reusable	See data processing.	Workflow notebooks (e.g., [27])

<https://doi.org/10.1371/journal.pcbi.1010476.t001>

ESEMPIO DI METADATI MINIMI
SECONDO LE DIVERSE FASI DEL
CICLO DELLA RICERCA



F = Findable. Standards di metadati

Metadata

RDA | Metadata Directory

RDA Metadata directory

[Edit this page](#)

[View the standards](#)

[View the extensions](#)

[View the tools](#)

[View the use cases](#)

[Browse by subject areas](#)

[Contribute](#)

[Add standards](#)

[Add extensions](#)

[Add tools](#)

[Add use cases](#)

Arts and Humanities [Edit](#)

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

Engineering [Edit](#)

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

Life Sciences [Edit](#)

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

Physical Sciences & Mathematics [Edit](#)

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

Social and Behavioral Sciences [Edit](#)

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

General Research Data [Edit](#)

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

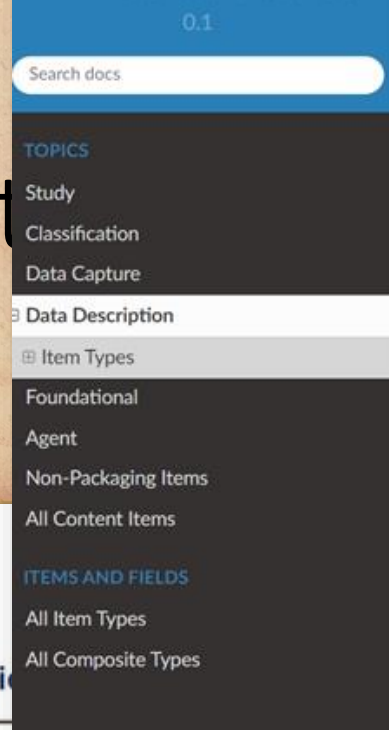
F = Findable - Met

STANDARD PER LE
SCIENZE SOCIALI



<https://ddialliance.org/>

Learn ▾ Products ▾ Membership ▾ Events ▾ Publications ▾



Data Description

Item Types

DDI has 13 item types related to Data Description

- ConceptualVariable
- RepresentedVariable
- Variable
- PhysicalInstance
- DataRelationship
- ManagedDateTimeRepresentation
- ManagedMissingValuesRepresentation
- ManagedNumericRepresentation
- ManagedScaleRepresentation
- ManagedTextRepresentation
- NCube
- VariableStatistics
- Weighting

Document, Discover and Interoperate

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding interpretation, and use -- by people, software systems, and computer networks. Use DDI to **D**ocument, **D**iscover, and **I**nteroperate!

Why Use DDI?

- ✓ Generate interactive codebooks
- ✓ Implement data catalogs
- ✓ Build question banks
- ✓ Create concordance mappings
- ✓ Harmonize and compare data
- ✓ Manage longitudinal data sets

Find Out More!



<https://ddialliance.org/>

Learn ▾ Products ▾ Membersh

What is DDI?

What does DDI stand for?

- Data Documentation Initiative

What is it?

- A free suite of standards/products originating from data professionals (within the behavioral, social and economic sciences) who realized
 - **Document** your data to make it more **Discoverable** and **Interoperable** between systems
 - Offer a framework to describe the research data produced and the processes used
 - Are platform-agnostic
 - Facilitate understanding, interpretation, and use by people, software systems, and computer networks
 - Have a great community of researchers who want their research data to be **FAIR** (Findable, Accessible, Interoperable, Reusable)
 - Offer lots of [free training materials](#) to support new and existing users
 - Allow for cross-domain usage (New! See below)

How many DDI standards are there?

- There are three different DDI standards:
 - [DDI Codebook \(DDI-C\)](#)
 - Intended to document simple or single survey data
 - Has the same information describing a dataset that
 - Used in the Dataverse data repository
 - Based on XML
 - [DDI Lifecycle \(DDI-L\)](#)
 - Can be used to document each stage of the lifecycle
 - Can be used for longitudinal datasets
 - Can describe more than one dataset
 - Based on XML
 - [DDI Cross-Domain Integration \(DDI-CDI\)](#)

Training Materials

These training materials were developed by the DDI Alliance Training Working Group, whose mission is to develop and maintain training materials for working with DDI. Only training slides approved by the DDI Alliance are made available here (and are

For DDI training materials created by others, please see the Zenodo [DDI Training Group Community](#).

DDI Basics: What is Metadata?

This is a brief module introducing the basic concept of metadata and describing the different types of metadata. It is a good introduction.

External Link:

<https://doi.org/10.5281/zenodo.5180480>

DDI Basics: What is DDI?

This is a brief general introduction to the major work products (DDI Codebook, DDI Lifecycle, etc.) and the DDI Alliance. It is a good result.

External Link:

<https://doi.org/10.5281/zenodo.5180509>

Table 1: DataCite Mandatory Properties

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional given name, family name, name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M



2017

DataCite Metadata Schema
Documentation for the Publication and
Citation of Research Data

Citation:

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with optional given name, family name, name identifier and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
10	ResourceType	1	A description of the resource.	The format is open, but the preferred format is a single term of some detail so that a pair can be formed with the sub-property. Text formats can be free-text OR terms from the CASRAI Publications resource type list.¹⁷ *** Examples: Dataset/Census Data, where 'Dataset' is resourceTypeGeneral value and 'Census Data' is

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
10.1	resourceTypeGeneral	1	The general type of a resource.	Controlled List Values: Audiovisual Collection DataPaper Dataset Event Image InteractiveResource Model PhysicalObject Service Software Sound Text¹⁸ Workflow Other See Appendix for definitions and examples.

F=Findable. Metadata



OpenRefine

A free, open source,
powerful tool for working
with messy data



Home

Community
Documentation
Download
Data Privacy
Contact Us
Blog

Openrefine

Welcome!

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with services and external data.

OpenRefine always keeps your data private on your own computer. You can share or collaborate. Your private data never leaves your computer. (It works by running a small server on your computer and you use a web browser to interact with it)

OpenRefine is available in more than 15 languages.

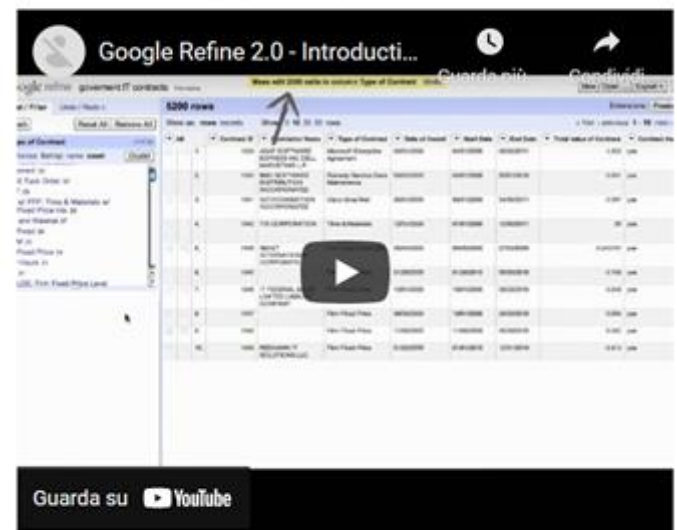
OpenRefine is part of [Code for Science & Society](#).

Introduction to OpenRefine

1. Explore Data

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out about this functionality by watching the video below.



2. Clean and Transform Data



F = findable. Metadata tools

What CEDAR does

<https://metadatacenter.org/>

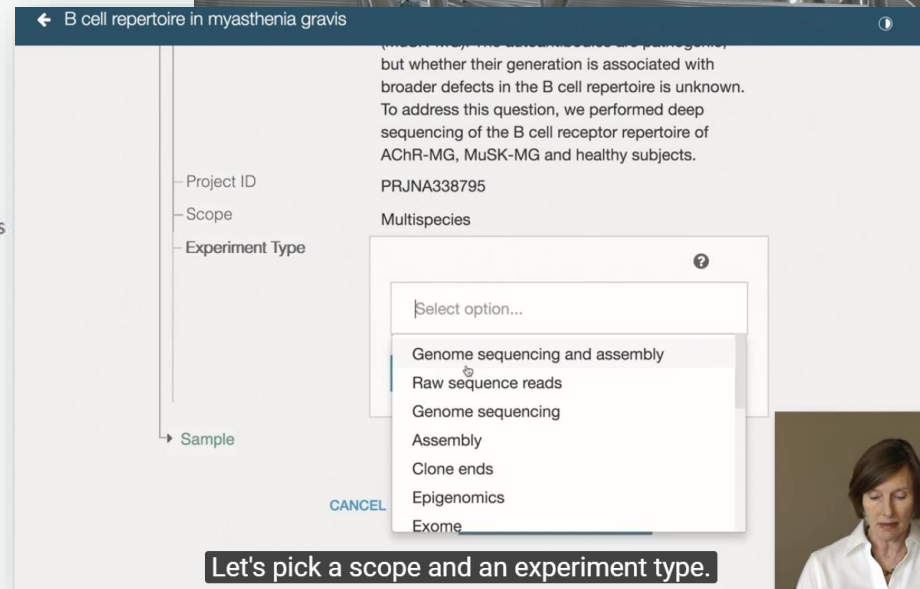
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools will create a metadata record to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools help users collect metadata, and download the information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

What can CEDAR do for me already?

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

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

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



Let's pick a scope and an experiment type.

Findable — Metadata cre

FAIRcookbook

Search this book...

FAIR Cookbook

FOREWORD

Introduction

Ethical values of FAIR

Glossary

RECIPES

Findability

Accessibility

Interoperability

1. Interlinking data from different sources

2. Identifier mapping with BridgeDb

3. Introduction to terminologies

9. Creating a metadata profile

Recipe Overview

Reading Time
20 minutes

Executable Code
Yes

Difficulty
★★★★

Creating a metadata profile

Recipe Type
Hands-on

Audience
Principal Investigator, Data Manager, Data Scientist

Cite me with FCB026

9.1. Graphical Overview

Defining a Metadata Requirement Profile: Transcriptomics Data

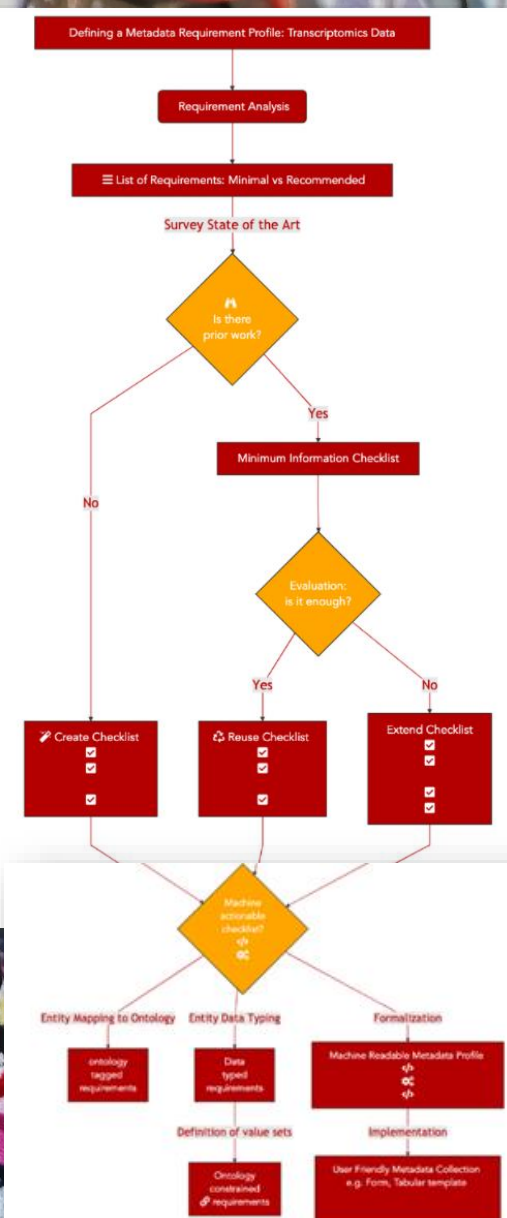


Fig. 9 Building a minimal metadata profile.

Metadati e dati:



RO-crate

RO-Crate 1.1

Background

Community

Examples

Outreach and Publications

Profiles

RO-Crate In Use

Specification

Tools

Search Research Object Crate

Research Object Crate (RO-Crate)

Permalink: <https://w3id.org/ro/crate>

TABLE OF CONTENTS

- 1 What is RO-Crate?
- 2 Where did RO-Crate come from?
- 3 Who is it for?
- 4 When can I use it?
- 5 How can I use it?
- 6 RO-Crate in use
- 7 Current
- 8 Contribute to RO-Crate community
- 9 Cite RO-Crate
 - a Cite RO-Crate as project/approach
 - b Cite RO-Crate Specification (any version)
 - c Other citations



Other citations

News: RO-Crate Metadata specification 1.1 released

What is RO-Crate?

RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata. It is based on schema.org annotations in JSON-LD, and aims to make best-practice in formal metadata description accessible and practical for use in a wider variety of situations, from an individual researcher working with a folder of data, to large data-intensive computational research environments.

Where did RO-Crate come from?

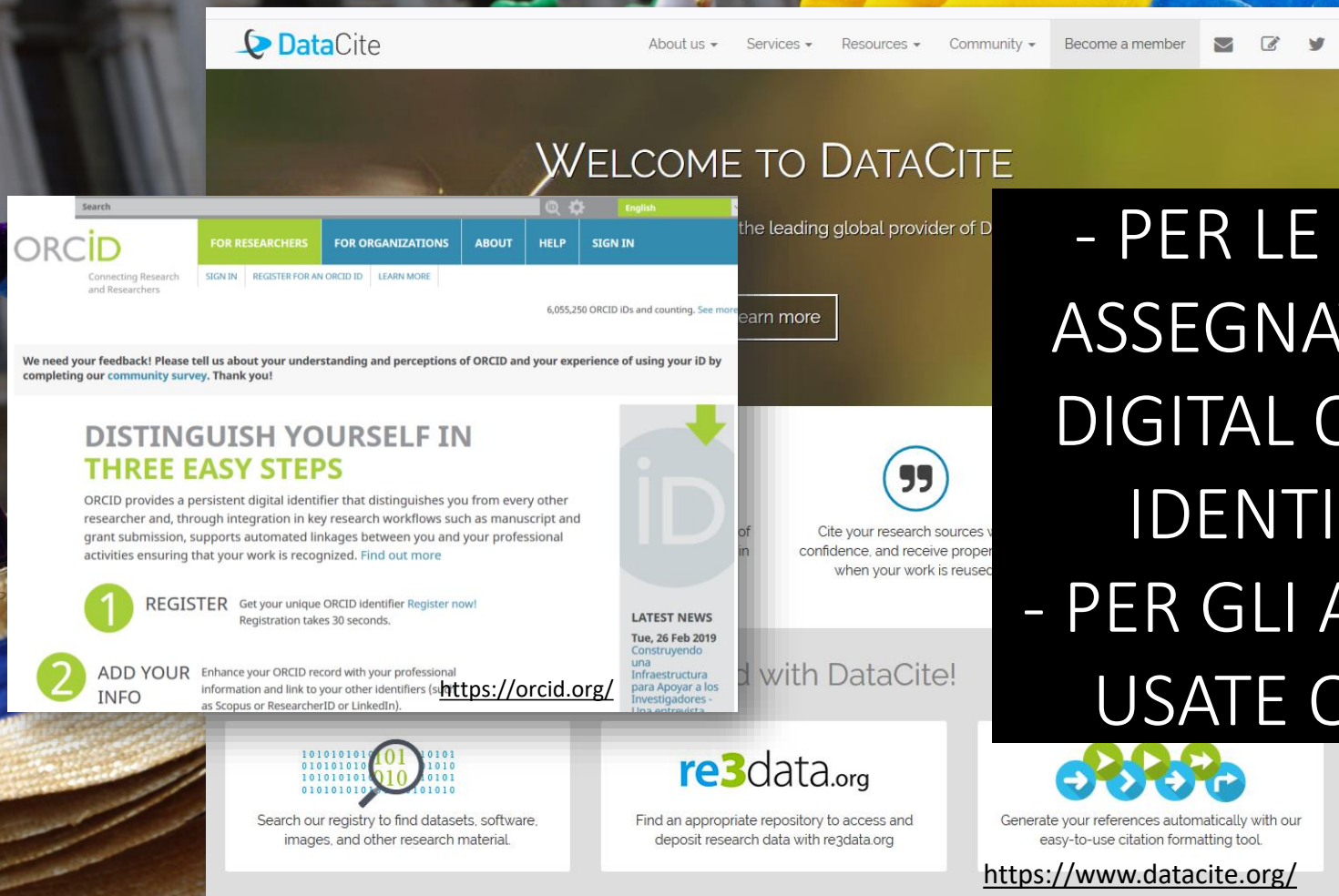
RO-Crate is the marriage of [Research Objects](#) with [DataCrate](#). It aims to build on their respective strengths, but also to draw on lessons learned from those projects and similar research data packaging efforts. For more details, see [background](#).

Who is it for?

The RO-Crate effort brings together practitioners from very different backgrounds, and with different motivations and use-cases. Among our core target users are: a) researchers engaged with computation and data-intensive, workflow-driven analysis; b) digital repository managers and infrastructure providers; c) individual researchers looking for a straight-forward tool or how-to guide to "FAIRify" their data; d)

STRUMENTO PER
«IMPACCHETTARE» DATI E
METADATI E RENDERE I DATI
FAIR

F = Findable. Identificativi persistenti



DataCite About us Services Resources Community Become a member

WELCOME TO DATACITE

the leading global provider of Digital Object Identifiers (DOIs)

ORCID Connecting Research and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

SIGN IN REGISTER FOR AN ORCID ID LEARN MORE

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

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

DISTINGUISH YOURSELF IN THREE EASY STEPS

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

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

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

3 LINK YOUR RESEARCH Link your research to your ORCID record. [Find out more](#)

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

re3data.org

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

Generate your references automatically with our easy-to-use citation formatting tool.

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

- PER LE COSE:
ASSEGNATE DOI
DIGITAL OBJECT
IDENTIFIER
- PER GLI AUTORI:
USATE ORCID

ORCID – ID. Tutto intorno a me

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



ROR per istituzioni



The background image shows the facade of the University of Rome Tor Vergata building, featuring the inscription "STVDIVM VRBIS" above the entrance. The ROR website is overlaid on the lower half of the image.

ROR  <https://ror.org/> ABOUT REGISTRY COMMUNITY BLOG **SUPPORT**



The banner image features a lion's head in the center, surrounded by colorful geometric patterns (triangles) in shades of orange, red, and purple.

ROR is a global, community-led registry of open persistent identifiers for research organizations

ACCESSIBLE



A = Accessible

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

- **Open access**

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

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

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

- **Restricted access**

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

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

- **Embargo**

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

Accessible

TRUSTED REPOSITORIES IMPORTANCE OF CONTEXT

For data to be managed over the long term, and made accessible in a continuous and sustained way, it should be deposited in a location that ensures trusted, ongoing stewardship of the data. Researchers depositing their data and those accessing it for reuse should be assured that data sets are authentic, retrievable, annotated sufficiently to understand the context of their creation, and assigned licence information that clarifies the conditions of reuse.

DIGITAL PRESERVATION

However, storage is not the same as preservation, because digital data are fragile and subject to corruption and degradation over time. File formats or the software and hardware required to access them may also become obsolete. Data published on websites can become inaccessible when links break, pages are moved, or the website disappears. Over time, technology, human actions (or inaction) and environmental factors challenge the integrity of data, so simply 'backing up' that data is not sufficient: it must be preserved. Digital preservation is not a single action, but a process that is designed to ensure digital data are continuously accessible into the future, through all of the changes that time and technology can inflict.



RECOMMENDATIONS

ALLEA report, p. 28

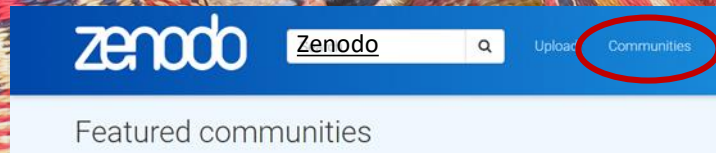


- » To ensure the best possible stewardship of your data, choose to deposit it in a digital repository that is certified by a recognised standard such as the CoreTrustSeal. The Registry of Research Data Repositories (re3data) provides a good starting point, noting disciplines, standards, content types, certification status and more. FAIRsharing (manually curated information on standards, databases, policies and collections) allows you to search databases by subject, and includes entries tagged 'Humanities and Social Sciences'.
- » Use disciplinary repositories where they exist, as they are more likely to be developed around domain expertise, disciplinary practices and community-based standards, which will promote the findability, accessibility, interoperability and ultimately the reuse and value of your data. The level of curation available in a repository is key to data quality and reusability.
- » Datasets should be assigned persistent identifiers (PID). Most repositories that are designed for long-term preservation will automatically assign or 'mint' persistent identifiers for your datasets, so choosing a quality repository will automate this step. Consider as well signing up for ORCID, a free service that assigns persistent identifiers to individuals/authors.
- » To facilitate findability of all research outputs, bidirectional links should be created between publications related outputs, such as data (using PIDs).
- » Include the richest metadata possible with your deposited data so that others can find it, understand the parameters under which it was created, and understand the conditions under which they can access and/or reuse it. See recommendations in this report in the sections on Licences and Metadata for more information.

BIDIRECTIONAL LINKS

PIDs also facilitate citation, and for increased findability, links should be created between publications and their associated datasets (bidirectional linking) These links are often created

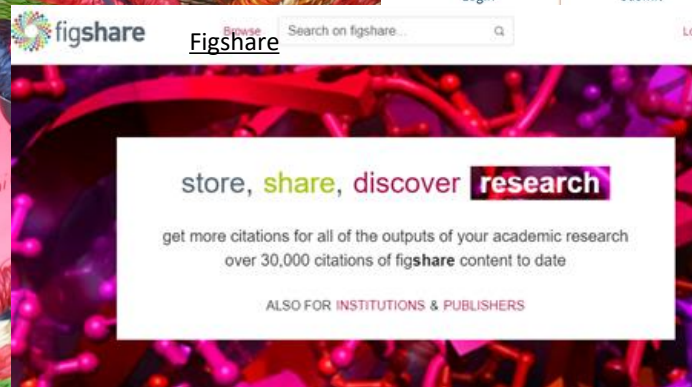
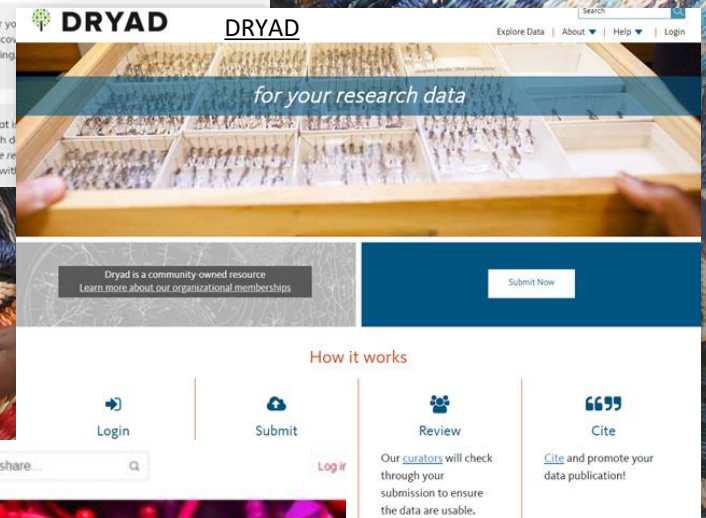
A = Accessible — Data repositories



Why use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

POSSIBILE CREARE
«COMUNITÀ»



A = Accessible. Data lab [in progress]



Save your data during research

Data lab

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

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

DataverseNL



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

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

Why DataverseNL?

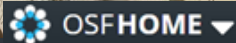
- Organization of data files in dataverses and datasets
- Addition of metadata and documentation
- Version management
- Management of access rights
- Easy collaboration with fellow researchers or project partners, even beyond your own university or research institute
- Centralized professional storage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

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

Costs

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

A = Accessible. Data repositories +



Search

The place to share your research

OSF is a free, open platform to support your research and enable collaboration.

Get started

Discover public research

Discover projects, data, materials, and collaborators on OSF that might be helpful to your own research.

Q Search discipline, author...

AMBIENTE INTEGRATO CONTIENE DATI, CODICE,
PROTOCOLLI E PERMETTE PUBBLICAZIONE PREPRINT
[SCEGLIERE SERVER AMBURGO PER GDPR]

How OSF supports your research



>



>



>



Search and Discover

Find papers, data, and materials to inspire your next research project. Search public projects to build on the work of others and find new collaborators.

Design Your Study

Start a project and add collaborators, giving them access to protocols and other research materials. Built-in version control tracks the evolution of your study.

Collect and Analyze Data

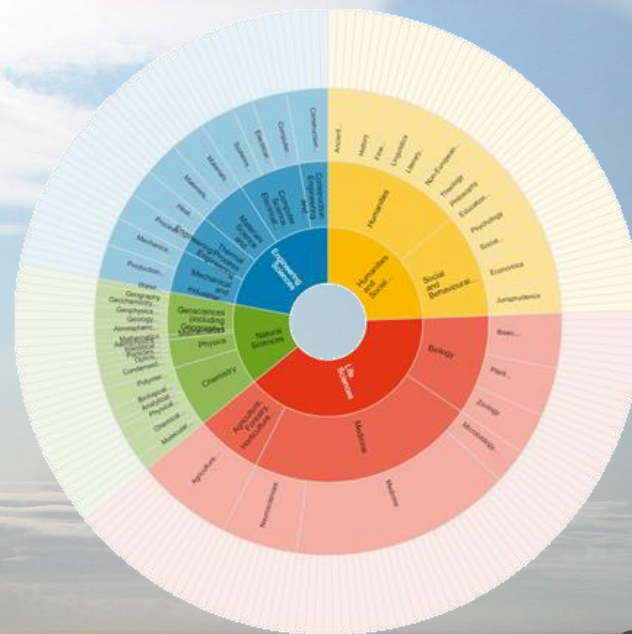
Store data, code, and other materials in OSF Storage, or connect your Dropbox or other third-party account. Every file gets a unique, persistent URL for citing and sharing.

Publish Your Reports

Share papers in OSF Preprints or a community-based preprint provider, so others can find and cite your work. Track impact with metrics like downloads and view counts.

Open Science Framework

A = Accessible. Cercate un archivio?



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](#)

A = Accessible. Data repositories

Checklist: is it the right repository for your data?

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

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

[DCC checklist](#)

CHECKLIST PER LA SCELTA

Legal terms and conditions

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) ☐

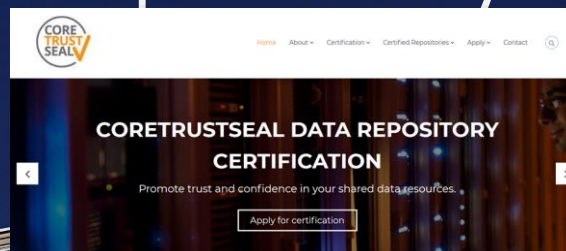
By agreeing to the terms and conditions the depositor will not be in breach of **copyright**, or any contract terms covering **Intellectual Property** in the research, (e.g. the grant conditions or a consortium agreement) ☐

Anything deposited that is not publicly accessible can be retrieved by the institution in response to a valid **Freedom of Information** request ☐

Findable, accessible and interoperable

Level 1	Level 2	Level 3
Metadata publishing: Data collections are catalogued in a repository according to funder expectations so that they are discoverable by title, creator, and date of deposition ☐	Repository publishes other pertinent information as metadata fields to enhance cross-disciplinary discovery ☐	Metadata is catalogued to enhance reuse according to sector-leading standards, or to fulfil domain-specific purposes ☐
Stable identifiers: Enables a DOI or other open standard identifier to be assigned to a landing page for each ingested dataset/ collection ☐	Supports assignment of related persistent IDs per dataset/ collection ☐	Supports assignment of multiple persistent IDs at different levels of granularity within dataset/ collection ☐
Discovery metadata: Provides Datacite mandatory metadata and exposes it according to open access repository protocols ☐	Provides metadata elements to enable broader discovery (e.g. geo-spatial) to reflect best practice changes and local needs ☐	Exposes discovery metadata as Linked Open Data to optimise automatic discovery ☐
Metadata harvesting: Sufficient information can be harvested about data deposited with third-party repositories, to meet funders' needs for metadata on ☐	Metadata can be routinely harvested with links to data producer IDs (e.g. ORCID), any grant information and related outputs, enabling it to meet the ☐	Metadata on the externally held research data is sufficiently structured and organized ☐

Criteria for the selection of a trustworthy repository



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

Title	URL	Charge	Notes for authors (N.B. we suggest checking in particular for policy on submission of data already published)	Publisher	Notes on Subject Area
Journal of Open Archaeology Data	http://openarchaeologydata.metajnl.com/		http://openarchaeologydata.metajnl.com/about/submissions	Ubiquity Press	Archaeology
Open Health Data	http://openhealthdata.metajnl.com/		http://openhealthdata.metajnl.com/about/submissions#authorGuidelines	Ubiquity Press	Public Health
Journal of Open Psychology Data	http://openpsychologydata.metajnl.com/		http://openpsychologydata.metajnl.com/about/submissions#onlineSubmissions	Ubiquity Press	Psychology
Scientific Data	http://www.nature.com/scientificdata/				
Genomics Data	http://www.journals.elsevier.com/genomics-data/				
Geoscience Data Journal	http://www.geoscience-data.com/				
Earth System Science Data	http://earth-system-science-data.net/				
Ecological Archives	http://esapubs.org/archive/				
Journal of Chemical and Engineering Data	http://pubs.acs.org/journal/jceaax				

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 potsherds counted during 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/10.5284/1000371>).

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

Data journals list



A = Accessible. Data journals

UKSG Insights June 10, 2020

Reading: Data journals: incentivizing data access and documentation within the scholarly communication system

Share: [f](#) [t](#) [s](#) [in](#)

Research Articles

Data journals: incentivizing data access and documentation within the scholarly communication system

Author: William H. Walters [✉](#)

Abstract

Data journals provide strong incentives for data creators to verify, document and disseminate their data. They also bring data access and documentation into the mainstream of scholarly communication, rewarding data creators through existing mechanisms of peer-reviewed publication and citation tracking. These same advantages are not generally associated with data repositories, or with conventional journals' data-sharing mandates. This article describes the unique advantages of data journals. It also examines the data journal landscape, presenting the characteristics of 12 data journals in

PORTANO I DATI E LA LORO
DOCUMENTAZIONE ALL'INTERNO
DELLA COMUNICAZIONE SCIENTIFICA

[UCL Home](#) » / [Open@UCL Blog](#) » / Data journals and data reports – don't miss out

Data journals and data reports – don't miss out on this useful publishing format! Aug. 2021

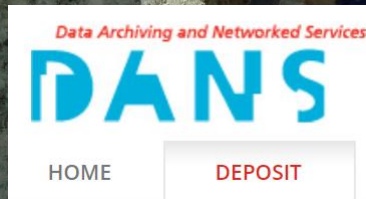
By Kirsty, on 17 August 2021

Guest post by [James Houghton](#) – Research Data Support Officer

Why not publish a data report article?

Publishing with a data journal offers several benefits. First, a data report article is more formal than a publication of data files in a repository and is a peer reviewed publication which then contributes to a researcher's publication record which is important for CVs and advancement for many. Second, they allow a more detailed explanation of a dataset and any analysis or code related to it than is usually otherwise possible. Third, the appearance of an article in a recognised journal can help to drive visibility of a dataset for other researchers. In practice it may often be the case that a repository will be used to host material which is discussed at length in a paper.

A = Accessible. Formati



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

A – Accessible – Format

National Archives



NATIONAL ARCHIVES

Appendix A: Tables of File Formats

Quick Links

Computer Aided Design

Digital Audio

Digital Moving Images

Digital Cinema

Digital Video

Digital Photographs

Scanned Text

Geospatial Formats

Presentation

Structured Data Formats

Email

Calendars

Navigational

Symbol Key

Preferred Formats ● ● ●

Acceptable Formats ● ●

Geospatial Formats

Geospatial records include digital cartographic data files and aerial photography that are created and processed in Geographic Information Systems (GIS) or other software applications for spatial analysis.

● ● ● Preferred Formats

Preferred Formats	Format Versions	Format Specifications
Geospatial Tagged Image File Format	1.8.2	Geo TIFF Format Specification: (http://geotiff.maptools.org/spec/geotiffhome.html)
Geographic Markup Language	2.0 through 3.2	ISO 19136:2007 & Version 3.2, OGC document 07-036: (http://www.opengeospatial.org/standards/iso)
Topologically Integrated Geographic Encoding and Referencing Files	2006 Second Edition	2006 Second Edition TIGER/Line®: (https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation.html)
Keyhole Markup Language	2.2	Open Geospatial Consortium Inc. OGC 07-147r2: (http://www.opengeospatial.org/standards/kml/)

● ● Acceptable Formats

Acceptable Formats	Format Versions	Format Specifications
Vector Product Format		MIL-STD-2407: (http://earth-info.nga.mil/publications/specs/printed/2407/2407_VPF.pdf)
ESRI ARC/INFO Interchange File Format		Reverse engineered specification: (http://avce00.maptools.org/docs/v7_e00_cover.html)
TerraGo Geospatial PDF	GeoPDF Encoding Best Practice Version 2.2	Open Geospatial Consortium Inc. OGC 08-139r2: (http://www.opengeospatial.org/standards/iso)
ESRI Shapefile (Compound)	1997 – current version	ESRI Shapefile Technical Description: (http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf)

● Acceptable for Imminent Transfer Formats

A = Accesible - Formati

FAIRcookbook

Search FAIR Cookbook

FAIR Cookbook

Introduction

Assessing FAIR

Infrastructure for FAIR

Improving Findability

Improving Accessibility

Improving Interoperability

How to interlink data from different sources?

Identifier mapping with BridgeDB

Which vocabulary to use?

Requesting terms addition to terminology artefacts

Tools for ontology-associated operations

Building an application ontology with Robot

Creating a Metadata Profile

From proprietary to open standard data format

File format validation - an example

Table of Data Standards

Data Formats Terminologies Models

mzML

PSI-MS

Ingredients

Tools and Software:

- github
- docker
- python

Converting Mass Spectrometry data to mzML format: a Step by Step Process.

Step 1: obtain the dataset

In the case of the IMI RESOLUTE project, the data is released via the University of Luxembourg server (assuming you have access resolved):

```
$> sftp fairplus@NNN.000.000.NNN
>get RESOLUTE_Targeted_Metabolomics_of_parental_cell_lines.tar.gz
>exit
```

From proprietary to open standard data format



Recipe metadata

Identifier: RX.X version: v1.0



Difficulty level



Reading Time

20 minutes

Recipe Type

Hands-on

Executable Code

Yes



Intended Audience

Principal Investigator

Data Manager

Data Scientist

Main Objectives

- Document how to convert raw data from a propriatory, vendor specific format to an open standard format.
- Apply the approach to an IMI dataset, more specifically a targeted metabolic profiling using Biocrates kit produced by IMI Resolute project.

INTEROPERABLE



DATI CON
SOFTWARE/MACCHINE

DATI CON ALTRI DATI

Interoperability



RECOMMENDATIONS

ALLEA Report p. 20

» Prefer human and machine-readable systems: coding of data models and of the actual data that is both human and machine-readable in a unified way provides better sustainability and long-term accessibility than machine-readable only code (binary codes), that may use different formats for data model description and the actual data. For both, hierarchical data models and graph-based data, various serialisations (file formats) are available that fulfil this condition (XML, TEI/XML, Turtle, N3, RDF/XML), whereas SQL based technologies need bigger efforts.

» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies are needed as well, but an overall ontology for the humanities has not yet been established. The ontology CIDOC-CRM and especially some extensions are well on their way to become a reference model for cultural heritage information.

» Data models follow the data management plan (DMP): when establishing a data model, researchers should keep the whole lifecycle of their data in mind, as it should be outlined in a DMP. Therefore, an extensive documentation of the data model, its software and tools are highly relevant and facilitates the transfer of data in a secure and trusted repository in order to keep them accessible. The same is true here: the more you use open standards for your data model, the easier this task becomes.

» Data models go FAIR: the FAIR Guiding Principles, correctly applied, ensure data are findable, accessible, interoperable and reusable. Data modelling should take this into account by using formal, easily accessible languages for knowledge representation, providing persistent identifiers, open standards, well documented Application Programming Interfaces (API), generic user interfaces and rich metadata. The FAIRification process developed by the GO FAIR initiative offers a system on how to shape the data modelling.

» Use open standards, and whenever possible, standardised technologies and procedures should be used. The World Wide Web Consortium W3C maintains several standards relevant for data models like XML and RDF. Within XML the Text or Music Encoding Initiative TEI/MEI or specific expressions of them have become standards for text or music editions. The query language SPARQL and the representation tool for linked data JSON-LD are common standards for RDF (refers to FAIR principle I.1).



I = Inteoperable. Standards



» Prefer human and machine-readable systems: coding of data models and of the actual data that is both human and machine-readable in a unified way provides better sustainability and long-term accessibility than machine-readable only code (binary codes), that may use different formats for data model description and the actual data. For both, hierarchical data models and graph-based data, various serialisations (file formats) are available that fulfil this condition (XML, TEI/XML, Turtle, N3, RDF/XML), whereas SQL based technologies need bigger efforts.

» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies are needed as well, but an overall ontology for the humanities has not yet been established. The ontology CIDOC-CRM and especially some extensions are well on their way to become a reference model for cultural heritage information.

NORMALIZE, USE
AUTHORITY FILES (es.
VIAF)

- ✓ Avoid disambiguity
- ✓ Easy tagging
- ✓ Independent from spelling
- ➡ Linked data and unique identifiers: Use **authority files** (Normdaten)!
- ➡ BARTOC (Basel Register of Thesauri, Ontologies & Classifications): <https://bartoc.org/>
 - **GND** (German National Library)
 - **ORCID** (Open Researcher and Contributor ID)
 - **GeoNames**
 - **Wikidata**
 - **Getty Union List of Artist Names**
 - **VIAF** (Virtual International Authority File)

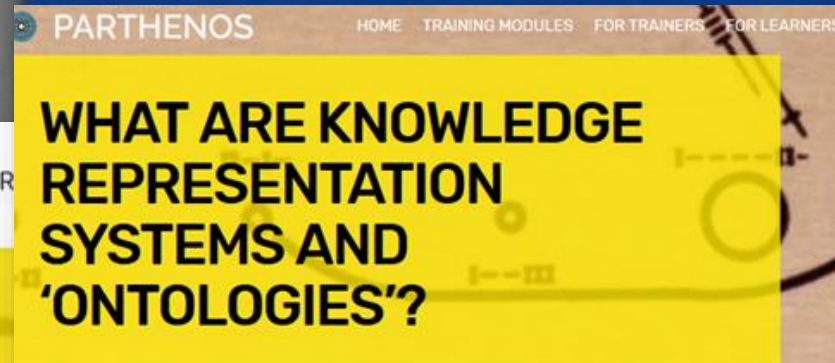
I = Interoperable. Standards



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

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

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



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



What is Metadata?

What are Standards?

What Are Knowledge Representation Systems and 'Ontologies'?

Sustainability

Methods and Tools

Networks

I = Interoperable. Standards

Standardization Survival Kit

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



Browse scenarios



Add a new scenario



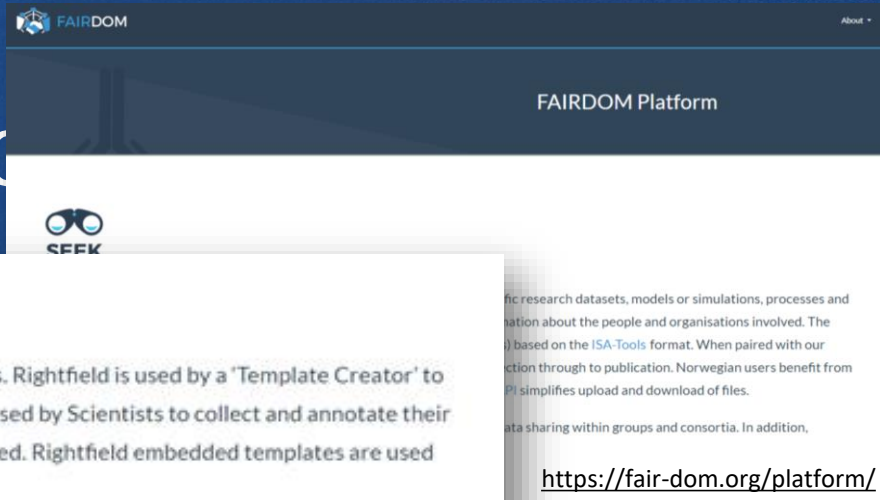
About the SSK

Increase efficiency, interoperability and sustainability by using standards

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

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

I= Interoperable – a



RightField

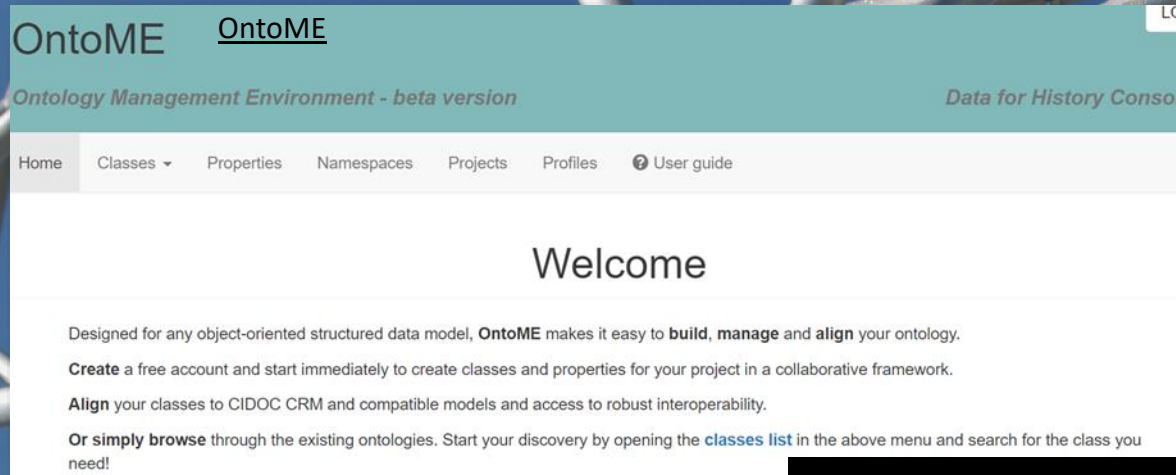
Rightfield is an open-source tool for adding ontology term selection to Excel spreadsheets. Rightfield is used by a 'Template Creator' to create semantically aware Excel spreadsheet templates. The Excel templates are then reused by Scientists to collect and annotate their data; without any need to understand,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

[MORE INFORMATION](#)

[Rightfield](#)

Metadata	Values (examples)	Notes
Asset Title		The name of the data file
Uploader		The person submitting the asset to SEEK
Uploader SEEK ID		If you add your own SEEK ID, this will help us link this asset with your profile
Project	Project	The project that the asset belongs to
ASSAY		
Assay SEEK ID		If referring to an existing Assay, you can link to it via the Assay SEEK ID.
Assay Title		The title of an existing assay
Assay_type	ExperimentalAssayType	The assay_type describes the type of experiment you are performing
Technology_type	amplification by-product_formation catabolic_response cell_growth_optimisation cell_size Comparative_genomic_hybridization comparative_genomics continuousEnzymatic	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	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

I = Inteoperable. Ontologies



The screenshot shows the OntoME web interface. At the top, there's a teal header with 'OntoME' and a link to 'OntoME'. Below it, a subtitle reads 'Ontology Management Environment - beta version' and 'Data for History Consortium'. A navigation bar includes links for Home, Classes (with a dropdown arrow), Properties, Namespaces, Projects, Profiles, and a User guide icon. The main content area has a 'Welcome' heading, followed by a paragraph explaining the platform's purpose: 'Designed for any object-oriented structured data model, **OntoME** makes it easy to **build, manage** and **align** your ontology.' It then lists three main actions: 'Create' (a free account to build a collaborative ontology), 'Align' (classes to CIDOC CRM and compatible models for interoperability), and 'Or simply browse' (existing ontologies via a 'classes list' in the menu).

ONTOME
ESEMPIO DI APPLICAZIONE
ALLE SCIENZE STORICHE



The screenshot shows an article from IOS Press. The header includes the 'IOS Press' logo, 'IOS Press Content Library', and the year '2021'. A navigation bar has links for Home and Journals. The article title is 'A challenge for historical research: Making data FAIR using a collaborative ontology management environment (OntoME)'. The author is 'Francesco Beretta', and the affiliation is 'Laboratoire de recherche historique Rhône-Alpes, CNRS – Université de Lyon, 14 avenue Berthelot, 69363 Lyon cedex 07, France'.

Abstract

This paper addresses the issue of interoperability of data generated by historical research and heritage institutions in order to make them re-usable for new research agendas according to the FAIR principles. After introducing the symogih.org project's ontology, it proposes a description of the essential aspects of the process of historical knowledge production. It then develops an epistemological and semantic analysis of conceptual data modelling applied to factual historical information, based on the foundational ontologies *Constructive Descriptions and Situations* and DOLCE, and discusses the reasons for adopting the CIDOC CRM as a core ontology for the field of historical research, but extending it with some relevant, missing high-level classes. Finally, it shows how collaborative data modelling carried out in the ontology management environment OntoME makes it possible to elaborate a communal fine-grained and adaptive ontology of the domain, provided an active research community engages in this process. With this in mind, the *Data for history* consortium was founded in 2017 and promotes the adoption of a shared conceptualization in the field of historical research.

I = Interoperable.

FORMAL ONTOLOGIES: A COMPLETE NOVICE'S GUIDE

[Formal ontologies](#)

Formal ontologies are proposed as a solution to the data heterogeneity problem because they describe broad ranges of data in a manner that is intellectually consistent and able to cover both general and particular levels of knowledge. Put another way, a formal ontology attempts to build a schema that applies to anyone and no one in a particular field.

A successful formal ontology must:

- accurately represent the most common information points of interest to a particular domain and the relationships that users want to trace between entities.
- offer sufficient abstract classes and relations in order to allow the representation of characteristic states of lack of knowledge. That is to say, create information structures which allow representing not just highly accurate information but also more general, uncertain information.

THE DATA HETEROGENEITY PROBLEM

[Data heterogeneity](#)

however, they reach their technical limits for large scale data integration and another solution is needed: a formal ontology.

Thesauri and Authority Files – why do we need them?

Before turning to formal ontologies, however, it is useful to quickly point to the role of **thesauri** and **authority files** in the process of standardization at the data level. Regardless of the standardization method chosen for the schema level, data integration is only fully achieved when harmonization is carried out also on the data value level. Enabling such standardization are thesauri and authority files. These are curated lists of either **controlled terminologies** or **controlled references**.

Controlled thesauri are generally curated by a specific community and provide a list of terms and their (un)official spellings for those concepts that are recognized and used for describing some aspect of reality. A classic example is the [Getty Art and Architecture Thesaurus](#).

+ Thesaurus Exercise (click to expand)

- Basic Formal Ontology: originally used in modelling of medical data, provides a complete methodology for data modelling
- DOLCE: was constituted to aid in modelling common sense natural language
- CIDOC CRM: originally designed in the museological community to account for cultural heritage and e-sciences data

On the other hand, other ontologies are designed to address very specific needs, ignoring the general aim of interoperability in favour of a more focused problem level. Examples of such focussed ontologies include:

- FOAF: an ontology for tracking social relations
- SPAR: for organizing citation data, article structure and content
- NeMO: for tracking scholarly process

CIDOC CRM Class Declarations	
E1 CRM Entity	
E2 Temporal Entity	
E3 Condition State	
E4 Period	
E5 Event	
E6 Destruction	
E7 Activity	
E8 Acquisition	
E9 Move	
E10 Transfer of Custody	
E11 Modification	
E12 Production	
E13 Attribute Assignment	
E14 Condition Assessment	
E15 Identifier Assignment	
E16 Measurement	
E17 Type Assignment	
E18 Physical Thing	
E19 Physical Object	
E20 Biological Object	
E21 Person	
E22 Human-Made Object	
E24 Physical Human-Made Thing	
E25 Human-Made Feature	
E26 Physical Feature	
E27 Site	

I = Interoperable



1. Interlinking data from different sources

2. Identifier mapping with BridgeDb

3. Introduction to terminologies and ontologies

4. Selecting terminologies and ontologies

5. Requesting new terms

6. Ontology-related tools and services

7. Building an application ontology with ROBOT

8. Creating a data/variable dictionary

9. Creating a metadata profile

10. Converting from proprietary to open format

11. An inventory of tools for converting your data to RDF

12. File format validation,

1. Interlinking data from different sources

Recipe Overview

 Reading Time
30 minutes

 Executable Code
No

 Difficulty
🔥🔥🔥

Interlinking data from different sources

 Recipe Type
Background information

 Audience
Principal Investigator, Data Manager, Data Scientist

Cite me with FCB016

1.1. Main Objectives

The FAIR principles, under [Interoperability](#) state that:

13. (Meta)data include qualified references to other (meta)data

[FAIR cookbook](#)

[FAIRsharing. Per essere interoperabili]

FAIRsharing.org
standards, databases, policies

search through all content

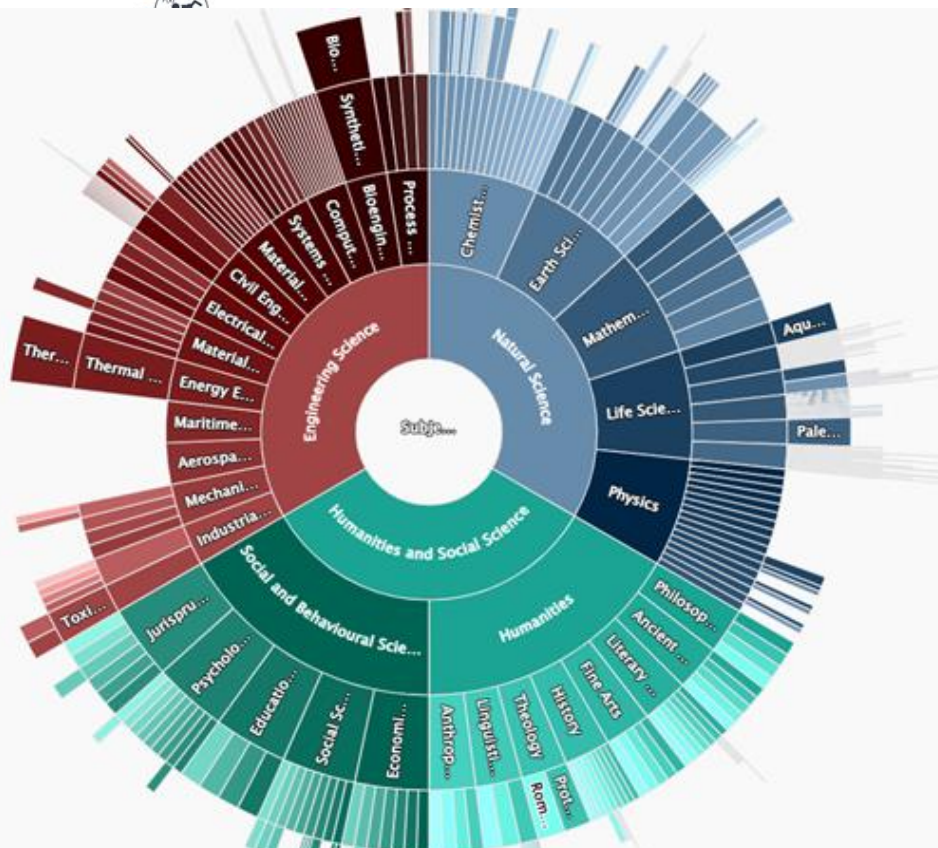
STANDARDS DATABASES POLICIES COLLECTIONS ADD CONTENT STATS LOGIN

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

RESEARCHERS DEVELOPERS & CURATORS JOURNAL PUBLISHERS LIBRARIANS & TRAINERS SOCIETIES & ALLIANCES FUNDERS

<https://fairsharing.org/>



FAIRSHARING
REGISTRO DI
STANDARD
SUDIDIVISO PER
DISCIPLINE

REUSABLE



R = Reusable. Documentazione

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



Project-level documentation



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

detail:

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

Data-level documentation

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



⊖ Quantitative data

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



Data-level documentation for quantitative data

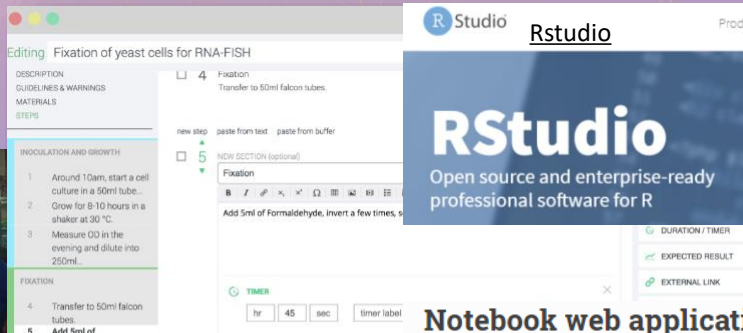
For quantitative data document the following:

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

R = Reusable. Documentazione

✓ protocols.io

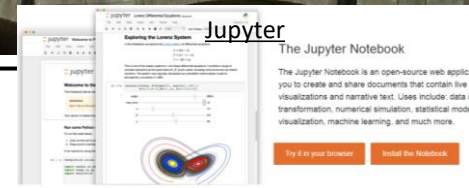
Make your science more reproducible
protocols.io is the #1 open access repository for science methods



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.



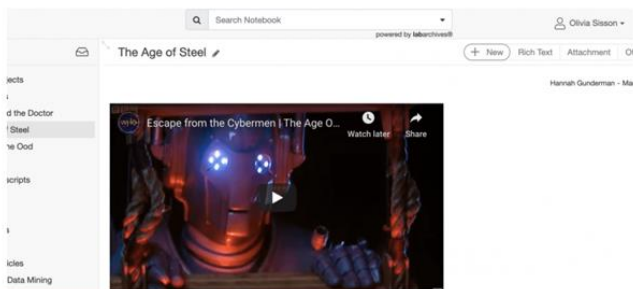
labarchives™
Better Science

Products Enterprise Integrations

Lab notebook for Humanities

Using an ELN for Humanities

by Olivia Sisson | May 21, 2020 | Data Management, Research | 0 comments



web application enables users to:

the browser, with automatic syntax highlighting, indentation, and tab

in the browser, with the results of computations attached to the code which

is of computations with rich media representations, such as HTML, LaTeX,

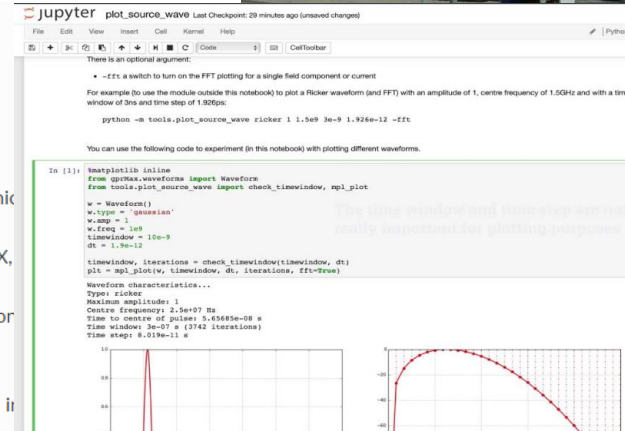
interactive JavaScript widgets, which bind interactive user interface con

live text using the Markdown markup language.

ematical equations using LaTeX syntax in Markdown, which are rendered in

...WHY NOT?

- PROTOCOLS.IO PER I PROTOCOLLI
- OPEN LAB NOTEBOOK TIENE TRACCA DI TUTTO IL LAVORO (DIFFICILE LA PRIMA VOLTA, POI...)



[documentation examples]

Structured tabular data should have as documentation (where applicable):

- variable names, labels and descriptions (maximum 80 characters)
- units of measurement for variables
- reference to the question number of a survey or questionnaire

Example: variable 'q11hexw' with label 'Q11: hours spent taking physical exercise in a typical week' — the label gives the unit of measurement and a reference to the question number (Q11)

- value code labels

Example: variable 'p1sex' = 'sex of respondent' with codes '1=female', '2=male', '8=don't know', '9=not answered'

- coding and classification schemes explained, with a bibliographic and dated reference (some standards change over time)

Examples: Standard Occupational Classification, 2000 — a series of codes to classify respondents' jobs; ISO 3166 alpha-2 country codes — an international standard of 2-letter country codes

- codes for missing data, with reason data are missing (blanks, system-missing or '0' values are best avoided)

Example: '99=not recorded', '98=not provided (no answer)', '97=not applicable', '96=not known', '95=error'

- deviating universe information for variables in case of skipped cases or questions
- derived or constructed variables created after collection, giving code, algorithm or command files used to create them — simple derivations, such as grouping age data into age intervals, can be explained in the variable and value labels; complex derivations can be described by providing the algorithms, logical statements or functions used to create derived variables, such as the SPSS or Stata command files

hse09ai.sav [DataSet2] - PASW Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing
175	quala10	Numeric	2	0	Which of the qualifications on this card do you have? 10	{-9, No ans...	-.99 - -1
176	actvb	Numeric	2	0	Activity status for last week	{-9, No ans...	-.99 - -1
177	empstat	Numeric	2	0	Manager/Foreman	{-9, No ans...	-.99 - -1
178	everjob	Numeric	2	0	Ever had paid employment or self-employed	{-9, No ans...	-.99 - -1
179	ftptime	Numeric	2	0	Full-time or part-time	{-9, No ans...	-.99 - -1
180	howlong	Numeric	2	0	How long have you been looking	{-9, No ans...	-.99 - -1
181	wkstr2	Numeric	2	0	Able to start work within 2 weeks (Government training scheme)	{-9, No ans...	-.99 - -1
182	wklook4	Numeric	2	0	Looking paid work/govt scheme last 4 weeks	{-9, No ans...	-.99 - -1
183	nempee	Numeric	2	0	Number employed at place of work	{-9, No ans...	-.99 - -1
184	nssec	Numeric	5	1	NS-SEC - long version (harmonised)	{-9.0, No a...	-.99.0 - -1.0
185	othpaid	Numeric	2	0	Ever had other employment (waiting to start work)	{-9, No ans...	-.99 - -1
186	payage	Numeric	3	0	Age when last had a paid job	{-9, No ans...	-.99 - -1
187	paylast	Numeric	4	0	Year left last paid job	{-9, No ans...	-.99 - -1
188	paymon	Numeric	2	0	Month last left paid job	{-9, No ans...	-.99 - -1
189	sclass	Numeric	2	0	Social Class	{-9, No ans...	-.99 - -1
190	seg	Numeric	2	0	Socio-Economic Group	{-9, No ans...	-.99 - -1
191	snempee	Numeric	2	0	Self employed, how many employees	{-9, No ans...	-.99 - -1
192	age	Numeric	3	0	Age last birthday	{-9, No ans...	-.99 - -1

Data View Variable View

PASW Statistics Processor is ready

G18

	A	B	C	D	E	F
	Site	Location	Type	Instrument Numb	From	
2	Beckingham	Beckingham & Idle Baro	Barometer	73937	7/2/2007	18/10/07
3	Beckingham	Beckingham Ditch	Diver	80137	7/2/2007	16/1/07
4	Beckingham	Beckingham Fld Centre	Diver	80136	7/2/2007	16/1/07
5	Beckingham	Beckingham Fld Edge	Diver	80129	7/2/2007	16/1/07
6	Bushley	Bushley Barometer	Barometer	77599	14/2/2007	4/11/07
7	Bushley	Bushley Ditch	Diver	63017	14/2/2007	23/1/07
8	Bushley	Bushley Fld Centre	Diver	53632	14/2/2007	23/1/07
9	Bushley	Bushley Fld Edge	Diver	53194	14/2/2007	12/4/07
10	Cuddycharch Sough	Cuddycharch Sough Baro	Barometer	62943	10/5/2007	30/1/07
11	Cuddycharch Sough	Cuddycharch Sough Fld Centre	Barometer	62963	10/5/2007	30/1/07
12	Cuddycharch Sough	Cuddycharch Sough Fld Edge	Barometer	62959	10/5/2007	30/1/07
13	Cuddycharch Sough	Wedholme Sough (River)	Diver	48432	10/5/2007	30/1/07
14	Idle	Idle Ditch	Diver	80133	7/2/2007	7/1/07
15	Idle	Idle Fld Centre	Diver	80131	7/2/2007	16/1/07
16	Idle	Idle Fld Edge	Diver	80132	7/2/2007	16/1/07
17	Idle	Idle Upland	Barometer	77531	8/2/2007	18/10/07
18	Morda	Morda Baro	Barometer	62975	31/5/2007	29/1/07
19	Morda	Morda Ditch	Barometer	62970	31/5/2007	29/1/07

Instrument details / Field schematics / Beckingham / Bushley / Cuddycharch / Idle / Morda

R= Reusable. License

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

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

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

RICORDA: NESSUN
COPYRIGHT SUI DATI
(NON CREATIVI)

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

on the legal protection of databases

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

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?

semplici dati e
informazioni

nessuna tutela

database
non creativo

solo diritto
sui generis

database
creativo

diritto sui generis
+ diritto d'autore

livello diritto
d'autore

livello diritto
sui generis

[webinar]

[DMP]

OpenAIRE 2019 SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

More Information about the 2019 webinar series.
data management plan | OA to research data | open science

Aspetti legali nella gestione dei dati della ricerca

Thomas Margoni
University of Glasgow - CREATE
OpenAIRE project

RDA
RESEARCH DATA ALLIANCE
EUROPE

IT NATIONAL NODE

OpenAIRE

UNIVERSITÀ DI TRENTO
Facoltà di Giurisprudenza

Paolo Guarda

IL REGIME GIURIDICO DEI DATI DELLA RICERCA SCIENTIFICA

sett. 2021

APOGEO

PROGRAMMAZIONE MARKETING DES

Home Libri Software Licensing & Data Governance

Software Licensing & Data Governance

Tutelare e gestire le creazioni tecnologiche

APOGEO

Italian Computing and Data Infrastructure

16 SETTEMBRE

PROTEZIONE E RIUSO. RIFLESSIONI TRA DIRITTO D'AUTORE E DATI DELLA RICERCA

Thomas Margoni, Università KU Leuven

Open Science
Café

OS café

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

2020

OpenAIRE Legal Policy Webinars

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

2020

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

R = Reusable – Aspetti legali

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

European Union (EU) law defines “databases”, but not data sets or, at least for copyright purposes, data. Databases that meet the legal definition^① can be protected by copyright if they are original. Data sets, if they correspond to the definition of database, are protected by copyright otherwise not. Data as such are normally excluded from copyright protection [2,3]. It is important to understand that copyright protects original expressions in the “literary and artistic” domain^②, an expression that has historically included works such as books, musical works, choreographies, cinematographic works, drawings, etc [4]. Ideas, procedures, methods of operation or mathematical concepts as such, news of the day and miscellaneous facts are excluded from copyright protection [4,5,6].



MIT Press Direct



2020

Data Intelligence

Volume 2, Issue 1-2

Winter-Spring 2020



[< Previous Article](#) [Next Article >](#)

Article Contents

Abstract

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

2. SUITABLE OPTIONS FOR LICENSING DATA AND DATABASE RIGHTS

January 01 2020

Licensing FAIR Data for Reuse

Ignasi Labastida , Thomas Margoni

[> Author and Article Information](#)

Data Intelligence (2020) 2 (1-2): 199–207.

https://doi.org/10.1162/dint_a_00042



Cite



PDF



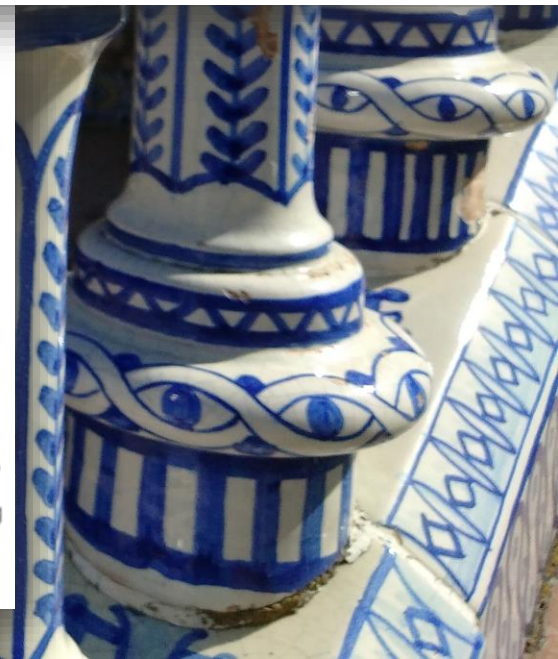
Permissions



Share

Abstract

The last letter of the FAIR acronym stands for Reusability. Data and metadata should be made available with a clear and accessible usage license. But, what are the choices? How can researchers share data and allow reusability? Are all the licenses available for sharing content suitable for data? Data can be covered by different layers of copyright protection making the relationship between data and copyright particularly complex. Some research



R = Reusable - Aspetti legali

 **OpenAIRE** How do I know SERVICES SUPPORT

Guides for Researchers

How do I know if my research data is protected?

Learn more about what is research data and their protection by intellectual property rights

 **OpenAIRE** SERVICES SUPPORT

How do I license

Guides for Researchers

How do I license my research data?

Learn more about licenses for research data and how to apply it

- WHAT IS RESEARCH DATA?
- PROTECTION OF RESEARCH DATA
- SUI GENERIS DATABASE RIGHT (SGDR)
- COPYRIGHT
- TRAINING MATERIALS

What is Research Data?

Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as 'relational' or 'functional' components of research, thus signalling that their identification and value lies in whether and how researchers use them as evidence for claims. They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.

LICENSES FOR RESEARCH DATA

HOW TO APPLY LICENSES FOR RESEARCH DATA

SPECIFICATIONS OF LICENSING RESEARCH DATA

TRAINING MATERIALS

Licenses for Research Data

What licence should be applied to the research data?

It depends on what rights protect your research data, if at all. In the light of what is explained in the guide "[How do I know if my research data is protected?](#)":

- If your research data qualifies as a work (literary work such as a journal article or a software), then CC BY 4.0 is usually the best choice. The use of the Share Alike (SA) is also compatible with the Open Access definition and reinforced in Plan S licensing guidance for publications. Non-commercial should be avoided as it is not Open Access compliant. Non-derivative is a tricky issue and should be avoided, especially if you do not know what you are doing. That said, it may not be incompatible with the Open Access definition.
- If your research data is a database or a dataset (unstructured data that do not meet the database definition) usually the best option is a CC0, which waives all your rights in the database.

Keep in mind that CC licences only deal with copyright and copyright related matter. Personal data are not included in CC and are analysed separately.

What is a Creative Commons licence?

- How can a protected dataset be used? +
- Where are licences found? +
- Interoperability and stacking +
- What happens if I use 'Share Alike' (SA) licensed material in my work? Does that mean I have to make my work available under the same SA licence? +
- Can a dataset be used if there is no licence? +
- What are the risks of using a dataset without a licence? +
- Training materials +

 **OpenAIRE** Can I use SERVICES SUPPORT

Guides for Researchers

Can I reuse someone else's research data?

Learn more on how to reuse research data

R = Reusable - Aspetti legali

EOSC-Pillar
Coordination and implementation of National & Thematic initiatives to support EOSC

2022

Results Use Cases Resources News & Events Th

Phase1
Research Proposal

Phase2
Research Implementation

Phase3
Research review

EOSC-Pillar
Coordination and implementation of National & Thematic initiatives to support EOSC

Legal Compliance

Guidelines for Researchers:
a Checklist

Check whether there is background information, data and intellectual property rights brought into the project. More specifically

Clarify who brings what

Identify the member state law applicable to each background material. Pay particular attention to the territorial applicability of each right

Make sure to secure clearance for each use by

- Obtaining any authorisation needed for re-use and protection of derived data and other intellectual outputs,
- Agree on rules of ownership attribution on derived data and other intellectual outputs.

Aim at avoiding secrecy and at allowing re-use

Define Clearly

The ownership and/or co-ownership of each research output stemming from

- The use and re-use of pre-existing background information, data and IPRs,
- Single or joint research activities within the framework of the project,
- Single or joint research activities partially within OR outside the framework of the project, if building or depending on project activities.

Creative Commons

CC Factsheet  creative commons UK

FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE v0.1

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

LICENZA CC0:
LEGALMENTE LA PIÙ
CORRETTA

What is Open Science?

[Open Science](#) is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

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

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

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



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

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

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

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

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

PUBLIC DOMAIN MARK LOGO



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

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



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



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

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

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

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

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

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

Commons Open

USARE UNA CC0

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

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

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

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

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

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

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

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

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

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

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

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



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



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



[non suoniamo tutti la stessa musica]

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

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

- **Diversity in copyright owner**

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

[CESSDA guide](#)

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

- **Diversity in copyright protection**

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

CHIARITE SUBITO CON I PARTNER
(SOPRATTUTTO ESTERI) SE CI SONO
OBBLIGHI DIVERSI DI LEGGE O
REGOLAMENTI INTERNI DA
RISPETTARE

[regole chiare dall'inizio]

Don't even
think of park-
ing here! 😊

- ... FISSATE REGOLE CHIARE DA SUBITO
- CHI HA I DIRITTI SUI DATI (SE ESISTE)
- CHI DEVE PROVVEDERE A CONSERVARE
 - CHI PUÒ SFRUTTARE

Licenze per software

ATTENZIONE A POSSIBILI
RESTRIZIONI

Open Software Licenses

Freedom of choice

- Does your institution let you choose an open license freely?
- Are there other legal restrictions?
 - Export restrictions
 - Dual-use of the software
 - Privacy

When in doubt, consult your legal advisor

Using these preferred licenses

saves you a lot of extra work.

Also, legal restrictions may apply.

First,

software that is published under an open license

is actually exported to all countries in the world.

So, by sharing your software openly,

you might be in conflict with an export restriction

to a specific country.

Second,

if your software

could also be applicable

to the production of weapons of mass destruction,

you are not allowed to publish it.

Third,

Meta Keijzer - de Ruijter, Project Manager at TUDelft – EdX MOOC on Open Science



Licenze per software 2

ESCLUDE SUCCESSIVA PROPRIETARIZZAZIONE

	Free			Non-free		
	Public domain & equivalents	Permissive license	Copyleft (protective license)	Noncommercial license	Proprietary license	Trade secret
Description	Grants all rights	Grants use rights, including right to relicense (allows proprietization, license compatibility)	Grants use rights, forbids proprietization	Grants rights for noncommercial use only. May be combined with share-alike.	Traditional use of copyright ; no rights need be granted	No information made public
Software	PD, Unlicense, CC0	BSD, MIT, Apache, MPL	GPL, AGPL	JRL, AFPL	Proprietary software, no public license	Private, internal software
Other creative works	PD, CC0	CC-BY	CC-BY-SA	CC-BY-NC	Copyright, no public license	Unpublished

<https://en.wikipedia.org/wiki/Copyleft>

Licenze per software 3

WIZARD CHE VI AIUTA AD
ASSOCIARE LA LICENZA
ADATTA ALLE VOSTRE
ESIGENZE

Licence differentiator

This tool attempts to help its users understand their own preferences in relation to free and open source software licences. It is no substitute for reading the licences themselves, and before placing any of your property under one of these licences, it is essential that you fully read and understand your chosen licence. The classifications of licence type that enable this tool to work are by necessity somewhat reductive, and therefore output of this tool cannot and must not be thought of as legal advice.

REMEMBER: ALWAYS READ AND UNDERSTAND YOUR CHOSEN LICENCE.

Choice One

Do you want to limit the results to licences that the Open Source Initiative describe as being "popular and widely used or with strong communities"?

This will guarantee that the licence will be 'mainstream' at the possible expense of some more esoteric but possibly useful characteristics.

Please choose

Choice Two (a)

All Free and Open Source licences will allow others to make modified versions of your code, and to make these modified versions available to others. Your licence can make conditions about how this happens - specifically what licences can be used on these modified versions. These conditions can help keep your code free, but they can also put some people off reusing your code.

Do you want to include licensing conditions on reuse?

If not, your licence will be one of the so-called 'permissive' licences.

Please choose

Choice Three

[Licence differentiator](#)



Academic Free License 3.0	[No score]
Adaptive Public License	[No score]
Affero GNU Public License	[No score]
Apache License 2.0	[No score]
Artistic License 2.0	[No score]
Attribution Assurance Licenses	[No score]
Boost Software License	[No score]
Common Development and Distribution License	[No score]
Common Public Attribution License 1.0	[No score]
Common Public License 1.0	[No score]
Eclipse Public License	[No score]
Educational Community License Version 2.0	[No score]
Eiffel Forum License v2.0	[No score]
European Union Public License	[No score]
Fair License	[No score]
GNU General Public License	[No score]
GNU General Public License v3.0	[No score]

3. DATI OPEN



Perché i dati aperti?



Wilma van Wezenbeek

@wvanwezenbeek

Following

#osc2018 @sjDCC I really like what Sarah said just now "There is more risk in losing your data than sharing your data #openscience"

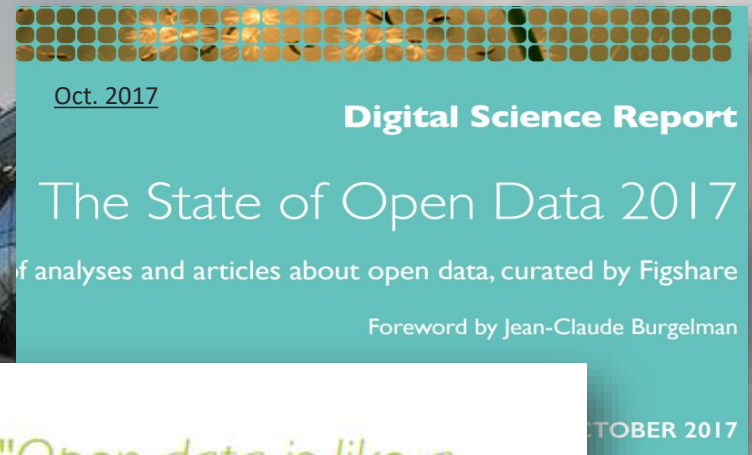
Traduci il Tweet

11:14 - 13 mar 2018

10 Retweet 10 Mi piace



<https://twitter.com/wvanwezenbeek/status/973502457115537408>



Sharing data: good for science, good for you



FA BENE ALLA
SCIENZA, FA BENE
AGLI AUTORI

Sharing data: good for science, good for you

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."

Open data saves lives

Digital Science Report

The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

Foreword by Natasha Simons, Australian Research Data Commons (ARDC)

Nov. 29, 2021

November 2021

Open data saves lives. The global pandemic has highlighted beyond anything that came before it the importance of data sharing in solving the big challenges of our time. COVID-19 data may be the most visualized data in history and it was made publicly available on a daily basis to people all over the world. The urgent need to better understand and treat the virus in 2020 brought unprecedented collective and collaborative action from all research stakeholders on an international scale to bring down barriers to research and speed up analysis and testing. These efforts, combined with support from governments and industry, resulted in not one but many vaccines made available by the end of the year. This gives us a glimpse of what incredible research outcomes are possible when we start with collaboration to address a common threat. Imagine how much more we could do, how many more lives we could save, if research data was routinely made open and shared. So, why isn't data sharing the norm? The answers lie in the harmony needed between policies, infrastructure, and practices.

Perché i dati aperti

Sharing Data Why share data

Sharing data

2. Why share data?



Better research

- Demonstrates research integrity, as there is transparency and accountability in the production of the data being released
- Encourages research enquiry and debate
- Promotes innovation and potential new data uses
- Encourages the improvement of research methods
- Prevents research fraud

Better impact

- Enables peer scrutiny of the research findings, validating the work carried out
- Increases the visibility of the research
- Provides credit for the creation of the data in its own right
- Can lead to new collaborations
- Produces a public record of the research

Better value

- Avoids duplication of effort in data creation
- Provides resources for use in teaching and learning
- Meets funder requirements
- Ensures data can be re-visited for future research
- Maximises return on research investment
- Preparing data for sharing also prepares it well for preservation

UNA RICERCA MIGLIORE

- INTEGRITÀ
- DIBATTITO
- RIUSO INEDITO

UN MAGGIORE IMPATTO

- VISIBILITÀ
- CREDITO
- COLLABORAZIONI

PIÙ VALORE

- EVITA DUPLICAZIONI
- MASSIMO RITORNO SU INVESTIMENTI

Can I get the data
associated with your **SCIENTIFIC PAPER** ?

Maybe later?



"Data" is
available upon
reasonable
request.



Repository name,
but no link.



Ok, but it's A LOT.



It was all in this
Github repository!



doi.org/something



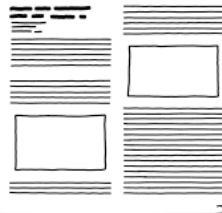
Some is here,
some is there, ask
us for the rest?



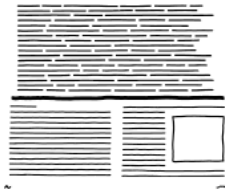
Data is "available"
upon reasonable
request.



Maybe it's in the
article/supplement?



Data is available
upon "reasonable"
request.



Only under these
specific terms.



Sorry, but nope.



La realtà

People will contact me to ask about stuff

Christopher and Alex (C&A) say: "This is usually an objection of people who feel overworked and that [data sharing] isn't part of their job..." I would also learn from each other – if a researcher is open to datasets, collaborating with others, and generally they should be outed by their community as a person who

COMUNICAZIONE
SCIENTIFICA=
GRANDE
CONVERSAZIONE

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that misinterpret the data need help." From the UK Data Archive: "Provide contextual information for your repository users so they can correctly use and understand your data."

IMPOSSIBILE, SE
NELLA «R» DI FAIR
AVETE
DOCUMENTATO

It's worth mentioning, however, a small number of people may actually be useful to counter willful misrepresentation (e.g., Freedom of Information legislation), as one can frequently point to the repository on the web to refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how interesting your data is to people that care about them." I'd also note that a dataset has value to future research. For example, "climate change" was a research topic that required documenting and understanding

EHM... ALLORA
PERCHÉ USA
FONDI PUBBLICI?

I might want to use it in a research paper

Anyone who's discussed data sharing with a researcher is familiar with this excuse. The operative word here is *might*. How many papers have we all considered writing, only to have them shift to the back burner? This is a real concern.

EMBARGO
PERFETTAMENTE
COMPATIBILE
CON FAIR

C&A suggest the embargo route: "One could require an embargo; require people to archive the data and make it public after X months. You could even go further and require things that are no longer cared about be archived. Eventually everything can become open. I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository."

I would also add that, as the original collector of the data, you are at a huge advantage compared to others that might want to use your dataset. You have knowledge about your system, the conditions during collection, the nuances of your methods, et cetera that could never be fully described in the best metadata.

I'm not sure I own the data

My data is too complicated.

C&A: "Don't be too smug. If it turns out it's not that complicated, professional [standing]." I would add that if it's too complicated to reproduce, which means it's arguably not worth sharing, it can be solved by more documentation.

IMPOSSIBILE, SE
NELLA «R» DI FAIR
AVETE
DOCUMENTATO

My data is embarrassingly bad

C&A: "Many eyes will help you improve your data. I accept your data for what it is." I accept the making the sausage. We know it's not perfect. Plus it helps you strive will be at the end of the collection phase.

MA COME SI FA A
FARE RICERCA
CON DATI
«BRUTTI»???

It's not a priority and I'm busy

SEMPRE MAGGIOR NUMERO DI FINANZIATORI E
DI RIVISTE LO CHIEDE...
STA DIVENTANDO UNA PRIORITÀ

...eppure

Digital Science Report The State of Open Data 2021

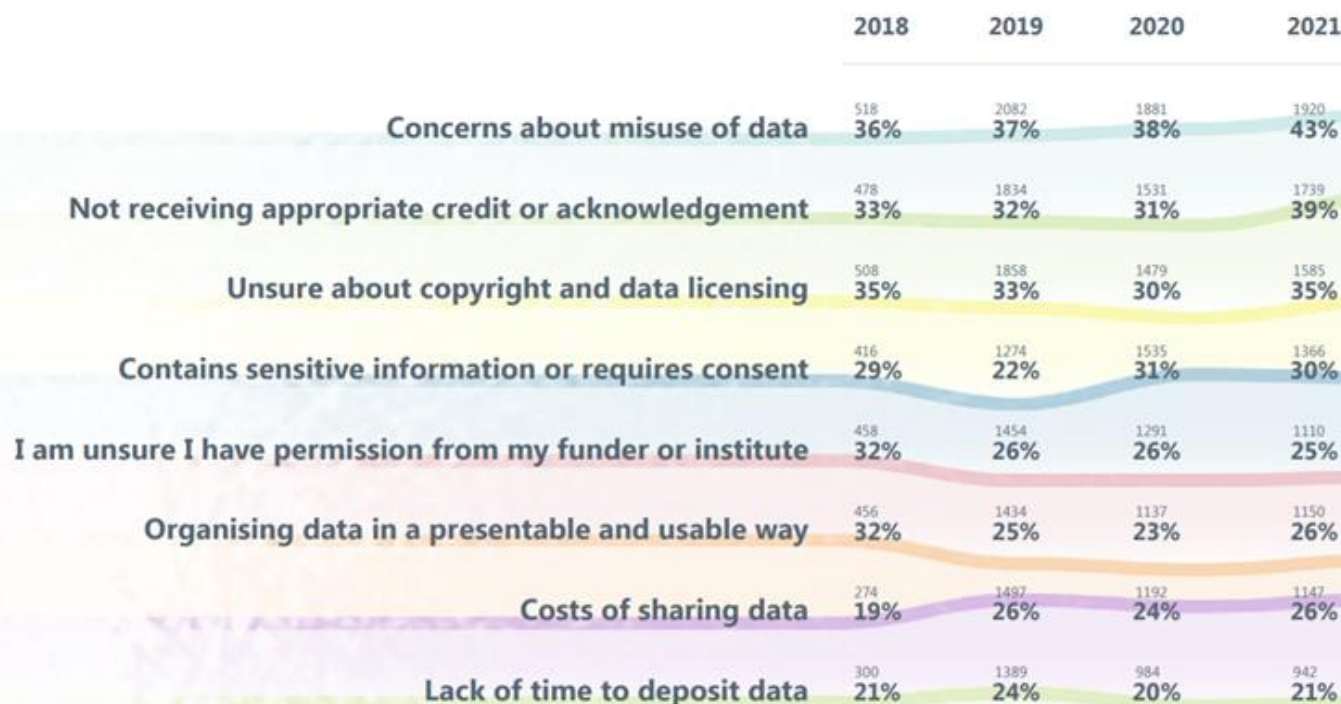
The longest-running longitudinal survey and analysis on open data

Foreword by Natasha SIMONS, Australian Research Data Commons (ARDC)

Nov. 29, 2021

November 2021

What is most striking about this year's State of Open Data report is that while researchers' familiarity and compliance with the FAIR data principles is greater than ever before, there is also more concern about sharing datasets than ever before. In their article on the three key findings of this year's State of Open Data report, Dr. Greg Goodey and Megan Hardeman stress that concern has risen in several key areas, one of which is not receiving enough credit or acknowledgement for data sharing. This points to the uncomfortable tension between the increasing ubiquity of data management and data availability policies and the rareness of rewards and recognition for data sharing. Clearly, the reward and recognition structures of academia are misaligned with the transparency of research



- PREOCCUPAZIONI
- USO IMPROPRIO
 - NESSUN CREDITO
 - CONSENSO

Pro e c

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	It is! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardener's diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.
2	I want to publish my work before anyone else sees my data.	Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh. If you are still concerned you can embargo your data for a specific period of time.
3	I have not got the time or money to prepare data for sharing	It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhancing the quality of the data for your use too.
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	Don't assume that participants will not participate because data sharing is discussed. Talk to them - they may be less reluctant than you might think, or less concerned over data sharing! Make it clear that it is entirely their decision, whereby they can decide whether their data can be shared, independent of them participating in the research. Explain clearly what data sharing means, and why it may be important. But they are still free to consent or not. You can always explain what data archiving means in practice for their data. If you have not asked permission to share data during the research, then you can always return to gain retrospective permission from participants.
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymisation procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time. Also data that is held in the UK Data Archive is not publically available. Only registered researchers can gain access to the data.
6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	Quantitative data can be anonymised through processes of aggregation, top coding, removal of variables, or controlled access to certain variables (i.e. postcodes).
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	Visual data can be anonymised through blurring faces or distorting voices, but this can be time consuming and costly to carry out. It can mean losing much of the value of the data. It is better to ask for consent to share data from participants in an unanonymised form,
8	I have made promises to destroy my data once the project finishes.	Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice from? You may need to negotiate with research ethics committee or ethics boards about this agreement.

CI SONO BUONI
ARGOMENTI CHE
SMONTANO LE
PAURE RISPETTO
ALLA
CONDIVISIONE

Pro e contro / 2

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING UK Data service p. 11
9	My data have been gathered under complete assurances of confidentiality.	Again why was such an assurance made? It is best to avoid unnecessary promises. Anonymisation procedures can be implemented to protect identities, but confidentiality can never be completely guaranteed. You can also consider controlling access to the data.
10	My data collection and resulting transcripts are in a foreign language.	This should not be a problem. The UK Data Archive can accept foreign language transcripts although translations into English are preferred.
11	It is impossible to anonymise my transcripts as too much useful information is lost.	Get in touch with us at the UK Data Archive. We may be able to help and it might not be as difficult as it looks. Also, access controls on the data may be a better solution than anonymisation if too much useful information would be lost.
12	My data collection contains data which I have purchased and it cannot be made public.	It is important to know who holds the copyright to the data you are using and to obtain the relevant permissions. You need to be aware of the licence conditions of the data you are using and what you can and cannot do with the data.
13	Other researchers would not understand my data at all - or may use them for the wrong purpose.	Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data.
14	There is IPR in the data.	This should not be a problem if you seek copyright permission from the owner of the intellectual property rights. This is best done early on in the research project, but could be sought retrospectively.

CI SONO BUONI
ARGOMENTI CHE
SMONTANO LE
PAURE RISPETTO
ALLA
CONDIVISIONE

Molto Pro!

1. You can't lose data that's in a public data repository

Have you ever lost track of a dataset? Maybe you've upgraded your computer or moved to a new institution. Maybe you deleted a file by mistake, or simply can't remember the name of the file you're looking for. No matter the cause, lost data can be embarrassing and time consuming. You're unable to supply requested information to journals during the submission process or to readers after publication. Future meta analyses or systematic reviews are impossible. And you may end up redoing experiments in order to move forward with your line of inquiry. With data securely deposited in a repository with a unique DOI for tracking, archival standards to prevent loss, and metadata and readme materials to make sure your data is used correctly, fulfilling journal requests or revisiting past work is easy.

2. Public data repositories support understanding, reanalysis and reuse

Transparently posting raw data to a public repository supports trustworthy, reproducible scientific research. Insight into the data and analysis gives readers a deeper understanding of published research articles. Offering the opportunity for others to interpret results demonstrates integrity and opens new avenues for discussion and collaboration. Machine-readable data formatting allows the work to be incorporated into future systematic reviews or meta analyses, expanding its usefulness.

3. Public data repositories facilitate discovery

Even the best data can't be used unless it can be found. Detailed metadata, database indexing, and bidirectional linking to and from related articles helps to make data in public repositories easily searchable—so that it reaches the readers who need it most, maximizing the impact and influence of the study as a whole.

PLOS BLOGS

The Official PLOS Blog

Oct. 2021

[About This Blog](#) [Contact](#)

[Browse all PLOS Blogs](#)

What a difference a data repository makes: Six ways depositing data maximizes the impact of your science

October 18, 2021 / [Lindsay Morton](#) / [Open Access](#) [Open Data](#)

4. Public data repositories reflect the true value of data

Data shouldn't be treated like an ancillary by-product of a research article. Data *is* research. And researchers deserve academic credit for collecting, capturing and curating the data they generate through their work. Public repositories help to illustrate the true importance and lasting relevance of datasets by assigning them their own unique DOI, distinct from that of related research articles—so that datasets can accumulate citations in their own right.

5. Public data demonstrates rigor

There's no better way to illustrate the rigor of your results than explaining exactly how you achieved them. Sharing data lets you demonstrate your credibility and inspires confidence in readers by contextualizing results and facilitating reproducibility.

6. Research with data in public data repositories attracts more citations

A 2020 study of more than 500,000 published research articles found articles that link to data in a public repository have a 25% higher citation rate on average than articles where data is available on request or as Supporting Information. The precise reasons for the association remain unclear. Are researchers who deposit carefully curated data in a repository also more likely to produce rigorous, citation-worthy research? Are researchers with the time and resources to devote to data curation and deposition more established in their careers, and therefore more highly cited? Are readers more likely to cite research when they trust that they can verify the conclusions with data? Perhaps some combination?

... «as open as possible»...

Commission européenne
Europese Commissie



Carlos Moedas ✓

@Moedas

Segui

2/4 "Open as possible, as closed as necessary" is the new principle for all [#data](#) from publicly funded [#research](#) in Europe [#openaccess](#)

RETWEET
76

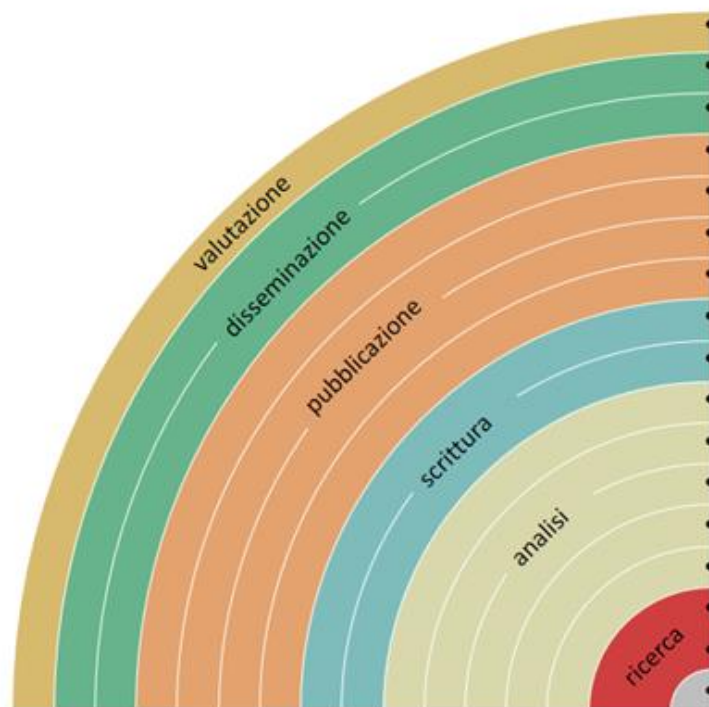
MI PIACE
32



Open science: perché solo i dati?

OS rainbow

Come rendere Open ogni passo della ricerca...



- aggiungendo misure di impatto alternative, es. [altmetrics](#)
- comunicando sui social media, es. [Twitter](#)
- condividendo poster e presentazioni, es. su [FigShare](#)
- utilizzando licenze aperte, es. [Creative Commons BY](#)
- depositando in [archivi](#) o pubblicando su [riviste Open](#)
- provando la open peer review, es. [PubPeer](#) o [F1000](#)
- condividendo preprints, su [OSFpreprint](#), [arXiv](#) o [biorXiv](#)
- con formati leggibili dalle macchine, es. [Jupyter](#) o [CoCalc](#)
- con la scrittura collaborativa, es. [Overleaf](#) o [Authorea](#)
- condividendo protocolli e workflow, es. su [Protocols.io](#)
- condividendo note di laboratorio, es. [OpenLabNotebook](#)
- condividendo software, es. su [GitHub](#) con licenza [GNU/MIT](#)
- condividendo i dati, es. su [Dryad](#), [Zenodo](#) o [Dataverse](#)
- pre-registrando esperimenti, es. [OSFregistry](#) o [AsPredicted](#)
- commentando pagine web, es. su [Hypothes.is](#) o [Pund.it](#)
- usando bibliografie condivise, es. su [Zotero](#)
- condividendo progetti di ricerca, es. su [RIO Journal](#)



Bianca Kramer & Jeroen Bosman <https://101innovations.wordpress.com> DOI: 10.5281/zenodo.1147025

Traduzione: Elena Giglia



DOI: 10.5281/zenodo.1195648

ORIZZONTE DI RIFERIMENTO RESTA LA OPEN SCIENCE...
TUTTO IL CICLO VA APERTO, NON SOLO I DATI

A wooden bench made of thick, weathered planks sits on a brick-paved surface. A sign made of four vertical wooden planks is leaning against the front of the bench. The sign has black text printed on it. The background shows more of the brick pavement and the legs of the bench.

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

... grazie e... ora tocca a voi!