Innovation Acta FOR GlyCanDrug

FAIR data basics

Elena Giglia

Soft Skills courses

Sevilla 3rd conference – November 20th/21st, 2025

FAIR data basics

Innovation Acta

GlyCanDrug

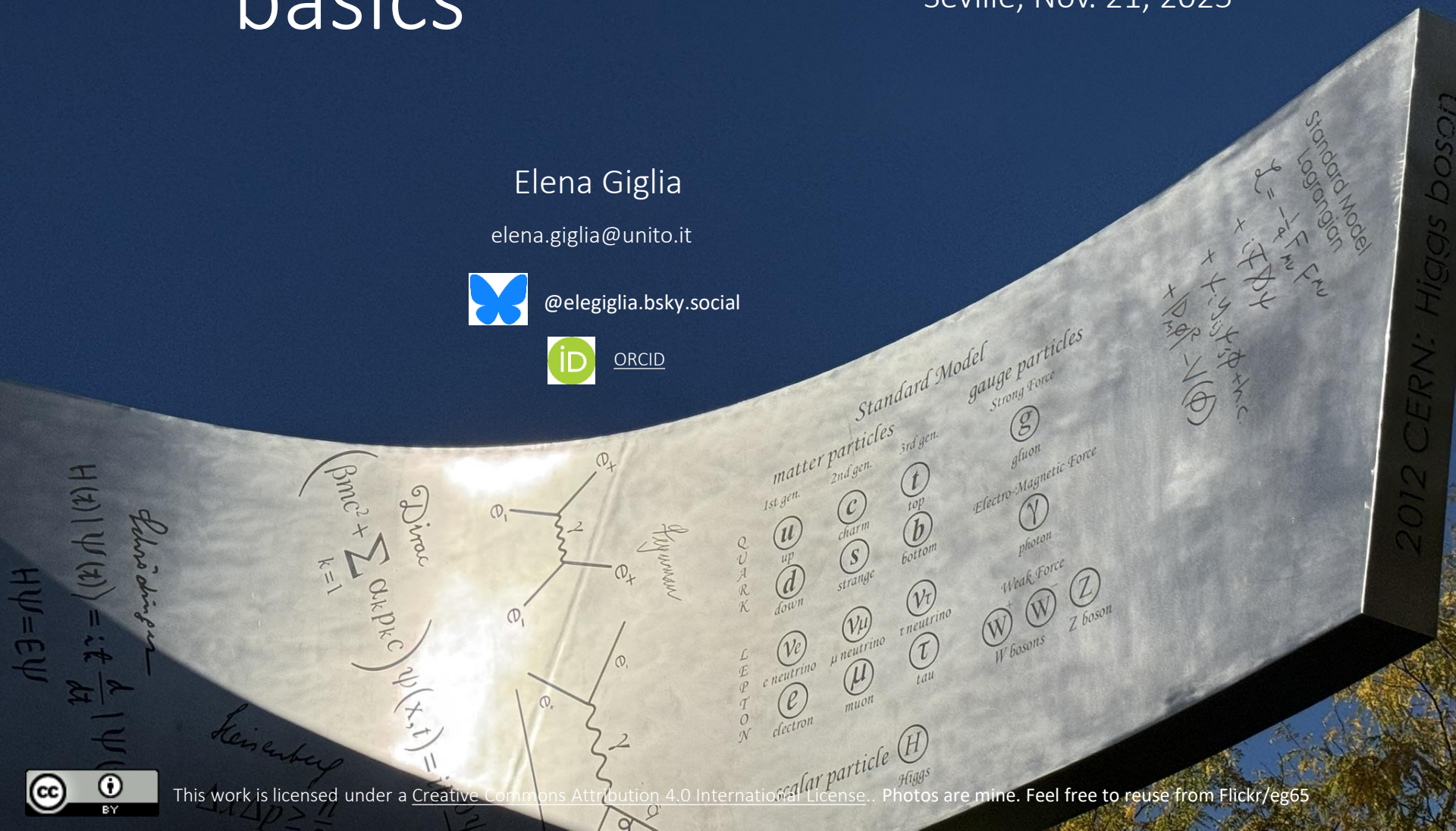
Seville, Nov. 21, 2025

Elena Giglia

elena.giglia@unito.it

@elegiglia.bsky.social

ORCID



This work is licensed under a Creative Commons Attribution 4.0 International License.. Photos are mine. Feel free to reuse from Flickr/eg65

STEP #2
DATA FAIR BY DESIGN

How-to

THIS IMAGE CREATED BY
Noa, Anna, Lilian e Charlotte
PERFECTLY SHOWS WHAT
«MAKING DATA FAIR» MEANS

…please remind 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:

MANAGING DATA IN A RESPONSIBLE WAY ACCORDING TO THE FAIR PRINCIPLES
IS ONE OF THE MANDATORY PRACTICES
YOU NEED TO GIVE AN OVERVIEW OF YOU DMP (1 PAGE) IN THE EXCELLENCE SECTION OF THE PROPOSAL

…FAIR means [for machines]

FAIR guide, Nature, March 2016

FINDABLE

- IDENTIFIERS
- METADATA

ACCESSIBLE

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

INTEROPERABLE

- STANDARDS
- ONTOLOGIES

REUSABLE

- LICENSES
- DOCUMENTATION

MACHINE-READABLE

FAIR principles

Data Intelligence 2020

Issues Online Early About Submit

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

FAIR Principles: Interpretations and Implementation Conside

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronal
Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizza
Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imm
Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastide
Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Philippe Rocca-Serra,
Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn

Previous Article Next Article

Article Contents

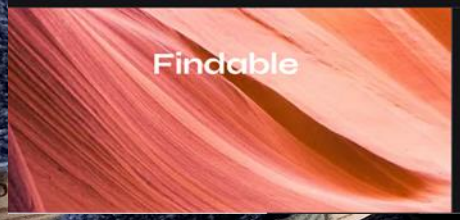
NO MISTAKES!

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

GO FAIR Foundation 2022

Home FAIR Criteria Qualification Fellow

The FAIR Guiding Principles



The first step in (re)using
should be easy to find for
readable metadata are e
datasets and services, sc
the FAIRification process

F1 F2 F3 F4 A1 A1.1 A1.2 A2 I1 I2 I3 R1 R1.1 R1.2 R1.3

Interpretation of F1

Principle F1 states that digital resources, i.e., data and metadata, must be assigned a globally unique and persistent identifier which serves as a permanent machine interpretable reference. The GO FAIR Foundation emphasises the need for persistence and global uniqueness, as well the property of resolvability of the identifiers (see also A1). Globally unique means that the identifier is guaranteed to unambiguously refer to the intended resources (where versal' as there are described digital assets outside the 'world'). Therefore, it is

F1 F2 F3 F4 A1 A1.1 A1.2 A2 I1 I2 I3 R1 R1.1 R1.2 R1.3

Interpretation of A1.2

This principle clearly demonstrates that following the FAIR guiding principles is not equal to making all data 'open'.

ORIGINAL INTERPRETATION

FAIR in a nut

FAIR data training
- Findable
- Accessible
- Interoperable
- Reusable

FAIR for Developers

FAIR data self-assessment tool

+SHARE

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

There are many resources created by the ARDC on the topic of metadata

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different leve

- Persistent identifiers: awareness level
- Persistent identifiers: working level
- Persistent identifiers: expert level

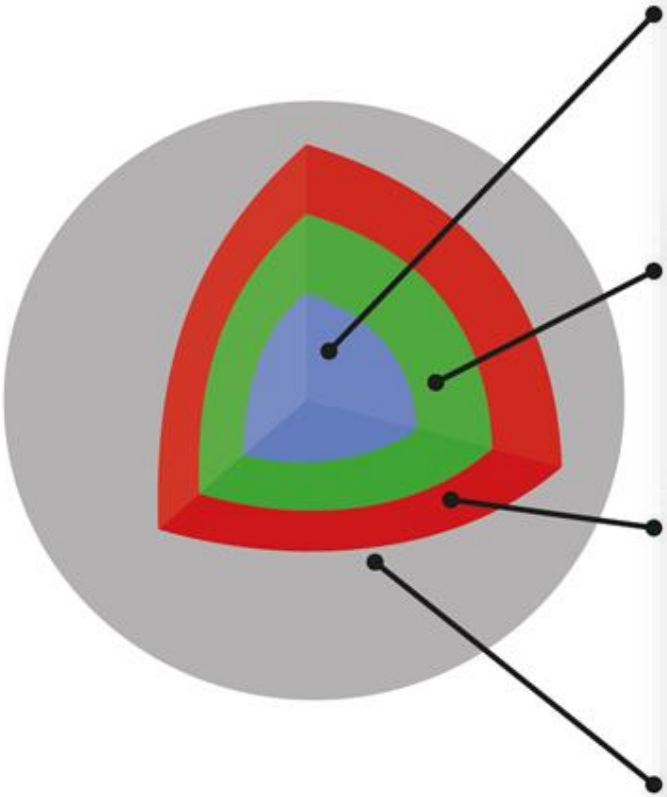
It is also a provider of services for minting persistent identifiers of many di of the data being identified):

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

Complementary to the assignment of persistent identifiers is their proper

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

FAIR object



DIGITAL OBJECT

Data, code and other research outputs

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

IDENTIFIERS

Persistent and unique (PIDs)

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

STANDARDS & CODE

Open, documented formats

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

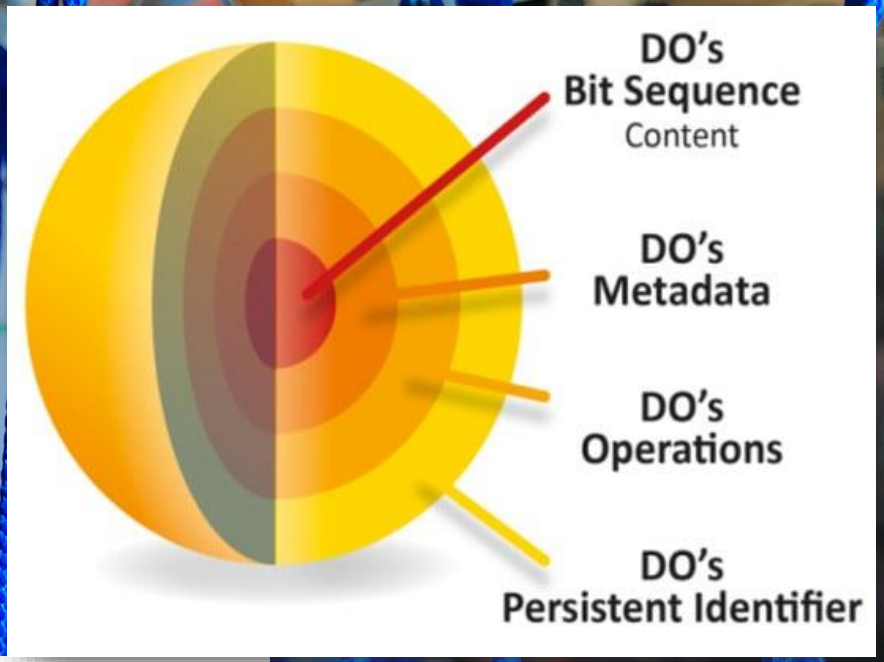
METADATA

Contextual documentation

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

ECD, 2015

FAIR digital object



FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units

2020

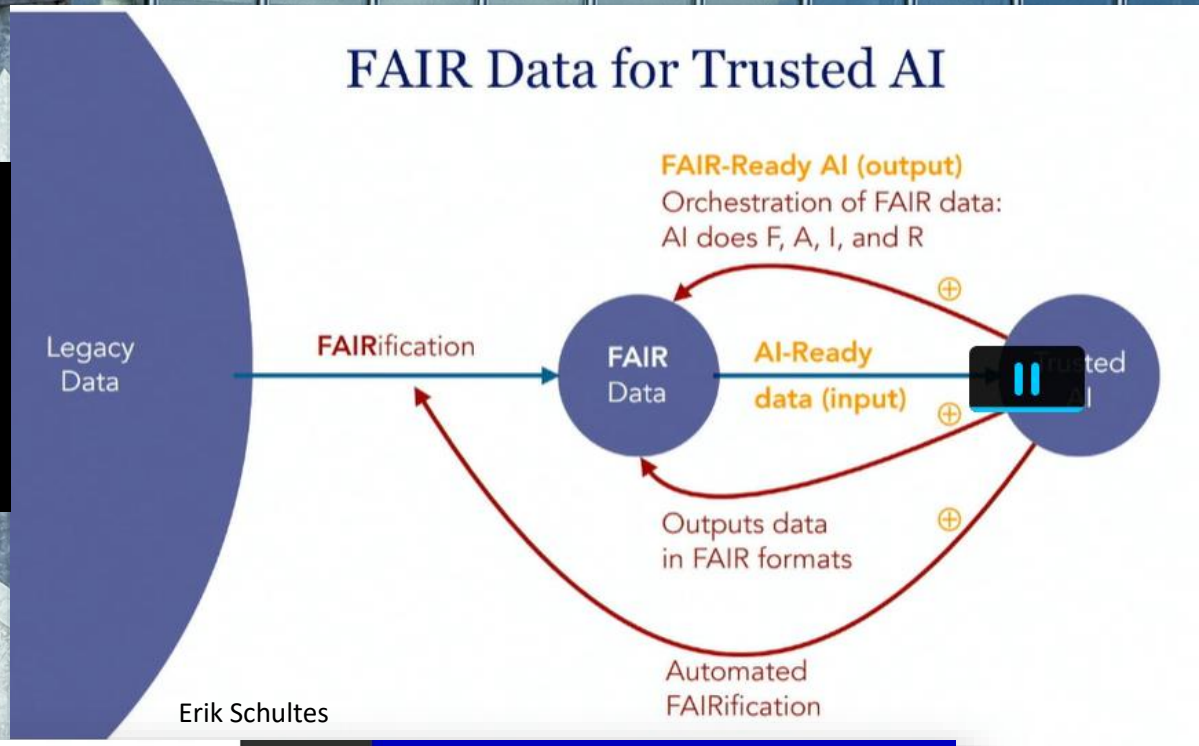
by Koenraad De Smedt [1,*], Dimitris Koureas [2] and Peter Wittenburg [3]

Abstract

Data science is facing the following major challenges: (1) developing scalable cross-disciplinary capabilities, (2) dealing with the increasing data volumes and their inherent complexity, (3) building tools that help to build trust, (4) creating mechanisms to efficiently operate in the domain of scientific assertions, (5) turning data into actionable knowledge units and (6) promoting data interoperability. As a way to overcome these challenges, we further develop the proposals by early Internet pioneers for Digital Objects as encapsulations of data and metadata made accessible by persistent identifiers. In the past decade, this concept was revisited by various groups within the Research Data Alliance and put in the context of the FAIR Guiding Principles for findable, accessible, interoperable and reusable data. The basic components of a FAIR Digital Object (FDO) as a self-contained, typed, machine-actionable data package are explained. A survey of use cases has indicated the growing interest of research communities in FDO solutions. We conclude that the FDO concept has the potential to act as the interoperable federative core of a hyperinfrastructure initiative such as the European Open Science Cloud (EOSC).

AI e FAIR

- FAIR DATA AS INPUT FOR TRUSTED AI
- FAIR AI CREATES FAIR OUTPUTS



Erik Schultes

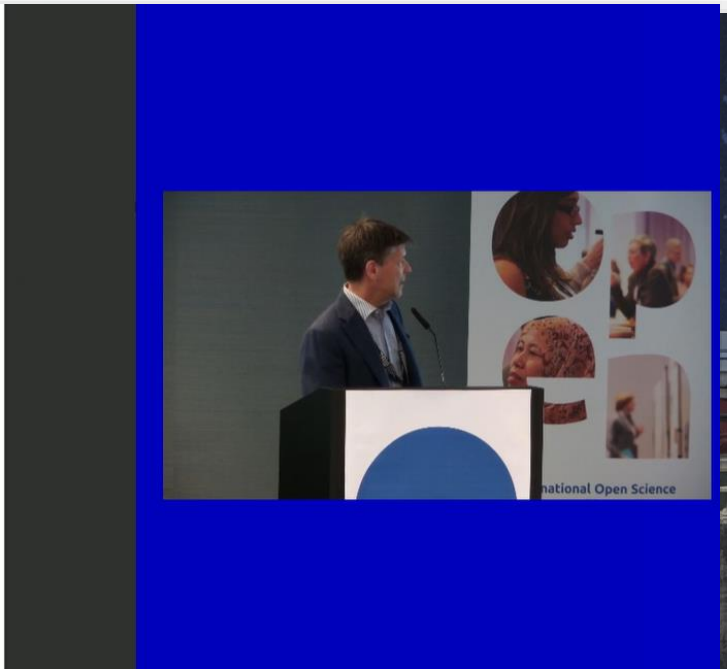
Proposal:

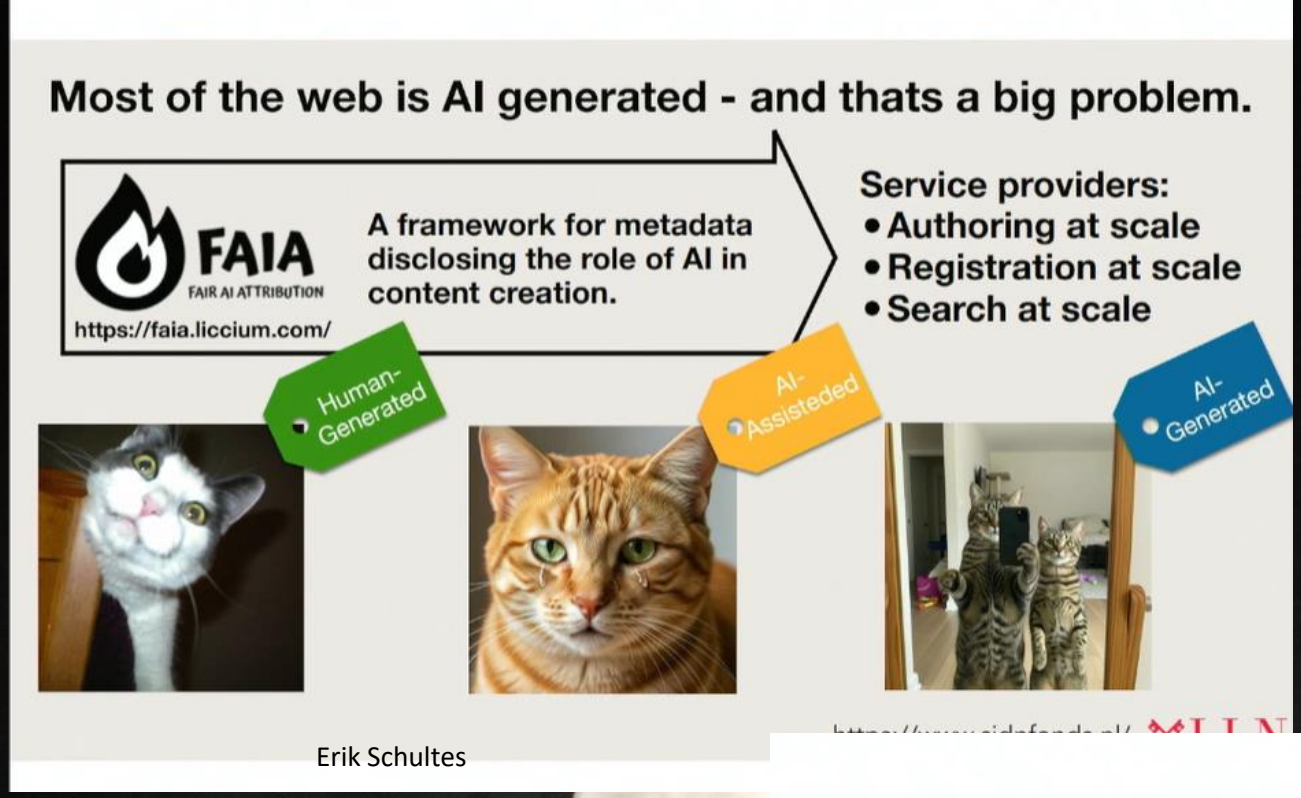
FAIR data as the standard for both AI input and output.

Trusted AI will consume and produce FAIR Data

AI-Ready Data is for FAIR AI inputs

FAIR-Ready AI will create FAIR AI outputs





Erik Schultes

FAIA, UNO STRUMENTO PER DICHIARARE L'USO DI AI NEGLI OGGETTI DIGITALI

Why is FAIA necessary?

https://faia.liccium.com/

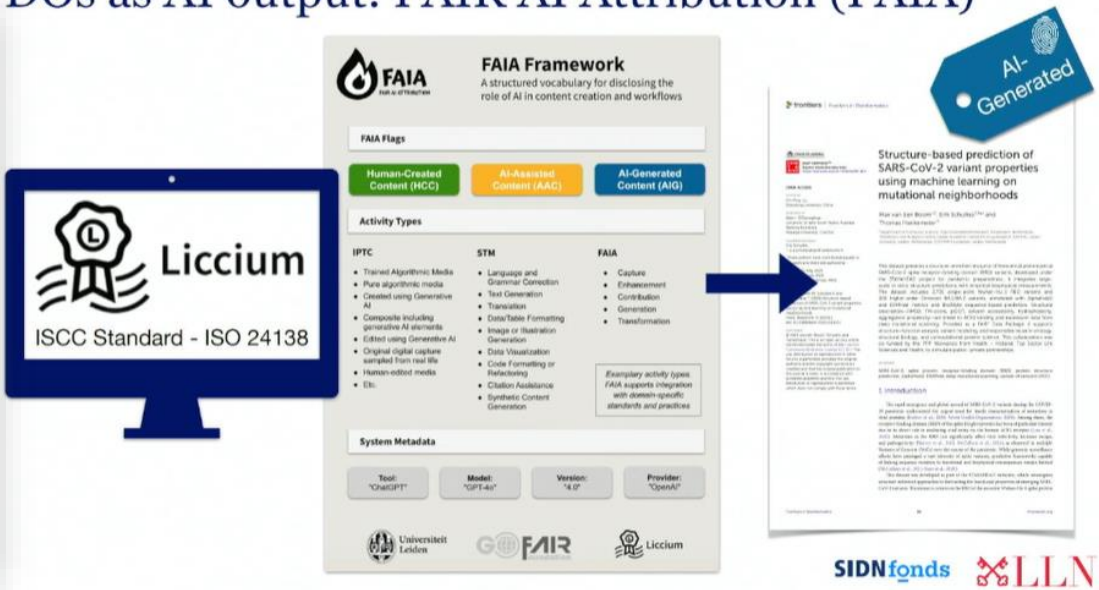
As AI tools increasingly shape the production of text, images, audio, and video, it is becoming harder to distinguish between content made by humans and that produced or modified by machines. This lack of transparency creates serious challenges:

- **Loss of trust** in digital media, journalism, and scientific publishing
- **Legal and regulatory gaps**, particularly under frameworks like the EU AI Act
- **Misrepresentation risks** for researchers, creators, and platforms
- **Inability to trace provenance** or evaluate reproducibility of AI-influenced work

FAIA addresses these issues by making AI involvement visible, verifiable, and portable. It enables compliance with disclosure obligations, supports ethical publishing practices, and builds a foundation for content authenticity and accountability in the age of generative AI.

Erik Schultes

FDOs as AI output: FAIR AI Attribution (FAIA)



FAIR: technology VS domain

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

FAIR GENERIC VS DOMAIN SPECIFIC STRICTLY INTERLINKED

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

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

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

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

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

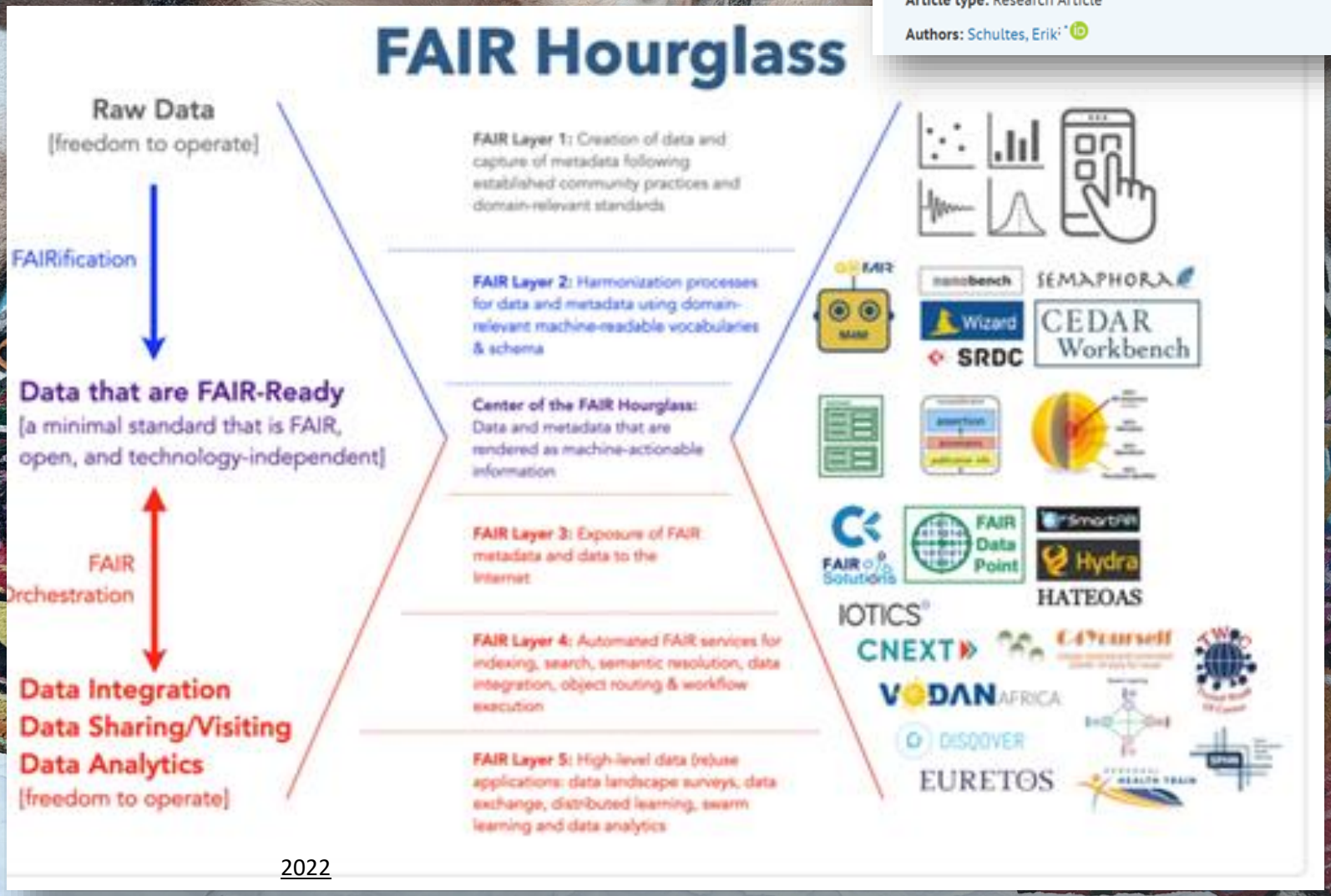
E.Schultes, 2019

Where are we

The FAIR hourglass: A framework for FAIR implementation

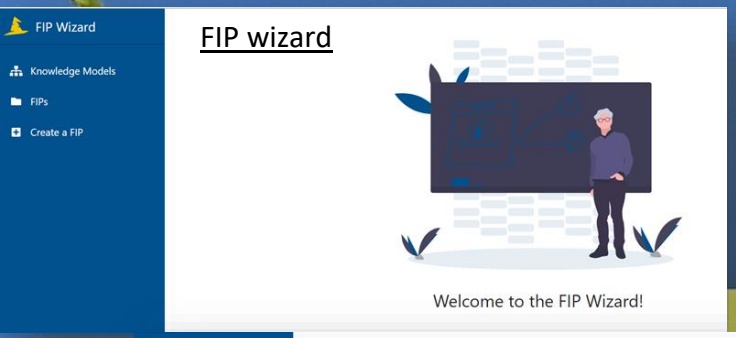
Article type: Research Article

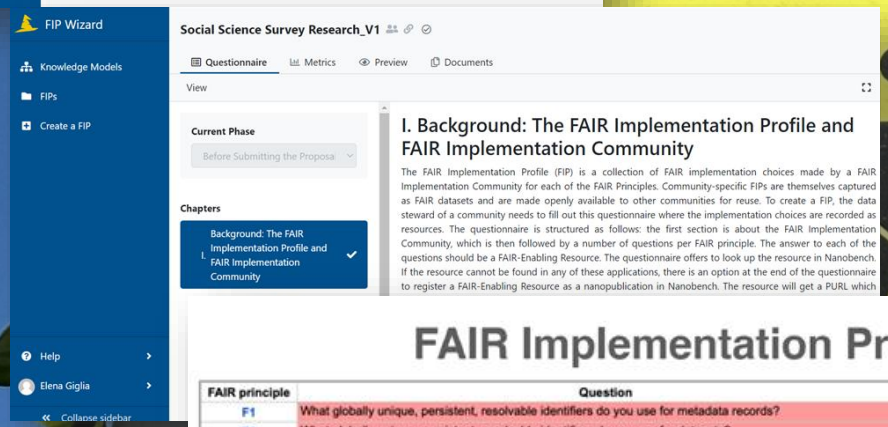
Authors: Schultes, Erik



2022

How / FAIR Implementation profiles





FAIR Implementation Profile

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

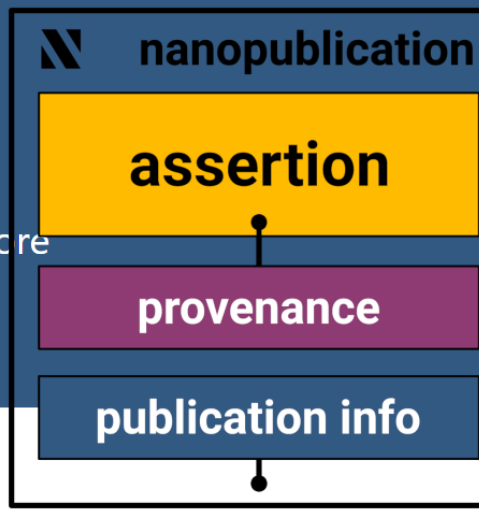
CREATE FAIR IMPLEMENTATION PROFILES REUSBALE BY YOUR COMMUNITY - KEYWORD: CONVERGENCE

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

Nanopublications Documentation Nano Sessions Contact https://nanopub.net/ Search

Nanopublications

FAIR data containers for scientific results, and more



What is a Nanopublication?

A nanopublication is a small knowledge graph snippet with metadata that is treated as an independent (scientific) publication. This information in a nanopublication can be about anything, for example a relation between a gene and a disease or an opinion. Nanopublications are expressed in a knowledge graph format that is formal and machine-interpretable. With nanopublications, it is possible to disseminate individual data as independent publications with without an accompanying research article. Furthermore, because nanopublications can be attributed and cited, they provide incentives for researchers to mal their data available in standard formats that drive data accessibility and interoperability. Nanopublications have the following general structure:

1. **Assertion**: The assertion is the main content of a nanopublication in the form of an small atomic unit of information
2. **Provenance**: This part describes how the assertion above came to be. This can include the scientific methods that were used to generate the assertion, for example a reference to the kind of study that was performed and its parameters.
3. **Publication Info**: This part contains metadata about the nanopublication as a whole, such as when and by whom it was created and the license terms for its reuse.

Nanopublications are implemented in the language RDF and come with an evolving ecosystem of tools and systems. They can be published to a decentralized server network, for example, and then queried, accessed, reused, and linked. To find out more, see the guidelines, the examples, or the published articles about nanopublications.

How to / FAIR connect

FSRs FIPs FICs SIPs

Search... SEARCH

Filter by type (All FSRs selected)

If you want to see a list of all FSRs labelled by a certain Type, go to the **dashboard**

Glossary

FSRs — **FAIR Supporting Resources (FSRs)** are digital objects that support FAIR Data Stewardship, represented as FAIR Digital Objects with Globally Unique, Persistent, Resolvable Identifiers (GUPRI) resolving to machine-readable metadata.

FIPs — **FAIR Implementation Profiles (FIPs)** are collections of FAIR implementation choices made by a community of practice for each of the FAIR Principles.

FICs — **FAIR Implementation Communities (FICs)** are self-identified groups of stakeholders that share a common interest in creating FAIR data and services.

SIPs — **Semantic Interoperability Profiles (SIPs)** are a list of FAIR Supporting Resources chosen by a community to support semantic interoperability of (meta)data.

GFF✓ — **Resources marked with the GFF qualification badge** have been quality-checked by experts trained by the GO FAIR Foundation

FAIR Connect

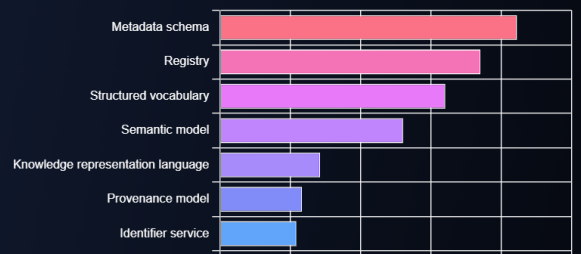
SEARCH FAIR SUPPORTING RESOURCES DASHBOARD ABOUT

Dashboard

1188 Published FSRs

Click on the bars to see a list of FSRs labelled by the selected Type

Sort by: Number of FSRs A-Z



456 Published FIPs

Click on the bars to see FIPs that are used by the community in a Domain

Sort by: Number of FIPs A-Z

FAIR cookbook

RECIPES TO MAKE YOUR DATA FAIR

FAIRcookbook

FAIRCOOKBOOK

GITHUB

The FAIR Cookbook
for FAIR doers

An online, open and live resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable; in one word

FAIRCOOKBOOK

GITHUB

FOREWORD
Introduction
Introducing the FAIR Principles
Reflecting on the ethical values of FAIR
Introducing our FAIRification framework
Prioritizing projects for FAIRification
Framing FAIR and the notion of metadata
Understanding the relation between FAIR and Knowledge Graphs
Training for FAIRification with open or synthetic biomedical datasets
Raising Awareness in Public Knowledge Graphs for Life Sciences
Reflecting on Practical

1. Unique, persistent identifiers

Recipe Overview
Reading Time 30 minutes
Executable Code No
Difficulty

Introducing unique, persistent identifiers
Recipe Type Background information
Audience Principal Investigator, Data Manager, Data Scientist
Maturity Level & Indicator
DSM-1-C0

Cite me with FCB006

The **FAIR principles**, under the Findability and the Accessibility chapters respectively, state that:

> F1. (Meta)data are assigned a globally unique and persistent identifier

FAIRCOOKBOOK

GITHUB

RECIPE

Biomedical datasets of relevance for training in FAIRification

Recipe Overview
Reading Time 15 minutes
Executable Code No
Difficulty

Training for FAIRification with open or synthetic biomedical datasets
Recipe Type Guidance
Audience Principal Investigator, Data Manager, Terminology Manager, Data Scientist
Maturity Level & Indicator

A Accessibility
EXEMPLAR RECIPES
- Transferring data with SFTP
- Downloading data with Aspera
LEARN MORE

I Interoperability
EXEMPLAR RECIPES
- Selecting terminologies and ontologies
- Creating a metadata profile
LEARN MORE

R Reusability
EXEMPLAR RECIPES
- Data licenses
- Declaring data's permitted uses
LEARN MORE

Infrastructure
LEARN MORE

Assessments
LEARN MORE

Applied Examples
LEARN MORE

Maturity model
LEARN MORE

FAIR toolk

FAIR Toolkit

PISTOIA ALLIANCE

Home | FAIR & Life Science Industry | Use Cases | Methods | About | Contact Us

Practical Support for FAIR Data

An overview of how the FAIR Toolkit provides practical support for implementation of FAIR data management through numerous use

Clinical trials

FAIRification of clinical trial data – Roche

Hear more

- Lessons
- Set up in

FAIR4clin

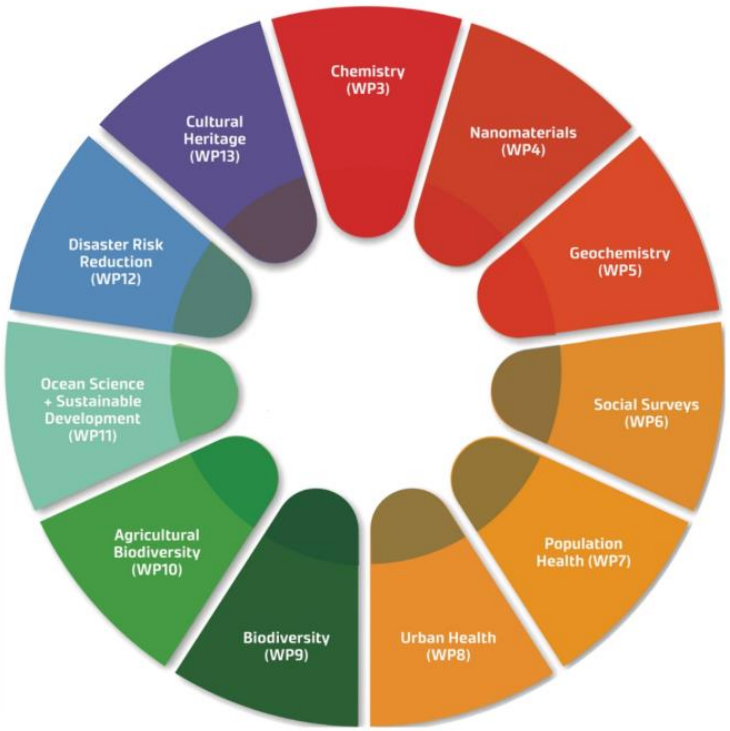
FAIR4Clin

A guide about FAIR for Clinical trial and healthcare data

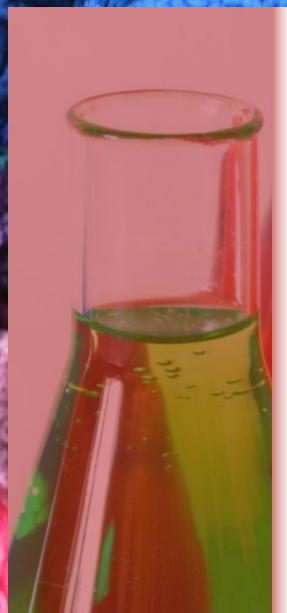
How / FAIRification m

WorldFAIR *disciplinary* and *cross disciplinary* case studies

The core of the WorldFAIR project are the 11 case studies, which represent a wide range of sciences, communities and challenges, with global geographical coverage. Each of the case studies are described below, with links to the relevant items, documents, people, and organisations related to their work.



https://worldfair-project.eu/



Chemistry Featured Outputs

IUPAC FAIR CHEMISTRY PROTOCOL SERVICES

A new service prototype for supporting standard programmatic chemical data exchange and validation is also now available through the **IUPAC FAIR Chemistry Protocol Services** project. Read more.

DIGITAL RECOMMENDATIONS FOR CHEMISTRY FAIR DATA POLICY AND PRACTICE (D3.1)

A review of some of the critical and persistent issues around documentation of chemical information. It also considers documentation requirements to achieve FAIR sharing of chemistry data in ways that are Reliable, Interpretable, Processable, and Exchangeable (RIPE). Read more.

TRAINING PACKAGE: FAIR CHEMISTRY COOKBOOK (D3.2)

UTILITY SERVICES FOR CHEMISTRY STANDARDS (D3.3)

To check your FAIRness

FAIRassist.org https://fairassist.org/#!/

Help you discover resources to measure and improve FAIRness.

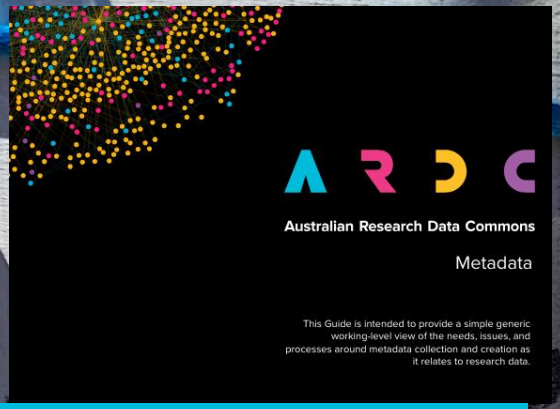
FAIRassist is the new, under development, educational component of the well established FAIRsharing resource.

Resource	Execution Type	Key Features	Organisation	Target Objects	Reading Material
5 Star Data Rating Tool	Manual - questionnaire	Based on rating systems and maturity models	CSIRO OzNome	Datasets	
AutoFAIR	Semi-automated	A portal for automating FAIR assessments for bioinfo	Department of Computer		
Data Stewardship Wizard	Predictive; based on a manually filled questionnaire	Helps researchers to design a data stewardship proc highest reasonable FAIR data.			
F-UJI	Automated	The REST API support a programmatic assessment o objects based on a set of core metrics developed by t metrics specification is available at https://doi.org/10			
FAIR Data Self-Assessment Tool	Manual - questionnaire	Educational and informational purposes			
FAIR Evaluator	Automated	1. Core universal maturity indicators 2. Compliance tests 3. Evaluation tool			
FAIR enough	Automated	1. Core universal maturity indicators and community compliance tests 2. Stable and fast evaluations execution (less than 1min for most evaluated resources, no commercial license required) 3. Library for defining, publishing and registering new maturity indicators 4. Supports ORCID authentication for creating collections and authoring evaluations			Maastricht Uni
FAIR-Aware	Manual - questionnaire	1. Online self-assessment that helps to assess current level of awareness on making datasets FAIR before depositing them in a data repository. 2. Added guidance texts explain the what, why, and how of each FAIR practice. 3. Trainer functionality allows flexible use of the tool for your own purpose			FAIRsFAIR (D
FAIR-Checker	Automated	FAIR-Checker is a web interface to evaluate FAIR metrics (as implemented through the FAIR Evaluation Service APIs https://fairsharing.github.io/FAIR-Evaluator-FrontEnd) and to provide developers with technical FAIRification hints. It's also a Python framework aimed at easing the implementation of FAIR metrics.			IFB (ELIXIR-
FAIRdat	Manual - questionnaire	A 5-star rating of the FAIR principles			DANS
FAIRness self-assessment grids	Manual - checklist	1. Assessment grids: quick and extensive 2. Designed as a decision tree 3. Researcher focused			RDA-SHARC
FAIRshake	Manual - questionnaire,	1. FAIR metrics (questions) and rubrics (collection of metrics)			NIH Data Com

How-to

THIS IMAGE CREATED BY
Noa, Anna, Lilian e Charlotte
PERFECTLY SHOWS WHAT
«MAKING DATA FAIR» MEANS

F = Findable. Metadata



What is metadata?

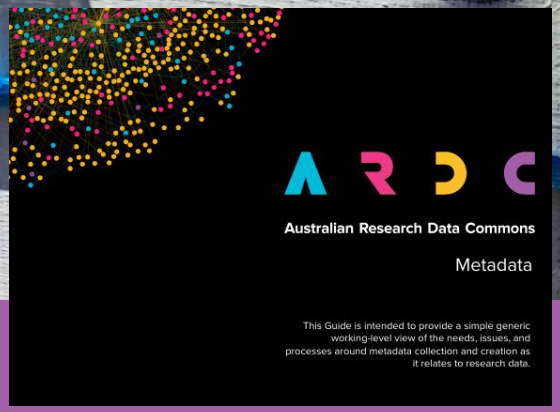
While generally 'meta-data' is summarised as 'data about data', the following provides examples of what this actually means:

- Metadata is information about an object or resource that describes characteristics of that object, such as content, quality, format, location and access rights.
- Metadata can be used to describe physical objects (e.g. pot shards and specimens) as well as digital objects (e.g. documents, images, data sets and software)
- Metadata can take many different forms, from free text (e.g. a read-me file) to standardised, structured, machine-readable, extensible content.
- Metadata is analogous to any other form of data, in terms of how it is created, managed, linked and stored.
- Metadata is associated with the data it describes. It can be embedded within the data file, or recorded a separate text/ spreadsheet file that is linked to the collection of data files it describes, or contained in a catalogue record that points to the research data collection.
- Metadata enables and enhances the discovery and reuse of data.

DATA THAT DESCRIBE OTHER DATA (PHYSICAL OR DIGITAL)

Metadata

F = Findable. Metadata



DIFFERENT LEVELS OF METADATA (COLLECTION, ITEM, SUB ITEM)

Levels of metadata

There are three levels of data groupings: data objects (**Items and Sub-items**) can be collected into groups (**Collections**):

1. **Sub-item level** - sometimes single objects can be made up of component parts: a book has chapters, a documentary has scenes, a relational database has tables, and so on. To search for a particular sub-item component requires not just the item-level metadata but also its sub-item or component level metadata.
2. **Item level** metadata describes individual objects.
3. **Collection level** metadata describes the collection as a whole. For example, Australian Urban Water Collection (https://researchdata.ands.org.au/australian-urban-water-collection/267201).

The collection approach and subsequent level of description impacts on discoverability, and cost and effort of management. Metadata ideally should be based on the needs of those for whom the collection is created.

For more information about collections metadata please read ARDC's Research Data Australia Collections policies (https://documentation.ands.org.au/display/DOC/Research+Data+Australia+Collection+Policies).

Metadata

F=Findable – Metadata

Metadata Standards Catalog Search Sign in

Metadata standars catalog

Metadata Standards Catalog

The RDA Metadata Standards Catalog is a collaborative, open directory of metadata standards applicable to research data. It is offered to the international academic community to help address infrastructure challenges.

Index of subjects

- Multidisciplinary
- Science
  - Atmospheric sciences
    - Climatology
    - Meteorology
  - Biological sciences
    - Biochemistry
      - Biochemicals
        - Proteins
      - Metabolism
    - Biology

OpenRefine Download Documentation Community Blog Donate

Openrefine

OpenRefine

OpenRefine is a powerful free, open source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

Download

Faceting
Drill through large datasets using facets and apply operations on filtered views of your dataset.

Clustering
Fix inconsistencies by merging similar values thanks to powerful heuristics.

Reconciliation
Match your dataset to external databases via reconciliation services.

Infinite undo/redo
Rewind to any previous state of your

Privacy
Your data is cleaned on your machine,

Wikibase
Contribute to Wikidata, the free

FRICTIONLESS DATA Introduction Projects Universe Adoption People Fellows Development Work With Us Blog

Frictionless

Data software and standards

Frictionle
brings sin
whether
engineeri
Why Fric

FAIRCOOKBOOK

FAIRcookbook

FOREWORD
Introduction
Introducing the FAIR Principles
Reflecting on the ethical values of FAIR
Introducing our FAIRification framework
Prioritizing projects for FAIRification
Framing FAIR and the notion of metadata
Understanding the relation between FAIR and Knowledge Graphs
Training for FAIRification with open or synthetic biomedical

11. Creating a metadata profile

Recipe Overview
Reading Time 20 minutes
Executable Code No
Difficulty

Creating a metadata profile
Recipe Type Hands-on
Audience Principal Investigator, Data Manager, Data Scientist
Maturity Level & Indicator
not applicable
Cite me with FCB026

CEDAR

CEDAR

Better metadata means better science

"The CEDAR platform provides an easy-to-use solution for creating and reusing FAIR metadata. CEDAR's metadata modeling, flexible but rigorous semantics, and ability to quickly produce structured metadata makes it perfect for ongoing Metadata 4 Machines workshops and emerging FAIR training courses."

Erik Schultes

F = Findable
Persist

Open Funder Registry (OFR)

https://www.crossref.org/services/funder-registry/

Home > Find a service > Open Funder Registry (OFR)

The Open Funder Registry (OFR, formerly FundRef) and associated funding metadata allows transparency into research funding and its outcomes. It's an open and unique registry of pe giving organizations around the world.

https://ror.org/

ROR

ABOUT SCOPE FACTS SUPPORTERS RESOURCES

Welcome to the Research Organization Registry Community

ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organization in the world.

DataCite

About us Services Res

WELCOME TO DATACITE

with the leading global provider of DOIs for re

Learn more

Cite your research sources with confidence, and receive proper credit when your work is reused.

ted with DataCite!

Search our registry to find datasets, software, images, and other research material.

re3data.org

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

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

https://www.datacite.org/

ORCID

Connecting Research and Researchers

FOR RESEARCHERS | FOR ORGANIZATIONS | ABOUT | HELP | SIGN IN

SIGN IN | REGISTER FOR AN ORCID ID | LEARN MORE

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

We need your feedback! Please tell us about your understanding and perceptions of ORCID and your experience of using your iD by completing our community survey. Thank you!

DISTINGUISH YOURSELF IN THREE EASY STEPS

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

1 REGISTER Get your unique ORCID identifier Register now! Registration takes 30 seconds.

2 ADD YOUR INFO Enhance your ORCID record with your professional information and link to your other identifiers (s as Scopus or ResearcherID or LinkedIn).

LATEST NEWS

Tue, 26 Feb 2019 Construyendo una Infraestructura para Apoyar a los Investigadores -

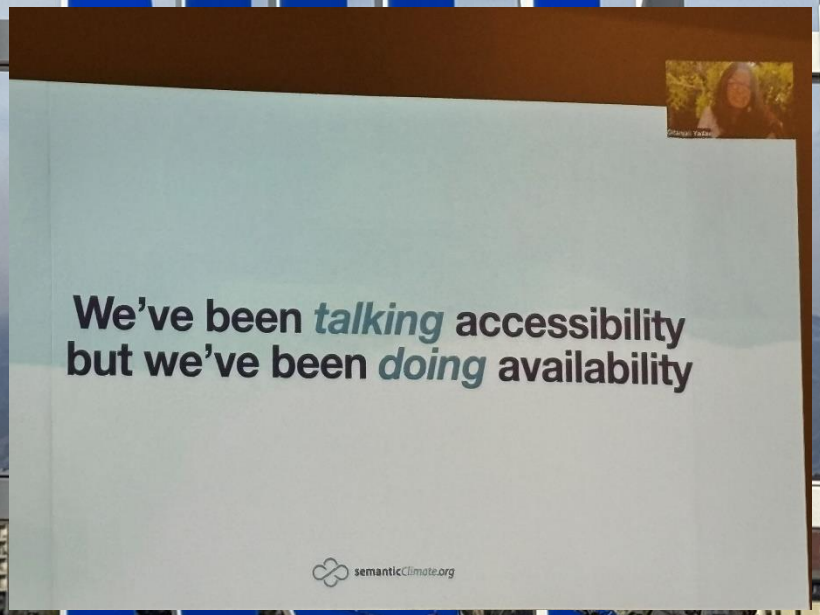
https://orcid.org/

- THINGS: DOI DIGITAL OBJECT IDENTIFIER
- PEOPLE: ORCID
- INSTITUTIONS: ROR
- FUNDERS: OFR

Accessible or Available?

THERE IS A [HUGE] DIFFERENCE

Available ≠ Accessible



zenodo Zenodo Upload Communities

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

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

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

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

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

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

By Kirsty, on 17 August 2021

Guest post by James Houghton – Research Data Support Officer

Why not publish a data report article?

DATA JOURNALS

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

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

Read more

Three new DOI Fabrica features to simplify account management

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

Read more

One step closer towards instant DOI search results

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

Read more

https://www.re3data.org/

[Time's up!

CoARA
COARA

Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is

The Commitments

1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research

2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators

3. Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index

4. Avoid the use of rankings of research organisations in research assessment

TIME'S UP!!!

- THE REFORM OF RESEARCH ASSESSMENT IS ONGOING
- COARA LAUNCHED IN 2022, WORKING GROUPS AND NATIONAL CHAPTERS ACTIVE
- 805 SIGNATORIES [NOV.2024]
- COMMITTIMENT: NO LONGER IMPACT FACTOR OR RANKING

Italy National Chapter

The main aims of the Italian National Chapter are to (i) enable mutual learning, share best practices, and raise awareness of best responsible assessment practices and indicators in the national community on the ongoing research assessment reform (CoARA commitments 7-8), and (ii) foster the discussion about the reviewing and development of assessment criteria, tools and processes for assessing research institutions, individual researchers and projects (CoARA commitment 6).
This outreach effort will support the implementation of the reform at the national level and will contribute to attract more institutions and stakeholders to sign the agreement.
The main activities will be focused on:
1) creating an active network among Italian institutions promoting the alignment of the

Signatories

anvur

Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR)

2024/2027 2024

CoARA Action Plan

Action Plan by ANVUR to implement the ten commitments of the Agreement on Reforming Research Assessment (ARRA)

YES BUT… WE ARE STILL EVALUATED ON IMPACT FACTOR

A = Accessible. Formats and protocols

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

Data Archiving and Networked Services

DANS

HOME DEPOSIT

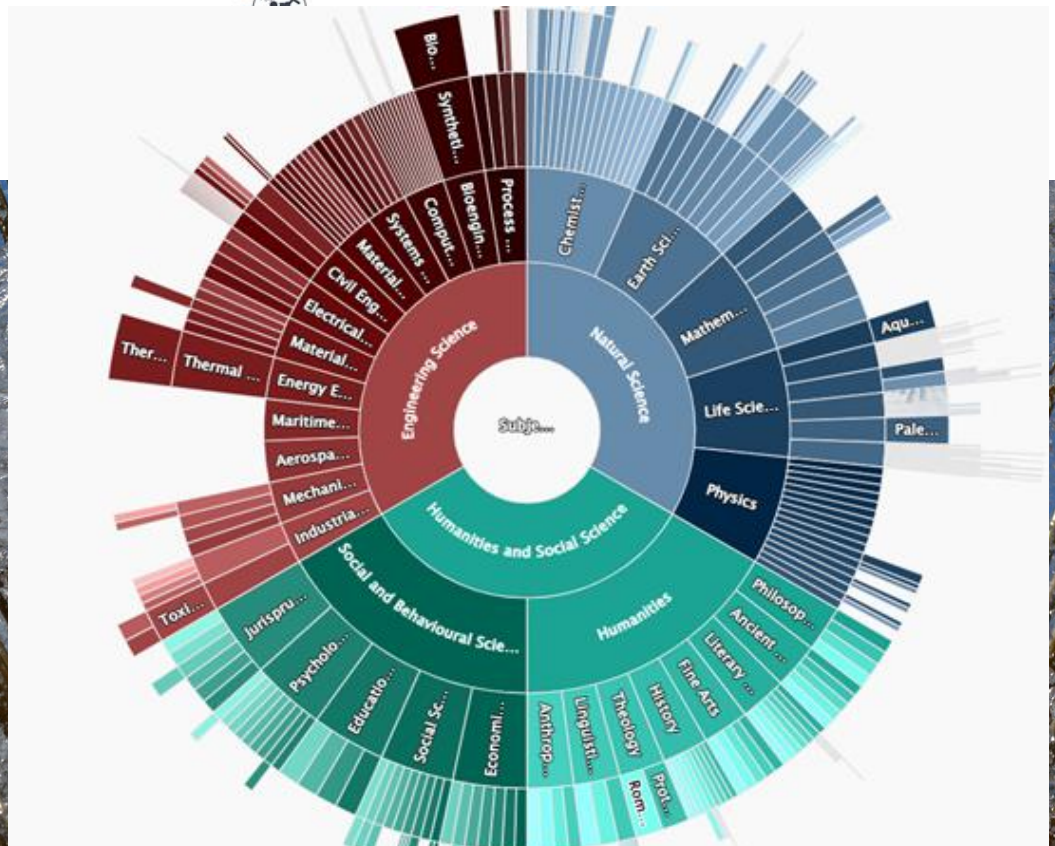
If your data are stored in other formats than those mentioned below, please contact DANS.

Type	Preferred format(s)	Non-preferred format(s)
Text documents	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
Plain text	• Unicode text (.txt)	• Non-Unicode text (.txt)
Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml) • Markdown (.md)
Programming languages	• MATLAB • NetCDF • Text-Fabric • Python	
Spreadsheets	• ODS (.ods) • CSV (.csv)	• Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)

https://dans.knaw.nl/en/file-formats/

Interoperable - standards

https://fairsharing.org/



FAIRSHARING [NEW VERSION] STANDARD REGISTRY

I – Interoperable - ontologies

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

2020

Ontology, Ontologies and the "I" of FAIR

Giancarlo Guizzardi

Check for updates

Author and Article Information

Data Intelligence (2020) 2 (1-2): 181–191.

https://doi.org/10.1162/dint_a_00040

WHY WE NEED ONTOLOGIES FOR THE «I» IN FAIR

Previous Article Next Article

Article Contents

Abstract

1. INTRODUCTION

2. THE DATAVERSE IS A WORLD OF SILOS

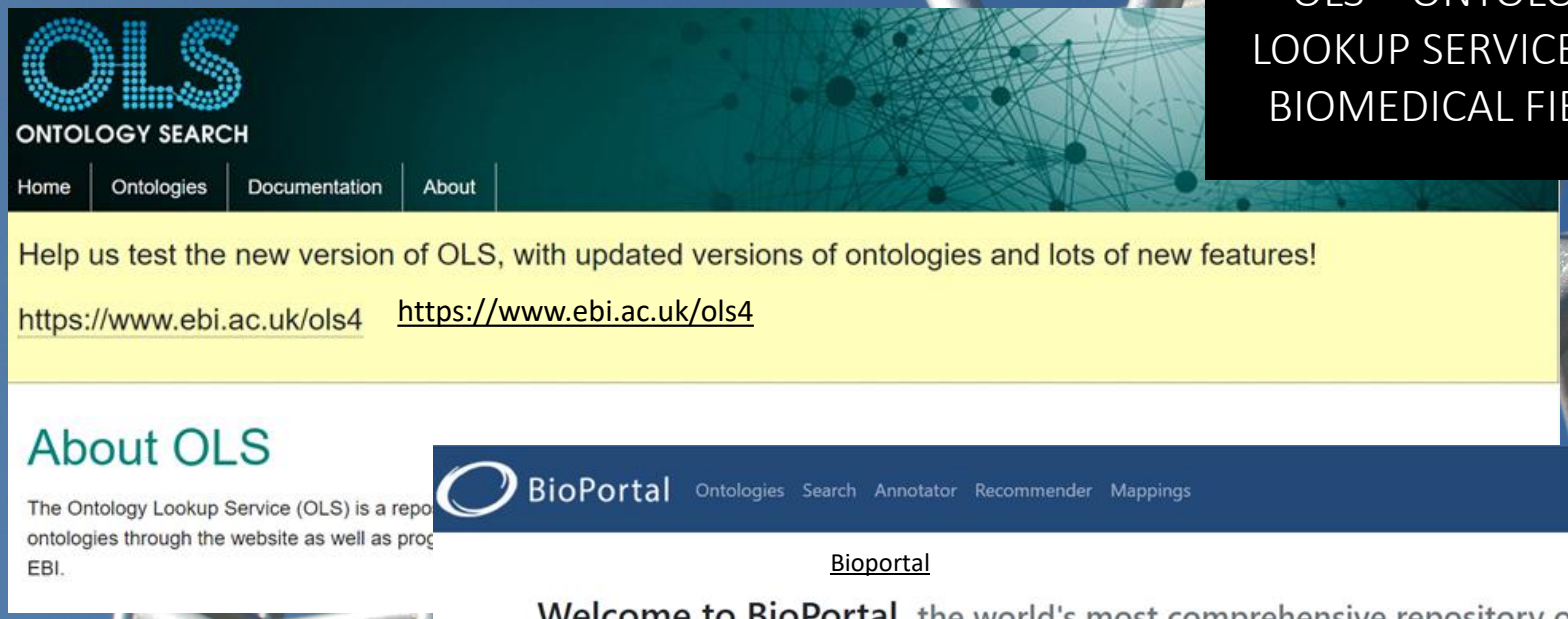
3. INFORMATION STRUCTURES AND SEMANTIC INTEROPERABILITY

Abstract

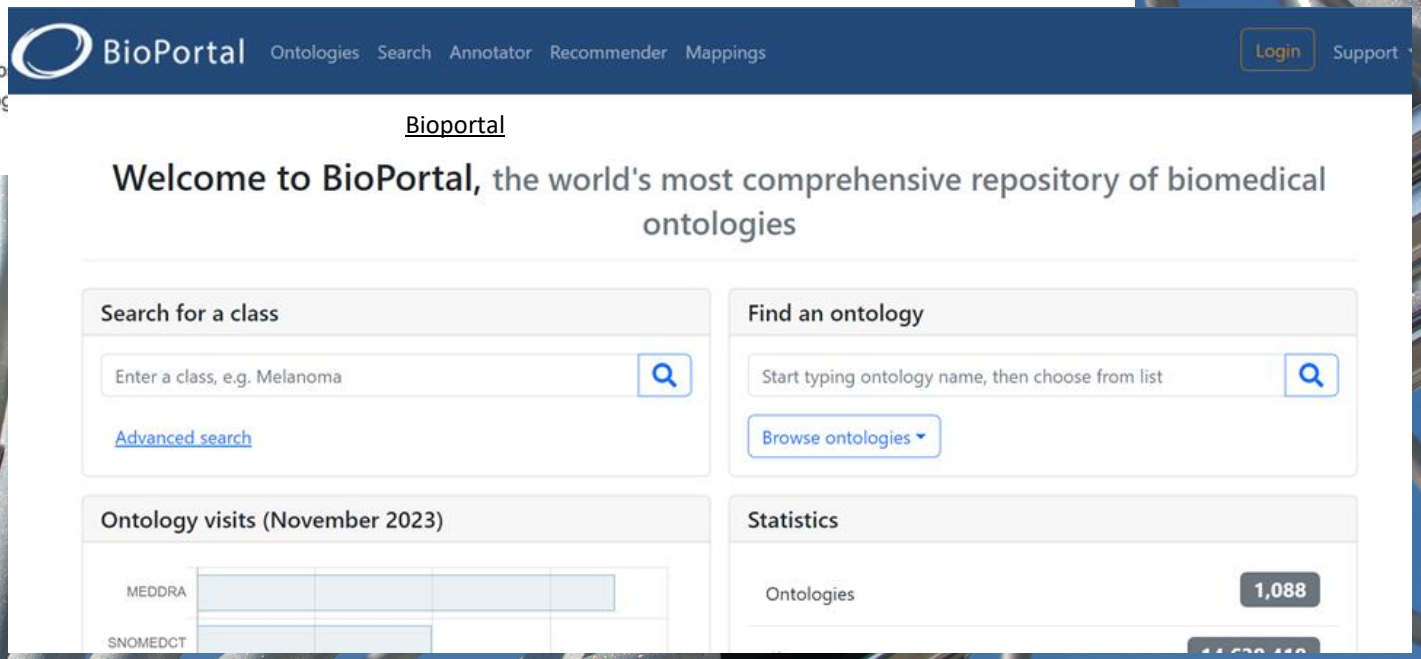
According to the FAIR guiding principles, one of the central attributes for maximizing the added value of information artifacts is interoperability. In this paper, I discuss the importance, and propose a characterization of the notion of *Semantic Interoperability*. Moreover, I show that a direct consequence of this view is that Semantic Interoperability cannot be achieved without the support of, on one hand, (i) *ontologies*, as *meaning contracts* capturing the conceptualizations represented in information artifacts and, on the other hand, of (ii) *Ontology*, as a discipline proposing formal methods and theories for clarifying these conceptualizations and articulating their representations. In particular, I discuss the fundamental role of formal ontological

I = Inteoperable. Ontologies

OLS – ONTOLOGY LOOKUP SERVICE FOR BIOMEDICAL FIELDS



https://www.ebi.ac.uk/ols4



Bioportal

What is an ontology?

Ontologies are frameworks for representing shareable and reusable knowledge across a domain.

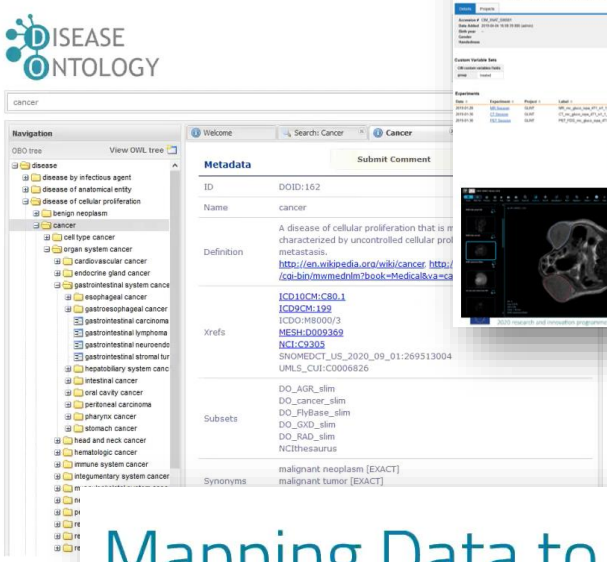
Their ability to describe relationships and their high interconnectedness make them the bases for **modeling high-quality, linked and coherent data**.

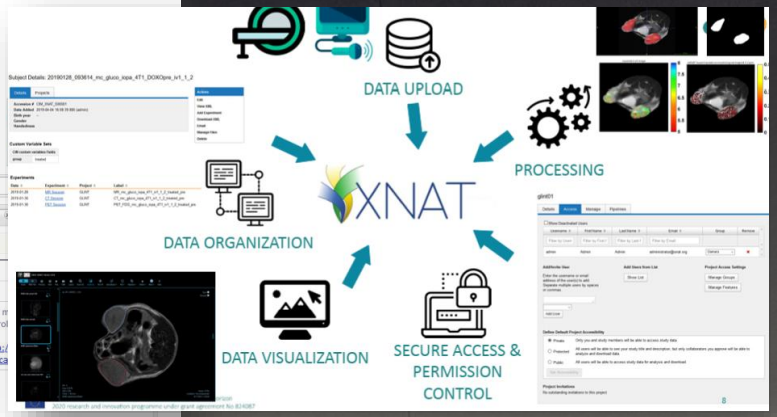
Some examples:

- Human Disease Ontology (DO)
- Experimental Factor Ontology (EFO)
- Uber-anatomy ontology (UBERON)
- Chemical Entities of Biological Interest (CHEBI)
- ...

What ontologies are for?

- Different data sources that have been annotated with ontologies are easier to integrate, instead of free text!





Mapping Data to Ontologies

	CHEMICAL		
Intervention sub	Antineoplastic agent	CHEBI:35610	antineoplastic agent
	Antiangiogenic agent		
	Anti-inflammatory agent	CHEBI_67079	Anti-inflammatory agent
	Anesthetic agent	CHEBI:38867	anaesthetic
	Contrast agent	CHEBI:37334	diagnostic imaging agent
	Tyrosin-Kinase Inhibitors (TKI)	CHEBI:38637	tyrosine kinase inhibitor
	Antibiotic agent	tbc	
	Other...		
Drug	Esomeprazole	CHEBI:50275	Esomeprazole
	Imatinib	CHEBI_45783	Imatinib
	Heparin	CHEBI:28304	Heparin
	Lansoprazole	CHEBI_6375	Lansoprazole
	Amiloride	CHEBI_2639	Amiloride
	Dichloroacetate	CHEBI_28240	Dichloroacetate
	Doxorubicin	CHEBI_28748	Doxorubicin
	Regorafenib	CN518174	Regorafenib
	Cariporide	NCIT_C76028	Cariporide
	Bicarbonate	NCIT_C37918	Bicarbonate Ion
	Saline	EFO_0002677	Saline
	Isoflurane	CHEBI_6015	Isoflurane
	Sevoflurane	CHEBI_9130	Sevoflurane
	Xylazine (Rompun)	NCIT_C75049	Xylazine
	Tiletamine/Zolazepam (Zoletil)	tbc	
	Rimadyl	tbc	
	Topiramate	CHEBI_63631	Topiramate
	Other...		
Contrast Agent	Glucose	CHEBI_17234	Glucose
	3-O-methyl-D-glucose (3OMG)	CHEBI_73918	3-O-methyl-D-glucose
	Fluorodeoxyglucose F 18 (18F-FDG)	NCIT_C964	Fludeoxyglucose F-18
	Gadoteridol	CHEBI_31643	Gadoteridol
	Iopamidol	CHEBI_31711	Iopamidol
	Gadolinium complexes	NCIT_C39765	Gadolinium complexes

Sara Zullino in Giglia, E.
FAIR DATA Basics 3rd Module (ISPAS)_part_2.mp4 (video)

From 1.40.00

https://doi.org/10.5281/zenodo.5537468

R = Reusable. Documentation

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

RDM elixir

RDMkit | Data management | About | Contribute

Data management
- Data life cycle
- Your role
- Your domain
- Your tasks
  - Compliance monitoring
  - Costs of data management

Your tasks

Data provenance

How to document and track your data provenance?

Description

Provenance is the documentation of why and how the data (but also datasets, computational ana and other research output) was produced, where, when and by whom. Data provenance is often definition might slightly differs in sor

CESSDA guide

cessda TRAINING

Project-level documentation

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

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

RDM Elixir

RDMkit | Data management | About | Contribute | Gi

Data management
- Data life cycle
- Your role
- Your domain
- Your tasks
  - Compliance monitoring

Your tasks

Documentation and metadata

How can you document data during the project?

ELN

Electronic Lab Notebooks

Collect & Analyze

Using an Electronic Lab Notebook

— Legal as

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?

How can a protected dataset be used? +

Where are licences found? +

Interoperability and stacking +

What happens if I use 'Share Alike' (SA) licensed material in my work? Does that mean I have to make my work available under the same SA licence? +

Can a dataset be used if there is no licence? +

What are the risks of using a dataset without a licence? +

Training materials +

OpenAIRE Can I use SERVICES

Guides for Researchers

Can I reuse someone else's research data?

Learn more on how to reuse research data

R – Reusable – Legal aspects

2024

Supporting eosc

Skills 4 eosc

D3.7 - Coordinated set of guides, fact-sheets and FAQs on ELSI aspects for Civil Servants and Policy Makers

COMPREHENSIVE GUIDE ON LEGAL ASPECTS

Skills 4 eosc

Coordinated set of guides, fact-sheets and FAQs on ELSI aspects for Civil Servants and Policy Makers

Contents

1 Introduction 5
1.1 Background: Challenges facing Policy Makers and Civil Servants in the connected age 5
1.2 Links to relevant Minimum Viable Skillset career profiles 7
1.3 Deliverable overview 9
2 Personal Data Protection 12
2.1 Fact Sheets 12
2.2 FAQ 15
2.3 Useful terms 19
3 Intellectual Property Rights and other non-personal data. 22
3.1 Fact Sheets 22
3.2 FAQ 23
4 Ethics in research and open science 62
4.1 FAIR Principles 63
4.2 Responsible Research and Innovation (RRI) 65
4.3 European Code of Conduct for Research Integrity 66
4.4 Responsible Open Science for Europe (ROSiE) 68
5 Conclusions 70
6 References 71
Appendix A: Infographs 78
Appendix B: Practical materials and templates 79

2023

Supporting eosc

Skills 4 eosc

Deliverable 3.6.

Coordinated set of guides, fact-sheets and FAQs on ELSI aspects for Civil Servants and Policy Makers

What are High-Value Datasets?
According to art 2(10) of the ODD

> means documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets.

According to the Annex I of the ODD, the current thematic categories for the high-value datasets are the following: (i) Geospatial; (ii) Earth observation and environment; (iii) Meteorological; (iv) Statistics; (v) Companies and company ownership; (vi) Mobility. "Such specific high-value datasets shall be":

> (a) available free of charge, subject to paragraphs 3, 4 and 5;
> (b) machine readable;

FAQs on the following topics:

- Personal Data Protection;
- Intellectual property and other non-personal data protection, including the following:
  - Intellectual Property (general aspects);
  - Copyright;
    - Intersection between Copyright and Open Science;
    - Licensing;
  - Sui Generis Database Rights;
  - Trade Secrets;
  - Non-personal data and Public Sector Bodies;
    - Data under the Open Data Directive;
    - Data under the Data Governance Act;
    - Data under EU-Funded Projects.

What is the Open Data Directive?
The Open Data Directive[43] aims to "promote the use of open data and stimulate innovation in products and services" by setting "minimum rules governing the re-use and the practical arrangements for facilitating the re-use of" certain data.[44]

Which kind of data is covered by the ODD?
The data covered by ODD are the following:

> (a) existing documents held by public sector bodies of the Member States;
> (b) existing documents held by public undertakings that are:
> (i) active in the areas defined in Directive 2014/25/EU;
> (ii) acting as public service operators pursuant to Article 2 of Regulation (EC) No 1370/2007;
> (iii) acting as air carriers fulfilling public service obligations pursuant to Article 16 of Regulation (EC) No 1008/2008; or
> (iv) acting as Community shipowners fulfilling public service obligations pursuant to Article 4 of Regulation (EEC) No 3577/92;
> (c) **research data** pursuant to the conditions set out in Article 10. (emphasis added)

What is understood as 'research data' under the ODD?

Do CC licenses cover other IP rights or personality rights?
No. Generally, CC licenses are meant to deal with copyrights and, in the case of the version 4.0, sui generis database rights (SGDR). As stated by the Creative Commons, '[t]he license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.'[34] Also, section 2(b)(1)(2) of the CC-BY 4.0. Provides that:

> 'Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise. Patent and trademark rights are not licensed under this Public License.'

3.2.3 Sui Generis Database Rights (SGDR)

What are the sui generis database rights?
Despite being in the same Directive as the copyright protection to databases, these rights are different in their scope and justification.

STEP #3
OPEN DATA

UNESCO, AI and Open data

OPEN DATA, A MUST FOR AI

In order to fight the epidemic, timely, relevant, and high-quality open data w
globally shared. This cooperative and efficient undertaking could serve as a bl
print for a better and more sustainable future for all.

Having raw data is the first essential step in processing and transforming them i
actionable information. But the required data need to have certain characterist
such as being accurate, timely and reliable. **Further critical aspects of data are t
they are findable, accessible, interoperable and reusable (FAIR) by anyone for a
purpose.**

The aim of these guidelines is to apprise Member States of the value of op
data, and to outline how data are curated and opened. These guidelines spe
concrete steps that Member States can take to open their data in three phase
Prepare, Open, Follow-up

These guidelines follow up on the **UNESCO Recommendation on the Ethics of Artificial Intelligence**, which, among other topics, includes a call for open data for AI. These guidelines will also play a crucial role in supporting the **UNESCO Recommendations on Open Science** by facilitating data sharing, enhancing reproducibility and transparency, promoting data interoperability and standards, supporting data preservation and long-term access.

120

Countries without Open Data policies

unesco

2023

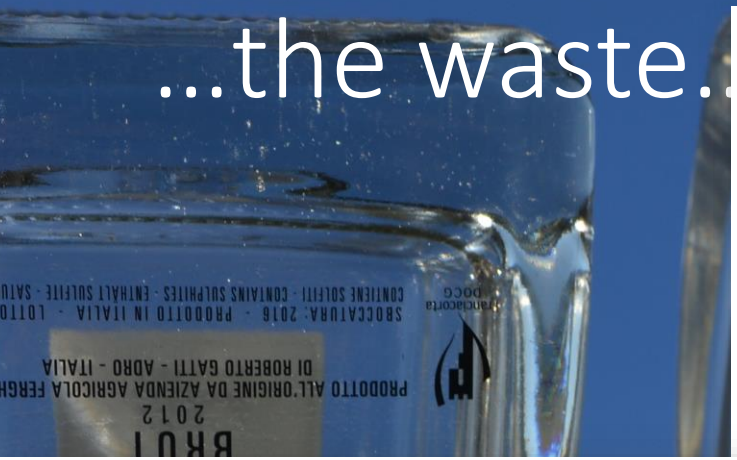
Open data for AI

What now?

1. DATA IN THE FIGHT OF A GLOBAL PANDEMIC - A CASE STUDY
   - 1.1 INTRODUCTION
   - 1.2 AI CHALLENGES
   - 1.3 DATA CHALLENGES
   - 1.4 WAY FORWARD
   - 1.5 SUMMARIZED LESSONS LEARNED TOWARDS OPEN DATA
2. VISION
   - 2.1 INTRODUCTION
   - 2.2 OPEN DATA
     - 2.2.1 BACKGROUND
     - 2.2.2 DEFINITION, PRINCIPLES, FIVE STAR PLAN AND LICENCES
     - 2.2.3 CATEGORIZATIONS OF OPEN DATA
     - 2.2.4 PROS AND CONS OF OPEN DATA
     - 2.2.5 EXISTING GENERAL OPEN DATA INITIATIVES AND RESOURCES
       - 2.2.5.1 SDG RELATED OPEN DATA INITIATIVES
       - 2.2.5.2 GOVERNMENT OPEN DATA INITIATIVES
       - 2.2.5.3 SCIENCE OPEN DATA INITIATIVES
       - 2.2.5.4 PRIVATE OPEN DATA INITIATIVES
       - 2.2.5.5 COVID-19 OPEN DATA INITIATIVES
     - 2.2.6 INDICES AND BAROMETERS
   - 2.3 AI AND OPEN DATA
     - 2.3.1 BACKGROUND
     - 2.3.2 SYNERGIES BETWEEN AI AND OPEN DATA TO TACKLE GLOBAL CHALLENGES
     - 2.3.3 POTENTIAL INCREASED AI RISKS DUE TO OPEN DATA
FURTHER TOPICS
3. GUIDELINES
   - 3.1 PREAMBLE
   - 3.2 PREPARATION
   - 3.3 OPEN THE DATA
   - 3.4 FOLLOW UP FOR REUSE AND SUSTAINABILITY

➢ Create an open-data-driven culture

Experience has shown that in addition to opening data legally and
technically at least two more things have to be accomplished to reach
an open-data-driven culture: Often staff, especially in Governmen
departments, are not used to sharing data, but rather have been
trained to keep a silo mentality. Moreover, data should become, i
possible, the exclusive basis for decision-making; in other words
decisions should be anchored in data. Both topics are also raised by
"Step 3: Develop open data capacities", yet in addition to capacities
cultural changes are required by all involved staff. There is a need
to foster proactive data disclosure, which can guarantee that data is
made available even before a demand for the data is placed. This can
be ensured through the inclusion of such provisions in the Access to
Information Laws.[104]



...the waste...

Unshared Data: The Unjustifiable Carbon Footprint of Science

2024

Sean Hill · Follow

8 min read · Apr 2, 2024

The Necessity of Sustainable Data Preservation

Confronting the environmental footprint of scientific research brings the critical role of sustainable data preservation into stark relief. The enormous consumption of energy and resources for generating scientific data underscores an urgent need: not just to preserve this data but also to ensure it's accessible for future research. This strategy is crucial, both for maximizing the return on our initial resource investment and for significantly reducing environmental impacts down the line.

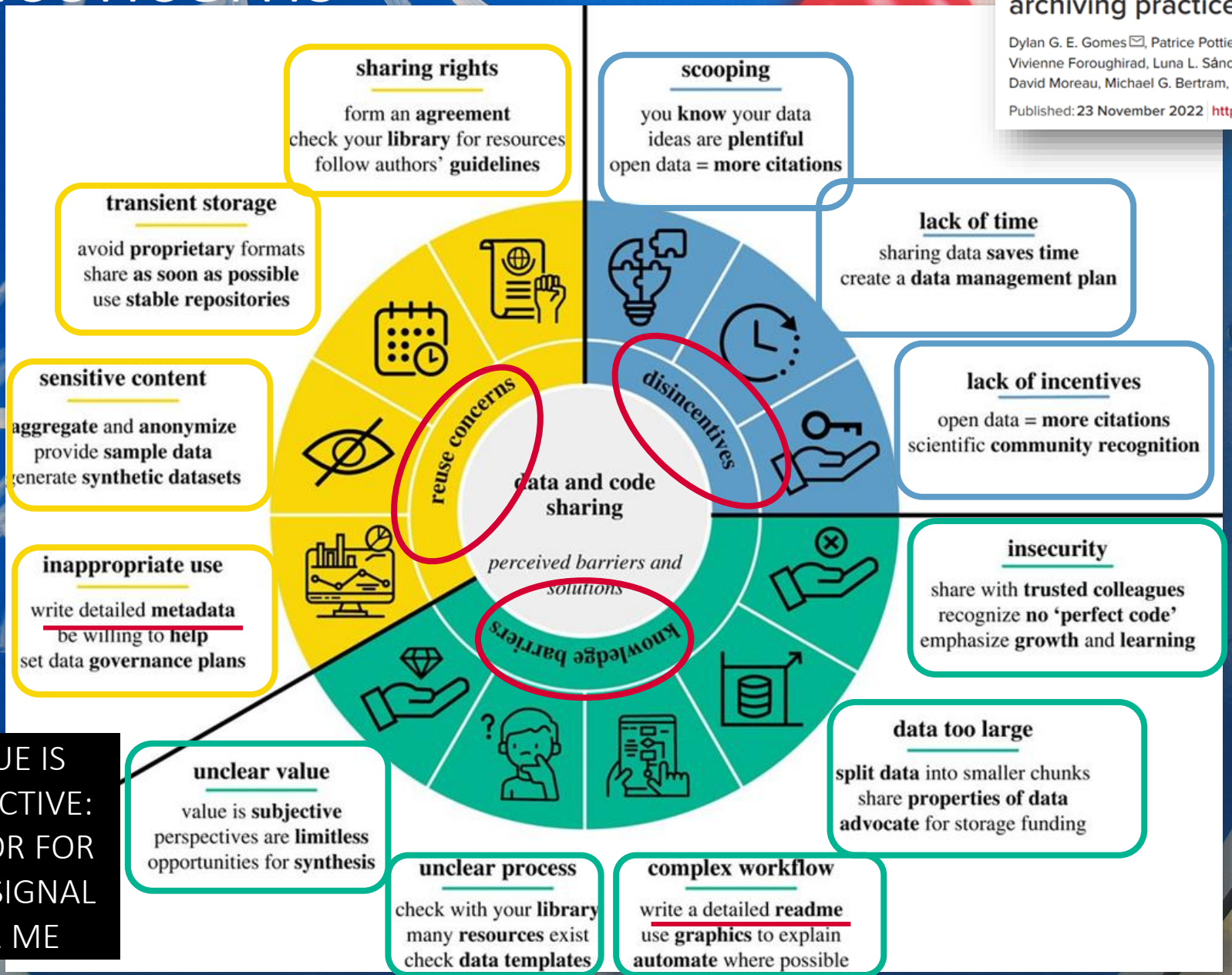
...concerns

Biological science practices

Why don't we share data a
Perceived barriers and ben
archiving practices

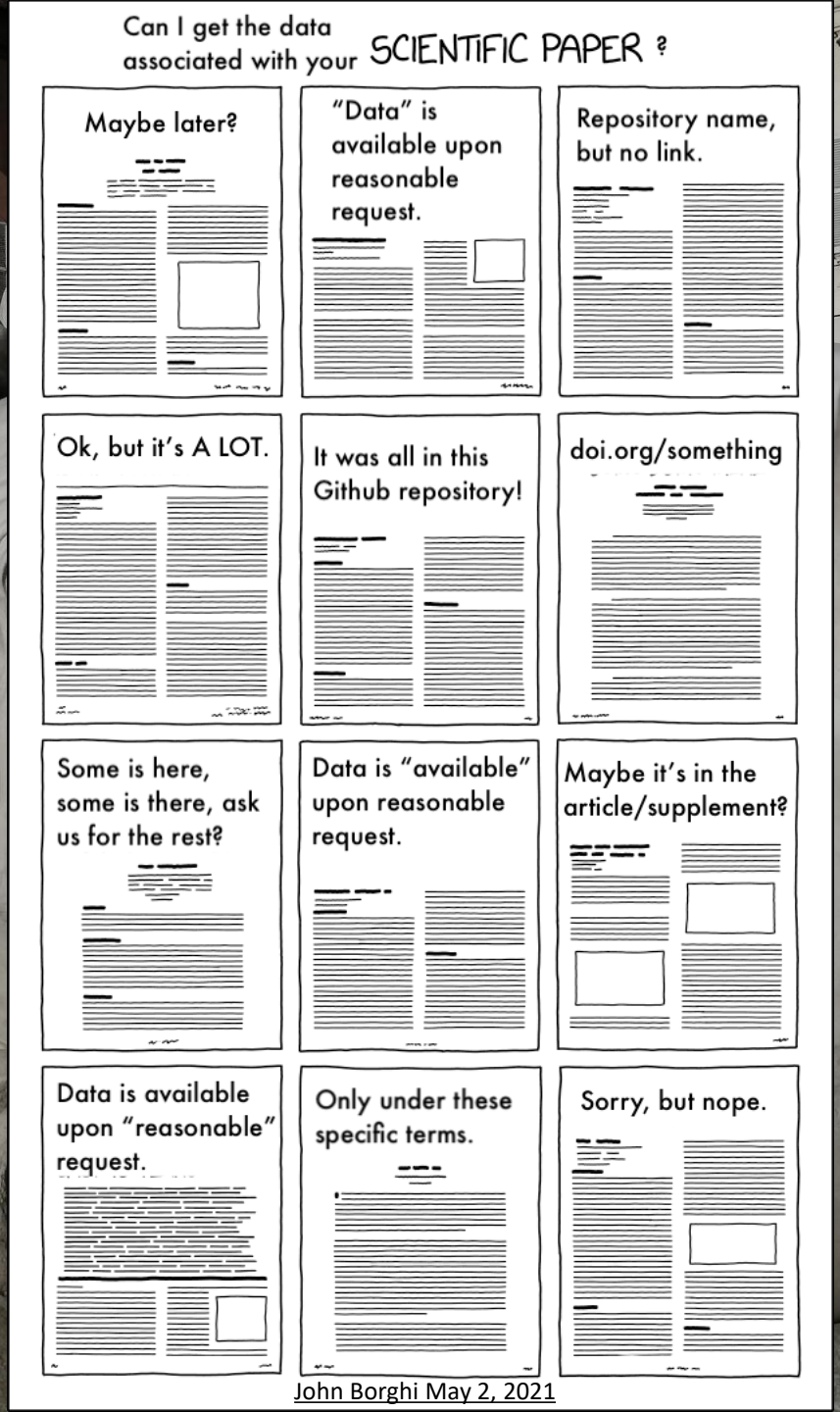
Dylan G. E. Gomes, Patrice Pottier†, Robert Crystal-O
Vivienne Foroughirad, Luna L. Sánchez-Reyes, Rachel
David Moreau, Michael G. Bertram, Cooper A. Smout a

Published: 23 November 2022 https://doi.org/10.1098



VALUE IS SUBJECTIVE: RUMOR FOR YOU, SIGNAL FOR ME

[reality]



John Borghi May 2, 2021

[real

FOR THIS FORM I NEED TO GIVE A REASON WHY YOU'RE NOT SHARING YOUR DATA
I CAN'T SHARE MY DATA BECAUSE THIS GROUP IN BELARUS WILL SEND IT TO MY EX STUDENT IN AUCKLAND WHO WILL INVENT EVERYTHING BEFORE ME. ALSO THE ICELANDIC ARE WATCHING!
OKAY... I'LL JUST TICK THE BOX LABELLED "IRRATIONAL CONSPIRACY"
ERRANTSCIENCE.COM

SCHOLARLY COMMUNICATION IS A CONVERSATION

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 a
learning from each other – if a researcher is opposed to the idea of discussing their datasets, collaborating with others, and generally being a good science citizen, then they should be outed by their community as a poor participant.

People will misinterpret the data

IMPOSSIBLE, IF IN «R» IN «FAIR» YOU DOCUMENTED

C&A suggest this: "Document how it s
correct such people; those that misint
help." From the UK Data Archive: "Pro
contextual information for your resea
correctly use and understand your da

It's worth mentioning, however, a second point C&A make: "Publishing may actu
be useful to counter willful misrepresentation (e.g. of data acquired through Free
of Information legislation), as one can quickly point to the real data on the web t
refute the wrong interpretation."

My data is not very interesting

EHM… SO WHY ARE WE FUNDING YOU WITH PUBLIC MONEY?

C&A: "Let others judge how interesti
people that care about them." I'd als
dataset has value to future research
"climate change" was a research top
documenting and understanding the phenomenon. From the UK Data Archive: "

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 PERFECTLY «FAIR»

C&A suggest the embargo route: "One
embargo; require people to archive the
es
public after X months. You could even
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.

IMPOSSIBLE, IF IN «R» IN «FAIR» YOU DOCUMENTED

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.

My data is embarrassingly bad

HOW CAN YOU DO RESEARCH WITH «BAD» DATA???

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.

It's not a priority and I'm busy

A GROWING NUMBER OF FUNDERS AND JOURNALS ASK FOR DATA…
IT'S A PRIORITY NOW

CARLY STRASSER

Open Science + Data Sharing Advocate

Home
Blog Posts
Previous Research

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

Closed Data… Excuses, Excuses

…and still

Digital Science Report

The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

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

Nov. 29, 2021

November 2021

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

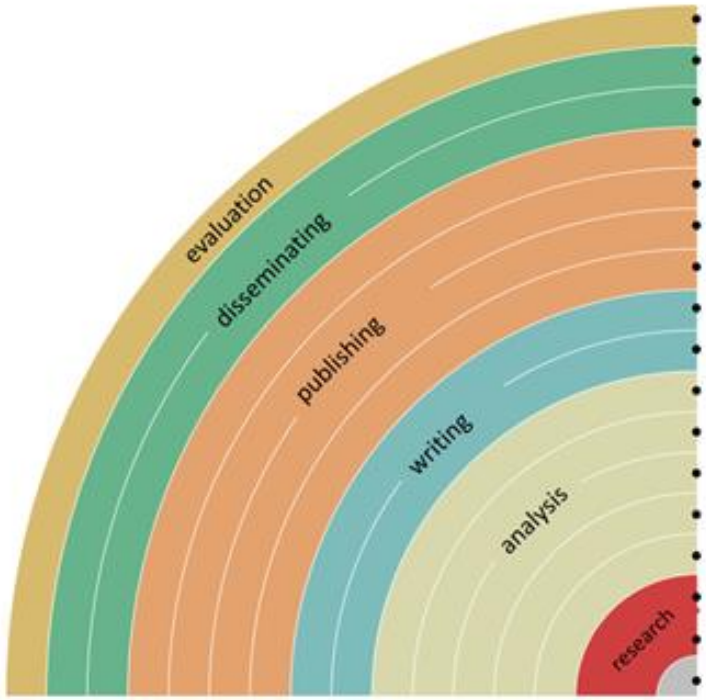
	2018	2019	2020	2021
Concerns about misuse of data	518 36%	2082 37%	1881 38%	1920 43%
Not receiving appropriate credit or acknowledgement	478 33%	1834 32%	1531 31%	1739 39%
Unsure about copyright and data licensing	508 35%	1858 33%	1479 30%	1585 35%
Contains sensitive information or requires consent	416 29%	1274 22%	1535 31%	1366 30%
I am unsure I have permission from my funder or institute	458 32%	1454 26%	1291 26%	1110 25%
Organising data in a presentable and usable way	456 32%	1434 25%	1137 23%	1150 26%
Costs of sharing data	274 19%	1497 26%	1192 24%	1147 26%
Lack of time to deposit data	300 21%	1389 24%	984 20%	942 21%

CONCERNS

- MISUSE
- NO CREDIT
- COPYRIGHT

Open Science: why just your data?

YOU CAN MAKE YOUR WORKFLOW MORE OPEN BY...



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

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

2021

Nothing has really happened until it has been described.
— Virginia Woolf

THANKS!