Principi FAIR
(con un pizzico di EOSC)

Elena Giglia
elena.giglia@unito.it
@egiglia

ispas
paths to successful innovations

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006544

Cosa vedremo in questo modulo

1. Cosa significa FAIR
2. Perché servono i dati FAIR BY DESIGN nella ricerca quotidiana, in EOSC e in Horizon Europe

MESSAGGI CHIAVE

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

[i tre passi fondamentali]

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

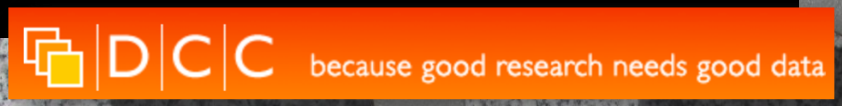
2. MA SE I DATI NON SONO «FAIR», APRIRLI COMPORTA RISCHI (USO SCORRETTO, CATTIVE INTERPETAZIONI, …)

3. MA SE I DATI NON SONO CORRETTAMENTE GESTITI, RENDERLI «FAIR» COSTA TROPPO TEMPO E DENARO. CON EOSC, DATI GESTITI E DATI FAIR TENDONO A COINCIDERE, FAIR BY DESIGN

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

1. Gestione dei dati / perché

1. SONO IL FONDAMENTO DI UNA RICERCA SOLIDA



2. IL COVID HA DIMOSTRATO CHE SERVONO I DATI, NON SOLO GLI ARTICOLI, E SERVONO SUBITO

3. I DATI SONO FRAGILI, SI PERDONO

4. ALCUNI SONO UNICI E NON POSSONO ESSERE RIPRODOTTI (METEO, TERREMOTI…)

5. POSSONO ESSERE MANIPOLATI, GESTIRLI GARANTISCE INTEGRITÀ

6. PERMETTONO VALIDAZIONI E RIPRODUCIBILITÀ

7. I DATI CREANO PONTI FRA LE DISCIPLINE

8. PERCHÉ POSSONO ESSERE RIUTILIZZATI (IN MODO INEDITO)

1. Gestione dei dati / come

DESCRIZIONE (metadati)

ORGANIZZAZIONE (file naming, folders, versioning…)

BACKUP E STORAGE

CONSERVAZIONE SUL LUNGO PERIODO

ASPETTI LEGALI

LUNGO TUTTO IL CICLO DI VITA

Serve formazione?

CESSDA expert guide

Data Management Expert Guide

1. Plan
2. Organise & Document
3. Process
4. Store
5. Protect
6. Archive & Publish
7. Discover

cessda TRAINING

Plan

In this introductory tour, you will become aware of what data management and a data management plan (DMP) are and why they are important. General concepts such as social science data and FAIR data will be explained. Based on our recommendations and good practice examples, you will be able to start writing your DMP.

Organise & Document

If you are looking for good practices in designing an appropriate data file structure, naming, documenting and organising your data files within suitable folder structures, this chapter is for you.

Adapt your DMP: part 6

This is the sixth 'Adapt your DMP' section in this tour guide. To adapt your DMP, consider the following elements and corresponding questions:

7 MODULI FREE ONLINE CON SEZIONE ADAPT YOUR DMP

DMP?

UN MODO STRUTTURATO DI PENSARE AI DATI

È UN «LIVING DOCUMENT», CRESCE COL PROGETTO

REGOLE CHIARE=MENO ERRORI DA SUBITO

…CHIARIAMO:
IL PROBLEMA NON È «IMPARARE» A FARE UN DMP MA IMPARARE A GESTIRE I DATI IN MODO FAIR E RESPONSABILE

NEL DMP SEMPLICEMENTE «DICHIARO» COME TRATTERÒ I MIEI DATI

NON È UNA FORMALITÀ MA UNA RESPONSABILITÀ (E UNO STRUMENTO PREZIOSO)

Perché serve FAIR

I DATI SONO DIFFICILI DA TROVARE

UNA VOLTA TROVATI, DIFFICILE ACCEDERE

UNA VOLTA AVUTO ACCESSO, DIFFICILI DA CAPIRE

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

PER QUESTO SERVE FAIR

…FAIR SIGNIFICA
[anche e soprattutto per le macchine]


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

- IDENTIFICATIVI
- METADATI

ACCESSIBLE

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

INTEROPERABLE

- STANDARDS
- ONTOLOGIE

REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE

IL TUTTO, LEGGIBILE DALLE MACCHINE

…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)

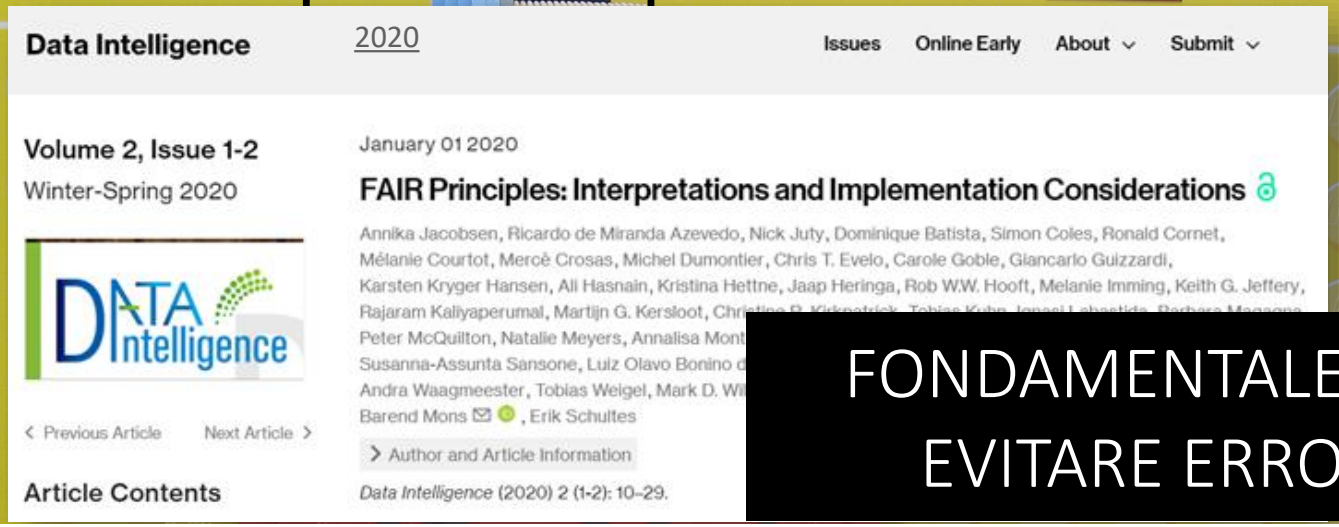
ACCESSIBLE

- CONSERVA I DATI ORA E PER IL FUTURO

REUSABLE

- EVITA SPRECO DI RISORSE
- CREA FIDUCIA

…prima di partire per FAIR



Data Intelligence 2020

Issues Online Early About Submit

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

FAIR Principles: Interpretations and Implementation Considerations

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 Mont… Susanna-Assunta Sansone, Luiz Olavo Bonino d… Andra Waagmeester, Tobias Weigel, Mark D. Wi… Barend Mons, Erik Schultes

Author and Article Information

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

Previous Article Next Article

Article Contents

FONDAMENTALE PER EVITARE ERRORI

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

FAIR train – GoFAIR video

FARM DATA TRAIN

Focus

ASSICURARE IL RIUSO

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

DATA Intelligence

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

Previous Article Next Article

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

Obiettivo

DATI
FAIR BY DESIGN

FAIR principles

To be Findable:

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

TO BE ACCESSIBLE:

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

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

Force 11

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

FAIR/Open

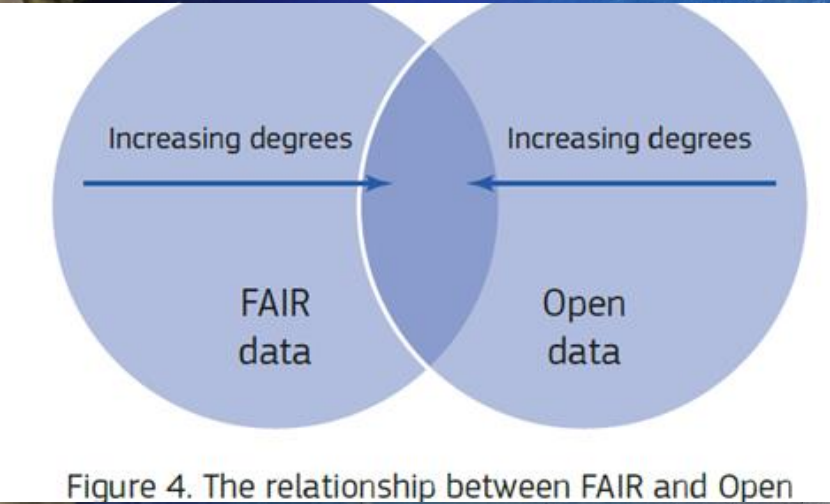


Figure 4. The relationship between FAIR and Open

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

FAIR in sintes

FAIR data training
- Findable
- Accessible
- Interoperable
- Reusable

FAIR for Developers

FAIR data self-assessment tool

+SHARE

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

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 leve

- 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 di of the data being identified):

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

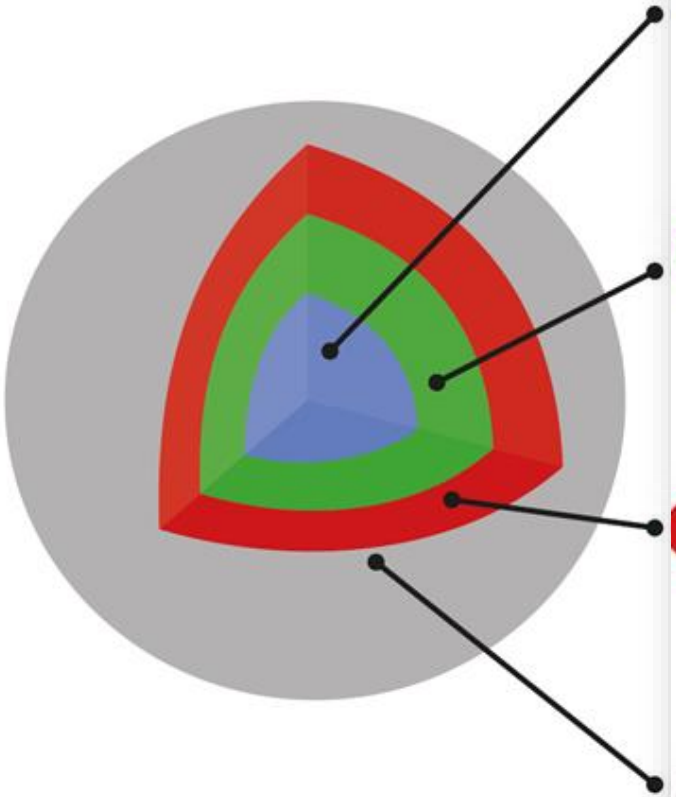
Complementary to the assignment of persistent identifiers is their proper

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

Nov. 20, 2018

Define	Implement			Embed and sustain	
Concepts for FAIR implementation	FAIR culture	FAIR ecosystem	Skills for FAIR	Incentives and metrics for FAIR data and services	Investment in FAIR
Rec. 1: Define FAIR for implementation	Rec. 4: Develop Interoperability frameworks	Rec. 7: Support semantic technologies	Rec. 10: Professionalise data science & stewardship roles	Rec. 12: Develop metrics for FAIR Digital Objects	Rec. 14: Provide strategic and coordinated funding
Rec. 2: Implement a Model for FAIR Digital Objects	Rec. 5: Ensure data management via DMPs	Rec. 8: Facilitate automated processing	Rec. 11: Implement curriculum frameworks and training	Rec. 13: Develop metrics to certify FAIR services	Rec. 15: Provide sustainable funding
Rec. 3: Develop components of a FAIR ecosystem	Rec. 6: Recognise & reward FAIR data & stewardship	Rec. 9: Certify FAIR services	Above line = priority recommendations		
Rec. 16: Apply FAIR broadly	Rec. 18: Cost data management	Rec. 22: Use information held in DMPs	Below line = supporting recommendations	Rec. 25: Implement and monitor metrics	Rec. 27: Open EOSC to all providers but ensure services are FAIR
Rec. 17: Align and harmonise FAIR and Open data policy	Rec. 19: Select and prioritise FAIR digital objects	Rec. 23: Develop components to meet research needs		Rec. 26: Support data citation and next generation metrics	
	Rec. 20: Deposit in Trusted Digital Repositories	Rec. 24: Incentivise research infrastructures to support FAIR data			
	Rec. 21: Incentivise reuse of FAIR outputs				

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 use to process and analyse the data.

METADATA

Contextual documentation

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

ECD, 2015

FAIR: technology VS domain

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

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

Box 2 | The FAIR Guiding Principles

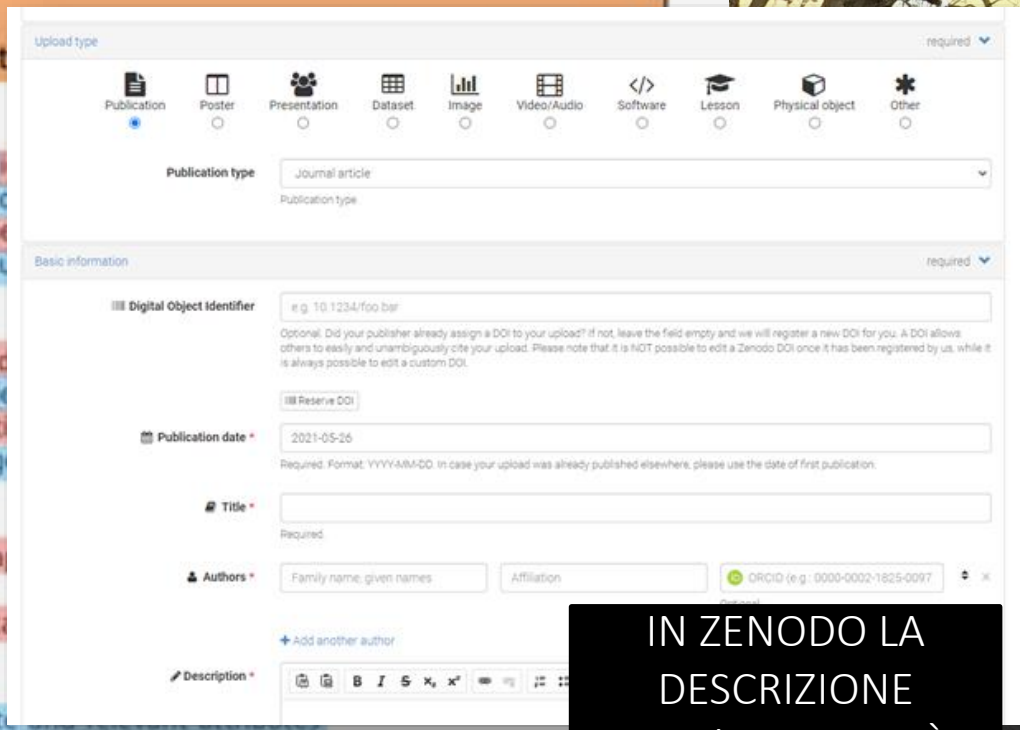
To be Findable:
F1. (meta)data are assigned a globally unique and persistent
F2. data are described with rich metadata (defined by R1 bel
F3. metadata clearly and explicitly include the identifier of th
F4. (meta)data are registered or indexed in a searchable resou

To be Accessible:
A1. (meta)data are retrievable by their identifier using a stand
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorizat
A2. metadata are accessible, even when the data are no long

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly a
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 accurat
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



IN ZENODO LA DESCRIZIONE DELL'OGGETTO È GIÀ IMPOSTATA

FAIR for dummies

COSA DEVE FARE IL RICERCATORE

COSA FA IL REPOSITORY

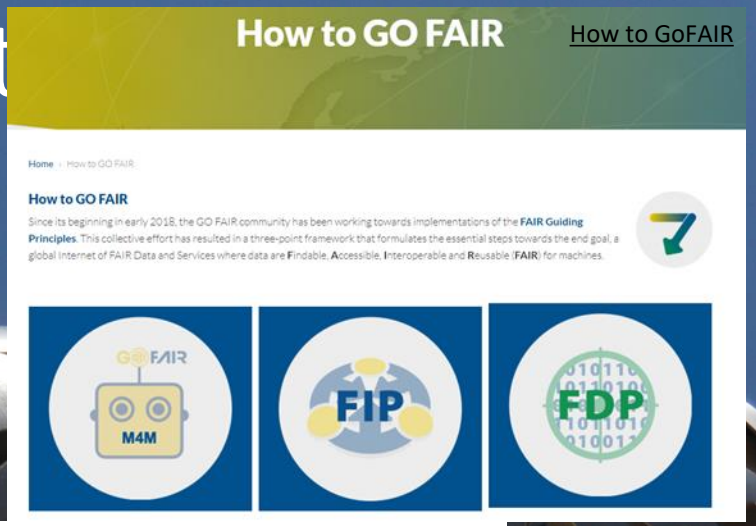
2019

Explanation of the FAIR data principles

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 - tecnicament



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.

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

FAIR Implementation profiles

CREARE FAIR IMPLEMENTATION PROFILES RIUSABILI ANCHE DA ALTRI NELLA COMUNITÀ DISCIPLINARE

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

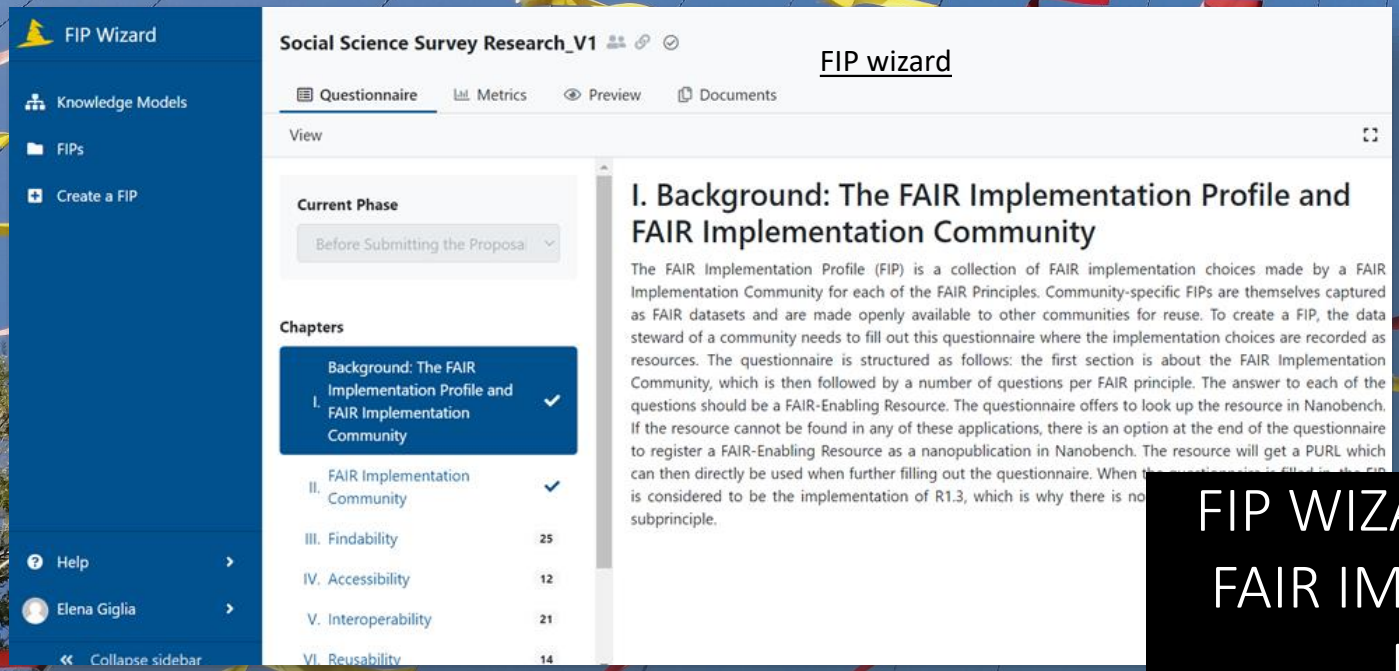
2020

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

PAROLA CHIAVE: CONVERGENZA



FIP wizard

FIP WIZARD PER CREARE FAIR IMPLEMENTATION PROFILES

[FAIR all'opera]

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

VODAN Africa&Asia

Making data and metadata FAIR ensures that these data are discoverable on the Internet of FAIR Data and Services. Central tot his 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.

OAI12

PRESENTATO A OAI12

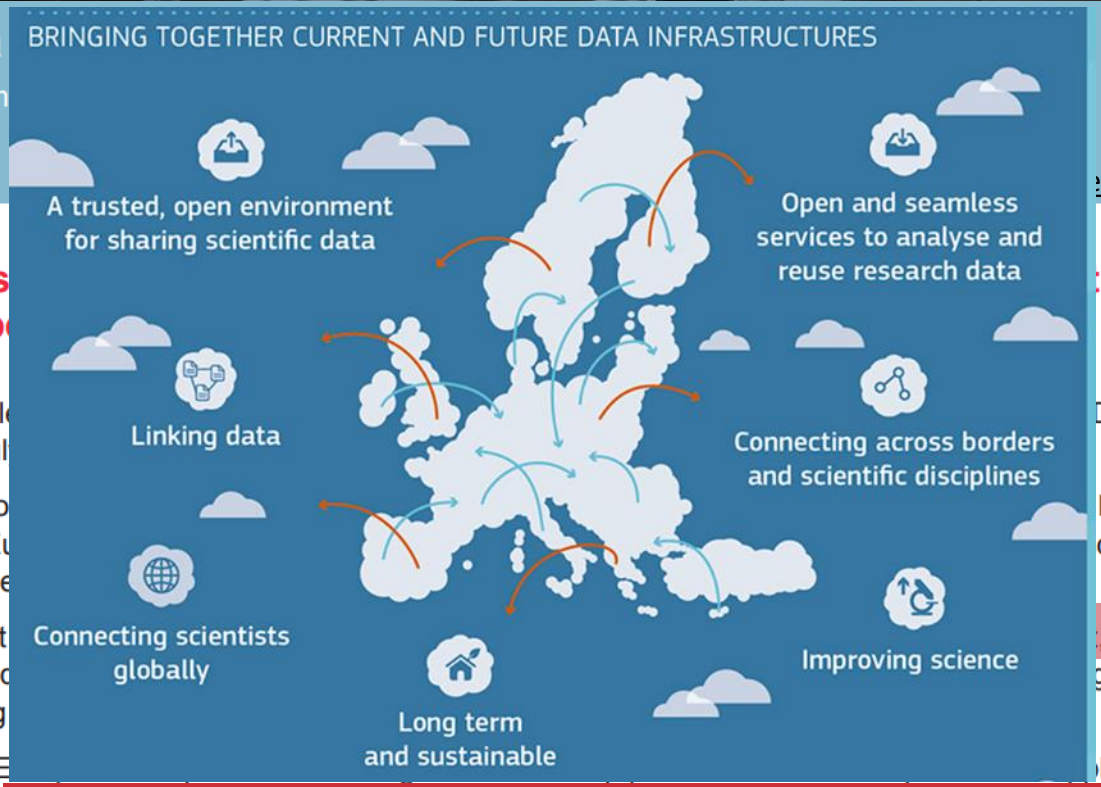
[perché c'è EOSC!]

The Vienna

Vienna, 23 Novem

u 1 8
· a t

enna, Nov.23, 2018



We, Ministers
European Op

: of the

1. Recall the chall
Brussels on 10 Jul

Declaration" signed in

2. Reaffirm the po
the vision of the Eu
States, sustainable

Europe. Confirm that
ciplines and Member

3. Recognise that
iterative and based
consensus among

, by its nature
gue to build trust and

4. Highlight that E
services for Scien
reaching out over

lication of cloud
the world,

5. Recall that the

d the federated

ACCESSO TRASPARENTE A DATI FAIR
«AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»

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
EOSC NON È UN REPOSITORY O UN SERVIZIO «CLOUD»
NON SI FA «UPLOAD» DEI DATI DENTRO EOSC
SI RENDONO I DATI FAIR IN MODO CHE I *SERVIZI* IN EOSC POSSANO TROVARLI («FINDABLE»)
E FORNIRE UN ACCESSO «SEAMLESS» A 20 MILIONI DI RICERCATORI

[parlare di dati significa anche data stewards]

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;

DATA STEWARDS SONO UNO DEI FATTORI DI SUCCESSO DI EOSC

SERVONO 500.00 DATA STEWARDS

Report, 2016

The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation[9]. The success of the EOSC depends upon it.

Ma chi è il data steward?

Data stewardship is the responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimising the usability, reusability and reproducibility of the resulting data.

It differs from data management, in the sense that data management concerns all actual, operational data-related activities in any phase of the data lifecycle, while data stewardship refers to the assignment of responsibilities in, and planning of, data management.

PIANIFICA E METTE IN ATTO
PRIMA, DURANTE E DOPO IL PROGETTO DI RICERCA,
LE AZIONI VOLTE A OTTIMIZZARE
L'USABILITÀ, RIUSABILITÀ E RIPRODUCIBILITÀ DEI DATI

Ma chi è il data steward?

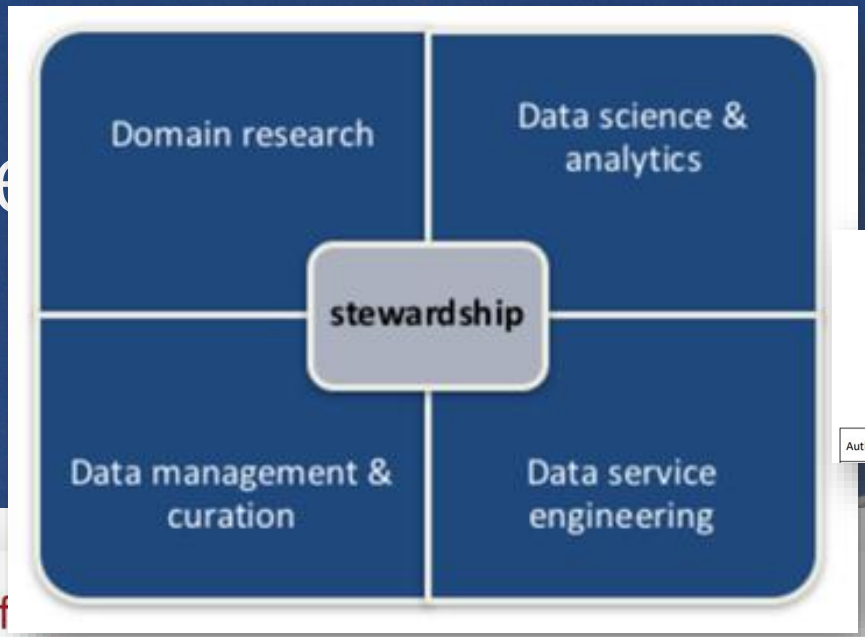
steward role has eight competence areas:

- Policy/strategy
- Compliance
- Alignment with FAIR data principles
- Services
- Infrastructure
- Knowledge management
- Network
- Data archiving

2019

DATA STEWARD HA 8 AREE DI COMPETENZA UNA DELLE FUNZIONI CHIAVE È DI FARE DA «PONTE» FRA DIVERSE DISCIPLINE E PROFESSIONI

Il profilo de



2018

EOSC pilot
The European Open Science Cloud for Research Pilot Project

D7.3: Skills and Capability Framework

Author(s)	Angus Whyte, Jerry de Vries, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi, Kevin Ashley

KØBENHAVNS UNIVERSITET

06/04/2020 8

Competence Prof

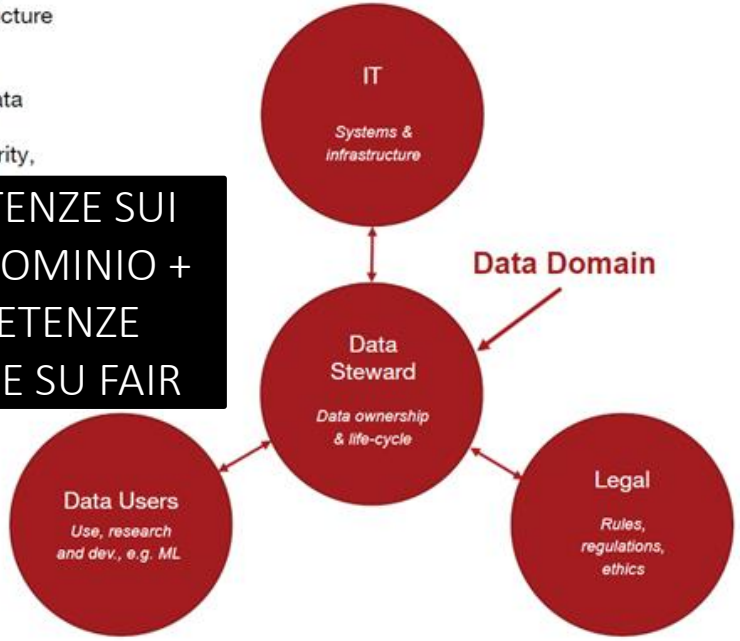
A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

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 [illegible] 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.

COMPETENZE SUI DATI DI DOMINIO + COMPETENZE TECNICHE SU FAIR



Copenhagen Univ. June 17 2020

FAIRifi

Jan 2020

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

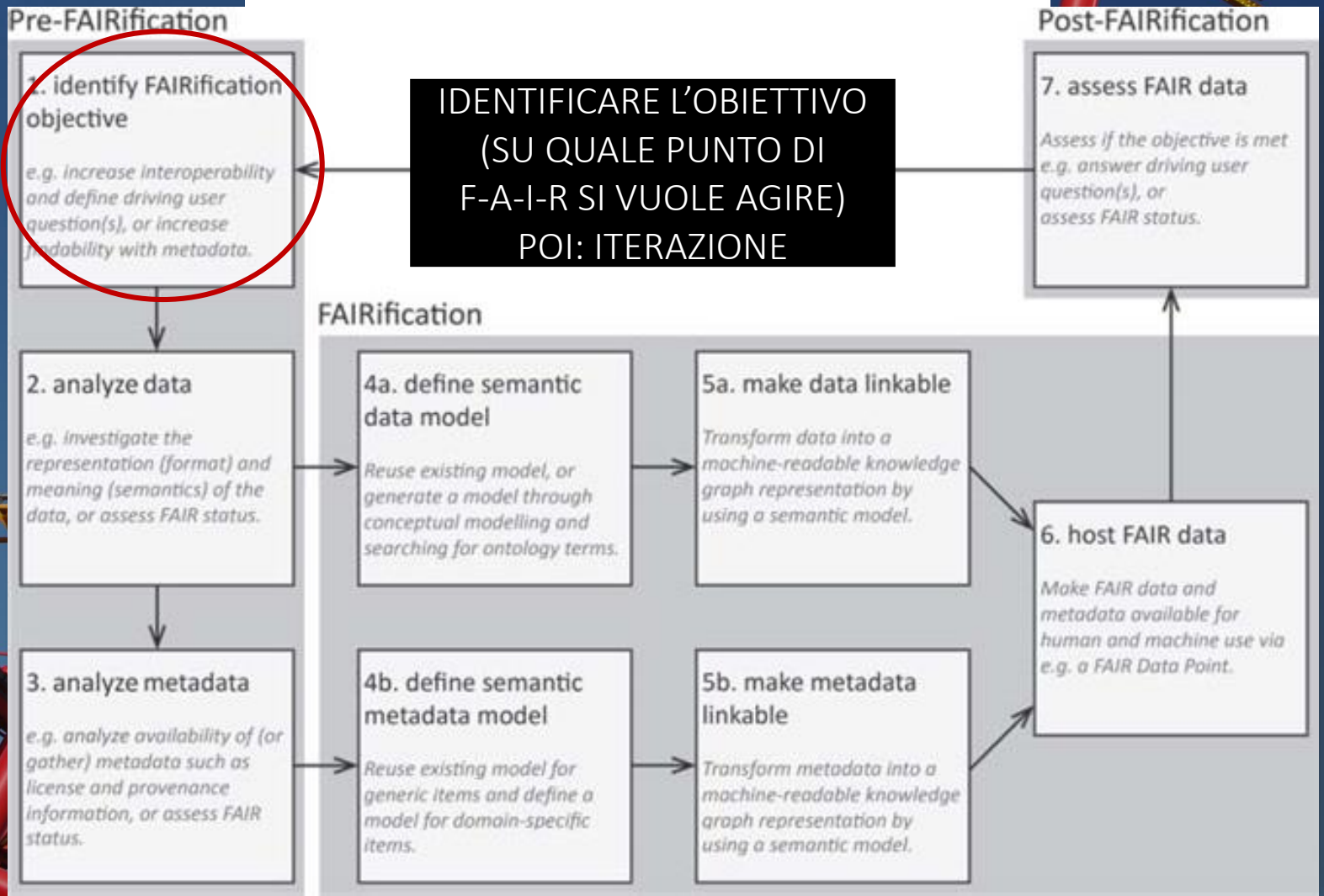
A Generic Workflow for the Data FAIRification Process

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



IDENTIFICARE L'OBIETTIVO
(SU QUALE PUNTO DI
F-A-I-R SI VUOLE AGIRE)
POI: ITERAZIONE

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

EOSCFAIR
Executive Board Working Group

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

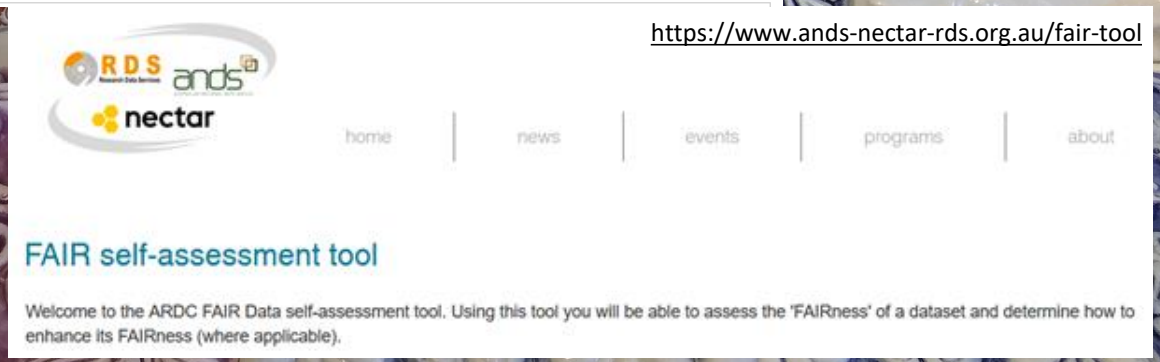
How accessible is the data? — No access to data or metadata

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

UTILISSIMO PER PORSI LE DOMANDE GIUSTE… MA SOGGETTIVO…

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



FAIR aware

- DOMANDE
- VERIFICA INTENZIONI
- VERIFICA CONOSCENZA
- FORNISCE INFORMAZIONI

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.

FAIR questions

FINDABLE

1. Are you aware that a data(set) should be assigned a global persistent and resolvable identifier when deposited with a dat repository?

2. Are you aware that when you deposit a data(set) in a data you will need to provide discovery metadata in order to make data(set) findable, understandable and reusable to others?

3. Are you aware that the data repository providing access to data(set) should make the metadata describing your data(set) in a format readable by machines as well as humans?

ACCESSIBLE

4. Are you aware that access to your data(set) may need to b and that metadata should include licence information under w data(set) can be reused?

5. Are you aware that metadata should remain available over if the data(set) is no longer accessible?

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

What does this mean?

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

Why is this important?

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

How to do this?

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

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

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

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

- 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 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.

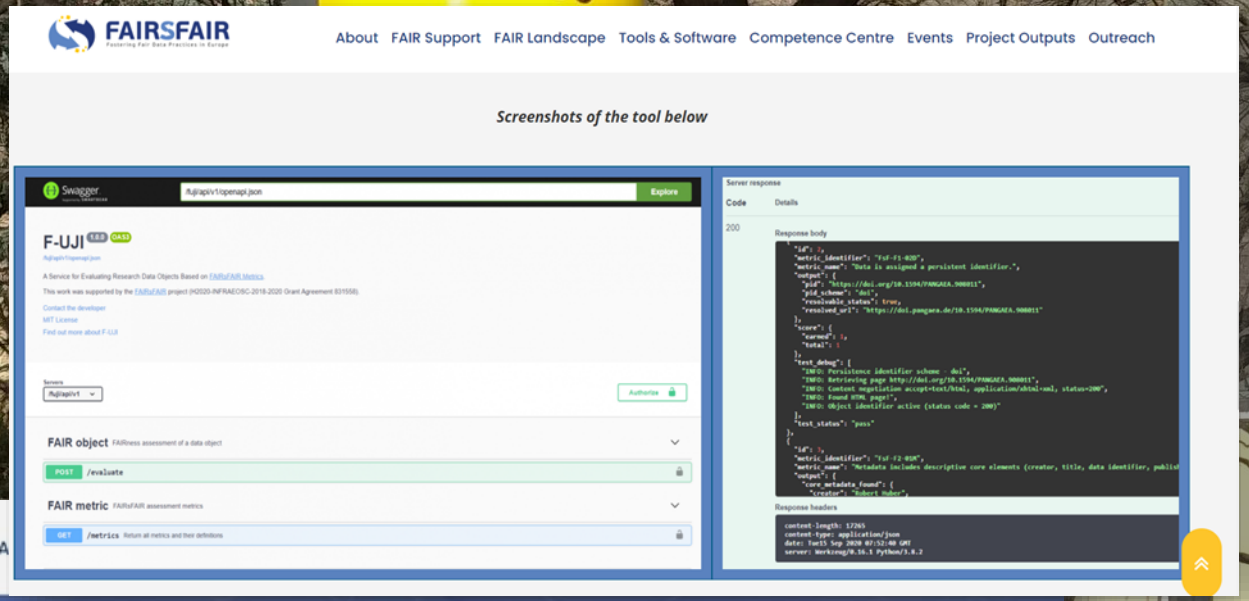
FAIR METRICS GEN2 - IDENTIFIER PERSISTENCE

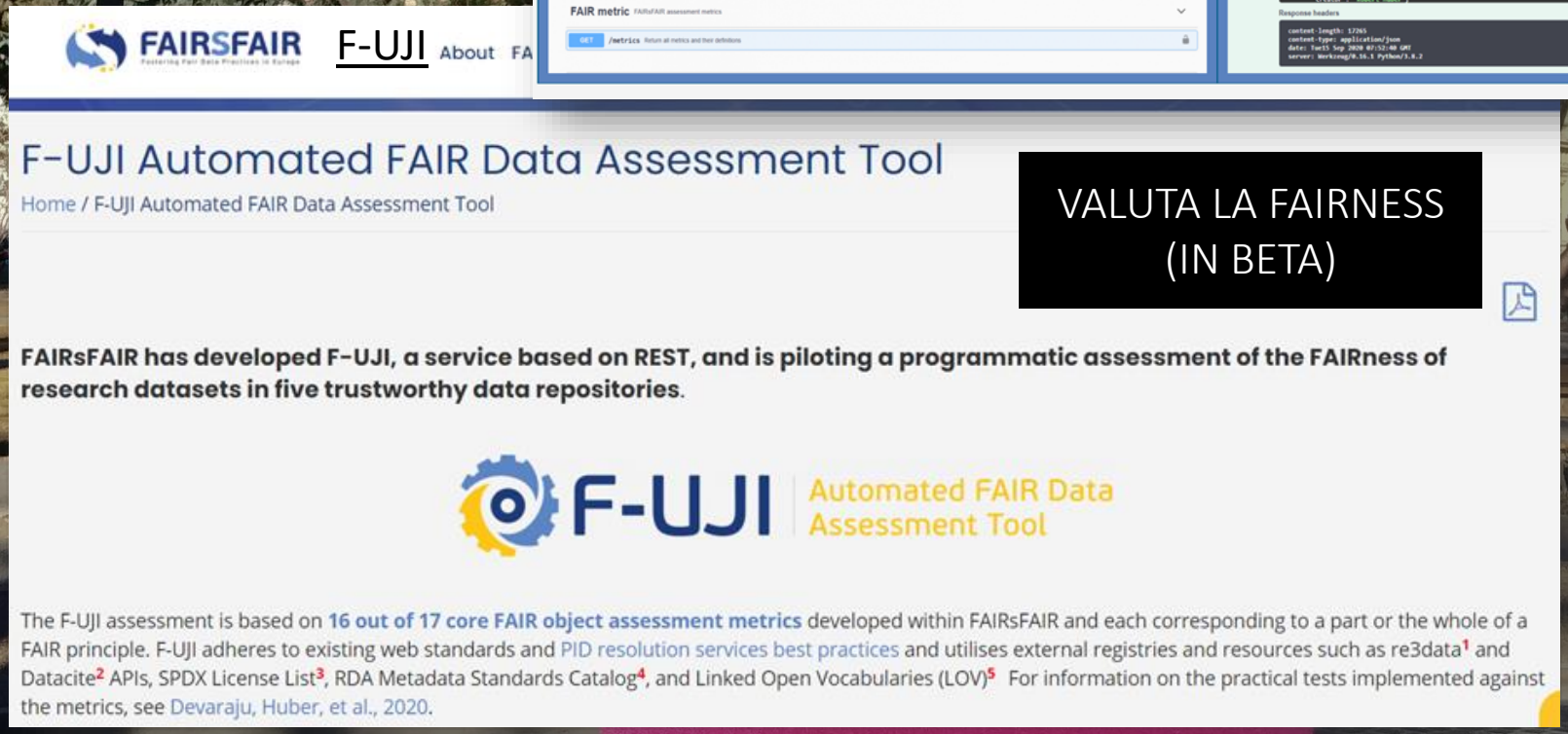
Status: Failure
Principle tested: F1
Description: Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema). For URLs that don't follow 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

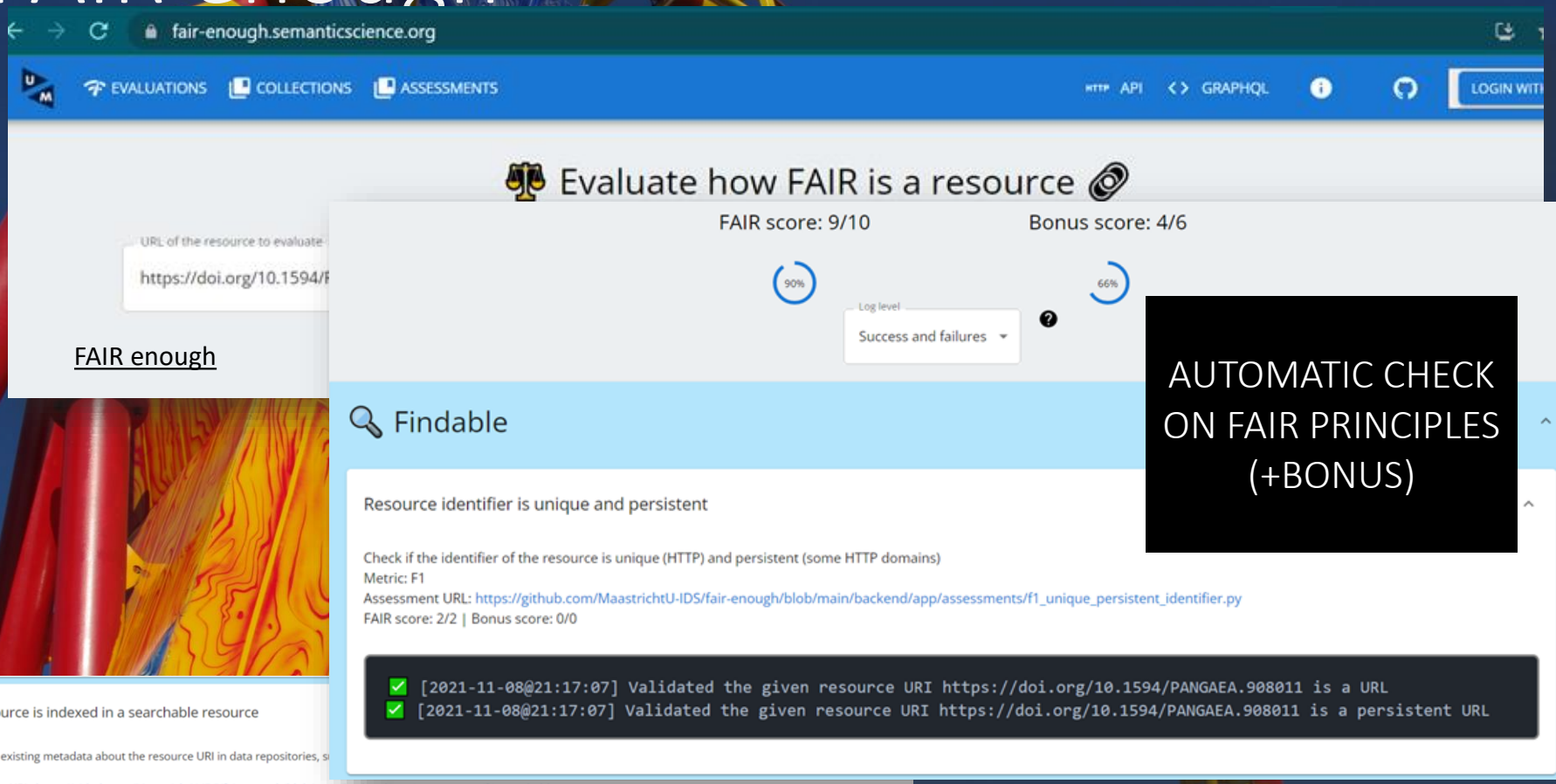
INFO: The metadata GUID appears to be a URL. Testing known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).
FAILURE: The metadata GUID does not conform with any known permanent-URL system.

F-UJI


FAIRSFAIR
About FAIR Support FAIR Landscape Tools & Software Competence Centre Events Project Outputs Outreach
Screenshots of the tool below
Swagger
fuji/api/v1/openapi.json
Explore
F-UJI
A Service for Evaluating Research Data Objects Based on FAIRsFAIR Metrics
Contact the developer
MIT License
Find out more about F-UJI
Servers
fuji/api/v1
Authorize
FAIR object FAIRness assessment of a data object
POST /evaluate
FAIR metric FAIRsFAIR assessment metrics
GET /metrics Return all metrics and their definitions
Server response
Code Details
200
Response body
Response headers

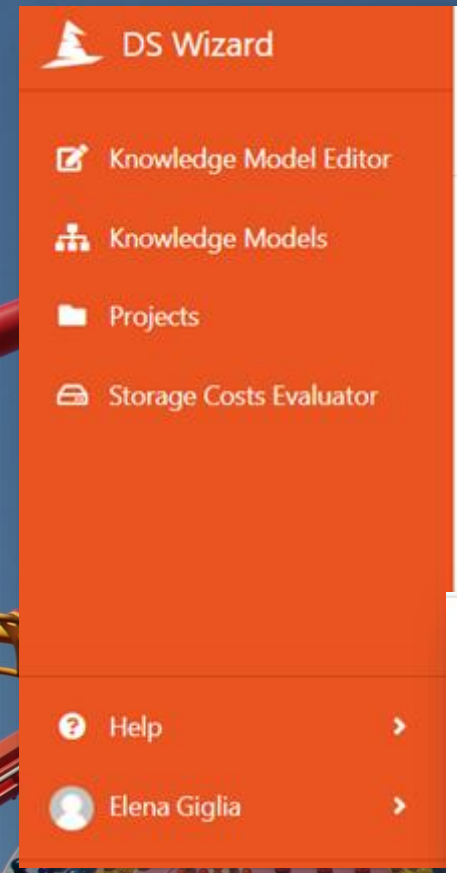

FAIRSFAIR
F-UJI
About
F-UJI Automated FAIR Data Assessment Tool
Home / F-UJI Automated FAIR Data Assessment Tool
VALUTA LA FAIRNESS
(IN BETA)
FAIRsFAIR has developed F-UJI, a service based on REST, and is piloting a programmatic assessment of the FAIRness of research datasets in five trustworthy data 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.

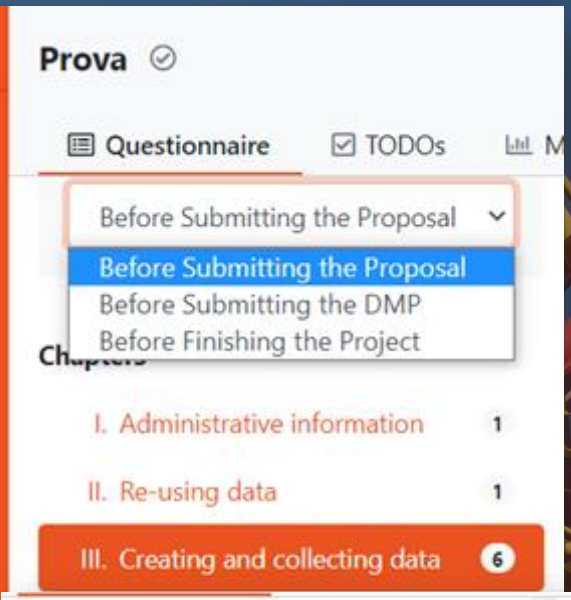
FAIR enough

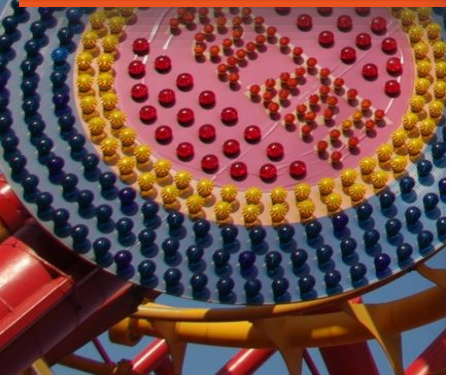


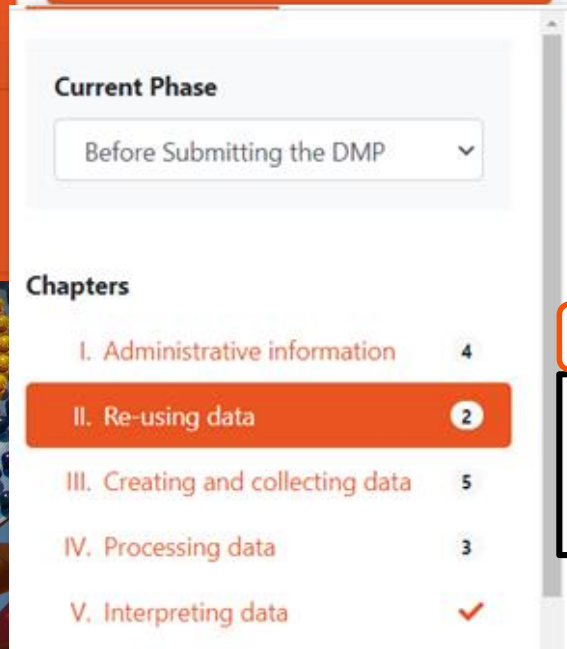
AUTOMATIC CHECK ON FAIR PRINCIPLES (+BONUS)

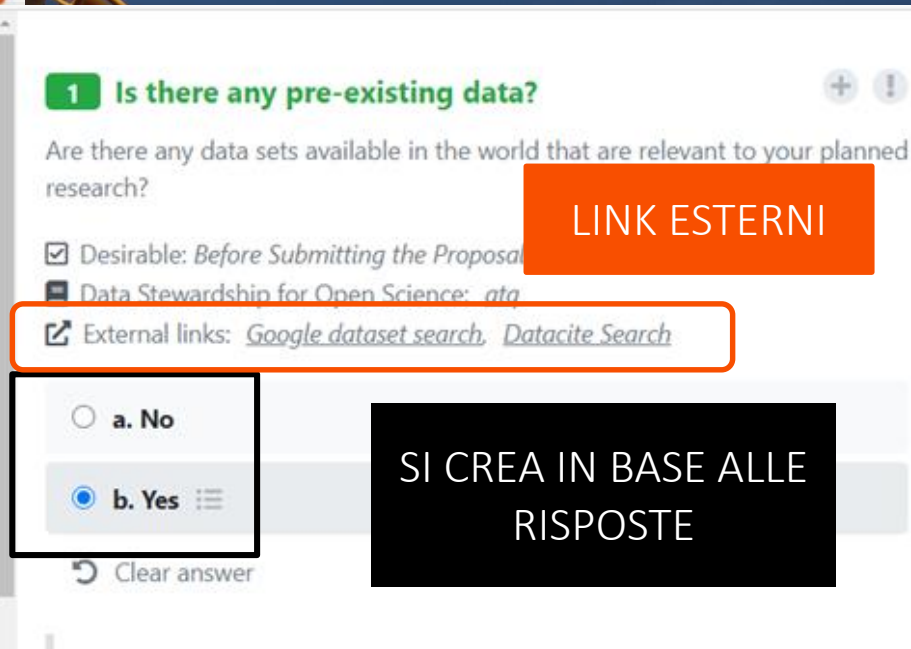
FAIR Wizard





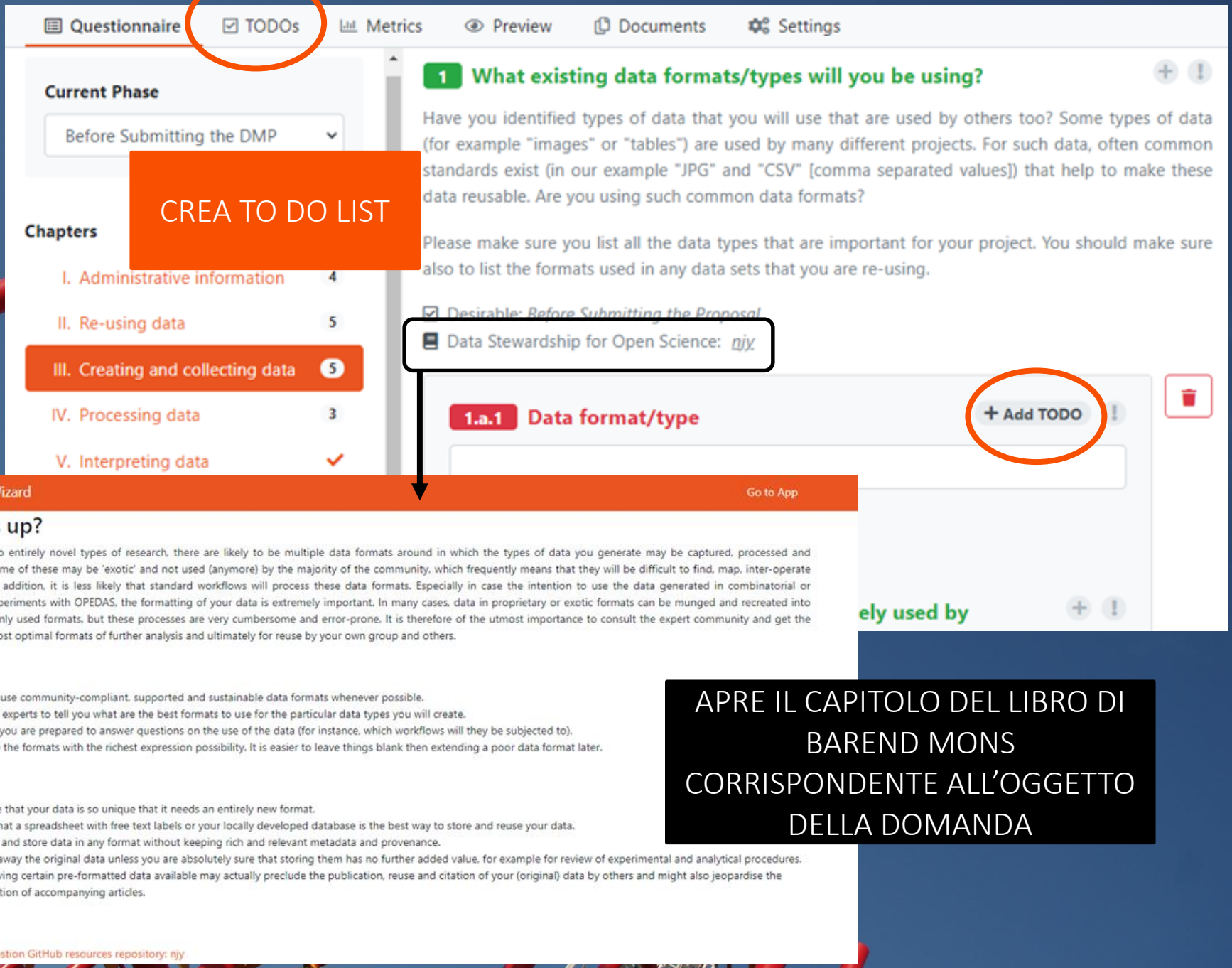






LINK ESTERNI

SI CREA IN BASE ALLE RISPOSTE


Questionnaire
TODOs
Metrics
Preview
Documents
Settings
Current Phase
Before Submitting the DMP
CREA TO DO LIST
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: njy
1.a.1 Data format/type
+ Add TODO
ely used by
DS Wizard
Go to App
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: njy
APRE IL CAPITOLO DEL LIBRO DI BAREND MONS CORRISPONDENTE ALL'OGGETTO DELLA DOMANDA

FAIR Wizard

ALLA FINE IL SISTEMA ESTRAE LE INFORMAZIONI RILEVANTI PER IL DMP E LO COMPILA AUTOMATICAMENTE

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.

IN PRATICA

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 context/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).

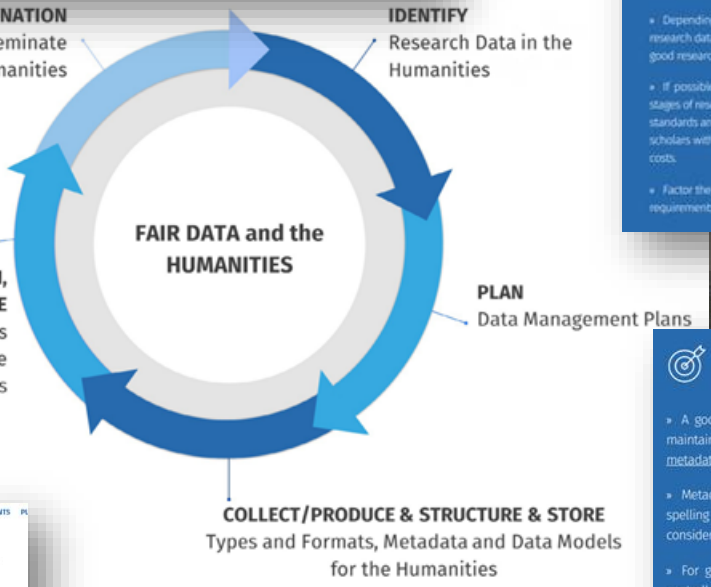
» Consider the right questions to be answered in your DMP that can account for discipline-specific requirements. The DMP templates suggested by funders are quite high level and provide generic guidance for file naming or versioning conventions, database structuring, and can be a good start. Tools like the dmponline.dcc.ac.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.



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 music editions. The query language SPARQL and the representation tool for linked data JSON-LD are common standards for RDF (refers to FAIR principle I1).

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

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

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

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. ...rary of Congress maintains a controlled vocabulary for subject headings: ...thorities/subjects.html

...e metadata should include a clear and explicit reference to the dataset ...h the inclusion of a PID in the metadata (see section on Trustworthy Data ...rsistent Identifiers).

...data as rich as possible in order to better contextualise your data and ...sider more detailed descriptions, and fuller provenance information, as ...ectrum of available metadata fields.

...be machine-readable.

Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | **February 2020**

February 2020

FINDABLE

Findable - Metadata

RECOMMENDATIONS

» A good starting point is to consult the Metadata Standards Directory, a community-maintained directory hosted by the Research Data Alliance: https://rd-alliance.github.io/metadata-directory/

» Metadata works best when terminology is consistent, e.g. naming conventions are followed, spelling is normalised, and so on. Depending on the complexity and size of your metadata, consider using a tool such as Open Refine to 'clean' your metadata.

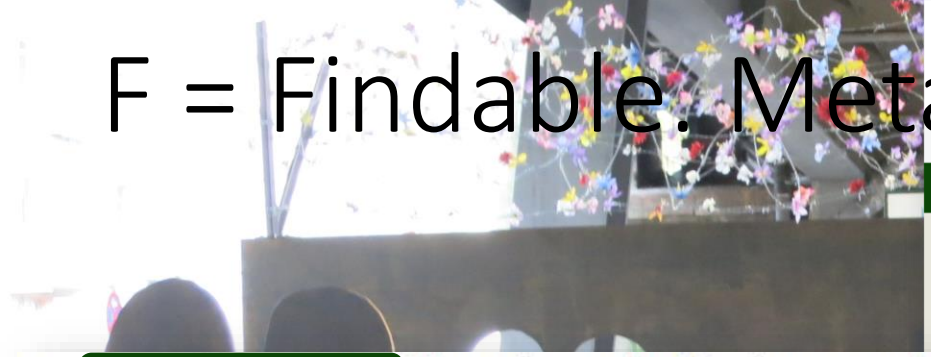
» For greater searchability and interoperability, researchers should also consider using controlled vocabularies to identify common terminology when populating metadata fields. For example, the Library of Congress maintains a controlled vocabulary for subject headings: http://id.loc.gov/authorities/subjects.html

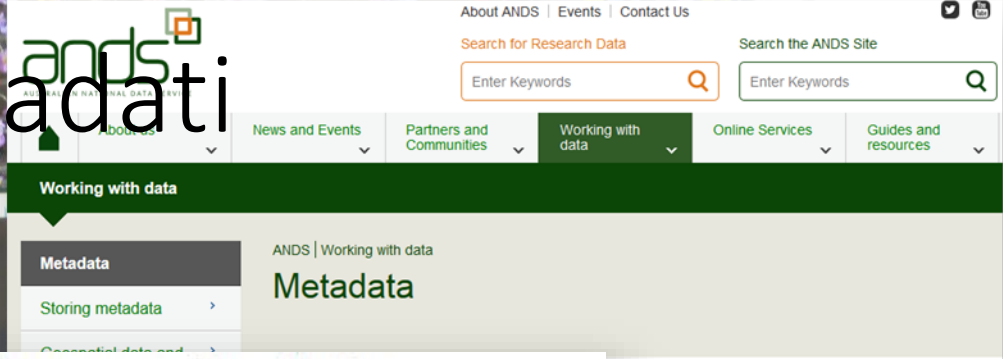
» For findability, the metadata should include a clear and explicit reference to the dataset it describes, through the inclusion of a PID in the metadata (see section on Trustworthy Data Repositories and Persistent Identifiers).

» Make your metadata as rich as possible in order to better contextualise your data and facilitate reuse. Consider more detailed descriptions, and fuller provenance information, as well as using the spectrum of available metadata fields.

» Metadata should be machine-readable.

F = Findable. Metadati

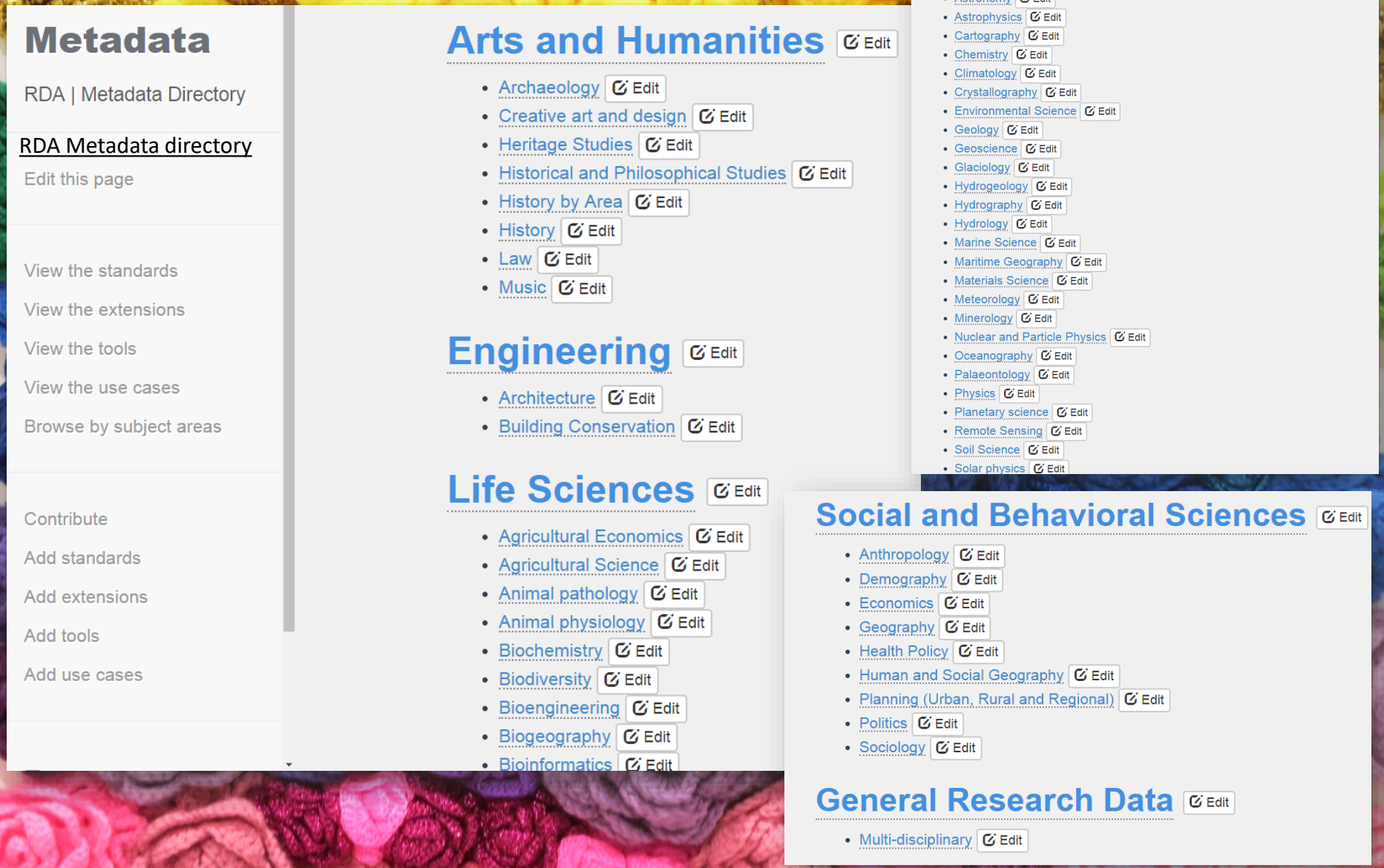




- **Descriptive metadata**: information required for discovery and assessment of the collection,
  - e.g. title, contributors, subject or keywords, study description, and location and dates of the study.
- **Provenance metadata**: this relates to the origins and processing of the data, and enables interpretation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



Metadata

RDA | Metadata Directory

RDA Metadata directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

Arts and Humanities Edit

- Archaeology Edit
- Creative art and design Edit
- Heritage Studies Edit
- Historical and Philosophical Studies Edit
- History by Area Edit
- History Edit
- Law Edit
- Music Edit

Engineering Edit

- Architecture Edit
- Building Conservation Edit

Life Sciences Edit

- Agricultural Economics Edit
- Agricultural Science Edit
- Animal pathology Edit
- Animal physiology Edit
- Biochemistry Edit
- Biodiversity Edit
- Bioengineering Edit
- Biogeography Edit
- Bioinformatics Edit

Physical Sciences & Mathematics Edit

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

Social and Behavioral Sciences Edit

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

General Research Data Edit

- Multi-disciplinary Edit

F = findable. Metadata tools

What CEDAR does

https://metadatacenter.org/

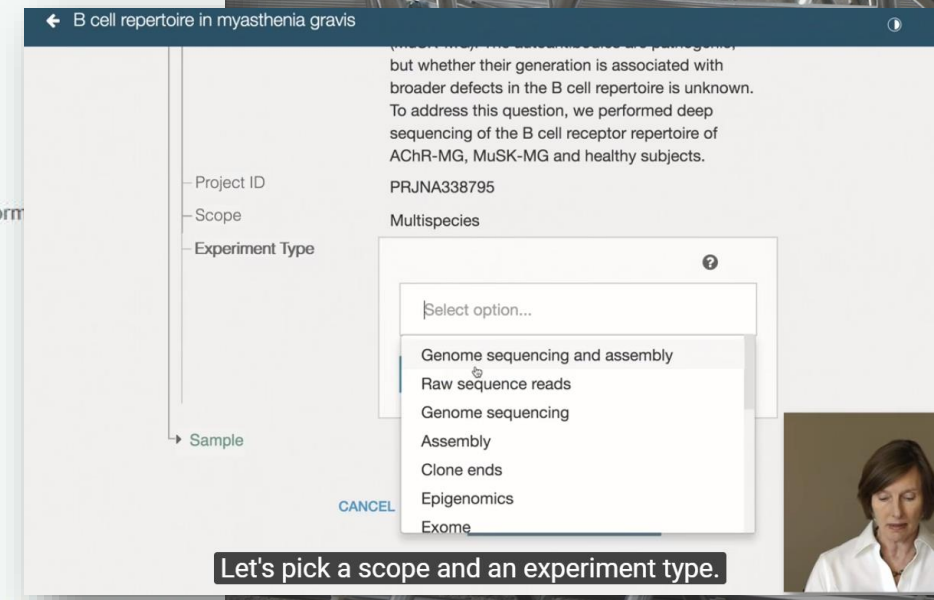
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools wil
metadata record is created to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools hel
to users, and download the information that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

What can CEDAR do for me already?

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

- create user-friendly, shareable forms for collecting metadata, with features like
  - nested and repeatable elements and fields
  - reusable elements
  - control over tool tips, field titles, and field descriptions
- share your forms and metadata
  - provide a link to your metadata editors, so they can enter metadata responses based on your form
  - 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.



Findable — Metadata cre

FAIR cookbook

Search this book...

FAIR Cookbook

FOREWORD

Introduction

Ethical values of FAIR

Glossary

RECIPES

Findability

Accessibility

Interoperability

1. Interlinking data from different sources

2. Identifier mapping with BridgeDb

3. Introduction to terminologies

9. Creating a metadata profile

Recipe Overview

Reading Time
20 minutes

Executable Code
Yes

Difficulty

Creating a metadata profile

Recipe Type
Hands-on

Audience
Principal Investigator, Data Manager, Data Scientist

Cite me with FCB026

9.1. Graphical Overview

FAIR Cookbook

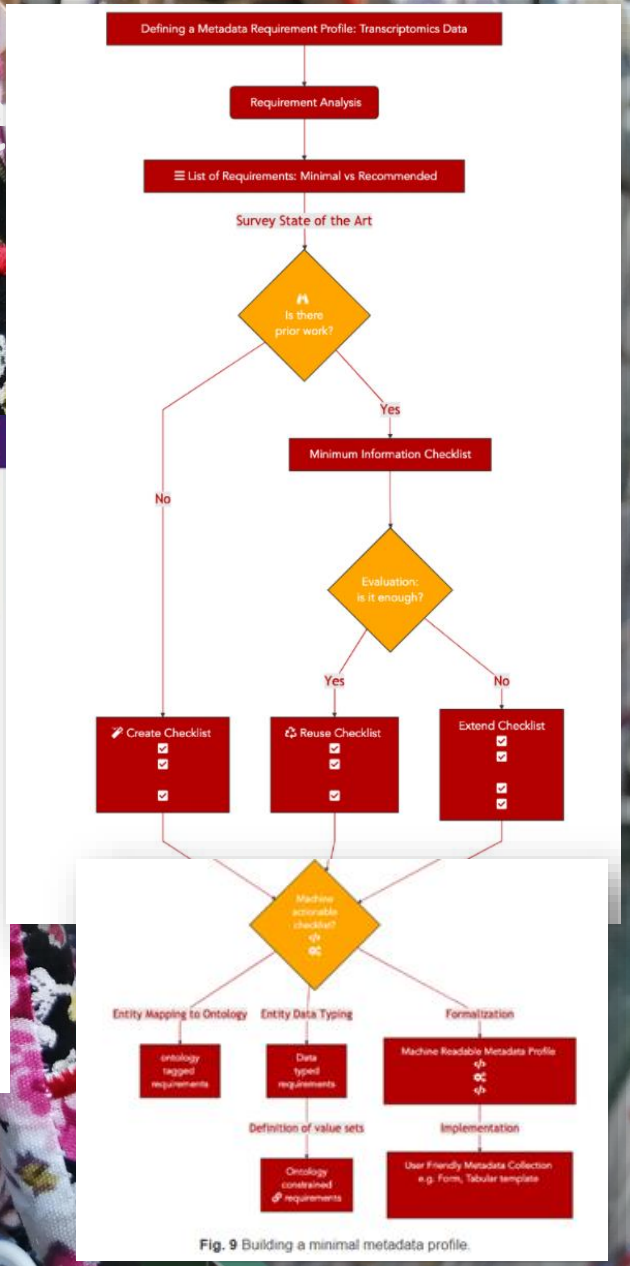
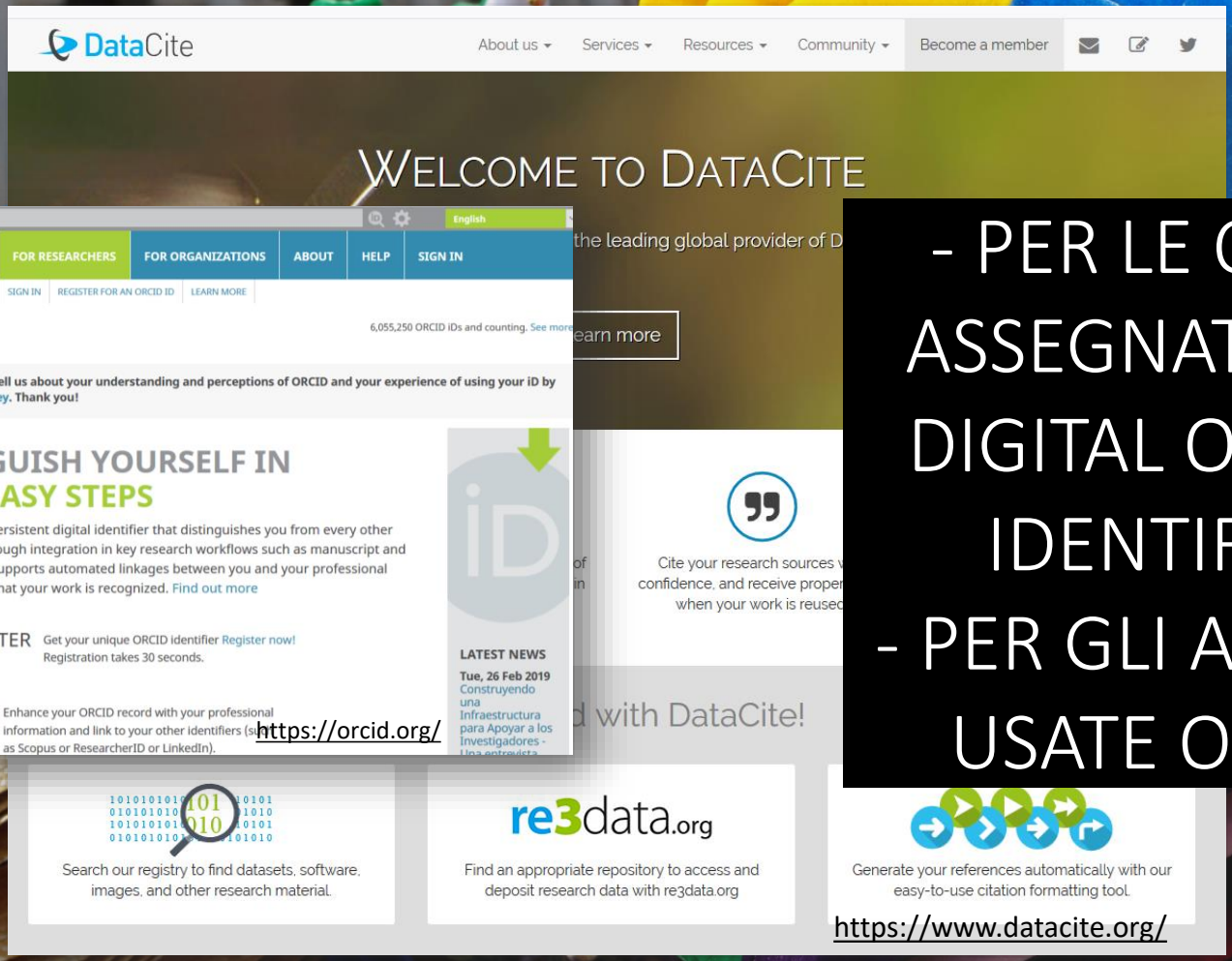


Fig. 9 Building a minimal metadata profile.

F = Findable. Identificativi persistenti



https://orcid.org/

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

https://www.datacite.org/

ACCESSIBLE

A = Accessible

- **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.

CESSDA Guide

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

Accessible

TRUSTED REPOSITORIES IMPORTANCE OF CONTEXT

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

DIGITAL PRESERVATION

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

RECOMMENDATIONS

ALLEA report, p. 28

» To ensure the best possible stewardship of your data, choose to deposit it in a digital repository that is certified by a recognised standard such as the CoreTrustSeal. The Registry of Research Data Repositories (re3data) provides a good starting point, noting disciplines, standards, content types, certification status and more. FAIRsharing (manually curated information on standards, databases, policies and collections) allows you to search databases by subject, and includes entries tagged 'Humanities and Social Sciences'.

» Use disciplinary repositories where they exist, as they are more likely to be developed around domain expertise, disciplinary practices and community-based standards, which will promote the findability, accessibility, interoperability and ultimately the reuse and value of your data. The level of curation available in a repository is key to data quality and reusability.

» Datasets should be assigned persistent identifiers (PID). Most repositories that are designed for long-term preservation will automatically assign or 'mint' persistent identifiers for your datasets, so choosing a quality repository will automate this step. Consider as well signing up for ORCID, a free service that assigns persistent identifiers to individuals/authors.

» To facilitate findability of all research outputs, bidirectional links should be created between publications related outputs, such as data (using PIDs).

» Include the richest metadata possible with your deposited data so that others can find it, understand the parameters under which it was created, and understand the conditions under which they can access and/or reuse it. See recommendations in this report in the sections on Licences and Metadata for more information.

BIDIRECTIONAL **LINKS**

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

A = Accessible – Data repositories

Zenodo

Featured communities

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»

Dataverse

Open source research data repository software

Researchers

Journals

Institutions

Developers

DRYAD

for your research data

How it works

Login

Submit

Review

Cite

Figshare

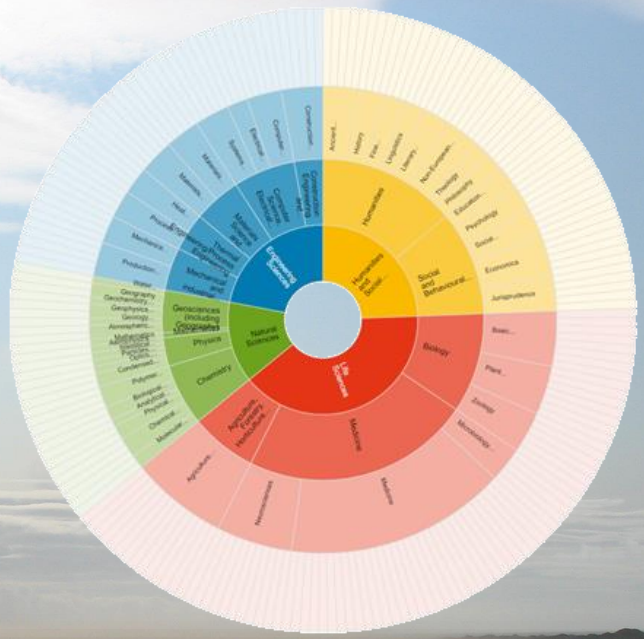
store, share, discover research

get more citations for all of the outputs of your academic research

over 30,000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

A = Accessible.
Cercate un archivio?



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
Search... https://www.re3data.org/ Search

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

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

Read more

Three new DOI Fabrica features to simplify account management

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

Read more

One step closer towards instant DOI search results

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

Read more

A = Accessible. Data 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://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
			nature.com/sdata/for-authors nature.com/sdata/for-authors#data-deposition	Nature	"open to submissions from a broad range of scientific disciplines, but

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

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

Aug. 2021

By Kirsty, on 17 August 2021

Guest post by James Houghton – Research Data Support Officer

Why not publish a data report article?

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

elsevier.com/journals/genomics-data/2213-5960/guide-for-

Dataset Description

Object Name

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

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

Reposit
UK Ar
10.5284

Publica
05/02/2012

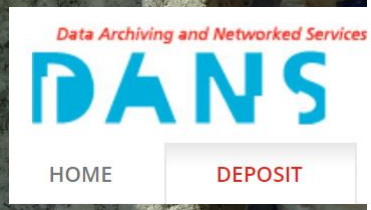
Language
English (a Greek language summary of the project methods and results can be found at www.ucl.ac.uk/asp/ or www.tuarc.trentu.ca/asp/).

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/10.5284/1000208, http://dx.doi.org/10.5284/1000103 and, to a lesser extent, also http://dx.doi.org/10.5284/1000351), albeit with due attention to the fact that the intensive methods used are not identical. The ASP data is particularly reusable because artefact locations, dates and identifications are recorded individually in the database rather than in aggregate. The standing structures and terraces from Antikythera are also the kinds

A = Accessible. Formati



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

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

INTEROPERABLE

DATI CON SOFTWARE/MACCHINE

DATI CON ALTRI DATI

I = Inteoperable. Standards

2020

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

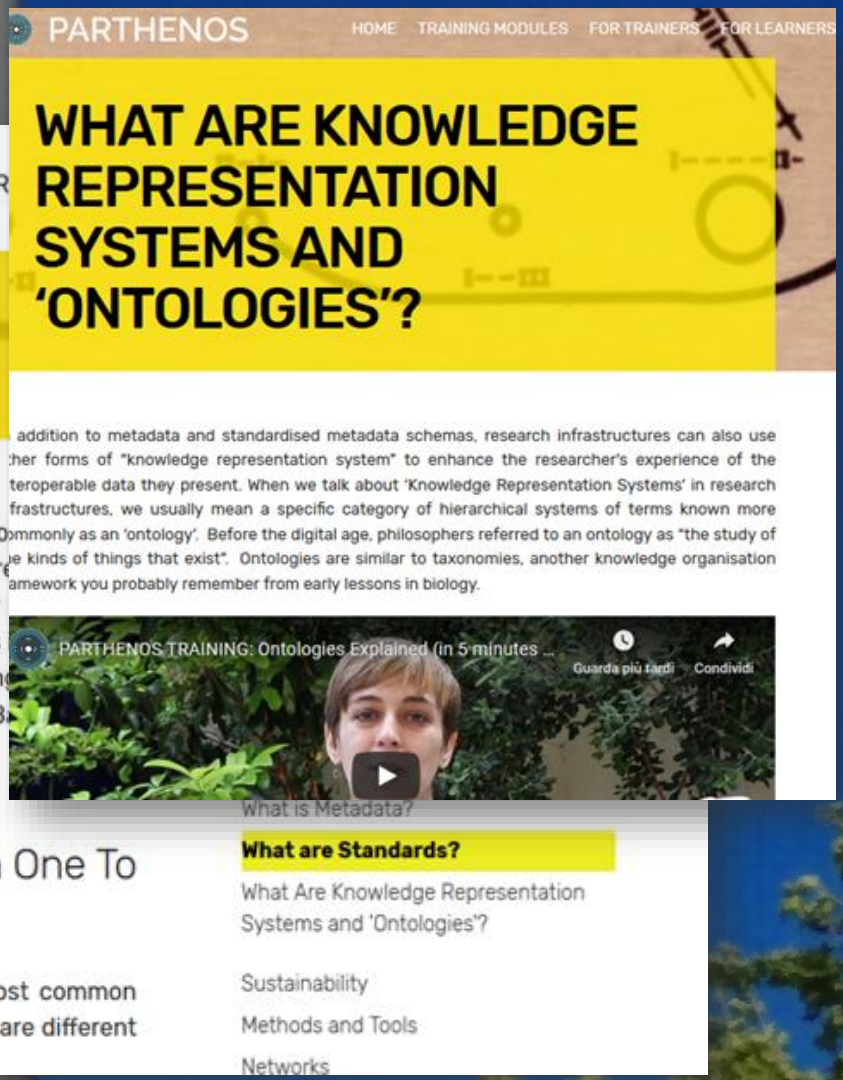
» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies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

NORMALIZZARE, USARE AUTHORITY FILES (es. VIAF)

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

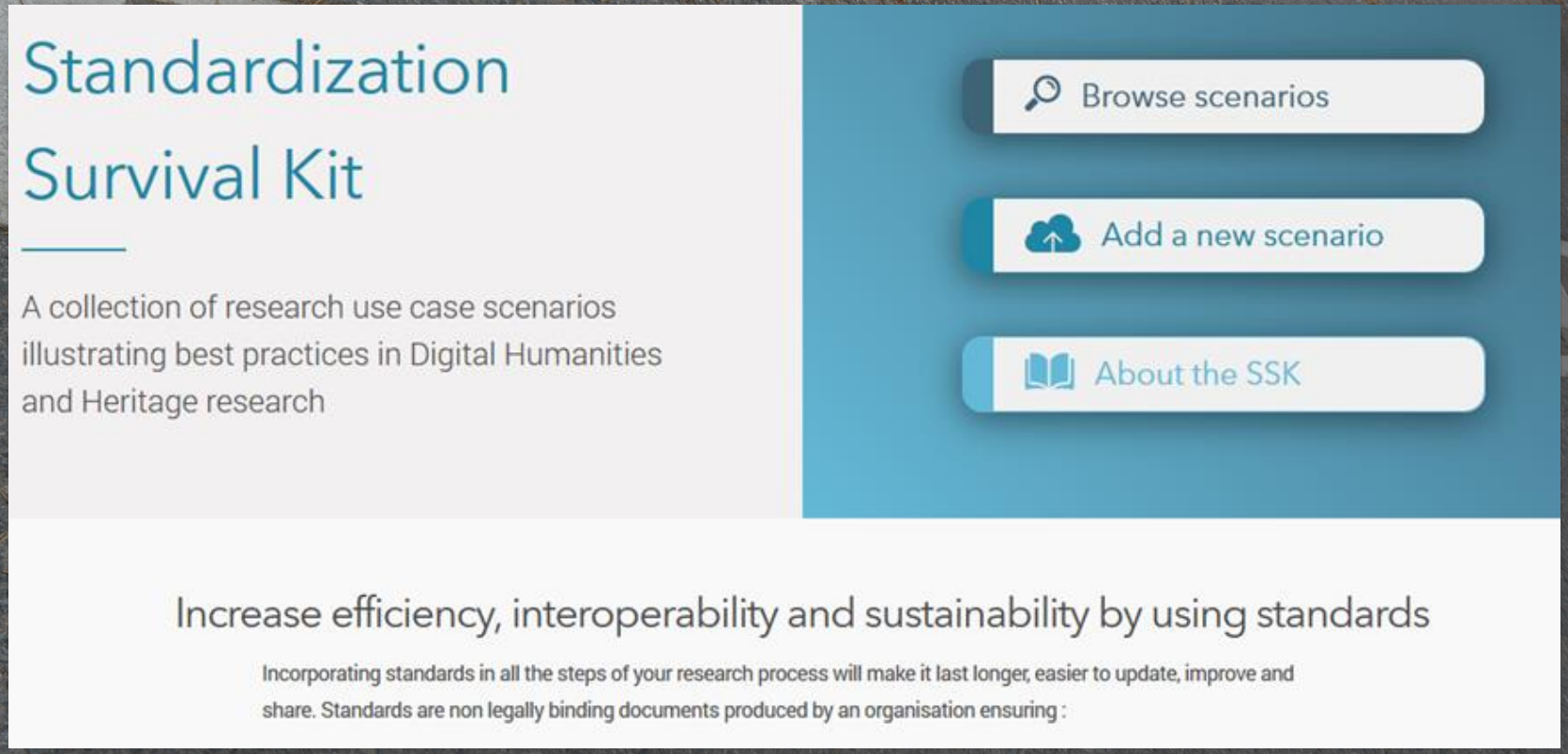
Wuttke 2018

I = Interoperable. Standards



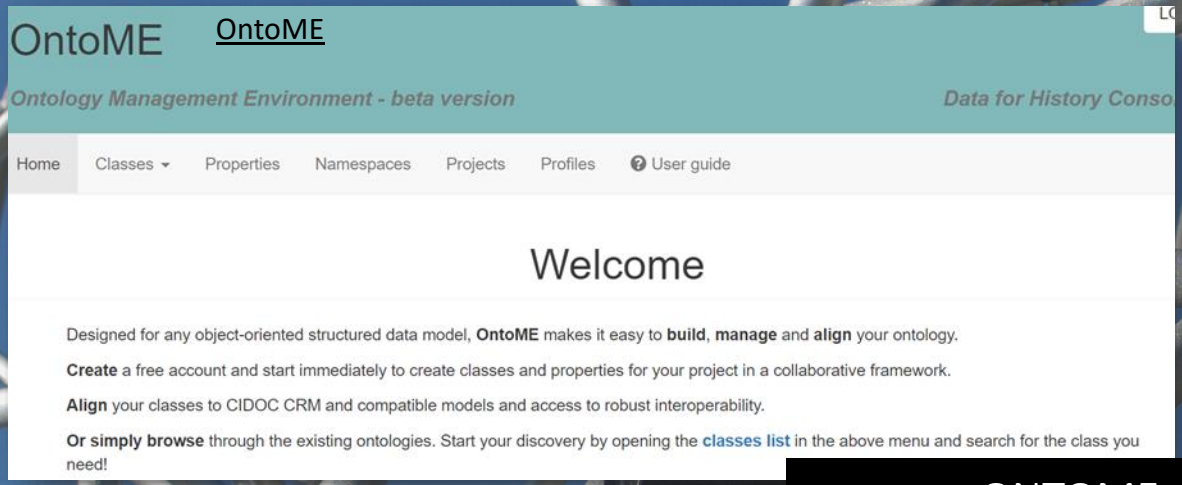
http://training.parthenos-project.eu/sample-page/intro-to-ri/interoperability/what-are-standards/

I = Interoperable. Standards



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

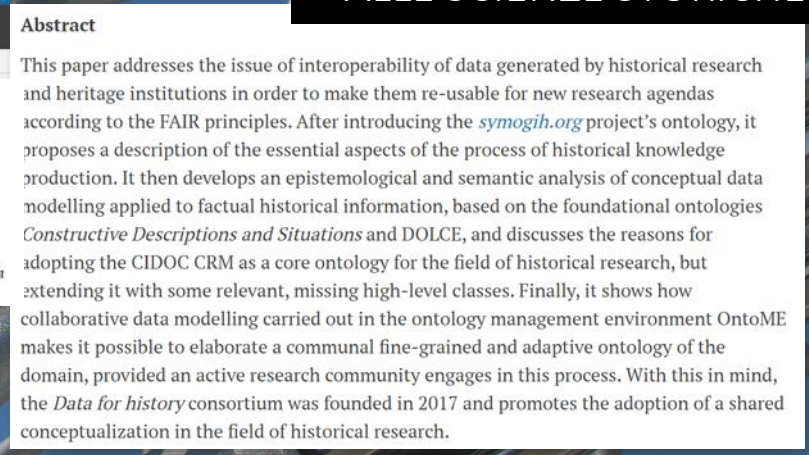
I = Inteoperable. Ontologies



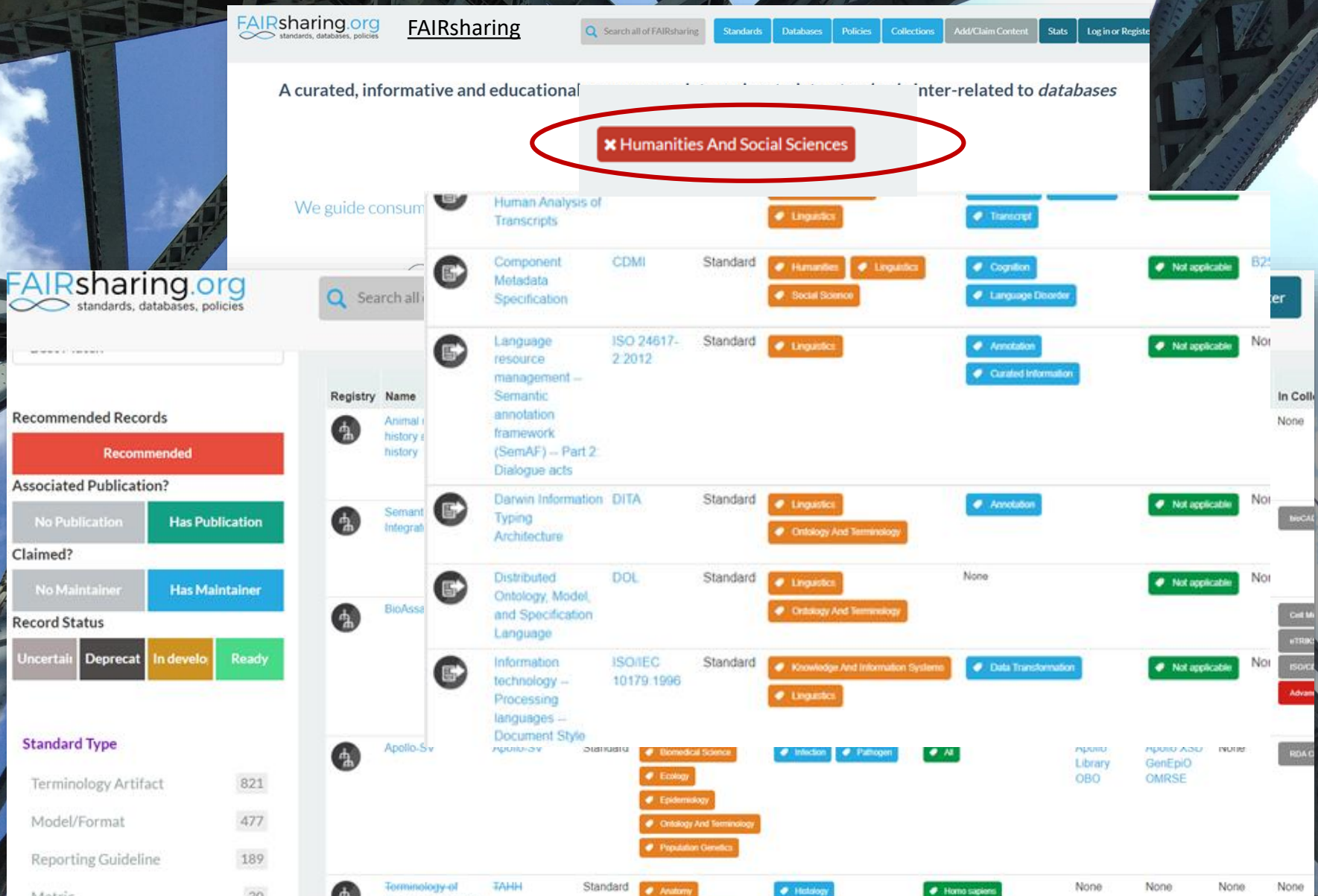
OntoME

ONTOME
ESEMPIO DI APPLICAZIONE
ALLE SCIENZE STORICHE

2021



[FAIRsharing. Per essere interoperabili]



REUSABLE

R = Reusable. Documentazione

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

cessda TRAINING

Project-level documentation

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

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

Data-level documentation

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

⊖ Quantitative data

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

Data-level documentation for quantitative data

For quantitative data document the following:

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

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

Reusable / licenses

Researchers are "prosumers" who produce and consume information and knowledge of other researchers. This section focuses on their role of producing knowledge and on ways to foster its diffusion by clear legal boundaries. In the humanities, texts are quite often closely intertwined with underlying data, which form an indispensable part of digital publications. Traditional conceptions of copyright like "All Rights Reserved" raise obvious problems for data sharing in the context of publications. In general: if machine readable data is to be shared, the recipient, in order to use the data effectively, will most likely need to make a local copy for analysis, or for merging with other data sets, or to extract some subset of the data. For this reason, our recommendation is to avoid applying any legal restrictions that do not embrace the principle of openness. The Reusability FAIR

RECOMMENDATIONS

ALLEA Report p. 26

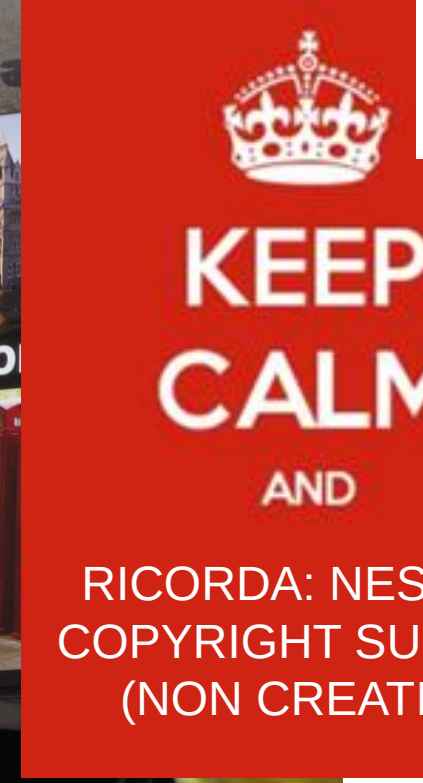
» Proper entitlement: first of all, identify who owns the data, i.e. whether you are entitled to license your work. You may only attribute a licence to a work of which you are the copyright holder. If there are co-authors, you have to agree with them on the licence. Furthermore, you are not allowed to license the works of the public domain. You should also be aware of whether there are any licensing requirements from the funding organisation or the data repository.

» Determine the necessary and sufficient level of access restrictions. Some data cannot be shared openly but can still be shared under certain restrictions while at the same time protecting the data. See for instance the CLARIN licensing framework for language data or the CESSDA access categories for qualitative and quantitative data (interviews, survey data etc).

» Use free and standardised licences: In order to benefit from the possibility of sharing data since the digital turn and to foster Open Science, use a licence as free as possible. The Open Knowledge Foundation and the Open Access Scholarly Publishers Association only

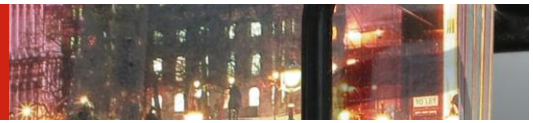
» For editors of journals and repositories managers: Avoid applying more restrictive licences like NC (non-commercial) or ND (no derivatives) just to be 'on the safe side'. NC can produce unintendedly limiting side-effects to potential re-users, as it is not quite clear whether the setting of a re-used work has commercial aspects or not. ND originates from the creative sector and is thought of as an instrument to protect the integrity of a work of art, such as a music composition. Many humanities scholars also want to protect their works from misuse and therefore are in favour of a ND licence. However, the risk of misuse through derivatives in the humanities is often quite low, so one has to balance this potential risk against the potentially unintended constraints imposed by ND, such as restrictions against reuse of publications in text and data mining procedures. Keep in mind that anybody deliberately deriving original content and thoughts by other scholars with misleading intention violates ethical scientific behaviour, whether a work is put under and ND licence or not.

R= Reusable. Licenze



RICORDA: NESSUN COPYRIGHT SUI DATI (NON CREATIVI)

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)

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

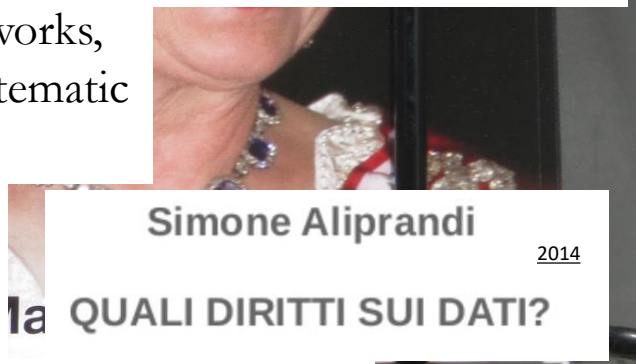
of 11 March 1996

on the legal protection of databases

NCIL OF THE EUROPEAN UNION,

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

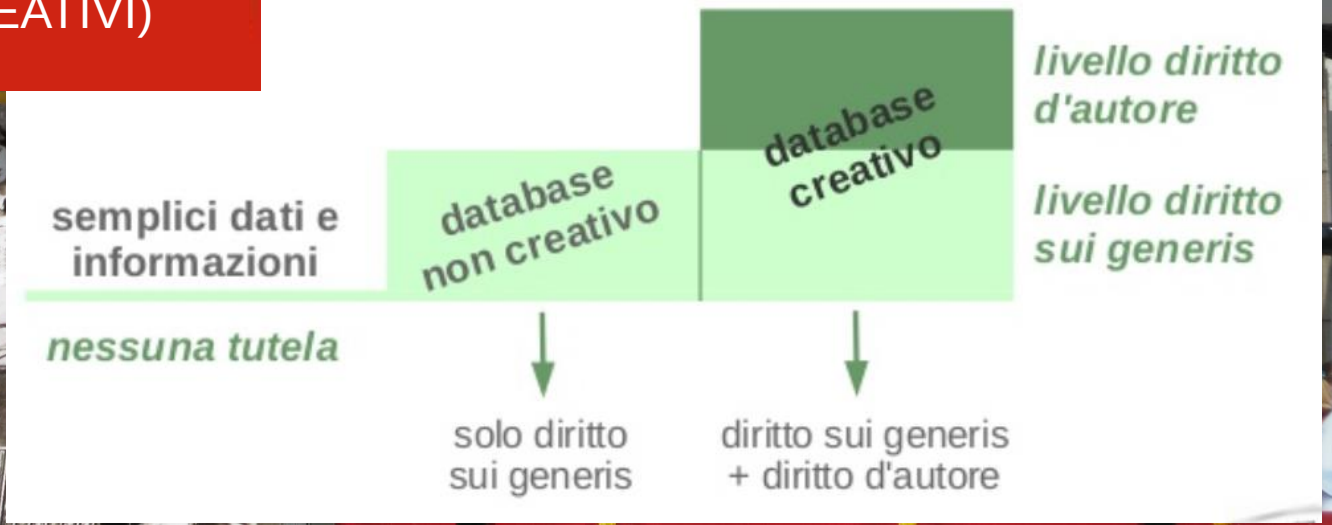
(1),



Simone Aliprandi

2014

QUALI DIRITTI SUI DATI?



[webinar]

sett. 2021

OS café

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

2020

2020

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 - Aspetti leg

OpenAIRE How do I know

SERVICES SUPPOR

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

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.

OpenAIRE How do I license

SERVICES SUPPORT

Guides for Researchers

How do I license my research data?

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

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?

OpenAIRE Can I use

SERVICES

Guides for Researchers

Can I reuse someone else's research data?

Learn more on how to reuse research data

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 +

Creative Com

CC Factsheet

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.

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

nmons 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)

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.

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

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

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

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

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

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

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

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

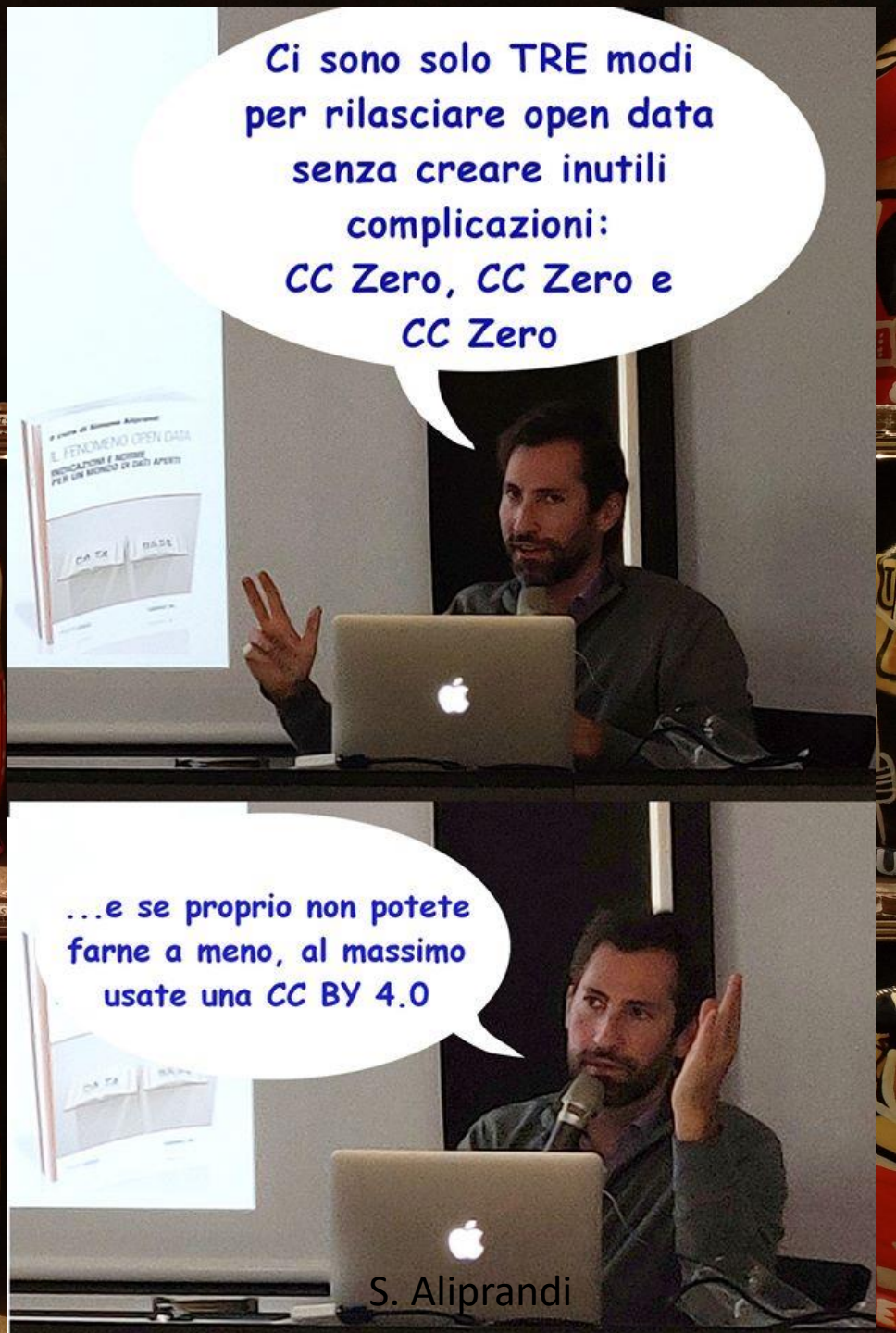
I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

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

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

...in sintesi...


Green Bay Packers
San Francisco 49ers
Ohio State Buckeyes
Auburn Tigers
Texas A&M
Texas Longhorns


Ci sono solo TRE modi per rilasciare open data senza creare inutili complicazioni: CC Zero, CC Zero e CC Zero
...e se proprio non potete farne a meno, al massimo usate una CC BY 4.0
S. Aliprandi

…e non dimenticate Horizon Europe

Proposal template Part B: technical description

1.2 **Methodology** *[e.g. 15 pages]*

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

ANNEX 5

SPECIFIC RULES

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

Open science: research data management

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

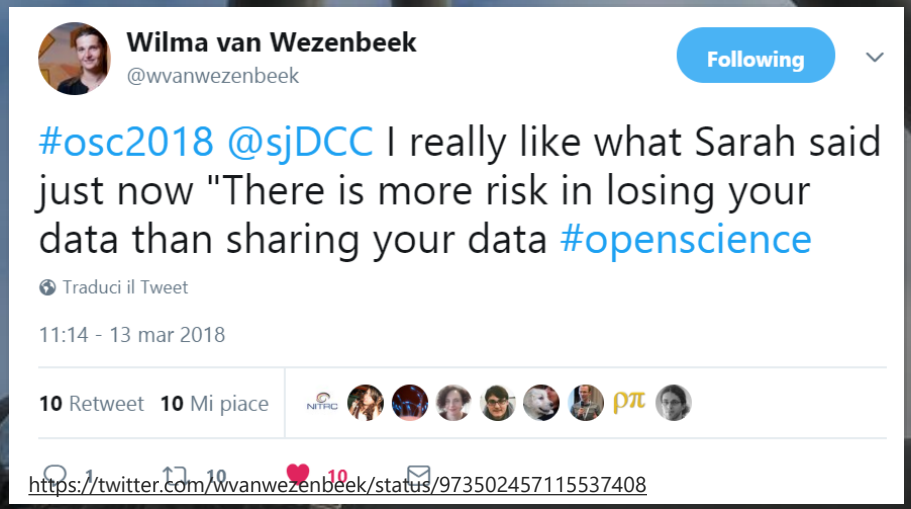
LA GESTIONE DEI DATI FAIR RIENTRA

- FRA LE PRATICHE RACCOMANDATE
- FRA LE PRATICHE OBBLIGATORIE

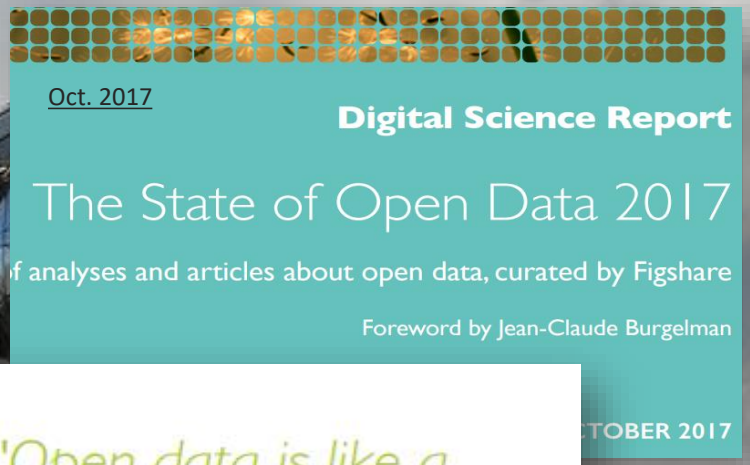
DOVETE DECLIANRLE ENTRAMBE IN MAX UNA PAGINA NELLA SEZIONE EXCELLENCE DELLA PROPOSTA


Dati aperti

Perché i dati aperti?



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



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

Open data saves lives

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

… «as open as possible»…

As open as possible, as closed as necessary

as a starting point. Active dissemination around data, once the data have been made FAIR, needs to become a key research data management best practice.

The rapid uptake of the FAIR data principles as part of a wider movement towards Open Science is changing how scientists and scholars collect, curate, preserve and share their research data. In particular the principle of "as open as possible, as closed as necessary" is aimed at guiding researchers in their efforts to strike a balance between sharing data and the need to account for issues around sensitive data/legal aspects. Overall this shift has also brought a focus on maximising data use and potential not only for future research but also in other areas (e.g. private sector) and for other categories of potential users (e.g. citizen scientists).

RECOMMENDATIONS

ALLEA Report, p. 32

» Humanities scholars are encouraged to take advantage of the frameworks, networks and resources that facilitate the discoverability and wider reuse of research:

- Domain registries, portals, harvesters, e.g. Re3data and FAIRsharing.org
- Platforms e.g. Europeana, AGATE
- Researcher profiles e.g. ORCID

» Share online your data and all supporting materials such as presentations, posters, blogs, data papers etc. and consider using social media for wider outreach, cite using persistent identifiers.

» Consider publishing a data paper either as a preprint or via a dedicated journal for data papers. An emerging practice supporting the FAIR principles, publishing data papers about data sets increases findability as well as reuse, as these provide the key information about specific datasets. e.g. Journal of Open Humanities Data, Research Data Journal for the Humanities and Social Sciences.

» Talk about your research outside academia, consider diverse audiences, such as journalists, policy makers, private companies or citizen scientists as Open Science is ultimately promoting the involvement of a wider audience in scientific research.

» Consider non-traditional channels and formats to present your data: infographics or interactive data visualisations, online exhibition or digital tours, websites or apps, executive summary/lay summary, also consider a wider use of national languages.

» Promote/prepare your datasets for use in class (schools or HEI) or for Hackathons (e.g. Coding Da Vinci).

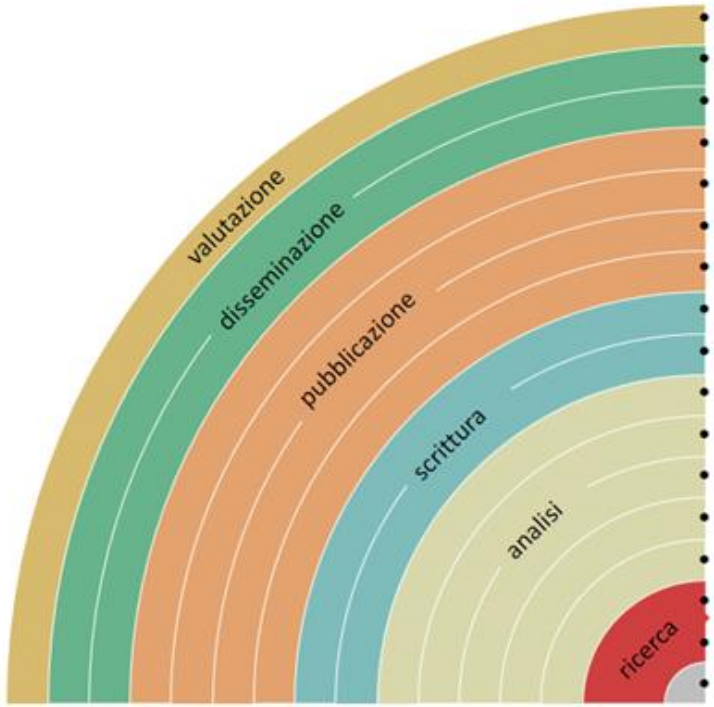
» As an institution, actively also showcase and provide institutional channels that researchers can leverage, and reward data dissemination.

» Encourage and support pedagogic approaches which include student production and curation of open research data, and use of existing open datasets as open educational resources (OER).

Open science: perché solo i dati?

OS rainbow

Come rendere Open ogni passo della ricerca...



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

Bianca Kramer & Jeroen Bosman https://101innovations.wordpress.com DOI: 10.5281/zenodo.1147025
Traduzione: Elena Giglia DOI: 10.5281/zenodo.1195648

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

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