DATI FAIR, DATI OPEN E DMP

CNR, maggio 2021

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

Impareremo in questo modulo

1. Cosa significa FAIR BY DESIGN nella ricerca quotidiana e in HorizonEurope
2. come scrivere un DMP, che diventerà obbligatorio [supporto!]

MESSAGGI CHIAVE

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

[i tre passi fondamentali]

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

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

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

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

…e non dimenticate Horizon Europe

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

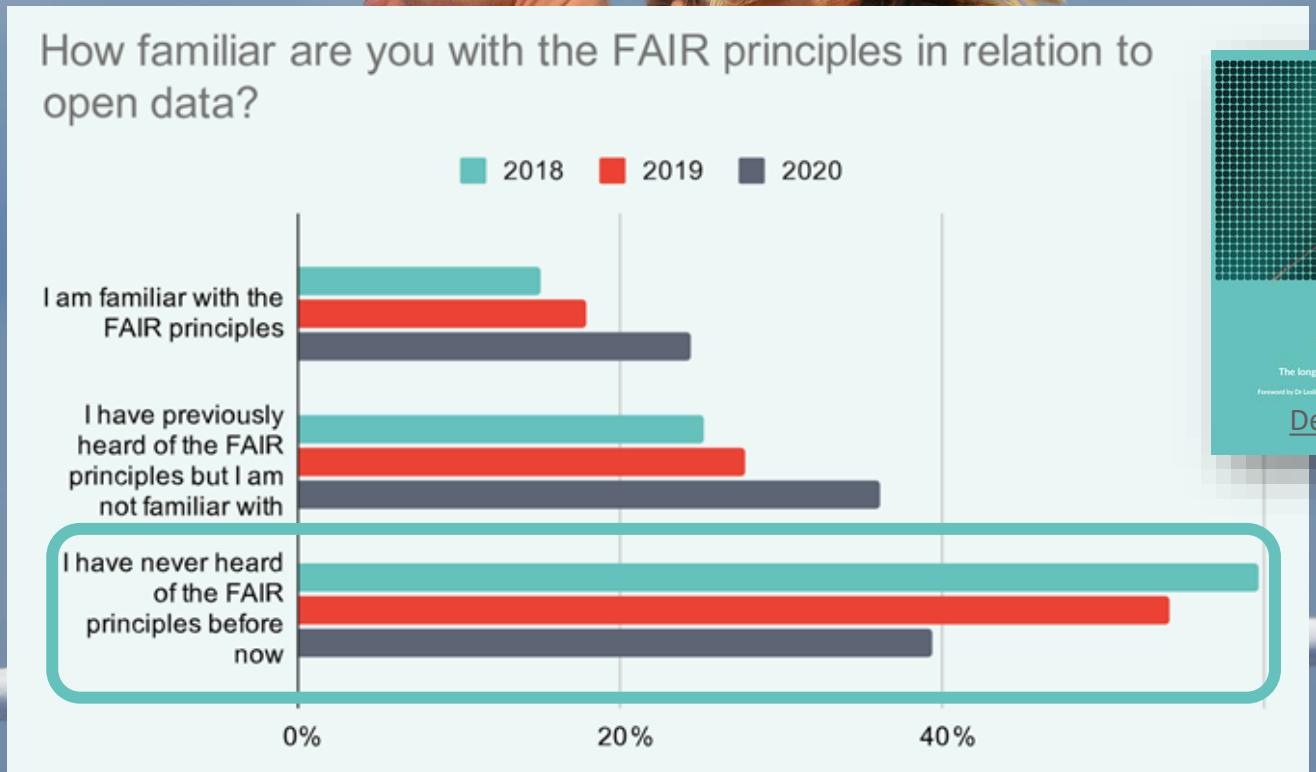
- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence 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 È PREVISTA

- NELLA PROPOSTA [VIENE QUINDI VALUTATA] COME SINTESI DI UNA PAGINA NELLA SEZIONE SCIENTIFIC EXCELLENCE
- SE IL PROGETTO VIENE APPROVATO, IL DMP VA PRESENTATO COME DELIVERABLE AL MESE 6 (ART. 17 ANNEX 5)

2. DATI FAIR BY DESIGN

Dati FAIR???


How familiar are you with the FAIR principles in relation to open data?
2018
2019
2020
I am familiar with the FAIR principles
I have previously heard of the FAIR principles but I am not familiar with
I have never heard of the FAIR principles before now
0%
20%
40%


Digital Science Report
The State of Open Data 2020
The longest-running longitudinal survey and analysis on open data
Dec.2020
December 2020

…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 E PROTOCOLLI APERTI

INTEROPERABLE

- STANDARDS
- ONTOLOGIE

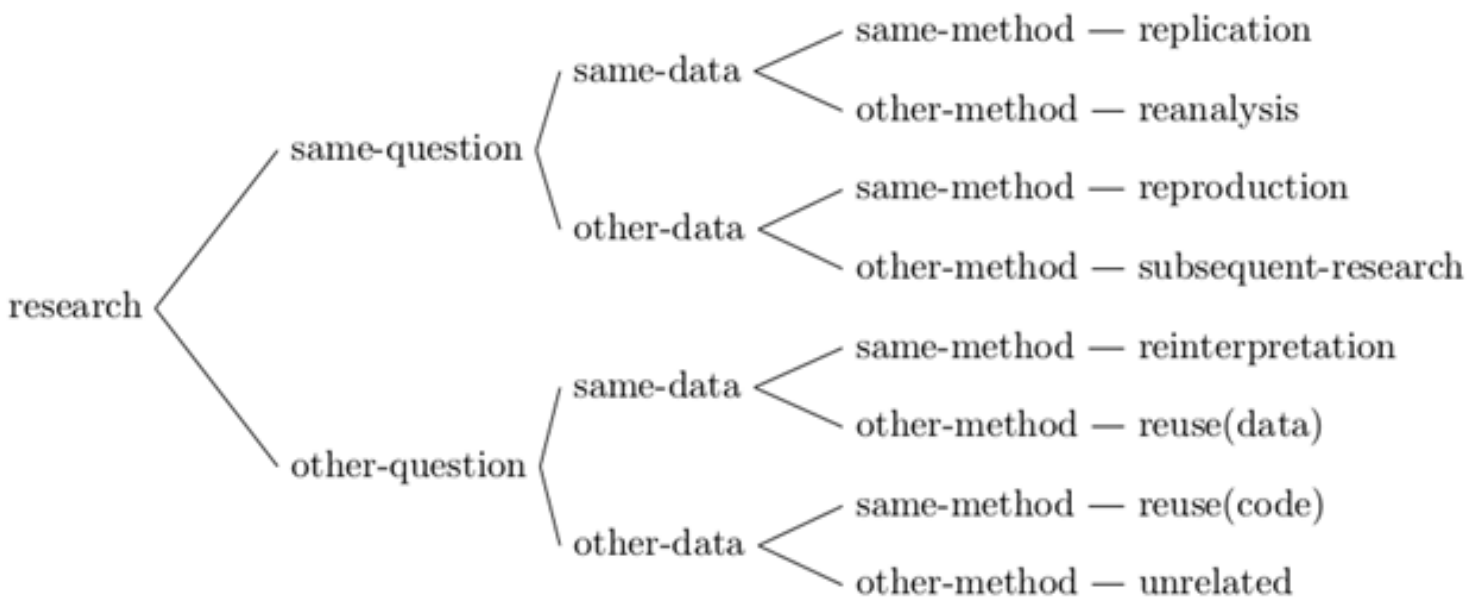
REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE

IL TUTTO, LEGGIBILE DALLE MACCHINE

Focus

ASSICURARE IL RIUSO



Van de Sandt et al. 2019

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

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

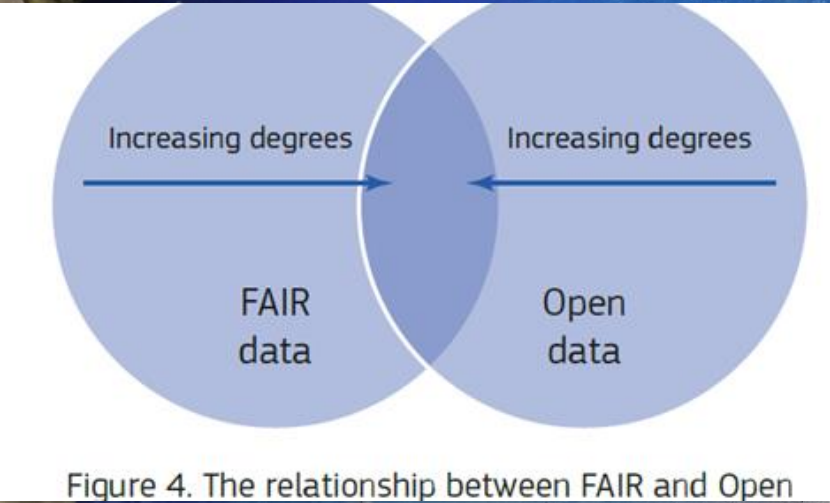


Figure 4. The relationship between FAIR and Open

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

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

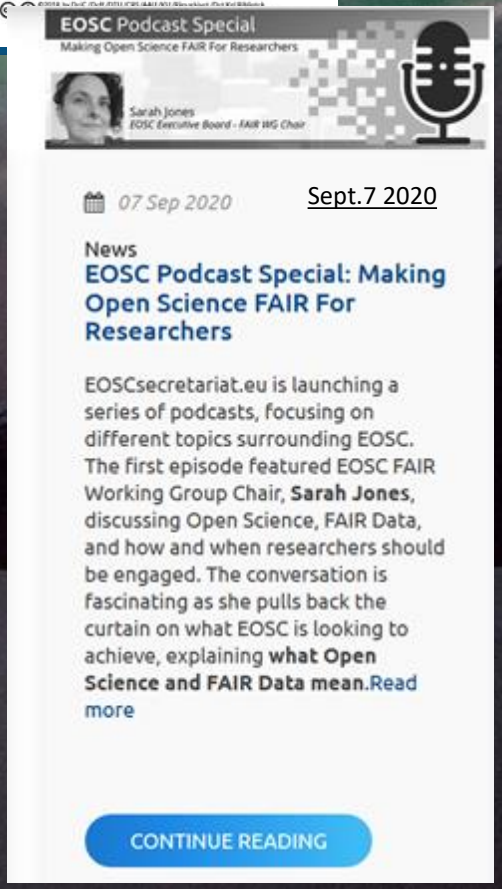
It is also a provider of services for minting persistent identifiers of many di of the data being identified):

- 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

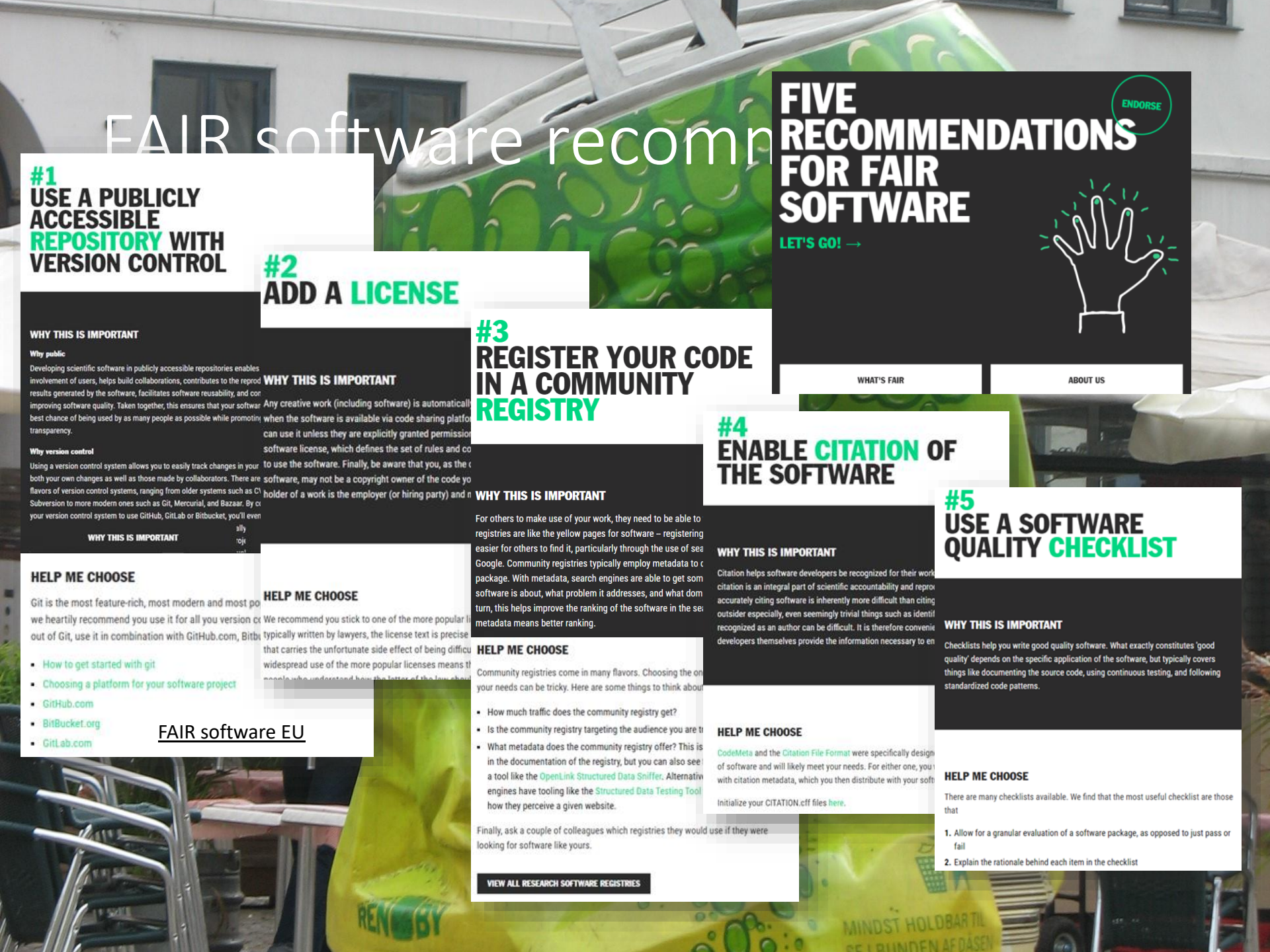
I miti su FAIR



Sept.7 2020

FAIR myths

FAIR software recomm



Nov. 20, 2018

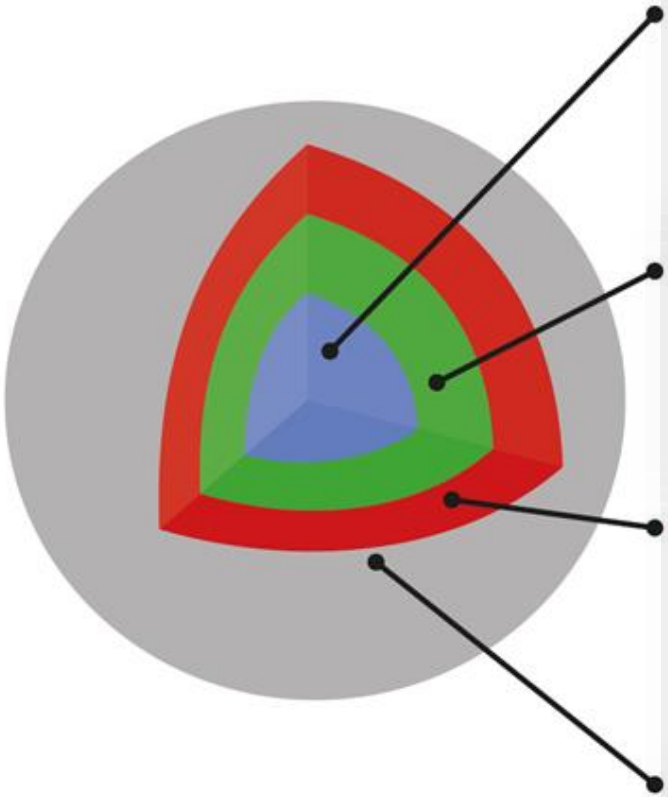
European Commission

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

TURNING FAIR INTO

Define	Implement			Embed and sustain	
Concepts for FAIR implementation	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

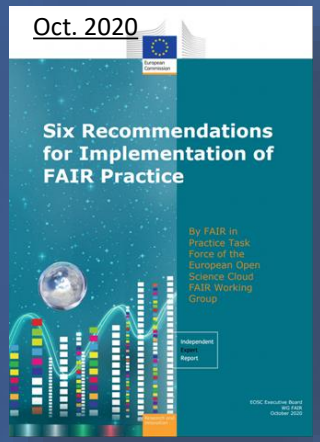
Contextual documentation

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

6 raccomandazioni

Oct. 2020



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

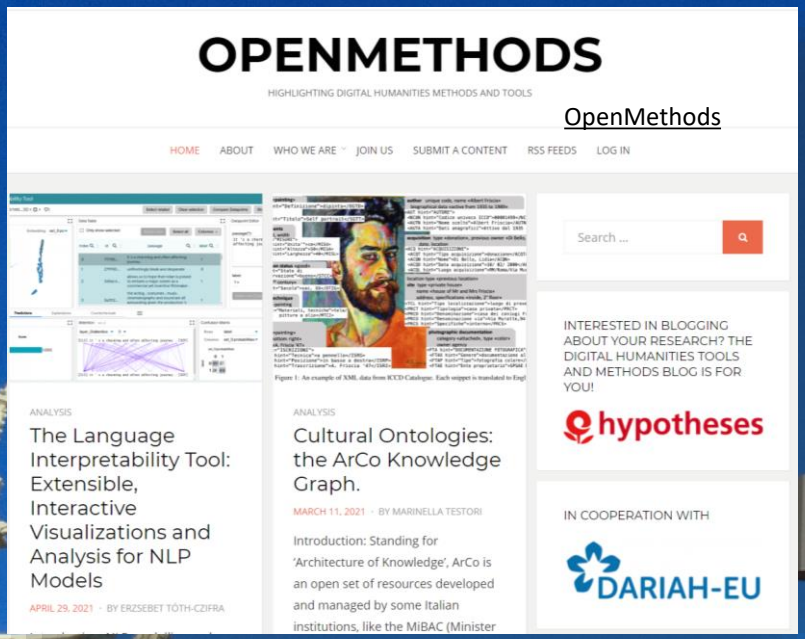
Recommendation 1: Fund awareness-raising, training, education and community-specific support.

Stakeholders: EOSC, Research funders, Institutions

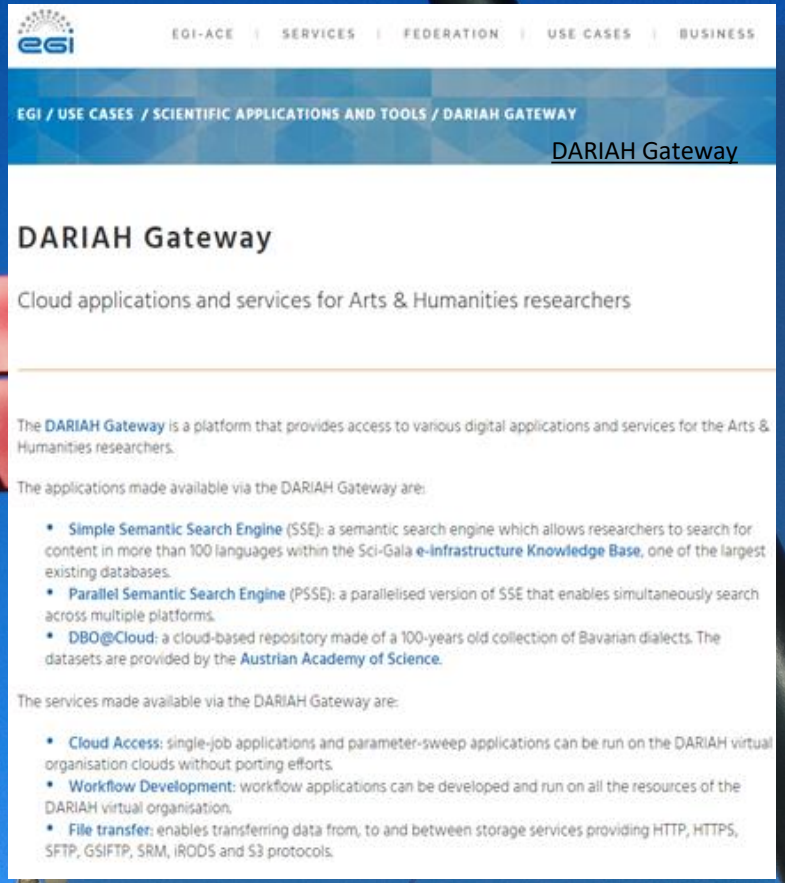
Rationale: Community-specific actions are needed because arguments and solutions which works for one community might not be the key drivers for another. Raising awareness is needed at all levels – from individual researchers through heads of institutions to policymakers – but in order to be meaningful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

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

…e le scienze umane?



OpenMethods



DARIAH Gateway

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

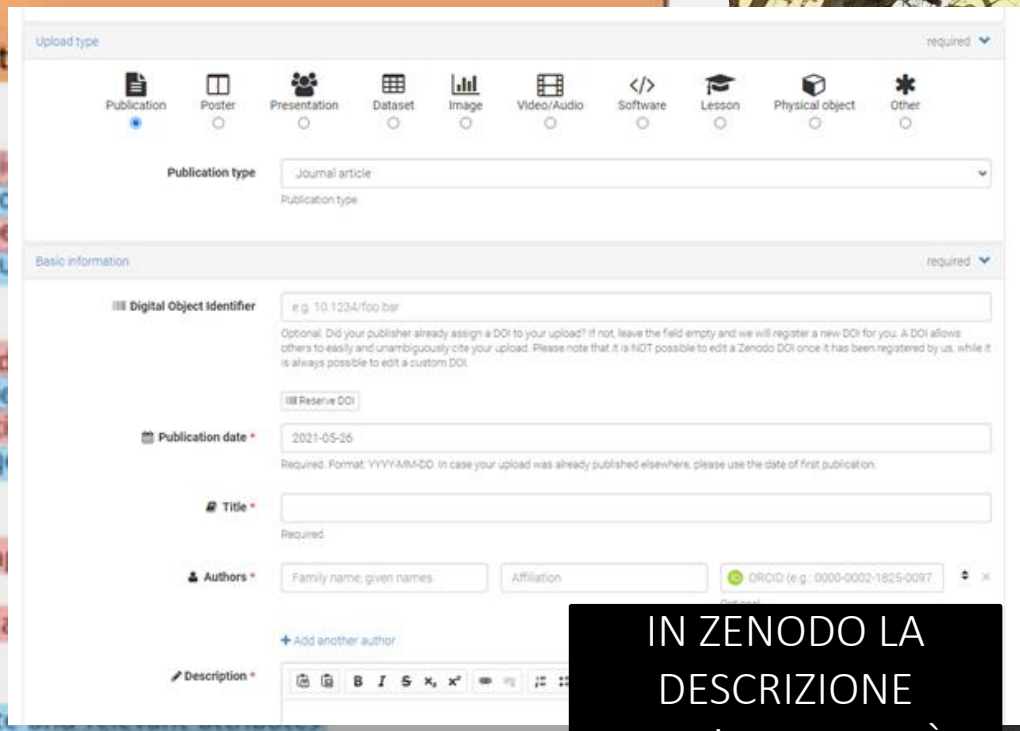
To be Findable:
F1. (meta)data are assigned a globally unique and persistent
F2. data are described with rich metadata (defined by R1 belo
F3. metadata clearly and explicitly include the identifier of the
F4. (meta)data are registered or indexed in a searchable resou

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 accurate
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

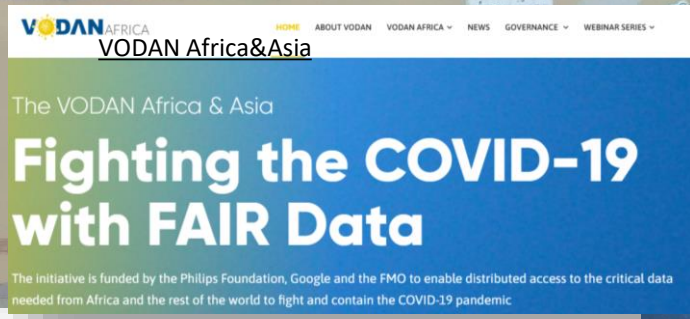
Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, doi:10.1038/sdata.2016.18

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 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.

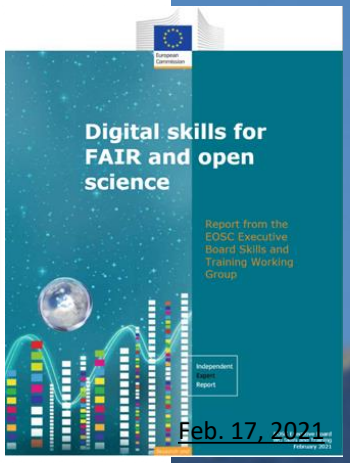
Mirjam van Reisen (IN Coordinator of **IN-Africa** and one of the coordinators of the **Ambassadors IN**) said:

> 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

SARÀ PRESENTATO A OAI12 (SETT.)

Digital skills for FAIR and open



INTORNO A FAIR/EOSC GIRANO DIVERSE COMPETENZE

Digital scienc

Policymakers & funders	Universities & research organisations	Competence centres	EOSC Association	EOSC projects
Understand the importance of addressing gaps in provision of digital skills for FAIR and open science				
Identify actors/roles that need to be supported in the EOSC ecosystem	Consider diversity of employment and training needs of staff	Identify knowledge to be taught and skills to be learned	Understand range of roles that initiatives must address	Ensure projects consider appropriate roles for both their staff and users
Consider how to support competence centres and encourage	Consider how to establish competence centres	Learn about governance, business models and alignment	Consider how to encourage and support compet-	Ensure projects consider FAIR and open science training
Recognise the need for investment in hubs for training resources	Know how to provide and access learning and training resources	Learn how to maximise interoperability, FAIRness and usage of learning and training resources	Facilitate interoperability and discovery of learning and training resources	Identify and engage key stakeholders, and build learning and training catalogues according to their goals and actions
Identify actions to advance national policies on FAIR and open science	Consider how to encourage and/or implement national approaches	Understand the importance of aligning with national programmes and policies	Identify ways to increase international alignment	Not relevant

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

Professionalizzare i data stewards

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

nationaal programma open science

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

Dutch roadmap towards national implementation of FAIR data stewardship

Feb. 1, 2021

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

"Invest 5% of research funds in ensuring data are reusable. It is irresponsible to support research but not data stewardship", said Barend Mons recently in a Nature article[1]. "Students in PhD programmes spend up to 80% of their time on 'data munging', fixing formatting and minor mistakes to make data suitable for analysis — wasting time and talent. With 400 such students,

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

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

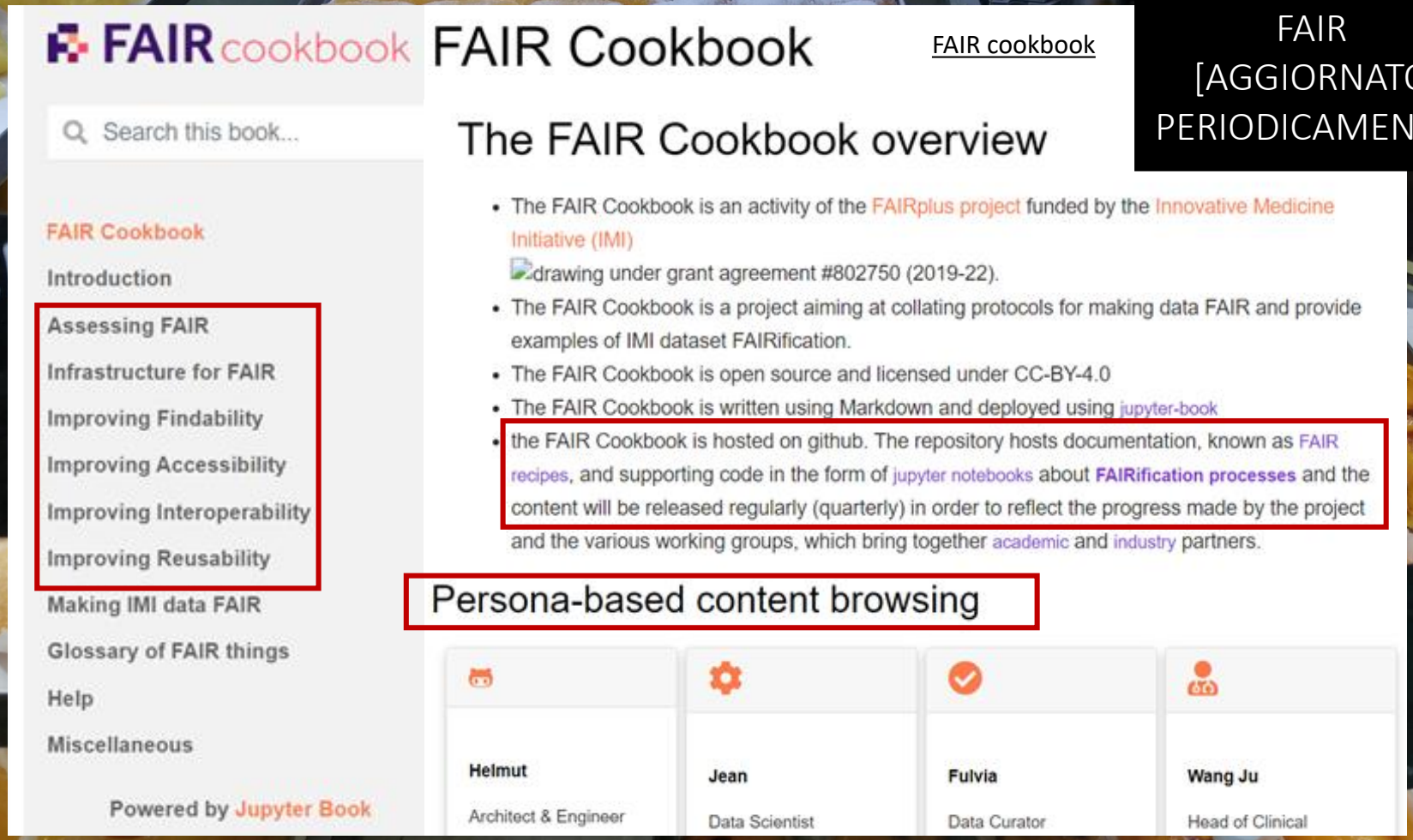
[supporto]

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

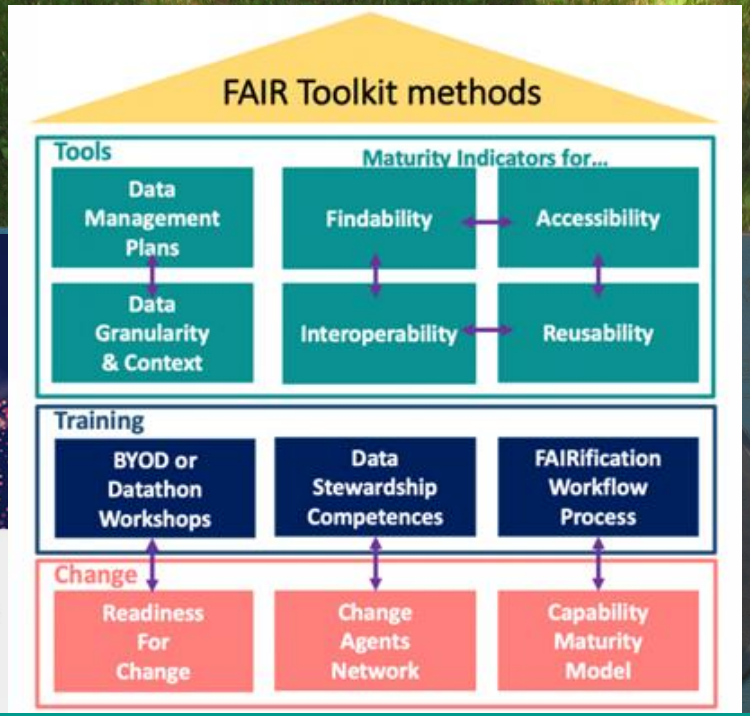
ICDI STA CREANDO COMPETENCE CENTER

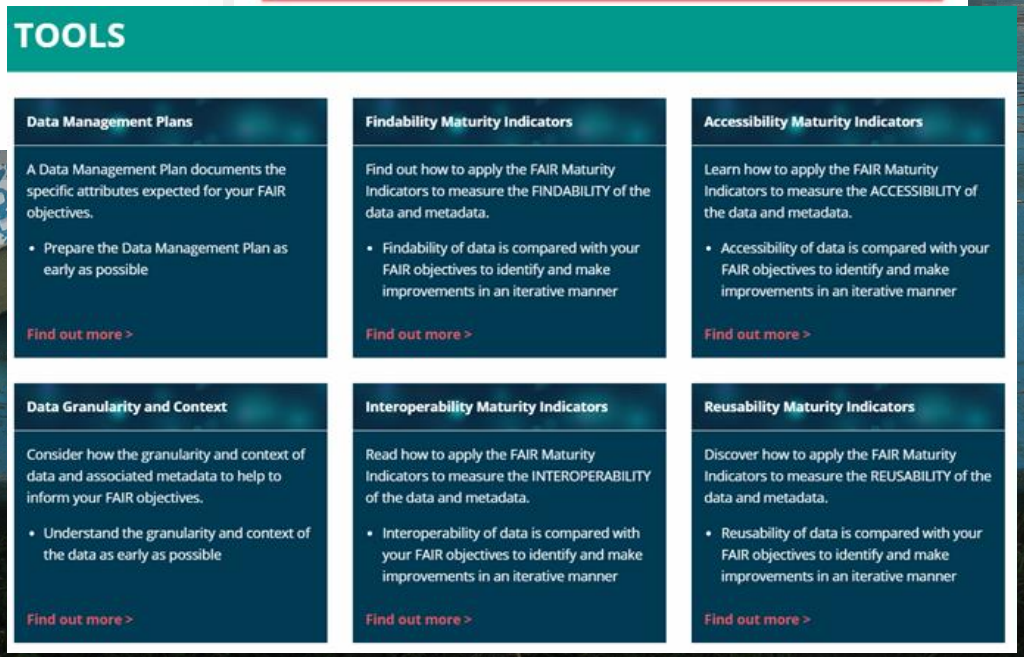
FAIR cookbook

RICETTE PER RENDERE I DATI FAIR [AGGIORNATO PERIODICAMENTE]



FAIR toolkit





FAIR TOOLKIT CON STRUMENTI, CASI D'USO, FORMAZIONE

FAIRifi

Jan 2020

Volume 2, Issue 1-2
Winter-Spring 2020

DATA Intelligence

January 01 2020

A Generic Workflow for the Data FAIRification Process

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

> Author and Article Information

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

https://doi.org/10.1162/dint_a_00028

FAIR è graduale

FAIR = CONTINUUM
«AS FAIR AS POSSIBLE»

Inclusiveness: consider FAIR as a process

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

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

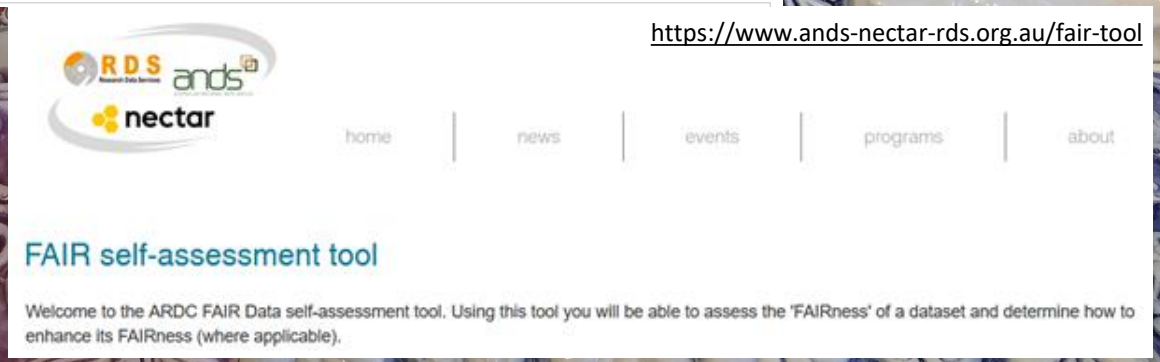
Accessible

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

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



UTILISSIMO PER PORSI LE DOMANDE GIUSTE… MA SOGGETTIVO…

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.

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

Selected datasets should be assigned a globally unique and persistent identifier (PID) so they can be located unambiguously by humans or machines on the web. Persistent identifiers are maintained and governed so that they remain stable and direct the users to the same relevant object consistently over time. Examples of PIDs include Digital Object Identifier (DOI), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK).

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

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

Want to know more?

F

FINDABLE

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

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

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

Yes
No

To what degree do you intend to comply with this?

Very likely 5 4 3 2 1 Very unlikely

FAIR maturity evaluator

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

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

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

Sept. 20, 2019

- 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.

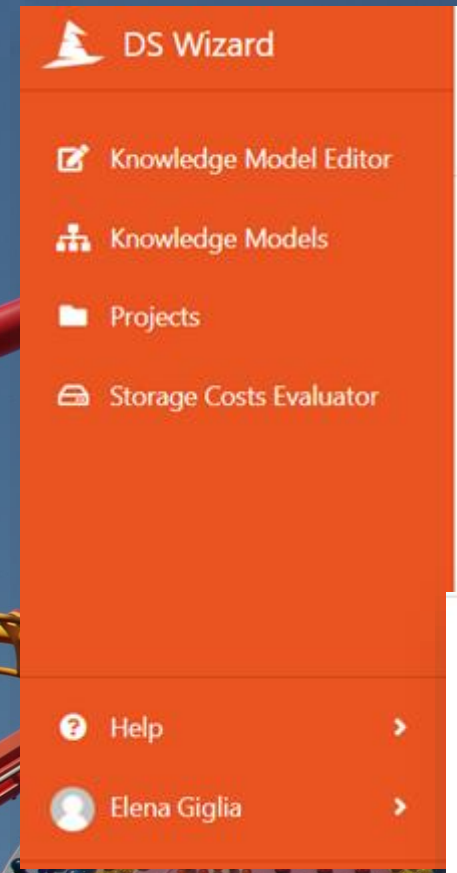
FAIR METRICS GEN2 - IDENTIFIER PERSISTENCE

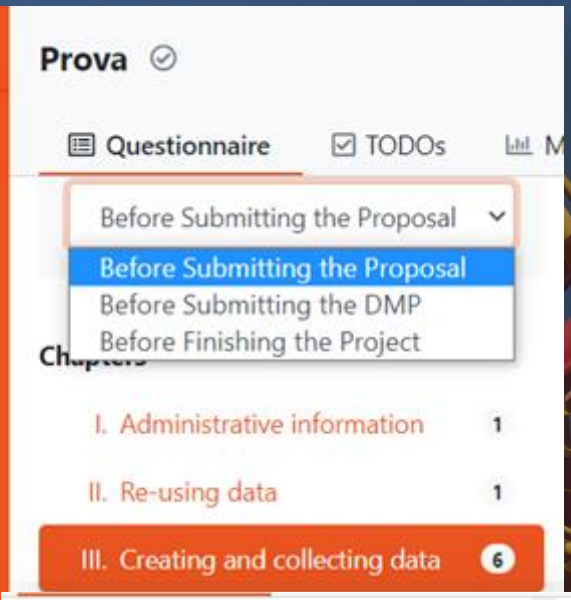
Status: Failure
Principle tested: F1
Description: Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema). For URLs that don't follow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

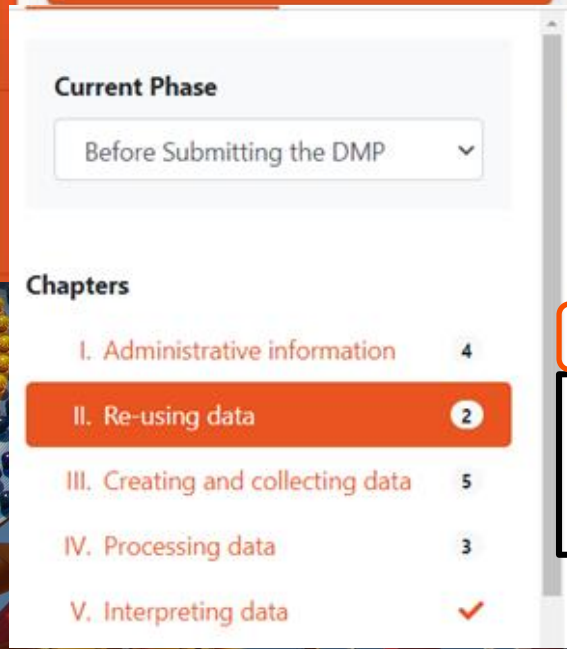
Test results

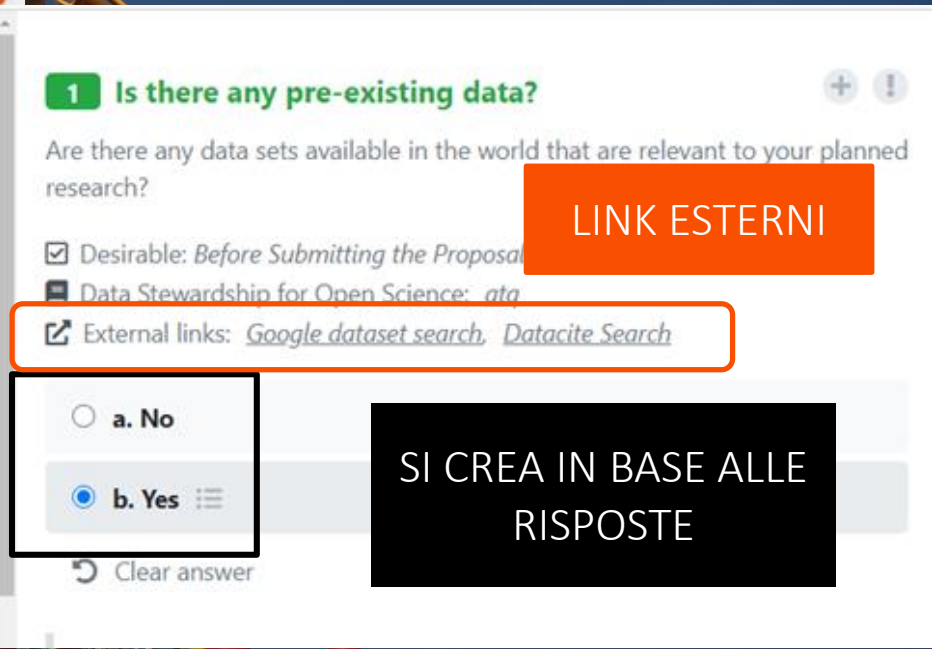
INFO: The metadata GUID appears to be a URL. Testing known URL persistence schemas (purl, oclc, fdlp, purlz, w3id, ark).
FAILURE: The metadata GUID does not conform with any known permanent-URL system.

FAIR Wizard



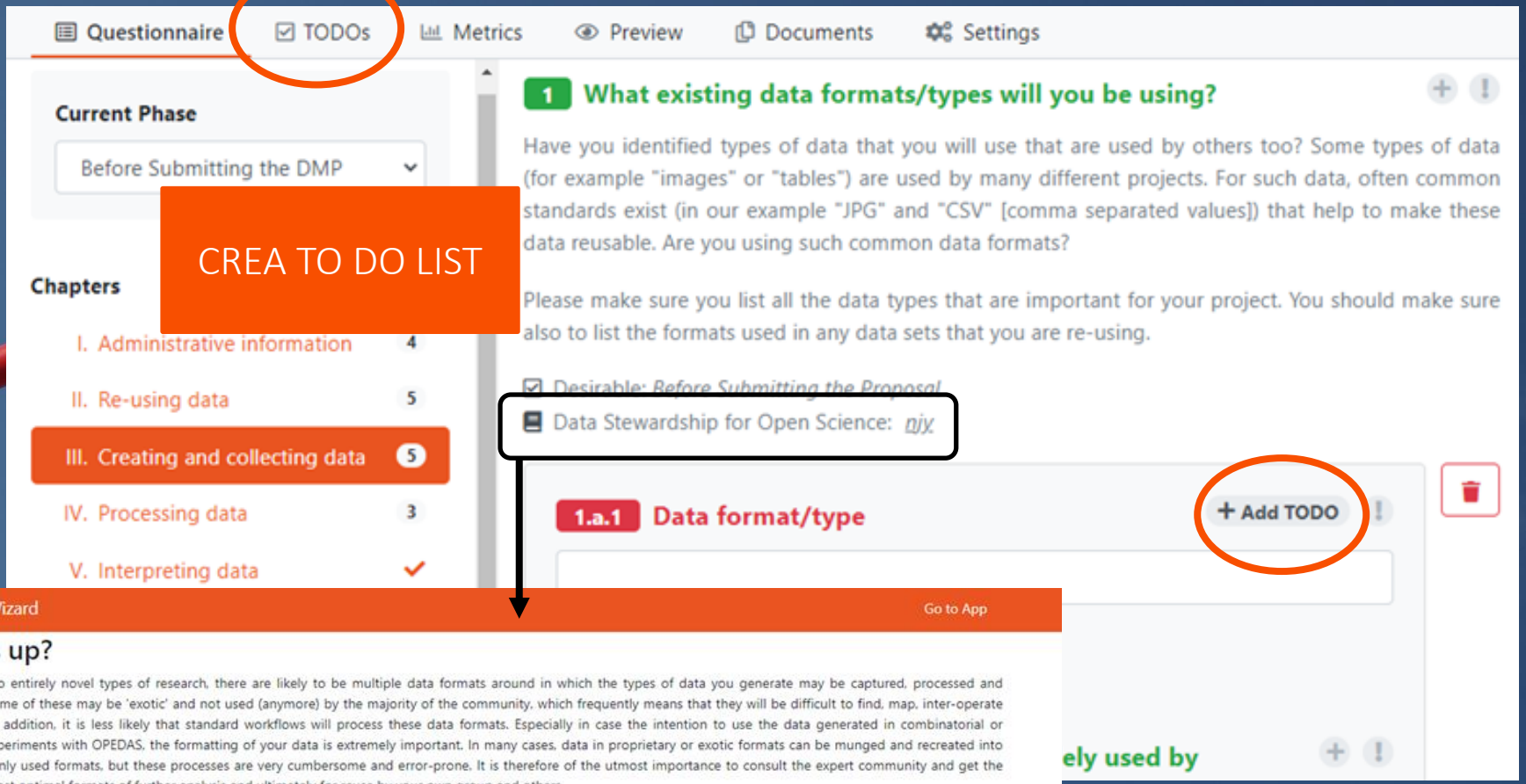






LINK ESTERNI

SI CREA IN BASE ALLE RISPOSTE



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 per enti di ricer

Oct. 2020

DO I-PASS FOR FAIR?

Policy: Does your organization (institute / university (of applied sciences)) have a FAIR research data policy?				
1. **Are the FAIR principles explicitly mentioned in the data (or research data) policy of your organization?**	The FAIR principles are not explicitly mentioned, but there is a reference to sustainable and long term storage.	The F and A principle are explicitly mentioned.	All FAIR principles are explicitly mentioned (or there is an explicit reference to the overarching Open Science concept).	☐ Beginner ☐ Intermediate ☐ Advanced
2. **Is the general research data policy translated into faculty/institutes specific d policies?**	Less than 20% of the faculties or sub-	20-80% of the faculties or sub-units have a specific policy.	More than 80% of the faculties or sub-units have	☐ Beginner ☐ Intermediate
3. **Is the general research dat policy supported by dedica service units (e.g. DCC) ?**				
4. **Which means and channels used to communicate the research data policy and services to the researchers**				

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

Sssssht…forse ALCUNI editori si stanno muovendo….

An Academic Publishers' GO FAIR Implementation Network (APIN)

Jan Velterop[a,*] and Erik Schultes[b,*]

Jan.6, 2021

Abstract

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

CREAZIONE DI ARTICOLI SECONDO I PRINCIPI FAIR

FAIR data nelle scienze umane

Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | **February 2020**

February 2020

RECOMMENDATIONS

» Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.

» Familiarise yourself with the FAIR Data Principles before you start collecting data and building corpora e.g. FORCE11: the FAIR Data Principles, GO-FAIR: FAIR Data Principles and discuss with colleagues and experts to build a better understanding.

» Digitally document all your research and data collection work -- at the beginning of a project it is difficult to judge which information of the research process will be important and valuable later on.

» Use well-established tools to facilitate your research work, as many of them allow data sharing e.g. MIT Libraries Digital Humanities: Tools and Resource Recommendations.

» Browse humanities datasets and consider whether your own assets could be published in a similar fashion (e.g. Humanities Commons, UK Data Archive, ARCHE re3data.org filtered for humanities).

» When you start producing data, keep this maxim of Open Science in mind: data should be 'as open as possible and as closed as necessary'.

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

SERVE CAMBIO DI PARADIGMA
CON INCENTIVI, SUPPORTO E RICONOSCIMENTO

RECOMMENDATIONS

» Clarify all legal issues at the beginning of your research project and include the findings of this process in the data management plan.

» Use checklists adequate to your research topic/discipline.

» Check the resources indicated by DARIAH, CLARIN. (see further reading).

» In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).

» Ask for consent to share anonymised data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities. (See for example the guide on informed consent in the CESSDA data management expert guide).

» Avoid collection of (sensitive and non-sensitive) personal data when possible.

» Get legal support (IPR, copyright, patents, trademarks etc.) from your home institution. If there is no dedicated office for this purpose, try to get information from your university library, as its staff are often confronted with such issues.

» If you need permission from the copyright holder in order to use sources like images for your publication, try to get one that covers both printed and digital copies.

» Finally, check the recommendations in the section on Licences, that are closely related to this section.

RECOMMENDATIONS

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

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

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

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

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

RECOMMENDATIONS

» If applicable, determine if the body funding your research has particular requirements for a DMP, or offers a template for framing your plan. If there is no required template, choose an existing appropriate one (e.g. via DMPOnline).

» Devise a DMP prior to collecting data. Define and plan for your data: all research projects deal with data. If your project includes the analysis of text corpora, for example, then the corpora themselves are data, and you should make sure they are clearly described, documented, and managed according to the FAIR principles so your research is reusable by others.

» Plan documentation of metadata: In order for your data to be comprehensible in the future and/or reusable by others, they will need descriptive metadata created according to a common schema to understand the 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).

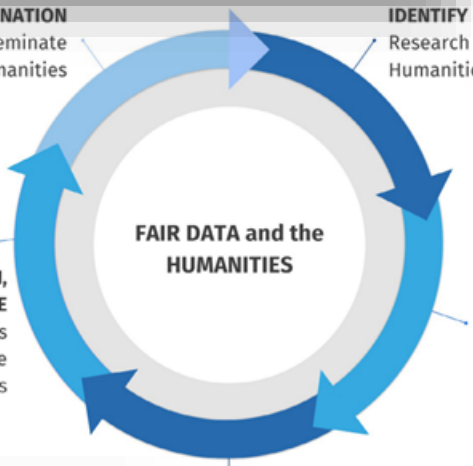
» Consider the right questions to be answered in your DMP that can account for discipline-specific requirements. The DMP templates suggested by funders are quite high level and provide generic guidance for file naming or versioning conventions, database structuring, and can be a good start. Tools like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

2020

RECOMMENDATIONS

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

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

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

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

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

» Metadata should be machine-readable.

FAIR data

Digital Humanities Methods and Tools 2019

An extension of the OpenMethods metablog for scholars who wish to share the methodological advancements of their research.

About

Credits

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

Search ...

"Here be dragons"

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

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

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

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

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

FAIR per collezioni di musei e bi

2018

code{4}lib JOURNAL

Mission Editorial Committee

Issue 40, 2018-05-04

FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections

Many heritage institutions would like their collections to be open and reusable but fail to achieve that situation because of organizational, legal and technological barriers. A set of guidelines and best practices is proposed to facilitate the process of making heritage collections reusable. These guidelines are based on the FAIR Principles for scholarly output (FAIR data principles [2014]), taking into account a number of accessible, interoperable and reusable. Archive and Museum Collections focus records. Clarifications and examples of recommendations for the assessment o principles.

by Lukas Koster, Saskia Woutersen-Windhouv

Access protocols

For access to both digital objects and digital metadata, universal access protocols must be made available on several levels. The first level comprises server level access protocols for copying files between servers and workstations (FTP/SFTP), direct server command line access (SSH) and downloading files and data through web browsers (HTTP). The second level consists of application level access protocols for accessing metadata, such as application programming interfaces (APIs), harvesting (OAI-PMH) and retrieving metadata (Z39.50, SRU) and linked data (SPARQL). Finally, there are the general formats and frameworks, other than the domain specific data formats (see *Standard metadata formats* below), such as XML, RDF, JSON, JSON-LD, which are used to represent the domain specific formats. For instance, MARC can be represented as MARC-XML, Dublin Core in XML, RDF, etc. In order to support reuse of metadata and digital objects as much as possible, options for selecting and downloading in bulk must be presented as well as options for retrieving and downloading records and objects individually.

Standard metadata formats

It must be possible to provide metadata records in various standard formats, other than the format the metadata is stored in. Examples are MARC, Dublin Core, MODS, METS. The stored metadata must be sufficient to describe the specific object type and support the intended reuse. In the linked data context vocabularies or ontologies in RDF are used. For this various standards are available, such as EDM (Europeana Data Model), CIDOC-CRM (for cultural heritage), BIBFRAME (the prospective linked data successor to MARC), as well as well-known formats like Dublin Core and schema.org.

Copyrights and licenses

In order to avoid the recurring necessity of requesting permission for reuse in every individual case, it should be instantly clear what actions are permitted with the objects and metadata. This is possible by including a license provided by the rights holder. The most used open content license is the Creative Commons (CC) license.[6] The CC license is a worldwide, irrevocable, non-exclusive license for the duration of copyright and similar rights, and sui generis database rights. The most used version is the 'CC-BY' (Creative Commons. Attribution 4.0 International [2018]). With a CC-BY, the rights holder gives permission that metadata or an object may be distributed, copied and adapted free of royalties under the condition of attribution. In addition

OPERAS

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

5 WORKSHOP NAZIONALI

GO FAIR FAIR Principles Implementation Networks News Even

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

Reports on Zenodo

CO-OPERAS

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

QUESTIONI APERTE

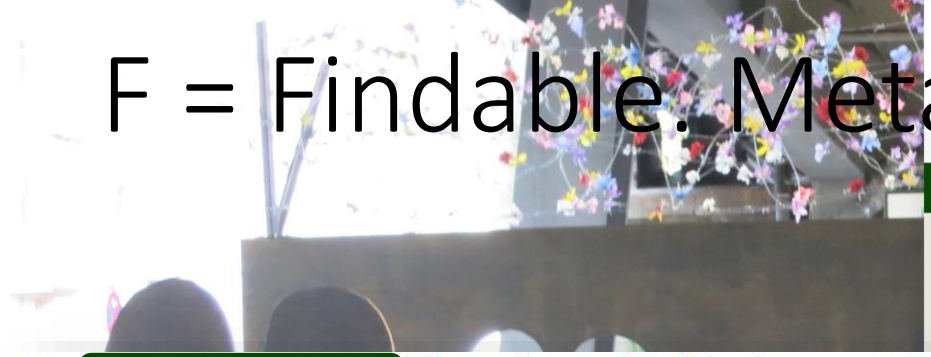
- CI VUOLE TEMPO, E NON VIENE RICONOSCIUTO
- MANCANO CONSAPEVOLEZZA E COMPETENZE
- **FINDABLE**: MANCANO METADATI (E COMPETENZE SUI)
- **ACCESSIBLE**: MANCA REPOSITORY CON GRANULARITÀ ADATTA; MANCA UNICO PUNTO DI ACCESSO [TRIPLE]
- **INTEROPERABLE**: METADATI SPESSO TROPPO SPECIFICI; LINGUE NAZIONALI,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

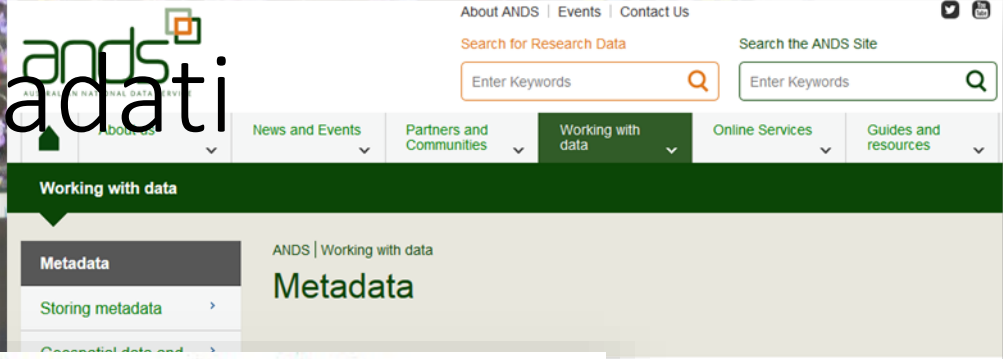
IN CORSO:

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

FINDABLE

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

Data Documentation and Metadata

FAIR Humanities!

- **TEI** (Text Encoding Initiative): www.tei-c.org
- **CEI** (Charter Encoding Initiative): http://www.cei.lmu.de/index.php
- **MEI** (Music Encoding Initiative): https://music-encoding.org/
- **CMDI** (Language Resources, CLARIN):
- **IIIF** (International Image Interoperability Framework): https://iiif.io/
- **EAD** (Encoded Archival Description, for finding aids): https://www.loc.gov/ead/
- **Dublin Core** (description of digital documents): http://dublincore.org/

Wuttke 2018

Metadata

RDA | Metadata Directory

RDA Metadata directory

Edit this page

View the standards

View the extensions

Arts and Humanities Edit

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

Social and Behavioral Sciences

- 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

DDI

Learn Products Membership Events Publications About

Document, Discover and Interoperate

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

innovation in metadata design, implementation & best practice

Dublin Core™ Metadata Initiative

URI	http://purl.org/dc/terms/creator
Label	Creator
Definition	An entity responsible for making the resource.
Comment	Recommended practice is to identify the creator with a URI. If this is not possible or
Type of Term	Property
Range Includes	• http://purl.org/dc/terms/Agent
Equivalent Property	• http://xmlns.com/foaf/0.1/maker
Subproperty of	• Creator (http://purl.org/dc/elements/1.1/creator) • Contributor (http://purl.org/dc/terms/contributor)

Term Name: date

URI	http://purl.org/dc/terms/date
Label	Date
Definition	A point or period of time associated with an event in the lifecycle of the resource.
Comment	Date may be used to express temporal information at any level of granularity. Reco ISO 8601-1 [ISO 8601-1] or a published profile of the ISO standard, such as the W3 Specification [EDTF]. If the full date is unknown, month and year (YYYY-MM) or just time specification in which start and end dates are separated by a '/' (slash) charact

TEI < Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

Text Encoding Initiative

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. In addition to the Guidelines themselves, the Consortium provides a variety of resources and training events for learning TEI, information on projects using the TEI, a bibliography of TEI-related publications, and software developed for or adapted to the TEI.

F=Findable. Metadata

Openrefine

OpenRefine

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

Home
Community
Documentation
Download
Data Privacy
Contact Us
Blog

Welcome!

OpenRefine (previously Google Refine) is a powerful tool for worki
cleaning it; transforming it from one format into another; and exte
services and external data.

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

OpenRefine is available in more than 15 languages.

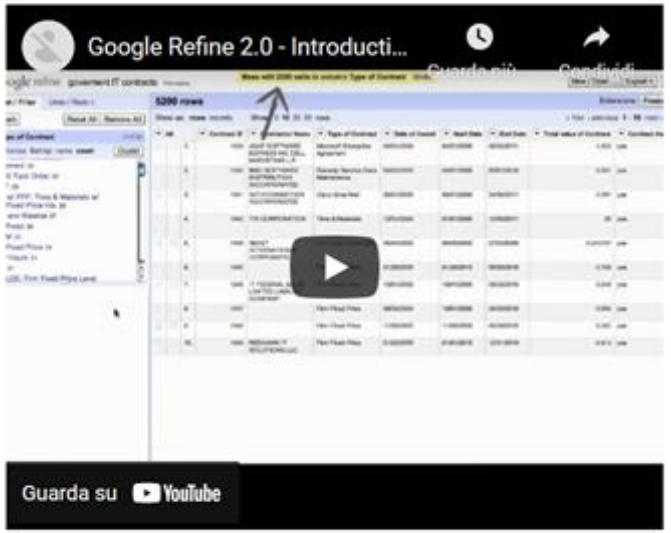
OpenRefine is part of Code for Science & Society.

Introduction to OpenRefine

1. Explore Data

1. Explore Data

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



2. Clean and Transform Data

F = Findable
Europeana data model

europeana

The Europeana Data Model for Cultural Heritage

#AllezCulture

What is the issue?

A vast number of Europe's cultural heritage objects are digitised by a wide range of data providers from the library, museum, archive and audio-visual sectors, and they all use different metadata standards. This data needs to appear in a meaningful way in a cross-cultural, multilingual context such as Europeana. Numerous cultural heritage resources such as thesauri exist worldwide and have the potential to add valuable content at low cost when re-used. Duplication of effort, however, needs to be avoided. The Linked Open Data environment lacks authoritative data from the cultural heritage community to contribute to the development of new knowledge.

What is the solution?

The Europeana Data Model (EDM) aims to bridge these gaps in the Europeana context. EDM is a major improvement on ESE, Europeana's first data model.

EDM transcends domain-specific metadata standards, yet accommodates the range and richness of community standards such as LIDO for museums, EAD for archives or METS for digital libraries.

It **facilitates Europeana's participation in the Semantic Web**, basing itself on an open, cross-dom semantic web-based framework.

EDM is a more developed data model that brings **more meaningful links to Europe's cultural herita data.** Data from partners or external information resources with references to persons, places, subje etc., will connect to other initiatives and institution This will result in sharing enriched content, adding and thereby generating more content in ways that single provider could achieve alone.

Parthenos training

europeana
think culture

Europeana Data Model – Mapping Guidelines v2.4

06/10/2017

2 Overview of the properties in each class

✔ = **Mandatory property**

➔ Blue = **at least** one of the blue properties should be present (and can be used alongside each other)

O Red = **at least** one of the red properties should be present (and can be used alongside each other)

♦ Green = **at least** one of the green properties should be present (and can be used alongside each other)

+ = recommended property

Detailed characteristics of all properties are given in the full tables following this summary.

Core Classes

Properties for edm:ProvidedCHO			
+	dc:contributor		dcterms:isReplacedBy
	dc:coverage		dcterms:isRequiredBy
+	dc:creator	+	dcterms:issued
+	dc:date		dcterms:isVersionOf
➔	dc:description		dcterms:medium
	dc:format		dcterms:provenance
+	dc:identifier		dcterms:references
✔	**dc:language** (if edm:type = TEXT)		dcterms:replaces
+	dc:publisher		dcterms:requires
	dc:relation	O	dcterms:spatial

F = findable. Metadata tools

What CEDAR does

https://metadatacenter.org/

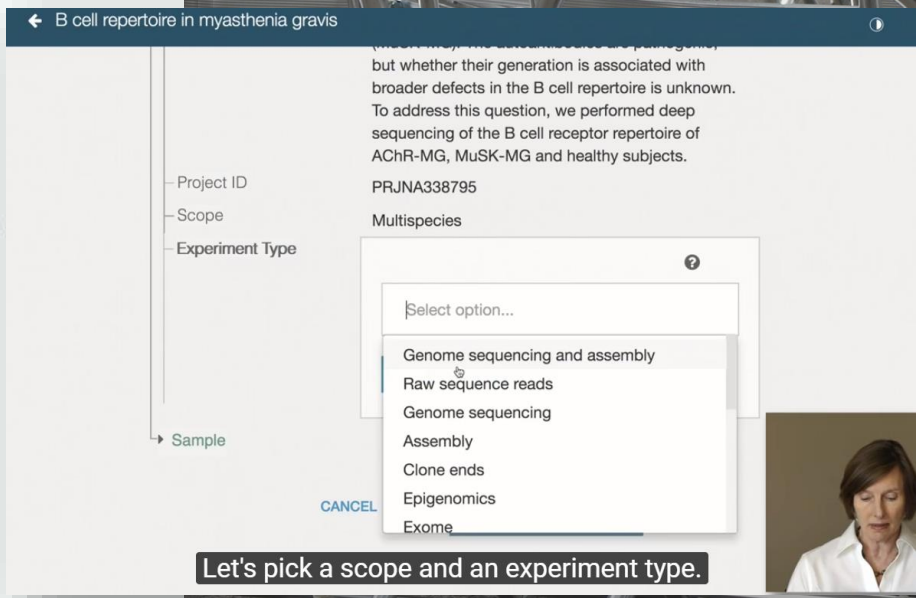
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools wil
metadata record is created to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools hel
to users, and download the information that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

What can CEDAR do for me already?

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

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

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



Findable — Metadata cre

FAIR Cookbook

FAIR cookbook

Search this book...

- FAIR Cookbook
- Introduction
- Assessing FAIR
- Infrastructure for FAIR
- Improving Findability
- Improving Accessibility
- Improving Interoperability
  - How to interlink data from different sources?
  - Identifier mapping with BridgeDB
  - Which vocabulary to use?
  - Requesting terms addition to terminology artefacts

Creating a Metadata Profile

Recipe metadata

identifier: RX.X version: v1.0

Difficulty level

Reading Time

20 minutes

Recipe Type

Hands-on

Executable Code

Yes

Intended Audience

Principal Investigator

Data Manager

Data Scientist

How to generate a metadata template

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

Step 1: Define competency questions

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

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

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

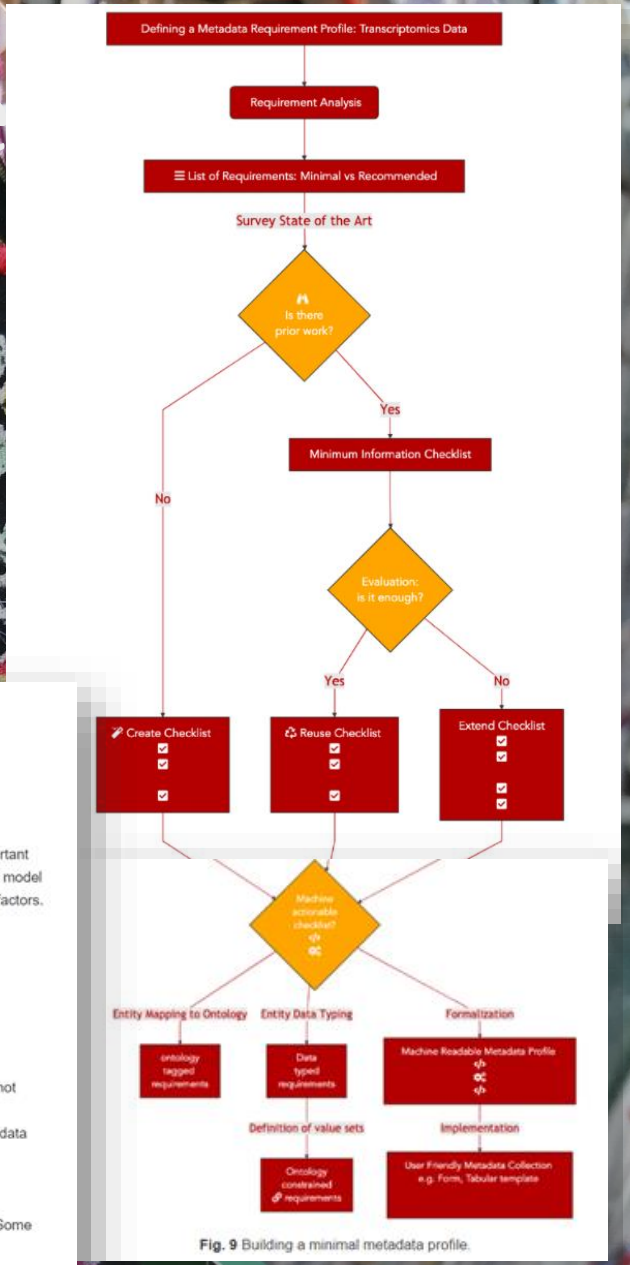
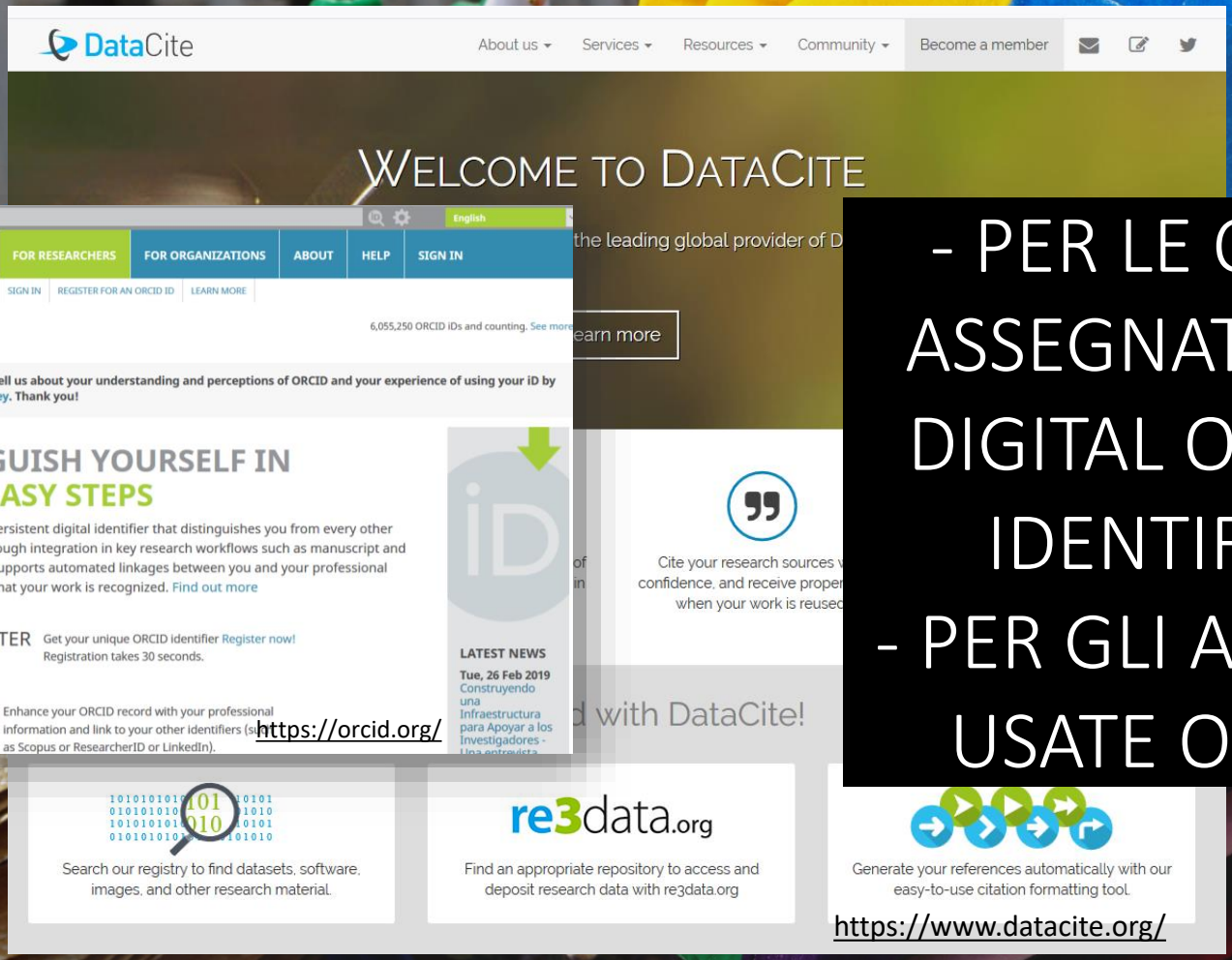


Fig. 9 Building a minimal metadata profile.

F = Findable. Identificativi persistenti



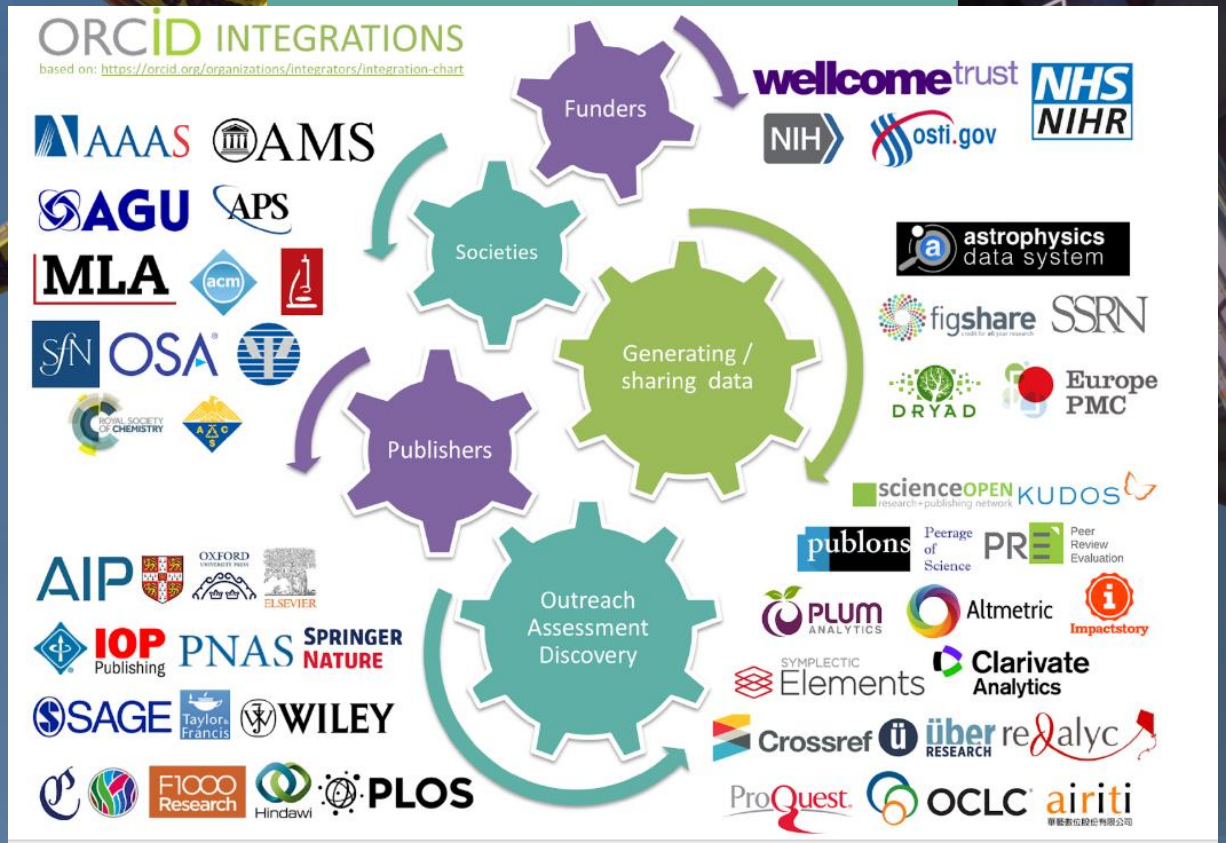
https://orcid.org/

https://www.datacite.org/

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

ORCID – ID. Tutto intorno a me

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



ACCESSIBLE

A = Accessible

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

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

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

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

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

CESSDA Guide

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

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. Data lab [in progress]

Data lab

Save your data during research

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

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

DataverseNL

Members of the three technical universities (Eindhoven, Twente and Delft) can draw upon DataverseNL through 4TU.ResearchData. The objective of this open source application is to make research data accessible to others. You can store data in the DataverseNL environment and grant multiple individuals controlled access to them.
Once you have requested the project environment or dataverse from 4TU.ResearchData, you are free to arrange this environment as you wish.

Why DataverseNL?

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

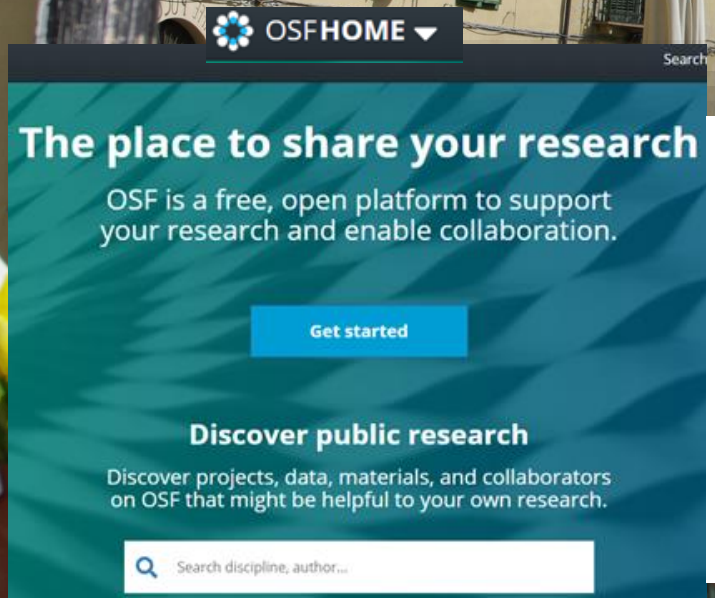
See our quick guide for getting started or contact us when you need additional support.

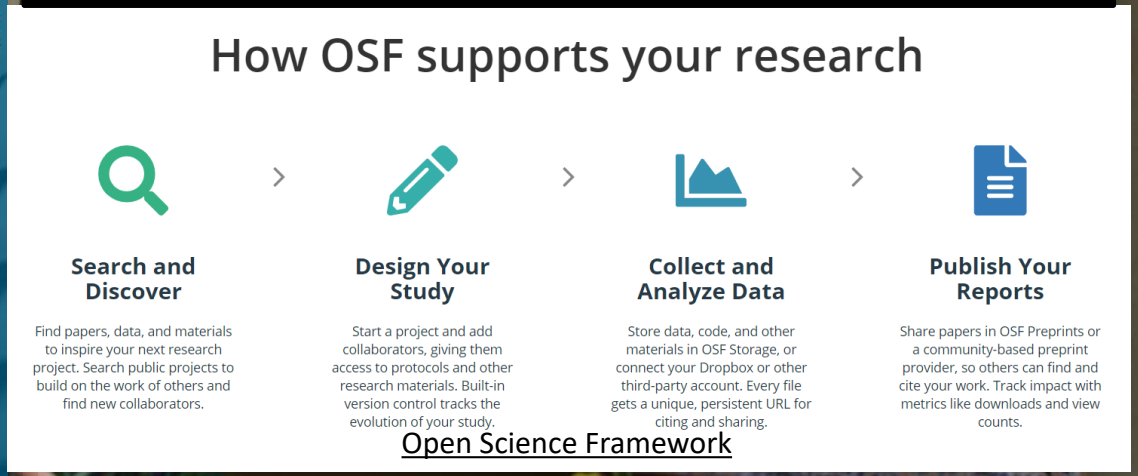
Costs

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

A = Accessible. Data repositories +

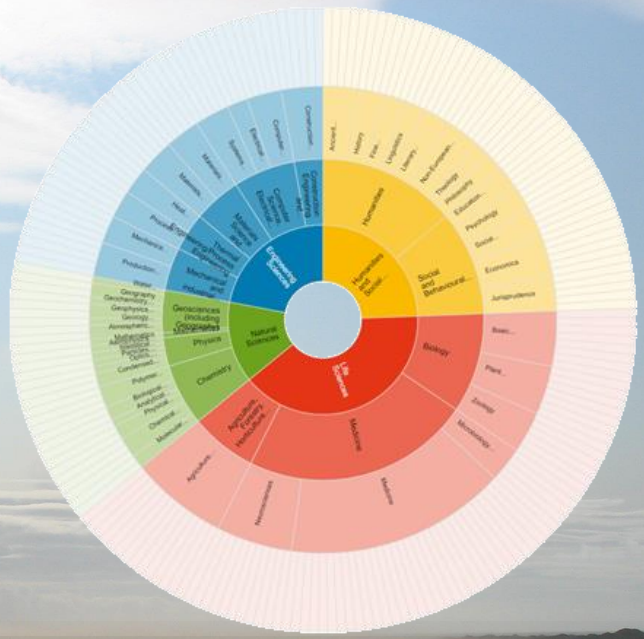
AMBIENTE INTEGRATO CONTIENE DATI, CODICE, PROTOCOLLI E PERMETTE PUBBLICAZIONE PREPRINT [SCEGLIERE SERVER AMBURGO PER GDPR]





Open Science Framework

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...

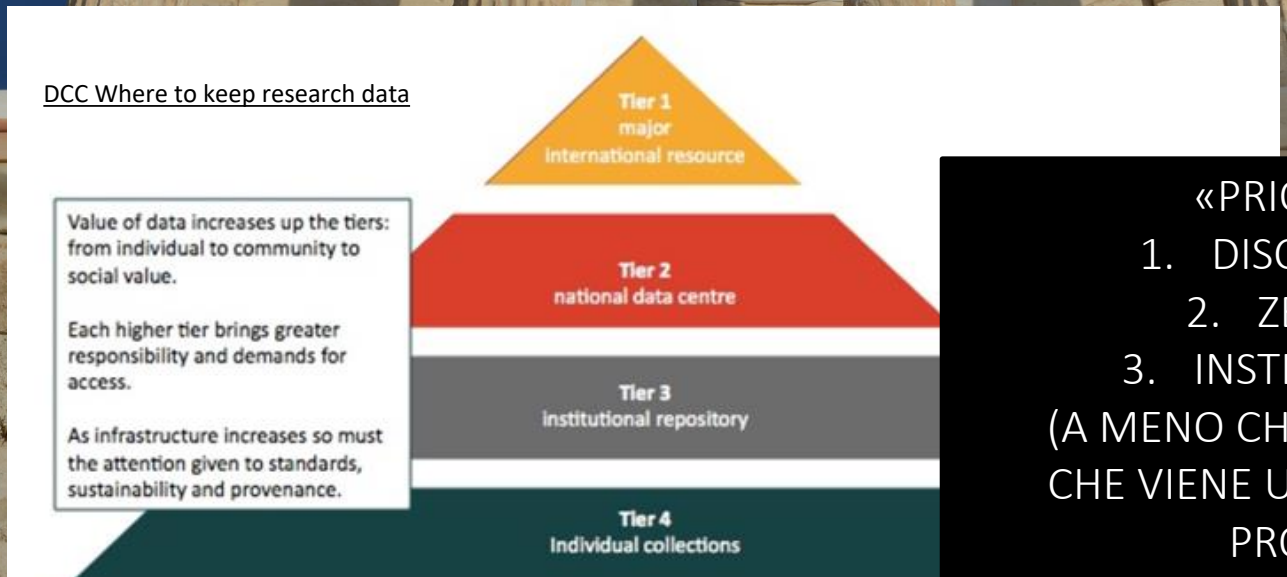
Read more

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 repositories



«PRIORITÀ»

1. DISCIPLINARY
2. ZENODO
3. INSTITUTIONAL

(A MENO CHE SIA DATAVERSE, CHE VIENE USATO DURANTE IL PROGETTO)

A = Accessible. Data repositories

Checklist: is it the right repository for your data?

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

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

DCC checklist

CHECKLIST PER LA SCELTA

Findable, accessible and interoperable

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

Legal terms and conditions

Personal data or data which may identify individuals when linked to other data should not be stored outside the European Economic Area, unless in a legal jurisdiction that ensures personal data is adequately protected ❑

By agreeing to the terms and conditions the depositor will not be breaching other **Data Protection** principles, or the terms of any confidentiality agreement with data subjects or owners (e.g. consent form, consortium agreement) ❑

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

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

Criteria for the selection of a trustworthy repository

Jan. 27, 2021

TRUSTWORTHY REPOSITORIES

Trustworthy repositories should meet the following minimum criteria:

1. **Provision of Persistent and Unique Identifiers (PIDs)**
   a. Allow data discovery and identification
   b. Enable searching, citing, and retrieval of data
   c. Provide support for data versioning

2. **Metadata**
   a. Enable finding of data
   b. Enable referencing to related relevant information, such as other data and publications
   c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
   d. Use metadata standards that are broadly accepted (by the scientific community)
   e. Ensure that metadata are machine-retrievable

3. **Data access and usage licences**
   a. Enable access to data under well-specified conditions
   b. Ensure data authenticity and integrity
   c. Enable retrieval of data
   d. Provide information about licensing and permissions (in ideally machine-readable form)
   e. Ensure confidentiality and respect rights of data subjects and creators

4. **Preservation**
   a. Ensure persistence of metadata and data
   b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

A = Accessible. Data journals

Data journals list

Title	URL	Charge	Notes for authors (N.B. we suggest checking in particular for policy on submission of data already published)
Journal of Open Archaeology Data	http://openarchaeologydata.metajnl.com/		http://openarchaeologydata.metajnl.com/about/submissions
Open Health Data	http://openhealthdata.metajnl.com/		
Journal of Open Psychology Data	http://openpsychologydata.metajnl.com/		
Scientific Data	http://www.nature.com/scientificdata/		
Genomics Data	http://www.journals.elsevier.com/genomics-data/		
Geoscience Data Journal	http://www.geosciencedata.com/		

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.

les providing the data, metadata and field
s (.csv, .txt, .csvt respectively) for the main
non-ceramic and non-lithic finds.
files providing the data, metadata and field
s (.csv, .txt, .csvt respectively) for the main
standing remains, except for terraces.
or polygon dataset (.shp and associated
shape of Antkythera's coastline.
tor polygon dataset (.shp and associated
main bedrock units on Antkythera.
or polygon dataset (.shp and associated
main stage-one survey units.
or polygon dataset (.shp and associated
main stage-two survey units.
or line dataset (.shp and associated files)
able agricultural terraces (i.e. the location

· *other* –
assistanc
· *geology*
(UCL) an

Repository
UK Archae
10.5284/10

Publication Date
05/02/2012

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

License
Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. http://dx.doi.org/10.5284/1000271, http://dx.doi.org/10.5284/1000208, 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

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

6.2. Data journals in humanities

Data journals SSH

In addition, or in parallel, to making your data available in a repository, you can maximize the credit that you gain for them by publishing them in a data journal. Data journals, usually built on the top of data repository services, are designed to introduce, describe and contextualize data sets to facilitate their online exploration. One of the important benefits of publishing your data in terms of research data management is that it undergoes a peer-review process, which means that experts validate its quality, provide templates for its proper description and offer guidance on where to deposit it.

- Data journals in the humanities:
  - Research Data Journal for the Humanities and Social Sciences
  - Journal of Open Humanities Data Dataverse
  - Journal of Cultural Analytics
  - Journal of Open Archaeology Data
  - Journal of the Text Encoding Initiative
  - (Video) Journal of Embodied Research

A = Accessible. Formati

Data Archiving and Networked Services

DANS

HOME DEPOSIT

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

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

A – Accessible - Formati

NationalArchives

NATIONAL ARCHIVES

Appendix A: Tables of File Formats

Quick Links

Computer Aided Design	Digital Audio	Digital Moving Images
Digital Cinema	Digital Video	
Digital Photographs	Scanned Tex	
Geospatial Formats	Presentation	
Structured Data Formats	Email	
Calendars	Navigational	

Symbol Key

Preferred Formats

Acceptable Formats

Geospatial Formats

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

Preferred Formats

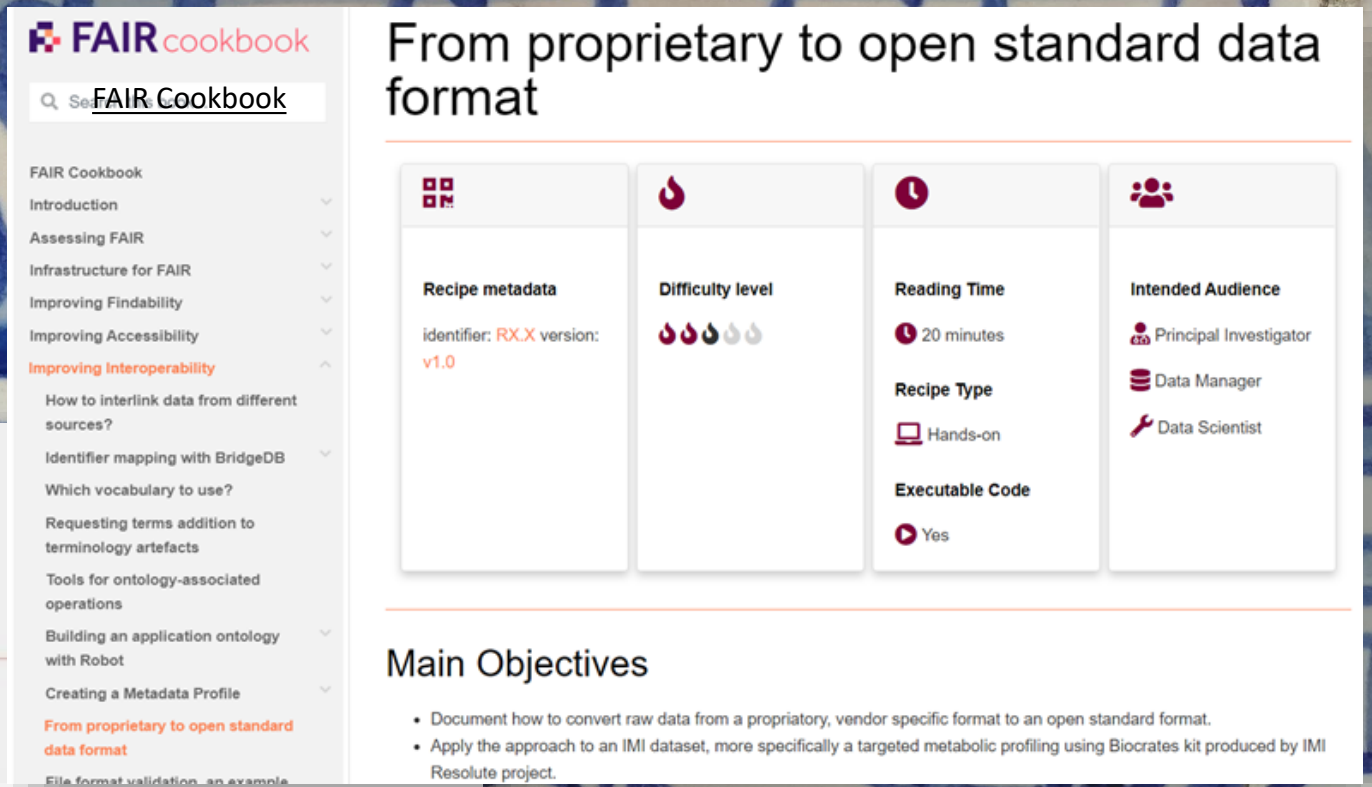
Preferred Formats	Format Versions	Format Specifications
Geospatial Tagged Image File Format	1.8.2	Geo TIFF Format Specification: (http://geotiff.maptools.org/spec/geotiffhome.html)
Geographic Markup Language	2.0 through 3.2	ISO 19136:2007 & Version 3.2, OGC document 07-036: (http://www.opengeospatial.org/standards/is)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

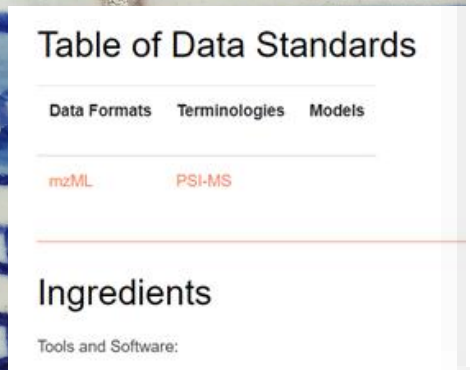
Acceptable Formats

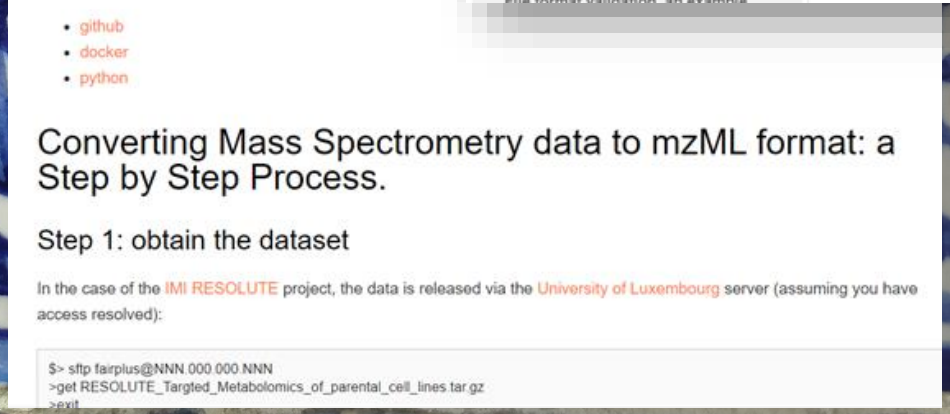
Acceptable Formats	Format Versions	Format Specifications
Vector Product Format		MIL-STD-2407: (http://earth-info.nga.mil/publications/specs/printed/2407/2407_VPF.pdf)
ESRI ARC/INFO Interchange File Format		Reverse engineered specification: (http://avce00.maptools.org/docs/v7_e00_cover.html)
TerraGo Geospatial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

Acceptable for Imminent Transfer Formats

A = Accesible - Formati

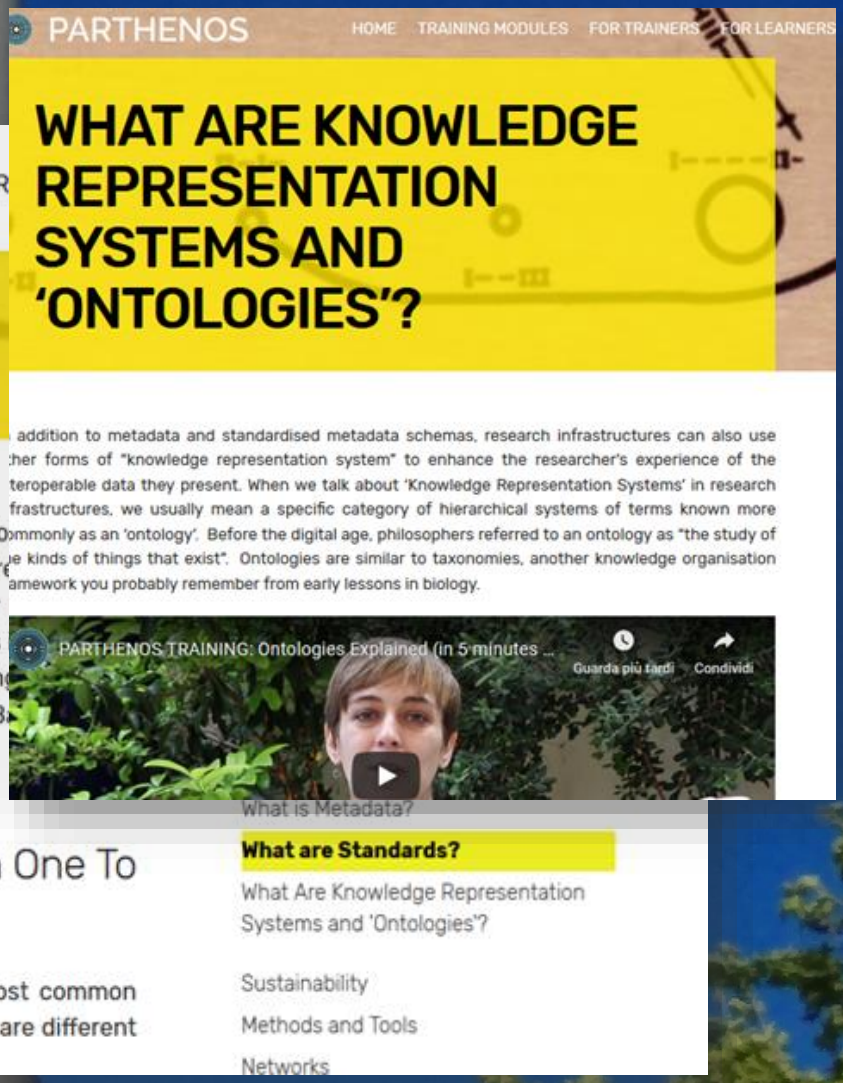






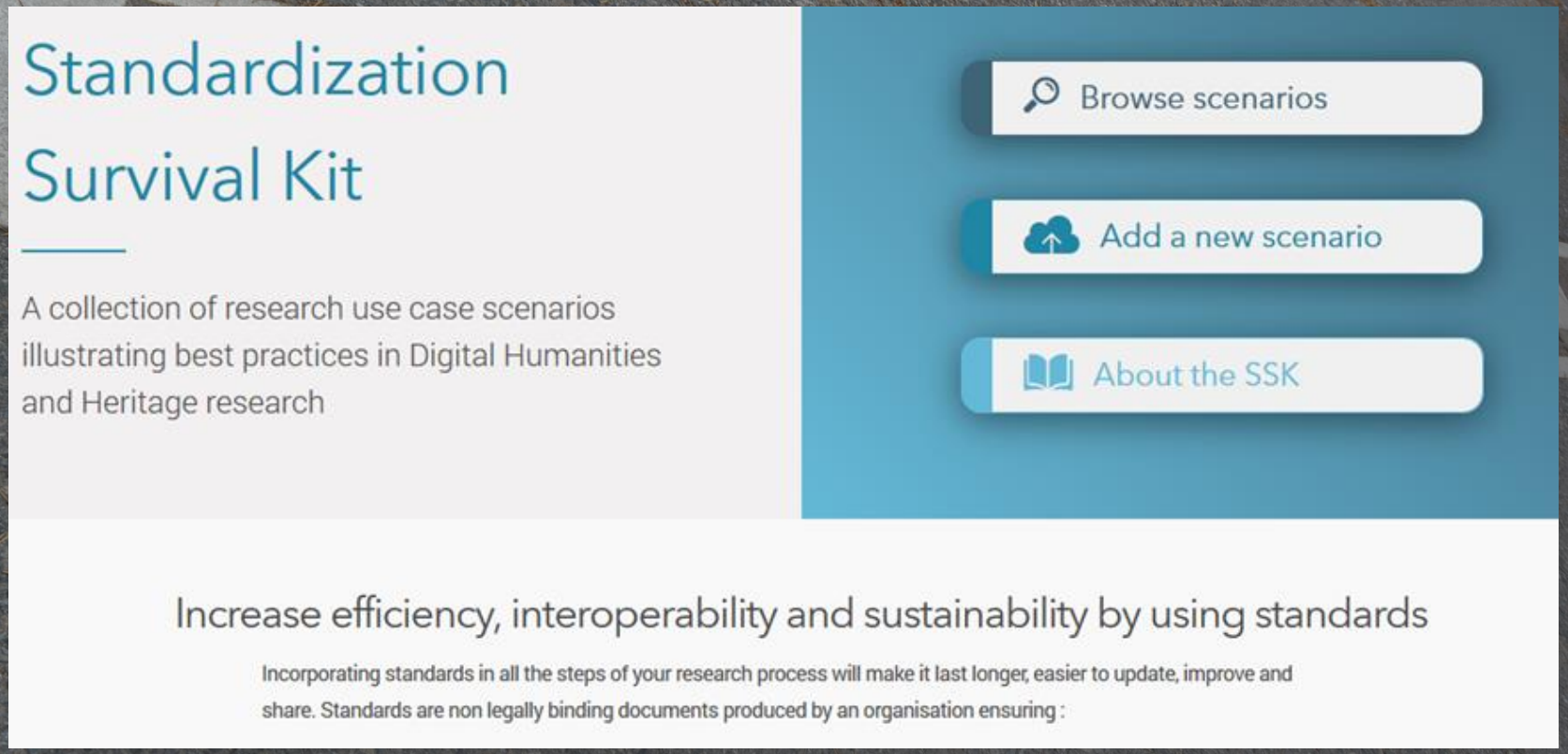
INTEROPERABLE

I = Interoperable. Standards



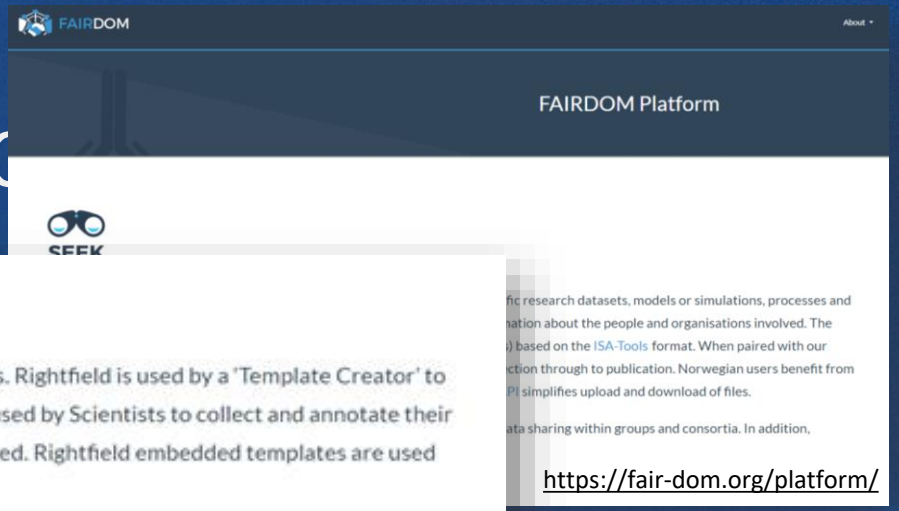
http://training.parthenos-project.eu/sample-page/intro-to-ri/interoperability/what-are-standards/

I = Interoperable. Standards



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

I= Interoperable – ac



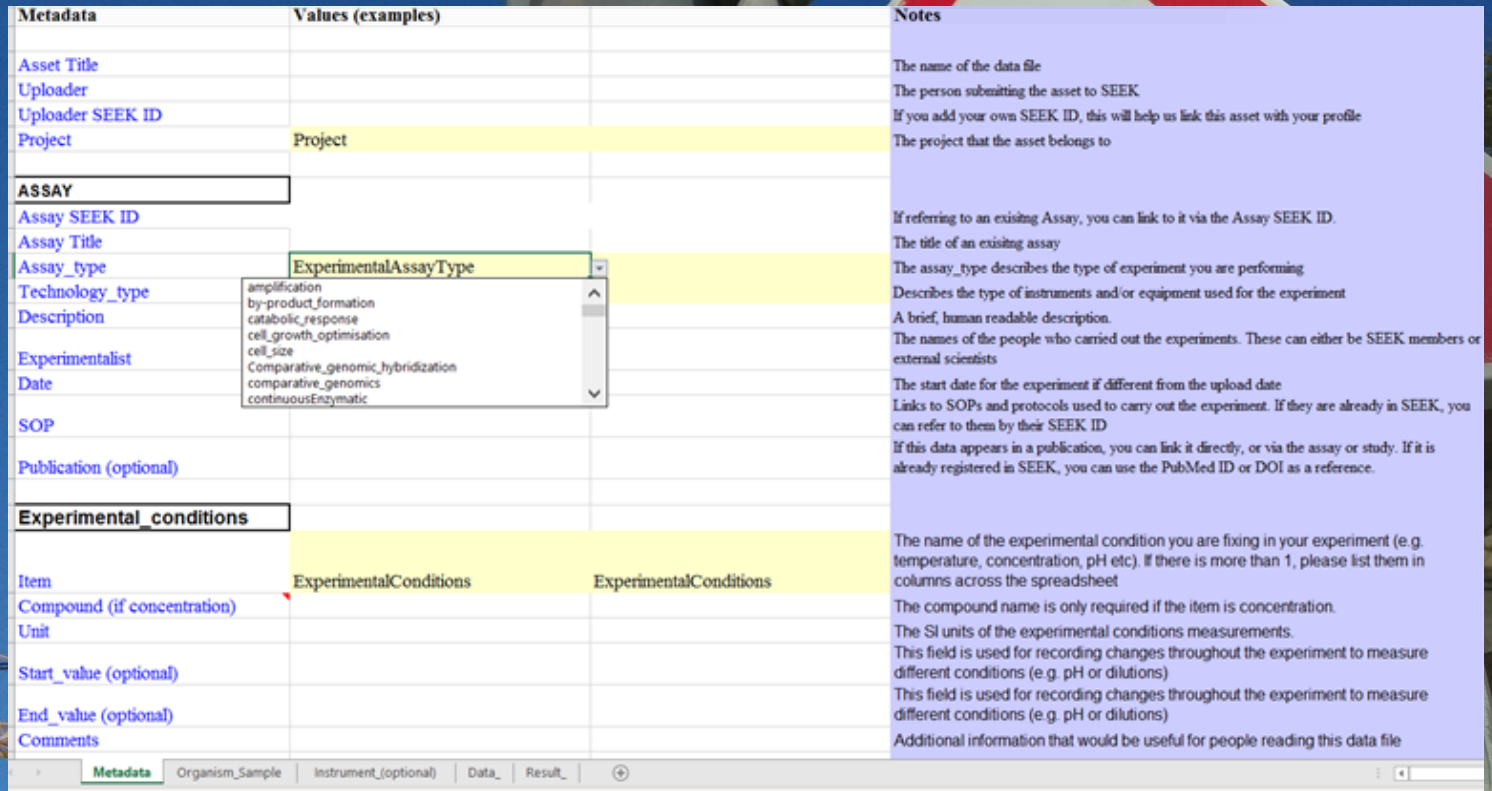
https://fair-dom.org/platform/

RightField

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

MORE INFORMATION

Rightfield



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

amplification
by-product_formation
catabolic_response
cell_growth_optimisation
cell_size
Comparative_genomic_hybridization
comparative_genomics
continuousEnzymatic

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

PARTHENOS

HOME TRAINING MODULES FOR TRAINERS FOR LEAF

Parthenos

What formal ontologies exist that are relevant to digital humanities?

The potential topics of interest in digital humanities are vast and potentially unlimited, so it is not really possible to state what formal ontologies may be of use in any given research project. The appropriateness of an ontology for application in a particular research programme depends on the data and information you want to include in it, and the research question brought to it. Your research question, and the kind of data you want to model should be compared against the scope of potential ontologies in order to see whether the proposed standard is suitable.

It is of interest, however, to point out the distinction between top-level ontologies and domain or application ontologies. **Top-level ontologies** allow integration of data on an extremely high level, sometimes including logical rules within the structure that allow for automated reasoning over datasets. Some of the more well known top-level ontologies include:

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

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

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

Unfortunately, there is no single collection of ontology resources that would allow the perusal of all potentially applicable ontologies. Sites such as schema.org and bartoc.org however do provide tools to find potential specific schema and ontology representations that may suit your needs.

I = Inteoperable. Ontologies

Opscidia

Opscidia's ontology generator

Written on 03 March 2021.

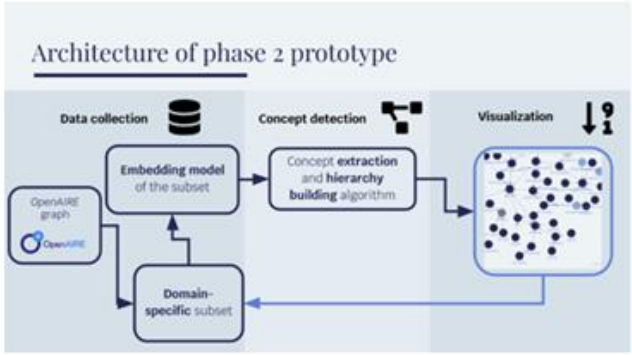
OPSCIDIA

The value of knowledge

GENERATORE DI ONTOLOGIE

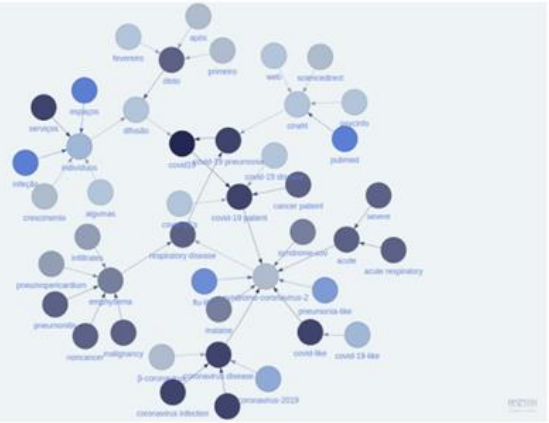
The solution proposed by Opscidia is an ontology generator that consists in three layers:

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

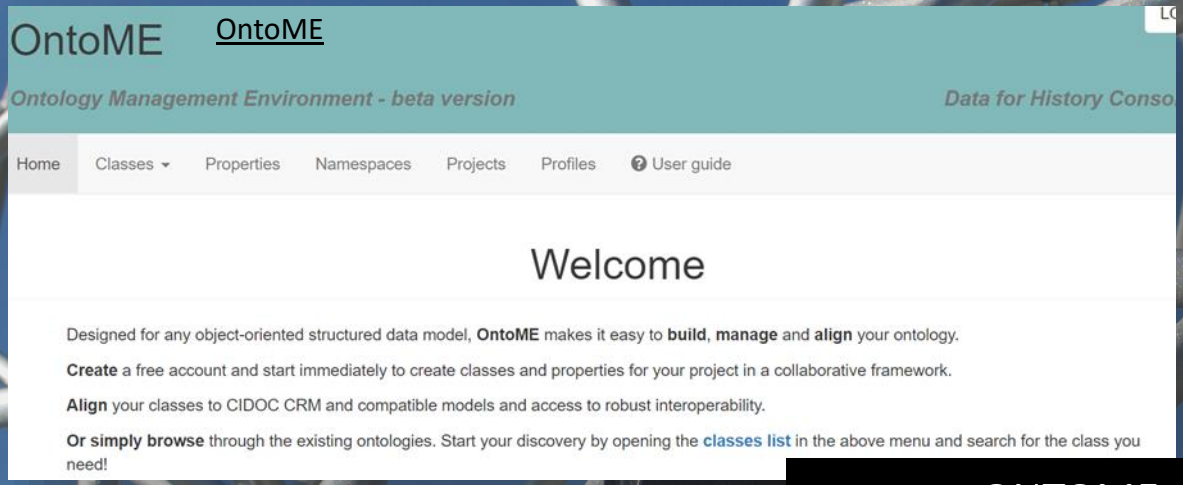


The results of the Ontology Generator

A simple tool for semi-automatic domain specific ontology creation has been built.
It takes a concept as an input and extracts from a subset of OpenAIRE graph a hierarchical list of concepts associated with the user input. This list is displayed using a simple visualization layer and linked back to the scientific literature through OpenAIRE graph.



I = Inteoperable. Ontologies



OntoME

ONTOME
ESEMPIO DI APPLICAZIONE
ALLE SCIENZE STORICHE

2021

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

Francesco Beretta
Laboratoire de recherche historique Rhône-Alpes, CNRS – Université de Lyon, 14 avenue Berthelot, 69363 Lyon cedex 07, France

Abstract

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

I = Inteoperable. 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.

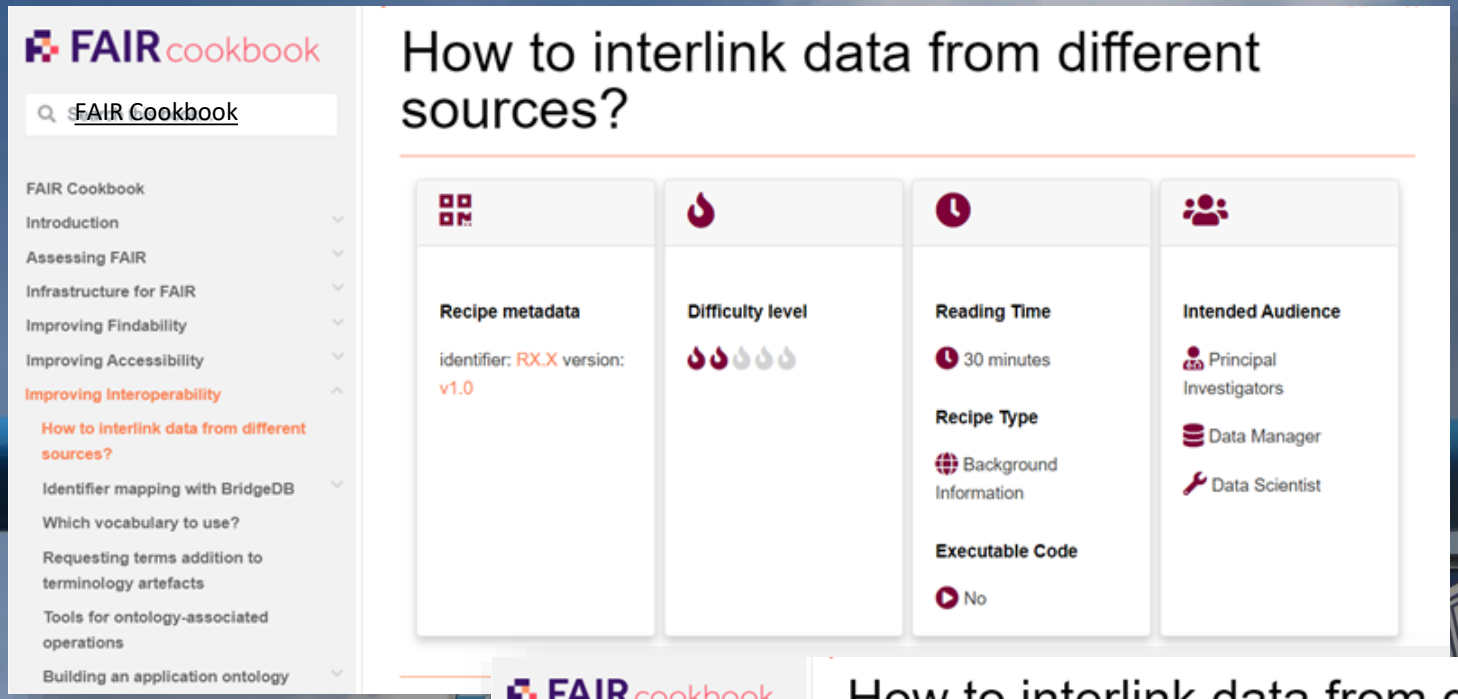
» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies 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.

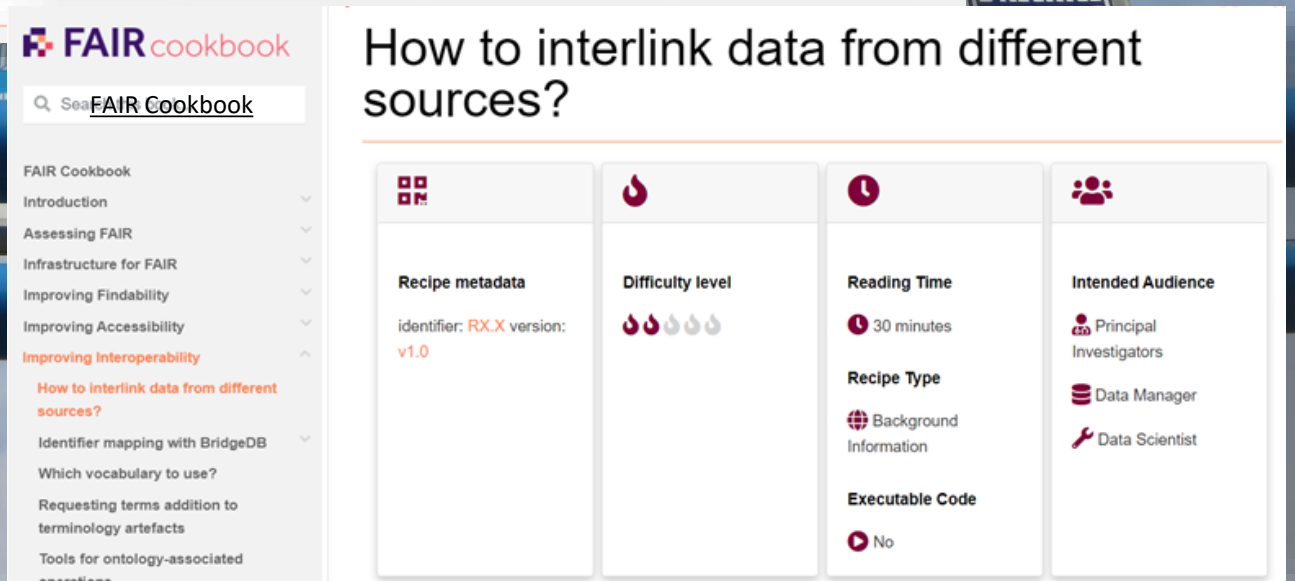
NORMALIZZARE, USARE AUTHORIY FILES (es. VIAF)

- ✓ 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

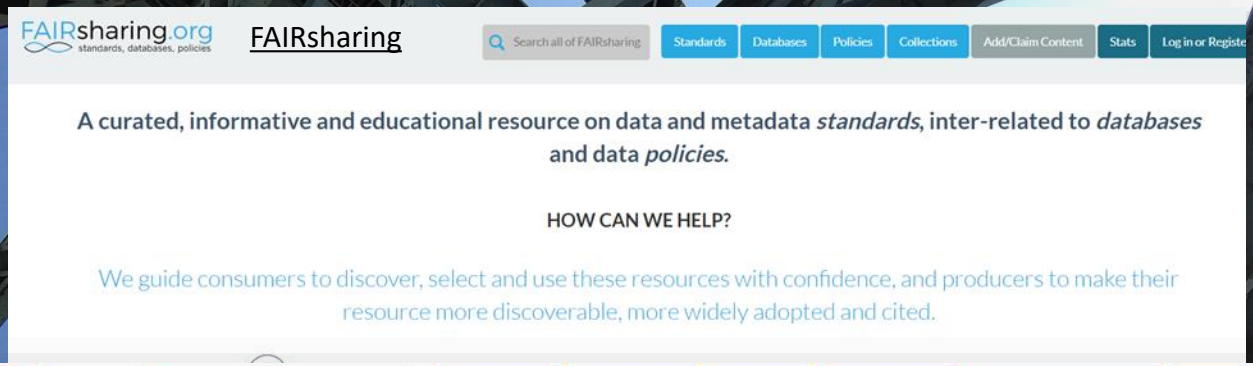
I = Interoperable

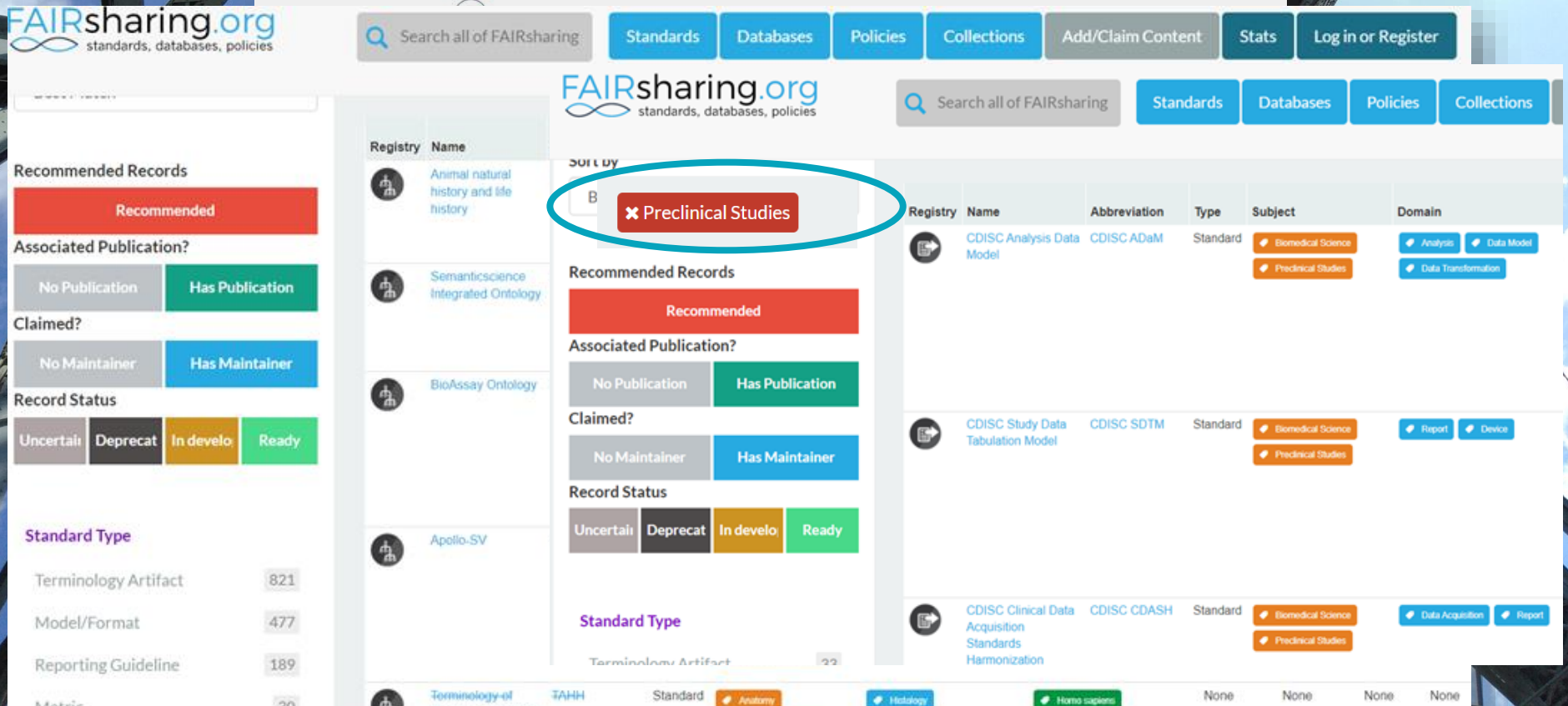




[FAIRsharing. Per essere interoperabili]

FAIRsharing





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

[es. di cosa documentare]

It can get pretty complex, though…

An **information unit** consists of - e.g. in the case of interviews:

- the **audio file** of the interview
- the **interview transcript** in the form of a digital text file
- the discussion guide or **questionnaire**, which explains the methodological approach and is necessary for the comprehensibility of the results of the study.
- the **project explanation** as well as the **declaration of consent** of the interviewee, which documents compliance with the legal provisions of the Federal and State Data Protection Act
- the **codebook**, which e.g. documents the development categories and variables used
- the **documentation** of the procedure for anonymization and pseudonymization
- the indexing information (**metadata**), which guarantees the citation ability of the interview and its findability

Wuttke 2018

[es. di cosa documentare]

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

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

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

- value code labels

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

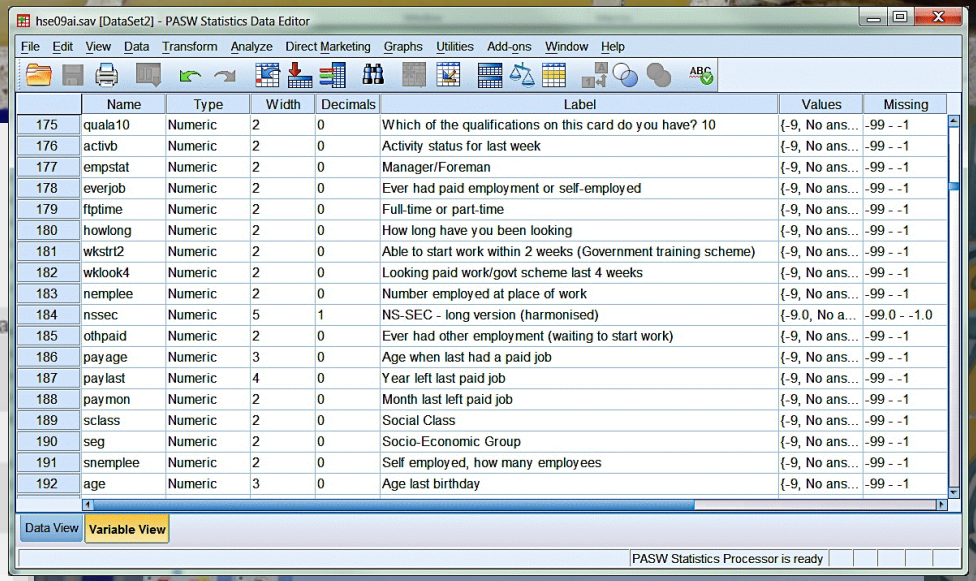
- coding and classification schemes explained, with a bibliographic and dated reference (some standards change over time)

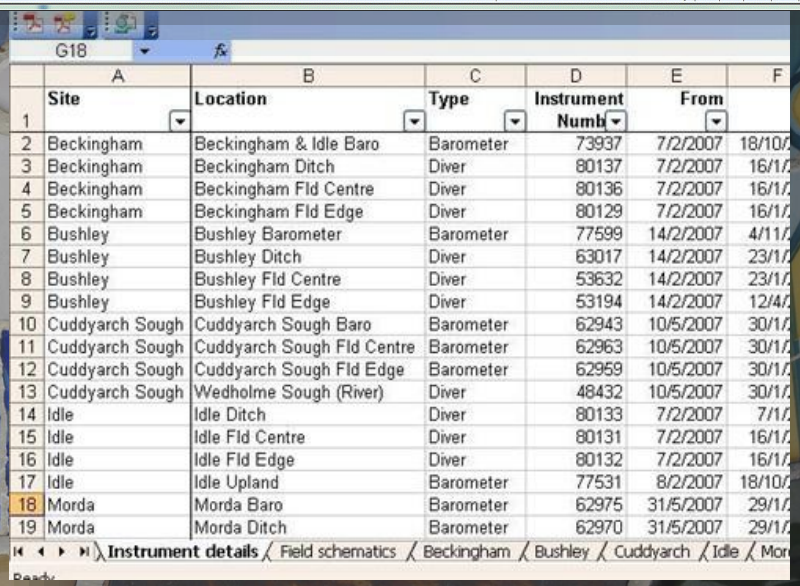
Examples: Standard Occupational Classification, 2000 —— a series of codes to classify respondents' jobs; ISO 3166 alpha-2 country codes —— an international standard of 2-letter country codes

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

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

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



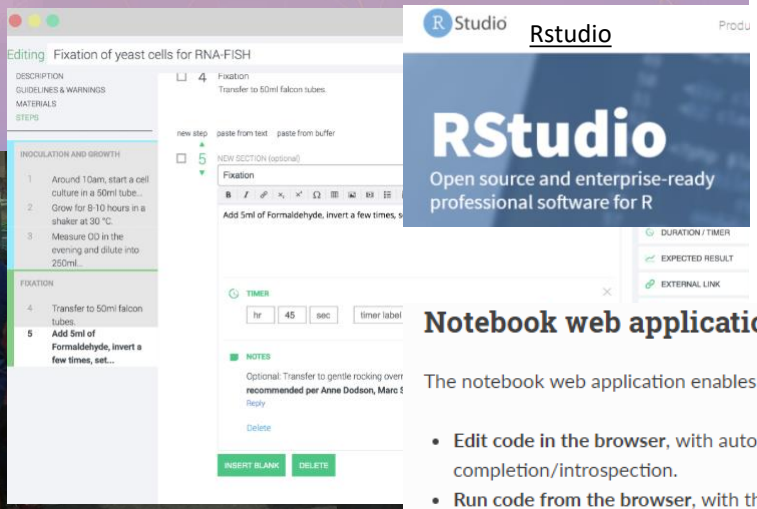


R = Reusable. Documentazione

protocols.io

Make your science more reproducible

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



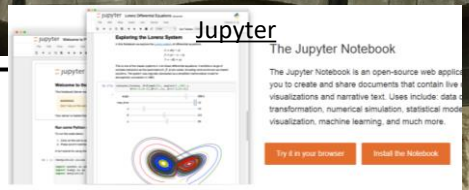
Rstudio

What is an Open Notebook?

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

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

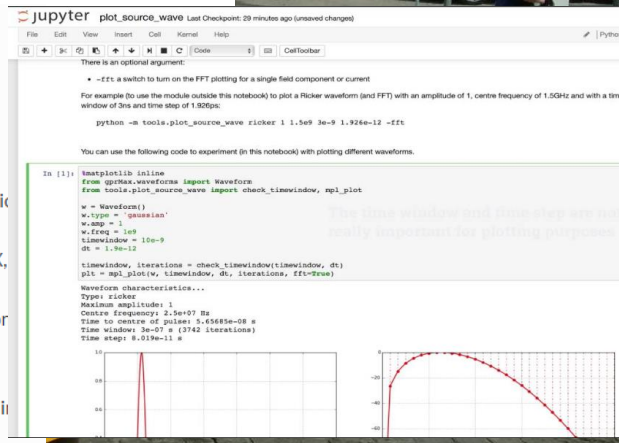
Jupyter



Notebook web application

The notebook web application enables users to:

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



…WHY NOT?

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

R= Reusable. Licenze

KEEP CALM AND

RICORDATE: 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,

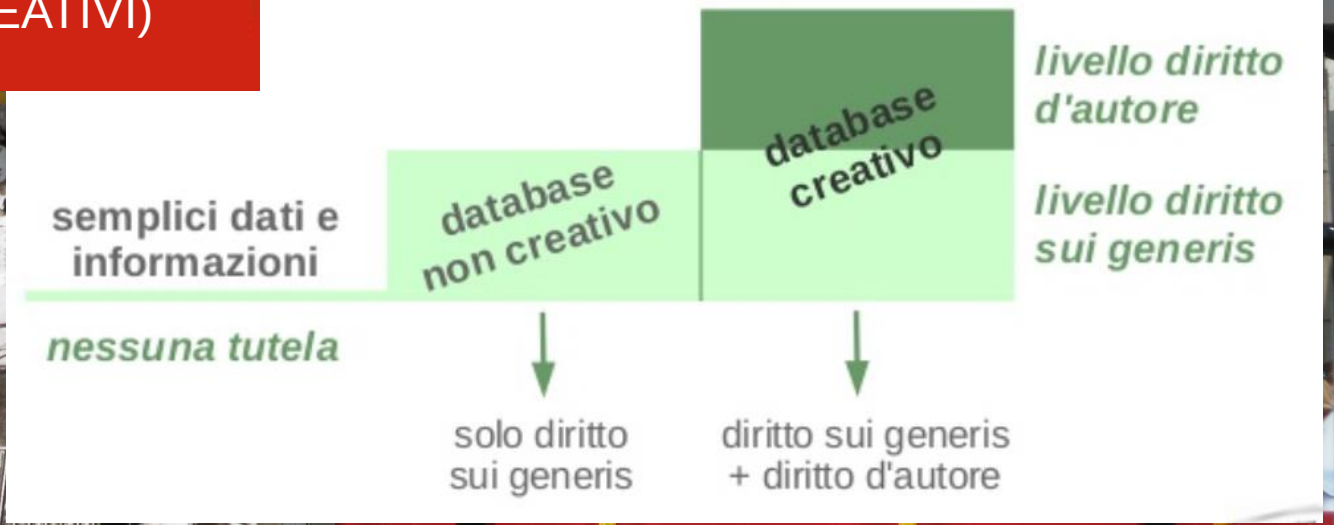
n Community, and in particular Article 57 (2), 66 and 100a thereof,

(1),

Simone Aliprandi

2014

QUALI DIRITTI SUI DATI?



R = Reusable. Licenze

ALLEA Report p. 26

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

Our recommendation is to avoid applying any legal restrictions that do not embrace the principle of openness. The Reusability FAIR principle recommends that data and metadata are released with a clear, human and machine readable data usage licence, in order to avoid legal ambiguity that could limit their reuse.

RECOMMENDATIONS

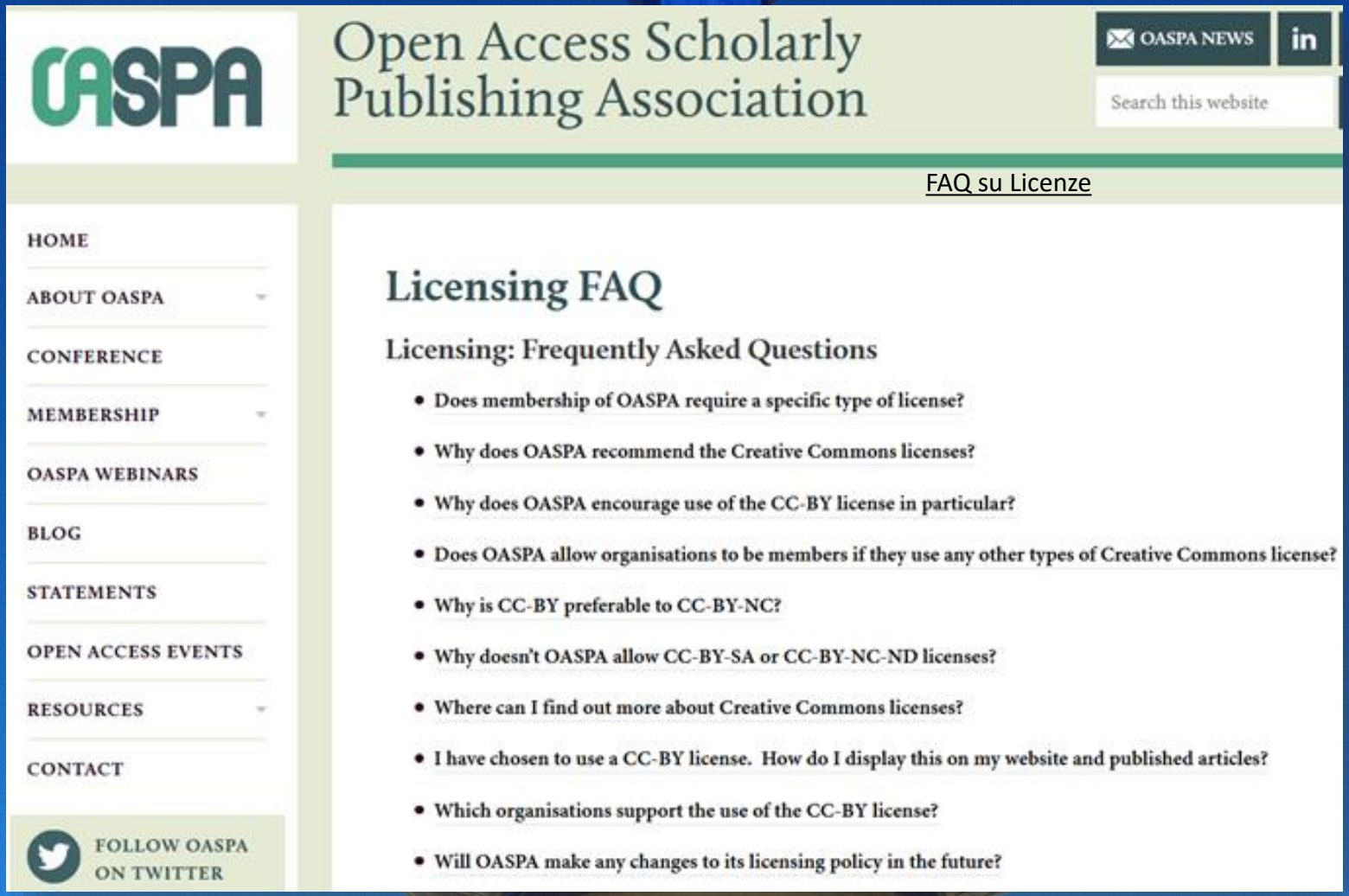
» Proper entitlement: first of all, identify who owns the data, i.e. whether you are entitled to license your work. You may only attribute a licence to a work of which you are the copyright holder. If there are co-authors, you have to agree with them on the licence. Furthermore, you are not allowed to 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



[webinar]

2013

2020

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

2020

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

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 +

R = Reusabl Hum

DEPOSIT, PRESERVE and SHARE

Legal Aspects

Data generated or collected in the e-humanities may potentially be subject to copyright in whole or in part. In addition, particular challenges may arise when some of the items within a data set themselves are subject to third party rights.

In practice, digital humanists can make good use of some checklists to determine whether and how data relevant for their research are subject to legal regulation. Some important questions that have to be solved concern topics like:

- Which national legislation applies to other researchers' work I use in my project?
- Do I have the right to collect, preserve and provide access to the data of my project?
- Is there sensitive information that could connect to some privacy issues?
- Are there risks of exposing the identity of human participants in my study?
- Am I allowed to digitally reproduce material and (re-)publish it in a digital reproduction?

2020

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).

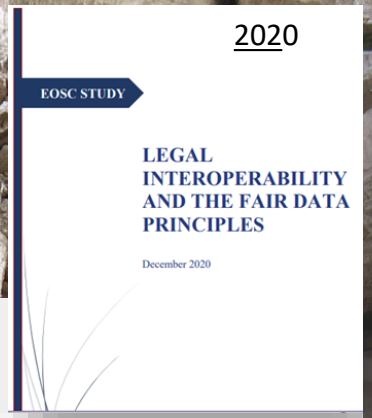
» In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).

» 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.

Legal interoperability

2020



Chapter two - addresses key thematic legal issues, including:

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

1. Open access to research data is an enabler of legal interoperability. The promotion of FAIR Principles should go hand-in-hand with efforts to make data open in accordance with the principle that data must be "as open as possible and as closed as necessary".

2. Regardless of whether the data is Open or not, all new data made available through the EOSC should be FAIR by design.

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

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

Creative 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

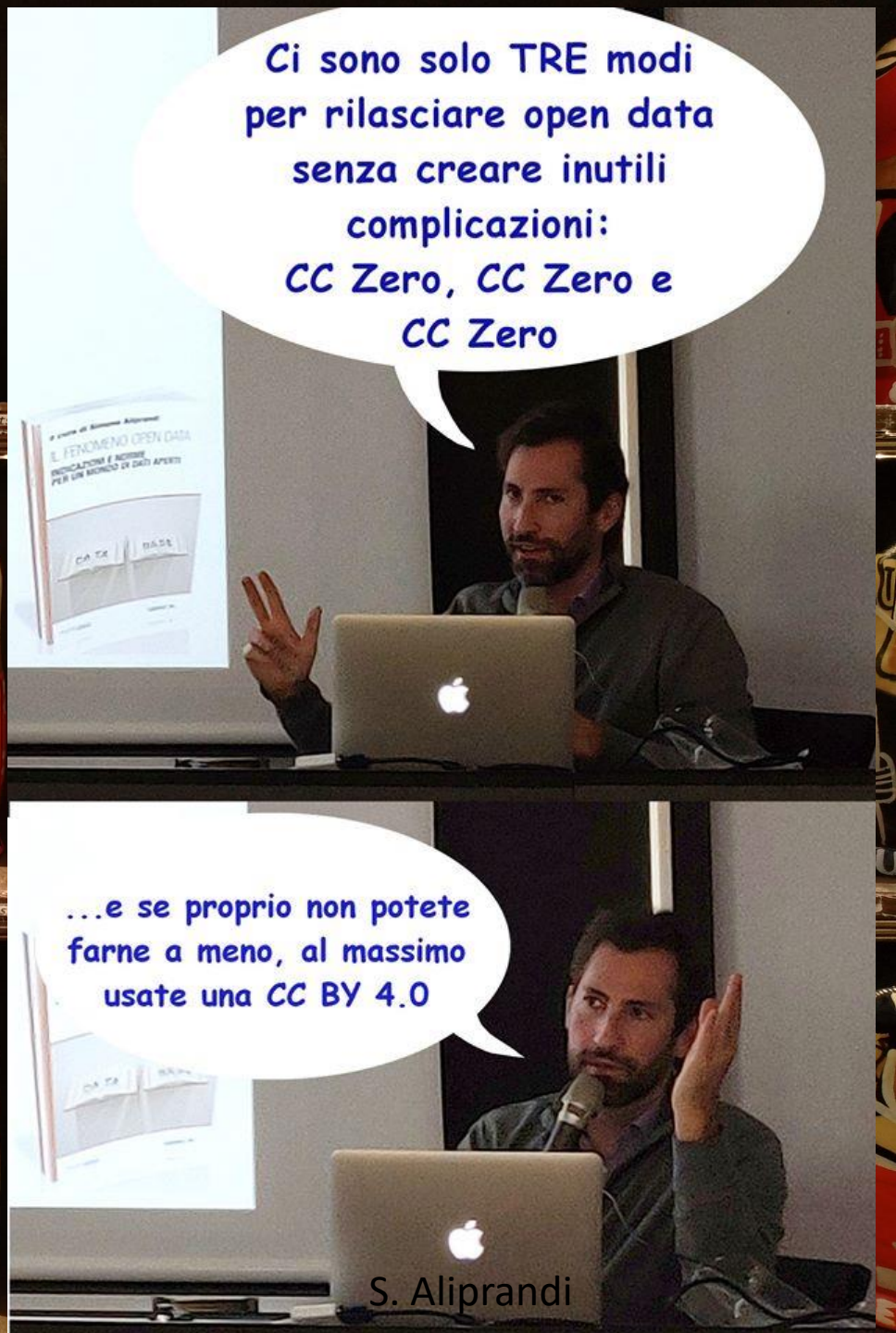
I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

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
SPURS
BULLS
GIANTS


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

Licenze per software 2

ESCLUDE SUCCESSIVA PROPRIETARIZZAZIONE

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

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

Licenze per software 3

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

Licence differentiator

OSSWATCH

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

REMEMBER: ALWAYS READ AND UNDERSTAND YOUR CHOSEN LICENCE.

Choice One

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

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

Please choose

Choice Two (a)

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

Do you want to include licensing conditions on reuse?

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

Please choose

Choice Three

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

License differentiator

[non suoniamo tutti la stessa musica]

⊕ Czech Republic

⊕ Finland

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

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

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

CESSDA guide

⊕ Switzerland

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

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

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

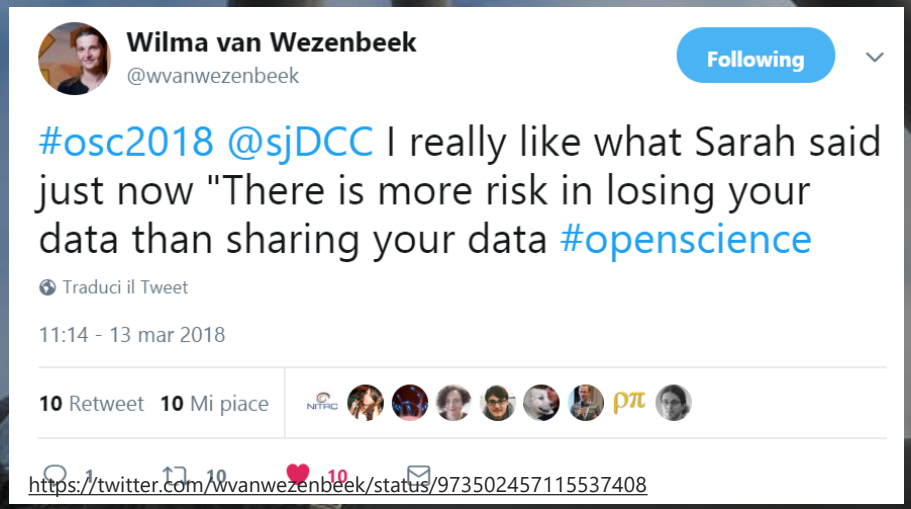
[regole chiare dall’inizio]

… FISSATE REGOLE CHIARE DA SUBITO

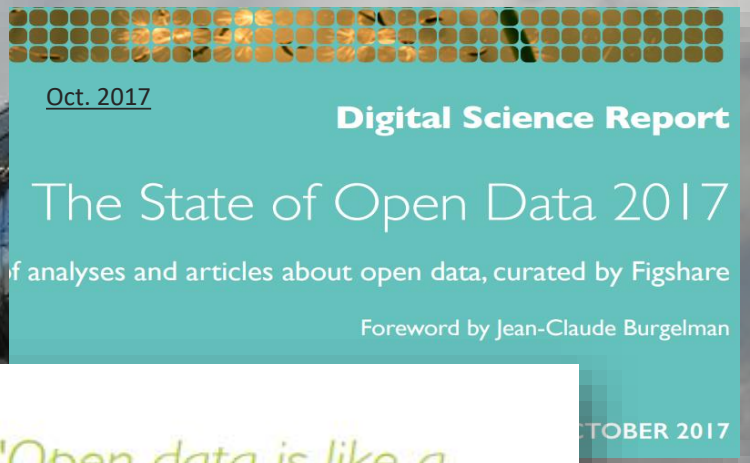
- CHI HA I DIRITTI SUI DATI (SE ESISTE)
- CHI DEVE PROVVEDERE A CONSERVARE
- CHI PUÒ SFRUTTARE

3. DATI OPEN

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."

Perché i dati ape

PORT postgraduate online research training | SCHOOL OF ADVANCED STUDY UNIVERSITY OF LONDON

Sharing Data Why share data

Sharing data

2. Why share data?

Better research

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

Better impact

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

Better value

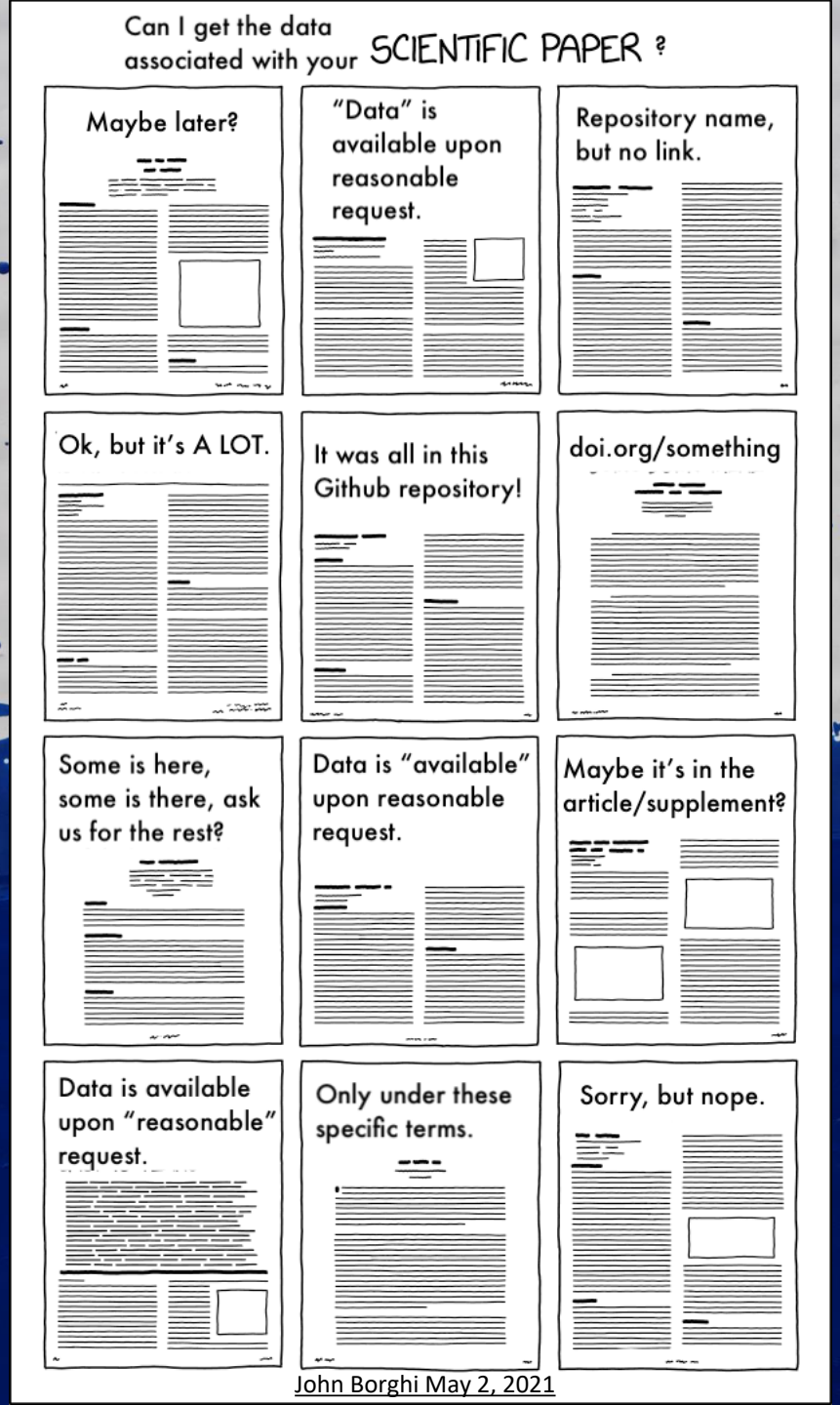
- Avoids duplication of effort in data creation
- Provides resources for use in teaching and learning
- Meets funder requirements
- Ensures data can be re-visited for future research
- Maximises return on research investment
- Preparing data for sharing also prepares it well for preservation

UNA RICERCA MIGLIORE
- INTEGRITÀ
- DIBATTITO
- RIUSO INEDITO

UN MAGGIORE IMPATTO
- VISIBILITÀ
- CREDITO
- COLLABORAZIONI

PIÙ VALORE
- EVITA DUPLICAZIONI
- MASSIMO RITORNO SU INVESTIMENTI

La realtà



John Borghi May 2, 2021

People will contact me to ask about stuff

Christopher and Alex (C&A) say: "This is usually
objection of people who feel overworked and th
[data sharing] isn't part of their job…" I would
learning from each other – if a researcher is opp
datasets, collaborating with others, and genera
they should be outed by their community as a p

COMUNICAZIONE SCIENTIFICA= GRANDE CONVERSAZIONE

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that mis r the help." From the UK Data Archive: "P ding contextual information for your res chers to correctly use and understand your

IMPOSSIBILE, SE NELLA «R» DI FAIR AVETE DOCUMENTATO

It's worth mentioning, however, a s actually be useful to counter willful misrepr Freedom of Information legislation), as one c refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how intere hav people that care about them." I'd neth dataset has value to future researc be "climate change" was a research t to documenting and understanding

EHM… ALLORA PERCHÉ USA FONDI PUBBLICI?

I might want to use it in a research paper

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

EMBARGO PERFETTAMENTE COMPATIBILE CON FAIR

C&A suggest the embargo route: "One embargo; require people to archive the es public after X months. You could even g ly things that are no longer cared about b eventually everything can become ope e of their datasets, but I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository.

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

I'm not sure I own the data

My data is too complicated.

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

IMPOSSIBILE, SE NELLA «R» DI FAIR AVETE DOCUMENTATO

My data is embarrassingly bad

C&A: "Many eyes will help you imp people will accept your data for what it is." I a ck end of making the sausage. We know it's accept that. Plus it helps you strive will be at m r next collection phase.

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

It's not a priority and I'm busy

CARLY STRASSER
Open Science + Data Sharing Advocate
Home
Blog Posts
Previous Research

http://carlystrasser.net/closed-data-excuses-excuses/

Closed Data… Excuses, Excuses

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

Pro e c

ANSWERS TWO

REASONS NOT TO SHARE DATA

UK Data service p. 11

1/2

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

CI SONO BUONI ARGOMENTI CHE SMONTANO LE PAURE RISPETTO ALLA CONDIVISIONE

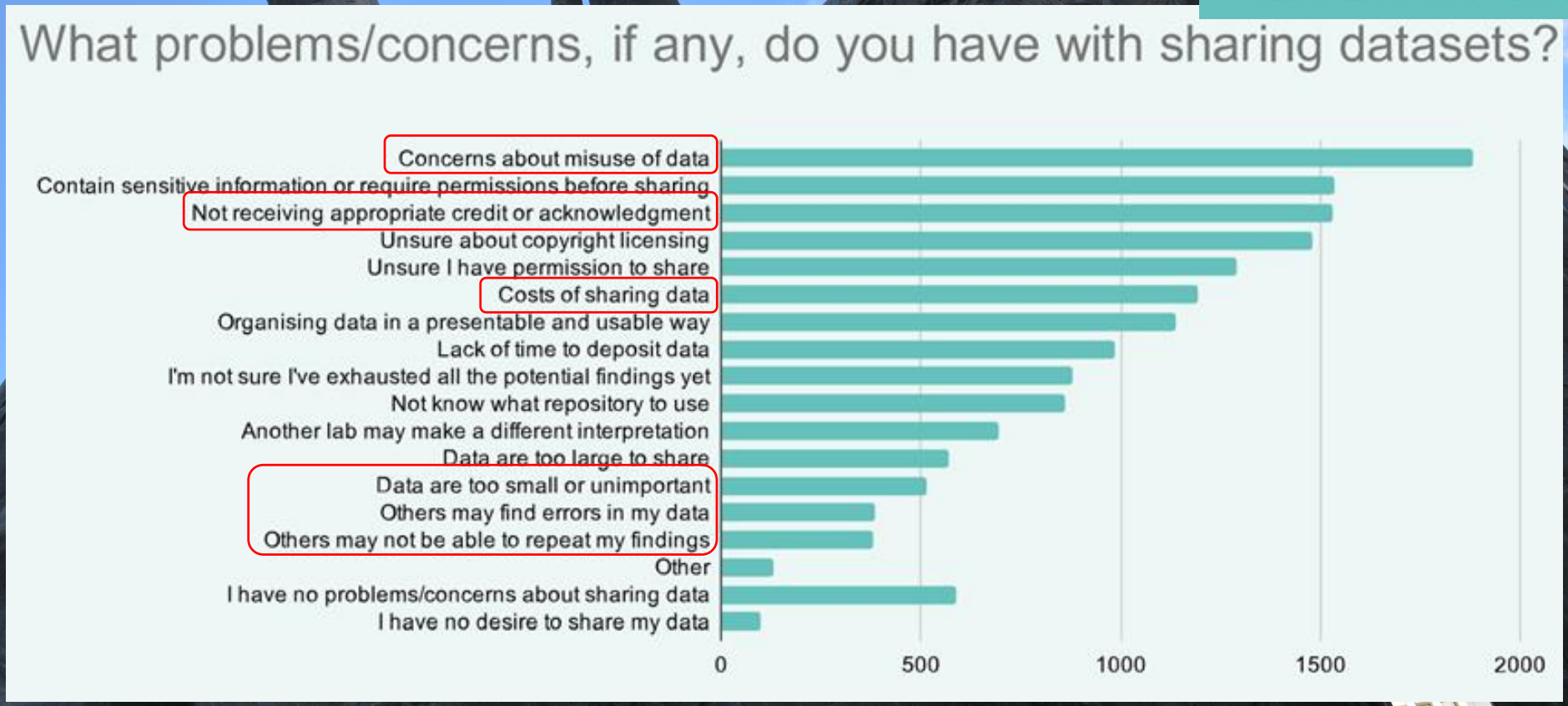
Pro e contro / 2

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

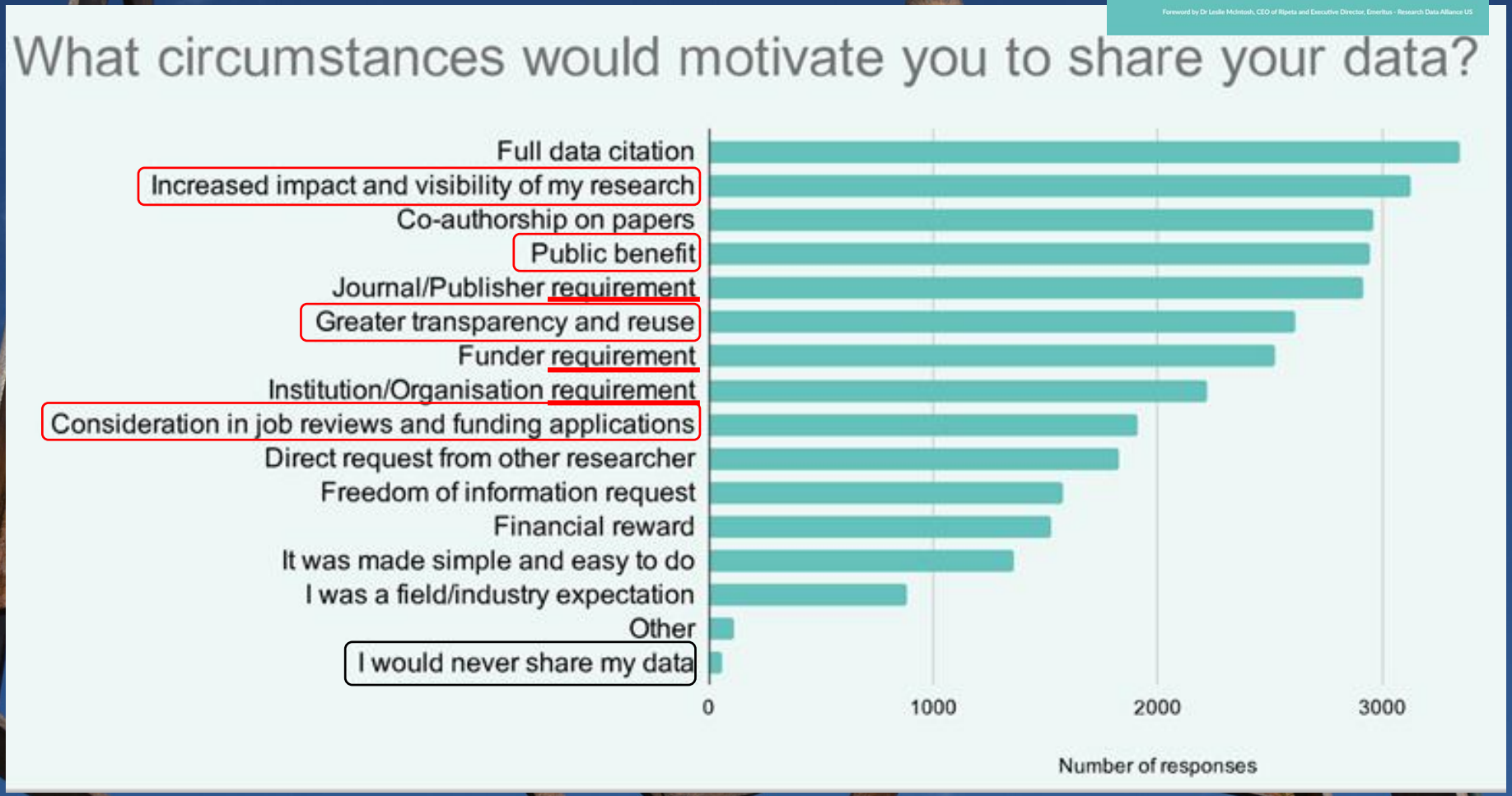
UK Data service p. 11

CI SONO BUONI ARGOMENTI CHE SMONTANO LE PAURE RISPETTO ALLA CONDIVISIONE

[scuse, scuse… aggiornato]



Motivi per condividere



… «as open as possible»…

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