FAIR PRINCIPLES

ICTP, November 23 2021

Elena Giglia
elena.giglia@unito.it
@egiglia

The road ahead

…but there is a lot more to discuss, in Open Science

1 Open Access (why, what and how)

2 FAIR principles

Why do we need FAIR

[BUT FIRST: EVER LOST YOUR OWN DATA/COULD NO LONGER ACCESS/UNDERSTAND?]

DATA ARE DIFFICULT TO FIND

ONCE FOUND, DIFFICULT TO ACCESS

ONCE ACCESSED, DIFFICULT TO INTERPRET

AND IF NOT FOUND YOU SPEND A LOT OF TIME IN RE-CREATING THEM

THAT'S WHY WE NEED FAIR

[the 3 steps]

1. DATA SHOULD BE «AS OPEN AS POSSIBLE»

2. BUT IF DATA ARE NOT «FAIR», OPENING IS RISKY (MISUSE, MISINTERPRETATION, …)

3. IF DATA ARE NOT PROPERLY MANAGED FROM THE BEGINNING, IT'S ALMOST IMPOSSIBLE TO MAKE THEM «FAIR»

[WITH EOSC MANAGED/FAIR INCREASINGLY OVERLAPPING, «FAIR BY DESING»]

AND MANAGING DATA PROPERLY IS IN THE PRIMARY INTEREST OF ANY RESEARCHER, AS THE WHOLE RESEARCH PROCESS RESULTS STREAMLINED AND MORE EFFECTIVE

Open Science in Horizon Europe

Open Science in Horizon Europe RIA/IA/CSA

Guide to Os in HEurope

IN THE METHODOLOGY YOU NEED TO ADDRESS BOTH:
1) HOW YOU WILL COMPLY WITH THE MANDATORY PRACTICES
2) HOW YOU WILL ADOPT RECOMENDED PRACTICES

RECOMMENDED PRACTICES

- IN THE **LIST OF ACHIEVEMENTS:** 5 RELEVANT OUTPUTS (publications, data) OPENLY ACCESSIBLE + PERSISTENT IDENTIFIER + «AS OPEN AS POSSIBLE» → **LIST OF ACHIEVEMENTS** Template PartA
- IN THE PROJECT **METHODOLOGY** 1) EMBEDDED OPEN SCIENCE PRACTICES 2) FAIR DATA MANAGEMENT + DMP SCHEMA → **EXCELLENCE** Template PartB
- **MAXIMIZING IMPACT** USING OPEN SCIENCE (OS IS AMONG KEY PATHWAY INDICATORS) + SCHEMA OF DISSEMINATION PLAN (DELIVERABLE M6) → **IMPACT** Template PartB
- OPEN SCIENCE PRACTICES/SKILLS IN PREVIOUS PROJECTS TO EVALUATE **QUALITY OF IMPLEMENTATION AND CONSORTIUM CAPACITY** → **QUALITY OF IMPLEMENTATION** Template PartB

MANDATORY PRACTICES

- **DEPOSIT+ IMMEDIATE ACCESS** (ZERO EMBARGO + CC BY) = 1. OPEN RESEARCH EUROPE 2. OA JOURNAL 3. TRADTIONAL JOURNAL [RETAINING RIGHTS] → **OPEN SCIENCE** Publications
- 1. RESPONSIBLE MANAGEMENT ACCORDING TO **FAIR PRINCIPLES** 2. DATA AND OTHER OUTPUTS **«AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»** 3. DATA MANAGMENT PLAN BY M6 → **OPEN SCIENCE** FAIR data
- **INFORMATION** ON OUTPUTS/TOOLS AND **ACCESS** TO DATA/RESULTS FOR **VALIDATION** OF RESEARCH → **ENSURE** REPRODUCIBILITY

PROJECT PROPOSAL **WILL BE EVALUATED ON**
a) HOW IT WILL ADOPT RECOMMENDED PRACTICES AND b) HOW IT WILL BE COMPLIANT TO MANDATORY ONES

OPEN SCIENCE IS A METHODOLOGY.
THAT'S WHY IN HORIZON EUROPE IT HAS BEEN MOVED TO THE **«SCIENTIFIC EXCELLENCE»** SECTION OF THE PROPOSAL TEMPLATE.
YOUR PROPOSAL WILL BE EVALUATED ALSO ON HOW YOU ADAPT/ADOPT OPEN SCIENCE PRACTICES AND A RESPONSIBLE FAIR DATA MANAGEMENT

…FAIR means [for machines]


SCIENTIFIC DATA

We'd like to understand how you use our websites in order to imp

Open Access | Published

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

FAIR guide, Nature, March 2016

FINDABLE

- IDENTIFIERS
- METADATA

ACCESSIBLE

- WHERE TO FIND THE DATA AND UNDER WHAT ACCESS CONDITIONS
- **NOT «OPEN»**
- OPEN FORMATS

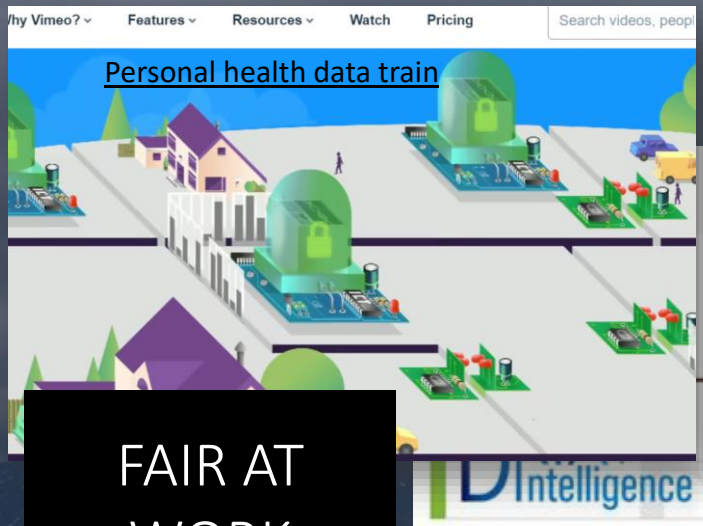
INTEROPERABLE

- STANDARDS
- ONTOLOGIES

REUSABLE

- LICENSES
- DOCUMENTATION

MACHINE-READABLE



FAIR AT WORK

2020

Issues Online Early About Submit

January 01 2020

FAIR Principles: Interpretations and Implementation Considerations

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagna, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Sch, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Egon L. Willighage, Barend Mons, Erik Schultes

Previous Article Next Article

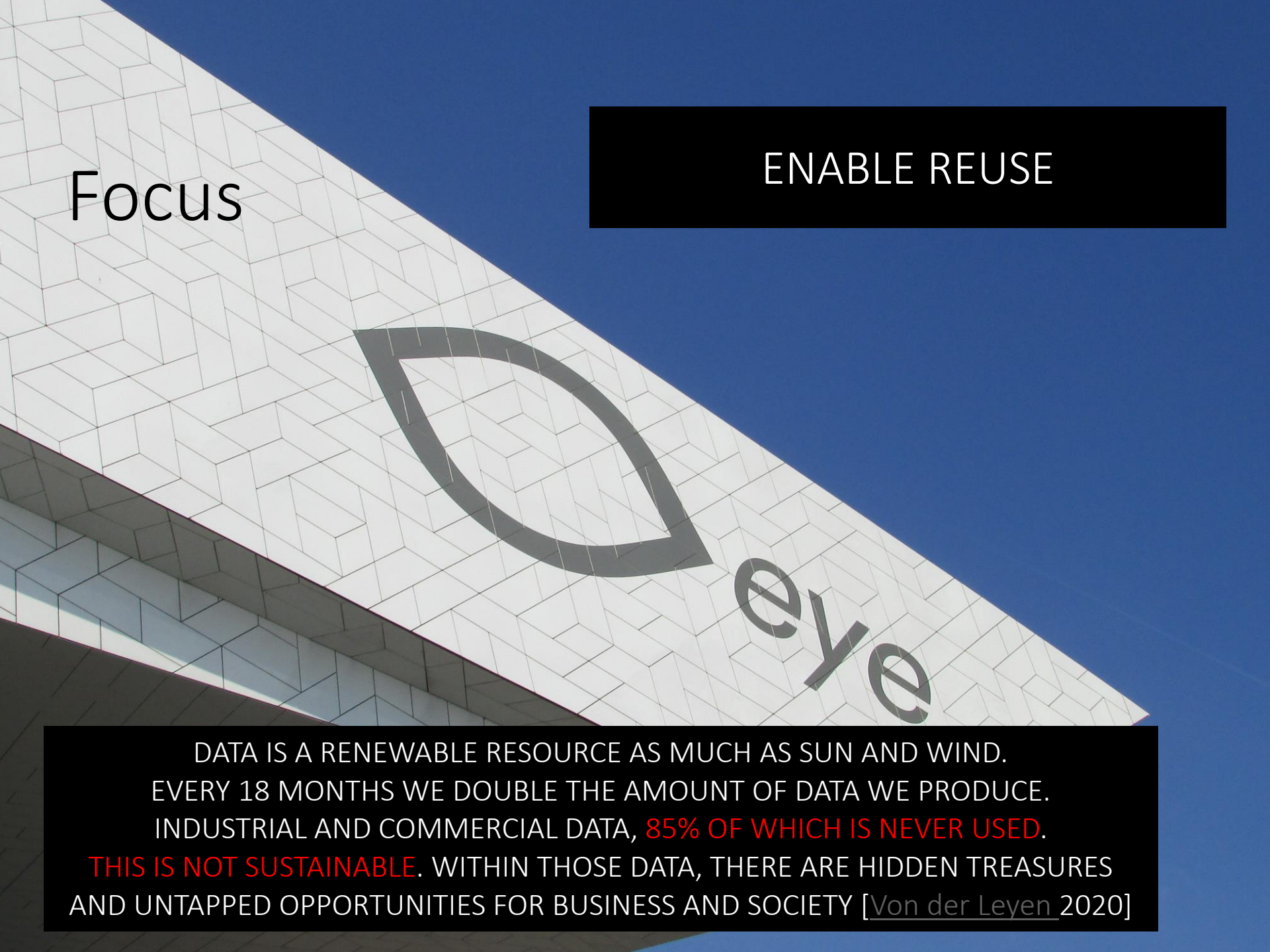
Author and Article Information

Article Contents

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

NO MISTAKES!

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

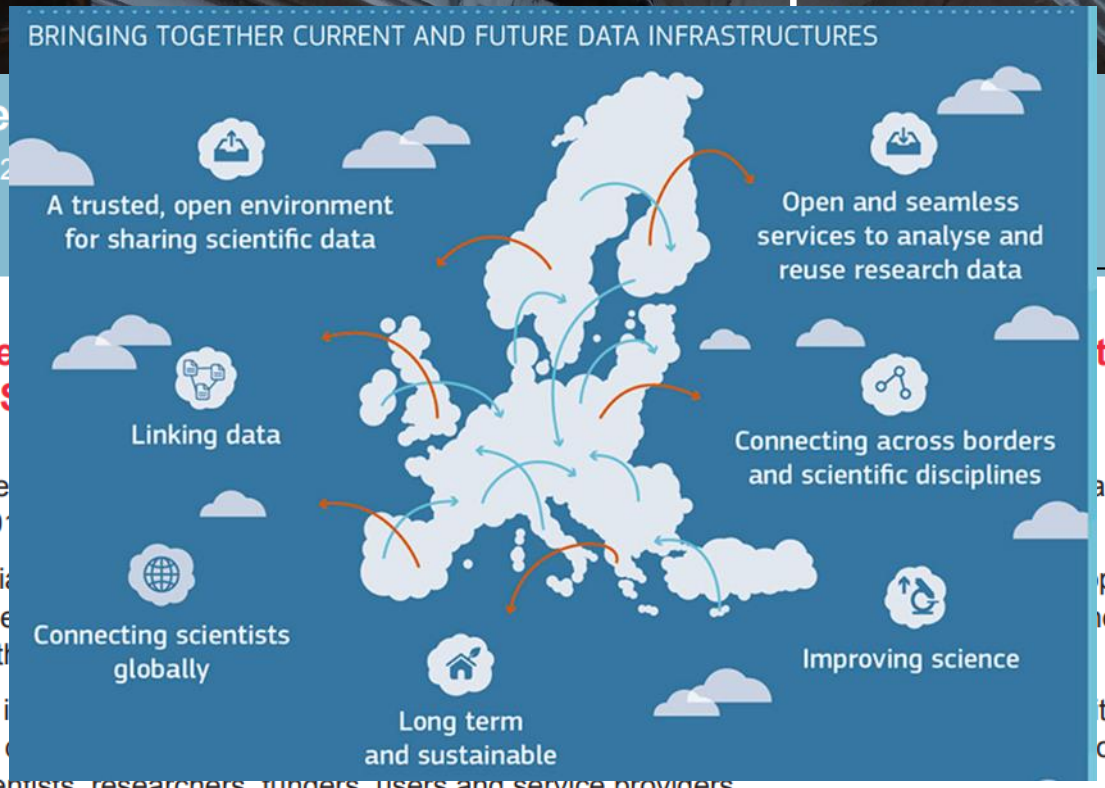


Focus

ENABLE REUSE

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]

[THAT'S WHY THE EOSC!]



The Vienna De

Vienna, 23 November 2

u 1 8 . a t

Nov.23, 2018

We, Ministers, de … the European Open S…

1. Recall the challenge… aration" signed in Brussels on 10 July 201…

2. Reaffirm the potenti… pe. Confirm that the vision of the Europe… es and Member States, sustainable in t…

3. Recognise that the i… ts nature iterative and based on … o build trust and consensus among scientists, researchers, funders, users and service providers.

4. Highlight that Europe is well placed to take a global leadership position in the development and application of cloud services for Science. Rea… and open to the world, reaching out over time to …

5. Recall that the Council … roadmap and the federated

SEAMLESS ACCESS
TO OPEN BY DEFAULT FAIR DATA

9. Call for the European Open Science Cloud to provide all researchers in Europe with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing and processing research data supported by FAIR data principles.

Science Cloud a reality, hinting at the need to further strengthen the ongoing dialogue across institutions and with stake-holders, for a new governance framework to be launched in Vienna, on 23 November 2018.

EOSC IS NOT A BIG BOX

THE EUROPEAN OPEN SCIENCE CLOUD? SOME NUANCES AND DEFINITIONS

Imagine a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes. Imagine that this all operates under well-defined and trusted conditions, supported by a sustainable and just value for money model. This is the environment that must be fostered in Europe and beyond to ensure that European research and innovation contributes in full to knowledge creation, meet global challenges and fuel economic prosperity in Europe. This we

EOSC IS NOT A REPOSITORY NOR A «CLOUD»

YOU MAKE YOUR DATA FAIR SO THAT EOSC *SERVICES* CAN «FIND» THEM…

A SUPPORTING ENVIRONMENT FOR OPEN SCIENCE AND NOT AN «OPEN CLOUD» FOR SCIENCE

YOU DON'T «UPLOAD» YOUR DATA INTO EOSC

AND GIVE SEAMLESS ACCESS TO 20 M EU RESEARCHERS

EOSC SRIA 1.0

FAIR principles

«ACCESSIBLE» DOES NOT MEAN «OPEN».

DATA CAN BE CLOSED, PROVIDED YOU – AND MACHINES - KNOW WHERE TO FIND THEM AND UNDER WHAT ACCESS CONDITIONS

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

FAIR/Open

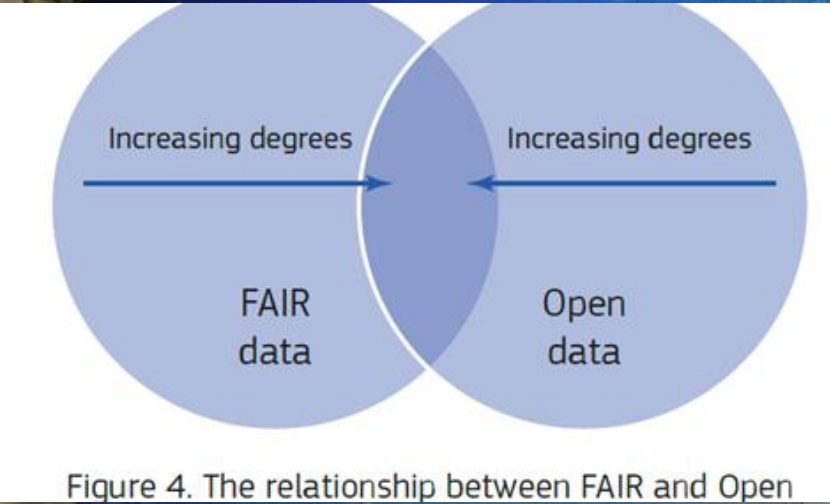


Figure 4. The relationship between FAIR and Open

THERE WILL BE AN INCREASING DEGREE IN OVERLAPPING.
BUT WE'LL ALWAYS HAVE PERFECTLY FAIR CLOSED DATA

Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud

2017

Article type: Research Article

Authors: Mons, Barend[a; b; c; *] | Neylon, Cameron[d] | Velterop, Jan[e] | Dumontier, Michel[f] | da Silva Santos, Luiz Olavo Bonino[b; g] | Wilkinson, Mark D.[h]

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 resourc achieve – possibly incrementally – their optimal discovery and scholarly reuse.

FAIR is not equal to RDF, Linked Data, or the Semantic Web The reference artic 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 guidel 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 fo authors' position clearly: "The re data publication autonomously, Principles. Computers are now a recent surveys, the time reported dealing with discovering and reu been pegged at 80% [19]. Were th deal with FAIR data and services, is today. The avoidance of time-v stewardship. To serve this potent and services should be actionable wherever possible.

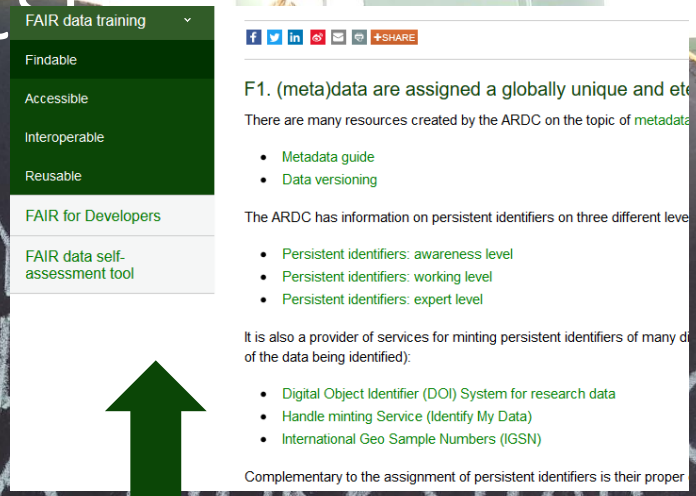
FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles,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

- PRINCIPLES, NOT STANDARD [IMPLEMENTATION NEEDED]
- NOT JUST FOR HUMANS
- NOT EQUAL TO LINKED DATA, RDF...
- NOT EQUAL TO «OPEN»

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.

FAIR in a nut



https://www.ands.org.au/working-with-data/fairdata/training

FAIR software

IOS Press Content Library

Towards FAIR principles for research software

Data Science, June 12 2020

Issue title: FAIR Data, Systems and Analysis

Guest editors: Paul Groth and Michel Dumontier

Article type: Position Paper

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

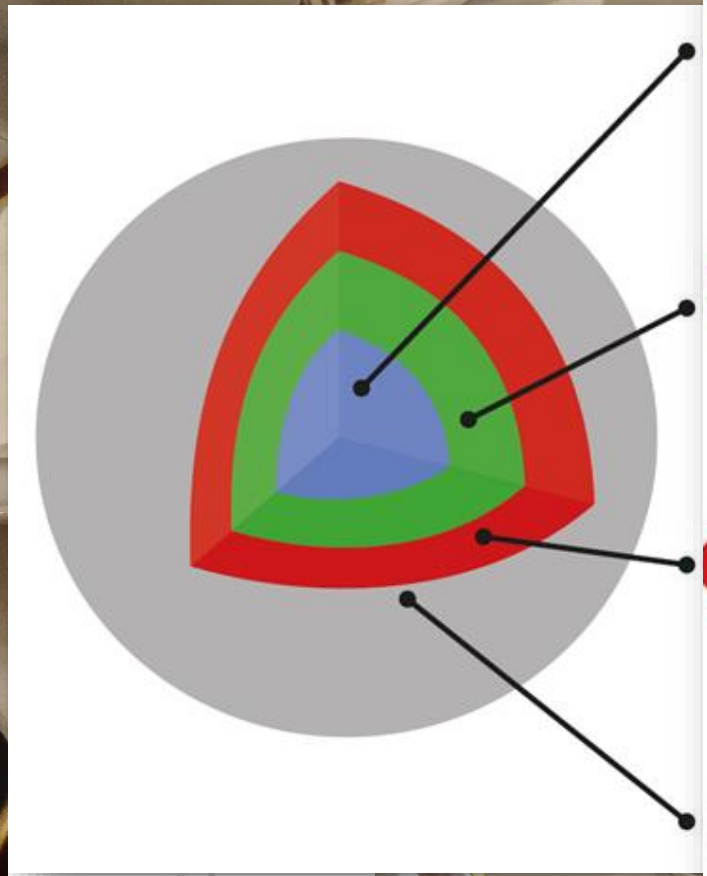
2. Software is not data

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 the
such as the possibility of having a Digit
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



DIGITAL OBJECT

Data, code and other research outputs

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

IDENTIFIERS

Persistent and unique (PIDs)

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

STANDARDS & CODE

Open, documented formats

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

METADATA

Contextual documentation

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

Nov. 2[illegible] 2018

… a report from Vienna

FAIR: technology VS domain

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

FAIR GENERIC VS
DOMAIN SPECIFIC
STRICTLY INTERLINKED

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

RESEARCHERS' RESPONSIBILITY

REPOSITORIES TAKE CARE OF…

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

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 fi les

COMMON RULES
FOR COMMUNITIES
- CONVERGENCE

International Conference on Conceptual Modeling
ER 2020: Advances in Conceptual Modeling pp 138-147 | Cite as

2020

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence

Authors | Authors and affiliations

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

Share

Questionnaire | Metrics | Preview | Documents | Settings

View | Comments | TODOs | Version history

Chapters

I. Background: The FAIR Implementation Profile and FAIR Implementation Community ✓
II. FAIR Implementation Community 2
III. Findability 6
IV. Accessibility 5
V. Interoperability 6
VI. Reusability 4
VII. Register a new resource as a nanopublication ✓

III. Findability

"Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process." - from FAIR Principles: Interpretations and Implementation Considerations

1 F1 What globally unique, persistent, resolvable identifiers do you use for metadata records?

Different identifier service providers offer different kinds of resolution services impacting F2 and I. One example would be a Digital Object Identifier (DOI).

You can answer by indicating **only a single resource at a time**. Please follow this priority:

1. *Consideration*: Please provide comments about the reasons why your community chose to use this resource
2. *Wikidata*: Check for existing resources from Wikidata (by typing in that field a keyword and clicking enter, you will find related resources). Alternatively, you may provide here an IRI for a resource from any other third party source. If not found there then follow point 3 below.
3. *Nanobench*: Check for existing resources from Nanobench (by clicking in that field, available resources

FIP Wizard

Knowledge Models
FIPs
Create a FIP

Social Science Survey Research_V1

Questionnaire | Metrics | P

View

Current Phase

Before Submitting the Proposal

Chapters

I. Background: The FAIR Implementation Profile and FAIR Implementation Community ✓
II. FAIR Implementation Community ✓
III. Findability 25

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

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

FAIR Implementation Profiles

FAIR Implementation Profile

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

[Data st… profile]

Strategic Research and Innovation Agenda (SRIA)
of the
European Open Science Cloud (EOSC)
Version 1.0 15 February 2021

SRIA 1.0

7.4. *Critical success factors*

The developments and expected impacts described above will not happen spontaneously. For these benefits to materialise a number of critical success factors (CSFs) must be in place. The following CSFs have been identified for EOSC:

- Researchers performing publicly funded research make relevant results available as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to help implement FAIR principles and support Open Science;

KØBENHAVNS UNIVERSITET 06/04/2020 8

Competence Profile

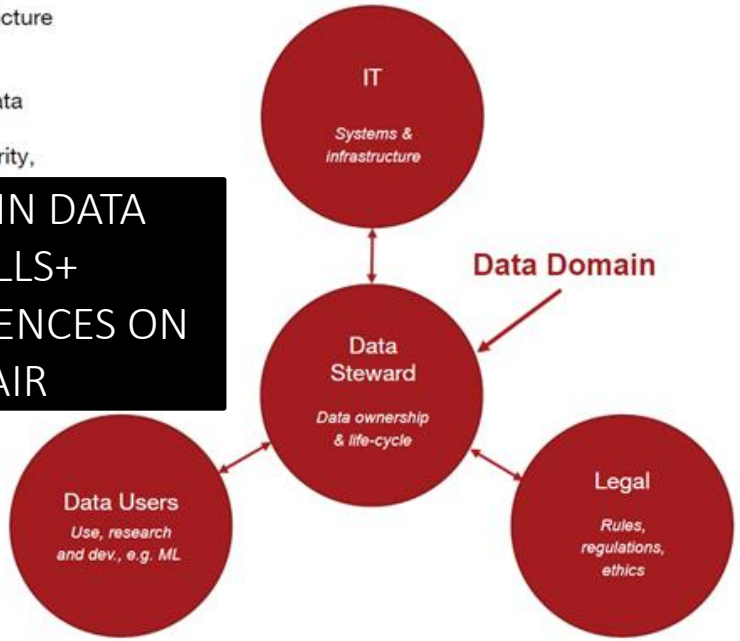
A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

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 … externally in the organisation.

Competence profile examples

- Domain-specific data understanding
- Ability to ensure that structured and unstructured data … 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.

DOMAIN DATA SKILLS+ COMPETENCES ON FAIR



Copenhagen Univ. June 17 2020

… FAIRy shades

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

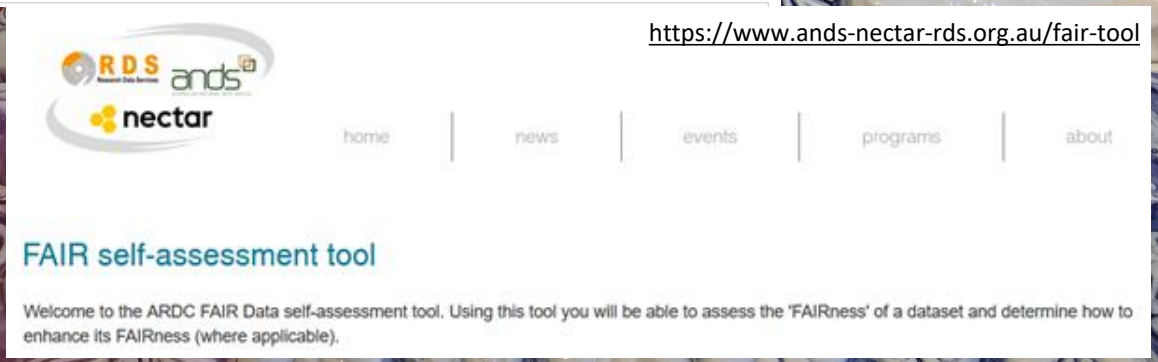
How accessible is the data? — No access to data or metadata

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

VERY USEFUL TO ASK THE RIGHT QUESTIONS BUT IT'S SUBJECTVE…

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



FAIR aware

- QUESTIONS
- TEST KNOWLEDGE
- TEST WILLINGNESS
- GIVES INFO

FAIR Aware

Data Archiving and Networked Services DANS

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.

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?

F

FINDABLE

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

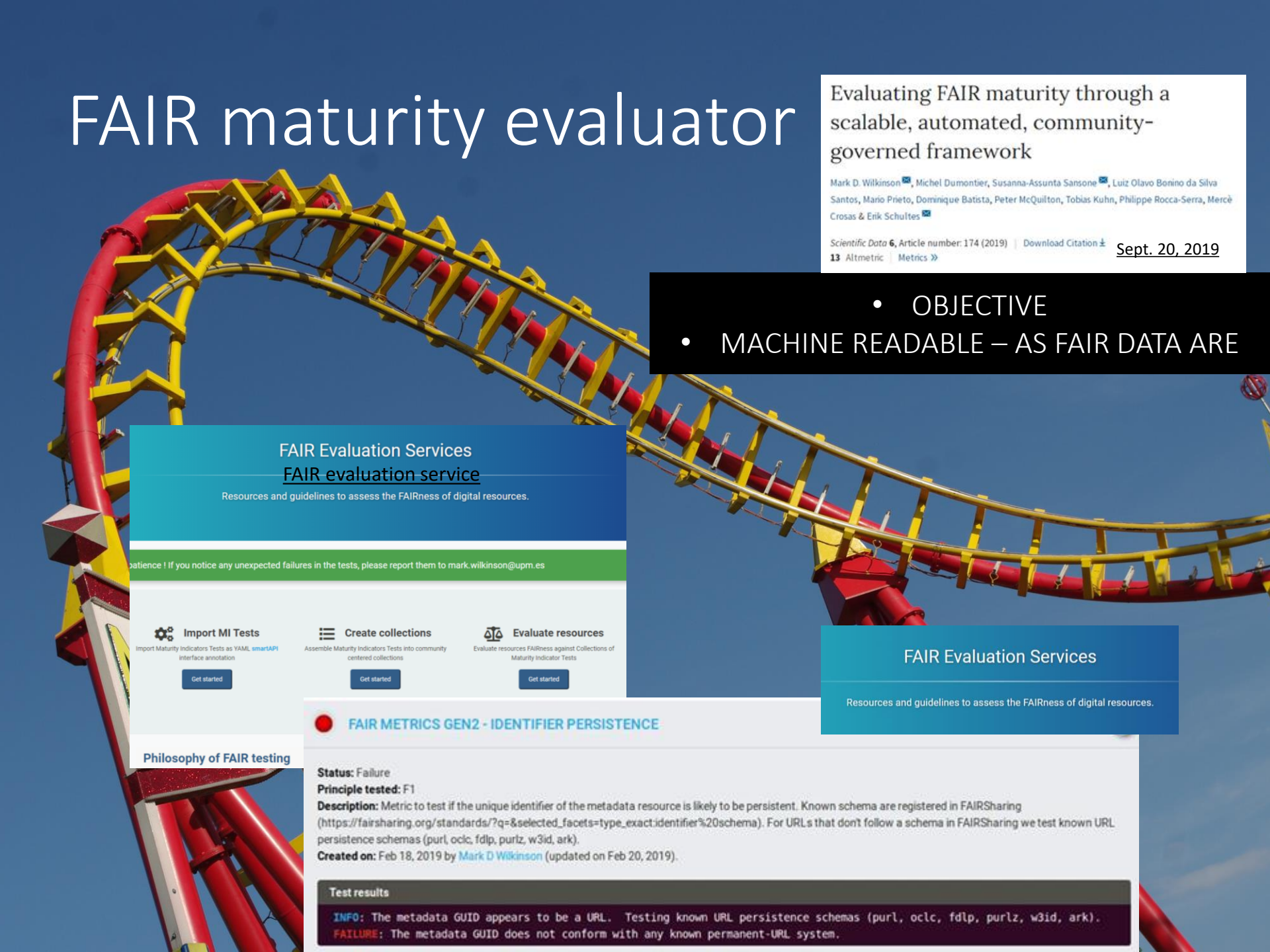
2. Are you aware that when you deposit a d
repository, you will need to provide some de
as discovery metadata) in order to make th
understandable and reusable to others?

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 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 & Erik Schultes

Scientific Data **6**, Article number: 174 (2019) | Download Citation
13 Altmetric | Metrics »

Sept. 20, 2019

- OBJECTIVE
- MACHINE READABLE – AS FAIR DATA ARE

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 smartAPI interface annotation
Get started

Create collections
Assemble Maturity Indicators Tests into community centered collections
Get started

Evaluate resources
Evaluate resources FAIRness against Collections of Maturity Indicator Tests
Get started

Philosophy of FAIR testing

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

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

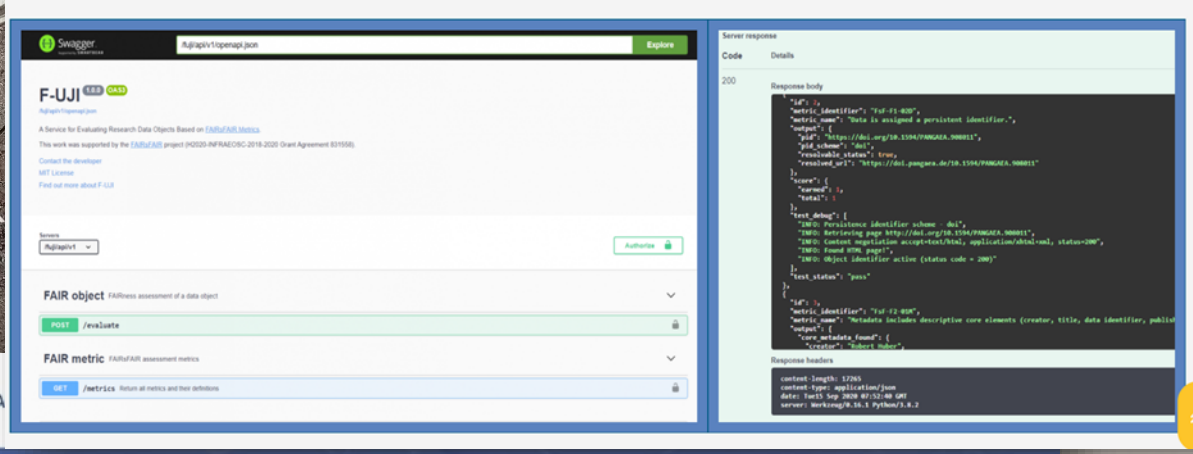
INFO: The metadata GUID appears to be a URL. Testing known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).
FAILURE: The metadata GUID does not conform with any known permanent-URL system.

F-UJI

FAIRSFAIR

About FAIR Support FAIR Landscape Tools & Software Competence Centre Events Project Outputs Outreach

Screenshots of the tool below



FAIRSFAIR F-UJI About FA

F-UJI Automated FAIR Data Assessment Tool

Home / F-UJI Automated FAIR Data Assessment Tool

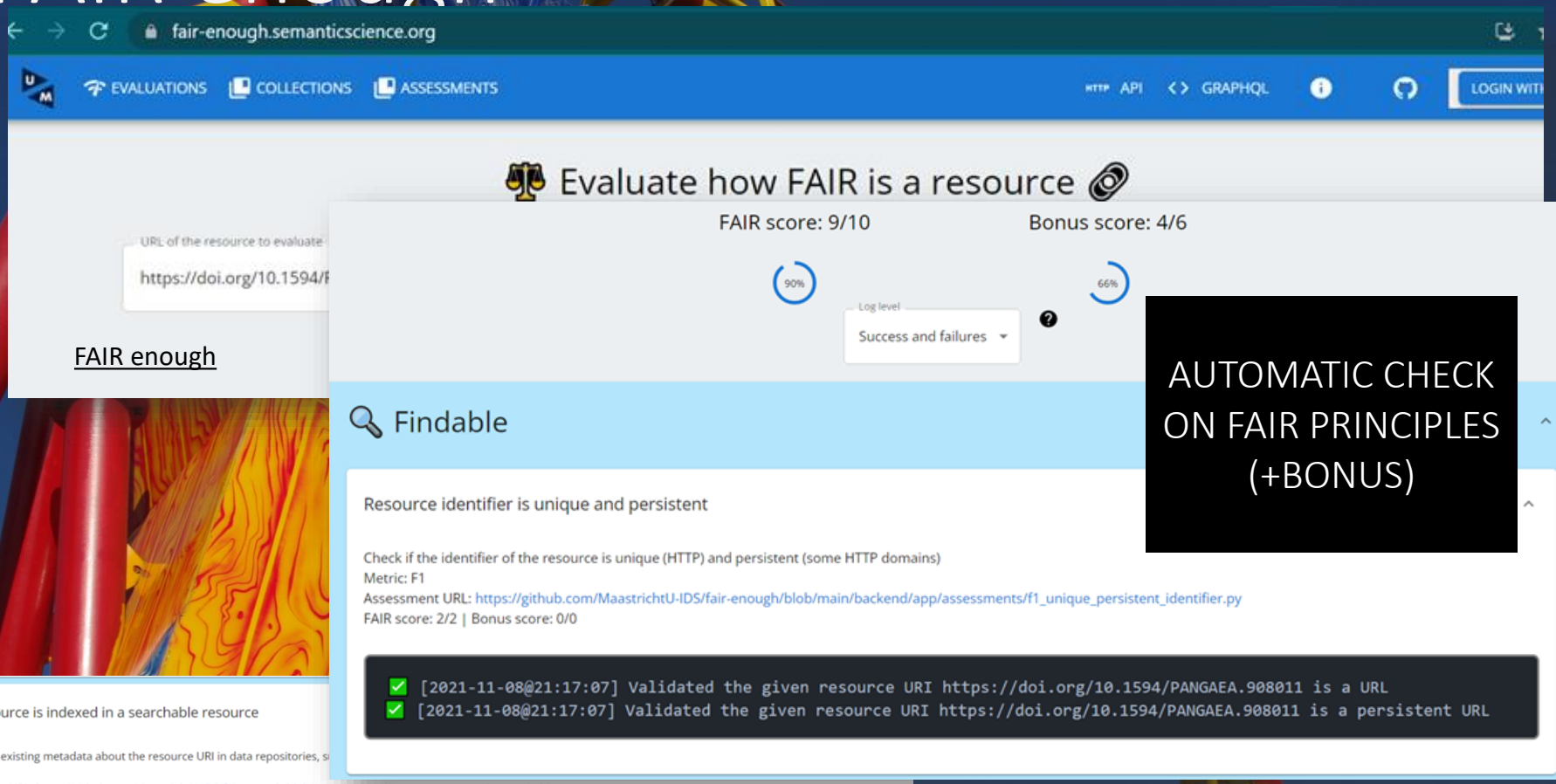
FAIRNESS EVALUATION (IN BETA)

FAIRsFAIR has developed F-UJI, a service based on REST, and is piloting a programmatic assessment of the FAIRness of research datasets in five trustworthy data repositories.

F-UJI Automated FAIR Data Assessment Tool

The F-UJI assessment is based on **16 out of 17 core FAIR object assessment metrics** developed within FAIRsFAIR and each corresponding to a part or the whole of a FAIR principle. F-UJI adheres to existing web standards and PID resolution services best practices and utilises external registries and resources such as re3data[1] and Datacite[2] APIs, SPDX License List[3], RDA Metadata Standards Catalog[4], and Linked Open Vocabularies (LOV)[5] For information on the practical tests implemented against the metrics, see Devaraju, Huber, et al., 2020.

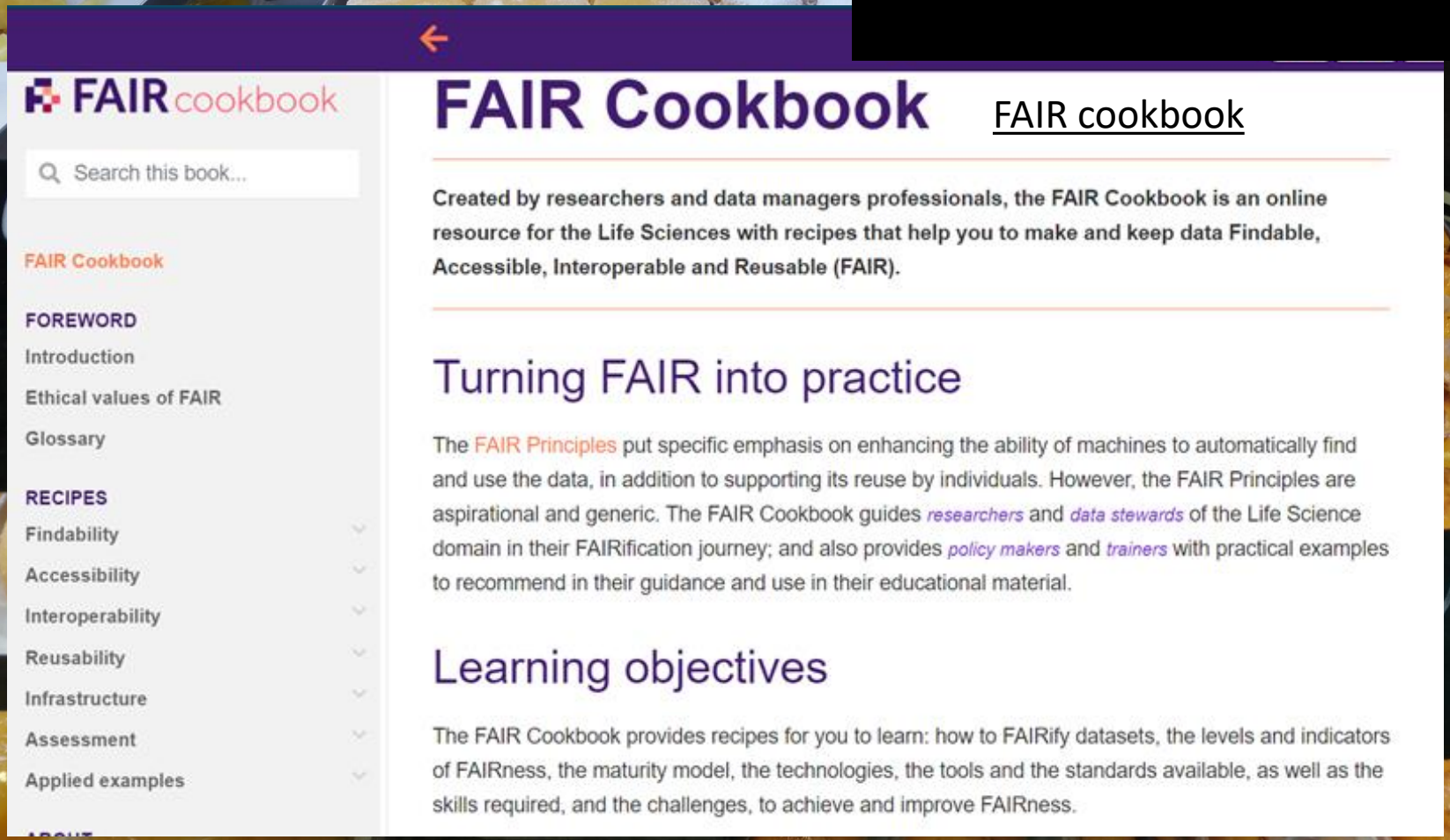
FAIR enough



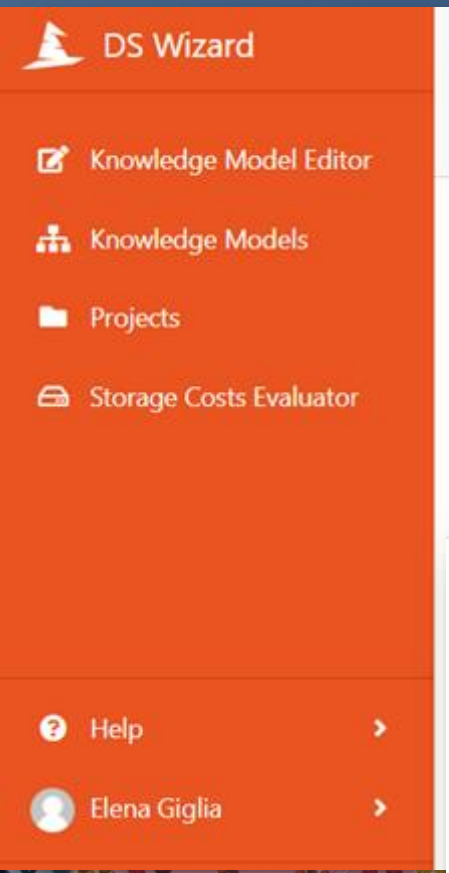
AUTOMATIC CHECK ON FAIR PRINCIPLES (+BONUS)

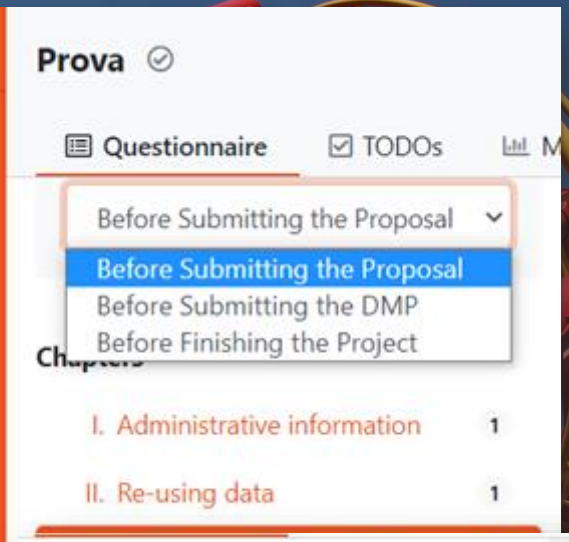
FAIR cookbook

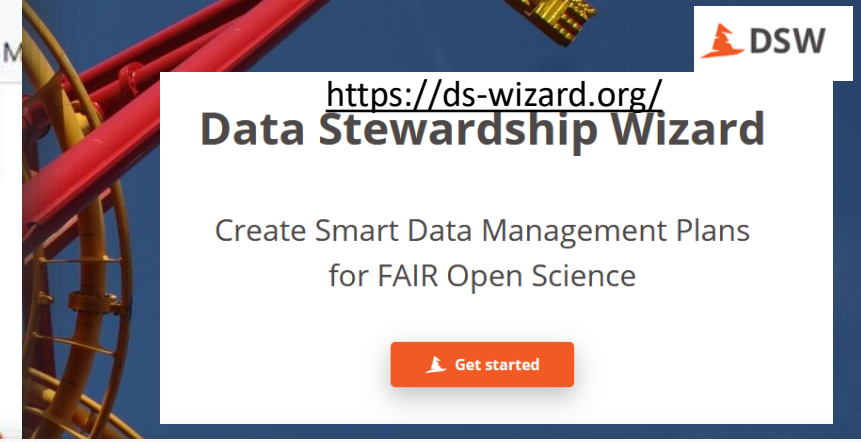
RECIPES TO MAKE YOUR DATA FAIR

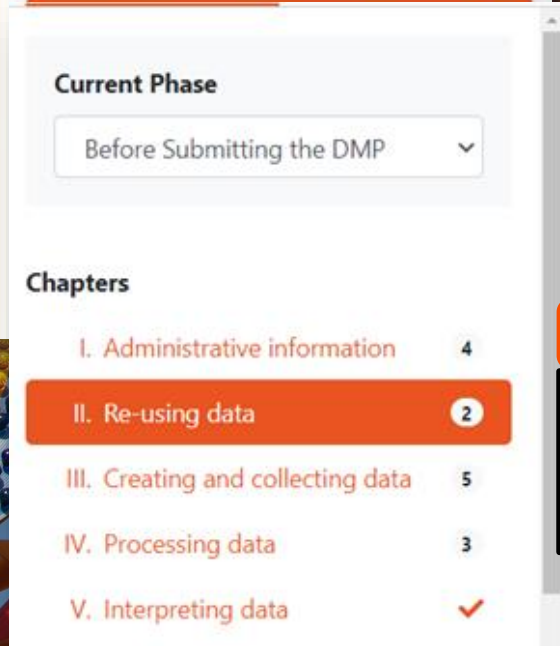


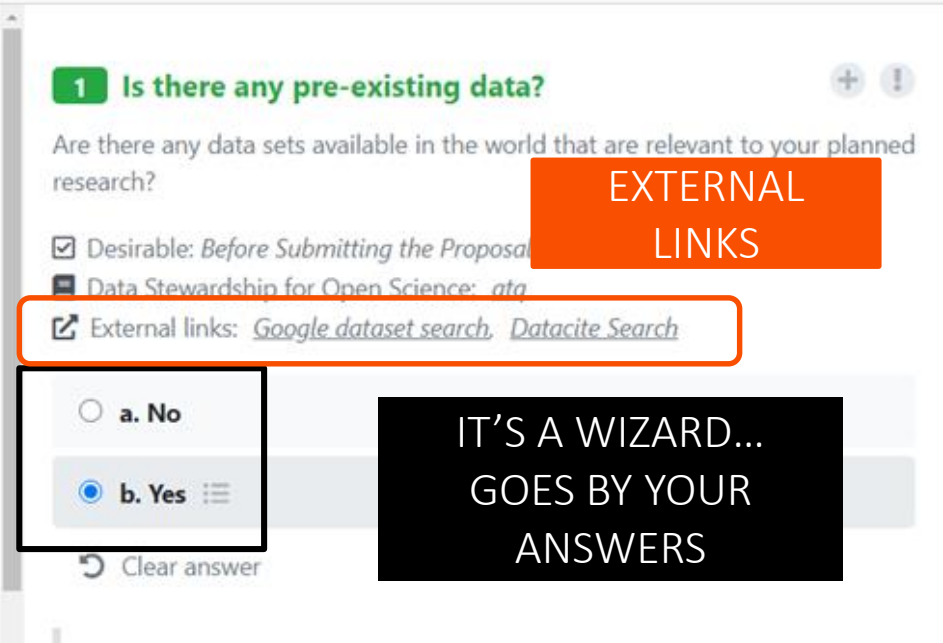
FAIR Wizard





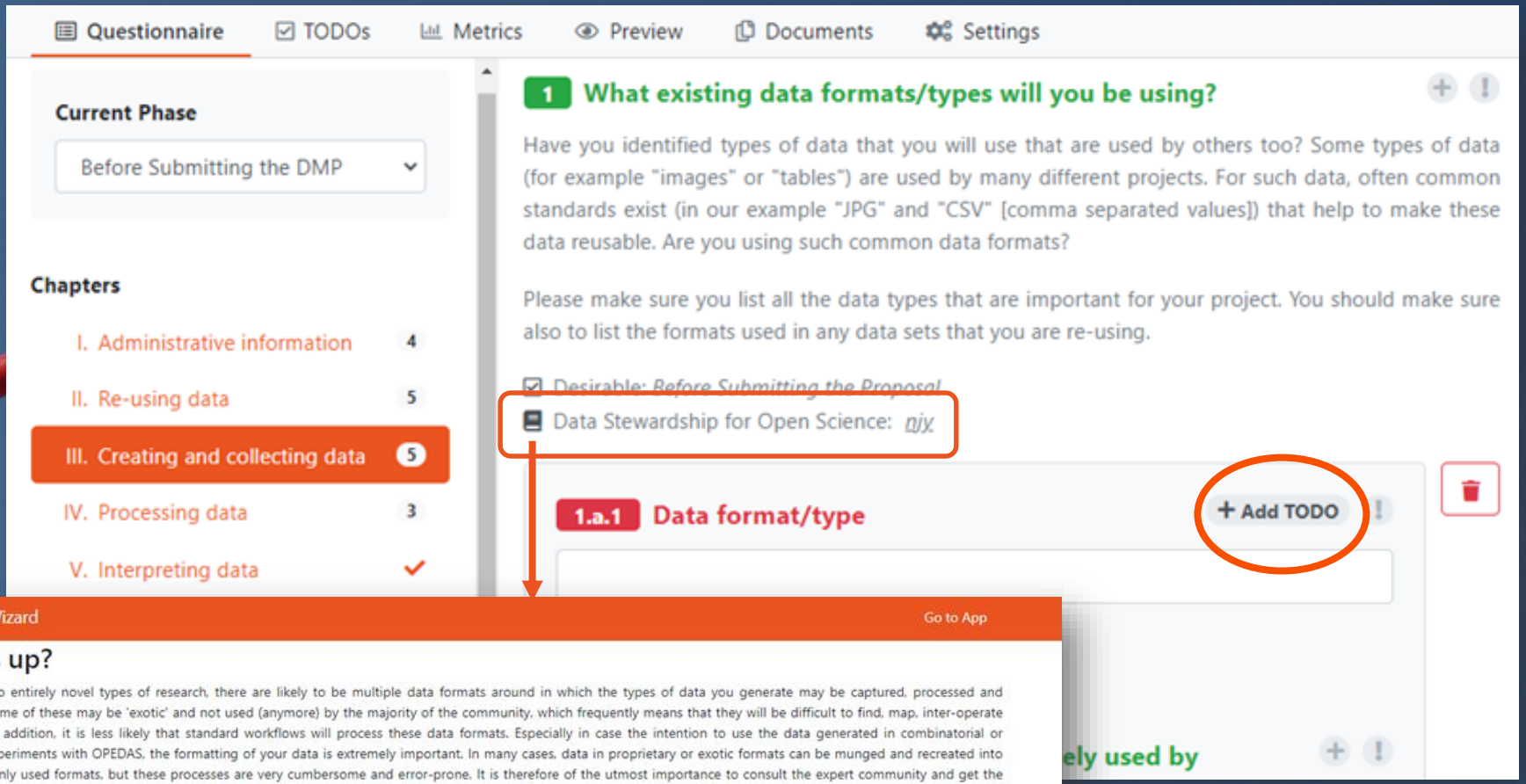


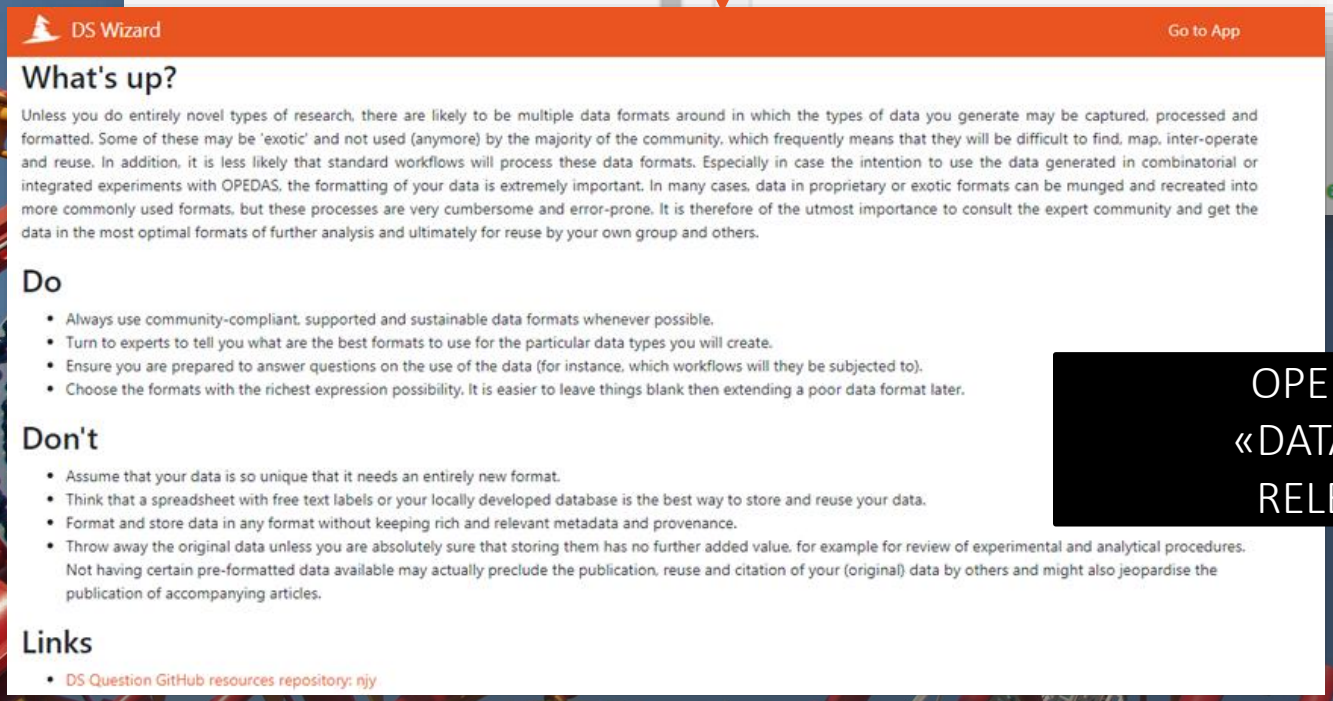




EXTERNAL LINKS

IT'S A WIZARD… GOES BY YOUR ANSWERS





OPENS PROF. MONS'
«DATA STEWARDSHIP»
RELEVANT CHAPTER

FAIR Wizard

IN THE END, THE SYSTEM EXTRACTS THE RELEVANT INFORMATION FROM YOUR ANSWER AND FILLS THE DMP IN

1. Data Summary

Re-used datasets

We will use the following reference datasets:

- **Leiden Bookseller database** (https://doi.org/10.5281/zenodo.5534110)

  We will use version "1.0" of this dataset. If a new version becomes available during the project, we will stay with the old version.

We will use the following already existing non-reference datasets:

- **VIAF**

  We will use its online version without downloading it.

Data formats and types

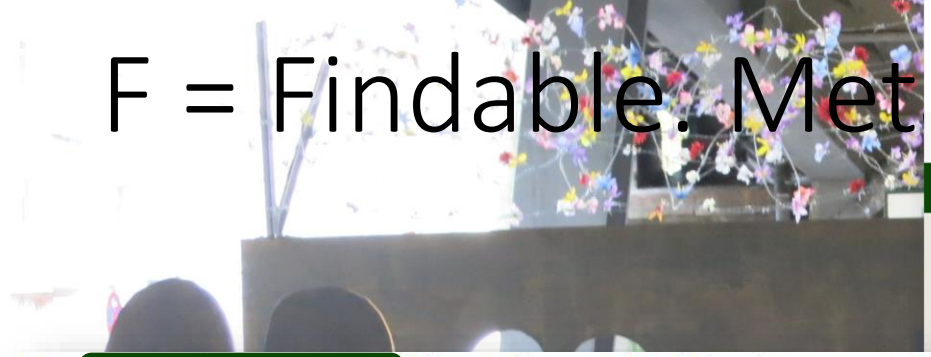
We will be using the following data formats and types:

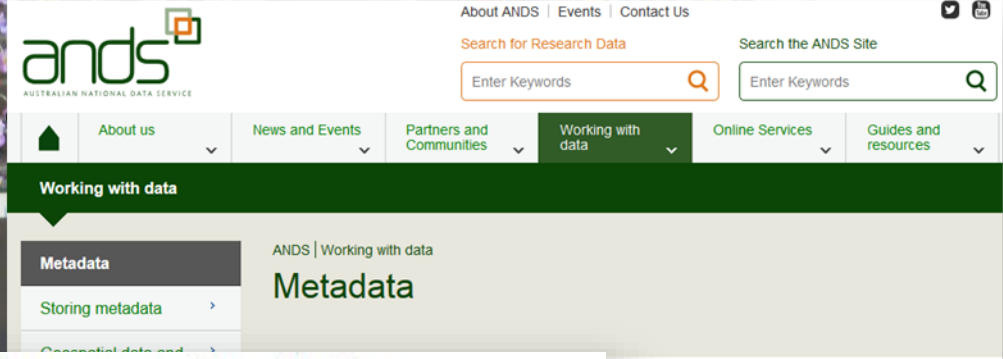
- **Comma-separated Values**

  It is a standardized format. This is a suitable format for long-term archiving.

FINDABLE

F = Findable. Met

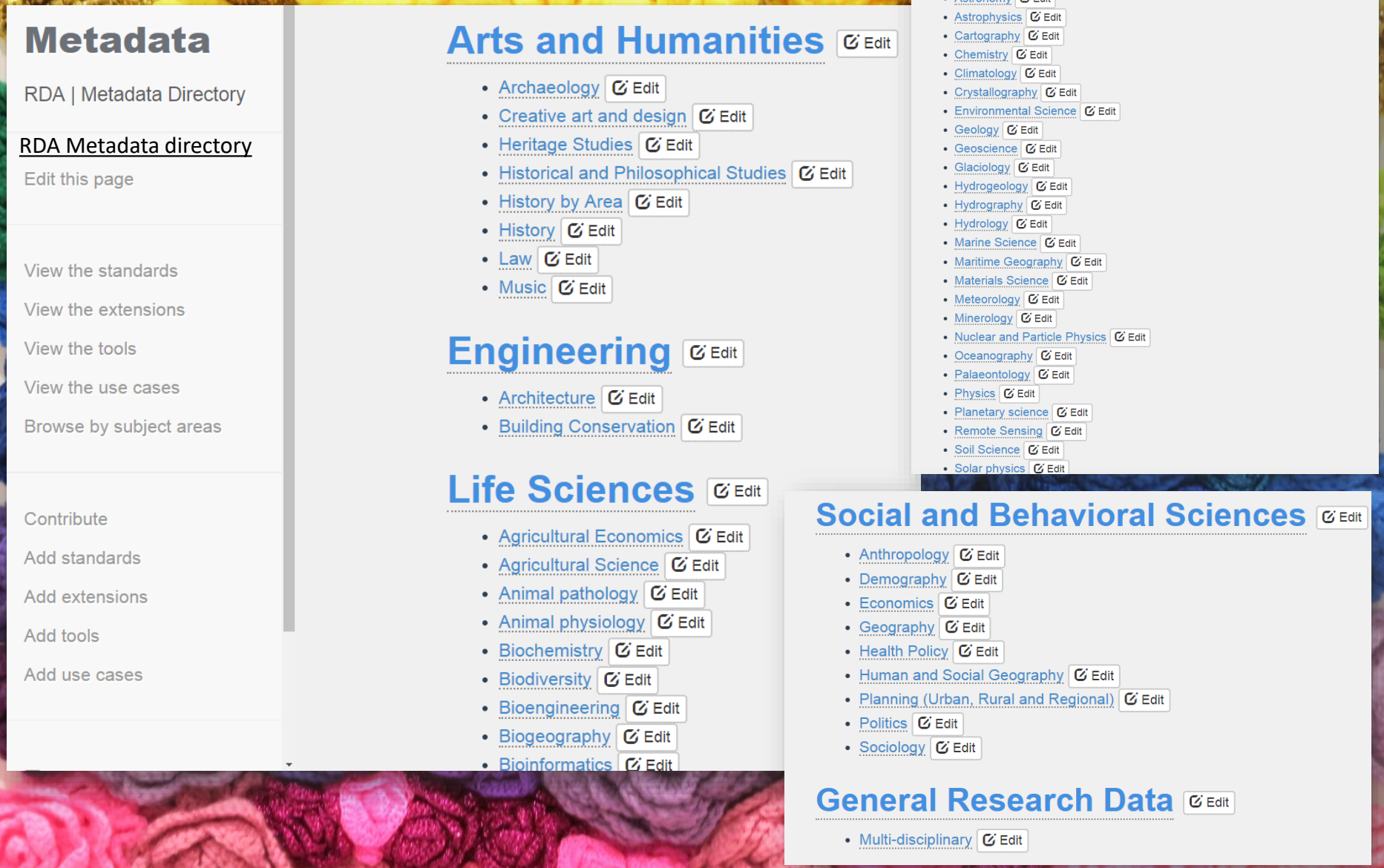




- **Descriptive metadata**: information required for discovery and assessment of the collection,
  - e.g. title, contributors, subject or keywords, study description, and location and dates of the study.
- **Provenance metadata**: this relates to the origins and processing of the data, and enables interpretation 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.

Metadata

F = Findable. Metadata standards



F = findable. Metadata tools

What CEDAR does

https://metadatacenter.org/

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

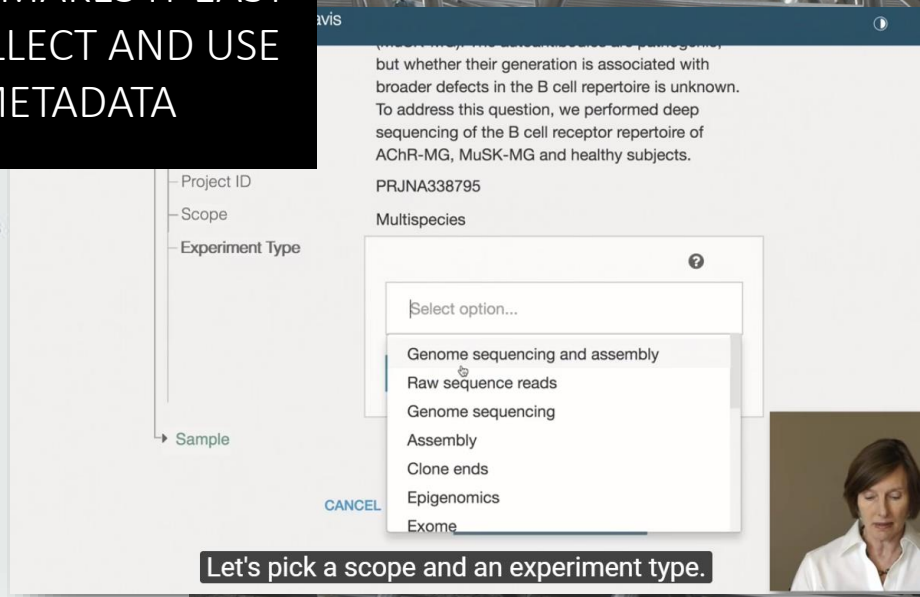
What can CEDAR do for me already?

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

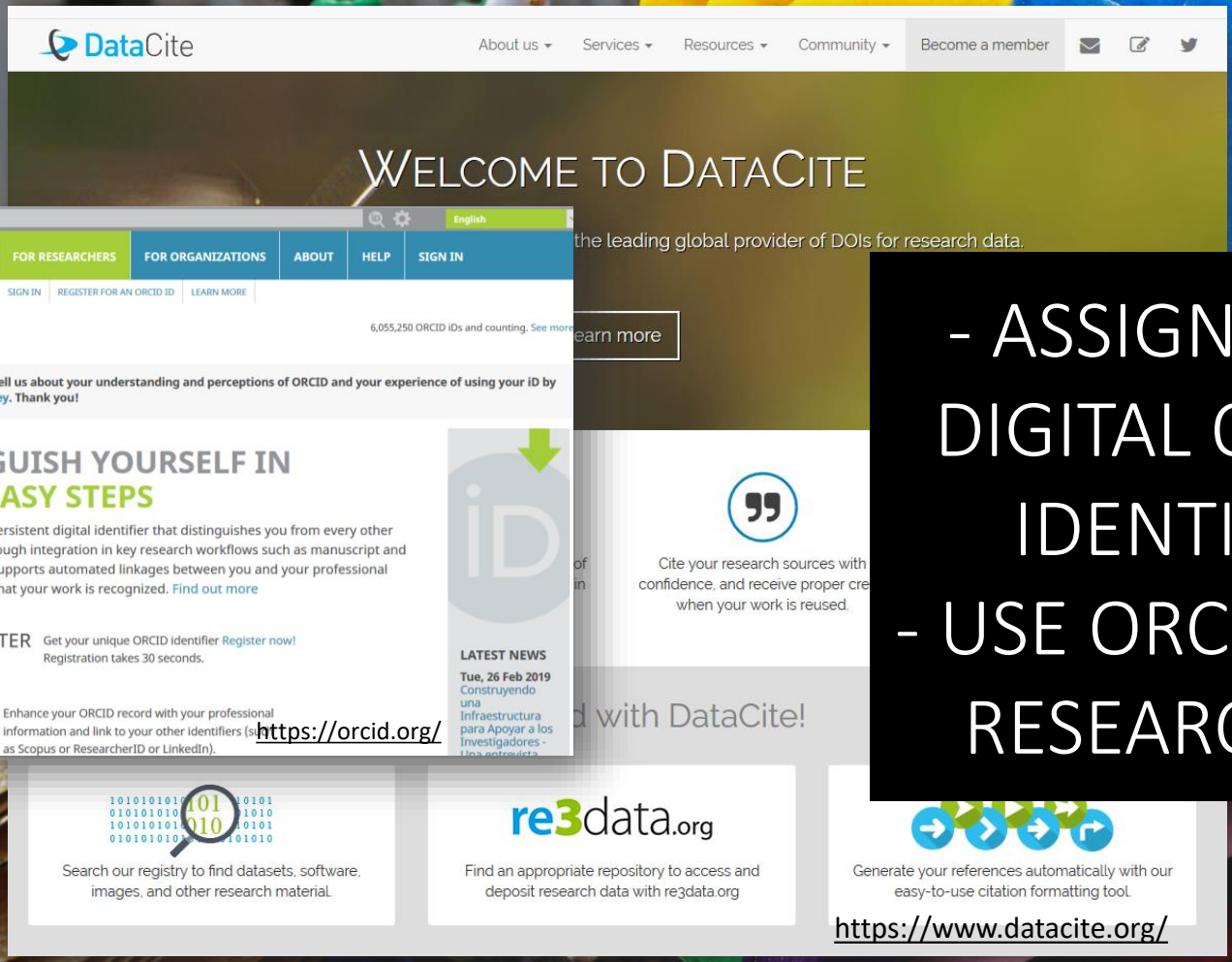
- create user-friendly, shareable forms for collecting metadata, with features like
  - nested and repeatable elements and fields
  - reusable elements
  - control over tool tips, field titles, and field descriptions
- share your forms and metadata
  - provide a link to your metadata editors, so they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

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

CEDAR MAKES IT EASY TO COLLECT AND USE METADATA



F = Findable. Persistent identifiers

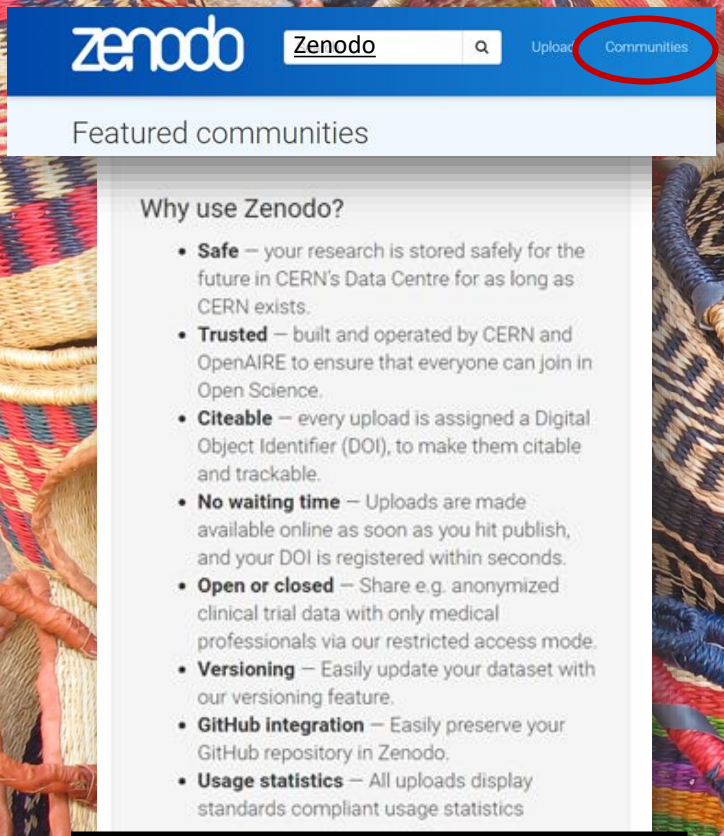


https://orcid.org/

- ASSIGN A DOI DIGITAL OBJECT IDENTIFIER
- USE ORCID-ID AS RESEARCHERS

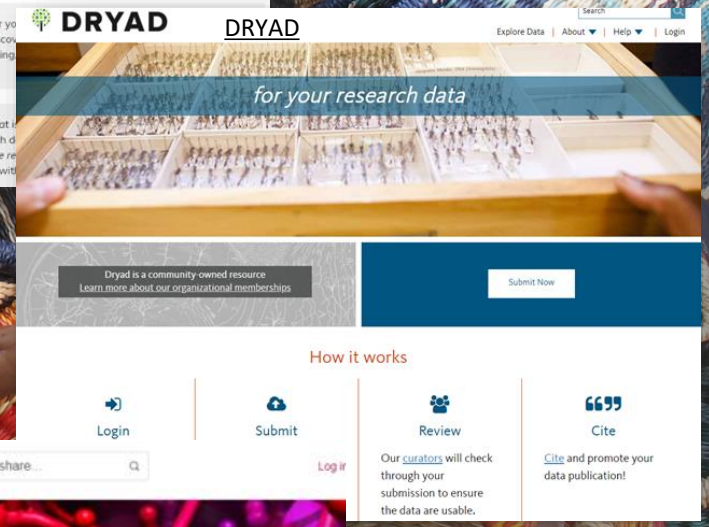
ACCESSIBLE

A = Accessible – Data repositories

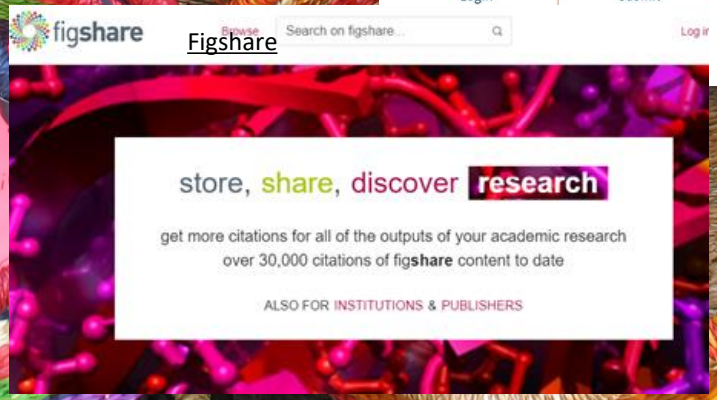


Zenodo

Dataverse



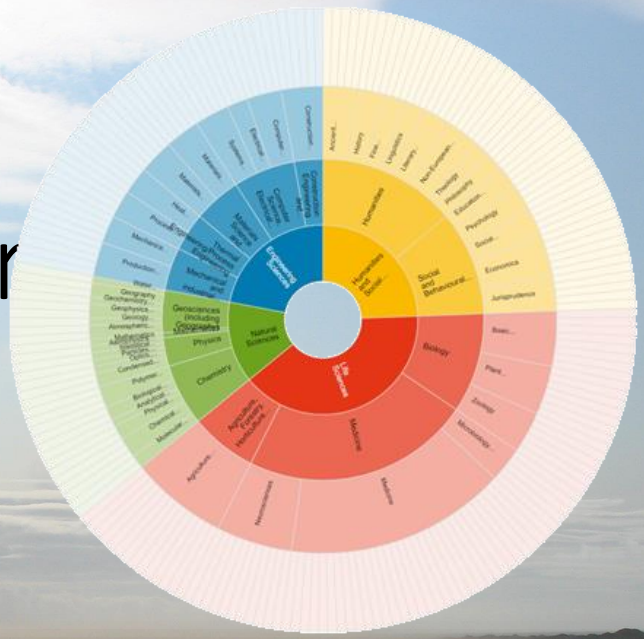
DRYAD



Figshare

YOU CAN CREATE A «COMMUNITY» [THE PROJECT?]

A = Accessible.
Looking for a repositor



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
Search... 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 Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to created it, did not exist. And the answer to what this illustrates – this is simply a distorted...

Read more

https://www.re3data.org/

A = Accessible. Data journals

Data journals list

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		
Journal of Open Psychology Data	http://openpsychologydata.metajnl.com/		http		

UCL Home » / Open@UCL Blog » / Data journals and data reports – don't miss ou

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

Aug. 2021

By Kirsty, on 17 August 2021

Guest post by James Houghton – Research Data Support Officer

Why not publish a data report article?

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

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.
- *unts* – three files providing the data, metadata and ld type definitions (.csv, .txt, .csvt respectively) for pot- herds counteddurin stage-one fieldwalking.
- *ttery* – three files providing the data, metadata and field pe definitions (.csv, .txt, .csvt respectively) for the main ttery database, assembled various artefact specialists.
- *trography* – three files providing the data, metadata d field type definitions (.csv, .txt, .csvt respectively) for ose sherds sampled for thin section petrography.
- *hics* – three files providing the data, metadata and field pe definitions (.csv, .txt, .csvt respectively) for the main hics database.
- *her* – three files providing the data, metadata and field pe definitions (.csv, .txt, .csvt respectively) for the main tabase of all non-ceramic and non-lithic finds.
- *ructs* – three files providing the data, metadata and field pe definitions (.csv, .txt, .csvt respectively) for the main tabase of all standing remains, except for terraces.
- *past* – a vector polygon dataset (.shp and associated les) with the shape of Antkythera's coastline.
- *ology* –a vector polygon dataset (.shp and associated les) with the main bedrock units on Antkythera.
- *tracts* – a vector polygon dataset (.shp and associated files) with the main stage-one survey units.
- *grids* – a vector polygon dataset (.shp and associated files) with the main stage-two survey units.
- *terraces* – vector line dataset (.shp and associated files) with all observable agricultural terraces (i.e. the location

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

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

Publication Date
05/02/2012

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

License
Creative Commons CC-BY 3.0

Reuse Potential
Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. http://dx.doi.org/10.5284/1000271 http://dx.doi.org/10.5284/1000 dx.doi.org/10.528 to the fact that th cal. The ASP data locations, dates a ally in the databas structures and ter

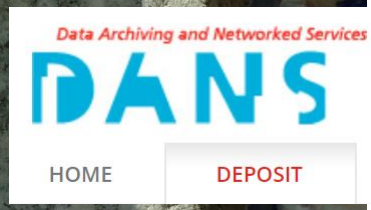
Data journals

Panayiota Polydoratou

Alexander Technological Educational Institute of Thessaloniki

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

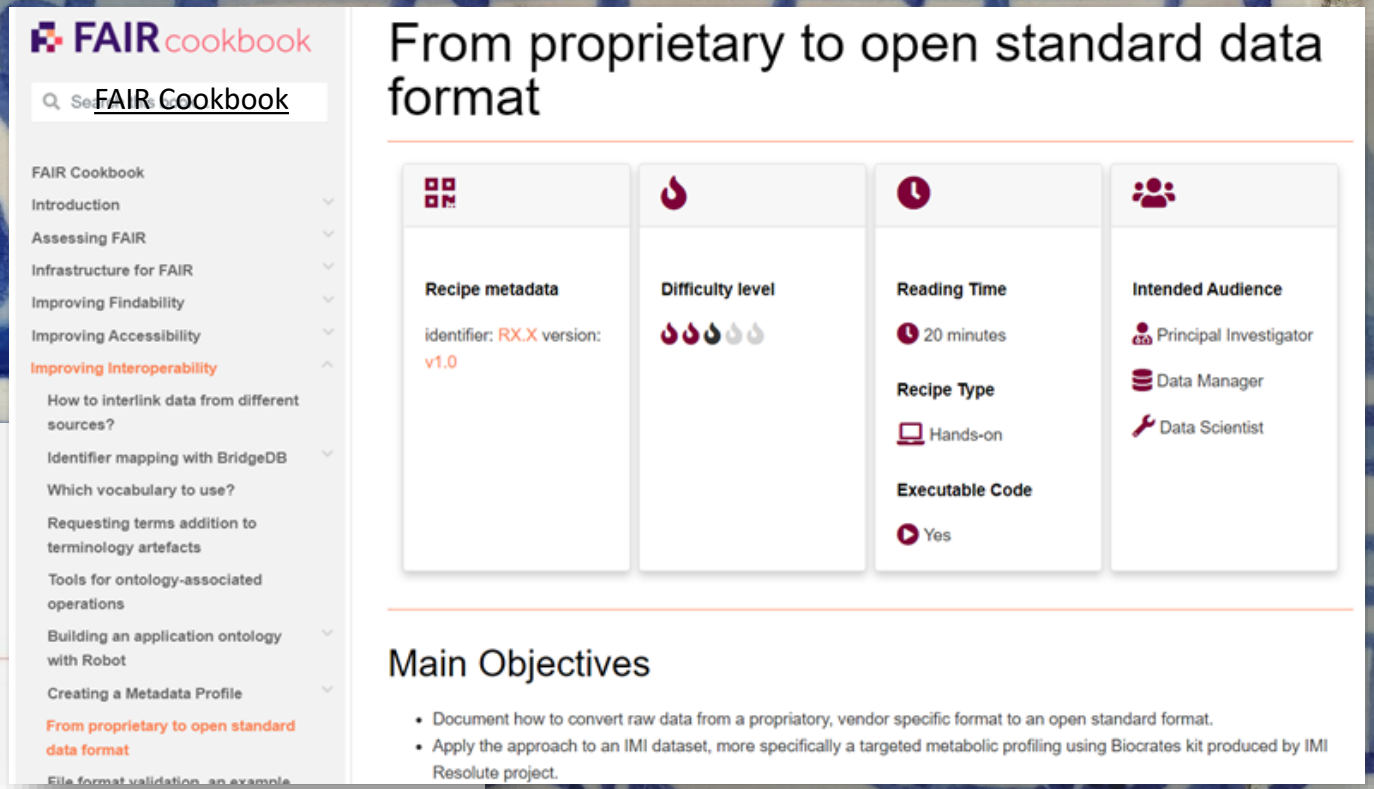
A = Accessible. Formats

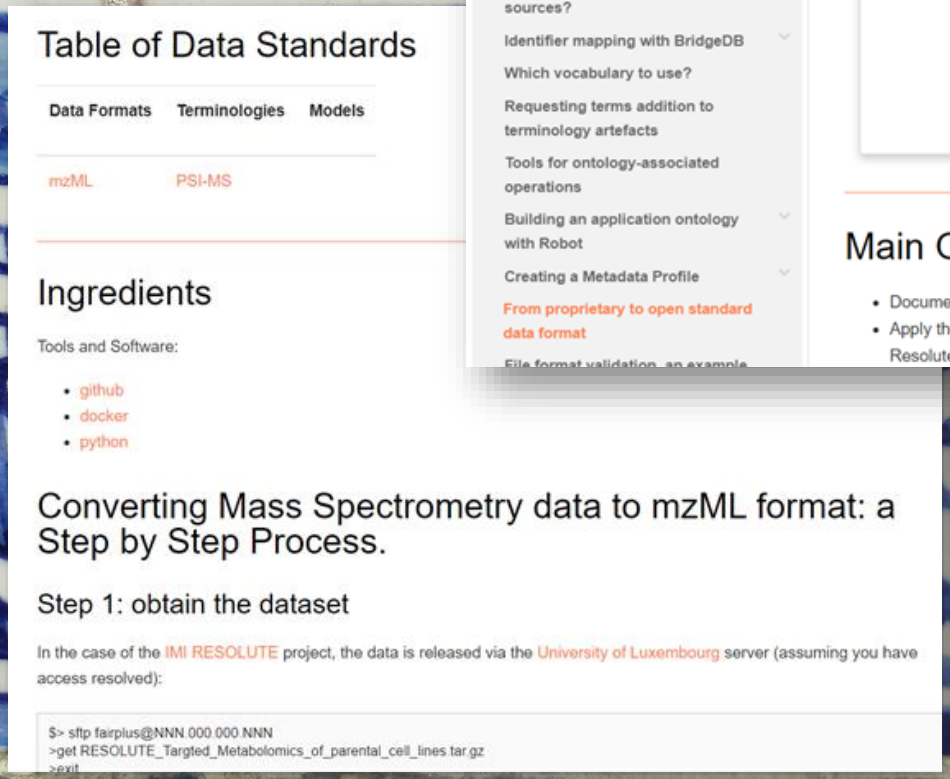


Type	Preferred format(s)	Non-preferred format(s)
Text documents	• PDF/A (.pdf)	• ODT (.odt) • MS Word (.doc, .docx) • RTF (.rtf) • PDF (.pdf)
Plain text	• Unicode text (.txt)	• Non-Unicode text (.txt)
Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml)
Spreadsheets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

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

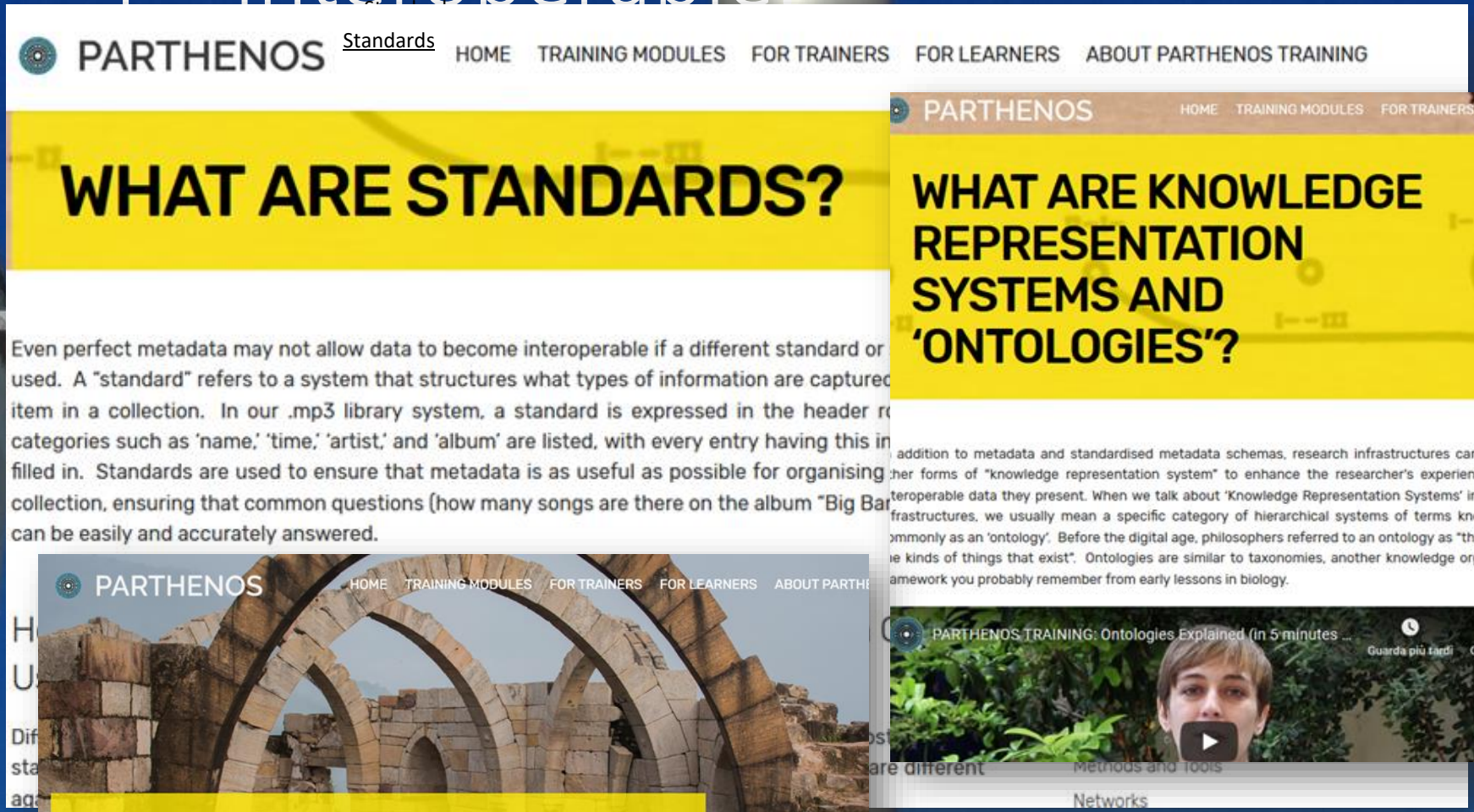
A = Accesible - Formats



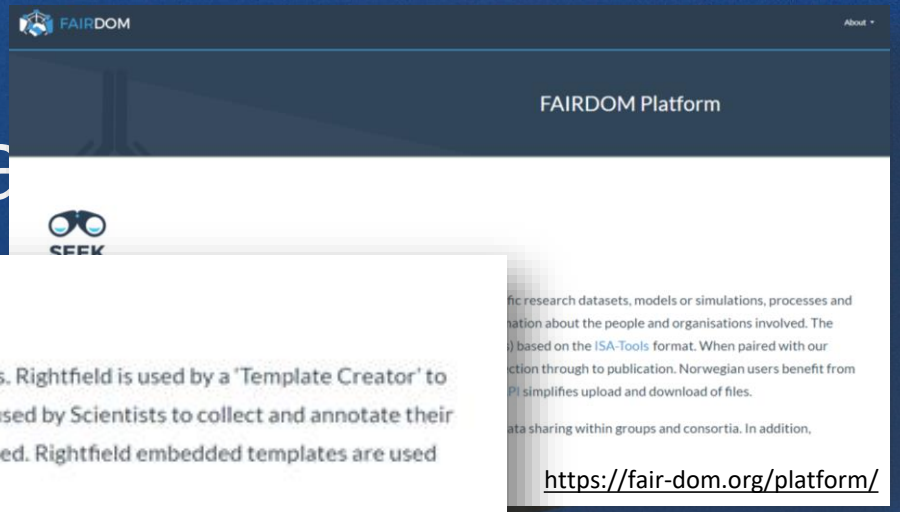


INTEROPERABLE

I = Interoperable



I= Interoperable – a



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

RightField

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

MORE INFORMATION

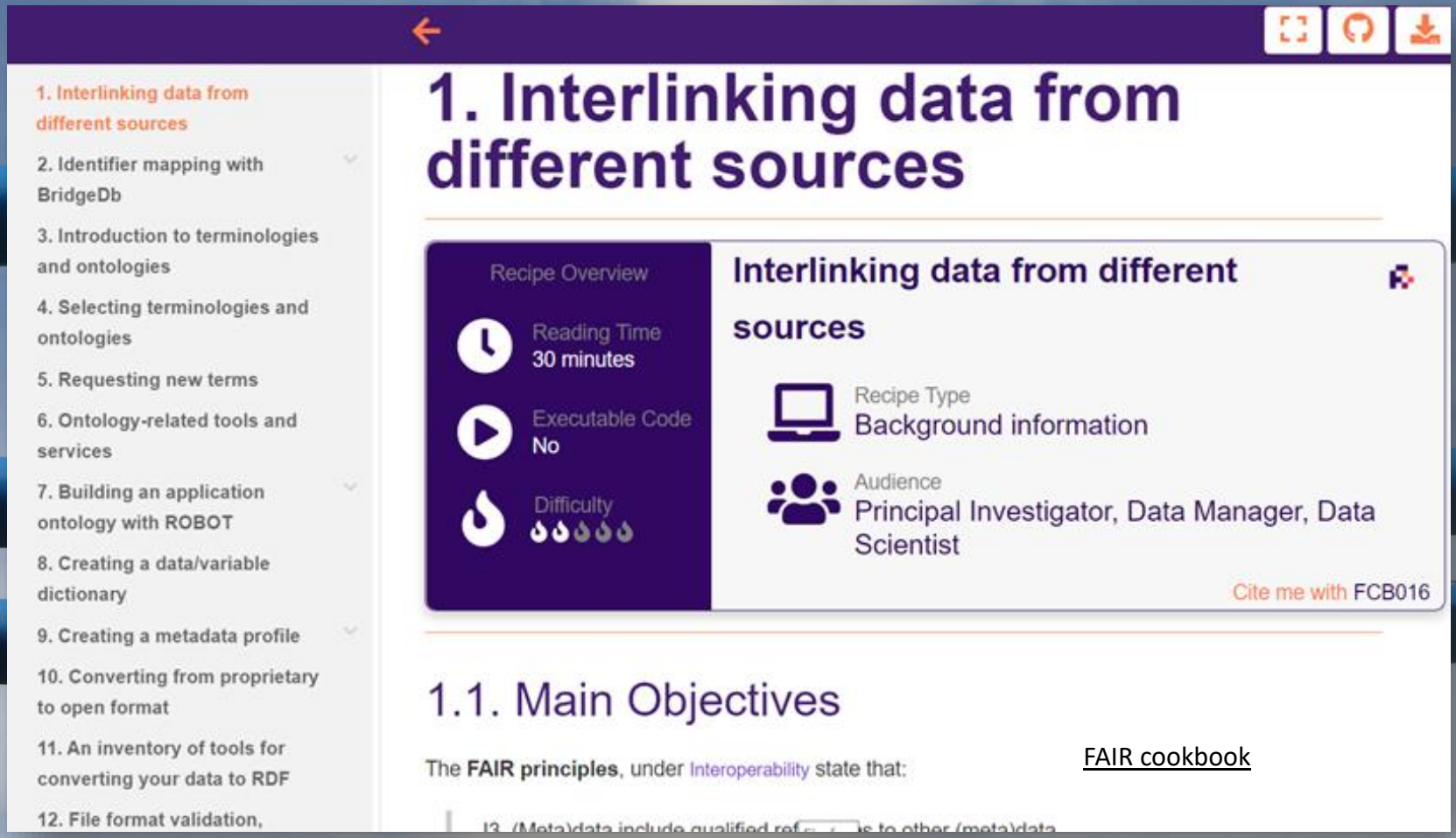
Rightfield

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

amplification
by-product_formation
catabolic_response
cell_growth_optimisation
cell_size
Comparative_genomic_hybridization
comparative_genomics
continuousEnzymatic

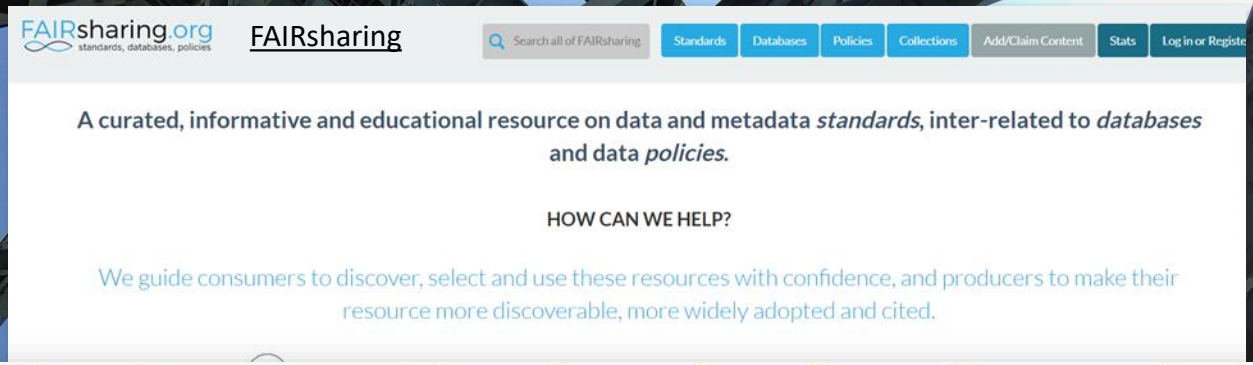
Metadata | Organism_Sample | Instrument_(optional) | Data_ | Result_

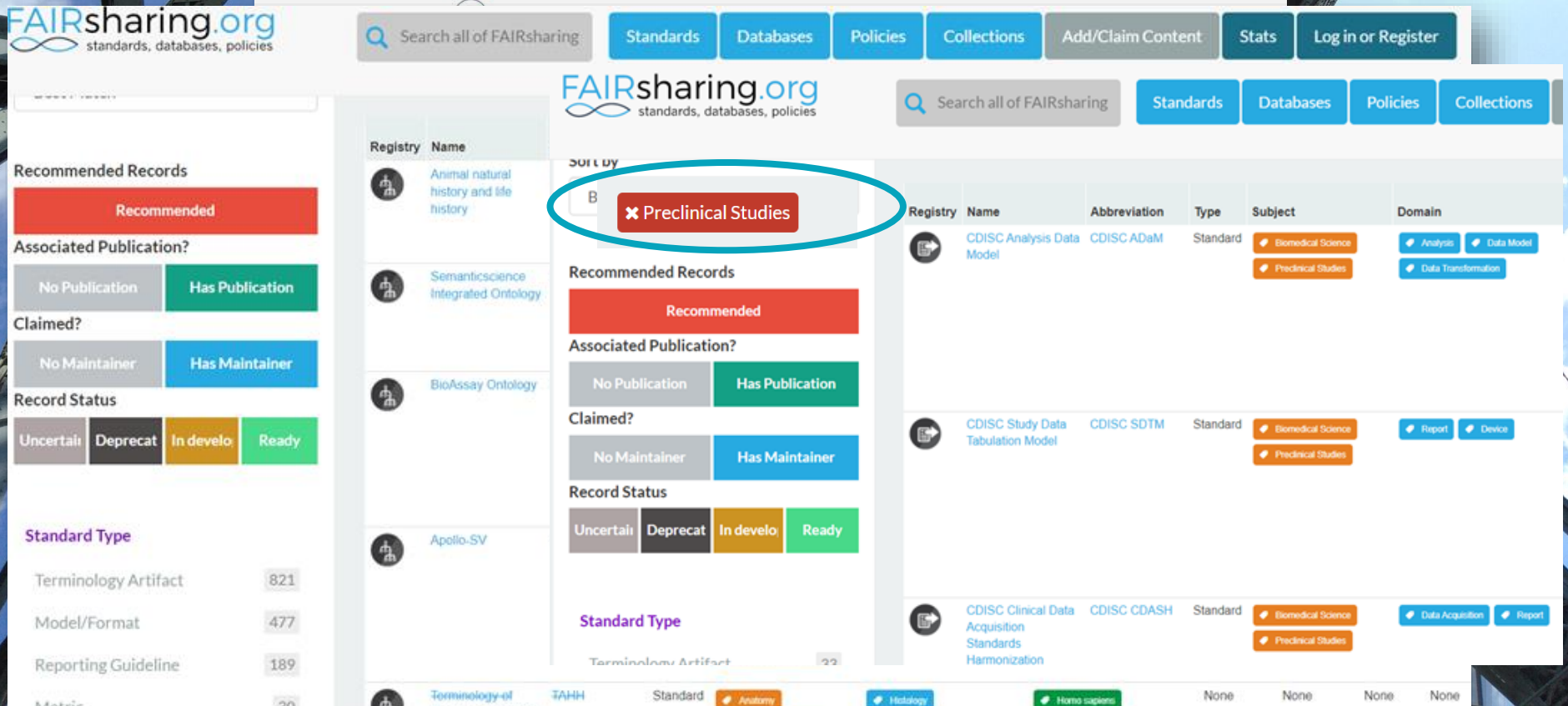
I = Interoperable



FAIR cookbook

[FAIRsharing. To be interoperable]





REUSABLE

R = Reusable. Documentation

DOCUMENTATION (README FILE) TO
- AVOID MISUSE/MISINTERPRETATION
- KEEP INTEGRITY

CESSDA guide

cessda TRAINING

Project-level documentation

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

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

Data-level documentation

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

⊖ Quantitative data

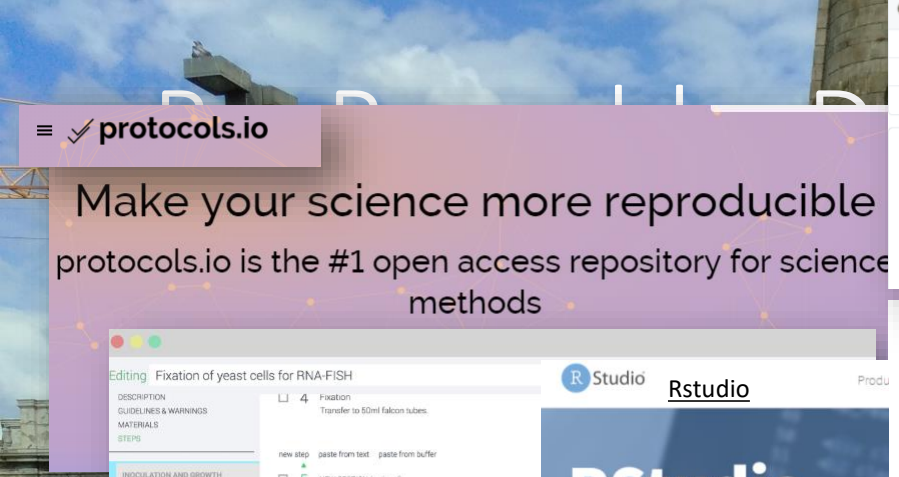
Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.

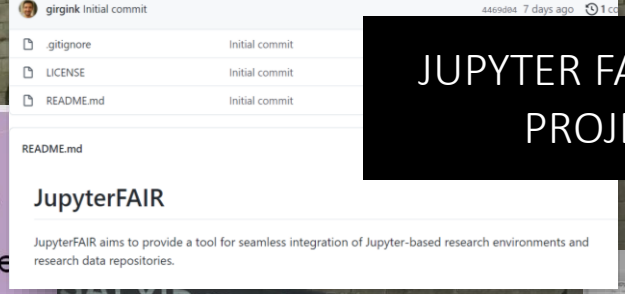
Data-level documentation for quantitative data

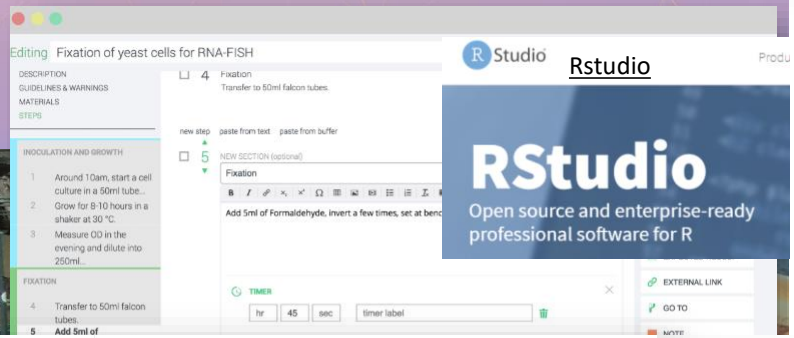
For quantitative data document the following:

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

JUPYTER FAIR, NEW PROJECT





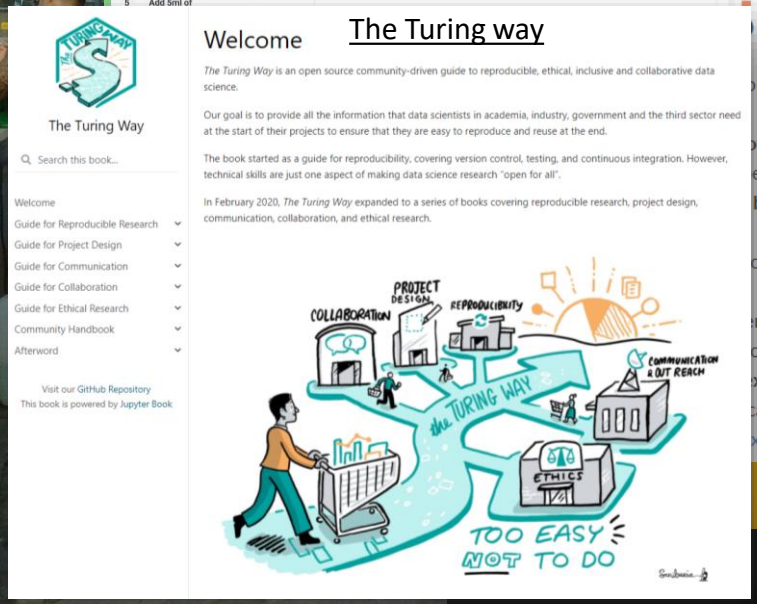


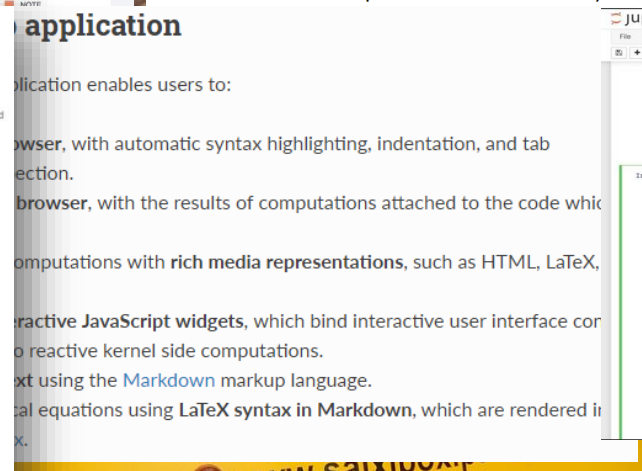
Jupyter

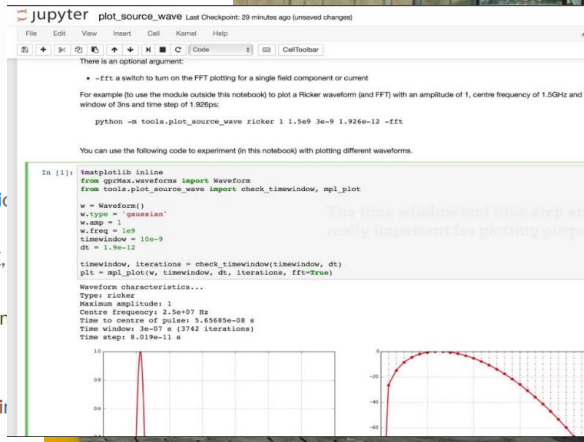
What is an Open Notebook?

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.







...WHY NOT?

- PROTOCOLS.IO TO DEPOSIT YOUR METHODS
- OPEN LAB NOTEBOOK TO TRACK ANYTHING YOU DO

[TIME CONSUMING THE FIRST TIME, THEN...]

R= Reusable. Licenses

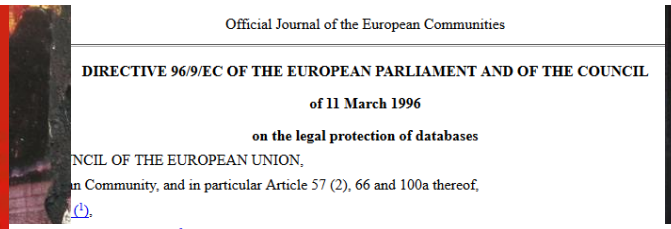
Copyright: protects the STRUCTURE, selection or arrangement of their contents" (Art. 3) NOT THE DATA

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

KEEP CALM AND

REMEMBER:
RAW DATA ARE NOT PROTECTED BY COPYRIGHT

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


Official Journal of the European Communities

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

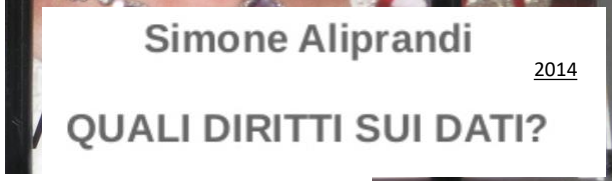
of 11 March 1996

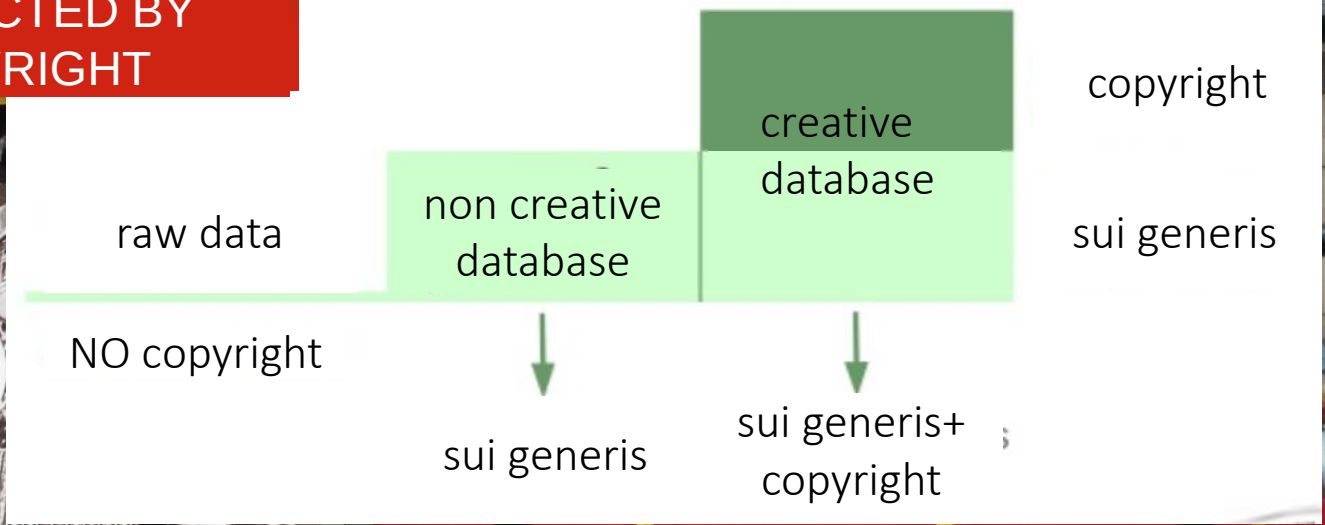
on the legal protection of databases

NCIL OF THE EUROPEAN UNION,

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

(1),





R = Reusable – Legal aspects

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

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

MIT Press Direct 2020

Data Intelligence

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

Licensing FAIR Data for Reuse

Ignasi Labastida, Thomas Margoni

Author and Article Information

Data Intelligence (2020) 2 (1-2): 199–207.
https://doi.org/10.1162/dint_a_00042

Cite PDF Permissions Share

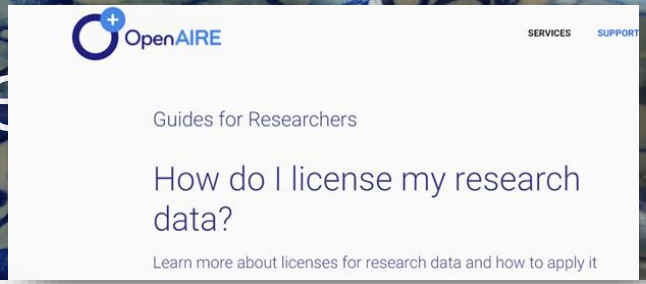
Previous Article Next Article

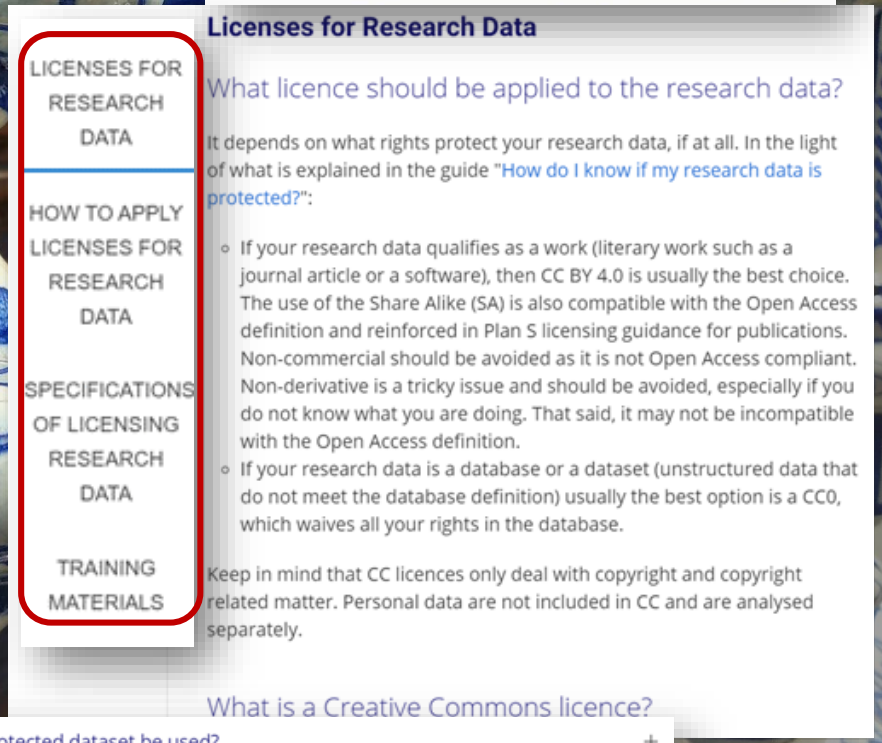
Article Contents

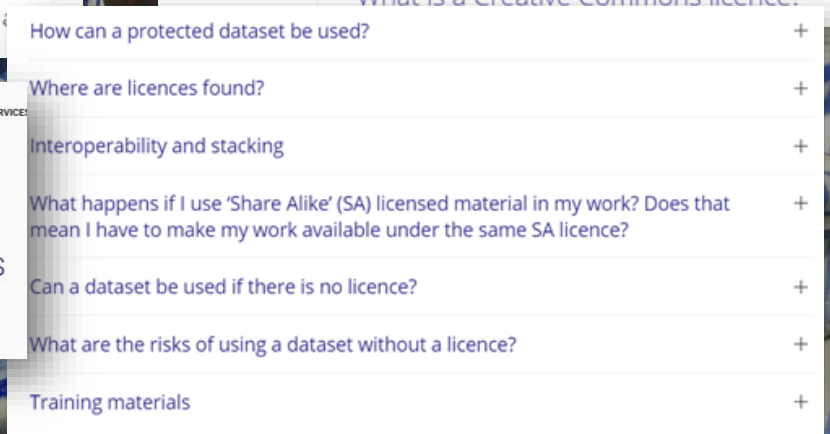
Abstract

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

R = Reusable – Legal aspe







Creative Co

CC Factsheet

creative commons UK

FACT SHEET ON
CREATIVE COMMONS & OPEN SCIENCE V.0.1

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

CC0: FROM A LEGAL POINT OF VIEW, THE ONE AND ONLY LICENSE

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

nmons and O

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

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

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

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

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

USE A CC0

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

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

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

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

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

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

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

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

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

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

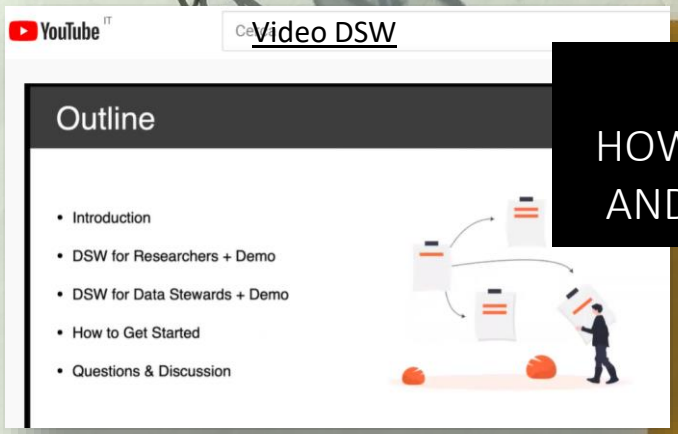
We recommend you avoid using a 'No Derivatives' licence. Here's why:

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

…Data Management Plan

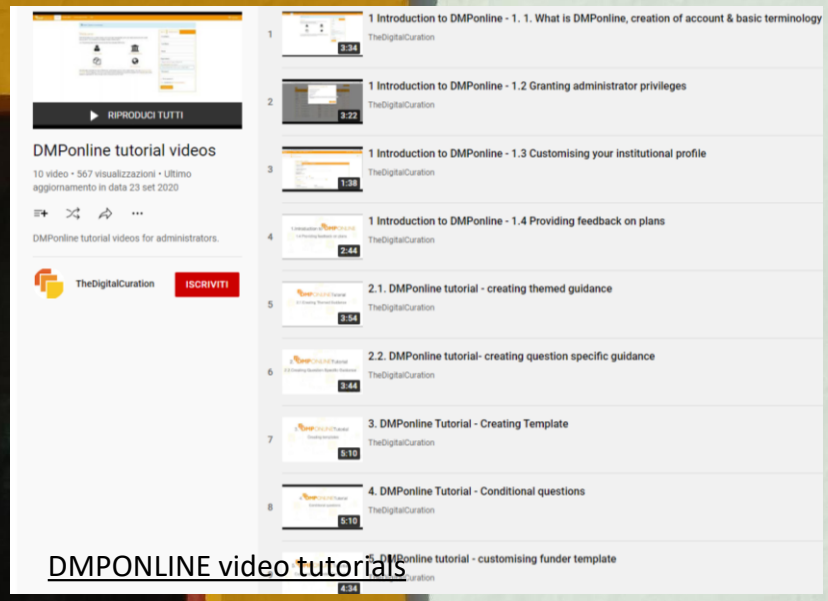
- A STRUCTURED WAY TO THINK TO YOUR RESEARCH FROM THE PERSPECTIVE OF YOUR DATA: collection, preservation, description, sharing
- CLEAR RULES FROM THE BEGINNNG
- LIVING DOCUMENT TO BE UPDATED
- …AND THEN… ACT ACCORDINGLY…

DMP online and Data Wizard video tutorials – to draft your DMP



Video DSW

VIDEO – HOW DMPONLINE AND DSW WORK



DMPONLINE video tutorials

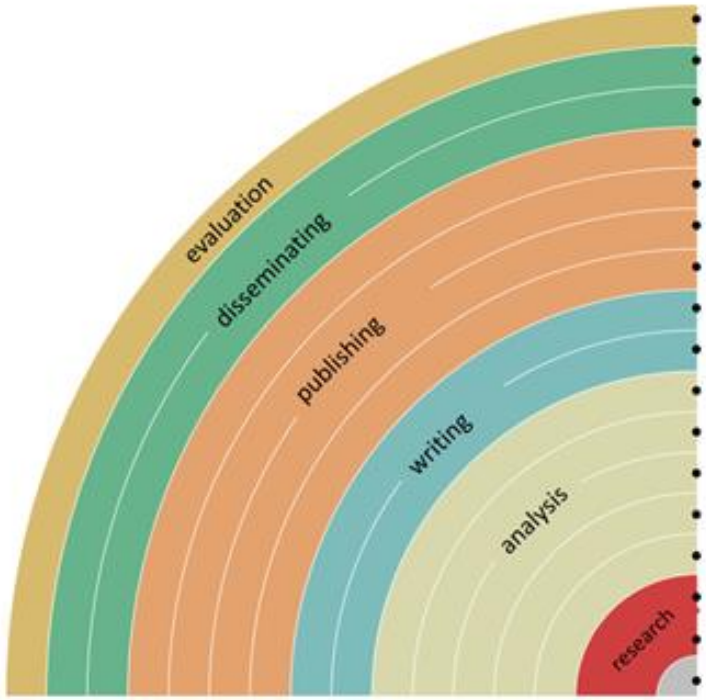
Tips&tricks

DMP tips

… «as open as possible»…

Open Science: why just your data?

YOU CAN MAKE YOUR WORKFLOW MORE OPEN BY...



- adding alternative evaluation, e.g. with. altmetrics
- communicating through social media, e.g Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. Creative Commons BY
- self archiving in archives or publishing on Open journals
- using open peer review, e.g. at PubPeer o F1000
- sharing preprints, e.g. at OSFpreprint, arXiv o biorXiv
- using actionable formats, e.g. with Jupyter o CoCalc
- open XML-drafting, e.g. at Overleaf o Authorea
- sharing protocols & workflows, e.g. at Protocols.io
- sharing notebooks, e.g. at OpenLabNotebook
- sharing code, e.g. at GitHub licensing GNU/MIT
- sharing data, e.g. at Dryad, Zenodo o Dataverse
- pre-registering, e.g. at OSFregistry o AsPredicted
- commenting openly, e.g. with Hypothes.is o Pund.it
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. with RIO Journal

2021

... thank you!