

OPEN SCIENCE DALLA A ALLA Z

6 – DATI FAIR, DATI OPEN E DMP



UniMOL, maggio 2021



Elena Giglia
Università di Torino
elena.giglia@unito.it



@egiglia



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

Fatevi sentire...

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

codice mentimeter: **81**

Impareremo in questo modulo

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

2. come scrivere un DMP, che
diventerà obbligatorio [supporto!]

MESSAGGI CHIAVE

- FAIR è il futuro
(se no si resta tagliati fuori)
- È più facile di quanto sembri...
- DMP online/Data Wizard sono
pronti all'uso, basta fare pratica

[i tre passi fondamentali]



OPEN FAIR GESTITI

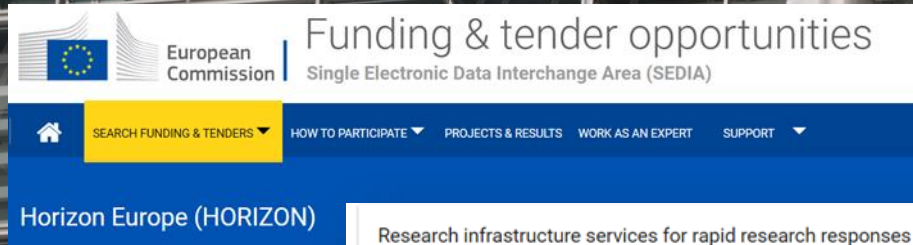
1. I DATI DEVONO ESSERE «AS OPEN AS POSSIBLE»

2. MA SE I DATI NON SONO «FAIR», APRIRLI COMPORTA RISCHI
(USO SCORRETTO, CATTIVE INTERPETAZIONI, ...)

3. MA SE I DATI NON SONO CORRETTAMENTE GESTITI, RENDERLI
«FAIR» COSTA TROPPO TEMPO E DENARO

E GESTIRE I DATI CORRETTAMENTE È NELL'INTERESSE PRIMARIO DI CHI FA RICERCA,
PERCHÉ L'INTERA RICERCA SCORRE PIÙ FLUIDA

...e non dimenticate Horizon Europe



The screenshot shows the top navigation bar of the European Commission website. It includes the European Commission logo, the text 'Funding & tender opportunities', and 'Single Electronic Data Interchange Area (SEDIA)'. Below this is a blue navigation bar with a home icon, 'SEARCH FUNDING & TENDERS', and several menu items: 'HOW TO PARTICIPATE', 'PROJECTS & RESULTS', 'WORK AS AN EXPERT', and 'SUPPORT'. A blue banner below the navigation bar reads 'Horizon Europe (HORIZON)'.

Research infrastructure services for rapid research responses to COVID-19 and other infectious disease epidemics HORIZON-INFRA-2021-EMERGENCY-02

Framework Programme: Horizon Europe (HORIZON)

Update: The submission session is now available for: HORIZON-INFRA-2021-EMERGENCY-02(HORIZON-RIA)

Date: 11 May 2021

FAIR and open data sharing in support to European preparedness for COVID-19 and other infectious diseases HORIZON-INFRA-2021-EMERGENCY-01

Framework Programme: Horizon Europe (HORIZON)

Update: The submission session is now available for: HORIZON-INFRA-2021-EMERGENCY-01(HORIZON-RIA)

Date: 11 May 2021

FAIR and open data sharing in support to European preparedness for COVID-19 and other infectious diseases HORIZON-INFRA-2021-EMERGENCY-01

Framework Programme: Horizon Europe (HORIZON)

Update:

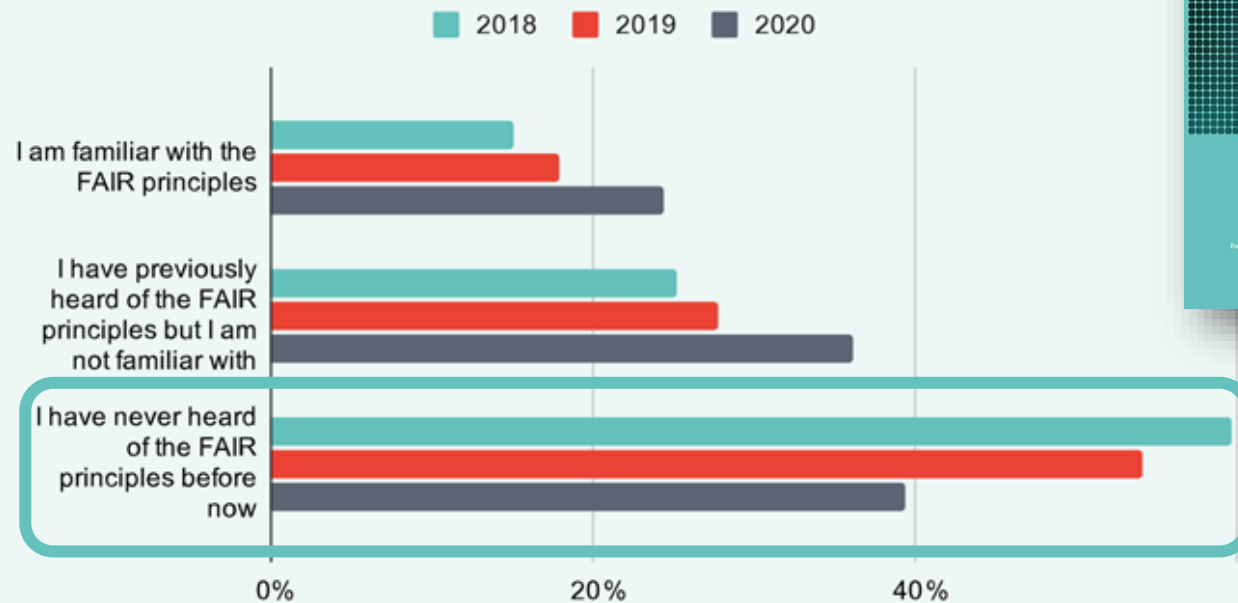
LA GESTIONE DEI DATI FAIR È PREVISTA (ART. 17 ANNEX 5) E IL DMP VA SINTETIZZATO IN UNA PAGINA NELLA SEZIONE SCIENTIFIC EXCELLENCE (OLTRE CHE PRESENTATO COME DELIVERABLE)

2. DATI FAIR BY DESIGN



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

INTEROPERABLE

- STANDARDS
- ONTOLOGIE

IL TUTTO, LEGGIBILE DALLE MACCHINE

ACCESSIBLE

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

REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE

...FAIR significa anche

FINDABLE

- RIDUCE IL RISCHIO DI PERDITA
- FA RISPARMIARE TEMPO

INTEROPERABLE

- FAVORISCE RICERCHE INTERDISCIPLINARI

I SINGOLI «BLOCCHI» DELLA RICERCA POSSONO ESSERE RICHIAMATI CON L'IDENTIFICATIVO, SENZA PERDERE TEMPO A RISCRIVERE (ES. UN PROTOCOLLO)

The FAIR Principles in a nutshell

Apr. 27, 2021

Findable



Accessible



Interoperable



Reusable



- Globally unique, resolvable, and persistent identifiers
 - *To retrieve and connect data*
- Community defined descriptive metadata
 - *To enhance discoverability*
- Common terminologies
 - *To use the same term mean the same thing*
- Detailed provenance
 - *To contextualize the data and facilitate reproducibility*
- Terms of access
 - *Open as possible, closed as necessary*
- Terms of use
 - *Clear licences, ideally to enable innovation and reuse*

ACCESSIBLE

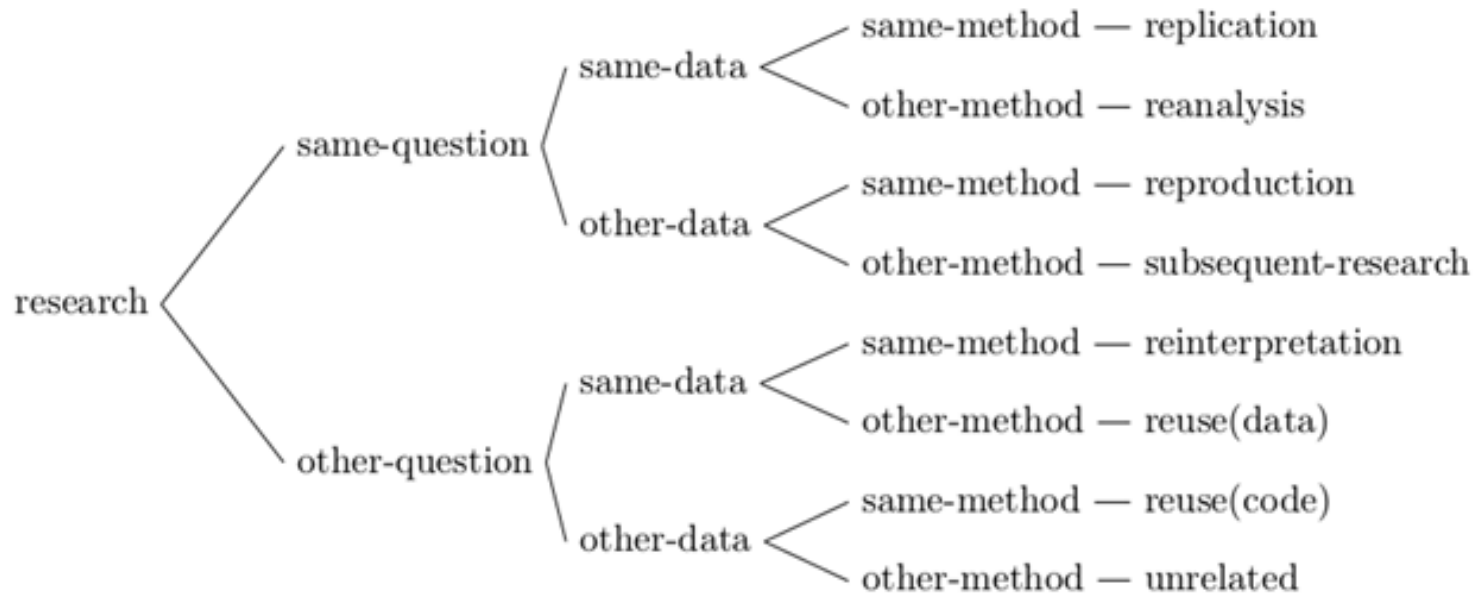
- CONSERVA I DATI ORA E PER IL FUTURO

REUSABLE

- EVITA SPRECO DI RISORSE
 - CREA FIDUCIA

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]

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.

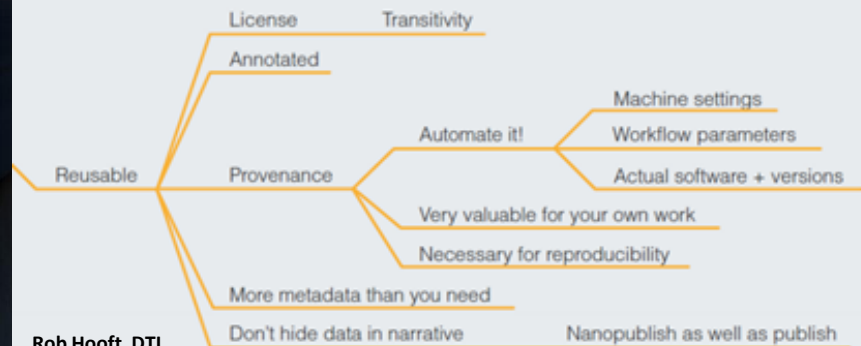
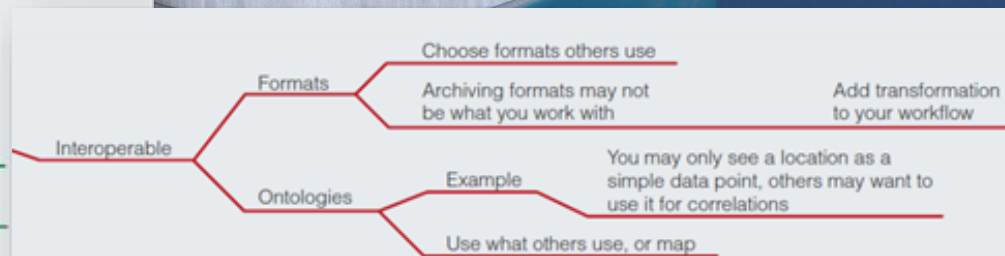
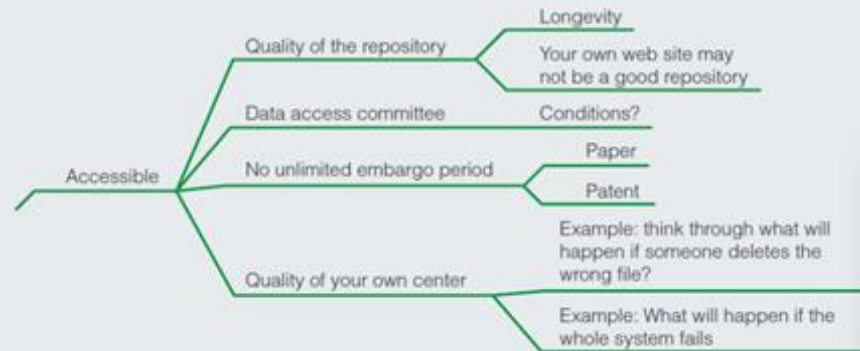
TO BE RE-USABLE:

- 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 principles



FAIR/Open

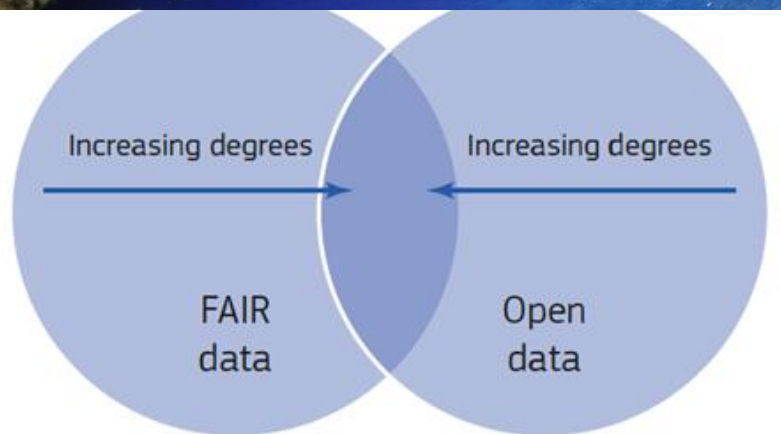


Figure 4. The relationship between FAIR and Open



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

4. ...and what FAIR is not

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 publication of the FAIR Principles states the authors' position clearly: "The release of the FAIR Principles is a landmark in the data publication autonomously, and the FAIR Principles. Computers are now able to process data autonomously, and recent surveys, the time reported for dealing with discovering and reusing data has been pegged at 80% [19]. Were the time to deal with FAIR data and services, as is today. The avoidance of time-consuming data stewardship. To serve this potential, data and services should be actionable wherever possible.

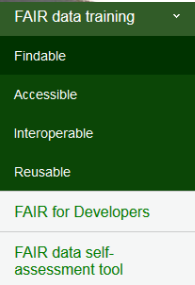
FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles,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

3. What FAIR is...

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 of data (the list 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 / DSI / DTU / CBS / AU / KU / Helsekyst / 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 / DSI / DTU / CBS / AU / KU / Helsekyst / 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 / DSI / DTU / CBS / AU / KU / Helsekyst / 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

Issue title: FAIR Data, Systems and Analysis

Guest editors: Paul Groth and Michel Dumontier

Article type: Position Paper

Authors: Lamprecht, Anna-Lena^a; ¹ | Garcia, Leyla^b ¹ | Kuzak, Mateusz^c; ¹ | Martinez, Carlos^d ¹ | Arcila, Ricardo^e ¹ | Martin Del Pico, Eva⁹ ¹ | Dominguez Del Angel, Victoria^h ¹ | van de Sandt, Stephanieⁱ ¹ | Ison, Joni¹ ¹ | Martinez, Paula Andrea^k ¹ | McQuilton, Peter^l ¹ | Valencia, Alfonso^m; ¹ | Harrow, Jennifer⁹ ¹ | Psomopoulos, Fotis⁹ ¹ | Gelpi, Josep LL⁹; ¹ | Chue Hong, Neil⁹; ¹ | Goble, Carole⁹ ¹ | Capella-Gutierrez, Salvador⁹; ¹

2. Software is not data

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

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

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

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

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

FAIR software recommendations

#1 USE A PUBLICLY ACCESSIBLE REPOSITORY WITH VERSION CONTROL

WHY THIS IS IMPORTANT

Why public

Developing scientific software in publicly accessible repositories enables involvement of users, helps build collaborations, contributes to the reproducibility of results generated by the software, facilitates software reusability, and contributes to improving software quality. Taken together, this ensures that your software has the best chance of being used by as many people as possible while promoting transparency.

Why version control

Using a version control system allows you to easily track changes in your software, both your own changes as well as those made by collaborators. There are many flavors of version control systems, ranging from older systems such as CVS to more modern ones such as Git, Mercurial, and Bazaar. By choosing a version control system to use GitHub, GitLab or Bitbucket, you'll even

HELP ME CHOOSE

Git is the most feature-rich, most modern and most popular version control system. We heartily recommend you use it for all your version control needs. If you are not familiar with Git, use it in combination with GitHub.com, Bitbucket.com or GitLab.com.

- [How to get started with git](#)
- [Choosing a platform for your software project](#)
- [GitHub.com](#)
- [BitBucket.org](#)
- [GitLab.com](#)

FAIR software EU

#2 ADD A LICENSE

WHY THIS IS IMPORTANT

Any creative work (including software) is automatically protected by copyright law when the software is available via code sharing platform. However, you can use it unless they are explicitly granted permission. A software license, which defines the set of rules and conditions under which you can use the software. Finally, be aware that you, as the creator of the software, may not be a copyright owner of the code you are using. The holder of a work is the employer (or hiring party) and not the creator.

HELP ME CHOOSE

We recommend you stick to one of the more popular licenses. Licenses are typically written by lawyers, the license text is precise and legally binding. The widespread use of the more popular licenses means that people who understand how the license of the law works.

#3 REGISTER YOUR CODE IN A COMMUNITY REGISTRY

WHY THIS IS IMPORTANT

For others to make use of your work, they need to be able to find it. Registries are like the yellow pages for software – registering your software makes it easier for others to find it, particularly through the use of search engines like Google. Community registries typically employ metadata to describe your software. With metadata, search engines are able to get some idea of what your software is about, what problem it addresses, and what domain it belongs to. In turn, this helps improve the ranking of the software in the search results. Metadata means better ranking.

HELP ME CHOOSE

Community registries come in many flavors. Choosing the one that best fits your needs can be tricky. Here are some things to think about:

- How much traffic does the community registry get?
- Is the community registry targeting the audience you are interested in?
- What metadata does the community registry offer? This is often found in the documentation of the registry, but you can also see a tool like the [OpenLink Structured Data Sniffer](#). Alternative engines have tooling like the [Structured Data Testing Tool](#) to help you understand how they perceive a given website.

Finally, ask a couple of colleagues which registries they would use if they were looking for software like yours.

[VIEW ALL RESEARCH SOFTWARE REGISTRIES](#)

FIVE RECOMMENDATIONS FOR FAIR SOFTWARE

LET'S GO! →

ENDORSE



WHAT'S FAIR

ABOUT US

#4 ENABLE CITATION OF THE SOFTWARE

WHY THIS IS IMPORTANT

Citation helps software developers be recognized for their work. Citation is an integral part of scientific accountability and reproducibility. Accurately citing software is inherently more difficult than citing a paper, even seemingly trivial things such as identifying the version of the software used. It is therefore convenient if software developers themselves provide the information necessary to enable citation.

HELP ME CHOOSE

[CodeMeta](#) and the [Citation File Format](#) were specifically designed for software and will likely meet your needs. For either one, you will need to add citation metadata, which you then distribute with your software.

Initialize your CITATION.cff files [here](#).

#5 USE A SOFTWARE QUALITY CHECKLIST

WHY THIS IS IMPORTANT

Checklists help you write good quality software. What exactly constitutes 'good quality' depends on the specific application of the software, but typically covers things like documenting the source code, using continuous testing, and following standardized code patterns.

HELP ME CHOOSE

There are many checklists available. We find that the most useful checklist is those that

1. Allow for a granular evaluation of a software package, as opposed to just pass or fail
2. Explain the rationale behind each item in the checklist

Research software

Netherlands
eScience center

E-sciencecenter

Enabling digitally enhanced research

The Netherlands eScience Center is the national center for the development and application of research software. Our software and tools enhance the use of digital methods in academic research, empowering researchers across all disciplines.

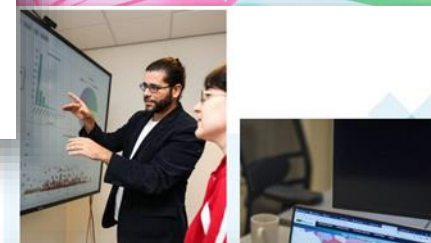
0



Our software

Our open-source software pushes the boundaries of academic research by allowing researchers to take full advantage of large amounts of complex data and powerful digital technologies.

[Browse our software solutions](#)



Start typing here to search for software

Tags -

- ☐ Big data (35)
- ☐ GPU (5)
- ☐ High performance computing (28)
- ☐ Image processing (5)
- ☐ Inter-operability & linked data (13)
- ☐ Machine learning (15)
- ☐ Multi-scale & multi model simulations (7)
- ☐ Optimized data handling (24)
- ☐ Real time data analysis (6)
- ☐ Text analysis & natural language processing (15)
- ☐ Visualization (25)
- ☐ Workflow technologies (29)

Organizations -

- ☐ Activinsights Ltd (1)
- ☐ Alfred Wegener Institute (2)

Netherlands eScience Center Python Template

Generic template for Python packages, so you can spend less time setting up and configuring, and comply with the Netherlands eScience Center Software Development Guide from the start.

23 hours ago

★ Featured

Noodles

Task-based parallel programming model in Python that offers the same intuitive interface when running complex workflows on your laptop or on large computer clusters.

17 days ago

★ Featured

Xenon

If you run codes remote machines to do your

GGIR

Converts raw data from wearables into insightful reports for researchers investigating human daily physical activity and sleep.

7 days ago

★ Featured

Kernel Tuner

Kernel Tuner greatly simplifies the development of highly-optimized and auto-tuned CUDA, OpenCL, and C code, supporting many advanced use-cases and optimization strategies that speed up the auto-tuning process.

28 days ago

★ Featured

mcfly

Notice, you find a suitable neural network



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

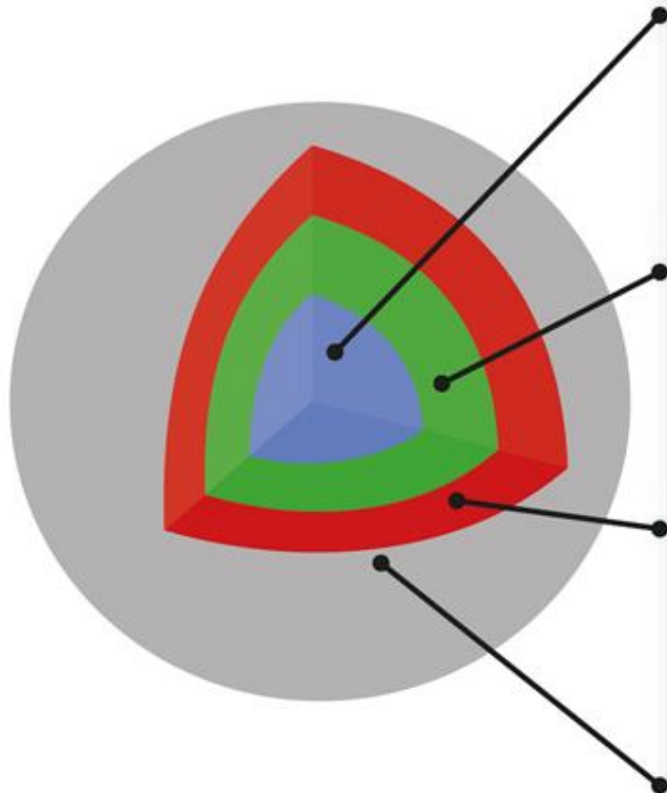
Investment in FAIR

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.

6 raccomandazioni

	EOSC	Research funders	Institutions	Policy-makers	Coordination fora	Standards bodies	Data service providers	Publishers
1. Fund awareness-raising, training, education and community-specific support	√	√	√					
2. Fund development, adoption and maintenance of community standards, tools and infrastructure	√	√			√	√	√	
3. Incentivise development of community governance	√	√			√			
4. Translate FAIR guidelines for other digital objects	√	√		√	√	√		
5. Reward and recognise improvements of FAIR practice	√	√	√	√				
6. Develop and monitor adequate policies for FAIR data and research objects	√	√	√	√				√

Recommendation 1: Fund awareness-raising, training, education and community-specific support.

Stakeholders: EOSC, Research funders, Institutions

Rationale: Community-specific actions are needed because arguments and solutions which works for one community might not be the key drivers for another. Raising awareness is needed at all levels – from individual researchers through heads of institutions to policymakers – but in order to be meaningful it must be based on adequate, community-specific arguments. Awareness raising, training, education and providing dedicated community-specific support take time and effort and thus such actions need to be financially supported. Funding pilot projects might be a useful mechanism to facilitate this.

Example: An initial pilot at TU Delft to fund data stewards with disciplinary knowledge helped communities realise the importance of FAIR practices, foster best practices and prompted them to appoint their data stewards as

Oct. 2020



Six Recommendations for Implementation of FAIR Practice

by FAIR in Practice Task Force of the European Open Science Cloud FAIR Working Group

Independent Report

EOSC Executive Board
10th March 2020

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

<https://www.nature.com/articles/sdata201618>

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 it describes
- 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 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 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

E.Schultes, 2019

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 - tecnicamente



A framework guiding FAIRification

The Three-point FAIRification Framework provides practical "how to" guidance to stakeholders seeking to go FAIR.

Moreover, by following this framework, stakeholders can rest assured that their efforts toward FAIRification will be optimally coordinated with the efforts of other stakeholders in the GO FAIR community. The three-point framework maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services.

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in **Metadata for Machines (M4M)** Workshops.
- The re-usable metadata schemata produced in the M4M compose part of the larger **FAIR Implementation Profile (FIP)**.
- The FAIR Implementation Profile in turn guides the choice and configuration of FAIR infrastructure, for example the use of **FAIR Data Points (FDP)** or **FAIR Digital Objects (FDO)** which contribute to a global Internet of FAIR Data and Services.

GO FAIR
**Metadata
4 Machines**

**FAIR
Implementation
Profile**

**FAIR
Data
Point**

1. METADATI (4 MACHINES)
2. FAIR IMPLEMENTATION PROFILES
3. FAIR DATA POINTS

[FAIR all'opera]

FAIR DATA POINT VISITATI
DA VIRTUAL MACHINES
I DATI NON SI MUOVONO
(CONFORMI GDPR)



Foundation. With a sense of urgency driven by the rapid developments on COVID-19 we came together to launch a GO FAIR Implementation Network to address the immediate challenges. For this epidemic, unfortunately, we have to 'FAIRify' COVID-19 data *'after the fact'* and use Chinese, Dutch, Swedish, etc. and English electronic (or even hand-written) health records to create proper FAIR data. The FAIRification will initially focus on the Clinical Research Form (CRF) model following the WHO standards. Multiple IN partners will create input forms that make it easy for local caregivers to create FAIR-CRF data in real time as a first step. As a second step, we will jointly develop (via online work sessions) localized **FAIR Data Points (FDP)**. FDP is a FAIR data repository with 'docking' capabilities as a 'station' for 'trains' (virtual machines (VMs)) that come to 'visit' the data locally, with a specific question to ask. The local data custodian (frequently a hospital or centre for disease control and prevention type of institution) grants permission to VMs to ask the question / run analyses. As the personal data of patients never leaves the underlying database of the local institution, GDPR issues are largely accommodated and in this way data can be 'shared' or rather 'visited' without violating any patient rights and, in the case of a disease outbreak, also governed by the laws and policies of the individual jurisdictions in which the outbreak manifests.

Trains (VMs) can visit multiple local FAIR Data Points to get their questions answered. For more information on the underlying technological approach please visit the [Personal Health Train IN pages](#). The data stewardship aspects of FAIR data will be addressed wherever possible with the [Data Stewardship Competence Centres IN](#).

[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)

SARÀ
PRESENTATO A
OAI12 (SETT.)



FAIR Data steward

Education core content

This 1-year degree should build upon students' educational/job background through domain specific data knowledge and leverage with theoretical and practical competences.

The education can be viewed as a Data Steward specialisation within the domain of their previous degree/jobs. The education contains **60 ECTS** and is expected to finish with a 15 ECTS project.

Preliminary Content

The 60 ECTS should be distributed among the following main areas:

- 22,5-30 ECTS: IT competences – including computational thinking, data modelling, data management, data harvesting, cleaning, and storing, infra-structure (storage & compute). An introduction to data science, machine learning, and their derived data needs.
 - 7,5-15 ECTS: Legal and ethical competences – including GDPR, FAIR, data security, and data & AI ethics.
 - 7,5-15 ECTS: Domain specific data competences – including knowledge about data, infrastructure, and practice within the students primary domain, e.g., health, life-science, finance/fintech, or the public sector.
 - 15 ECTS: Graduate project (possibly in collaboration with academia, industry, or the public sector)
- Competences such as project management, communication skills, and change management should be

Competence Profile

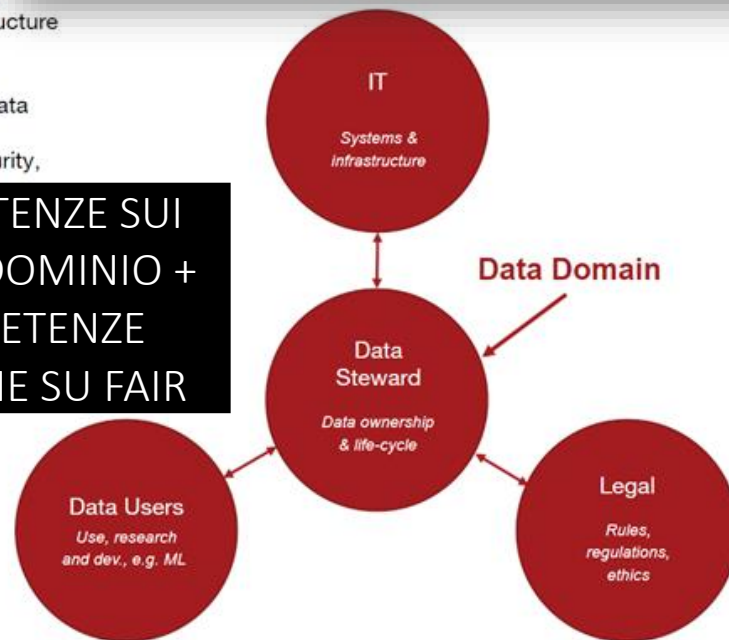
A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

The future Data Steward must assume ownership and responsibility for data, data quality, and the data life-cycle as their primary function. They should ensure collaboration and coherence between IT competences, quality assurance, security, rules & regulations, and facilitate the application and use of data externally in the organisation.

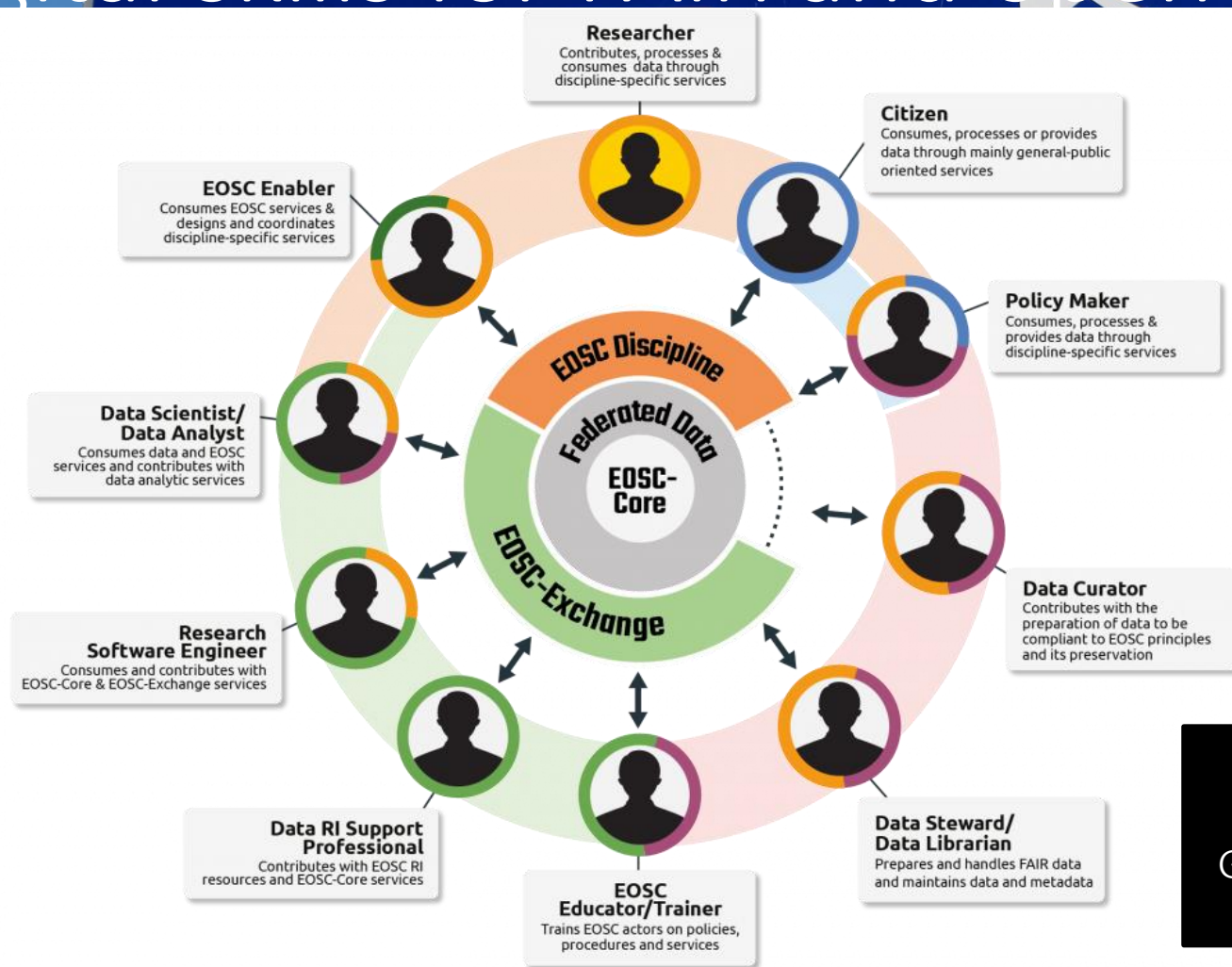
Competence profile examples

- Domain-specific data understanding
- Ability to ensure that structured and unstructured data is modelled, harvested, stored, and maintained in documented, and regulated fashion with focus and findability, accessibility, interoperability, and reusability.
- Competences to facilitate HPC (High Performance Computing) during development and research through handling of large-scale data in public and private enterprises.
- Understanding of and competences within legal, ethical and security aspects of data handling, data sharing, e.g., integrity and GDPR.

COMPETENZE SUI DATI DI DOMINIO + COMPETENZE TECNICHE SU FAIR



Digital skills for FAIR and open



INTORNO A
FAIR/EOSC
GIRANO DIVERSE
COMPETENZE



Digital science

Policymakers & funders Universities & research organisations Competence centres EOSC Association EOSC projects

Understand the importance of addressing gaps in provision of digital skills for FAIR and open science

Identify actors/roles that need to be supported in the EOSC ecosystem	Consider diversity of employment and training needs of staff	Identify knowledge to be taught and skills to be learned	Understand range of roles that initiatives must address	Ensure projects consider appropriate roles for both their staff and users
---	--	--	---	---

Consider how to support competence centres and encourage	Consider how to establish competence centres	Learn about governance, business models and alignment	Consider how to encourage and support compet-	Ensure projects consider FAIR and open science training
--	--	---	---	---

Recognise the need for investment in hubs for training resources	Know how to provide and access learning and training resources	Learn how to maximise interoperability, FAIRness and usage of learning and training resources	Facilitate interoperability and discovery of learning and training resources	Identify and engage key stakeholders, and build learning and training catalogues according to their goals and actions
--	--	---	--	---

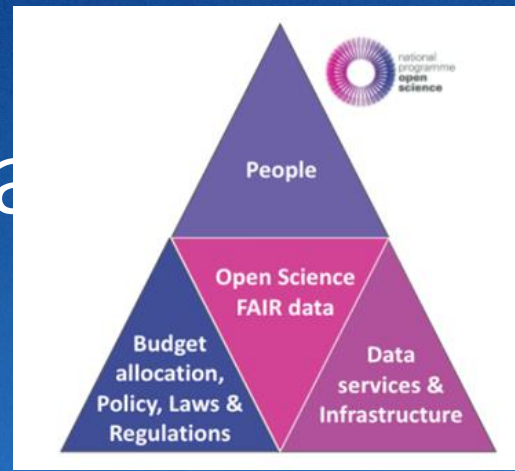
Identify actions to advance national policies on FAIR and open science	Consider how to encourage and/or implement national approaches	Understand the importance of aligning with national programmes and policies	Identify ways to increase international alignment	Not relevant
--	--	---	---	--------------



AZIONI RICHIESTE
 AI DIVERSI
 STAKEHOLDER
 ...LA PRIMA:
 CAPIRE
 L'IMPORTANZA!!!

Professionalizzazione e Inquadramento nelle Strutture (Non Legati al Singolo Progetto)

URGE PROFESSIONALIZZAZIONE
E INQUADRAMENTO NELLE
STRUTTURE (NON LEGATI AL
SINGOLO PROGETTO)



Professionalising data stewardship in the Netherlands: competences, training and education

Dutch roadmap towards national implementation of FAIR data stewardship

Feb. 1, 2021

Preamble: The urgency of a coherent approach towards professionalising data stewardship

“Invest 5% of research funds in ensuring data are reusable. It is irresponsible to support research but not data stewardship”, said Barend Mons recently in a Nature article¹. “Students in PhD programmes spend up to 80% of their time on ‘data munging’, fixing formatting and minor mistakes to make data suitable for analysis — wasting time and talent. With 400 such students,

Data stewardship is a new profession. To ensure a data steward’s position within a research institute, the profession should become part of HR planning of the organisation and include a vision on career development. In addition, research institutes should collaborate with policy makers and research-funding organisations to realise a common approach to recognise and reward data stewards as full members in research groups, and to make sufficient budgets available to maintain data steward positions in the long run.

The interplay between these three aspects determines whether researchers in a particular research setting are sufficiently equipped and supported to be able to perform data stewardship in the context of Open Science. By having a clear view on each of these aspects, it will be finally possible to answer the question that each research-performing institute should ask itself: **how many data stewards do we need where in the organisation with which competences?**

[supporto]



- ▶ una rete di esperti (disciplinari, settoriali, legali, ...);
- ▶ capacità di raccolta e disseminazione di buone pratiche, anche attraverso collegamenti alle attività dei propri membri e associati;
- ▶ possibilità di sfruttare relazioni internazionali consolidate per promuovere i risultati raggiunti a livello italiano e viceversa, valorizzando così l'attività nazionale in Europa;
- ▶ capacità di costruire una raccolta di risorse knowledge base sull'esempio di FAIRsFAIR;
- ▶ capacità di organizzare attività "hands on" e ambienti pre-ordinati per corsi o sandbox;
- ▶ diffusione di competenze legate ai FAIR data e alla scienza aperta attraverso l'organizzazione di corsi su OS e l'inserimento di nuovi insegnamenti e curricula grazie alla stretta relazione con le Università;
- ▶ incrementare la partecipazione e la presenza di Infrastrutture di Ricerca fisiche e non, che posseggono competenze nel FAIR Data Management;
- ▶ favorire l'accesso alle risorse condivise nei workflow di ricerca;
- ▶ Portale Open Science CNR, che sarà lanciato a breve e pensato come una sorta di one-stop shop con un catalogo di risorse accessibili online (dai documenti ufficiali alle risorse di training e informative);
- ▶ possibilità di visualizzare informazioni su andamenti statistici messe a disposizione da OpenAIRE.

ICDI STA CREANDO
COMPETENCE
CENTER

FAIR data Forum

FAIRdata Forum

FAIRdata Forum supports the FAIR Competence Centre as an open space for users to post queries, answers, comments and events notifications pertaining to FAIR. The FAIRdata Forum was initiated by the FAIRsFAIR "Fostering FAIR Data Practices in Europe" project which has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558.



all categories ▾

Categories

Latest

Top

Category

Topics Latest

General

Topics that don't need a category, or don't fit into any other existing category.


3   [Welcome to the FAIRdata Forum!](#) 14h[How to register \[Como registrarse\]](#) Aug '20**Training Resources**

Training materials refers to any digital object used in a training context. That includes slide presentations, software tools, datasets, etc.

-  [FAIRsFAIR Training Materials](#)
-  [Where to find training materials](#)
-  [Best practice for training materials](#)

13 [Introduction to Linked Data](#) 11d[Data access exercise](#) 29d[Data journals – publish and promote your data](#) 29d**Help Desk**

Do you have a question or need help related to FAIR? The Help Desk area of the FAIRdataforum is a place to post questions, share expertise and learn from the community.

-  [Queries/Answers](#)

0

User and Implementation Stories

A place to share and collate user or implementation stories.

0

Information Hub

The information hub contains background information for those new to FAIR. If you have ideas for new topics please add a comment and we will look into it.

-  [I am new to FAIR - what is it?](#)

1 [Recommended reading](#) Sep '20**Fair News**

Share and discuss FAIR-related news

0

Site Feedback

Discussion about this site, its organization, how it works, and how we can improve it.

1 [Enable users to register or sign into the forum using their ORCID](#)

FAIR cookbook

RICETTE PER
RENDERE I DATI
FAIR
[AGGIORNATO
PERIODICAMENTE]

 **FAIR**cookbook

FAIR Cookbook

FAIR cookbook

 Search this book...

FAIR Cookbook

Introduction

Assessing FAIR

Infrastructure for FAIR

Improving Findability

Improving Accessibility

Improving Interoperability

Improving Reusability

Making IMI data FAIR

Glossary of FAIR things

Help

Miscellaneous

Powered by **Jupyter Book**

The FAIR Cookbook overview

- The FAIR Cookbook is an activity of the **FAIRplus project** funded by the **Innovative Medicine Initiative (IMI)** drawing under grant agreement #802750 (2019-22).
- The FAIR Cookbook is a project aiming at collating protocols for making data FAIR and provide examples of IMI dataset FAIRification.
- The FAIR Cookbook is open source and licensed under CC-BY-4.0
- The FAIR Cookbook is written using Markdown and deployed using **jupyter-book**
- the FAIR Cookbook is hosted on github. The repository hosts documentation, known as **FAIR recipes**, and supporting code in the form of **jupyter notebooks** about **FAIRification processes** and the content will be released regularly (quarterly) in order to reflect the progress made by the project and the various working groups, which bring together **academic** and **industry** partners.

Persona-based content browsing



Helmut

Architect & Engineer



Jean

Data Scientist



Fulvia

Data Curator



Wang Ju

Head of Clinical

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

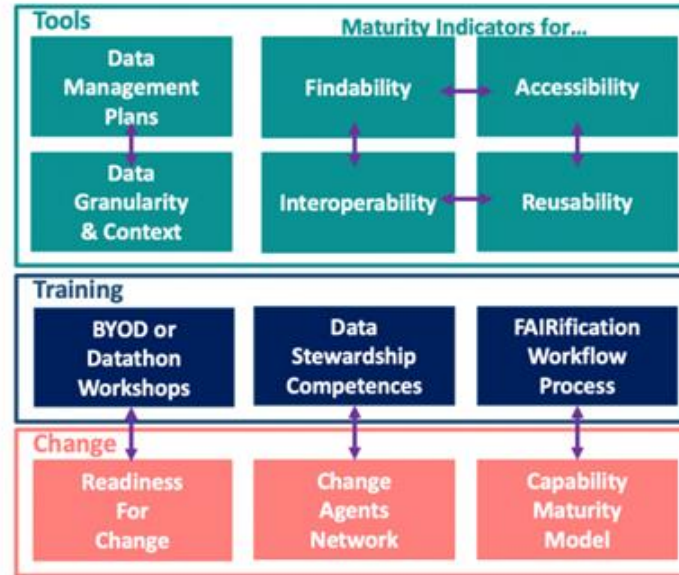
FAIR Toolkit

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 >](#)

[suggerimenti per «convincere»]

Readiness for Change

Find out how the data stakeholders can prepare for the changes necessary to make data FAIR.

- Readiness for change provides the means to engage a wider group of staff throughout an organisation.

OVERVIEW

The practicalities of FAIR implementation will involve much change in the work setting of industry life science. This method describes the importance of readiness of data stakeholders to engage with changes that will be necessary for making data FAIR, being Findable, Accessible, Interoperable and Reusable [1]. It places emphasis on people rather than technology through focus on understanding how the data stakeholders are ready for change.

Change management is a collective term for all approaches to prepare, support and help individual, teams and organisations in making organisational changes. Implementation of FAIR data management by life science organisations is likely to require radical and lasting change. A managed approach to change is designed to 1) minimise resistance to change, 2) increase speed of change and 3) increase the likelihood of sustained change. Numerous books and articles on change management can be found, which include numerous models and acronyms such as 1) 8-step Process for leading change [2] and 2) ADKAR which is *Awareness* of the need to change, *Desire* to participate and support change, *Knowledge* of how to change, *Ability* to implement required skills and behaviours and *Reinforcement* to sustain change [3].

This method is focussed readiness to change which is likely to be critical for better data management shaped by the FAIR guidelines. Readiness of all stakeholders for the necessary changes will provide the means to engage a wider group of staff throughout an organisation.

At a Glance

RELATED USE CASES

- All of the use cases of this FAIR Toolkit

RELATED METHODS

- Change agents
- Data Capability Maturity Model

SETTING

- Get "buy-in" from all stakeholders for the necessary changes as early as possible

TEAM

- Data Stewards
- Scientist collecting the data and metadata
- Management

TIMING

- Early and sustained

DIFFICULTY

- Medium to high

HOW TO

FAIR Toolkit

Below are a set of questions designed to understand the readiness of key stakeholders impacted by implementation of FAIR data management. These questions are accompanied by example answers to illustrate the application to FAIR data management.

Q1. What has the past experience been with change initiatives?

- **Potential scepticism** can be countered by **clear communication** of benefits, sharing workshops and early success.

Q2. Who are the key groups / people to be impacted by a change initiative?

- **Scientists** who collect or generate scientific data from laboratory experiments in house and external sources.
- **Middle managers** responsible for productivity goals.
- **Senior managers** for budget and strategic goals.
- **External parties** like technology providers, which have to adapt FAIR principles

Q3. How well do you feel the rationale for change has been communicated?

- Clearly communicate **why change brings benefits** to the scientists and workshops for sharing best practice.

Q4. How do relevant methods and principles fit with the strategic goals of your key groups?

- Need to show how the **benefits** support **performance and productivity goals**

Q5. Who are the key groups / people to be most resistant to a change initiative?

- **Middle managers** are likely to be most resistant to change. Their concern will be delivery of tactical goals.
- **Scientists** will be resistant if too much "red tape". Must have easy mechanisms for implementation.
- **Senior managers** must provide adequate budget for data management. Important to show value of change early.

Q6. What do you see as the major risks for this change initiative?

- Implementation requires **planning, education, tools and budget**. An important part of "business as usual".
- No "magical wand" after thought. Important to show early success for sustainability, to avoid "forgetfulness".

Change agents

DATA STEWARDS SONO CHIAVE DI
VOLTA NEL CAMBIAMENTO
(E SE SI MOSTRA CHE SI PUÒ FARE E
CHE I BENEFICI ARRIVANO, CHI
GOVERNA SI CONVINCE)

Change Agents

Learn how change agents, such as data stewards, play an important role to support FAIR data management and application.

- A network of change agents coordinate data management across the organisation to support the necessary changes.

OVERVIEW

The practicalities of FAIR implementation will involve much change in the work setting of industry life science. This method describes how data stewards can act as change agents are critical to support for making data FAIR, being Findable, Accessible, Interoperable and Reusable [1]. It places emphasis on people rather than technology through focus on those stakeholders who are act as the network of change agents.

Change management is a collective term for all approaches to prepare, support and help individual, teams and organisations in making organisational changes. Implementation of FAIR data management by life science organisations is likely to require radical and lasting change. A managed approach to change is designed to 1) minimise resistance to change, 2) increase speed of change and 3) increase the likelihood of sustained change. Numerous books and articles on change management can be found, which include numerous models and acronyms such as 1) Kotter's 8-step process for leading change [2] and 2) the ADKAR model which is *Awareness* of the need to change, *Desire* to participate and support change, *Knowledge* of how to change, *Ability* to implement required skills and behaviours and *Reinforcement* to sustain change [3].

This method is focussed on change agents who are likely to be critical for successful implementation of FAIR data principles and guidelines. A network of change agents such as data stewards can act in a coordinated manner across the organisation to support the necessary changes actively during daily work. In particular, they will also be important for understanding and communication of what aspects of change are working well and what requires attention. See the training method *and data stewardship* for further details about the responsibilities and

At a Glance

RELATED USE CASES

All of the use cases of this FAIR Toolkit

RELATED METHODS

- Data stewardship
- Readiness for change

SETTING

- Support the whole life cycle of FAIR data management

TEAM

- Data Stewards
- Scientist generating or collecting the data and metadata
- Management

TIMING

- Ongoing and sustained

DIFFICULTY

- Medium to high

HOW TO

FAIR toolikt

Below are a set of questions designed to build and harness a network of change agents who support the change actively as an important facet of their daily work. They will be able to understand and communicate what is well and what requires attention. These questions are accompanied by example answers to illustrate how change agents relate to the implementation of sustained FAIR data management.

Q1. Who would you identify as key change agents?

- **Data steward:** Appointed to each important group who will be a senior scientist familiar with the concepts and process of data stewardship.
- **Middle managers** must support common data policies which can be reused.
- **Senior managers** must invest appropriate levels of budget for data management training, workshops and data service provision.
- **Support service staff** in Business Technology and Informatics functions are also likely to be important.

Q2. How can the change agents help to drive adoption of the change?

- The **change agents**, especially the **data stewards** supported by **management** need to facilitate new or improved business processes.
- These will foster the attitude that **data sets and corresponding metadata are valuable corporate assets** which must be managed effectively.

Q3. How can the network of change agents help to overcome barriers to change?

- The **network of data stewards** will facilitate implementation of FAIR data management at an optimal level of capability, determined through feasibility studies.
- Iterative application of FAIR maturity indicators will show opportunities for improvement, and the resulting benefits.
- Success will include more reuse of the data, better reproducibility and realisation of value from data and more time for insightful data analytics.

Use cases will show case such benefits. This will be important to communicate the value of the

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



FAIRification workflow

PER RENDERE FAIR I DATI
OCCORRONO COMPETENZE
MULTIPLE
... E I DATA STEWARDS
SONO AL CENTRO

January 01 2020 Jan 2020
A Generic Workflow for the Data FAIRification Process
Anirika Jacobson, Rajaram Kalyanasundar, Luis Olavo Santos da Silva Santos, Bernd Mone, Erik Schultze, Marco Risse, Mark Thompson
Data Intelligence (2020) 2 (1-2): 56-65
<https://doi.org/10.1002/dint.v.00028>

Data FAIRification requires different types of expertise and should therefore be carried out in a multidisciplinary team guided by FAIR data steward(s). The different sets of expertise are on i) the data to be FAIRified and how they are managed, ii) the domain and the aims of the data resource within it, iii) architectural features of the software that is (or will be) used for managing the data, iv) access policies applicable to the resource, v) the FAIRification process (guiding and monitoring it), vi) FAIR software services and their deployment, vii) data modelling, viii) global standards applicable to the data resource, and ix) global standards for data access. A good working approach is to organize a team that contains or has access to the required expertise. The core of such a team may be formed by data stewards, with at least expertise of the local environment and of the FAIRification process in general.

FAIRification v

FORNISCE I TEMPI,
TEAM NECESSARIO
E USE CASES

FAIRification Workflow

Find out how a generic workflow can be deployed by workshops or action team to make important datasets FAIR.

- FAIRification as a retroactive workflow is common at this time
- FAIRification by design (data "born" FAIR) is far more desirable for the future

OVERVIEW

FAIR Toolkit

This method is a step-by-step, generic workflow for making data FAIR, often known as "FAIRification". It can be deployed by workshops or a team of subject matter experts guided by FAIR data stewards as published by Jacobsen *et al* at Leiden University and GO-FAIR International Support & Coordination Office [1]. These authors describe how it should be applicable generically to any type of data (and metadata) and how it has emerged from a series of "Bring Your Own Data" (BYOD) workshops. A related methodology has been developed by the IMI FAIRplus project who have deployed FAIRification squad teams [2] and collating specific protocols as "recipes" for FAIRification in the FAIR cookbook [3].

The figure below shows how the FAIRification workflow is divided into three "phases": Pre-FAIRification, FAIRification, and Post-FAIRification (dark grey boxes) that are further specified by "steps" indicating typical aspects of practical FAIRification (light grey boxes): 1) identify FAIRification objectives, 2) analyse data, 3) analyse metadata, 4) define the semantic model for data (4a) and metadata (4b), 5) make data (5a) and metadata (5b) linkable 6) host FAIR data and metadata and 7) assess FAIR data and metadata for meeting the objective.

The order of the steps in the workflow is not strict, can be iterative and some of the FAIRification steps maybe optional. The practical deployment of this workflow is shown in the figure below which will depend on the data source, use case, domain or FAIRification objective.

Pre-FAIRification

Post-FAIRification

At a Glance

RELATED USE CASES

- FAIRifying collaborative research on Real World Data - The Hyve
- Prospective FAIRification of data on the EDISON platform - Roche

RELATED METHODS

- BYOD datathon workshops
- Data Management Plans
- Data granularity and context
- Data Capability Maturity Model
- FAIR Maturity Indicator methods

SETTING

- Applicable to existing datasets

TEAM(S)

- Data Steward(s)
- Relevant Subject Matter Expert(s)

TIMING

- 2 days for pre and post-FAIRification
 - These might be workshops
- Time for FAIRification depends on
 - the FAIR objective
 - the complexity of the dataset(s)
 - the complexity of the use case

DIFFICULTY

- Low for pre and post-FAIRification
- Medium to high for FAIRification

RESOURCES

- FAIR Cookbook
- COLID - Quick start

Prospective FAIRification of Data on the EDISON platform - Roche

Roche embeds data standards and quality checks to harmonize, automate and integrate very heterogeneous and complex processes.

- Self-contained micro services deliver performance and scalability
- Scalable and flexible for data models in clinical and non-clinical

OVERVIEW

What if we can help make our data FAIR even before it enters our company? As a large pharmaceutical company, we asked ourselves this question as we realized that we spend too much time to find, access and pool data and that retrospective FAIRification of data is not sustainable for the long-term. We are building the EDISON platform to enable prospective FAIRification of data at the point of entry to Roche, by harmonizing, automating and integrating very heterogeneous & complex processes across multiple departments, building in data standards and quality checks at every step of the process.

The EDISON platform is built as an ecosystem of self-contained micro services to ensure maximum performances, scalability and low maintenance. The current scope of EDISON is clinical non-CRF data but the platform is scalable and flexible to cover large variety of data models, both clinical and non-clinical.

At a Glance

TEAM

- Project Manager
- Business Analysts
- Validation Team
- Software Developers

TIMELINE

- A validated productive release with additional functionality every 3 months
- The project under its current scope will run for 12 months

BENEFITS AND DELIVERABLES

- Enabling prospective FAIRification eliminating the need for resource

Data capability maturity model



PER LE ISTITUZIONI: MATURITÀ NEL PRODURRE DATI FAIR

HOW TO FAIR Toolkit

CMM applied to FAIR data management determines the maturation steps on a pathway to support Findability, Accessibility, Interoperability and Reusability at an optimal level of granularity for machines and humans.

The maturation steps or stages of FAIR data CMM are as follows:

1. **Initial:** Data reuse is not possible outside of the project or department who produced the data sets. No long-term solutions for data sustainability and access.
2. **Repeatable:** Data usage is limited, only possible with help of experts who are involved in the project and requires manual effort. Domain experts help are required to interpret the data. Data access is governed mostly by the project owners.
3. **Defined:** FAIR data sets can be utilized by other parties with a minimal effort. Organizational and community standards are utilized, variations are documented. Linking data sets can be achieved with some mapping effort. Access and sustainability processes are well defined.
4. **Managed (Capable):** Data is linked. Metadata follows (canonical) community standards, when not the variations are explicitly documented and mapped. Organization wide services are available for searching and accessing FAIR data sets. Machine accessibility and data linking is fully achieved.
5. **Sustainable (or Optimising for efficiency):** The data can be reused cross organizations/communities with a minimum of effort. FAIR data can be maintained FAIR throughout time. Usage of FAIR data is monitored.

The FAIR Maturity Indicators are applied to assess the starting level of FAIR maturity, as it is now. Following this, feasibility analysis is undertaken to take account of the different levels of data granularity: 1) data catalogue 2) data collection 3) data set 4) data content and 5) context for the data. See the Toolkit method on granularity and context for more detail about feasibility analysis. Selection of data for FAIR transformation should include satisfaction of scientific and business requirements or questions in the feasibility analysis to understand the target for a desired level of FAIR transformation. This will guide and inform the practical steps to improve the FAIRness of the data as part of a cycle of continuous improvement, as illustrated in **Figure 2** below.

Data Capability Maturity Model

Capability Maturity Model applied to data helps an organisation to determine the capability for making data assets FAIR.

- The method defines five levels of maturity for how an organisation makes and maintains FAIR data.

OVERVIEW

The phenomenal growth of Life Science data makes it more important than ever to invest in support for the life cycle of data management which is enhanced by making the data FAIR, being Findable, Accessible, Interoperable and Reusable [1]. This method applies the Capability Maturity Model (CMM) to the transformation of FAIR data which considers maturity levels as key process steps. This will help an organisation to determine the optimal level for FAIR transformation of data or making data FAIR by design.

The CMM method recognises five levels of maturity for making data FAIR data. It is used to develop and refine the process for software development, first described in *IEEE Software* [2] by Humphrey Watts. CMM has since been evolved and commercialised by Carnegie Mellon University as the CMMI Institute to better integrate with the strategic goals of an organisation [3]. CMM can aid in business processes generally and has also been used extensively worldwide in government offices, commerce, and industry [4]. It has been applied by the CMMI institute to provide the Data Management Maturity (DMM) program under commercial license [5]. Existing maturity assessment models applied to data stewardship has been reviewed as three types of maturity matrix for geographic and climatology data [6]. Most recently, CMM has been adapted by the FAIRplus IMI consortium [7] to improve an organisation's life science data management process, which is the basis for the method described here.

At a Glance

USE CASES

All of the use cases of this FAIR Toolkit

RELATED METHODS

- Data stewardship
- Change agents
- Readiness for change
- FAIRification workflow

SETTING

- An enabler for an organisation to determine its investment in capability to make and maintain FAIR data assets

TEAM

- Data stewards
- Scientists who collect the data
- Science management

TIMING



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



Training

Search for Research Data

Search the ANDS Site

Enter Keywords

Enter Keywords

About us

News and Events

Partners and Communities

Working with data

Online Services

Guides and resources

Working with data

The FAIR data principles

FAIR webinar series (Aug/Sep 2017)

FAIR data training

Findable

Accessible

Interoperable

Reusable

FAIR data training materials

ANDS | Working with data | The FAIR data principles




FAIR data training

[f](#)
[t](#)
[in](#)
[g](#)
[+](#)
[g+](#)
[e](#)

If you intend to run workshops on FAIR data, or include FAIR in training that you are already running here are some ideas and resources.

- A basic checklist (or more comprehensive breakdown) as a tool for discussing the components of FAIR
- 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
- Case studies of domain specific consideration of the principles

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



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

[home](#) | [news](#) | [events](#) | [programs](#) | [about](#)

FAIR aware



Data Archiving and Networked Services



FAIRaware

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.

- DOMANDE
- VERIFICA INTENZIONI
- VERIFICA CONOSCENZA
- FORNISCE INFORMAZIONI

1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository?

Selected datasets should be assigned a globally unique and persistent identifier (PID) so they can be located unambiguously by humans or machines on the web. 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), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK).

Identifiers are normally assigned by data repositories (or other service providers) when data and/or metadata are made available through their services. Repositories ensure that the identifier continues to point to the same data or metadata, according to the specified access terms and conditions. For example, you can search for data repositories providing DOIs on registries such as Re3data or FAIRsharing (see related databases).

It is worth noting here that 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.

[Want to know more?](#)

FINDABLE

1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository?

2. Are you aware that when you deposit a dataset in a data repository, you will need to provide some data (e.g. discovery metadata) in order to make the dataset understandable and reusable to others?

3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans?

- ☒ Yes
☐ No

To what degree do you intend to comply with this?

Very likely ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1
Very unlikely

FAIR aware

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 Principles² 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 FAIR4FAIR³ 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.

ACCESSIBLE

4. Are you aware that access to your dataset may need to be controlled and that metadata should include licence information under which the data can be reused? ⓘ ☐ Yes ☐ No

5. Are you aware that metadata should remain available over time, even if the data is no longer accessible? ⓘ ☐ Yes ☐ No

INTEROPERABLE

6. Are you aware that the metadata describing your datasets should use controlled vocabularies? ⓘ ☐ Yes ☐ No

REUSABLE

7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata? ⓘ ☐ Yes ☐ No




8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard? ⓘ ☒ Yes ☐ No

9. Are you aware that data should be deposited preferably in a file format that is open – to support reuse – and supported by the repository for long-term preservation? ⓘ ☐ Yes ☒ No

10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation? ⓘ ☐ Yes ☐ No

FAIR maturity evaluator

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

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

Scientific Data 6, Article number: 174 (2019) | Download Citation  | [Sept. 20, 2019](#)
13 Altmetric | Metrics 

- OGGETTIVO
- LEGGIBILE DALLE MACCHINE – COME I DATI FAIR

FAIR Evaluation Services


FAIR evaluation service

Resources and guidelines to assess the FAIRness of digital resources.

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



Import MI Tests

Import Maturity Indicators Tests as YAML  interface annotation

Get started



Create collections

Assemble Maturity Indicators Tests into community centered collections

Get started



Evaluate resources

Evaluate resources FAIRness against Collections of Maturity Indicator Tests

Get started

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

Philosophy of FAIR testing



FAIR METRICS GEN2 - IDENTIFIER PERSISTENCE

Status: Failure

Principle tested: F1

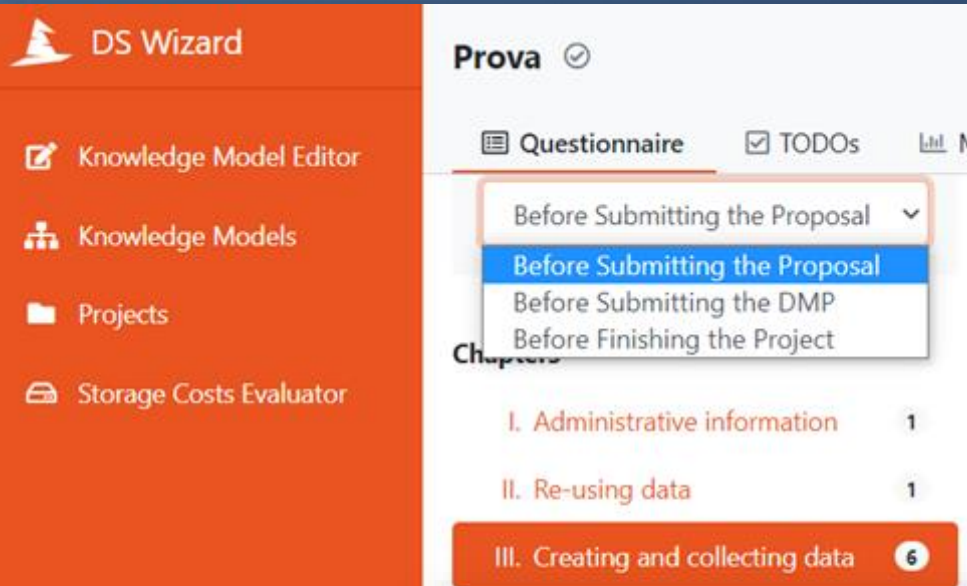
Description: Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema). For URLs that don't follow a schema in FAIRSharing we test known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

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

Test results

INFO: The metadata GUID appears to be a URL. Testing known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).
FAILURE: The metadata GUID does not conform with any known permanent-URL system.

FAIR Wizard



DS Wizard

- Knowledge Model Editor
- Knowledge Models
- Projects
- Storage Costs Evaluator

Prova

Questionnaire TODOs

Before Submitting the Proposal

- Before Submitting the Proposal
- Before Submitting the DMP
- Before Finishing the Project

Chapters

- I. Administrative information 1
- II. Re-using data 1
- III. Creating and collecting data 6



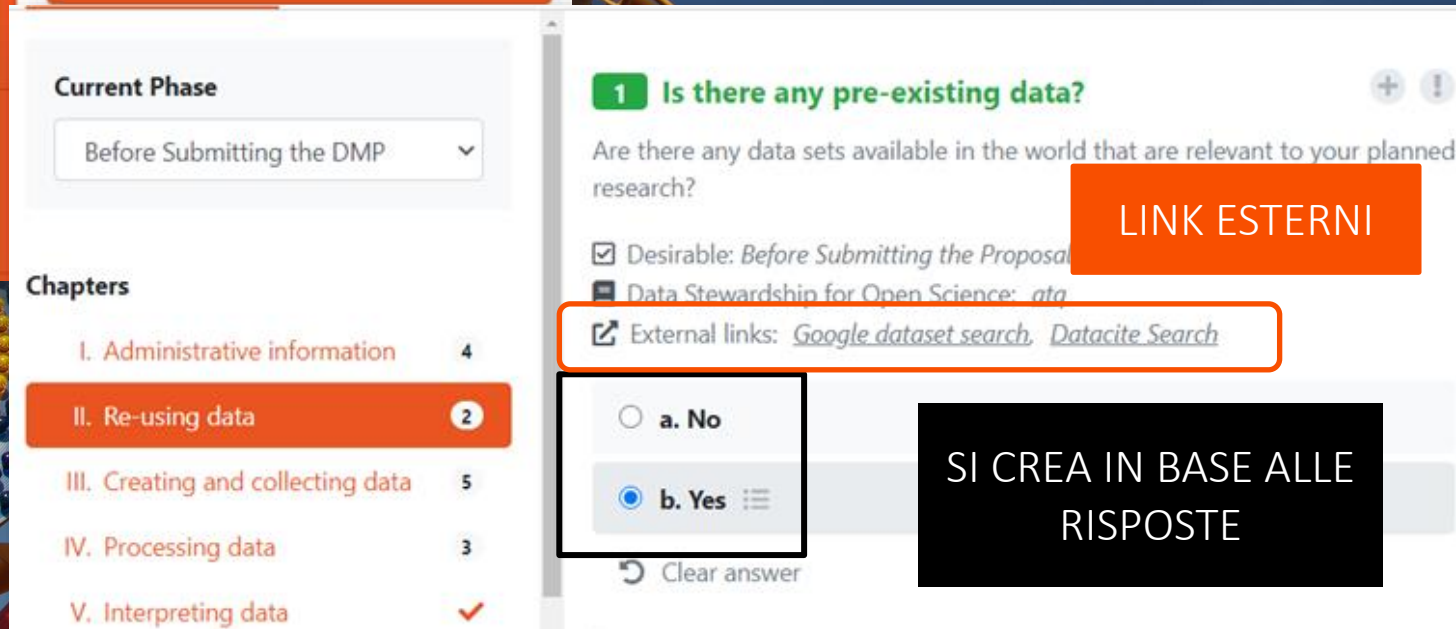
DSW

<https://ds-wizard.org/>

Data Stewardship Wizard

Create Smart Data Management Plans
for FAIR Open Science

Get started



Current Phase

Before Submitting the DMP

Chapters

- I. Administrative information 4
- II. Re-using data 2
- III. Creating and collecting data 5
- IV. Processing data 3
- V. Interpreting data

1 Is there any pre-existing data?

Are there any data sets available in the world that are relevant to your planned research?

☒ Desirable: Before Submitting the Proposal

☒ Data Stewardship for Open Science: [ata](#)

External links: [Google dataset search](#), [Datacite Search](#)

☐ a. No

☒ b. Yes

Clear answer

LINK ESTERNI

SI CREA IN BASE ALLE
RISPOSTE

Questionnaire **✓ TODOs** Metrics Preview Documents Settings

Current Phase
Before Submitting the DMP

Chapters

- I. Administrative information 4
- II. Re-using data 5
- III. Creating and collecting data 5**
- IV. Processing data 3
- V. Interpreting data ✓

1 What existing data formats/types will you be using?

Have you identified types of data that you will use that are used by others too? Some types of data (for example "images" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG" and "CSV" [comma separated values]) that help to make these data reusable. Are you using such common data formats?

Please make sure you list all the data types that are important for your project. You should make sure also to list the formats used in any data sets that you are re-using.

☒ Desirable: Before Submitting the Proposal

☒ Data Stewardship for Open Science: [nijx](#)

1.a.1 Data format/type

+ Add TODO

DS Wizard [Go to App](#)

CREA TO DO LIST

What's up?

Unless you do entirely novel types of research, there are likely to be multiple data formats around in which the types of data you generate may be captured, processed and formatted. Some of these may be 'exotic' and not used (anymore) by the majority of the community, which frequently means that they will be difficult to find, map, inter-operate and reuse. In addition, it is less likely that standard workflows will process these data formats. Especially in case the intention to use the data generated in combinatorial or integrated experiments with OPEDAS, the formatting of your data is extremely important. In many cases, data in proprietary or exotic formats can be munged and recreated into more commonly used formats, but these processes are very cumbersome and error-prone. It is therefore of the utmost importance to consult the expert community and get the data in the most optimal formats of further analysis and ultimately for reuse by your own group and others.

Do

- Always use community-compliant, supported and sustainable data formats whenever possible.
- Turn to experts to tell you what are the best formats to use for the particular data types you will create.
- Ensure you are prepared to answer questions on the use of the data (for instance, which workflows will they be subjected to).
- Choose the formats with the richest expression possibility. It is easier to leave things blank then extending a poor data format later.

Don't

- Assume that your data is so unique that it needs an entirely new format.
 - Think that a spreadsheet with free text labels or your locally developed database is the best way to store and reuse your data.
 - Format and store data in any format without keeping rich and relevant metadata and provenance.
 - Throw away the original data unless you are absolutely sure that storing them has no further added value, for example for review of experimental and analytical procedures.
- Not having certain pre-formatted data available may actually preclude the publication, reuse and citation of your (original) data by others and might also jeopardise the publication of accompanying articles.

Links

- [DS Question GitHub resources repository: nijx](#)

APRE IL CAPITOLO DEL LIBRO DI
BAREND MONS
CORRISPONDENTE ALL'OGGETTO
DELLA DOMANDA

FAIR Wizard

[About](#)[Features](#)[Resources](#)[Get](#)

Current Phase

Before Submitting the Proposal

Chapters

I. Administrative details1

II. Re-using data1

III. Creating and collecting data7

IV. Processing data3

V. Interpreting data1

VI. Preserving data6

VII. Giving access to data3

More

TODOs

Summary Report

III. Creating and collecting data

We will make sure that we know what data will be coming together in the project, when it will be coming. We also need to make sure that we have adequate storage space to deal with it, and that all the responsibilities have been taken care of.

1

What data formats/types will you be using?

Have you identified types of data that you will use that are used by others too? Some types of data (for example "images" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG" and "CSV") that help to make these data reusable. Are you using such common data formats?

You should make sure also to list the formats used in any data sets that you are re-using.

☒ Desirable: Before Submitting the Proposal

☒ Data Stewardship for Open Science: [nyx](#)

1.a.1

Data format/type

RDryXML Syntax Specification

FAIRsharing

<https://fairsharing.org/bsg-s001261>

☒ Desirable: Before Submitting the Proposal

1.a.2

Is this a standard data format used by others in this field?

☒ Desirable: Before Submitting the DMP

☐ a. No

LINK A
FAIRSHARING
REGISTRY

FAIR tools

The table also shows how the individual tool fits in with the FAIR dimensions. You can use the overview to find the right tools to make your research data (more) FAIR - depending on the data you have and what you want to do with it.

DK FAIR portal

Availability	Discipline	FAIR dimension	Phases in research life cycle	Service Name	Description
International	Generic (Tabular data)		Process/Analyze	OpenRefine	OpenRefine is a standalone open source desktop application for data cleanup and transformation to other formats (i.e. data wrangling)
International	Generic	FA(I)(R)	Publish/Disseminate, Archive, Discover & Re-use, Release, Preserve	Zenodo	Zenodo is a general-purpose open access research data repository, hosted by CERN (Switzerland) that provides a place for researchers to deposit datasets. Researchers in any subject area to are able to upload files up to 50 GB. It has an integration with GitHub to make code hosted in GitHub citable. Support: zenodo.org/support
International	Generic - Highly recognized in Social Sciences	FA(I)(R)	Publish/Disseminate, Archive, Discover & Re-use, Release, Preserve	Harvard Dataverse	Dataverse is a data repository that is widely used within the Social Sciences. Researchers can login with their institutional credentials via WAYF. Data can be made findable by applying discipline-specific metadata schemes and digital object identifiers (DOIs). Data is

International	Generic	FA(I)(R)	Publish/Disseminate, Archive	B2SHARE	B2SHARE is a secure and trustworthy way for researchers, scientific communities and citizen scientists to store and share small-scale research data from diverse contexts
International	Generic	FA(I)(R)	Discover & Re-use	B2FIND	B2FIND is a discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other repositories.
International	Generic (Tabular data)	FA(I)(R)	Document	Data Package Creator	Data Package is a simple container format used to describe and package a collection of data. It can be used to package any kind of data. At the same time, for specific

FAIR convergence matrix

The FAIR principles articulate the behaviors expected from digital artifacts that are Findable, Accessible, Interoperable and Reusable by machines and by people. Although by now widely accepted, the FAIR Principles by design do not explicitly consider actual implementation choices enabling FAIR behaviors. As different communities have their own, often well-established implementation preferences and priorities for data reuse, coordinating a broadly accepted, widely used FAIR implementation approach remains a global challenge. In an effort to accelerate broad community convergence on FAIR implementation options, the GO FAIR community has launched the development of the FAIR Convergence Matrix. The Matrix is a platform that compiles for any community of practice, an inventory of their self-declared FAIR implementation choices and challenges. The Convergence Matrix is itself a FAIR resource, openly available, and encourages voluntary participation by any self-identified community of practice (not only the GO FAIR Implementation Networks). Based on patterns of use and reuse of existing resources, the Convergence Matrix supports the transparent derivation of strategies that optimally coordinate convergence on standards and technologies in the emerging Internet of FAIR Data and Services.

FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

[Show all authors](#)

Hana Pergl Sustkova, Kristina Maria Hettne, Peter Wittenburg, Annika Jacobsen,

Posted Online January 31, 2020

https://doi.org/10.1162/dint_a_00038

© 2019 Chinese Academy of Sciences Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license

Data Intelligence

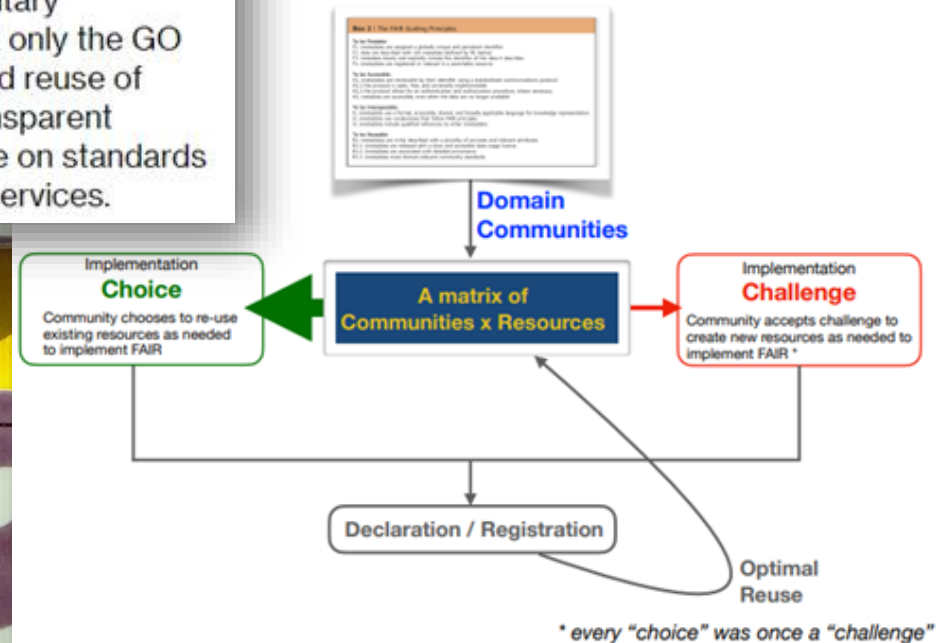
Volume 2 | Issue 1-2 | Winter-Spring 2020
p.158-170

2020

Cordoba

FAIR Implementation Profiles

2020



MATRICE CON LE RISORSE OPERATE
DALLE COMUNITÀ
(SCELTA O SFIDA A CREARNE UNA
NUOVA)
PAROLA CHIAVE: **CONVERGENZA**

...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).	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> 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	<input type="checkbox"/> Beginner <input type="checkbox"/> 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.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> 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.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> 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.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> 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	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced <input type="checkbox"/> Not yet

Ssssst...forse ALCUNI editori si stanno muovendo....

An Academic Publishers' GO FAIR Implementation Network (APIN)

Jan Velterop^{a,*} and Erik Schultes^{b,*}

Jan.6, 2021

Abstract

Presented here is a proposal for the academic publishing industry to get actively involved in the formulation of protocols and standards that make published scientific research material machine-readable in order to facilitate data to be findable, accessible, interoperable, and re-usable (FAIR). Given the importance of traditional journal publications in scholarly communication worldwide, active involvement of academic publishers in advancing the more routine creation and reuse of FAIR data is highly desired.

CREAZIONE DI
ARTICOLI SECONDO I
PRINCIPI FAIR

FAIR data nelle scienze umane



Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | February 2020

February 2020



RECOMMENDATIONS

- » Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.
- » Familiarise yourself with the FAIR Data Principles before you start collecting data and building corpora e.g. FORCE11: the FAIR Data Principles, GO-FAIR: FAIR Data Principles and discuss with colleagues and experts to build a better understanding.
- » Digitally document all your research and data collection work -- at the beginning of a project it is difficult to judge which information of the research process will be important and valuable later on.
- » Use well-established tools to facilitate your research work, as many of them allow data sharing e.g. MIT Libraries Digital Humanities: Tools and Resource Recommendations.
- » Browse humanities datasets and consider whether your own assets could be published in a similar fashion (e.g. Humanities Commons, UK Data Archive, ARCHE re3data.org filtered for humanities).
- » When you start producing data, keep this maxim of Open Science in mind: data should be 'as open as possible and as closed as necessary'.

Awareness of the FAIR principles and willingness to adopt them is not sufficient to transform data practices in any discipline. The paradigm shift requires effort, and this effort, which impacts on many roles in the research and higher education sectors, requires incentives, support, and recognition for adoption to be successful.

SERVE CAMBIO DI
PARADIGMA
CON INCENTIVI, SUPPORTO E
RICONOSCIMENTO

RECOMMENDATIONS

- Clarify all legal issues at the beginning of your research project and include the findings of this process in the data management plan.
- Use checklists adequate to your research topic/discipline.
- Check the resources indicated by DARIAH, CLARIN (see further reading).
- In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).
- Ask for consent to share anonymised data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities. (See for example the guide on informed consent in the CESSDA data management expert guide).
- Avoid collection of (sensitive and non-sensitive) personal data when possible.
- Get legal support (IPR, copyright, patents, trademarks etc.) from your home institution. If there is no dedicated office for this purpose, try to get information from your university library, as its staff are often confronted with such issues.
- If you need permission from the copyright holder in order to use sources like images for your publication, try to get one that covers both printed and digital copies.
- Finally, check the recommendations in the section on Licences, that are closely related to this section.

RECOMMENDATIONS

- 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.

RECOMMENDATIONS

- If applicable, determine if the body funding your research has particular requirements for a DMP or offers a template for framing your plan. If there is no required template, choose an existing appropriate one (e.g. via DMPOnline).
- Devise a DMP prior to collecting data. Define and plan for your data: all research projects deal with data. If your project includes the analysis of text corpora, for example, then the corpora themselves are data, and you should make sure they are clearly described, documented, and managed according to the FAIR principles so your research is reusable by others.
- Plan documentation of metadata: in order for your data to be comprehensible in the future and/or reusable by others, they will need descriptive metadata created according to a common schema to understand the content/purpose of the research. The richer the metadata, the more intelligible and useful the dataset (see section on Metadata).
- Use standardised terminology to increase interoperability. Consider employing vocabularies or ontologies that follow FAIR principles to increase interoperability and findability (e.g. see FAIRsharing.org).
- Consider the right questions to be answered in your DMP that can account for discipline-specific requirements. The DMP templates suggested by funders are quite high level and provide generic guidance for file naming or versioning conventions, database structuring and can be a good start. Tools like the dmponline.co.uk provide discipline specific examples that can be of further reference.
- DMP as living documents: Update your data management plan regularly in order to take into account any potential relevant changes such as using new data types and/or models, technology, new institutional data management policies, reassessing legal aspects or licences for legal compliance etc.
- Depending on the size of the organisation: think of providing institutional support for research data management (RDM), organise information sessions to raise awareness about good research data management, and the risks of not managing it early.
- If possible, consider involving library and/or repository support staff from the initial stages of research data management planning to discuss the best solutions, specifications, standards and protocols along which the repository operates. Repository staff can also assist scholars with understanding any specific data management requirements and associated costs.
- Factor the cost of research data management (time or human resources) into budgetary requirements at the point of application.

DISSEMINATION
What it means to disseminate data in the Humanities

IDENTIFY
Research Data in the Humanities

FAIR DATA and the HUMANITIES

PLAN
Data Management Plans

DEPOSIT for PRESERVATION, CITE & SHARE
License and Legal aspects
TDRs and PIDs for the Humanities

COLLECT/PRODUCE & STRUCTURE & STORE
Types and Formats, Metadata and Data Models for the Humanities

RECOMMENDATIONS

- 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 musical editions. The query language SPARQL and the representation tool for linked data JSON-LD are common standards for RDF (refers to FAIR principle 1).
- 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, ORCID, 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

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.



FAIR data



“Here be dragons”

#4 The phrase “Hic sunt dracones” (transl. “Here be Dragons”), is used on some old maps of the world to describe an area that was unknown to the cartographer. I found it quite appropriate to summarize the ambivalence of humanists towards data and all these “fancy” concepts discussed by “infrastructure people” like FAIR Data, the EOSC, or Research Data Management. First of all, it must be said that humanities researchers tend to be ambivalent about the concept of ‘data’^[15] and that “[t]here are issues surrounding [...] the acceptance of the ‘research data concept’”^[16]. In short, they just don’t use the word “data”, but talk about “sources”, “research materials” etc., which leads to the fact that the whole “data talk” doesn’t appeal to them. Additionally, an expeditionary survey conducted by PARTHENOS in 2017 among researchers in the domain of digital humanities, language studies, and cultural heritage showed that the FAIR Principles and the EOSC, concepts and recommendations, thriving among “infrastructure folks”, are relatively little known in the research communities themselves.^[17] Often, the publication of research data only comes as an afterthought (if at all).^[18] However, at the end of a project, it is often too late to publish the data in a meaningful way because of the lack of documentation and the lack of resources to prepare the data properly for publishing.

“Here be dragons”: Open Access to Research Data in the Humanities

#8 To sum up my observations so far: Humanities research data, in general, is rather heterogeneous, idiosyncratic, and complex^[30] and humanists are ambivalent about the term “data”. Digital practices are already part of the research activities of many humanists, especially in the Digital Humanities, but they are not equally fully developed. This leads to the fact that the potential of digital research data and methods is not fully exploited, because the digital research process is not carefully planned, with other words many research data already exist in digital form, but they are not findable, quality controlled, and reusable.^[31] All in all, the land of FAIR Research Data is still unknown territory for many humanists, or at least scary as if dragons would indeed live there. In the next part, therefore, I will argue for increased efforts for awareness raising and skills building and a “fellowship of the data”, a support system to facilitate the quest for FAIR data in the humanities.

To create this broad culture of FAIR data sharing in the humanities we have to roll up our sleeves, team up, and distribute hats:

1. Embrace Open principles,
2. bridge the gap between the digital and the humanities and look what we can learn from the Digital Humanities and other more data-savvy disciplines.

DATI SONO ETEROGENEI
METODI ANCORA POCO
SVILUPPATI...RECUPERARE IL
TEMPO PERSO!

5 WORKSHOP NAZIONALI

OPERAS

open scholarly communication in the european
research area for social sciences and humanities



FAIR Principles Implementation Networks News Events

**CO-OPERAS publishes a variety of workshop
reports on FAIRification efforts in the SSH**

[Reports on Zenodo](#)

CO-OPERAS

- COSA SONO I DATI NELLE SCIENZE
UMANE
- QUANTO SONO FAIR?

QUESTIONI APERTE

- CI VUOLE TEMPO, E NON VIENE RICONOSCIUTO
- MANCANO CONSAPEVOLEZZA E COMPETENZE
- **FINDABLE**: MANCANO METADATI (E COMPETENZE SUI)
- **ACCESSIBLE**: MANCA REPOSITORY CON GRANULARITÀ
ADATTA; MANCA UNICO PUNTO DI ACCESSO [TRIPLE]
- **INTEROPERABLE**: METADATI SPESSO TROPPO SPECIFICI;
LINGUE NAZIONALI, MANCANO ONTOLOGIE
- **REUSABLE**: DATI CON DIRITTI DIFFICILI DA DEFINIRE
(PATRIMONIO CULTURALE)

IN CORSO:

- FAIR PUBLISHING TOOLKIT
(FOCUS: IDENTIFICATIVI)
- FAIR IMPLEMENTATION
PROFILES
 - TRAINING
 - ADVOCACY

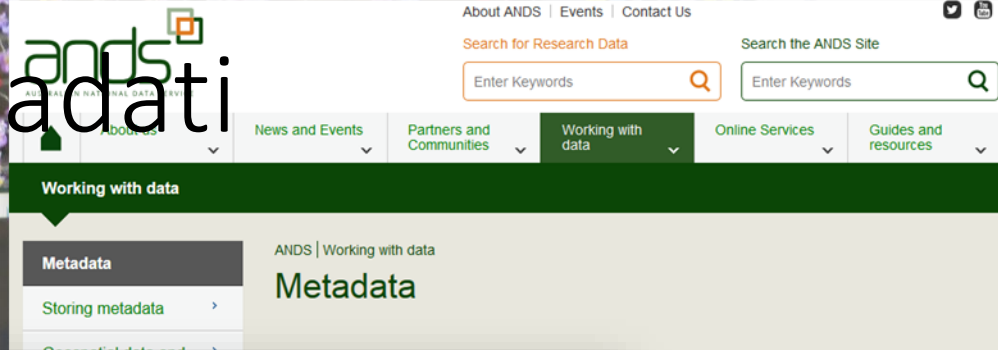
March 2020

Humanities and data:
listening to the communities
on the path towards FAIRness

FINDABLE



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. 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](#)

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.	<p>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.</p> <p>Text formats can be free-text OR terms from the CASRAI Publications resource type list.¹⁷</p> <p>***</p> <p>Examples:</p> <p>Dataset/Census Data, where 'Dataset' is resourceTypeGeneral value and 'Census Data' is</p>

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

F = findable. Metadata tools

What CEDAR does

<https://metadatacenter.org/>

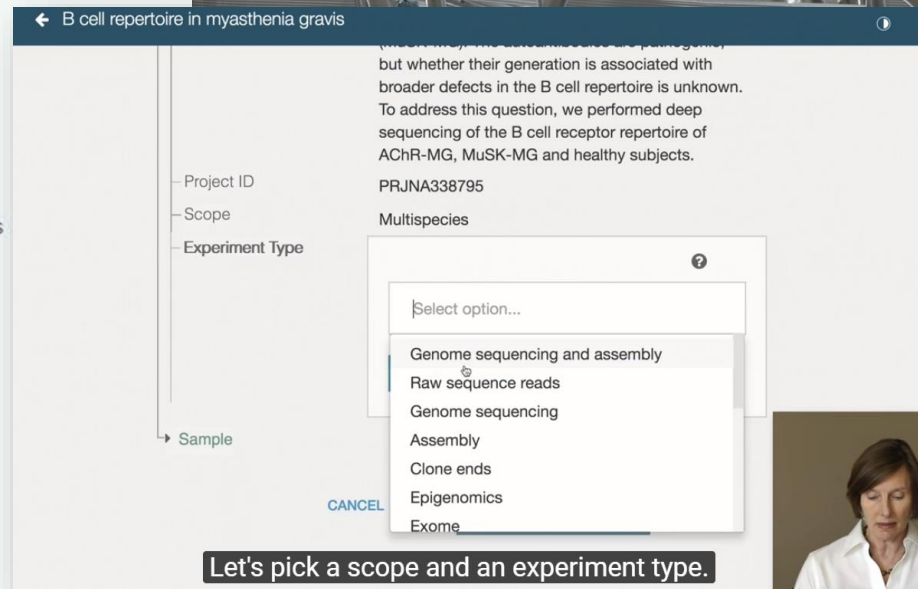
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools will create a metadata record to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools help users collect metadata, and download the information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

FAIR cookbook

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

Creating a Metadata Profile

			
Recipe metadata	Difficulty level	Reading Time	Intended Audience
identifier: RX.X version: v1.0	🔥🔥🔥	🕒 20 minutes	👤 Principal Investigator
		Recipe Type	📊 Data Manager
		💻 Hands-on	🔧 Data Scientist
		Executable Code	
		▶ Yes	

How to generate a metadata template

The following steps are intended as a starting point to guide the generation of a metadata template.

Step 1: Define competency questions

- What are the questions you would like to address with the template? Without a set of a competency questions, important variables may easily be forgotten. It is equally possible to collect too much metadata, making the resulting metadata model opaque and difficult to navigate. Competency questions serve as a guide to identify the most relevant experimental factors.

Step 2: Define a Minimal Set Of Metadata (MSOM) according to these questions

- Compile metadata from different sources
- Generate consolidated view on metadata by merging attributes as far as possible
- Differentiate metadata available for most of the studies from metadata occurring rarely (sparse matrix)
- Identify gaps in the metadata available for most of the studies comprising data that is considered important but has not been captured in the past
- Define a MSOM to be captured in the future from the metadata that is available for most of the studies and the metadata considered to be important
- Identify available community standards regarding minimal sets of metadata
- Add metadata attributes from those community standards to the MSOM, if they are not yet included
- Assign cardinality to the MSOM (identify mandatory metadata and how many times the attributes may be reported. Some metadata might not be mandatory but are still important to capture, if available)
- Identify appropriate ontologies representing your data and establish an application ontology (see recipe 4 of UC3)
- Assign, as far as possible, ontologies to the MSOM and the sparse matrix

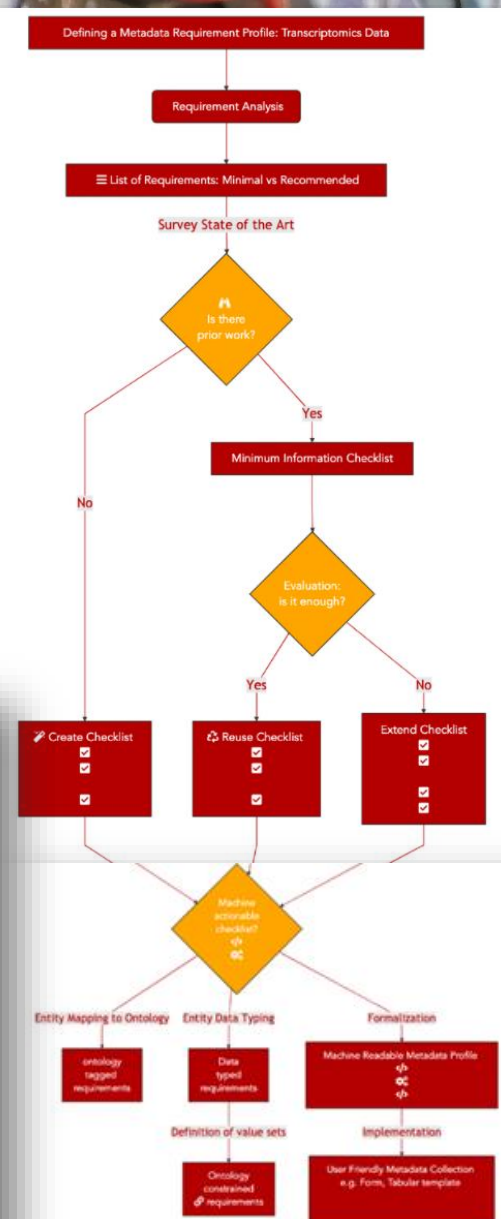


Fig. 9 Building a minimal metadata profile.

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 learn 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).
- 3 LINK YOUR RESEARCH** Link your research to your ORCID record.

<https://orcid.org/>

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!



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**

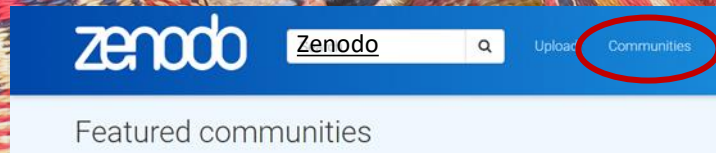
Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.

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.

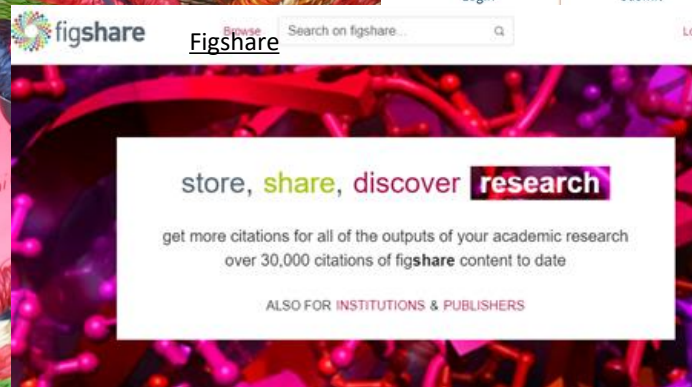
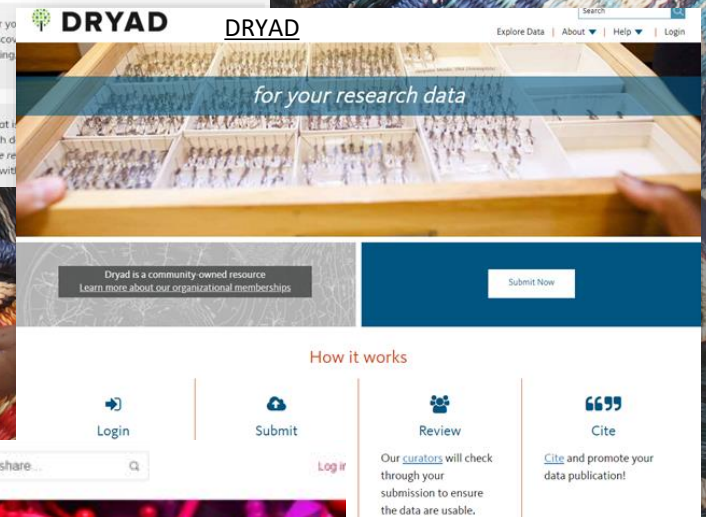
A = Accessible — Data repositories



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?

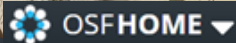
- Organization of data files in dataverses and datasets
- Addition of metadata and documentation
- Version management
- Management of access rights
- Easy collaboration with fellow researchers or project partners, even beyond your own university or research institute
- Centralized professional storage 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

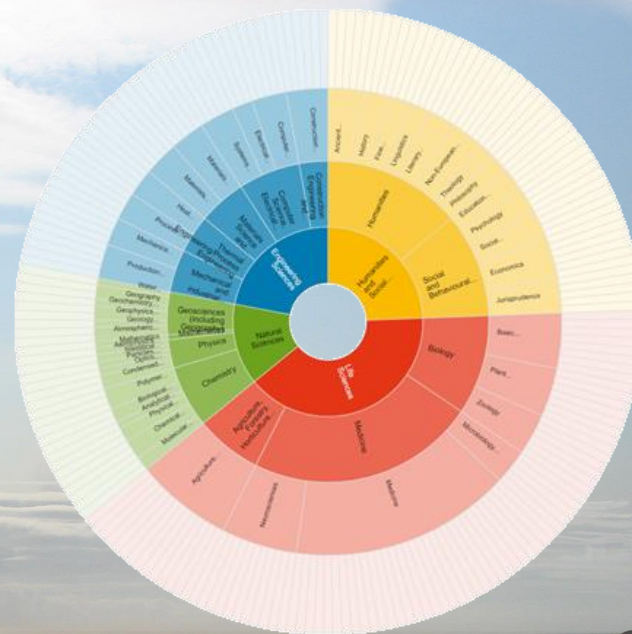
Store data, code, and other materials in OSF Storage, or connect your Dropbox or other third-party account. Every file gets a unique, persistent URL for citing and sharing.

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?



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

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



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? 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](#)

[gentle reminder] ...RISK ASSESSMENT

RICORDATEVI
ANCHE DELLA
SICUREZZA

Research Data Management @Harvard

RDM@Harvard

Home Vision Data Lifecycle Policies Resources Contacts

DATA LIFECYCLE

- Planning Data Management
- Data Acquisition and Collection
- Storage, Security, and Analysis
- Data Security
- Data Organization Best Practices
- Vendors and Tools
- Computing and Analytics Help
- Dissemination and Preservation

HOME / DATA LIFECYCLE /

Storage, Security, and Analysis

This table describes the 4 main data storage and research computing facilities offered at Harvard

	FAS Odyssey	HMS Q2	IQSS RCE	HBS RCS
Sensitive data support	Level 3, Level 4	Level 3	Level 3	Level 3, Level 4
User interface	Command line	Command line	Graphical User Interface	Graphical User Interface
Quick start tutorial	Quick Start Guide	Using Slurm	Contact: help@ia.harvard.edu	Quick Start
Integrated Applications (gui-based and executables)	List of 681 modules that each contain their own applications.	Lmod system of environment modules, with each module containing its own applications	List of statistical applications. Additional software packages can be installed	Software packages include R, Stata, Matlab, SAS, Stata/Transfer, and more. Users can install additional software.
Training available?	Yes, regularly scheduled workshops.	Yes, regularly scheduled classes.	Yes, regularly scheduled statistical workshops.	Yes, required training session upon sign-up, subsequent information sessions about best practices and new resources. Regular workshops in collaboration with IQSS.

Level	Data Classification and Examples (abridged version)
5	<p>Information that would cause severe harm to individuals or the University if disclosed.</p> <ul style="list-style-type: none"> Research information classified as Level 5 by an IRB or otherwise required to be stored or processed in a high security environment and on a computer not connected to the Harvard data networks Certain individually identifiable medical records and genetic information, categorized as extremely sensitive
4	<p>Information that would likely cause serious harm to individuals or the University if disclosed.</p> <ul style="list-style-type: none"> High Risk Confidential Information (HRCI) and research information classified as Level 4 by an IRB Personally identifiable financial or medical information Information commonly used to establish identity that is protected by state, federal, or foreign privacy laws and regulations Individually identifiable genetic information that is not Level 5 National security information (subject to specific government requirements) Passwords and Harvard PINs that can be used to access confidential information
3	<p>Information that could cause risk of material harm to individuals or the University if disclosed.</p> <ul style="list-style-type: none"> Research information classified as Level 3 by an IRB Information protected by the Family Educational Rights and Privacy Act (FERPA) to the extent it is not covered under Level 4 including non-directory student information and directory information about students who have requested a FERPA block HUIDs associated with names or any other information that could identify individuals Harvard personnel records (employees may discuss terms and conditions of employment with each other and third parties) Level 4 including non-directory student information and directory information about students who have requested a FERPA block HUIDs associated with names or any other information that could identify individuals Harvard personnel records (employees may discuss terms and conditions of employment with each other and third parties) Institutional financial records Individual donor information Other personal information protected under state, federal and foreign privacy laws not classified as Level 4 or 5
2	<p>Information the disclosure of which would not cause material harm, but which the University has chosen to keep confidential.</p> <ul style="list-style-type: none"> Unpublished research work and intellectual property not in Level 3 or 4 Research information classified as Level 2 by an IRB Patent applications and work papers, drafts of research papers Building plans and information about the University physical plant
1	<p>Public information.</p> <ul style="list-style-type: none"> Research data that has been de-identified in accordance with applicable rules Published research Published information about the University Course catalogs Directory information about students who have not requested a FERPA block Faculty and staff directory information

A = Accessible. Data repositories

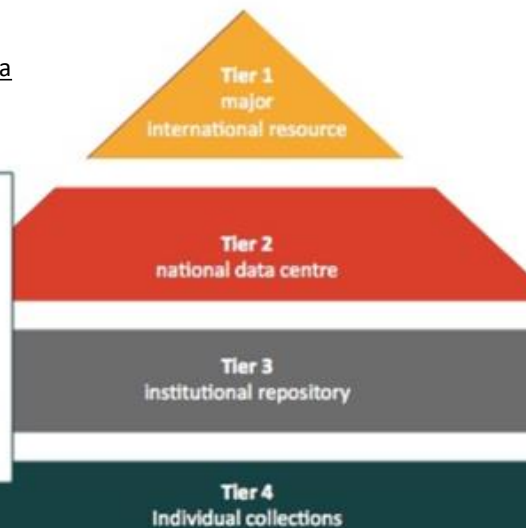


DCC Where to keep research data

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

Each higher tier brings greater
responsibility and demands for
access.

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



«PRIORITÀ»

1. DISCIPLINARY
 2. ZENODO
 3. INSTITUTIONAL
- (A MENO CHE SIA DATAVERSE,
CHE VIENE USATO DURANTE IL
PROGETTO)

A = Accessible. 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) ☐

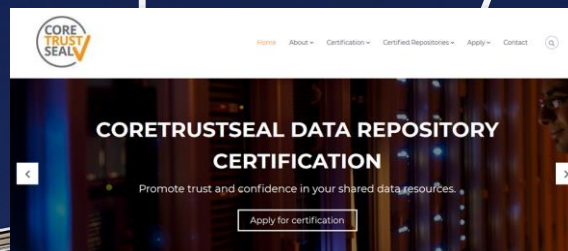
By agreeing to the terms and conditions the depositor will not be in breach of **copyright**, or any contract terms covering **Intellectual Property** in the research, (e.g. the grant conditions or a consortium agreement) ☐

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 <input type="checkbox"/>	Repository publishes other pertinent information as metadata fields to enhance cross-disciplinary discovery <input type="checkbox"/>	Metadata is catalogued to enhance reuse according to sector-leading standards, or to fulfil domain-specific purposes <input type="checkbox"/>
Stable identifiers: Enables a DOI or other open standard identifier to be assigned to a landing page for each ingested dataset/ collection <input type="checkbox"/>	Supports assignment of related persistent IDs per dataset/ collection <input type="checkbox"/>	Supports assignment of multiple persistent IDs at different levels of granularity within dataset/ collection <input type="checkbox"/>
Discovery metadata: Provides Datacite mandatory metadata and exposes it according to open access repository protocols <input type="checkbox"/>	Provides metadata elements to enable broader discovery (e.g. geo-spatial) to reflect best practice changes and local needs <input type="checkbox"/>	Exposes discovery metadata as Linked Open Data to optimise automatic discovery <input type="checkbox"/>
Metadata harvesting: Sufficient information can be harvested about data deposited with third-party repositories, to meet funders' needs for metadata on <input type="checkbox"/>	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 <input type="checkbox"/>	Metadata on the externally held research data is sufficiently structured and organized <input type="checkbox"/>

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.geosciencedata.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				

Data journals list

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>).

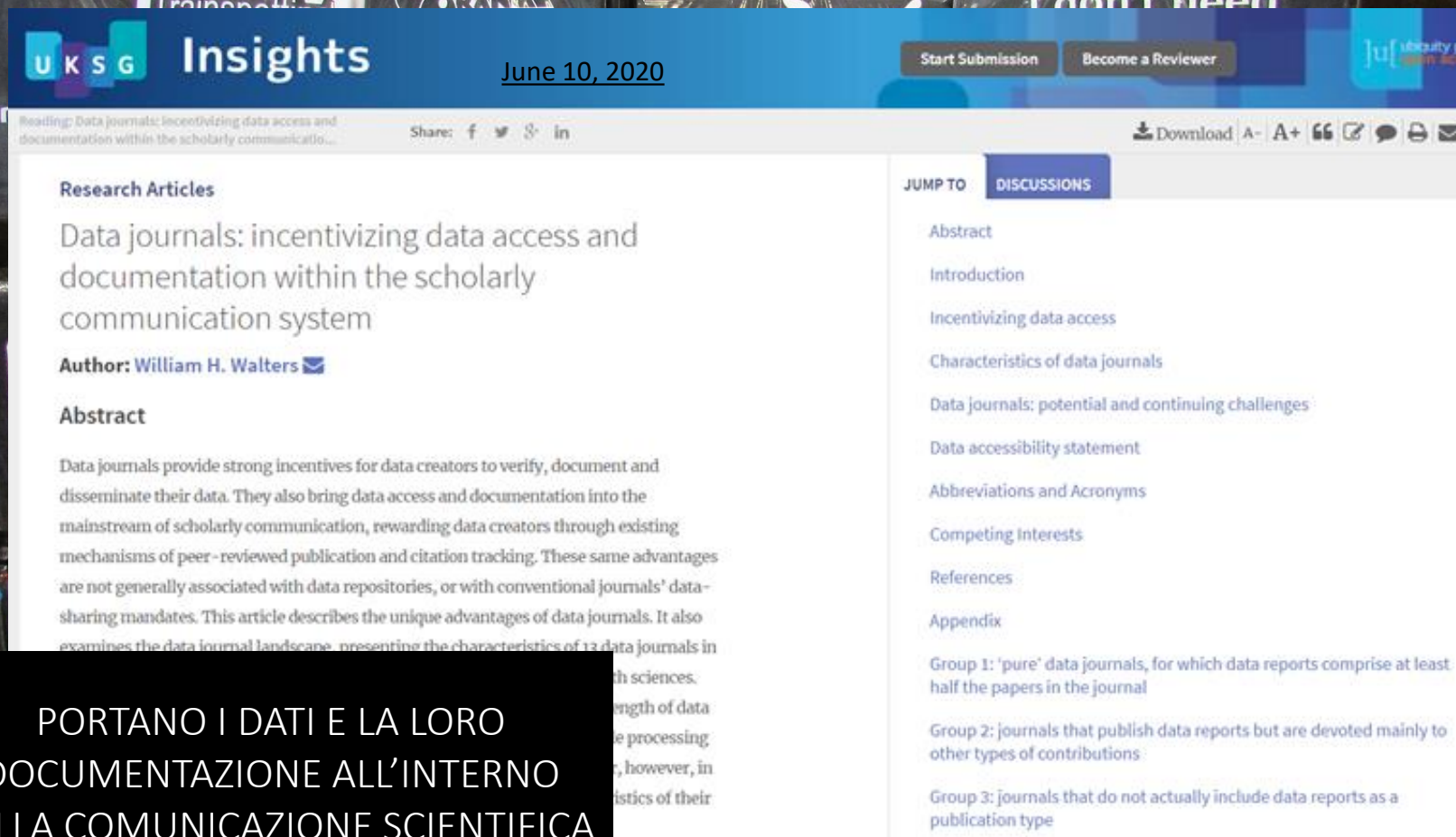
Data journals

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

A = Accessible. Data journals



The screenshot shows the UKSG Insights website interface. At the top, there's a blue header with the UKSG logo and the word 'Insights'. To the right of the header are buttons for 'Start Submission' and 'Become a Reviewer'. Below the header, the date 'June 10, 2020' is displayed. The main content area features an article titled 'Data journals: incentivizing data access and documentation within the scholarly communication system' by William H. Walters. The article's abstract is visible, discussing the incentives for data creators to verify, document, and disseminate their data. On the right side of the article, there's a 'JUMP TO' section with a 'DISCUSSIONS' tab, listing various sections of the article for navigation.

UKSG Insights June 10, 2020

Start Submission Become a Reviewer

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

Share: f t s in

Download A- A+ " " " "

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

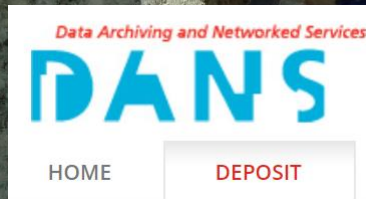
th sciences, length of data e processing, however, in istics of their

JUMP TO **DISCUSSIONS**

- Abstract
- Introduction
- Incentivizing data access
- Characteristics of data journals
- Data journals: potential and continuing challenges
- Data accessibility statement
- Abbreviations and Acronyms
- Competing Interests
- References
- Appendix
- Group 1: 'pure' data journals, for which data reports comprise at least half the papers in the journal
- Group 2: journals that publish data reports but are devoted mainly to other types of contributions
- Group 3: journals that do not actually include data reports as a publication type

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

A = Accessible. Formati



Type	Preferred format(s)	Non-preferred format(s)
Text documents	<ul style="list-style-type: none"> PDF/A (.pdf) 	<ul style="list-style-type: none"> ODT (.odt) MS Word (.doc, .docx) RTF (.rtf) PDF (.pdf)
Plain text	<ul style="list-style-type: none"> Unicode text (.txt) 	<ul style="list-style-type: none"> Non-Unicode text (.txt)
Markup language	<ul style="list-style-type: none"> XML (.xml) HTML (.html) Related files: .css, .xslt, .js, .es 	<ul style="list-style-type: none"> SGML (.sgml)
Spreadsheets	<ul style="list-style-type: none"> ODS (.ods) CSV (.csv) 	<ul style="list-style-type: none"> MS Excel (.xls, .xlsx) PDF/A (.pdf) OOXML (.docx, .docm)
Databases	<ul style="list-style-type: none"> SQL (.sql) SIARD (.siard) DB tables (.csv) 	<ul style="list-style-type: none"> MS Access (.mdb, .accdb) (v. 2000 or later) dBase (.dbf) HDF5 (.hdf5, .he5, .h5)
Statistical data	<ul style="list-style-type: none"> SPSS Portable (.por) SPSS (.sav) STATA (.dta) DDI (.xml) data (.csv) + setup (.txt) 	<ul style="list-style-type: none"> SAS (.7dat; .sd2; .tpt) R (* under examination)
Raster images	<ul style="list-style-type: none"> JPEG (.jpg, .jpeg) TIFF (.tif, .tiff) PNG (.png) JPEG 2000 (.jp2) 	<ul style="list-style-type: none"> 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



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



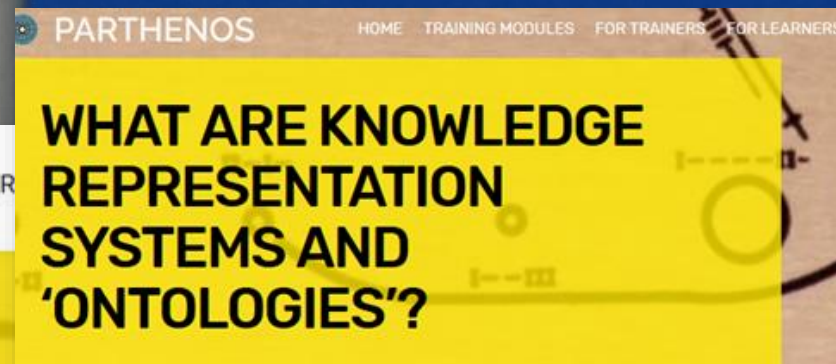
I = Interoperable. Standards



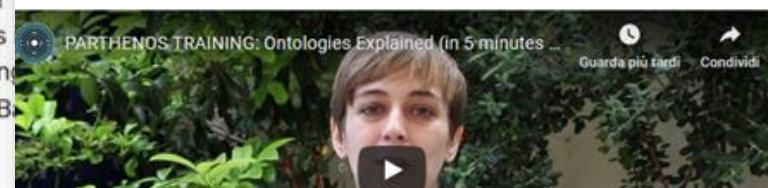
Even perfect metadata may not allow data to become interoperable if a different standard is used. A "standard" refers to a system that structures what types of information are captured in a collection. In our .mp3 library system, a standard is expressed in the header categories such as 'name,' 'time,' '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



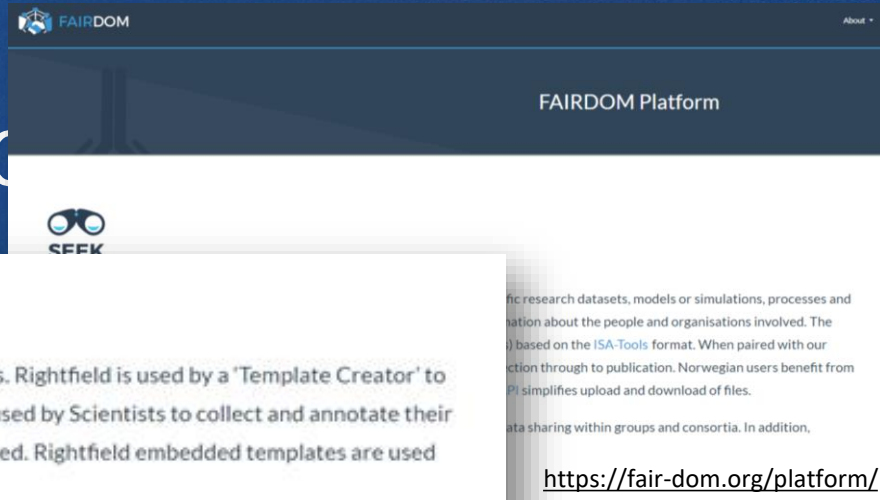
About the SSK

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



Opscidia's ontology generator

Written on 03 March 2021.

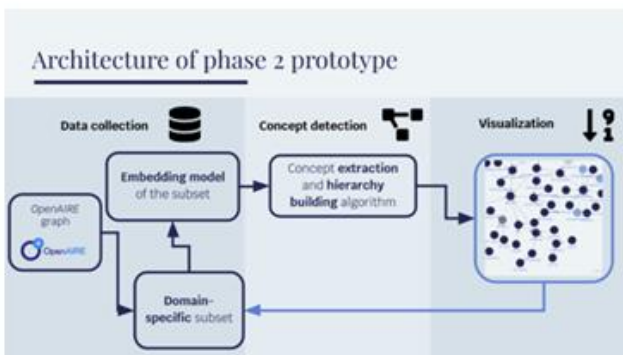
Opscidia



GENERATORE DI ONTOLOGIE

The solution proposed by Opscidia is an ontology generator that consists in three layers:

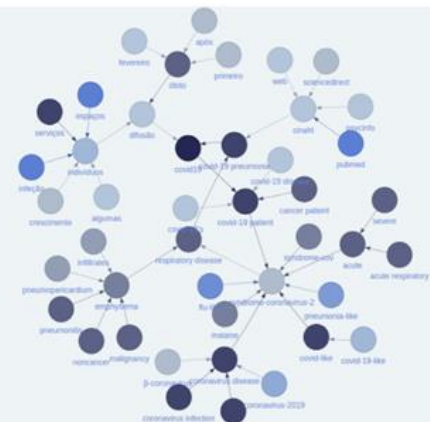
- **Data collection layer:** here it consists mostly in harvesting the resources (API or Dumps of specific OpenAIRE communities)
- **Concept detection layer:** A simple, unsupervised algorithm extracts and hierarchizes concepts related to seed concept entered by the user. It can easily scale-up both with the amount of data and with the amount of users / requests.
- **Visualization layer:** A visualization tool represents graphically the produced ontology and links it back to the documents of the corpus from which the ontology was created.



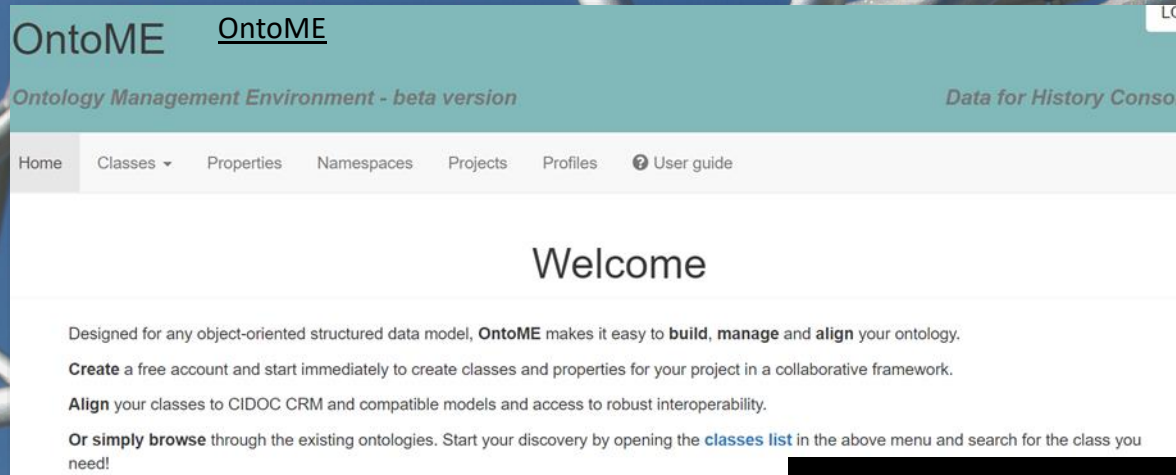
The results of the Ontology Generator

A simple tool for semi-automatic domain specific ontology creation has been built.

It takes a concept as an input and extracts from a subset of OpenAIRE graph a hierarchical list of concepts associated with the user input. This list is displayed using a simple visualization layer and linked back to the scientific literature through OpenAIRE graph.



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'. A navigation bar contains links: Home, Classes (with a dropdown arrow), Properties, Namespaces, Projects, Profiles, and a User guide icon. The main content area has a 'Welcome' heading. Below it, three bullet points describe the platform's capabilities: building, managing, and aligning ontologies; creating accounts for collaborative work; and aligning classes to CIDOC CRM for interoperability. It also mentions browsing existing ontologies.

OntoME [OntoME](#)

Ontology Management Environment - beta version Data for History Consortium

Home Classes ▾ Properties Namespaces Projects Profiles ? User guide

Welcome

Designed for any object-oriented structured data model, **OntoME** makes it easy to **build, manage** and **align** your ontology.

Create a free account and start immediately to create classes and properties for your project in a collaborative framework.

Align your classes to CIDOC CRM and compatible models and access to robust interoperability.

Or simply browse through the existing ontologies. Start your discovery by opening the [classes list](#) in the above menu and search for the class you need!

ONTOME
ESEMPIO DI APPLICAZIONE
ALLE SCIENZE STORICHE



The screenshot shows an article from the IOS Press Content Library, dated 2021. The article title is 'A challenge for historical research: Making data FAIR using a collaborative ontology management environment (OntoME)'. The author is Francesco Beretta, from the Laboratoire de recherche historique Rhône-Alpes, CNRS – Université de Lyon. The article is part of the 'Journals' section.

IOS Press IOS Press Content Library 2021

Home Journals


A challenge for historical research: Making data FAIR using a collaborative ontology management environment (OntoME)

Francesco Beretta
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 = Inteoperable. Ontologies

**BioPortal** [Ontologies](#) [Search](#) [Annotator](#) [Recommender](#) [Mappings](#)

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

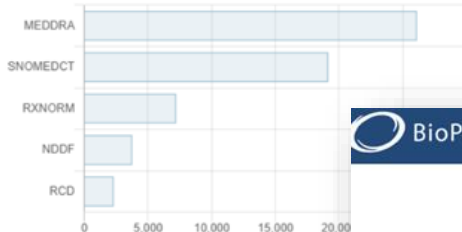
Search for a class

[Advanced Search](#)

Find an ontology

[Browse Ontologies](#)

Ontology Visits (April 2021)



Ontology	Visits (approx.)
MEDDRA	20,000
SNOMEDCT	18,000
RXNORM	8,000
NDDF	4,000
RCD	2,000

BioPortal Statistics

Category	Count
Ontologies	872
Classes	10,085,431
Properties	36,286

**BioPortal** [Ontologies](#) [Search](#) [Annotator](#) [Recommender](#) [Mappings](#)

Browse

Browse the library of ontologies [?](#)

Showing 29 of 1036 Sort: Popular

[Submit New Ontology](#)

Entry Type

☒ [Ontology \(29\)](#)

☐ [Ontology View \(2\)](#)

Uploaded in the Last

Category

☐ [All Organisms \(0\)](#)

☐ [Anatomy \(1\)](#)

☐ [Animal Development \(0\)](#)

☐ [Animal Gross Anatomy \(1\)](#)

☐ [Arabidopsis \(0\)](#)

Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)

MedDRA is an international medical terminology with an emphasis on use for data entry, retrieval, analysis, and display

Uploaded: 1/7/21

notes: 1 projects: 10 classes: 75,042

SNOMED CT (SNOMEDCT)

SNOMED Clinical Terms

Uploaded: 1/7/21

notes: 3 projects: 23 classes: 361,042

RxNORM (RXNORM)

RxNorm Vocabulary

Uploaded: 1/7/21

projects: 1 classes: 30,111


National Drug Data File (NDDF)

National Drug Data File Plus Source Vocabulary

projects: 1 classes: 30,111

BIOPORTAL,
ONTOLOGIE PER LE
SCIENZE BIOMEDICHE

I = Interoperable



FAIR Cookbook

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

How to interlink data from different sources?

 Recipe metadata identifier: <i>RX.X</i> version: <i>v1.0</i>	 Difficulty level 🔥🔥🔥	 Reading Time 🕒 30 minutes Recipe Type 🌐 Background Information Executable Code 🎥 No	 Intended Audience 👤 Principal Investigators 📊 Data Manager 🔧 Data Scientist
--	--	--	---



FAIR Cookbook

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

How to interlink data from different sources?

 Recipe metadata identifier: <i>RX.X</i> version: <i>v1.0</i>	 Difficulty level 🔥🔥🔥	 Reading Time 🕒 30 minutes Recipe Type 🌐 Background Information Executable Code 🎥 No	 Intended Audience 👤 Principal Investigators 📊 Data Manager 🔧 Data Scientist
--	--	---	---

[FAIRsharing. Per essere interoperabili]

FAIRsharing.org
standards, databases, policies

FAIRsharing

Search all of FAIRsharing

Standards Databases Policies Collections Add/Claim Content Stats Log in or Register

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

HOW CAN WE HELP?

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.

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards Databases Policies Collections Add/Claim Content Stats Log in or Register

Recommended Records

Recommended

Associated Publication?

No Publication Has Publication

Claimed?

No Maintainer Has Maintainer

Record Status

Uncertain Deprecat In develo Ready

Standard Type

Terminology Artifact 821

Model/Format 477

Reporting Guideline 189

Matrix 20

Registry Name

Animal natural history and life history

Semanticscience Integrated Ontology

BioAssay Ontology

Apollo-SV

Sort by

B

Preclinical Studies

Recommended Records

Recommended

Associated Publication?

No Publication Has Publication

Claimed?

No Maintainer Has Maintainer

Record Status

Uncertain Deprecat In develo Ready

Standard Type

Terminology Artifact 22

Registry Name Abbreviation Type Subject Domain

CDISC Analysis Data Model CDISC ADaM Standard Biomedical Science Analysis Data Model Data Transformation

CDISC Study Data Tabulation Model CDISC SDTM Standard Biomedical Science Report Device

CDISC Clinical Data Acquisition Standards Harmonization CDISC CDASH Standard Biomedical Science Data Acquisition Report

Terminology of FAHH Standard Anatomy Histology Home sapiens None None None None

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

[es. di cosa documentare]

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'

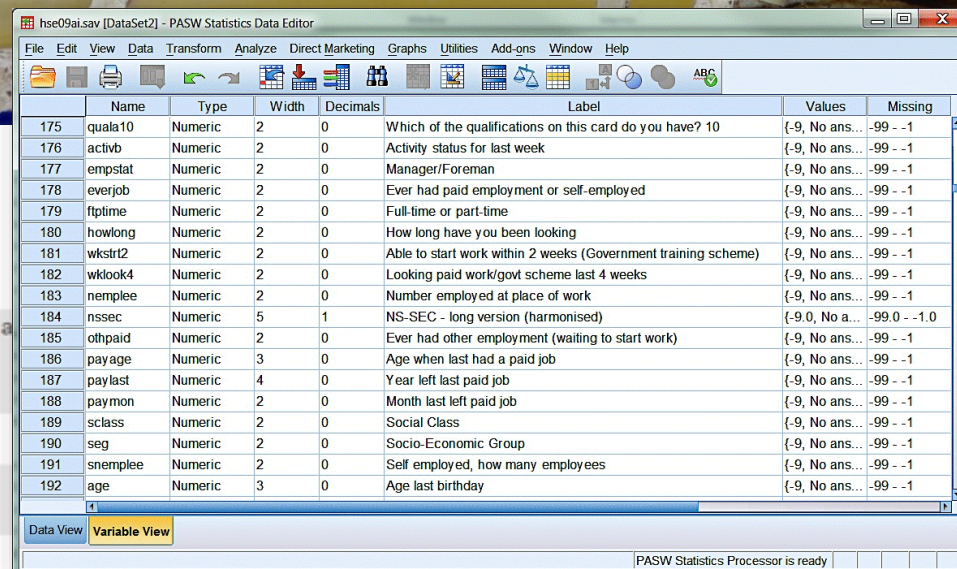
- coding and classification schemes explained, with a bibliographic and dated reference (some standards change over time)

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

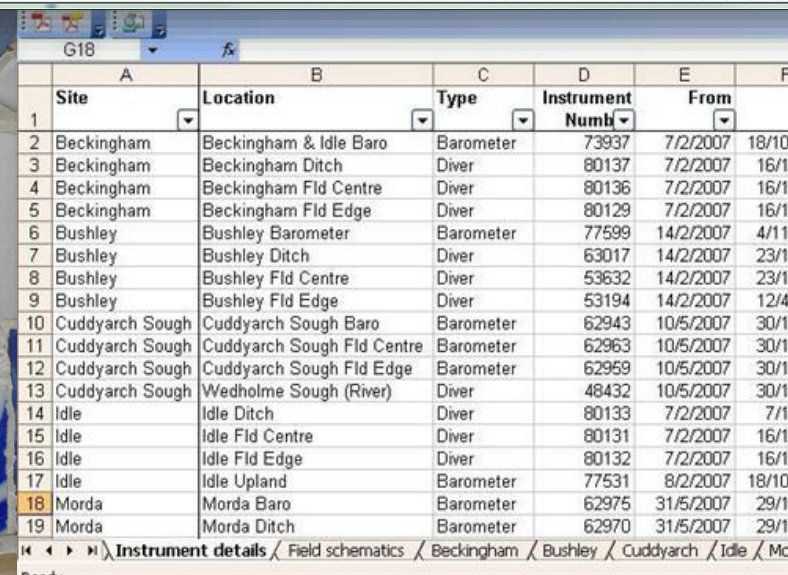


hse09ai.sav [DataSet2] - PASW Statistics Data Editor

	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	wkstr12	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

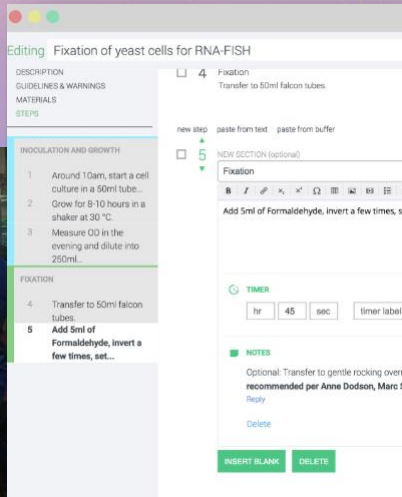
	A	B	C	D	E	F
1	Site	Location	Type	Instrument Numb	From	
2	Beckingham	Beckingham & Idle Baro	Barometer	73937	7/2/2007	18/10/07
3	Beckingham	Beckingham Ditch	Diver	80137	7/2/2007	16/1/07
4	Beckingham	Beckingham Fld Centre	Diver	80136	7/2/2007	16/1/07
5	Beckingham	Beckingham Fld Edge	Diver	80129	7/2/2007	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. Documentazione

≡  protocols.io

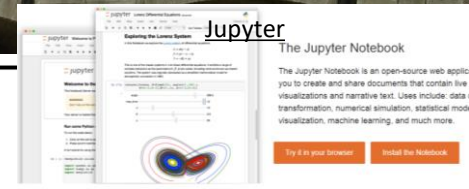
Make your science more reproducible
protocols.io is the #1 open access repository for science methods



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.

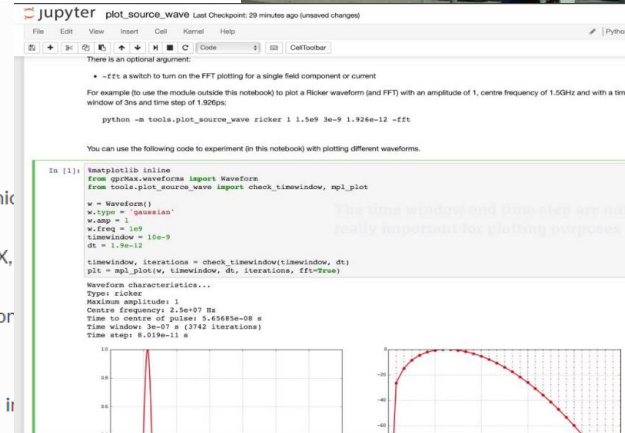
They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.



Notebook web application

The notebook web application enables users to:

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



...WHY NOT?

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

R= Reusable. License

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

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

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

RICORDA: NESSUN
COPYRIGHT SUI DATI
(NON CREATIVI)

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

on the legal protection of databases

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

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?

semplici dati e
informazioni

nessuna tutela

database
non creativo

solo diritto
sui generis

database
creativo

diritto sui generis
+ diritto d'autore

livello diritto
d'autore

livello diritto
sui generis

[webinar]



OpenAIRE 2019

SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

More Information about the 2019 webinar series.

data management plan | OA to research data | open science

Aspetti legali nella gestione dei dati della ricerca

Thomas Margoni
University of Glasgow - CREATE
OpenAIRE project

Support

RESOURCES
Open Science Primers
Guides
Factsheets
Use cases

HELPDESK
FAQs
Ask a Question

TRAINING
Webinars
Workshops
Community of Practice



2013



2020

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

2020

OpenAIRE Legal Policy Webinars

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

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

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

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

Legal interoperability

Chapter two - addresses key thematic legal issues, including:

- Copyright (including database rights) and licences;
- Other intellectual property rights (e.g. patents, trade secrets, neighbouring rights);
- Privacy and data protection (GDPR);
- Other restrictions and legitimate reasons (e.g., protection of sovereign genetic resources and traditional knowledge); and
- Private law considerations.

1. Open access to research data is an enabler of legal interoperability. The promotion of FAIR Principles should go hand-in-hand with efforts to make data open in accordance with the principle that data must be “as open as possible and as closed as necessary”.
2. Regardless of whether the data is Open or not, all new data made available through the EOSC should be FAIR by design.

9. Copyrightable **data** should be FAIR and, to the greatest extent possible, be made part of the public domain or assigned a permissive licence, unless legal or legitimate reasons apply. The Creative Commons No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL) or an equivalent statement of rights should be preferred. In cases where liability is a concern that cannot be addressed by other means, the CC BY 4.0 licence is an appropriate alternative.

11. From a licence compatibility perspective, attribution should be pursued by means of moral and ethical obligations e.g., the European Code of Conduct for Research Integrity or the development of Persistent Identifiers, or by way of a standard form of acknowledgement, rather than by means of a licence such as the CC BY 4.0. The CC0 is, in general, preferred over the CC BY 4.0, although both are generally permissive.

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 e 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"

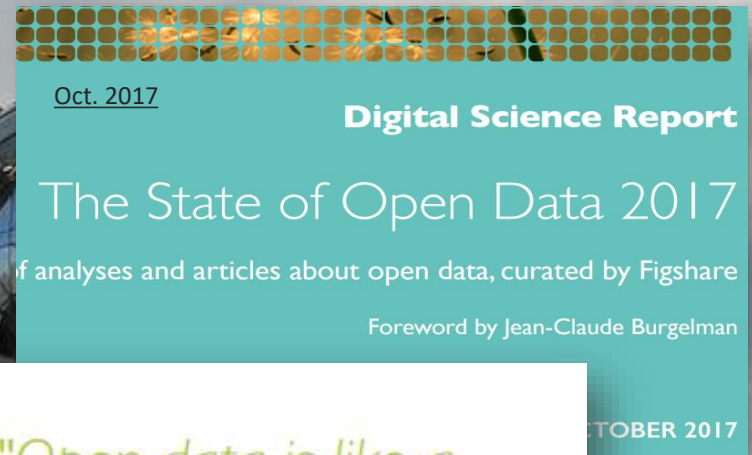
Traduci il Tweet

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."

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

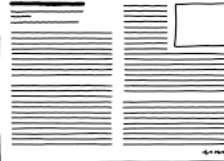
- EVITA DUPLICAZIONI
- MASSIMO RITORNO SU INVESTIMENTI

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



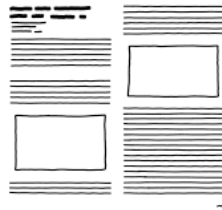
Some is here,
some is there, ask
us for the rest?



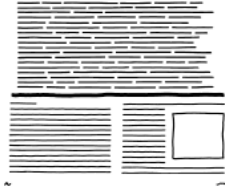
Data is "available"
upon reasonable
request.



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 researchers to correctly use and understand your data."

IMPOSSIBILE, SE
NELLA «R» DI FAIR
AVETE
DOCUMENTATO

It's worth mentioning, however, a small number of people will actually be useful to counter willful misrepresentation (Freedom of Information legislation), as one can quickly point to the record 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 say that a dataset has value to future research. "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 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 by the researcher to 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

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.

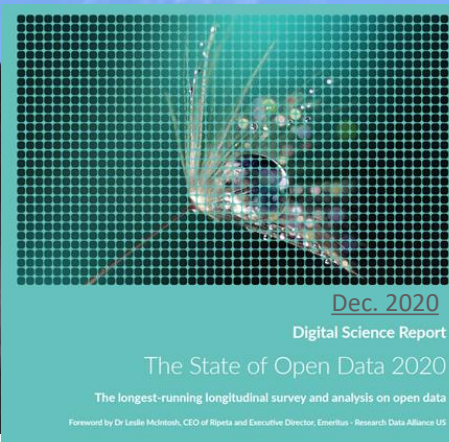
CI SONO BUONI
ARGOMENTI CHE
SMONTANO LE
PAURE RISPETTO
ALLA
CONDIVISIONE

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

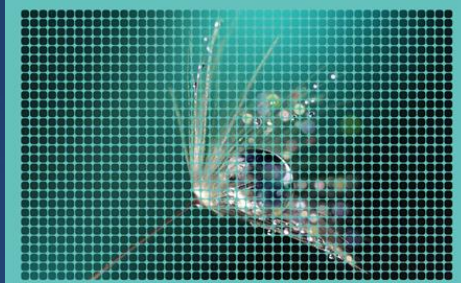
[scuse, scuse... aggiornato]



What problems/concerns, if any, do you have with sharing datasets?



Motivi per condividere



Dec. 2020

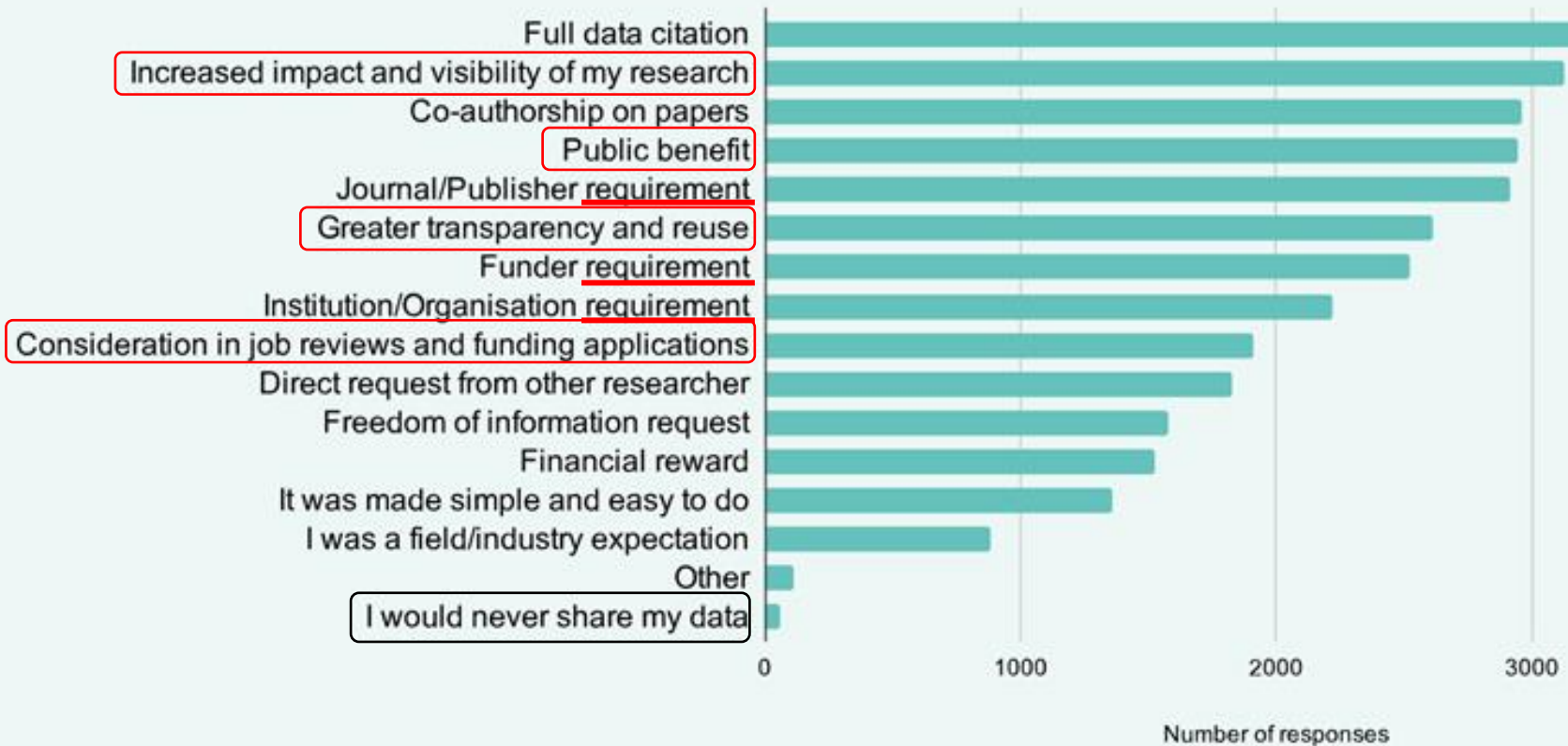
Digital Science Report

The State of Open Data 2020

The longest-running longitudinal survey and analysis on open data

Foreword by Dr Leslie McInnes, CEO of Ripeta and Executive Director, Emeritus - Research Data Alliance US

What circumstances would motivate you to share your data?



Cosa manca per condividere



Add a Preprint Search Support

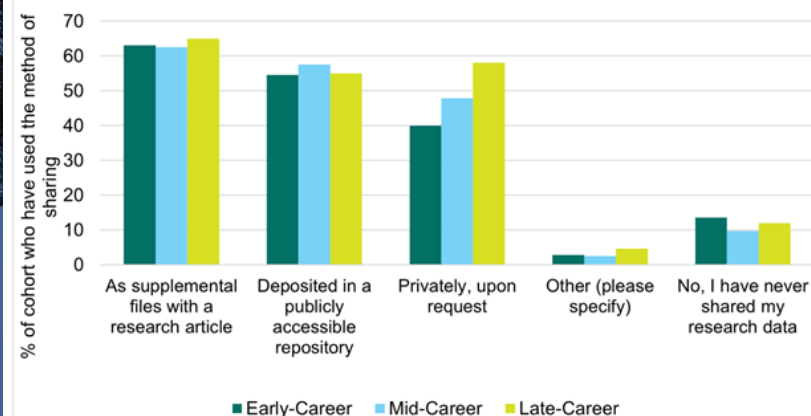
A survey of researchers' needs and priorities for data sharing

Feb 22, 2021

AUTHORS
Iain Hrynaskiewicz, James Harney, Lauren Cadwallader

	Importance ^a			Satisfaction ^a		
	mean ^a	±stdev ^a	n ^a	mean ^a	±stdev ^a	n ^a
Data Preparation^a						
Spend less time organizing my data files ^a	58.3 ^a	28.4 ^a	781 ^a	60.8 ^a	25.2 ^a	775 ^a
Spend less time deciding which datasets to share ^a	38.5 ^a	31.4 ^a	777 ^a	66.1 ^a	23.9 ^a	774 ^a
Spend less time describing my research data ^a	48.0 ^a	28.8 ^a	777 ^a	63.5 ^a	22.8 ^a	772 ^a
Prepare usage rights statement outlining conditions of use and acknowledgment ^a	55.1 ^a	31.2 ^a	743 ^a	53.1 ^a	25.7 ^a	734 ^a
Policy Requirements^a						
Spend less time preparing Data Management Plan(s) ^a	48.9 ^a	27.8 ^a	785 ^a	58.0 ^a	25.1 ^a	780 ^a
Comply with journal policies on data sharing ^a	69.1 ^a	27.9 ^a	744 ^a	67.7 ^a	25.6 ^a	738 ^a
Comply with funder policies on data sharing ^a	73.9 ^a	28.1 ^a	744 ^a	69.5 ^a	24.0 ^a	737 ^a
Comply with institutional policies on data sharing ^a	67.1 ^a	30.2 ^a	741 ^a	68.4 ^a	25.2 ^a	735 ^a
Meet funder requirements for data management plans ^a	63.1 ^a	29.5 ^a	742 ^a	65.1 ^a	23.5 ^a	735 ^a
Ensure funder knows my Data Management Plan has been followed ^a	52.9 ^a	30.0 ^a	741 ^a	61.7 ^a	22.9 ^a	732 ^a
Data Publishing^a						
Get help determining which datasets I have permission to share ^a	47.5 ^a	33.4 ^a	775 ^a	59.7 ^a	27.1 ^a	771 ^a
Spend less time finding a repository for my data ^a	43.1 ^a	31.4 ^a	774 ^a	61.3 ^a	27.7 ^a	768 ^a
Ability to place an embargo on my data ^a	44.1 ^a	34.7 ^a	740 ^a	60.6 ^a	24.8 ^a	731 ^a
Spend less time describing my supplemental files ^a	43.5 ^a	29.9 ^a	734 ^a	60.3 ^a	23.4 ^a	726 ^a
Ability to upload my data along with my article ^a	53.9 ^a	31.6 ^a	731 ^a	59.1 ^a	24.6 ^a	719 ^a
Spend less time creating a Data Availability Statement ^a	45.4 ^a	29.1 ^a	726 ^a	55.8 ^a	23.1 ^a	720 ^a
Ability to create a Data Availability Statement that includes links to my research data files ^a	50.9 ^a	30.4 ^a	727 ^a	54.3 ^a	23.8 ^a	720 ^a
Ability to create a Data Availability Statement that includes a description of each of my research data files ^a	47.0 ^a	28.7 ^a	727 ^a	53.6 ^a	22.6 ^a	720 ^a
Spend less time uploading my data files ^a	46.0 ^a	31.8 ^a	730 ^a	58.0 ^a	24.6 ^a	721 ^a
Choose an appropriate license for my data ^a	54.5 ^a	31.9 ^a	728 ^a	52.3 ^a	25.9 ^a	723 ^a
Increase the discoverability of my research data ^a	64.4 ^a	31.2 ^a	731 ^a	51.7 ^a	23.4 ^a	719 ^a
My research data has its own Digital Object Identifier (DOI) ^a	58.9 ^a	33.8 ^a	726 ^a	59.3 ^a	26.9 ^a	718 ^a
Reuse of my data^a						
Understand who is using my data sets ^a	54.9 ^a	32.5 ^a	742 ^a	46.7 ^a	27.4 ^a	733 ^a
Ability to control who can use my data ^a	40.0 ^a	35.6 ^a	742 ^a	54.3 ^a	26.6 ^a	733 ^a
Trust the researchers who request my data ^a	58.6 ^a	33.8 ^a	738 ^a	55.9 ^a	24.5 ^a	730 ^a
Increase my co-authorship opportunities ^a	53.0 ^a	34.1 ^a	741 ^a	54.1 ^a	24.7 ^a	734 ^a
Increase the likelihood that my research papers are cited ^a	70.0 ^a	27.4 ^a	728 ^a	52.8 ^a	21.9 ^a	721 ^a
Increase the likelihood that my research benefits science ^a	84.8 ^a	21.8 ^a	730 ^a	55.2 ^a	21.7 ^a	718 ^a
Ability to track downloads of my research data ^a	51.9 ^a	30.6 ^a	728 ^a	50.1 ^a	23.9 ^a	716 ^a
Ability to track citations of my research data ^a	65.2 ^a	28.3 ^a	728 ^a	54.4 ^a	25.9 ^a	719 ^a

Methods of data sharing previously used



... «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](#)

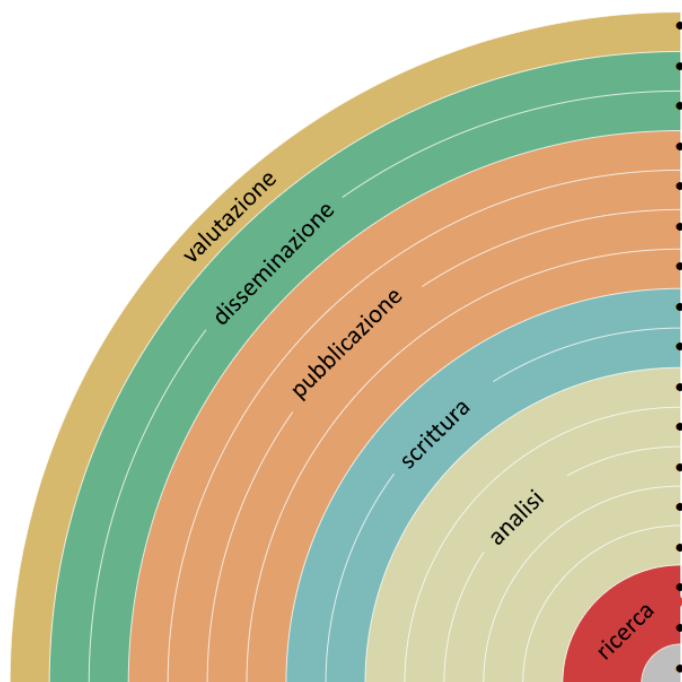
RETWEET
76

MI PIACE
32



Open science: perché solo i dati?

Come puoi 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, es. su OSF, 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. OpenNotebookScience
- condividendo software, es. su GitHub con licenza GNU/MIT
- condividendo i dati, es. su Dryad, Zenodo o Dataverse
- pre-registrando esperimenti, es. su OSF 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](https://doi.org/10.5281/zenodo.1147025)

<https://doi.org/10.5281/zenodo.1195647>

Traduzione: Elena Giglia



DOI: [10.5281/zenodo.1195648](https://doi.org/10.5281/zenodo.1195648)

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

UN MODO STRUTTURATO
DI PENSARE AI DATI

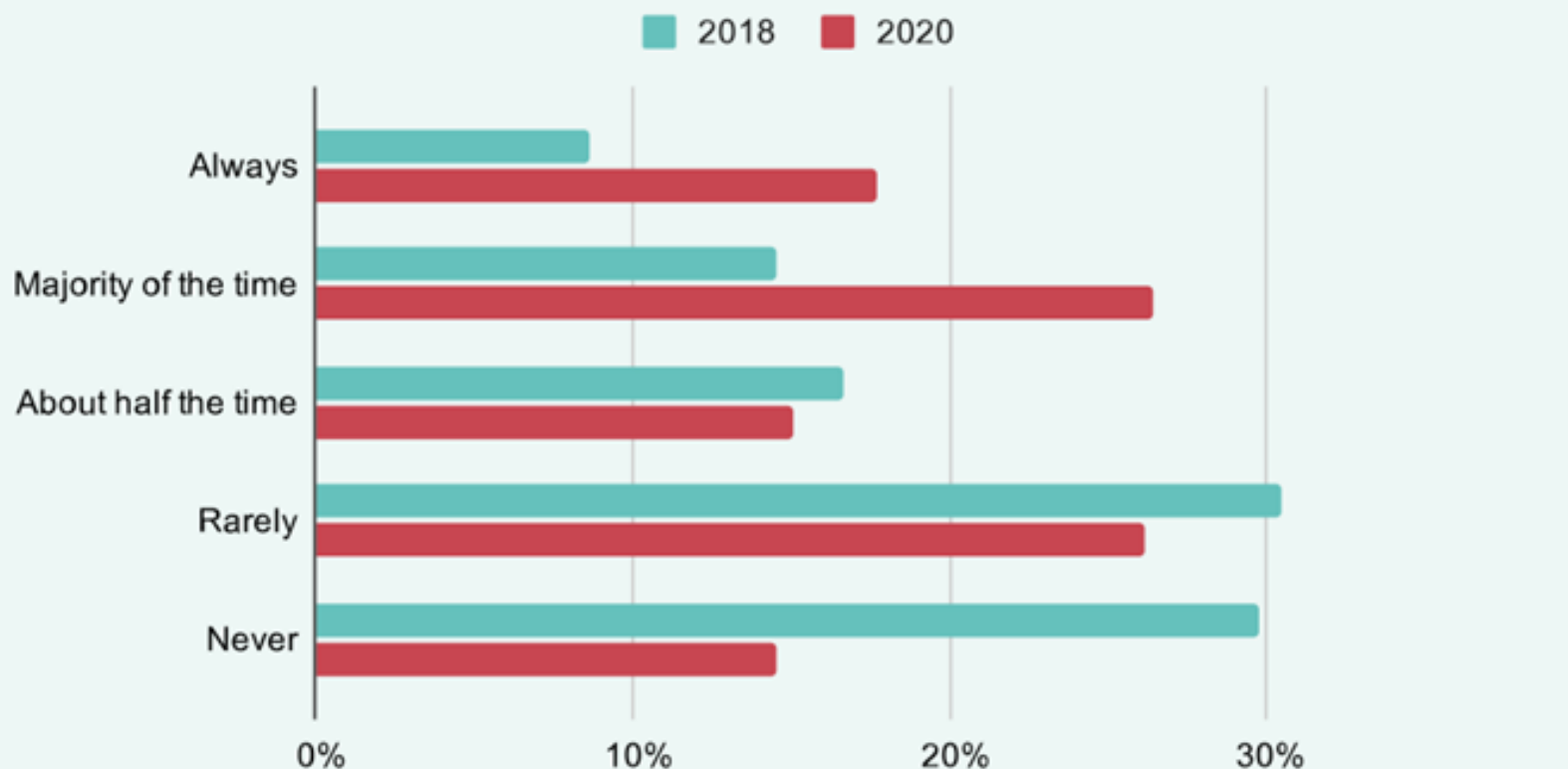
REGOLE CHIARE=MENO
ERRORI DA SUBITO

È UN «LIVING DOCUMENT»,
CRESCe COL PROGETTO

DOVE METTERE TUTTE QUESTE INFORMAZIONI?
NEL DATA MANAGEMENT PLAN

DMP in crescita

How often do you create a data management plan for the research you carry out? by Year



Trucchi e suggerimenti

SINTETICO E
SPECIFICO

Top tip - keep it short and specific!

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



NON COPIATE

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

ESSERE GENERICI NON SERVE
A NULLA
[we expect a huge size of data;
data will be available]

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

CIÒ CHE DICHIARATE NEL
DMP POI VA FATTO
VERAMENTE... QUINDI
A) NON FATE GLI
SPLENDIDI
B) NON IMPEGNATEVI A
FARE COSE CHE
SAPETE IMPOSSIBILI
Es. DATI PSEUDONIMIZED, non
ANONIMIZED

1

START EARLY

Read the guidance and ask for advice early on in the process, as writing a DMP may take some time

2

CONSIDER REUSE

Think about reusing existing data. Describe what you will need to know about your data five years from now

3

CHECK POLICIES

Talk to your supervisor or lab members about existing data management policies and standards

4

MAKE USE OF SUPPORT

Use your in-house support services like RDM Support, the Library, IT department or legal desk

5

THINK BROAD

Also address software code, algorithms and any other valuable research assets in your DMP

6

COPY WHERE YOU CAN

Look at other (submitted) plans and copy when appropriate

7

BE UNIQUE WHERE NEEDED

Since every research project is unique, so are the data it generates. Copying from sample DMPs is not sufficient

8

BE CONCRETE

Make your answers as concrete as possible. Show that you have consulted RDM experts

9

SAY SO IF YOU DON'T KNOW

Indicate what you do not yet know and how you will resolve these questions later

10

UPDATE

DMPs add to the planning of your research methods. Therefore define, carry out and update your DMP just as you would any method

Tips&tricks / 2

A data planning process ensures that all aspects of data management are holistically explored at the start of a project. Short-term and long-term aims can be balanced, so that decisions made early in a project do not negatively impact on the ability to find and use the research data in future.

Effective management of data provides researchers with many benefits, including

- time saved through reduced duplication of effort
 - decreased risk of loss, theft or inappropriate use of data
 - good research practice ensures the integrity and quality of data
 - data can be understood and used now and in the future
-
- helps researchers find and gain access to data management – expertise and infrastructure offered at the University
 - increased researcher profile through data dissemination and re-use.

A data planning process is particularly important in the context of collaborative research projects. Researchers may identify areas of potential difficulty or conflict, and these can be resolved with colleagues and collaborators before they escalate into issues. Clarifying ownership of data, and ensuring early agreement on technical standards and frameworks across institutions, are an important part of establishing trust and ensuring that a project runs smoothly.

EVITA SFORZI DI
DUPLICAZIONE
EVITA LA PERDITA
RISOLVE I
CONFLITTI

Vantaggi di un DMP

CESSDA Guide



⊖ Benefit 3. Clarifies needed budget


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

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

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

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

Guida al DMP



SCIENCE
EUROPE


Jan. 27, 2021

Contact us | Member log in | Q

ABOUT US OUR PRIORITIES WHAT'S GOING ON OUR RESOURCES

> Our resources

27.01.2021



SCIENCE EUROPE
PRACTICAL GUIDE TO
THE INTERNATIONAL ALIGNMENT OF
RESEARCH DATA MANAGEMENT
Extended Edition
with DMP Evaluation Rubric

Practical Guide to the International Alignment of Research Data Management - Extended Edition

This resource offers targeted guidance for organisations, scientific communities, as well as individual researchers, to organise research data and preserve it appropriately

Originally released in 2019, and following its successful uptake by many organisations, the extended edition features a brand-new rubric to facilitate the evaluation of a data management plan (DMP). The guide also presents core requirements for DMPs, criteria for the selection of trustworthy repositories, and guidance to comply with organisational requirements.

EDIZIONE AGGIORNATA
DELLA GUIDA 2018

DMP Core Requirements

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS



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



1. Data description and collection or re-use of existing data

- How will new data be collected or produced and/or how will existing data be re-used?
- What data (for example the kinds, formats, and volumes) will be collected or produced?



2. Documentation and data quality

- What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
- What data quality control measures will be used?



3. Storage and backup during the research process

- How will data and metadata be stored and backed up during the research process?
- How will data security and protection of sensitive data be taken care of during the research?



4. Legal and ethical requirements, codes of conduct

- If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
- How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
- How will possible ethical issues be taken into account, and codes of conduct followed?

SEZIONI MINIME IN UN DMP



5. Data sharing and long-term preservation

- How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
- How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
- What methods or software tools will be needed to access and use the data?
- How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?



6. Data management responsibilities and resources

- Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
- What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS



SCIENCE EUROPE

PRACTICAL GUIDE TO
THE INTERNATIONAL ALIGNMENT OF
RESEARCH DATA MANAGEMENT

Extended Edition
with DMP Evaluation Public

Jan. 27, 2021



DMP Core Requirements

Translating the Core Requirements into a DMP template

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

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

GENERAL INFORMATION

Administrative information

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

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

1 a

How will new data be collected or produced and/or how will existing data be re-used?

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

2 DOCUMENTATION AND DATA QUALITY

2 a

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

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

2 b

What data quality control measures will be used?

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

UTILI COME
SPUNTO PER LE
PRIME VOLTE,
PER CAPIRE
QUALI DOMANDE
FARSI



Jan. 27, 2021

DMP – Rubric for evaluation



3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS

Guidance for Researchers

3a

How will data and metadata be stored and backed up during the research?

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

Sufficiently Addressed The DMP...

- Clearly (even if briefly) describes:
 - › The location where the data and backups will be stored during the research activities.
 - › How often backups will be performed.
 - › The use of robust, managed storage with automatic backup (for example storage provided by the home institution).

or

- Explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.

Insufficiently Addressed The DMP...

- Provides no information or very vague reference to how data will be stored and backed up during the project.

Guidance for Researchers

1a

How will new data be collected or produced and/or how will existing data be re-used?

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

Sufficiently Addressed The DMP...

- Gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used.
- Explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use.

Insufficiently Addressed The DMP...

- Provides little or no details on where the data come from and what data will be collected or re-used.
- Does not, if applicable, provide sufficient rationale for generating new data.

DMP questions

CESSDA guide

Adapt your Data Management Plan

A list of Data Management Questions based on the Expert Tour Guide on Data Management



Overview

Title of the project

Date of this plan

Description of the project

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

Origin of Data

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

Principal researchers

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

Collaborating researchers (if applicable)

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

Funder (if applicable)

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

Data producer

- Which organisation has the administrative responsibility for the data?

Project data contact

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

Data owner(s)

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

Roles

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

Costs

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

UTILI COME
SPUNTO PER LE
PRIME VOLTE,
PER CAPIRE
QUALI DOMANDE
FARSI

Organising and documenting your data

Data collection

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

Data organisation

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

Data type and size

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

File format

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

Folder structure and names

- How will you structure and name your folders?

File structure and names

- How will you structure and name your files?

Documentation

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

Metadata

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

Metadata standard (if applicable)

- What metadata standard(s) will you use?

ORGANISE &
DOCUMENT

DMP questions

Processing your data

Versioning

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

Interoperability

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

If applicable:

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

Storing your data

Storage

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

Backup

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

Security

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

Protecting your data

Ethical review (if applicable)

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

Informed consent (if applicable)

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

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

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

Intellectual property rights (IPR)/Copyrights

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

Agreements (if applicable)

- What are the agreements with other stakeholders?

Restrictions (if applicable)

- Any other restrictions that need to be considered?

UTILI COME
SPUNTO PER LE
PRIME VOLTE,
PER CAPIRE
QUALI DOMANDE
FARSI

Basic Information.

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

Volume and Life Cycle of the Data.

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

For RAW data, please consider the following:

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

For Secondary and Published data, please consider the following:

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

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

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

Data Management Checklist

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

UTILI COME
SPUNTO PER LE
PRIME VOLTE,
PER CAPIRE
QUALI DOMANDE
FARSI

DMP online

PERSONALIZZABILE

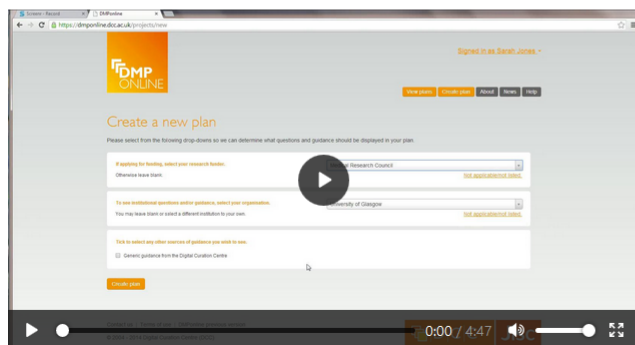


Home About Future plans Help Change language

Welcome.

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

Screencast on how to use DMPonline



Sign in

Veteran tapes

Project Details Plan overview Write Plan Share Download

expand all | collapse all

13/13 answered

Data Collection (2 / 2)

What data will you collect or create?

B *I*

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

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

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

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

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

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

Save

Guidance

Comments (1)

DCC

DCC guidance

Guidance

Questions to consider:

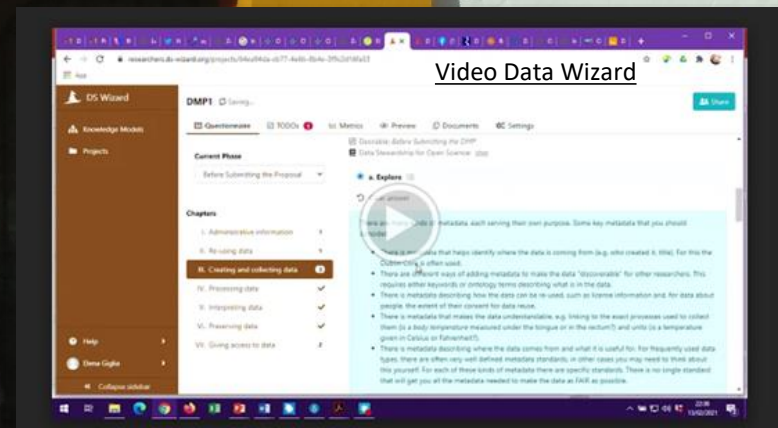
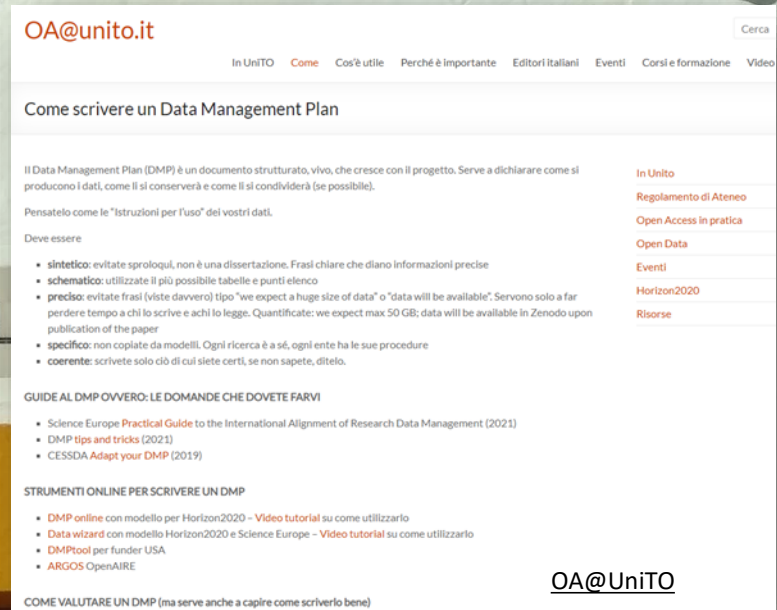
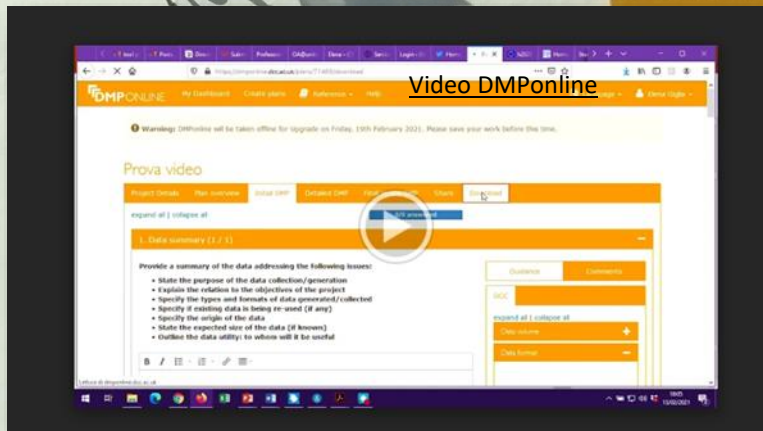
- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

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

GRATUITO
BASTA REGISTRARSI; POI SI
ACCEDE E SI TROVANO TUTTI I
PROPRI DMP NEL DESKTOP

DMP online e Data Wizard video



Create a new plan

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

* What research project are you planning?

☒ mock project for testing, practice, or educational purposes

* Select the primary research organisation

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

* Select the primary funding organisation

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

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

PREDISPOSTO PER
LE 3 VERSIONI

ALLA FINE, SCARICA
IL PDF

CONDIVISIBILE
PER
SCRIVERLO
INSIEME

IDENTIFICATIVI

GUIDA

Prova

Project Details

Plan overview

Initial DMP

Detailed DMP

Final review DM

Share

Download

* Project title

☒ mock project for testing, practice, or educational purposes

Funder

Grant number

Project abstract

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

[See the full list](#)

Save

Guidance

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

5 organisations to see their

Curation Centre

from additional organisations

ID

Principal Investigator

Name

ORCID iD

Email

Phone

Data Contact Person

☒ Same as Principal Investigator

Save

[Project Details](#)[Plan overview](#)[Initial DMP](#)[Detailed DMP](#)[Final review DMP](#)[Share](#)[Download](#)

Set plan visibility

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

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

PRIVATO O PUBBLICO?

Manage collaborators

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

Email address	Permissions
elena.giglia@unito.it	Owner

Invite collaborators

* Email

* Permissions

- ☐ Co-owner
- ☐ Editor
- ☐ Read only

SI POSSONO INVITARE
COLLABORATORI CON LIVELLI DI
PERMESSO DIVERSI

Horizon 2020 DMP

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

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

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

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

Template version 1, published on 16 May 2019

Initial DMP (6 sections, 9 questions) +

Detailed DMP (9 sections, 31 questions) +

Final review DMP (9 sections, 31 questions) +

Initial DMP (6 sections, 9 questions)

Instructions

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

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

Write plan

1. Data summary

Provide a summary of the data addressing the following issues:

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

2. FAIR data

2.1 Making data findable, including provisions for metadata:

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

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

2.2 Making data openly accessible:

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

2.3 Making data interoperable:

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

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

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

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

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

GUIDA E
SUGGERIMENTI

expand all | collapse all

0/9 answered

1. Data summary (0 / 1)

2. FAIR data (0 / 4)

3. Allocation of resources (0 / 1)

4. Data security (0 / 1)

5. Ethical aspects (0 / 1)

6. Other (0 / 1)

...E NON «A HUGE
AMOUNT OF
DATA»

Data volume

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).

Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?

COSTI

- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

expand all | collapse all

0/9 answered

1. Data summary (0 / 1)

GUIDA RAPIDA

Provide a summary of the data addressing the following issues:

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

= README
FILE

Guidance

DCC

expand all | collapse all

Data volume

Data format

Data description

Data format




TABELLA PER DATI DI
FORMATO DIVERSO

- Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tiff).

- Explain why you have chosen certain formats. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community.

- Using standardised, interchangeable or open formats ensures the long-term usability of data; these are recommended for sharing and archiving.

- See UK Data Service guidance on [recommended formats](#) or DataONE Best Practices for [file formats](#).

B I    

SCRIVETE
DIRETTAMENTE
(USATE TABELLE E
PUNTI ELENCO)

ALTRE GUIDE
SPECIFICHE

USATE FORMATI
STANDARD (CON
ELENCO)

Save

IN OGNI PASSAGGIO POTETE
SALVARE E CONTINUARE DOPO

2. FAIR data (0 / 4)

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

2.1 Making data findable, including provisions for metadata:

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

Guidance





Comments

EC

DCC

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





APRE LA
DIRECTORY DEI
METADATA
STANDARDS

B I    

Save

2.2 Making data openly accessible:

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

B I    

Save

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

Guidance

Comments

EC

DCC

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

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

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

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

APRE REGISTRY OF
REPOSITORIES

2.3 Making data interoperable:

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

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

Save

Guidance

Comments

EC

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

COMMENTI

Guidance

Comments

Add comments to share with collaborators

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

Save

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

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

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

Save

EC

DCC

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

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

29.3 Open access to research data

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

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

RICORDATE LORO CHE HANNO FIRMATO UN GRANT AGREEMENT...

3. Allocation of resources (0 / 1)

Explain the allocation of resources, addressing the following issues:

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

Guidance

Comments

EC

DCC

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

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

RICORDARE SEMPRE
CHE EVENTUALI
COSTI SONO
RIMBORSABILI

Save

4. Data security (0 / 1)

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

Guidance

Comments

EC

DCC

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

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

Save

5. Ethical aspects (0 / 1)

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

Guidance

Comments

EC

DCC

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

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

Save

6. Other (0 / 1)

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

Guidance

Comments

DCC

Related policies

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

QUI SAREBBE
URGENTE UNA
POLICY DI ATENEO

Save

Project Details

Plan overview

Initial DMP

Detailed DMP

Final review DMP

Share

Download

Download settings

Select phase to download Initial DMP ▼

Optional Plan Components

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

Format

pdf ▼

PDF formatting

Font

Face

Arial, Helvetica, Sans-Serif ▼

Size (pt)

10 ▼

Top

25 ▼

Bottom

20 ▼

Left

12 ▼

Right

12 ▼

Margin (mm)

Download Plan

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

DMP in Horizon Europe

IN HORIZON EUROPE
- NELLA PROPOSTA 1 PAGINA
SULLA GESTIONE DEI DATI
- DMP COME DELIVERABLE



⚠ *Proposals selected for funding under Horizon Europe will need to develop a detailed data management plan (DMP) for making their data/research outputs findable, accessible, interoperable and reusable (FAIR) as a deliverable by month 6 and revised towards the end of a project's lifetime.*

⚠ *For guidance on open science practices and research data management, please refer to the relevant section of the [HE Programme Guide](#) on the Funding & Tenders Portal.*

Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- establish a data management plan ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements

A wooden bench made of thick, weathered planks sits on a brick-paved surface. A sign made of four vertical wooden planks is placed in front of the bench. The sign has the text "IF YOU ARE NOT DOING WHAT YOU LOVE, YOU ARE WASTING YOUR TIME." written on it in bold, black, sans-serif capital letters. The background shows a brick wall and a paved area.

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

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