FAIR data management (and a bit of Open Science)

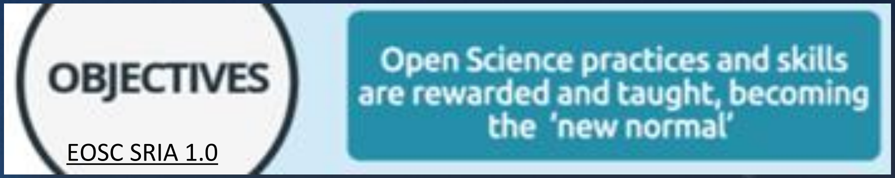
University of Turin, Dec.5, 2023

Elena Giglia

elena.giglia@unito.it

𝕏 @egiglia

[a bit of Open Science, because...]



OBJECTIVES

Open Science practices and skills are rewarded and taught, becoming the 'new normal'

EOSC SRIA 1.0

OPEN SCIENCE IS THE «NEW NORMAL»

Lessons learned from COVID

OPEN DATA SAVE LIVES

Digital Science Report

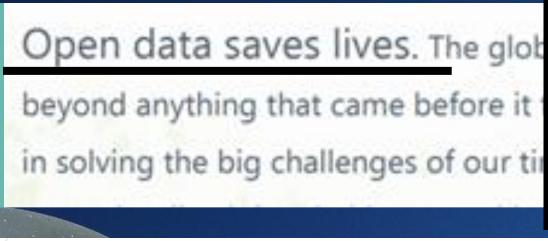
The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

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

November 2021

Nov. 29 2021



Open data saves lives. The glob

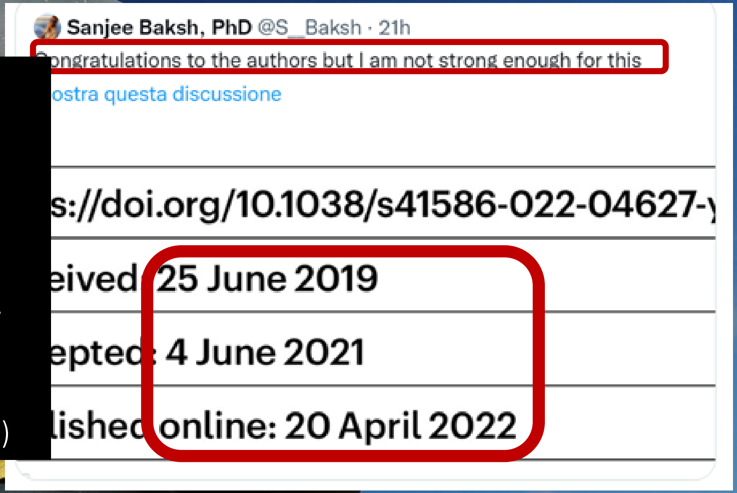
beyond anything that came before it

in solving the big challenges of our ti

WE NEED DATA [FAIR BY DESIGN] (AND NOT ONLY THE FINAL SYNTHESIS OF THE RESEARCH, I.E. THE ARTICLE)

… …AND WE NEED RESULTS IMMEDIATELY… TRADITIONAL SUBSCRITPION BASED JOURNALS: FIRST ARTICLES (WITH NO DATA) AT THE EARLIEST IN DEC. 2020

(9-18 MONTHS AVERAGE PUBLICATION TIME)



Sanjee Baksh, PhD @S_Baksh · 21h

ongratulations to the authors but I am not strong enough for this

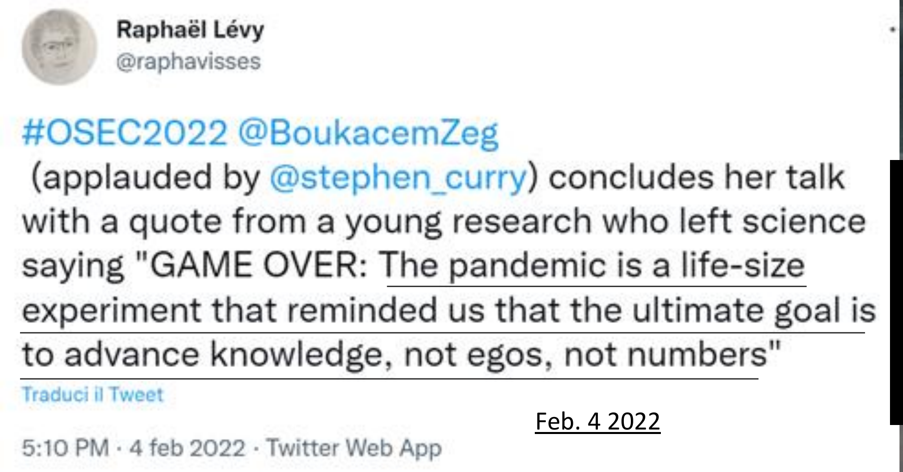
ostra questa discussione

s://doi.org/10.1038/s41586-022-04627-y

eived 25 June 2019

epted 4 June 2021

lished online: 20 April 2022



Raphaël Lévy
@raphavisses

#OSEC2022 @BoukacemZeg (applauded by @stephen_curry) concludes her talk with a quote from a young research who left science saying "GAME OVER: The pandemic is a life-size experiment that reminded us that the ultimate goal is to advance knowledge, not egos, not numbers"

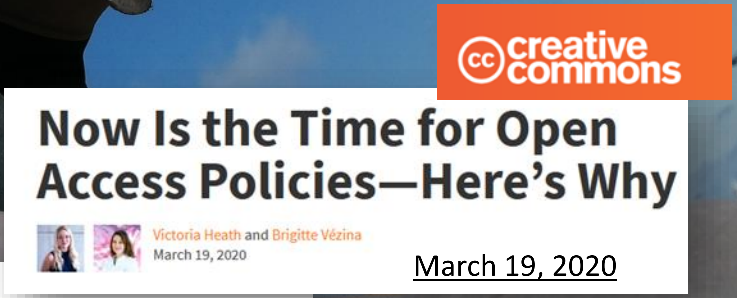
Traduci il Tweet

5:10 PM · 4 feb 2022 · Twitter Web App

Feb. 4 2022

THE PANDEMIC IS A LIFE-SIZE EXPERIMENT THAT REMINDED US THAT THE ULTIMATE GOAL IS TO ADVANCE KNOWLEDGE, NOT EGOS, NOT NUMBERS

Lessons learned from COVID



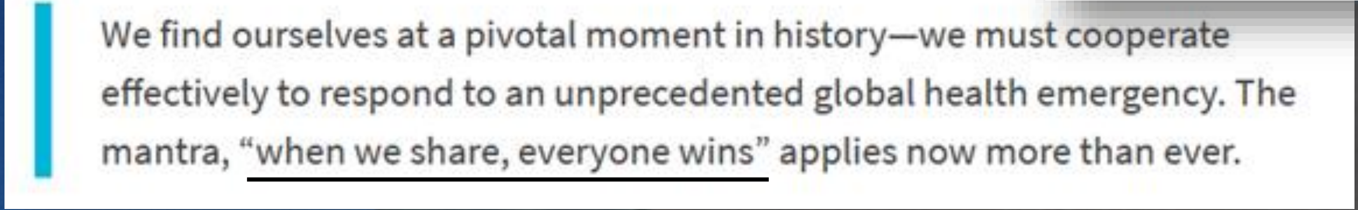
creative commons

Now Is the Time for Open Access Policies—Here's Why

Victoria Heath and Brigitte Vézina
March 19, 2020

March 19, 2020

SHARING IS CRUCIAL



> We find ourselves at a pivotal moment in history—we must cooperate effectively to respond to an unprecedented global health emergency. The mantra, "when we share, everyone wins" applies now more than ever.

… so what about the current system?

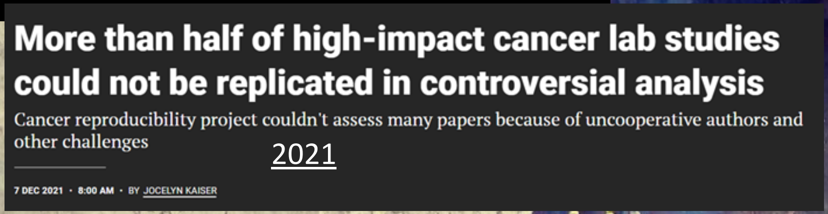
WE ARE STILL TOO FOCUSED ONLY ON PAPERS (FOR EVALUATION)

WE PAY 10 BN $ TO LOCK UP BEHIND PAYWALLS A CONTENT PRODUCED WITH PUBLIC MONEY AND GIVEN FOR FREE

…WITH AN AVERAGE PUBLICATION TIME OF 9-18 MONTHS…

…AND 179% INCREASE IN SELF-CITATIONS…

…AND 70% OF STUDIES WHICH ARE NOT REPRODUCIBLE…



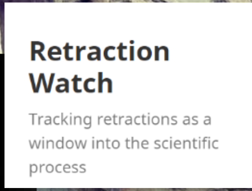
More than half of high-impact cancer lab studies could not be replicated in controversial analysis

Cancer reproducibility project couldn't assess many papers because of uncooperative authors and other challenges

7 DEC 2021 · 8:00 AM · BY JOCELYN KAISER

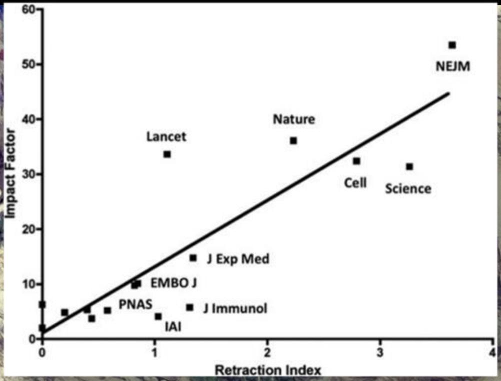
2021

… AND 43% RETRACTIONS FOR FRAUD, WITH A DIRECT CORRELATION BETWEEN THE #RETRACTIONS/JOURNAL IMPACT FACTOR



Retraction Watch

Tracking retractions as a window into the scientific process



WHY? BECAUSE EVALUATION BECAME AN OBSESSION, AND PEOPLE GAME THE SYSTEM AT EVERY LEVEL

GAMING THE METRICS

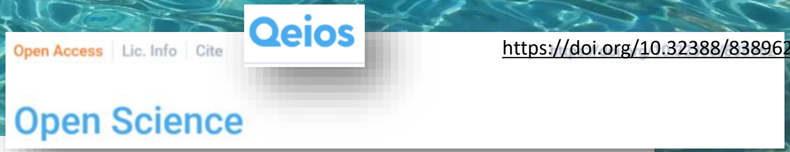
Misconduct and Manipulation in Academic Research

EDITED BY Mario Biagioli AND Alexandra Lippman

2019

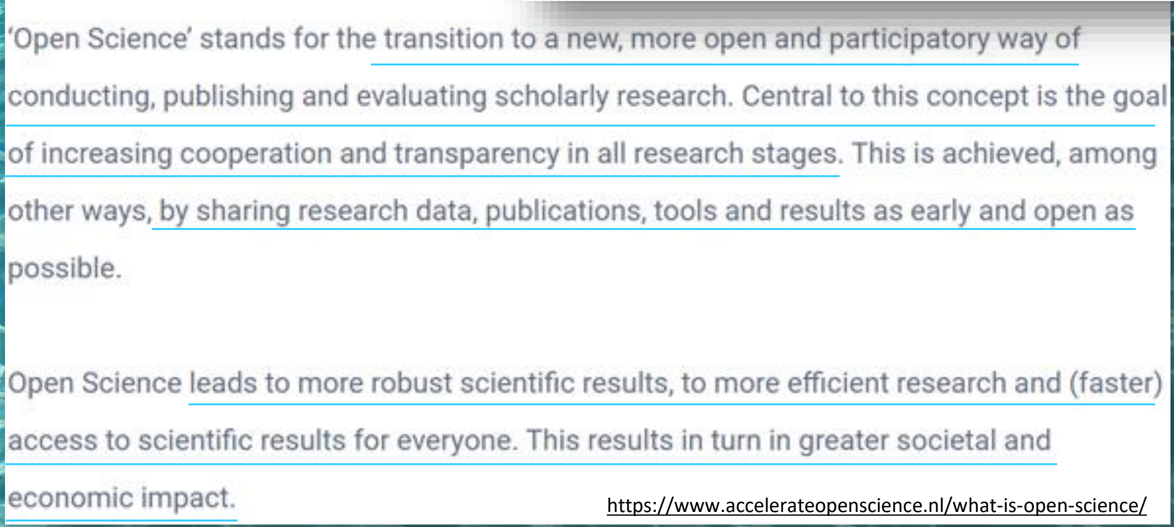
Open Science might help?

Open Science – definition



Open Science

https://doi.org/10.32388/838962



'Open Science' stands for the transition to a new, more open and participatory way of conducting, publishing and evaluating scholarly research. Central to this concept is the goal of increasing cooperation and transparency in all research stages. This is achieved, among other ways, by sharing research data, publications, tools and results as early and open as possible.

Open Science leads to more robust scientific results, to more efficient research and (faster) access to scientific results for everyone. This results in turn in greater societal and economic impact.

https://www.accelerateopenscience.nl/what-is-open-science/

WE ARE TALKING PUBLIC MONEY: PUBLICLY FUNDED RESEARCH SHOULD BE PUBLICLY AVAILABLE

NEW WAY OF
- CONDUCTING
- PUBLISHING
- EVALUATING
RESEARCH

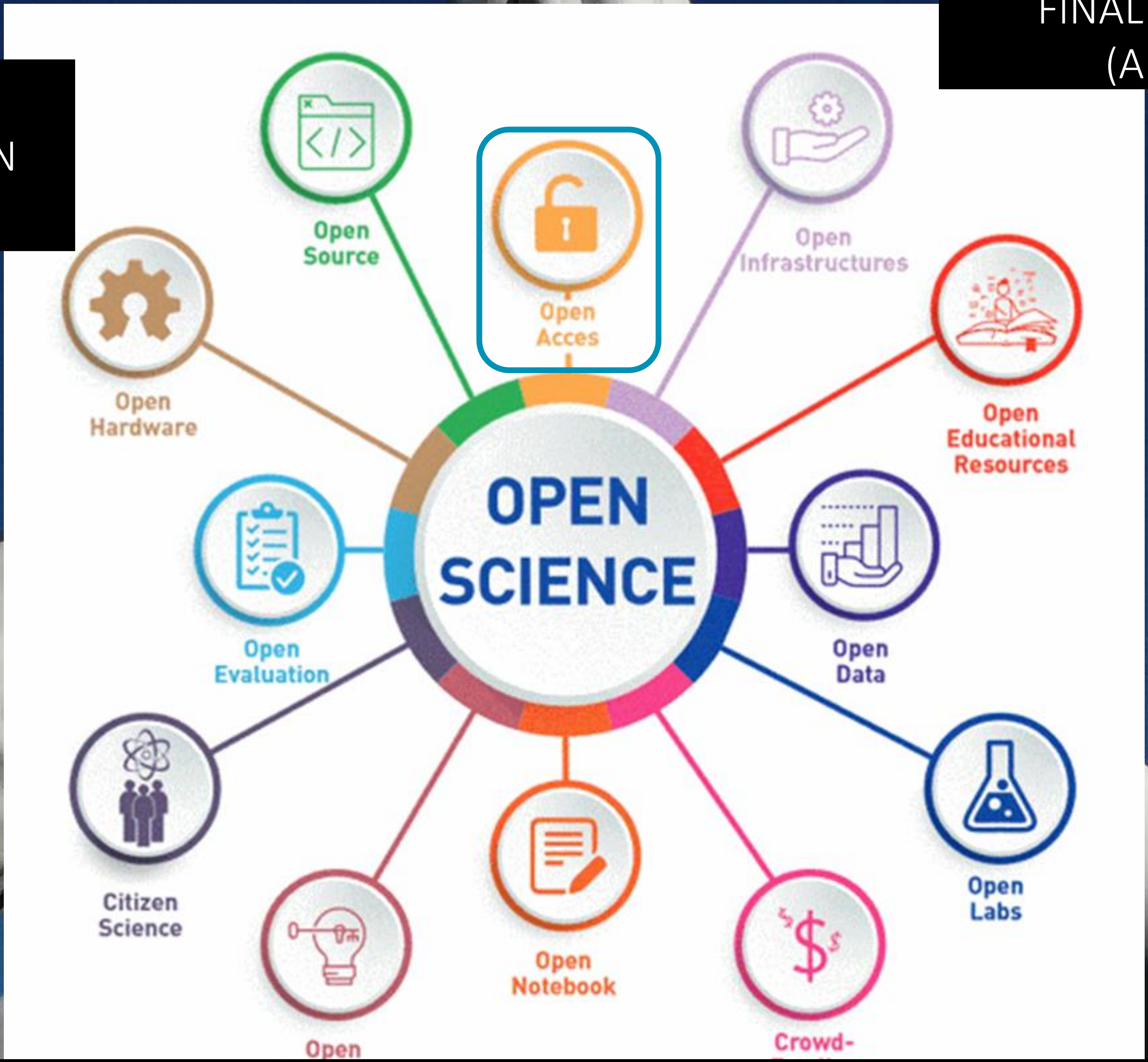
SHARING
- DATA/TEXTS
- TOOLS
- RESULTS...
AS EARLY AND OPEN AS POSSIBLE

OS LEADS TO MORE ROBUST SCIENTIFIC RESULTS, MORE EFFICIENT RESEARCH AND FASTER ACCESS
+ GREATER SOCIETAL AND ECONOMIC IMPACT

Open Science

OPEN SCIENCE≠OPEN ACCESS

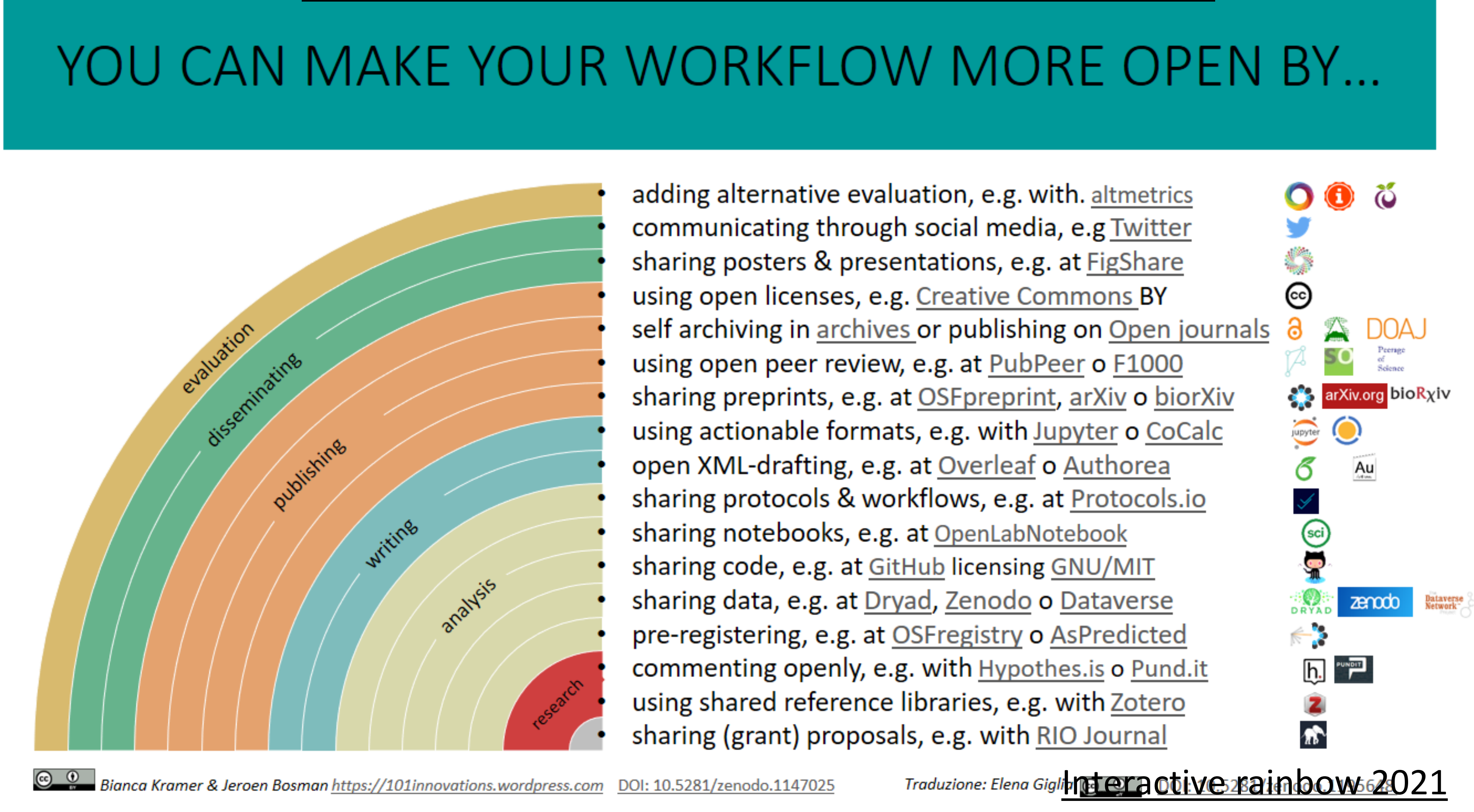
FOCUS ON THE ENTIRE PROCESS, NOT ONLY THE FINAL SYNTHESIS (ARTICLE)



ALL THESE COMPONENTS TO BE EMBEDDED IN THE PROPOSAL TEMPLATE, 1.2 EXCELLENCE-METHODOLOGY AND TO BE EVALUATED UNDER «SCIENTIFIC EXCELLENCE»

Oper

ANY COMPONENT OF THIS RAINBOW SHOULD COUNT AS «RESEARCH OUTPUT»



YOU CAN MAKE YOUR WORKFLOW MORE OPEN BY…

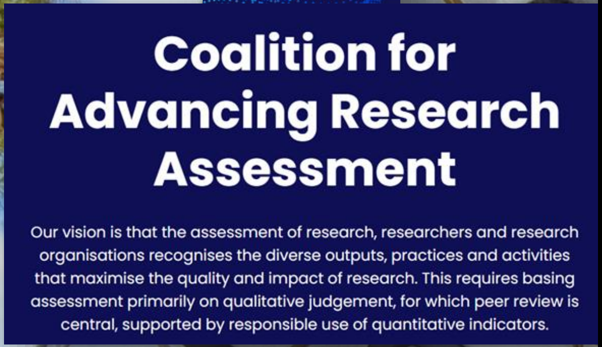
- adding alternative evaluation, e.g. with. altmetrics
- communicating through social media, e.g Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. Creative Commons BY
- self archiving in archives or publishing on Open journals
- using open peer review, e.g. at PubPeer 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 *Traduzione: Elena Giglia* DOI: 10.5281/zenodo.1195648

Interactive rainbow 2021

TECHNICALLY, IT'S THERE.
WHAT IS STILL NEEDED IS THE CULTURAL SHIFT…
AND YOUR FIRST STEP!

CoARA
COARA

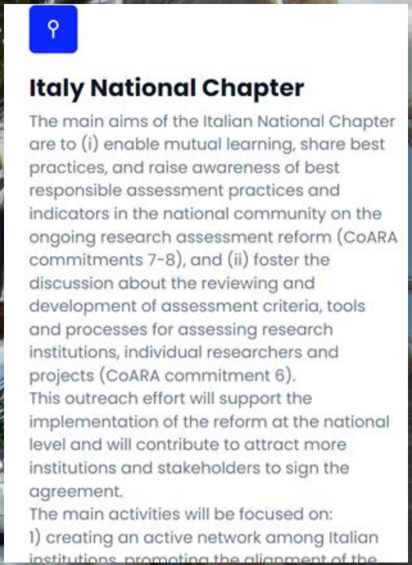


Coalition for Advancing Research Assessment

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 central, supported by responsible use of quantitative indicators.

TIME IS UP!!!

- THE REFORM OF RESEARCH EVALUATION HAS STARTED
- COARA LAUNCHED IN 2022, 644 SIGNATORIES
- ITALIAN CHAPTER IS ACTIVE
- COMMITTMENT: 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)

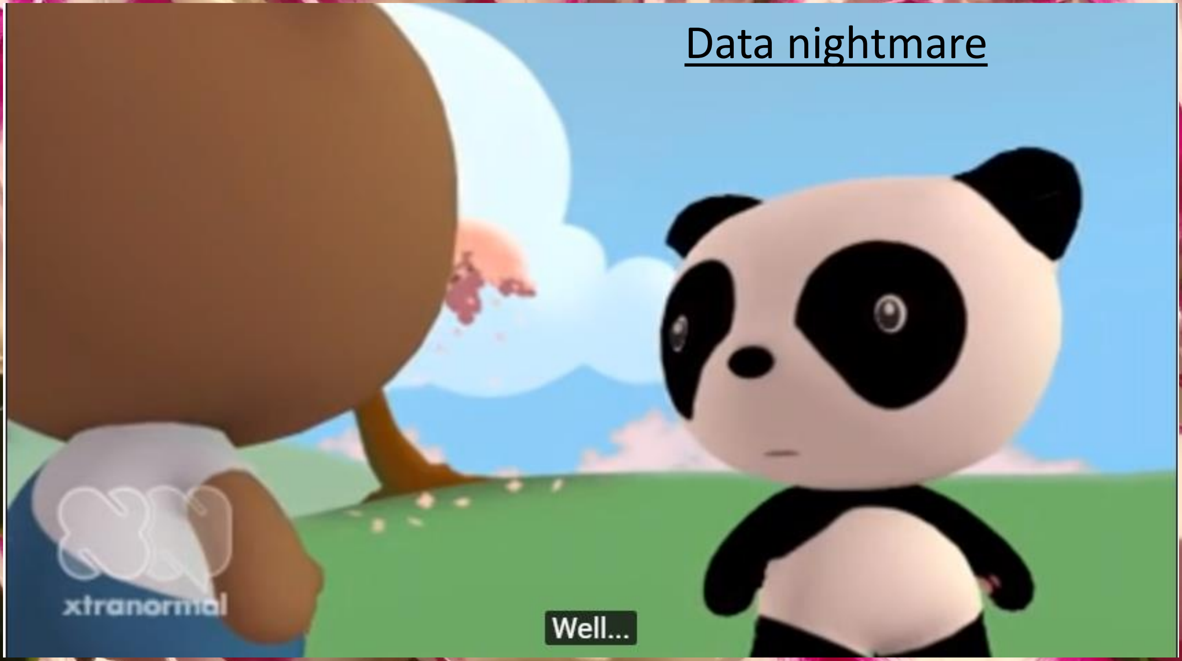
I believe in a research culture that recognises a diversity of contributions to science and society; that celebrates high quality and impactful research; and that values sharing, collaboration, integrity and engagement with society, transmitting knowledge from generation to generation.

Mariya Gabriel
Commissioner for Innovation, Research, Culture, Education and Youth

YES, BUT...
WE ARE STILL EVALUATED BY IMPACT FACTOR

Why should we care about data?

Why should you take care of your data?



Data nightmare

… THIS IS THE DATA STEWARD'S NIGHTMARE:

- NO BACKUP
- NO SOFTWARE
- - NO DATA LEGEND

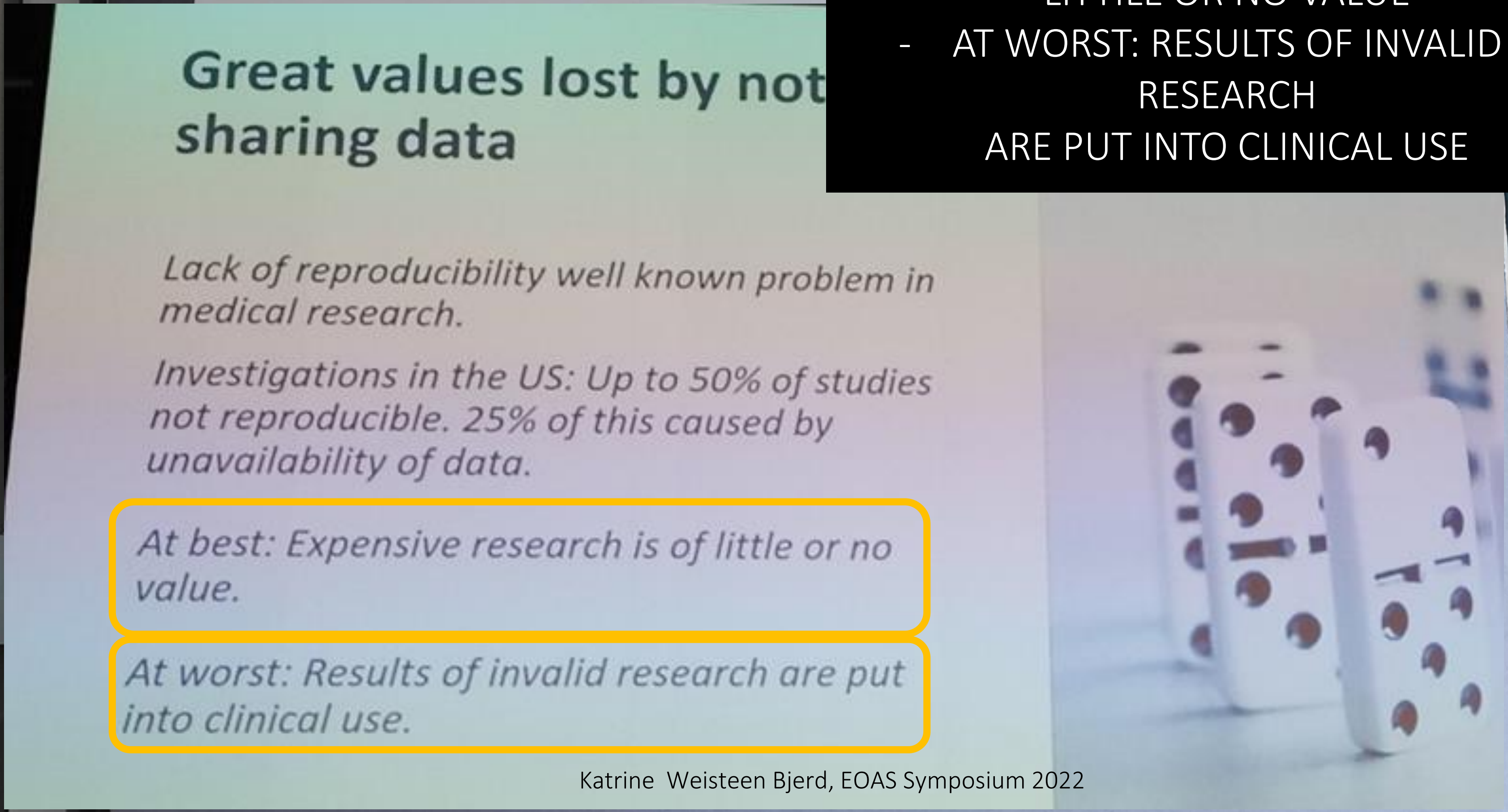
… AND:

- DATA GENERATED WITH PUBLIC FUNDS
- PUBLISHED IN «SCIENCE» (DATA POLICY)
- REQUESTED FROM A DIFFERENT DISCIPLINE

Why should we care about data?

LOST VALUE IF DATA ARE MISSING:

- AT BEST: EXPENSIVE RESEARCH IS OF LITTILE OR NO VALUE
- AT WORST: RESULTS OF INVALID RESEARCH
ARE PUT INTO CLINICAL USE



Great values lost by not sharing data

Lack of reproducibility well known problem in medical research.

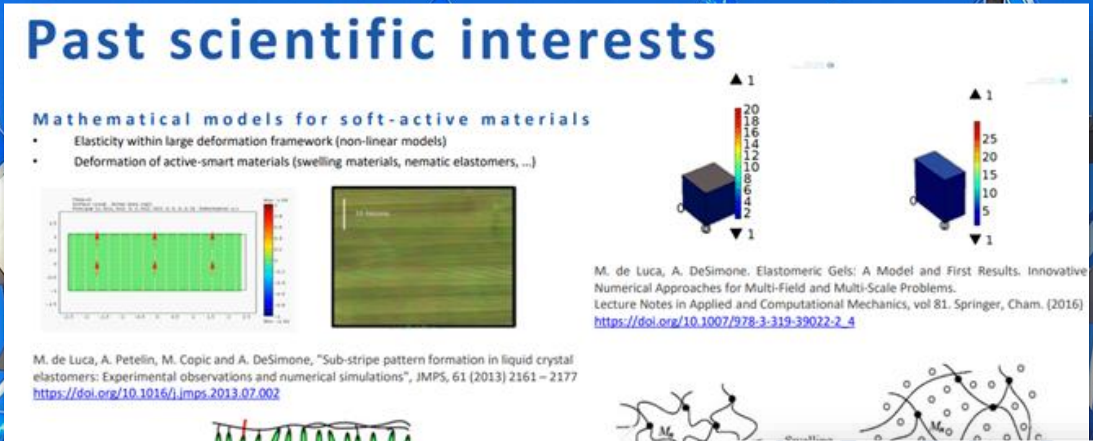
Investigations in the US: Up to 50% of studies not reproducible. 25% of this caused by unavailability of data.

At best: Expensive research is of little or no value.

At worst: Results of invalid research are put into clinical use.

Katrine Weisteen Bjerd, EOAS Symposium 2022

Why should we care about data?
A personal view



Past scientific interests

Mathematical models for soft-active materials

- Elasticity within large deformation framework (non-linear models)
- Deformation of active-smart materials (swelling materials, nematic elastomers, …)

M. de Luca, A. Petelin, M. Copic and A. DeSimone, "Sub-stripe pattern formation in liquid crystal elastomers: Experimental observations and numerical simulations", JMPS, 61 (2013) 2161 – 2177 https://doi.org/10.1016/j.jmps.2013.07.002

M. de Luca, A. DeSimone. Elastomeric Gels: A Model and First Results. Innovative Numerical Approaches for Multi-Field and Multi-Scale Problems. Lecture Notes in Applied and Computational Mechanics, vol 81. Springer, Cham. (2016) https://doi.org/10.1007/978-3-319-39022-2_4

10 YEARS ON…

- DO I HAVE ACCESS TO MY OWN PUBLICATIONS?
- WHERE ARE MY DATA?
- CAN I REPRODUCE MY SIMULATIONS?

[M.R. DE LUCA, PhD]

Research (FAIR) data management

2023

|Mariarita de Luca|
https://orcid.org/0000-0002-5507-968X
mariarita.deluca@areasciencepart.it

Institute for Research and Innovative Technologies (RIT)
AREA SCIENCE PARK

1° Workshop for National PhD in "Theoretical and Applied Neuroscience" , Bertinoro 18.10.2023

What about my data and my publications?

- Do I have access to my publications?
- Where are my data?
- Can I reproduce my numerical simulations?

Image by Elisa from Pixabay

Why should we care about data?

1. DATA ARE THE FOUNDATION OF GOOD RESEARCH

DCC because good research needs good data

2. COVID SHOWED THAT WE NEED DATA, AND WE NED THEM AS SOON AS POSSIBLE

3. DATA ARE FRAGILE. THEY GET LOST

4. SOME DATA ARE UNIQUE AND NOR REPRODUCIBLE (ATMOSPHERIC, EARTHQUAKES…)

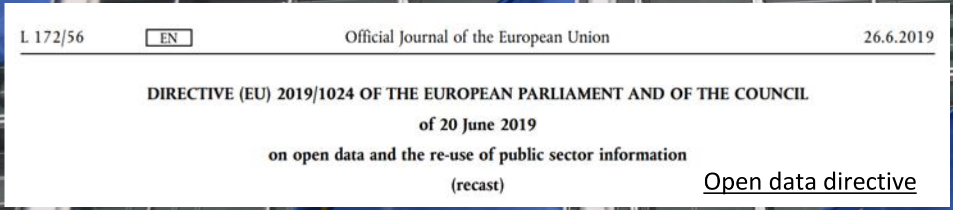
5. DATA CAN BE MANIPULATED, DATA MANAGEMENT PRESERVES INTEGRITY

6. TO ALLOW FOR CHECKS AND REPRODUCIBILITY

7. DATA CAN BE REUSED (IN UNEXPECTED WAYS)

Why should we care about data?

8.1 WE HAVE TO. OPEN DATA DIRECTIVE



L 172/56 EN Official Journal of the European Union 26.6.2019

DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 20 June 2019
on open data and the re-use of public sector information
(recast)

Open data directive

DIRECTIVE ENLARGED TO INCLUDE RESEARCH DATA

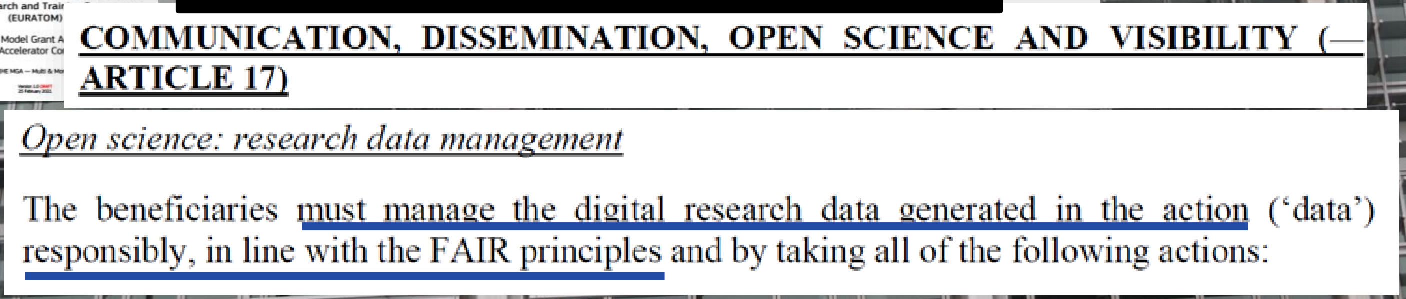
8.2 WE HAVE TO. IN HORIZON EUROPE YOU HAVE TO RESPONSIBLY MANAGING RESEARCH DATA ACCORDING TO FAIR PRINCIPLES (MANDATORY PRACTICE)

V.1 Feb 2021

Horizon Europe (HORIZON)
Euratom Research and Trai
(EURATOM)

General Model Grant A
EIC Accelerator Co

ANNEX 5



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

Open science: research data management

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

8.3 WE AVE TO. WE HAVE EOSC

EOSC Association
Advancing Open Science in Europe

8.4. WE HAVE TO. A GROWING NUMBER OF JOURNALS IS ASKING FOR DATA TO BE DEPOSITED UPON PUBLICATIONS (TRASPARENCY AND E REPRODUCIBILITY)

Why should we care about data?

Data creates a bridge between traditional disciplines, spawning discovery and innovation from the humanities to the hard sciences. Data dissolves barriers, opening up new channels of communication, lines of research, and commercial opportunities. Data will be the engine, the spark to create a better world for all.

World Economic Forum 2012

European Commission

Sept. 29, 2021

Communication from
the Commission to the
European Parliament, the Council,
the European Economic and
Social Committee and
the Committee of the Regions on

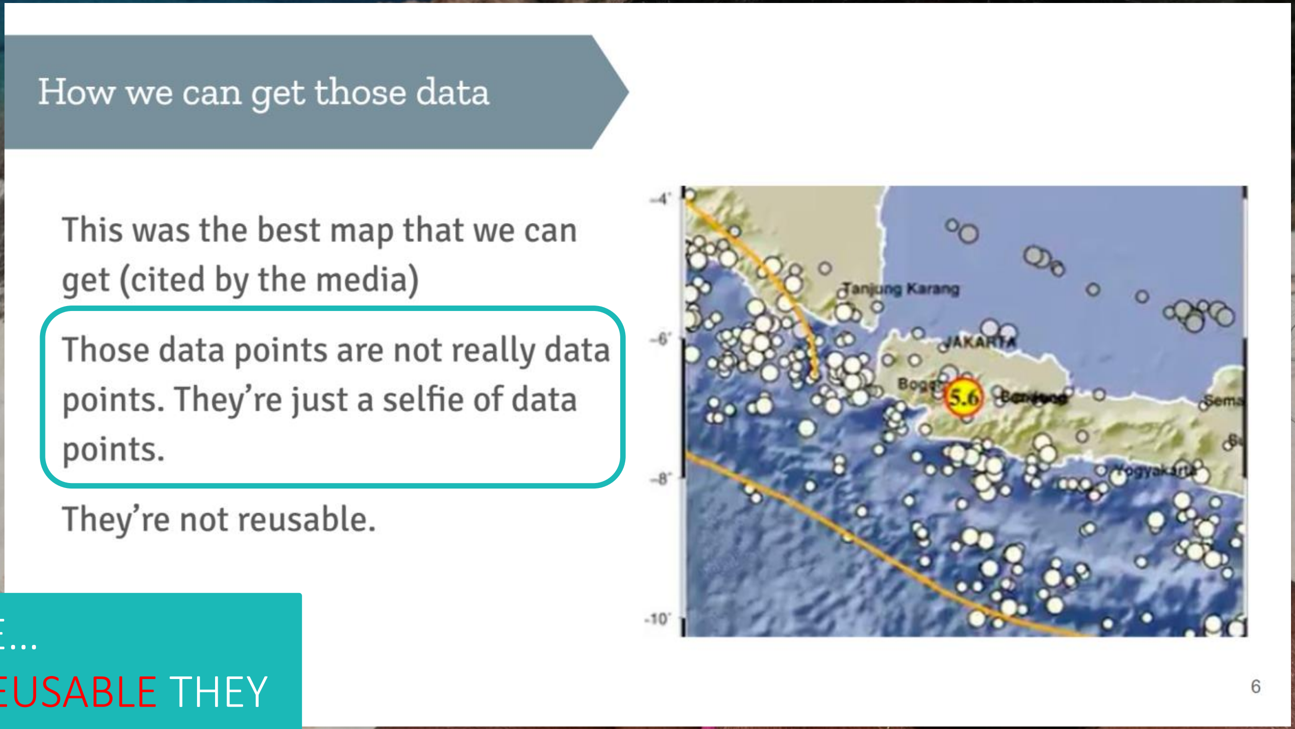
European Missions

9. DATA CREATES BRIDGES…

…REMIND: HORIZON EUROPE AND THE MISSIONS…

Why should we care about FAIR data?

...the selfie...



How we can get those data

This was the best map that we can get (cited by the media)

Those data points are not really data points. They're just a selfie of data points.

They're not reusable.

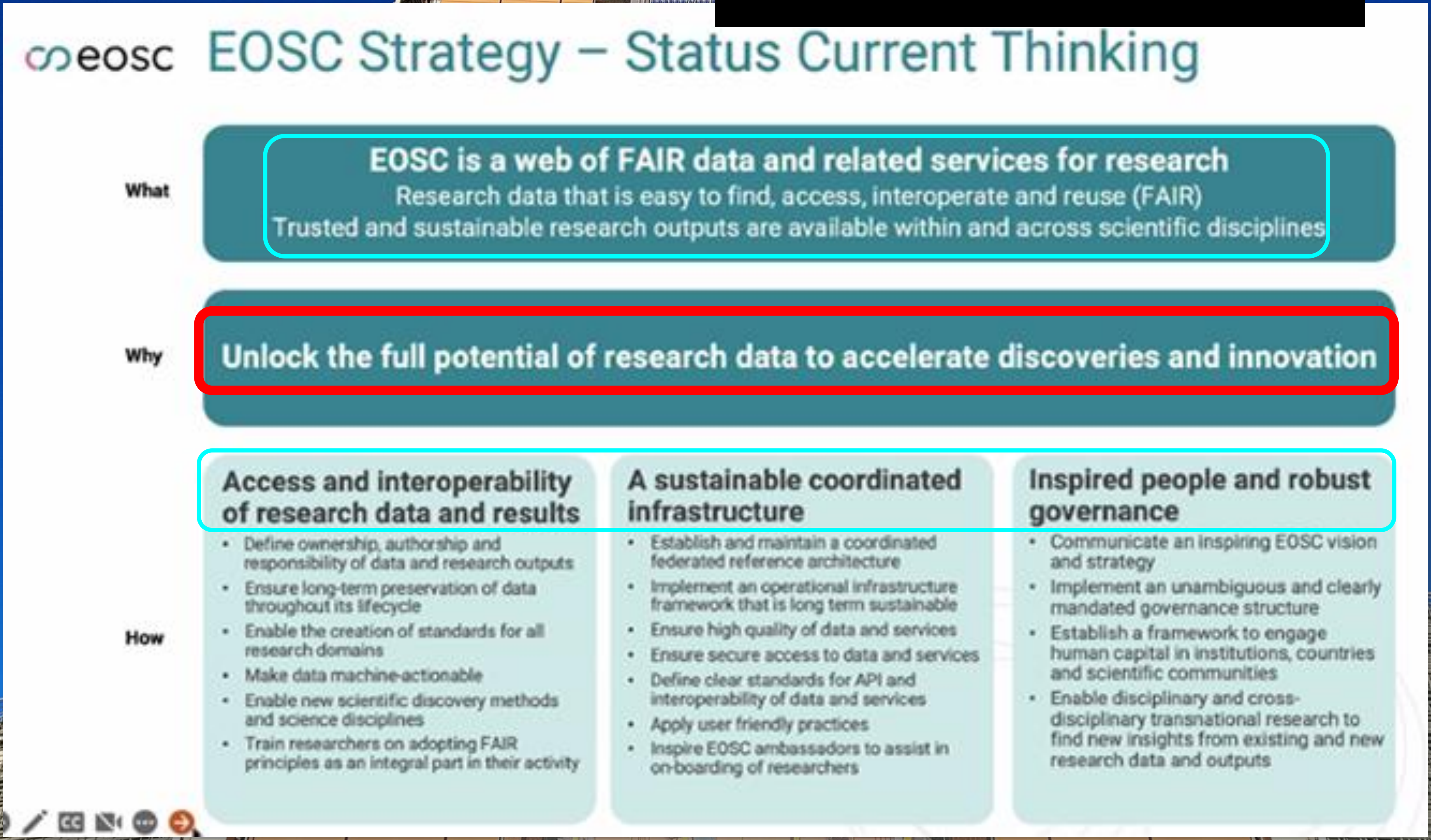
6

IN «FAIR» THE STRESS IS ON «R»

BEWARE…
IF DATA ARE NOT REUSABLE THEY ARE JUST A SELFIE OF DATA [USELESS]
[Dasapta Erwin Irawan]

…the EOSC

…VIRTUAL ENVIRONMENT TO UNLOCK THE FULL POTENTIAL OF RESEARCH DATA TO ACCELERATE DISCOVERIES AND INNOVATION



eosc EOSC Strategy – Status Current Thinking

What

EOSC is a web of FAIR data and related services for research

Research data that is easy to find, access, interoperate and reuse (FAIR)

Trusted and sustainable research outputs are available within and across scientific disciplines

Why

Unlock the full potential of research data to accelerate discoveries and innovation

How

Access and interoperability of research data and results

- Define ownership, authorship and responsibility of data and research outputs
- Ensure long-term preservation of data throughout its lifecycle
- Enable the creation of standards for all research domains
- Make data machine-actionable
- Enable new scientific discovery methods and science disciplines
- Train researchers on adopting FAIR principles as an integral part in their activity

A sustainable coordinated infrastructure

- Establish and maintain a coordinated federated reference architecture
- Implement an operational infrastructure framework that is long term sustainable
- Ensure high quality of data and services
- Ensure secure access to data and services
- Define clear standards for API and interoperability of data and services
- Apply user friendly practices
- Inspire EOSC ambassadors to assist in on-boarding of researchers

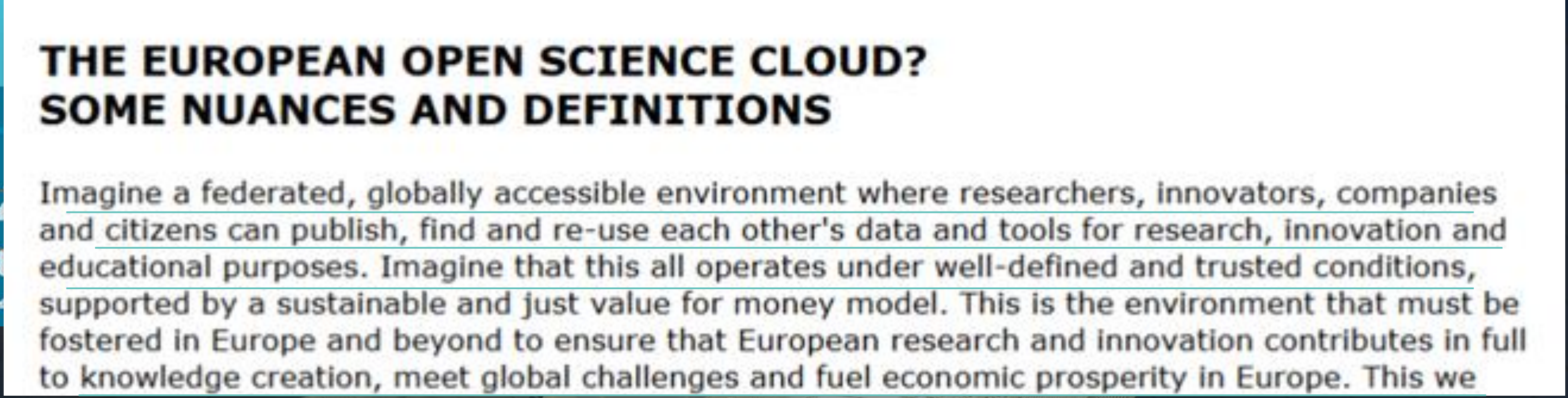
Inspired people and robust governance

- Communicate an inspiring EOSC vision and strategy
- Implement an unambiguous and clearly mandated governance structure
- Establish a framework to engage human capital in institutions, countries and scientific communities
- Enable disciplinary and cross-disciplinary transnational research to find new insights from existing and new research data and outputs

[EOSC IS NOT A BIG BOX]

2016

Realising the European Open Science Cloud



THE EUROPEAN OPEN SCIENCE CLOUD?
SOME NUANCES AND DEFINITIONS

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

EOSC IS NOT A REPOSITORY NOR A «CLOUD»

YOU DON'T «UPLOAD» YOUR DATA INTO EOSC

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

AND GIVE SEAMLESS ACCESS TO 20 M EU RESEARCHERS

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

OBJECTIVES

Open Science practices and skills are rewarded and taught, becoming the 'new normal'

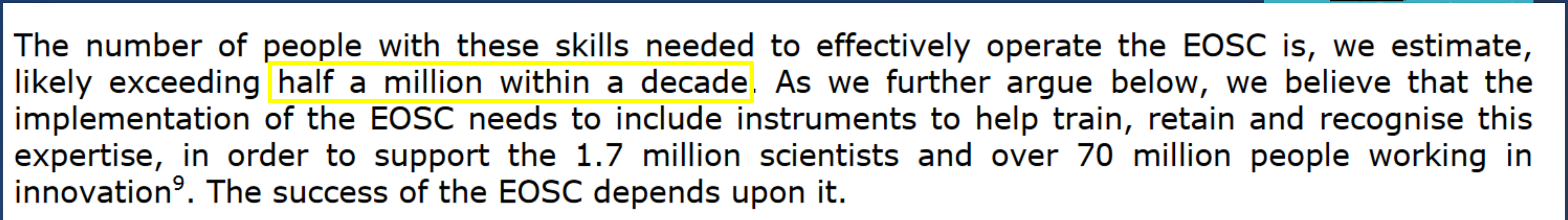
EOSC SRIA 1.0

[EOSC/FAIR is based on data stewardship]

European Commission

Realising the European Open Science Cloud

Report, 2016

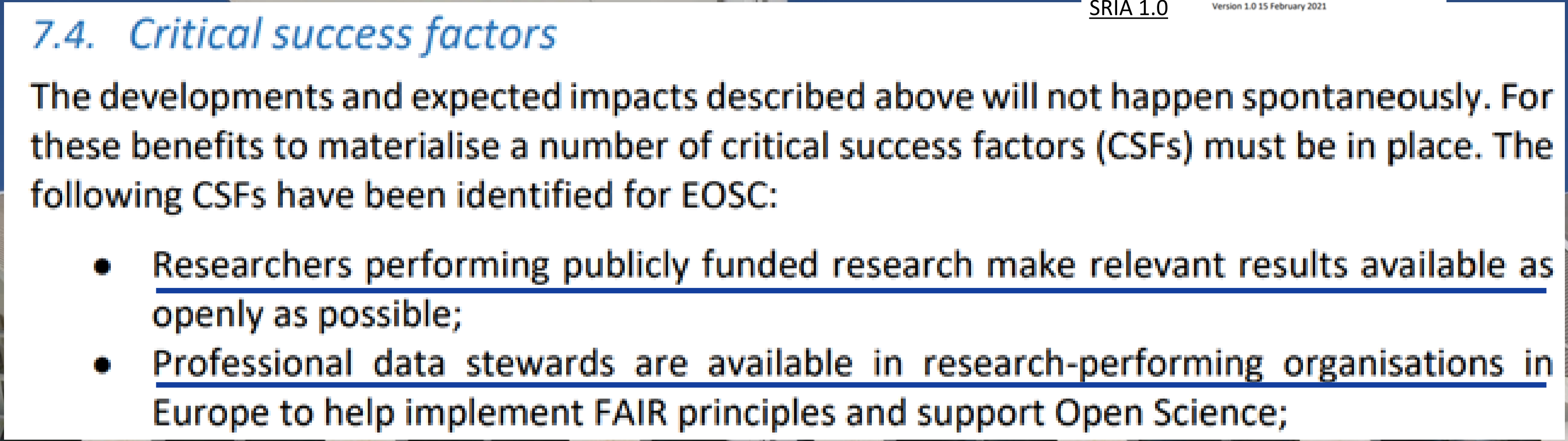

The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise, in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

- WE NEED 500.00 DATA STEWARDS
- DATA STEWARDS ARE ONE OF THE CRITICAL SUCCESS FACTORS OF EOSC

Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)

Version 1.0 15 February 2021

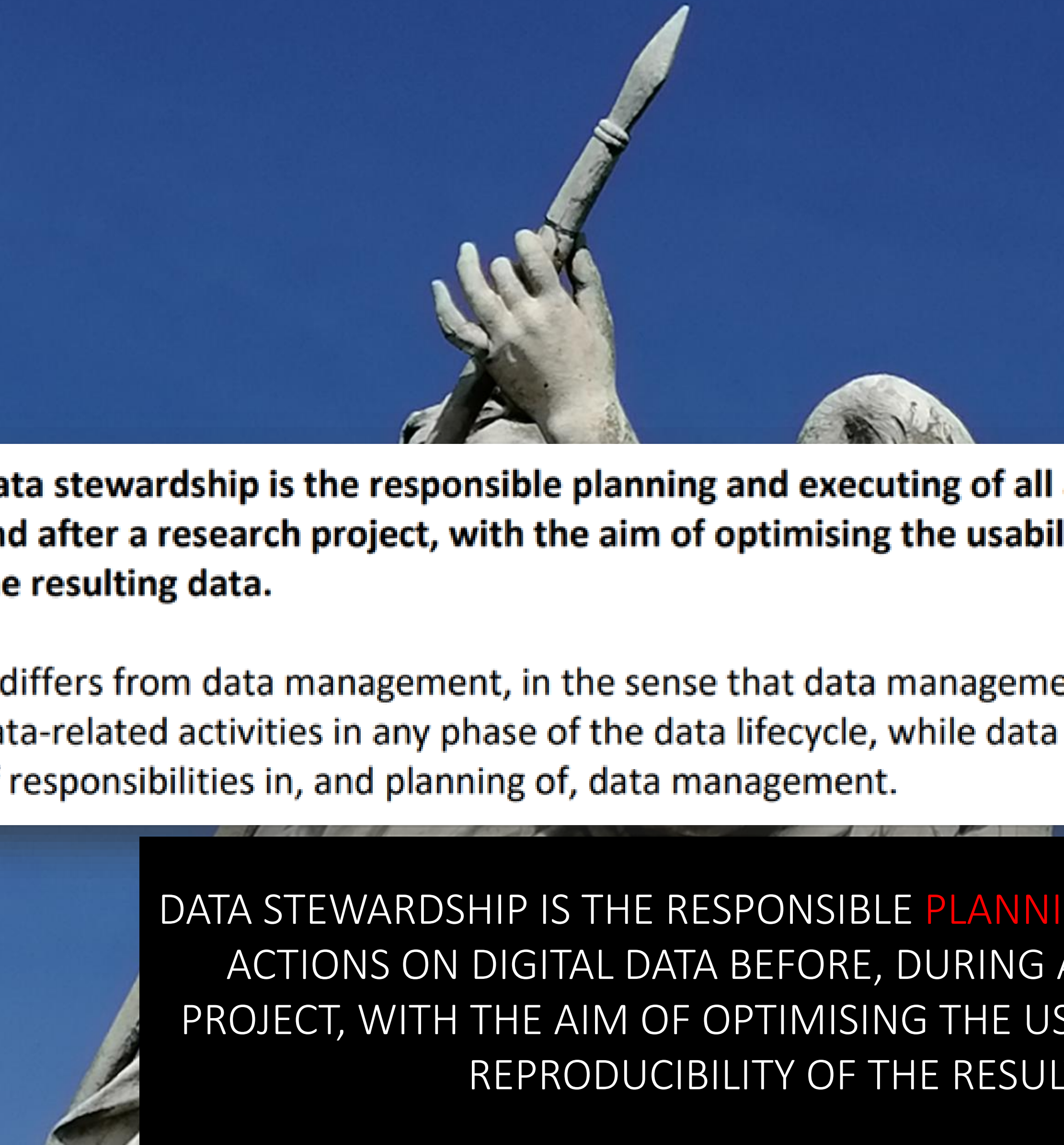
SRIA 1.0

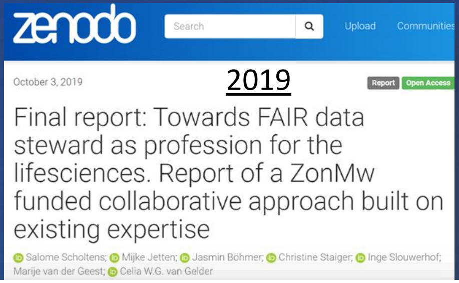

7.4. Critical success factors

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

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

What is data stewardship?





October 3, 2019

2019

Report Open Access

Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise

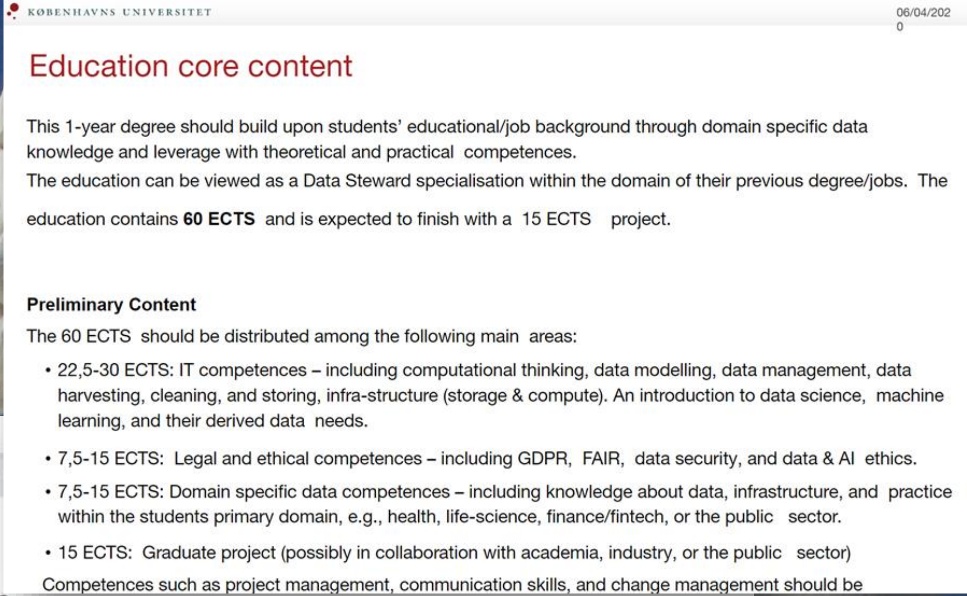
Salome Scholtens; Mijke Jetten; Jasmin Böhmer; Christine Staiger; Inge Slouwerhof; Marije van der Geest; Celia W.G. van Gelder

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

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

DATA STEWARDSHIP IS THE RESPONSIBLE PLANNING AND EXECUTING OF ALL ACTIONS ON DIGITAL DATA BEFORE, DURING AND AFTER A RESEARCH PROJECT, WITH THE AIM OF OPTIMISING THE USABILITY, REUSABILITY AND REPRODUCIBILITY OF THE RESULTING DTAA

[competence profi



KØBENHAVNS UNIVERSITET

06/04/202
0

Education core content

This 1-year degree should build upon students' educational/job background through domain specific data knowledge and leverage with theoretical and practical competences.
The education can be viewed as a Data Steward specialisation within the domain of their previous degree/jobs. The education contains **60 ECTS** and is expected to finish with a 15 ECTS project.

Preliminary Content

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

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

Competences such as project management, communication skills, and change management should be

KØBENHAVNS UNIVERSITET

Premi Esc per uscire dalla modalità

Competence Profile

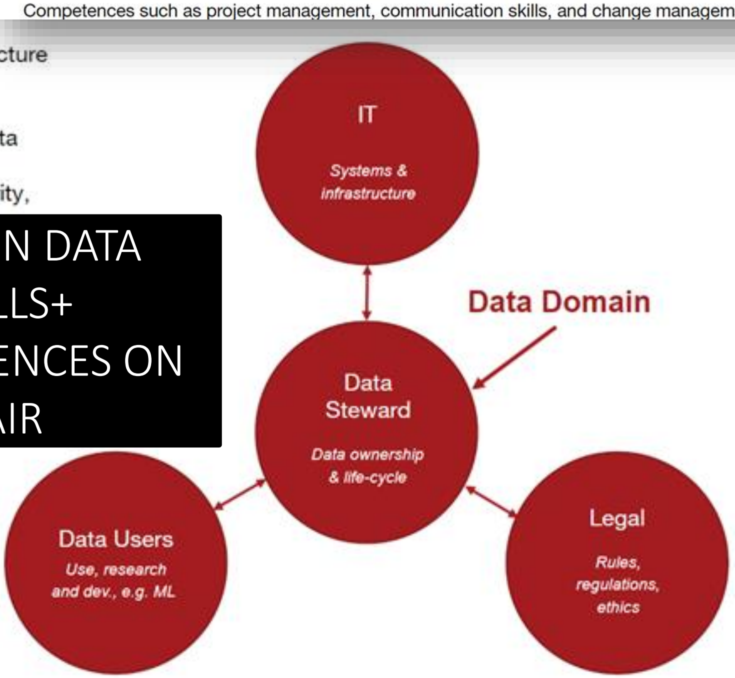
A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

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

Competence profile examples

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

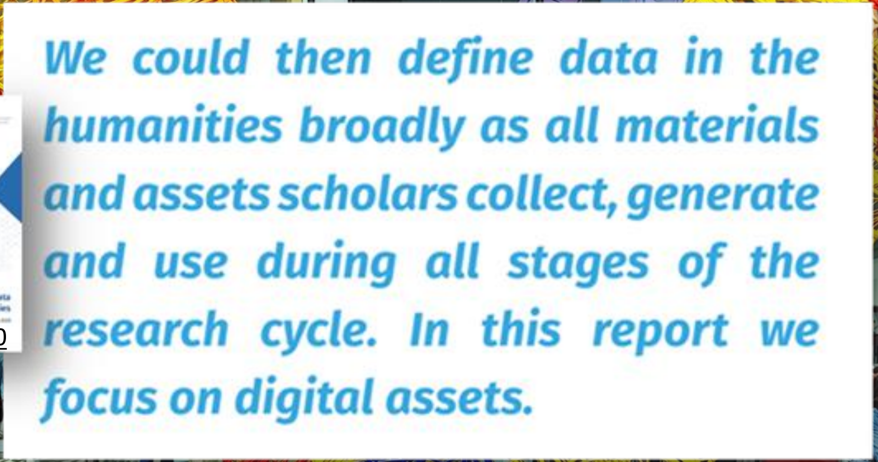
DOMAIN DATA SKILLS+ COMPETENCES ON FAIR


IT
Systems & infrastructure
Data Domain
Data Steward
Data ownership & life-cycle
Data Users
Use, research and dev., e.g. ML
Legal
Rules, regulations, ethics

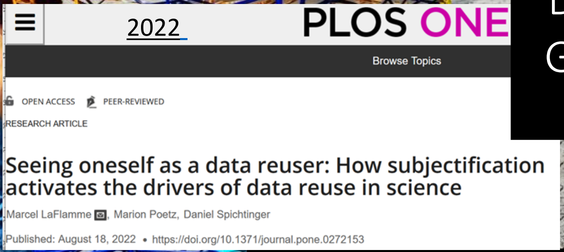
Copenhagen Univ. June 17 2020

Data

ALLEA 2020



We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets.



2022

PLOS ONE

RESEARCH ARTICLE

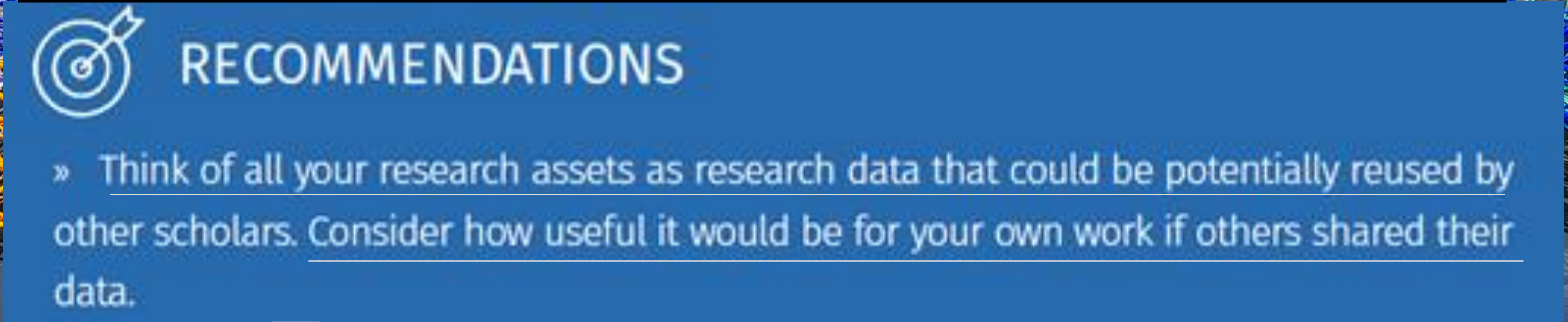
Seeing oneself as a data reuser: How subjectification activates the drivers of data reuse in science

Marcel LaFlamme, Marion Poetz, Daniel Spichtinger

Published: August 18, 2022 • https://doi.org/10.1371/journal.pone.0272153

DATA=ALL MATERIALS AND ASSETS COLLECTED, GENERATED AND USED DURING THE RESEARCH CYCLE

THINK OF ALL YOUR RESEARCH ASSETS AS RESEARCH DATA THAT COULD POTENTIALLY BE REUSED



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.

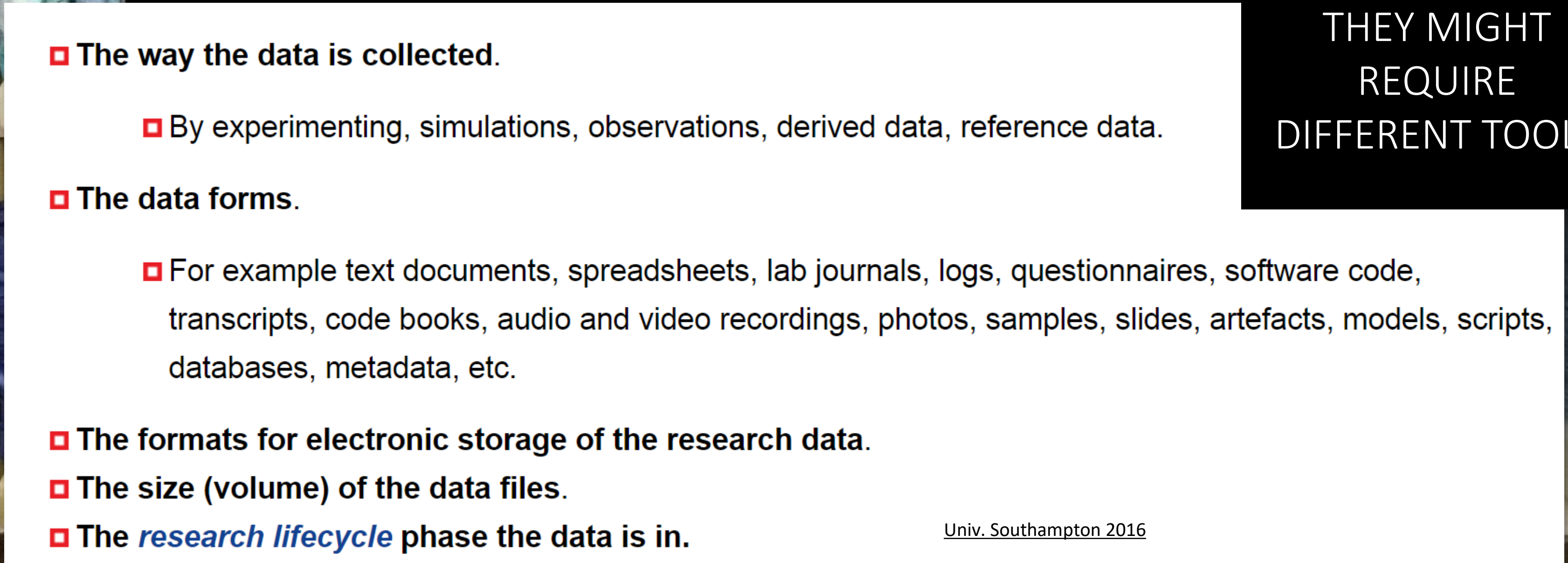
Data basics

[DMP]

5 WAYS TO THINK OF DATA :

- THE WAY DATA ARE COLLECTED
- THEIR FORM
- THEIR FORMAT
- THEIR SIZE/VOLUME
- THE WORKFLOW PHASE THEY ARE IN

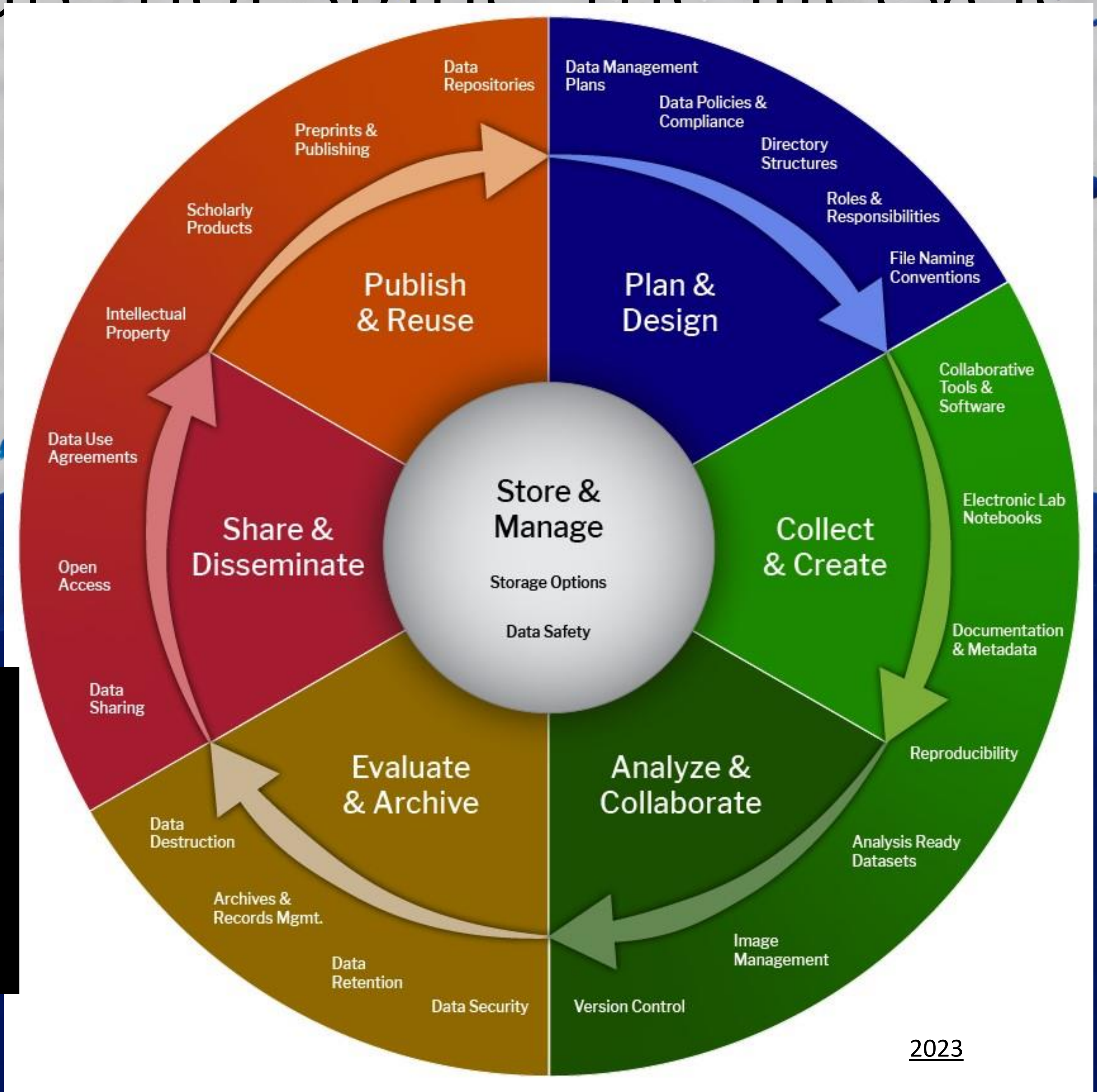
THEY MIGHT REQUIRE DIFFERENT TOOLS



- □ **The way the data is collected.**
  - □ By experimenting, simulations, observations, derived data, reference data.
- □ **The data forms.**
  - □ For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, slides, artefacts, models, scripts, databases, metadata, etc.
- □ **The formats for electronic storage of the research data.**
- □ **The size (volume) of the data files.**
- □ **The *research lifecycle* phase the data is in.**

Univ. Southampton 2016

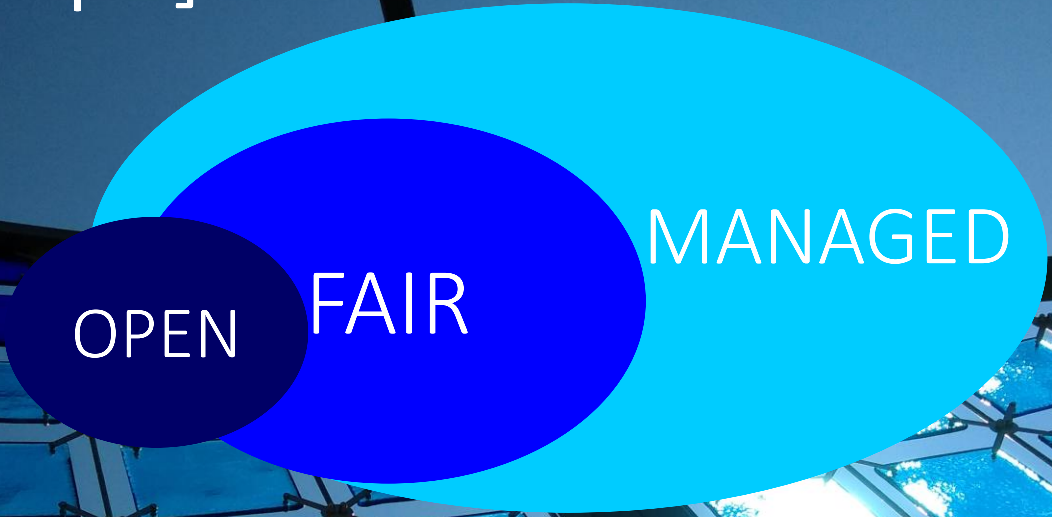
Data are not static: the lifecycle



2023

PLANNIG DATA MANAGEMENT IN EVERY STEP OF THE CYCLE IS CRUCIAL

[the 3 steps]



1. DATA SHOULD BE AS OPEN AS POSSIBLE

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

3. IF DATA ARE NOT PROPERLY MANAGED FROM THE BEGINNING, IT'S ALMOST IMPOSSIBLE TO MAKE THEM «FAIR» [WITH EOSC MANAGED/FAIR INCREASINGLY OVERLAPPING, «FAIR BY DESIGN»]

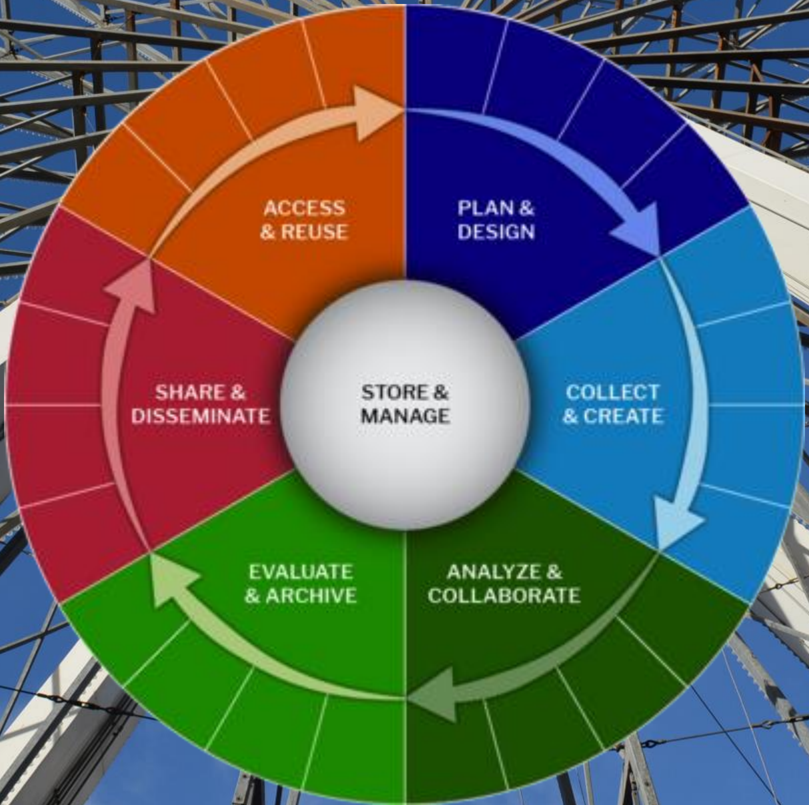
AND MANAGING DATA PROPERLY IS IN THE PRIMARY INTEREST OF ANY RESEARCHER,
AS THE WHOLE RESEARCH PROCESS RESULTS STREAMLINED AND MORE EFFECTIVE

1) Data management

DESCRIPTION FOR DISCOVERABILITY (metadata)

ORGANIZATION (file naming, folders, versioning…)

BACKUP AND STORAGE

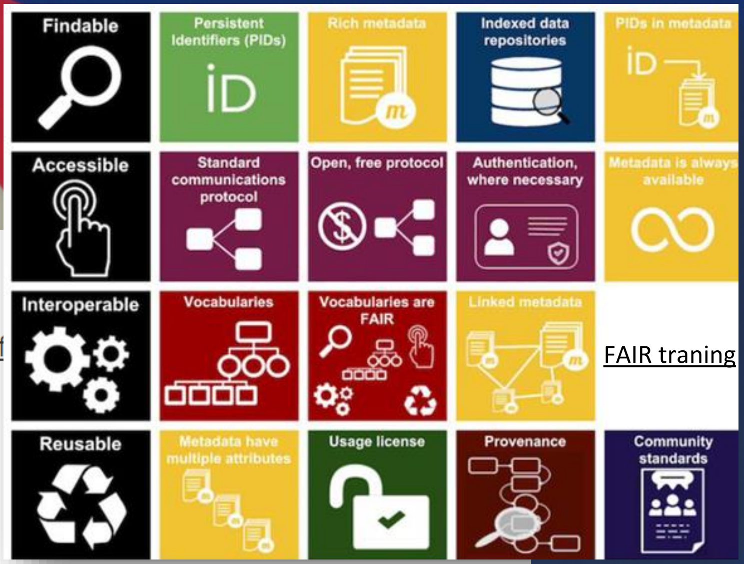


LONG TIME PRESERVATION

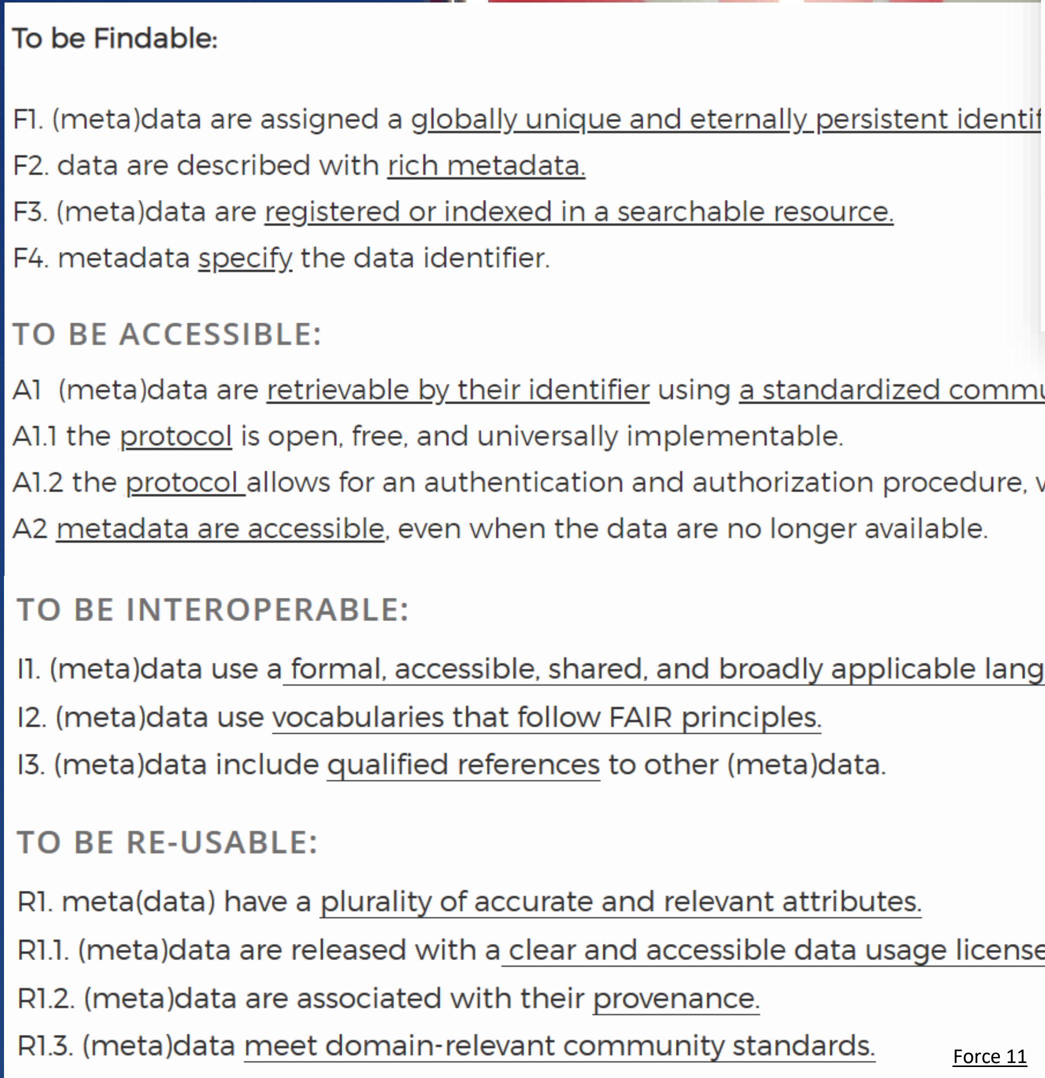
LEGAL ASPECTS

ALONG THE ENTIRE LIFE CYCLE

2) Make data FAIR


Findable
Persistent Identifiers (PIDs)
Rich metadata
Indexed data repositories
PIDs in metadata
Accessible
Standard communications protocol
Open, free protocol
Authentication, where necessary
Metadata is always available
Interoperable
Vocabularies
Vocabularies are FAIR
Linked metadata
Reusable
Metadata have multiple attributes
Usage license
Provenance
Community standards

FAIR traning



To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identif

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 kr

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» DOES NOT MEAN «OPEN».

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

3) Whenever possible, make them Open

YOU SAVE LIVES.

Digital Science Report

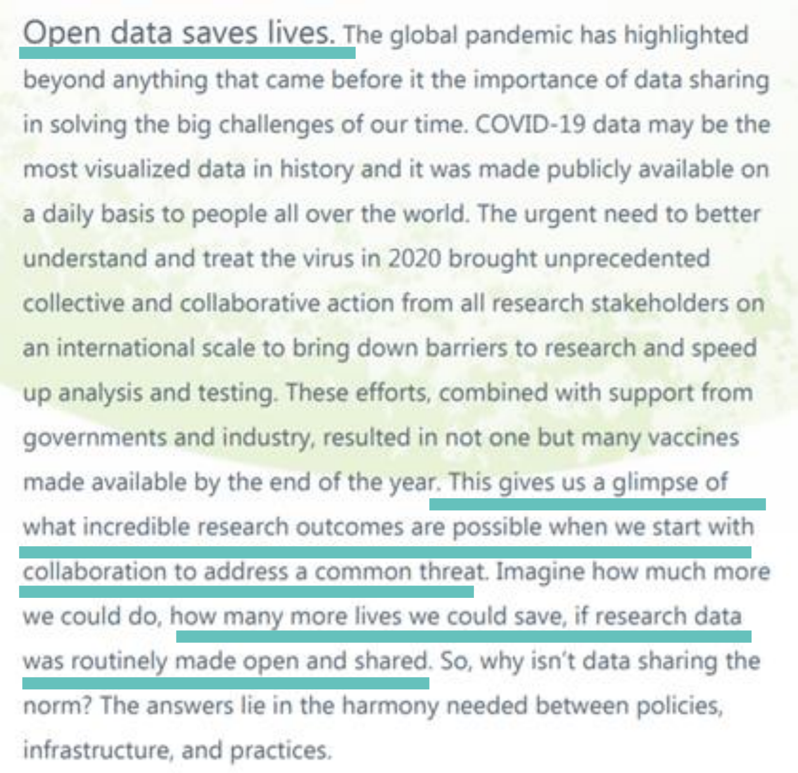
The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

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

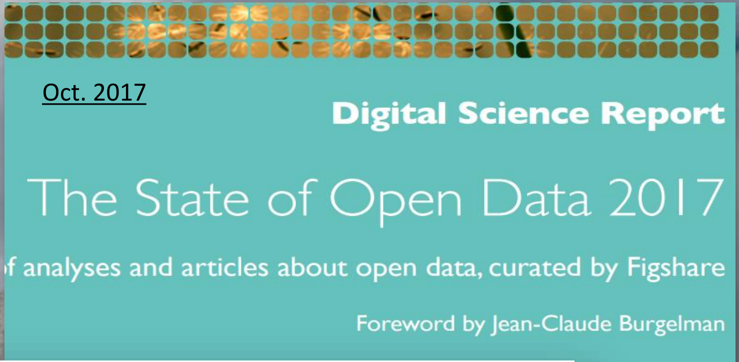
Nov. 29, 2021

November 2021



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

YOU CREATE VALUE



Oct. 2017

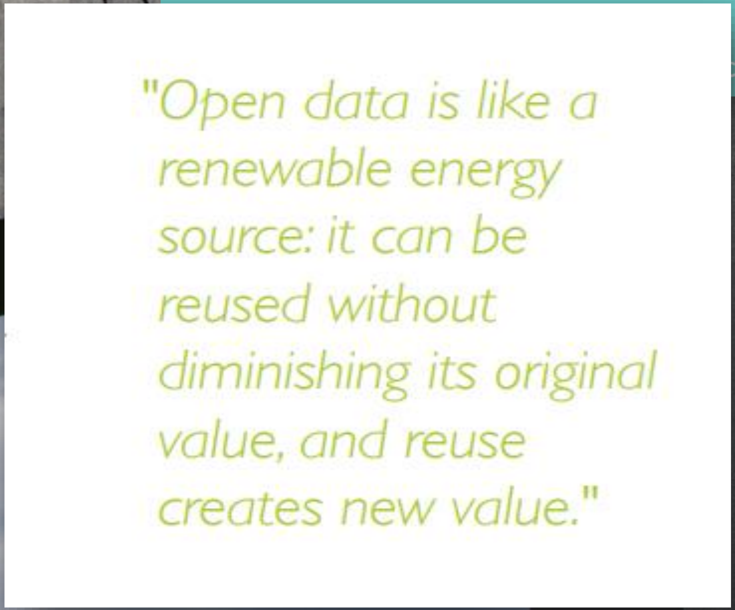
Digital Science Report

The State of Open Data 2017

f analyses and articles about open data, curated by Figshare

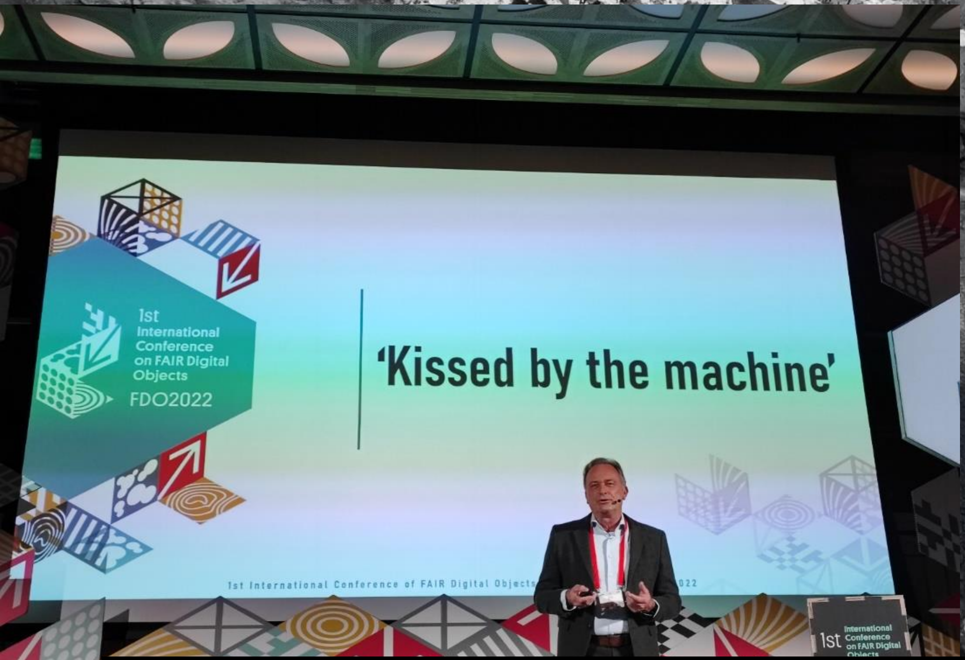
Foreword by Jean-Claude Burgelman

TOBER 2017



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

Kissed or missed?


1st International Conference on FAIR Digital Objects
FDO2022
'Kissed by the machine'
1st International Conference of FAIR Digital Objects

1st International Conference on FAIR Digital Objects
FDO2022
'Missed by the machine'
1st International Conference of FAIR Dig
October 2022

FAIR PRINCIPLES ARE

«MACHINE ACTIONABLE»

(MORE THAN READABLE)

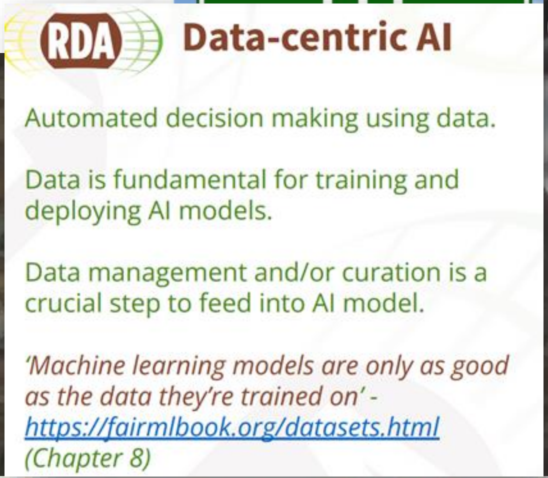
FAIR = FULLY AI READY

IF NOT… YOU'LL BE MISSED (INSTEAD OF KISSED) BY THE MACHINE

RDA

Decision making procedures in data management and data stewardship for Open Science

Connie Clare, PhD



Data-centric AI

Automated decision making using data.

Data is fundamental for training and deploying AI models.

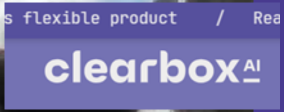
Data management and/or curation is a crucial step to feed into AI model.

'Machine learning models are only as good as the data they're trained on' - *https://fairmlbook.org/datasets.html* *(Chapter 8)*

Clearbox AI

Clearbox

We are on a mission to harness powerful AI technologies t improve businesses and society in a trustworthy and human-centered way.



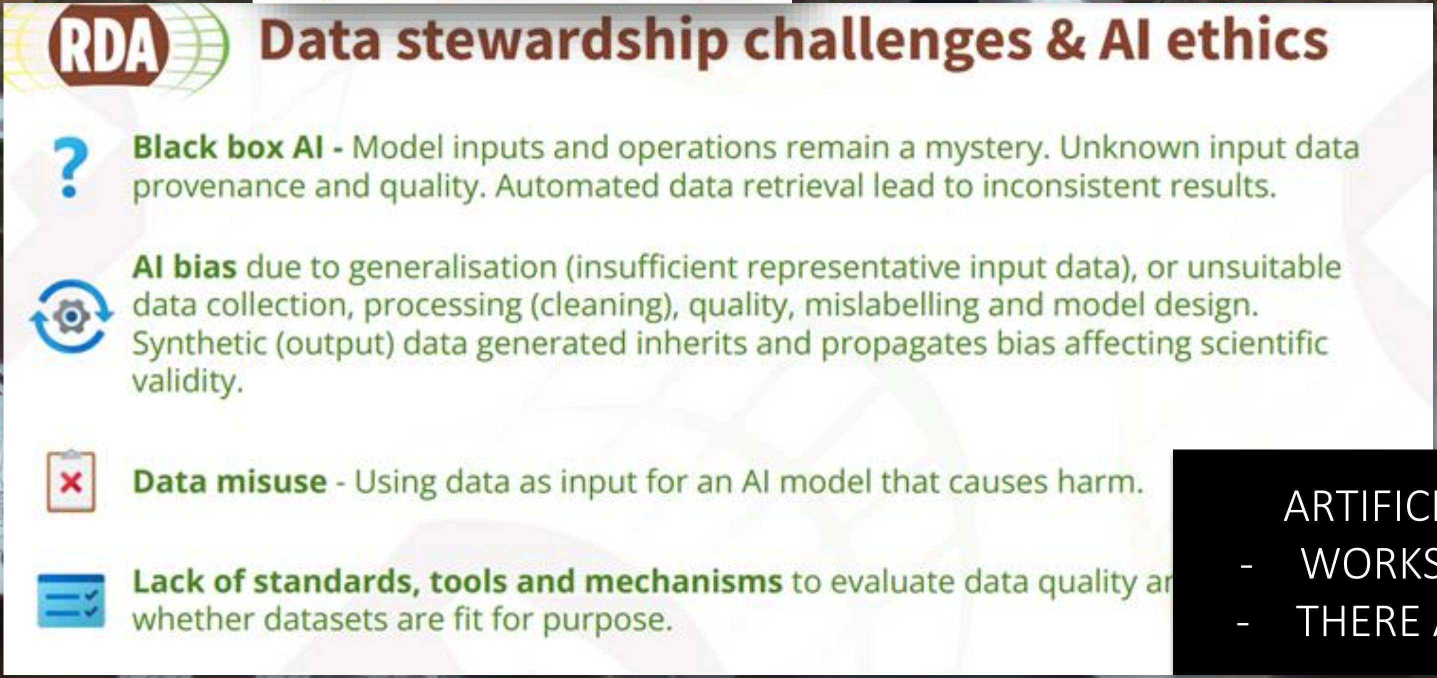
s flexible product / Rea

clearbox AI

Your

Synthetic Data

provider



Data stewardship challenges & AI ethics

Black box AI - Model inputs and operations remain a mystery. Unknown input data provenance and quality. Automated data retrieval lead to inconsistent results.

AI bias due to generalisation (insufficient representative input data), or unsuitable data collection, processing (cleaning), quality, mislabelling and model design. Synthetic (output) data generated inherits and propagates bias affecting scientific validity.

Data misuse - Using data as input for an AI model that causes harm.

Lack of standards, tools and mechanisms to evaluate data quality ar whether datasets are fit for purpose.

ARTIFICIAL INTELLIGENCE

- WORKS IF DATA ARE GOOD
- THERE ARE ETHICAL ISSUES

FAIR/Open

2018

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

TURNING FAIR INTO REALITY

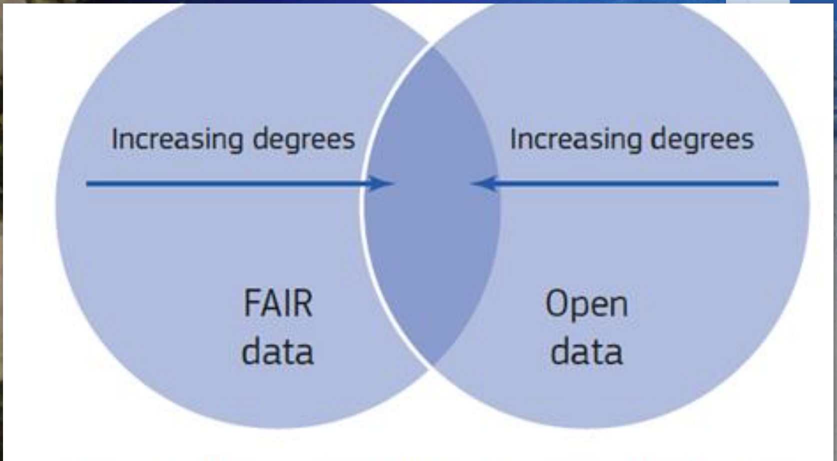
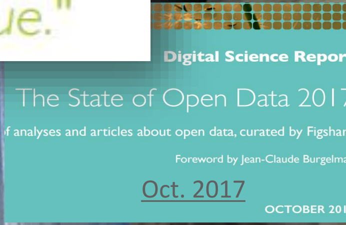


Figure 4. The relationship between FAIR and Open

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



Digital Science Repor

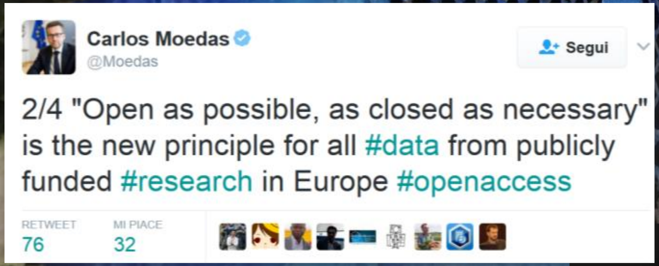
The State of Open Data 2017

f analyses and articles about open data, curated by Figshar

Foreword by Jean-Claude Burgelma

Oct. 2017

OCTOBER 20



Carlos Moedas
@Moedas

Segui

2/4 "Open as possible, as closed as necessary" is the new principle for all #data from publicly funded #research in Europe #openaccess

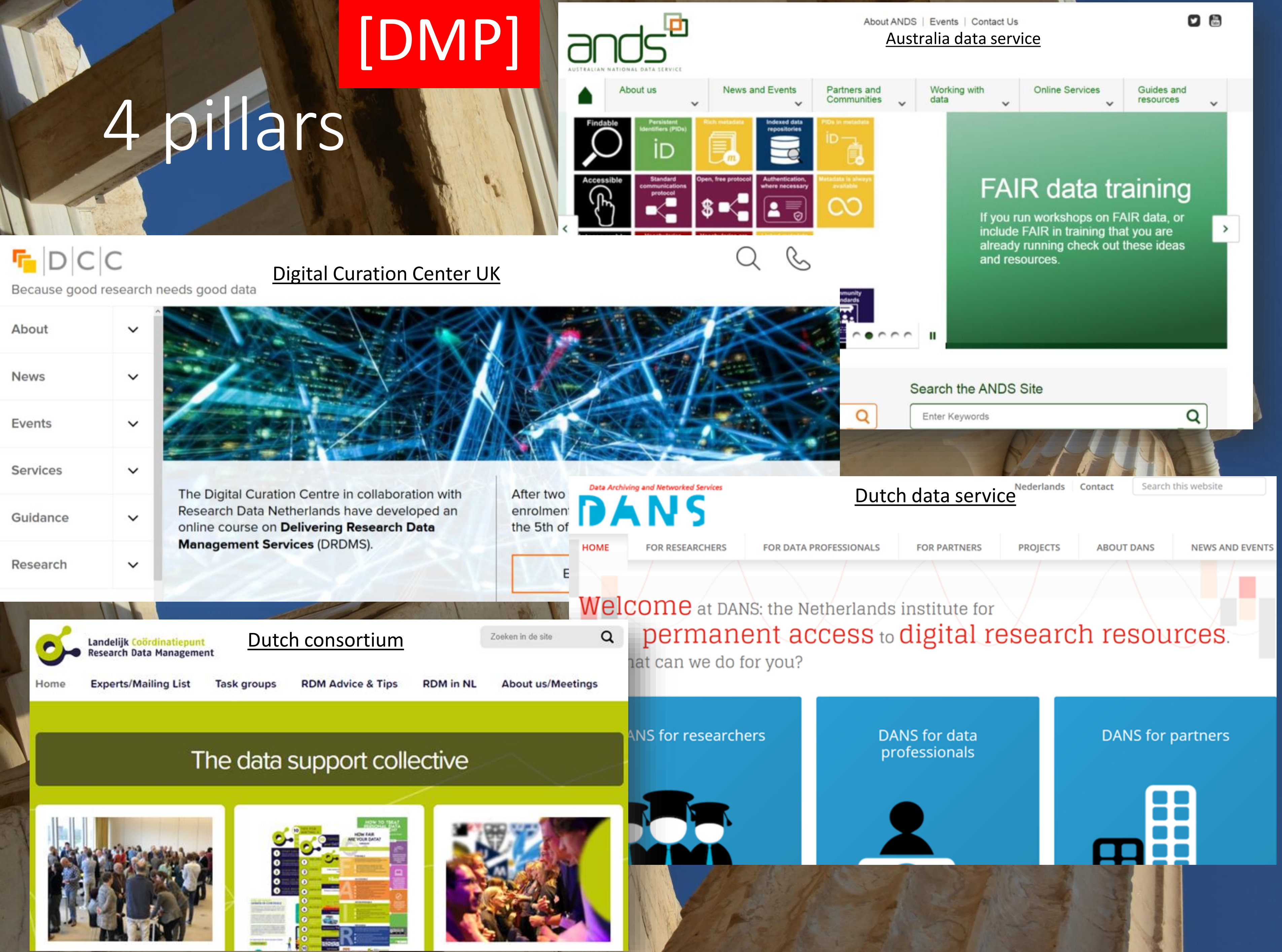
RETWEET 76 MI PIACE 32

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

STEP #1
DATA
MANAGEMENT

[DMP]

4 pillars



Australia data service

ands
AUSTRALIAN NATIONAL DATA SERVICE
About ANDS | Events | Contact Us
About us
News and Events
Partners and Communities
Working with data
Online Services
Guides and resources
Findable
Persistent Identifiers (PIDs)
Rich metadata
Indexed data repositories
PIDs in metadata
Accessible
Standard communications protocol
Open, free protocol
Authentication, where necessary
Metadata is always available
Search the ANDS Site
Enter Keywords

FAIR data training

If you run workshops on FAIR data, or include FAIR in training that you are already running check out these ideas and resources.

Digital Curation Center UK

DCC
Because good research needs good data
About
News
Events
Services
Guidance
Research

The Digital Curation Centre in collaboration with Research Data Netherlands have developed an online course on **Delivering Research Data Management Services** (DRDMS).

After two
enrolmen
the 5th of

Dutch data service

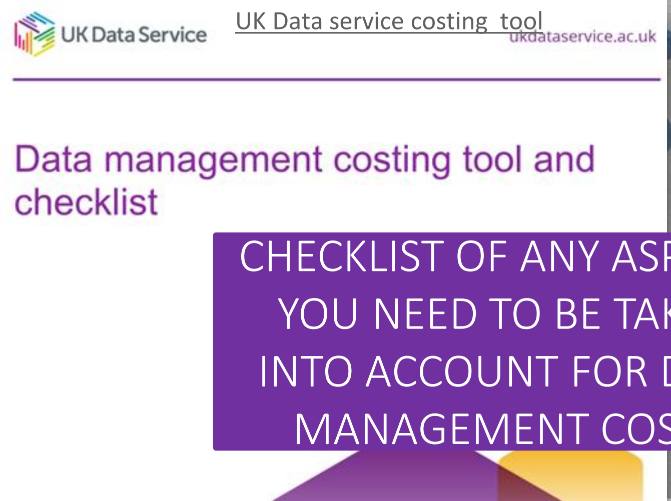
Welcome at DANS: the Netherlands institute for permanent access to digital research resources.

hat can we do for you?

Dutch consortium

The data support collective

Costs



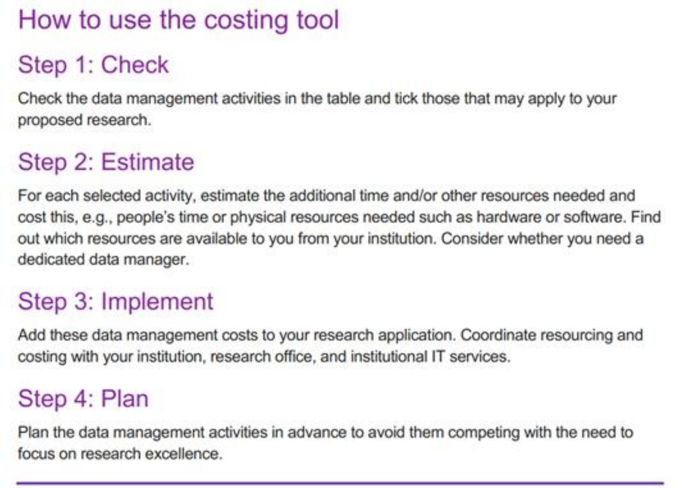
UK Data Service

ukdataservice.ac.uk

Data management costing tool and checklist

UK Data service costing tool

CHECKLIST OF ANY ASPECTS YOU NEED TO BE TAKEN INTO ACCOUNT FOR DATA MANAGEMENT COSTS



How to use the costing tool

Step 1: Check

Check the data management activities in the table and tick those that may apply to your proposed research.

Step 2: Estimate

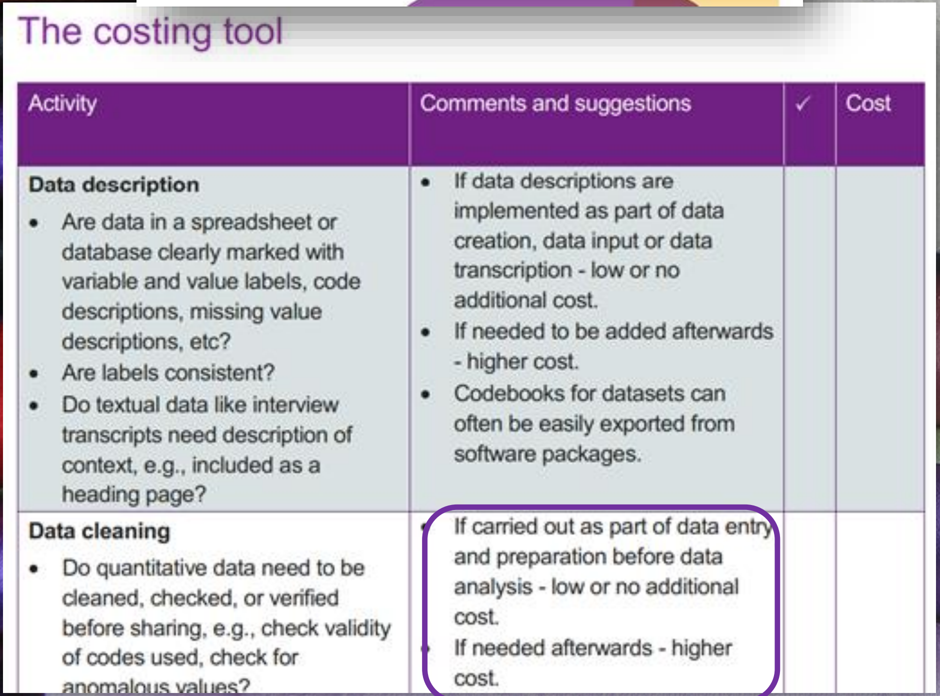
For each selected activity, estimate the additional time and/or other resources needed and cost this, e.g., people's time or physical resources needed such as hardware or software. Find out which resources are available to you from your institution. Consider whether you need a dedicated data manager.

Step 3: Implement

Add these data management costs to your research application. Coordinate resourcing and costing with your institution, research office, and institutional IT services.

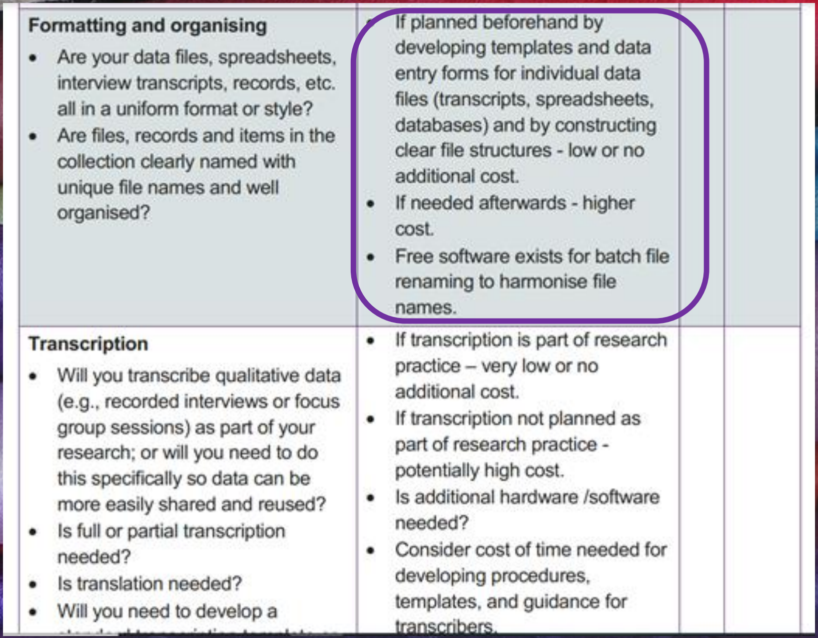
Step 4: Plan

Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.



The costing tool

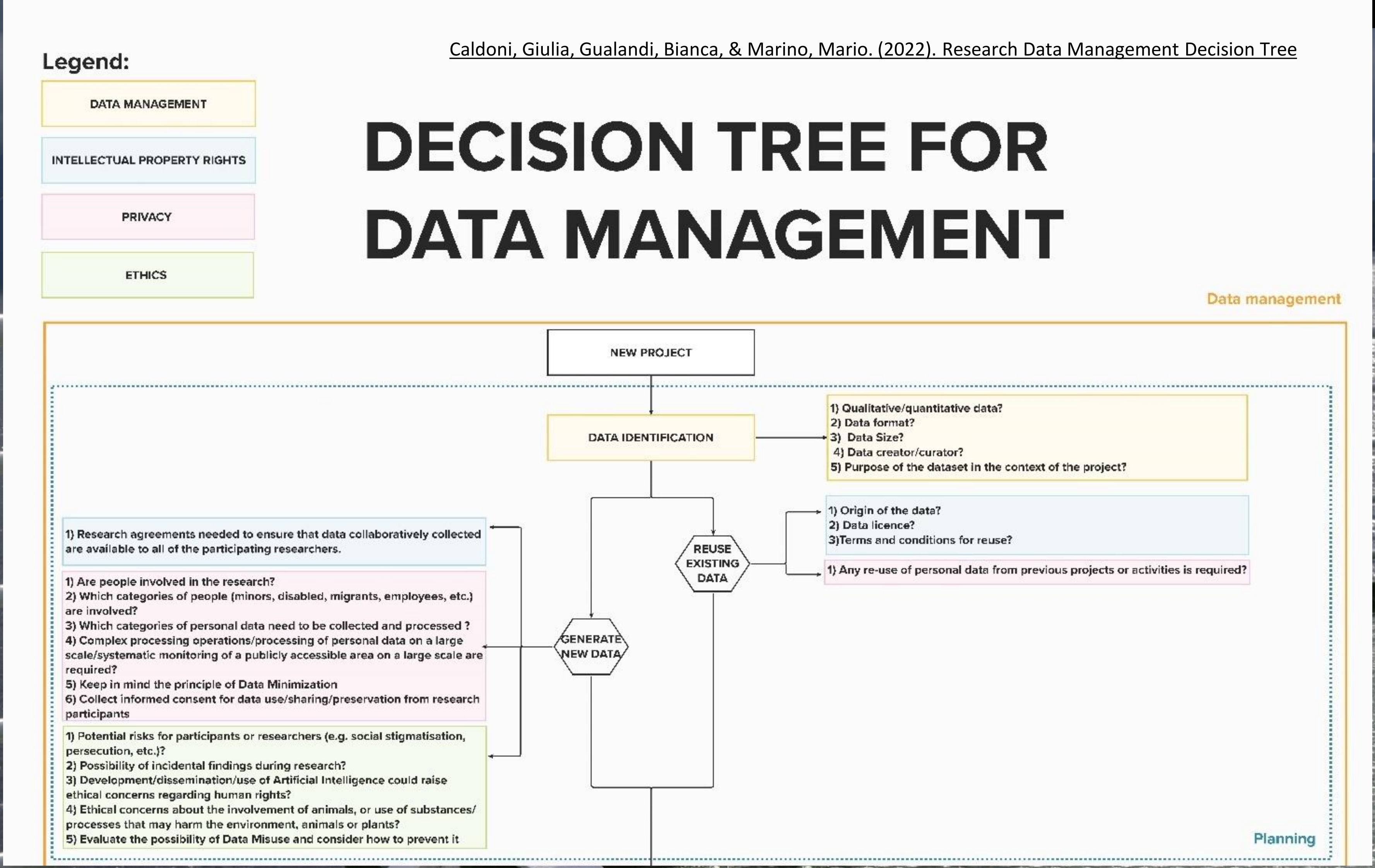
Activity	Comments and suggestions	✓	Cost
Data description • Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc? • Are labels consistent? • Do textual data like interview transcripts need description of context, e.g., included as a heading page?	• If data descriptions are implemented as part of data creation, data input or data transcription - low or no additional cost. • If needed to be added afterwards - higher cost. • Codebooks for datasets can often be easily exported from software packages.		
Data cleaning • Do quantitative data need to be cleaned, checked, or verified before sharing, e.g., check validity of codes used, check for anomalous values?	• If carried out as part of data entry and preparation before data analysis - low or no additional cost. • If needed afterwards - higher cost.		



Formatting and organising • Are your data files, spreadsheets, interview transcripts, records, etc. all in a uniform format or style? • Are files, records and items in the collection clearly named with unique file names and well organised?	• If planned beforehand by developing templates and data entry forms for individual data files (transcripts, spreadsheets, databases) and by constructing clear file structures - low or no additional cost. • If needed afterwards - higher cost. • Free software exists for batch file renaming to harmonise file names.		
Transcription • Will you transcribe qualitative data (e.g., recorded interviews or focus group sessions) as part of your research; or will you need to do this specifically so data can be more easily shared and reused? • Is full or partial transcription needed? • Is translation needed? • Will you need to develop a	• If transcription is part of research practice – very low or no additional cost. • If transcription not planned as part of research practice - potentially high cost. • Is additional hardware /software needed? • Consider cost of time needed for developing procedures, templates, and guidance for transcribers.		

Before boarding

USEFUL TOOL AS A FIRST APPROACH TO DATA MANAGEMENT [PLAN]



Caldoni, Giulia, Gualandi, Bianca, & Marino, Mario. (2022). Research Data Management Decision Tree

Legend:

- DATA MANAGEMENT
- INTELLECTUAL PROPERTY RIGHTS
- PRIVACY
- ETHICS

DECISION TREE FOR DATA MANAGEMENT

Data management

NEW PROJECT

DATA IDENTIFICATION

1) Qualitative/quantitative data?
2) Data format?
3) Data Size?
4) Data creator/curator?
5) Purpose of the dataset in the context of the project?

REUSE EXISTING DATA

1) Origin of the data?
2) Data licence?
3)Terms and conditions for reuse?

1) Any re-use of personal data from previous projects or activities is required?

GENERATE NEW DATA

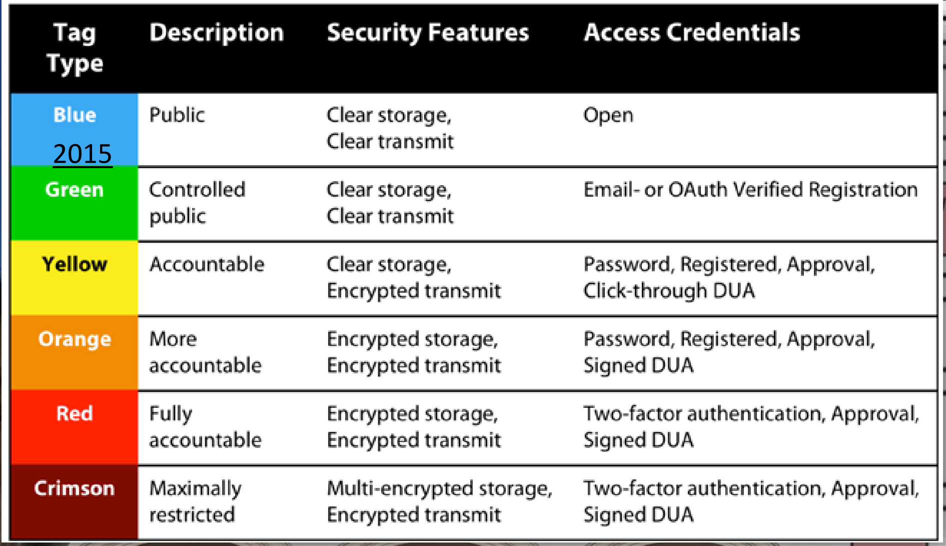
1) Research agreements needed to ensure that data collaboratively collected are available to all of the participating researchers.

1) Are people involved in the research?
2) Which categories of people (minors, disabled, migrants, employees, etc.) are involved?
3) Which categories of personal data need to be collected and processed ?
4) Complex processing operations/processing of personal data on a large scale/systematic monitoring of a publicly accessible area on a large scale are required?
5) Keep in mind the principle of Data Minimization
6) Collect informed consent for data use/sharing/preservation from research participants

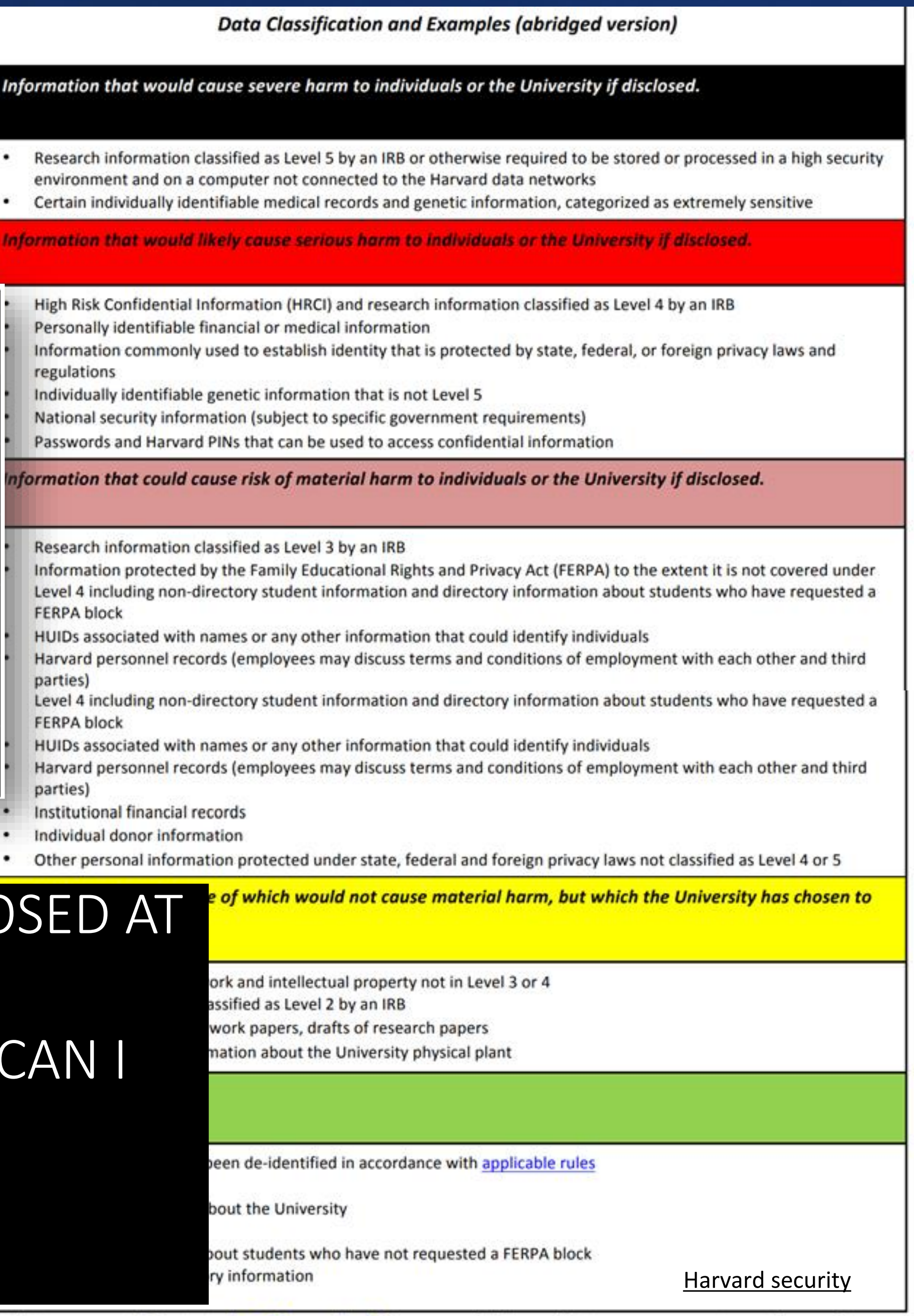
1) Potential risks for participants or researchers (e.g. social stigmatisation, persecution, etc.)?
2) Possibility of incidental findings during research?
3) Development/dissemination/use of Artificial Intelligence could raise ethical concerns regarding human rights?
4) Ethical concerns about the involvement of animals, or use of substances/processes that may harm the environment, animals or plants?
5) Evaluate the possibility of Data Misuse and consider how to prevent it

Planning

[remind: it’s not open/close at th



Tag Type	Description	Security Features	Access Credentials
Blue 2015	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA



Level	Data Classification and Examples (abridged version)
5	Information that would cause severe harm to individuals or the University if disclosed.
	• Research information classified as Level 5 by an IRB or otherwise required to be stored or processed in a high security environment and on a computer not connected to the Harvard data networks • Certain individually identifiable medical records and genetic information, categorized as extremely sensitive
4	Information that would likely cause serious harm to individuals or the University if disclosed.
	• High Risk Confidential Information (HRCI) and research information classified as Level 4 by an IRB • Personally identifiable financial or medical information • Information commonly used to establish identity that is protected by state, federal, or foreign privacy laws and regulations • Individually identifiable genetic information that is not Level 5 • National security information (subject to specific government requirements) • Passwords and Harvard PINs that can be used to access confidential information
	Information that could cause risk of material harm to individuals or the University if disclosed.
	• Research information classified as Level 3 by an IRB • Information protected by the Family Educational Rights and Privacy Act (FERPA) to the extent it is not covered under Level 4 including non-directory student information and directory information about students who have requested a FERPA block • HUIDs associated with names or any other information that could identify individuals • Harvard personnel records (employees may discuss terms and conditions of employment with each other and third parties) • Level 4 including non-directory student information and directory information about students who have requested a FERPA block • HUIDs associated with names or any other information that could identify individuals • Harvard personnel records (employees may discuss terms and conditions of employment with each other and third parties) • Institutional financial records • Individual donor information • Other personal information protected under state, federal and foreign privacy laws not classified as Level 4 or 5
	e of which would not cause material harm, but which the University has chosen to
	ork and intellectual property not in Level 3 or 4 assified as Level 2 by an IRB work papers, drafts of research papers mation about the University physical plant
	een de-identified in accordance with applicable rules bout the University out students who have not requested a FERPA block ry information

Harvard security

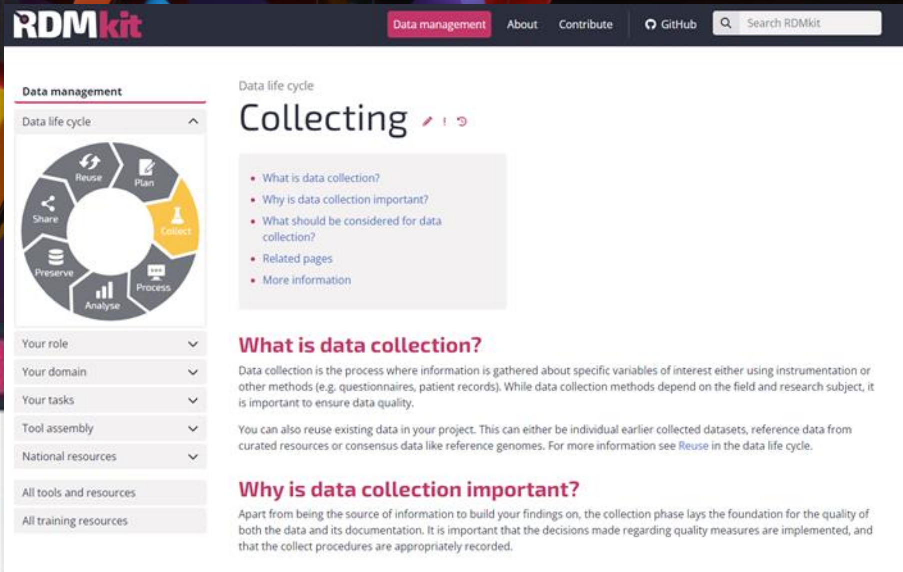
...THE ISSUE IS NOT JUST OPEN/CLOSED AT THE END.

DURING MY RESEARCH, WHERE CAN I SAFELY STORE THE DATA?

WHO CAN ACCESS THEM?

WHAT ABOUT SECURITY?

Training



Data life cycle

Collecting

- What is data collection?
- Why is data collection important?
- What should be considered for data collection?
- Related pages
- More information

What is data collection?

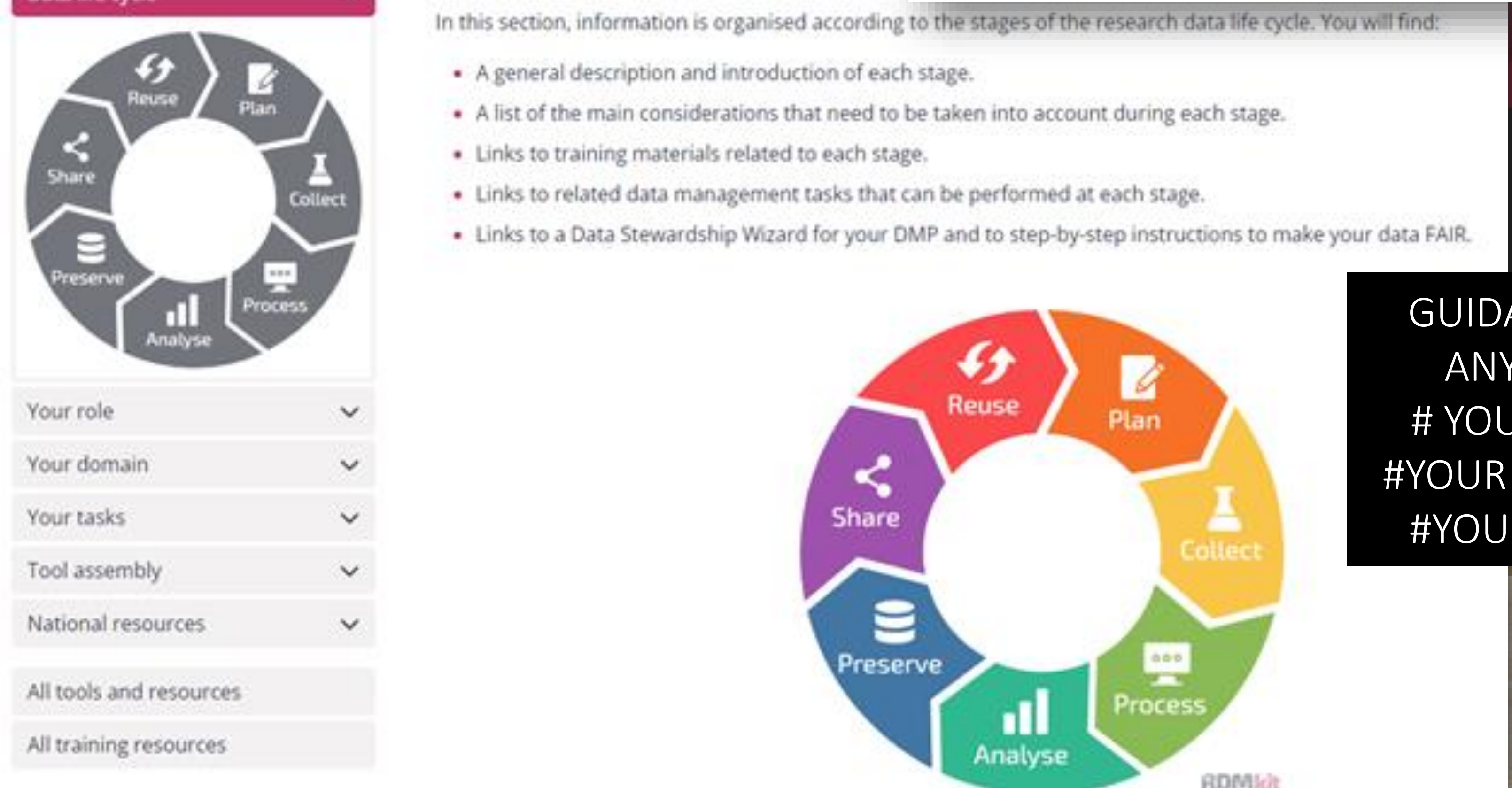
Data collection is the process where information is gathered about specific variables of interest either using instrumentation or other methods (e.g. questionnaires, patient records). While data collection methods depend on the field and research subject, it is important to ensure data quality.

You can also reuse existing data in your project. This can either be individual earlier collected datasets, reference data from curated resources or consensus data like reference genomes. For more information see Reuse in the data life cycle.

Why is data collection important?

Apart from being the source of information to build your findings on, the collection phase lays the foundation for the quality of both the data and its documentation. It is important that the decisions made regarding quality measures are implemented, and that the collect procedures are appropriately recorded.

RDM kit



Data life cycle

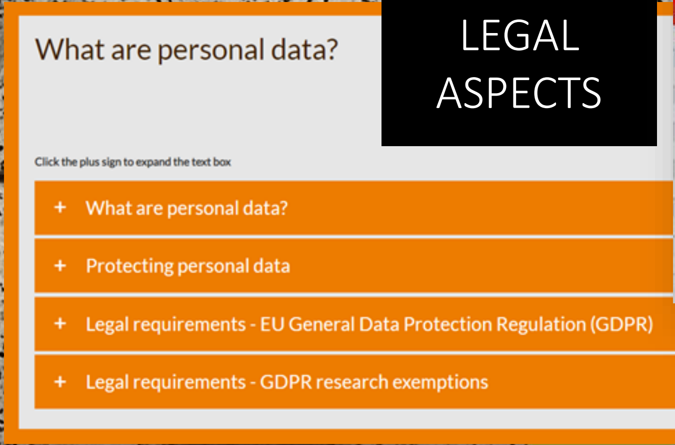
In this section, information is organised according to the stages of the research data life cycle. You will find:

- A general description and introduction of each stage.
- A list of the main considerations that need to be taken into account during each stage.
- Links to training materials related to each stage.
- Links to related data management tasks that can be performed at each stage.
- Links to a Data Stewardship Wizard for your DMP and to step-by-step instructions to make your data FAIR.

GUIDANCE IN ANY STEP
# YOUR ROLE
#YOUR DOMAIN
#YOUR TASKS

Data management ABC

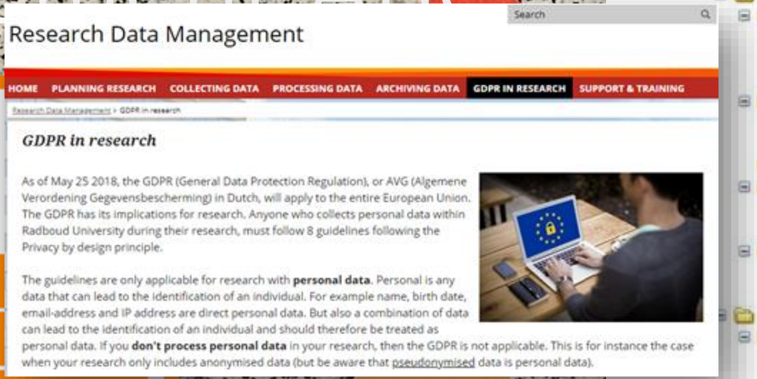
LEGAL ASPECTS



What are personal data?

Click the plus sign to expand the text box

- + What are personal data?
- + Protecting personal data
- + Legal requirements - EU General Data Protection Regulation (GDPR)
- + Legal requirements - GDPR research exemptions



Research Data Management

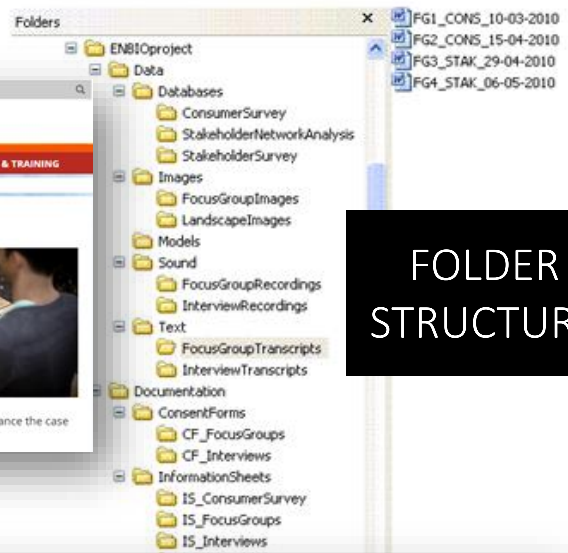
GDPR in research

As of May 25 2018, the GDPR (General Data Protection Regulation), or AVG (Algemene Verordening Gegevensbescherming) in Dutch, will apply to the entire European Union. The GDPR has its implications for research. Anyone who collects personal data within Radboud University during their research, must follow 8 guidelines following the Privacy by design principle.

The guidelines are only applicable for research with **personal data**. Personal is any data that can lead to the identification of an individual. For example name, birth date, email-address and IP address are direct personal data. But also a combination of data can lead to the identification of an individual and should therefore be treated as personal data. If you **don't process personal data** in your research, then the GDPR is not applicable. This is for instance the case when your research only includes anonymised data (but be aware that pseudonymised data is personal data).

FOLDER STRUCTURE

It helps to restrict the level of folders to three or four deep and not to have more than ter items in each list.



```
ENBIOproject
  Data
    Databases
      ConsumerSurvey
      StakeholderNetworkAnalysis
      StakeholderSurvey
    Images
      FocusGroupImages
      LandscapeImages
    Models
    Sound
      FocusGroupRecordings
      InterviewRecordings
    Text
      FocusGroupTranscripts
      InterviewTranscripts
  Documentation
    ConsentForms
      CF_FocusGroups
      CF_Interviews
    InformationSheets
      IS_ConsumerSurvey
      IS_FocusGroups
      IS_Interviews
```

```
FG1_CONS_10-03-2010
FG2_CONS_15-04-2010
FG3_STAK_29-04-2010
FG4_STAK_06-05-2010
```

FILE NAMING
CHOOSE A SCHEMA AND BE CONSISTENT!
[ALL THE MORE SO IF YOU HAVE PARTNERS]

File naming conventions

File naming

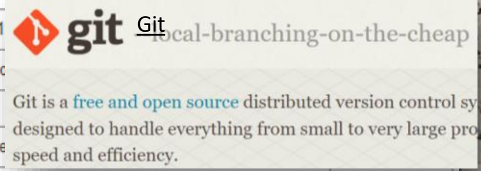
The conventions comprise the following 13 rules. Follow the links for examples and explanations of the rules.

1. Keep file names short, but meaningful
2. Avoid unnecessary repetition and redundancy in file names and file paths.
3. Use capital letters to delimit words, not spaces or underscores
4. When including a number in a file name always give it as a two-digit number, i.e. 01-99, unless it is a year or another number with m
5. If using a date in the file name always state the date 'back to front', and use four digit years, two digit months and two digit days: YY or YYYY or YYYY-YYYY.
6. When including a personal name in a file name give the family name first followed by the initials.
7. Avoid using common words such as 'draft' or 'letter' at the start of file names, unless doing so will make it easier to retrieve the rec
8. Order the elements in a file name in the most appropriate way to retrieve the record.
9. The file names of records relating to recurring events should include the date and a description of the event, except where the inclu of these elements would be incompatible with rule 2.
10. The file names of correspondence should include the name of the correspondent, an indication of the subject, the date of the corre whether it is incoming or outgoing correspondence, except where the inclusion of any of these elements would be incompatible wi
11. The file name of an email attachment should include the name of the correspondent, an indication of the subject, the date of the c 'attch', and an indication of the number of attachments sent with the covering email, except where the inclusion of any of these ele incompatible with rule 2.
12. The version number of a record should be indicated in its file name by the inclusion of 'V' followed by the version number and, where applicable, 'Draft'.
13. Avoid using non-alphanumeric characters in file names.

VERSIONING

Example version control table:

Title:	Vision screening tests in Essex nurseries		
File Name:	VisionScreenResults_00_05		
Description:	Results data of 120 Vision Screen Tests carried out in 5 nurseries in Essex during June 2007		
Created By:	Chris Wilkinson		
Maintained By:	Sally Watsley		
Created:	04/07/2007		
Last Modified:	25/1		
Based on:	Visio		
Version	Responsible	Note	
00_05	Sally Watsley	Version 00_03 and 00_04 compared and merged by SW	25/11/2007
00_04	Vani Yussu	Entries checked by VY, independent from SK	17/10/2007
00_03	Steve Knight	Entries checked by SK	29/07/2007
00_02	Karin Mills	Test results 81-120 entered	05/07/2007
00_01	Karin Mills	Test results 1-80 entered	04/07/2007



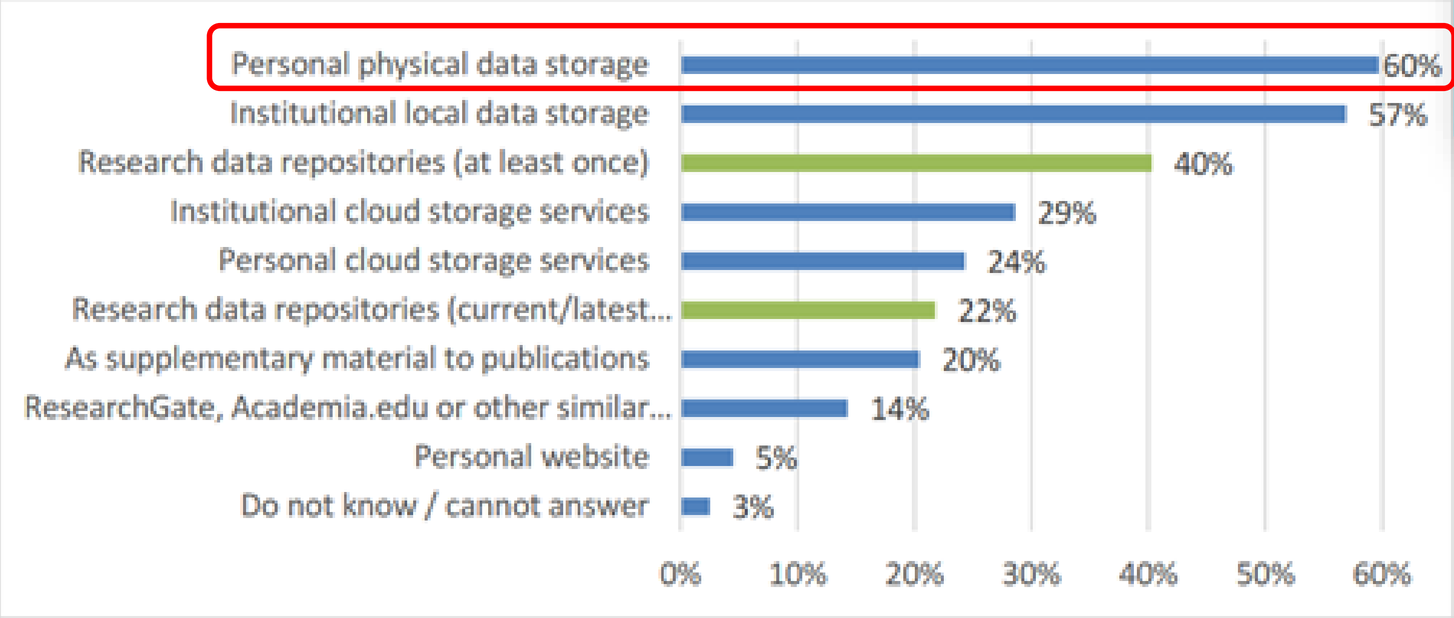
Git is a free and open source distributed version control sy designed to handle everything from small to very large pro speed and efficiency.

ervice

Data management ABC / storage

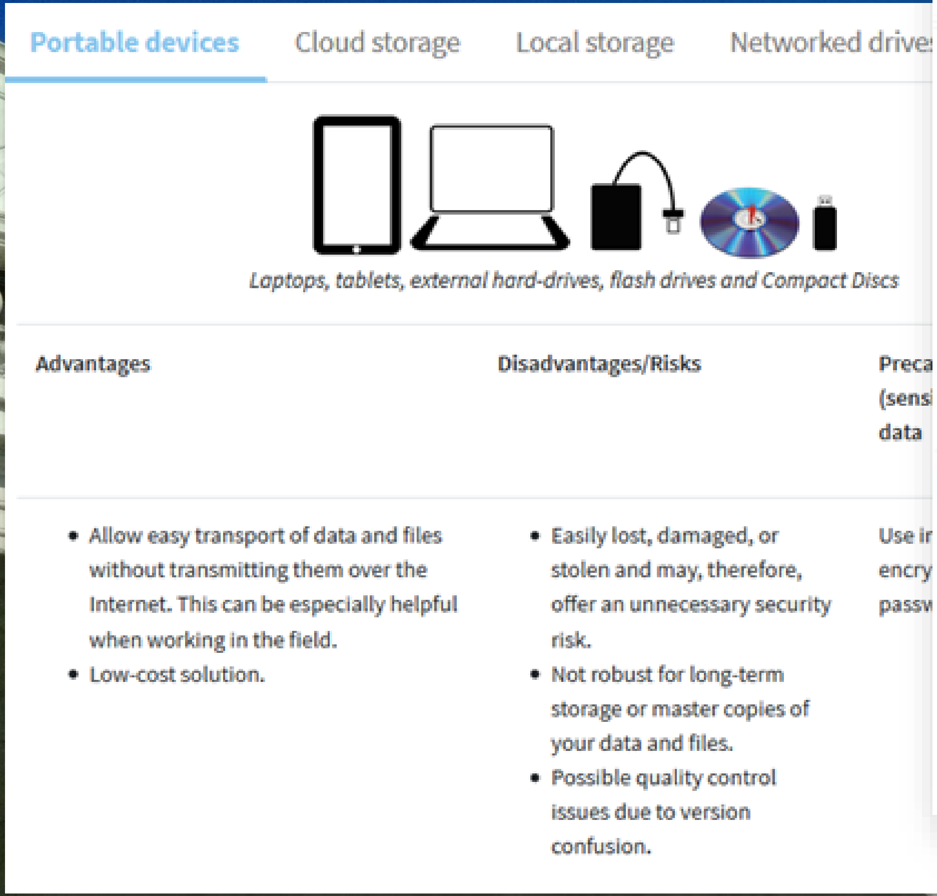
WHERE DO YOU STORE YOUR DATA?

Figure 12. Locations in which respondents or their research teams stored usable data during their current/most recent research activity



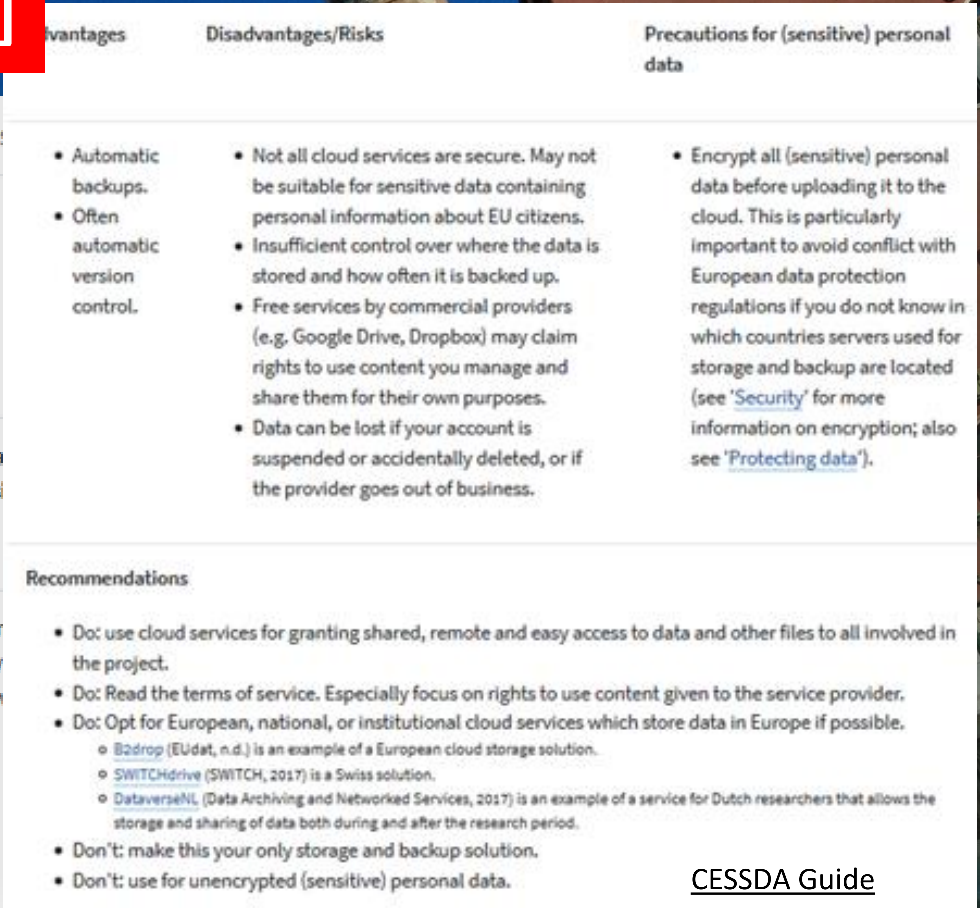
Note: Multiple answers could be selected by a single respondent. Results from the question 'Have you ever stored your research data in a research data repository?' have also been integrated into the figure. Only researchers who did not select research data repositories in the question 'Where have you or your research

Data Management ABC- backup and storage [DMP]



Laptops, tablets, external hard-drives, flash drives and Compact Discs

Advantages	Disadvantages/Risks	Preca (sensi data
• Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field. • Low-cost solution.	• Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk. • Not robust for long-term storage or master copies of your data and files. • Possible quality control issues due to version confusion.	Use ir encry passw



lvantages	Disadvantages/Risks	Precautions for (sensitive) personal data
• Automatic backups. • Often automatic version control.	• Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens. • Insufficient control over where the data is stored and how often it is backed up. • Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes. • Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.	• Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see 'Security' for more information on encryption; also see 'Protecting data').

Recommendations

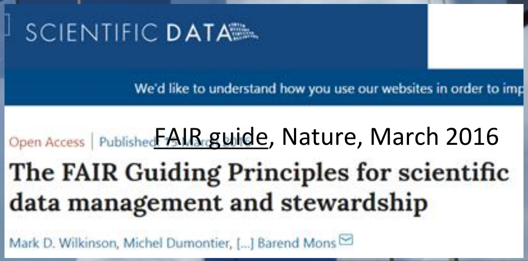
- Do: use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.
- Do: Read the terms of service. Especially focus on rights to use content given to the service provider.
- Do: Opt for European, national, or institutional cloud services which store data in Europe if possible.
  - B2drop (EUdat, n.d.) is an example of a European cloud storage solution.
  - SWITCHdrive (SWITCH, 2017) is a Swiss solution.
  - DataverseNL (Data Archiving and Networked Services, 2017) is an example of a service for Dutch researchers that allows the storage and sharing of data both during and after the research period.
- Don't: make this your only storage and backup solution.
- Don't: use for unencrypted (sensitive) personal data.

CESSDA Guide

DIFFERENT TOOLS FOR DIFFERENT STEPS OF THE RESEARCH CYCLE.
DURING THE EXPERIMENT YOU ALSO NEED TO COLLABORATE WITH THE TEAM

STEP #2
FAIR DATA

…FAIR means [for machines]



SCIENTIFIC DATA

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

Open Access | Published

FAIR guide, Nature, March 2016

The FAIR Guiding Principles for scientific data management and stewardship

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

Findable

- IDENTIFIERS
- METADATA

Accessible

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

Interoperable

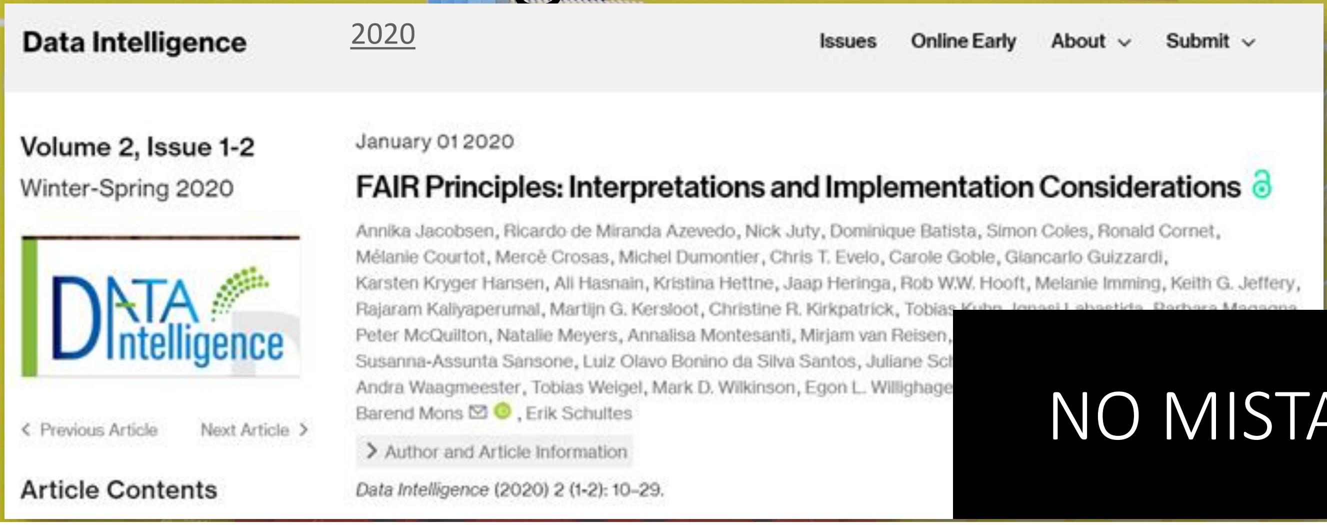
- STANDARDS
- ONTOLOGIES

Reusable

- LICENSES
- DOCUMENTATION

MACHINE-READABLE

…before starting for FAIR



Data Intelligence

2020

Volume 2, Issue 1-2

Winter-Spring 2020

Article Contents

January 01 2020

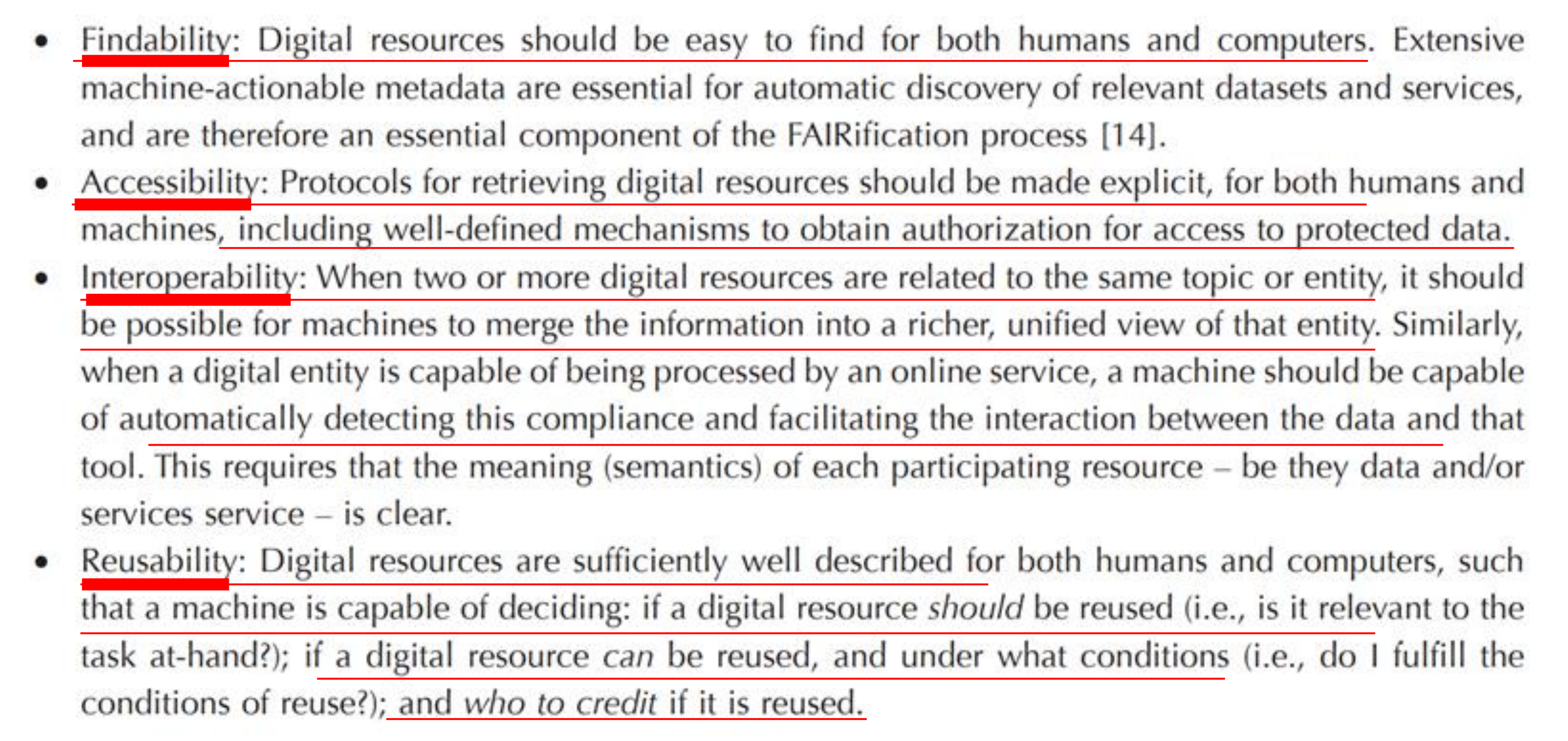
FAIR Principles: Interpretations and Implementation Considerations

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

› Author and Article Information

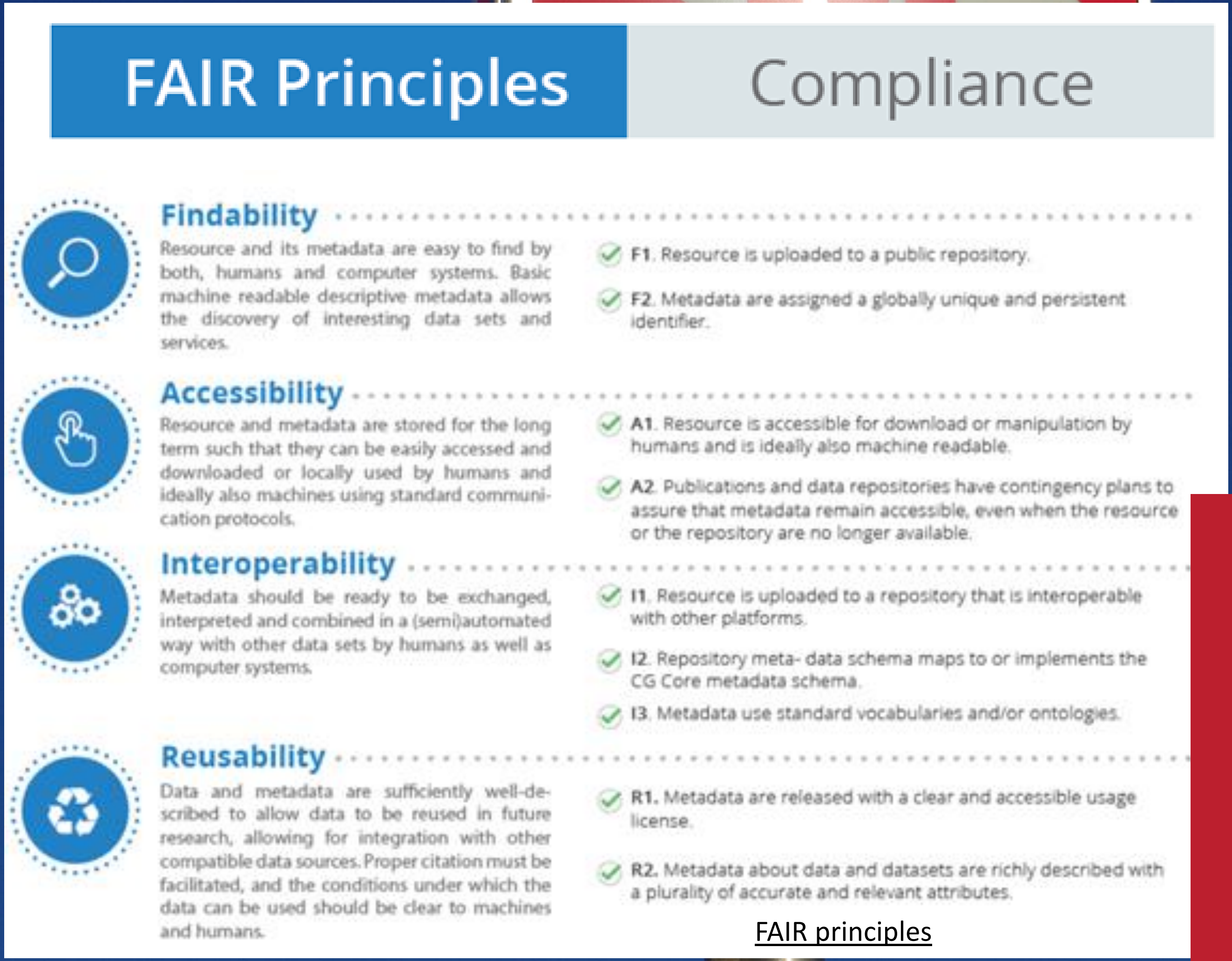
Data Intelligence (2020) 2 (1-2): 10–29.

NO MISTAKES!



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

FAIR principles



FAIR Principles | Compliance

Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ **F1**. Resource is uploaded to a public repository.
- ✓ **F2**. Metadata are assigned a globally unique and persistent identifier.

Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ **A1**. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ **A2**. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.

Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ **I1**. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ **I2**. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ **I3**. Metadata use standard vocabularies and/or ontologies.

Reusability

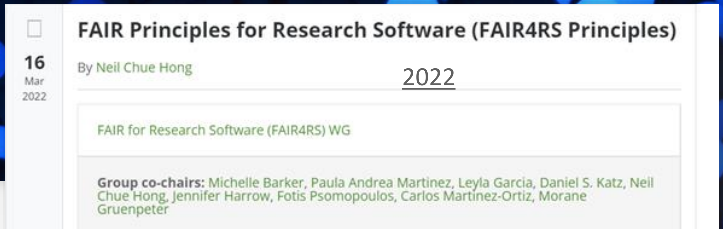
Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

- ✓ **R1.** Metadata are released with a clear and accessible usage license.
- ✓ **R2.** Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.

FAIR principles

«ACCESSIBLE»
DOES NOT MEAN
«OPEN».
DATA CAN BE CLOSED, PROVIDED YOU – AND MACHINES - KNOW WHERE TO FIND THEM AND UNDER WHAT ACCESS CONDITIONS

FAIR research software



16 Mar 2022

FAIR Principles for Research Software (FAIR4RS Principles)

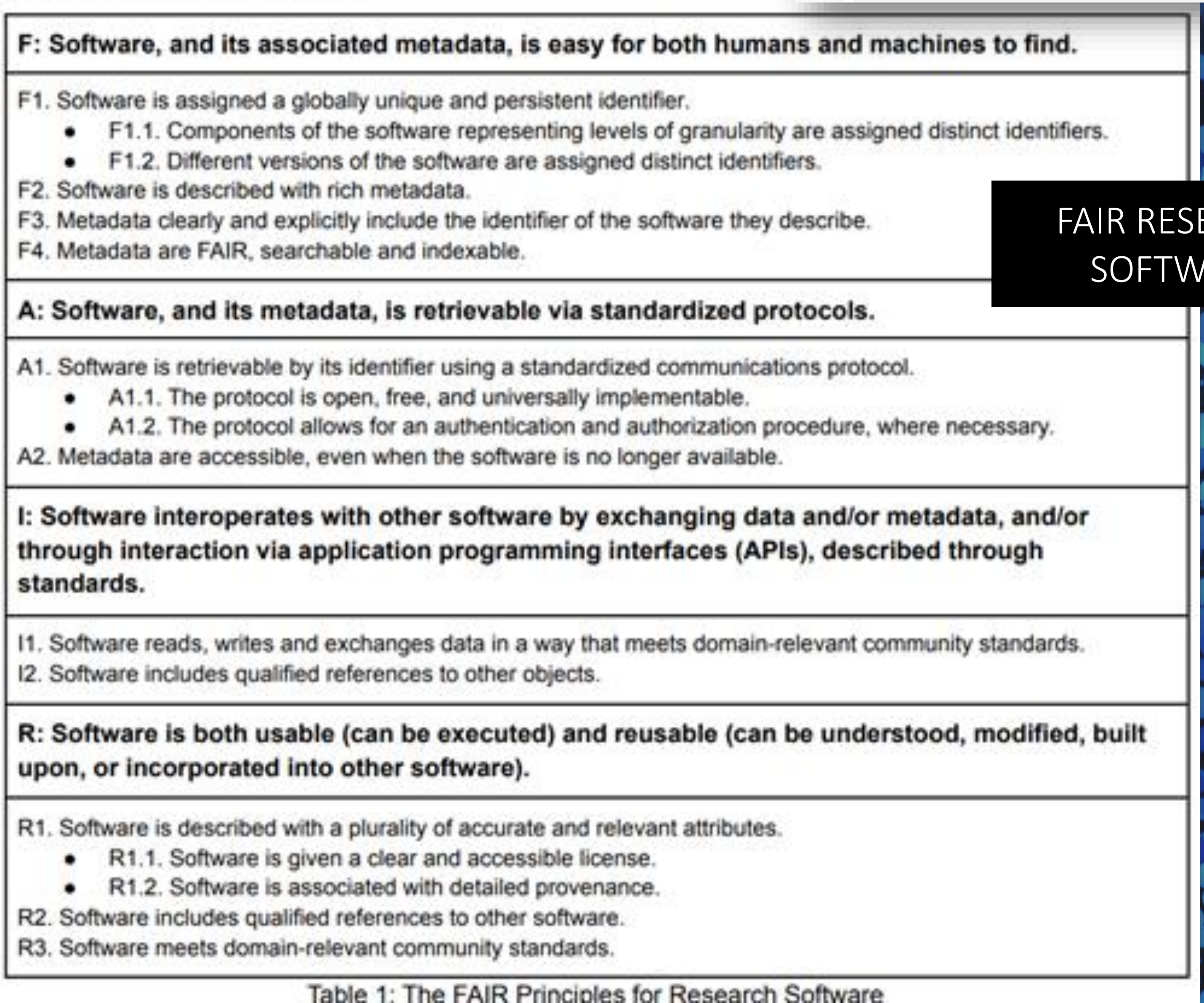
By Neil Chue Hong

2022

FAIR for Research Software (FAIR4RS) WG

Group co-chairs: Michelle Barker, Paula Andrea Martinez, Leyla Garcia, Daniel S. Katz, Neil Chue Hong, Jennifer Harrow, Fotis Psomopoulos, Carlos Martinez-Ortiz, Morane Gruenpeter

The FAIR4RS Principles are:



F: Software, and its associated metadata, is easy for both humans and machines to find.

F1. Software is assigned a globally unique and persistent identifier.
- F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.
- F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.
F3. Metadata clearly and explicitly include the identifier of the software they describe.
F4. Metadata are FAIR, searchable and indexable.

A: Software, and its metadata, is retrievable via standardized protocols.

A1. Software is retrievable by its 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 software is no longer available.

I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
I2. Software includes qualified references to other objects.

R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

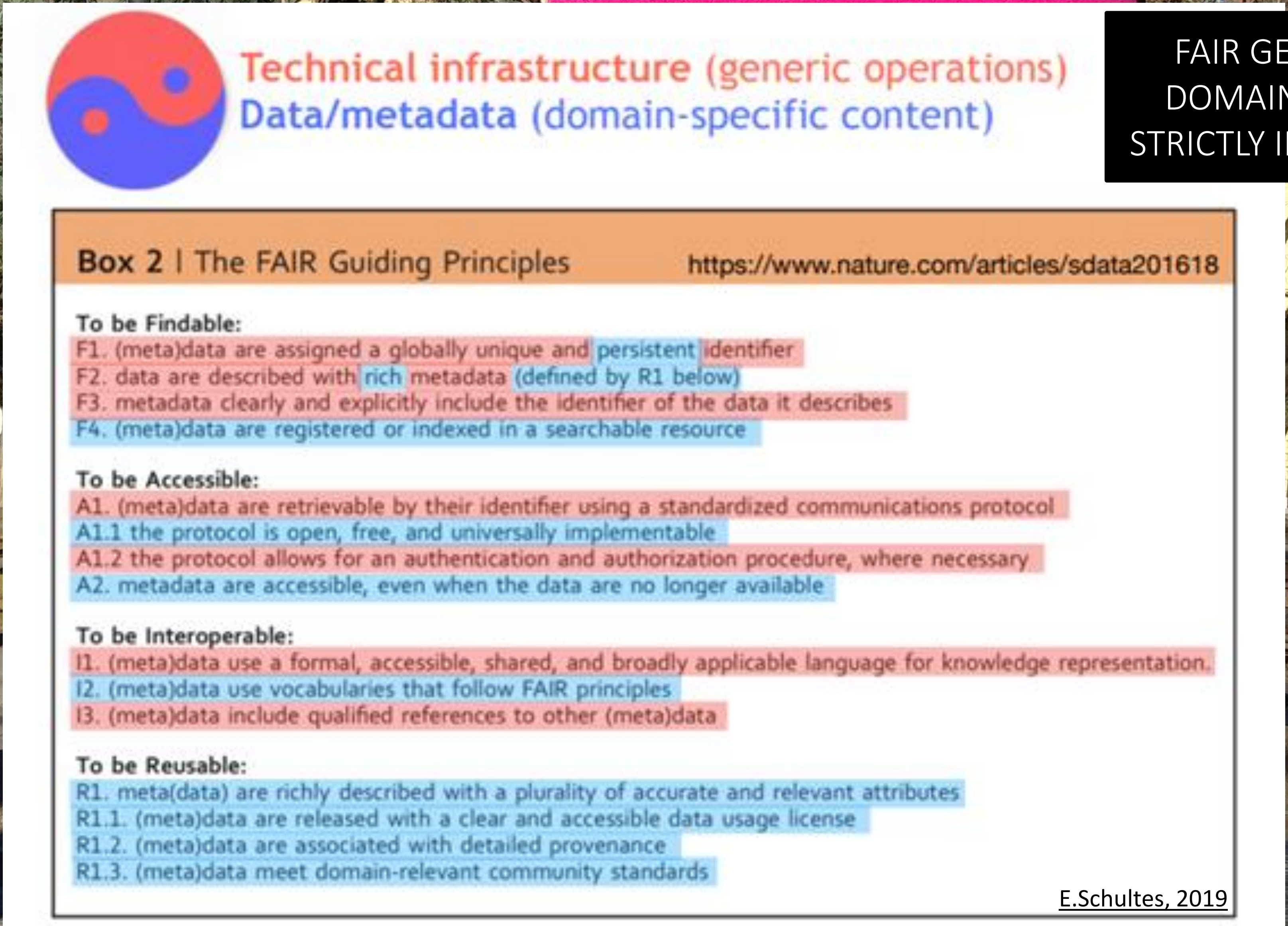
R1. Software is described with a plurality of accurate and relevant attributes.
- R1.1. Software is given a clear and accessible license.
- R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.
R3. Software meets domain-relevant community standards.

Table 1: The FAIR Principles for Research Software

FAIR RESEARCH SOFTWARE

FAIR: technology VS domain



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

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

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

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

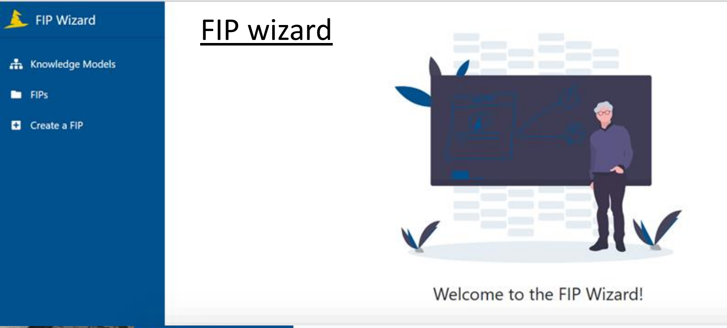
To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

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

E.Schultes, 2019

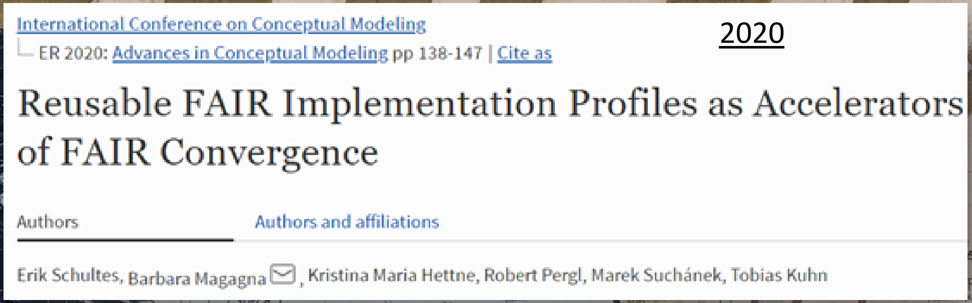
FAIR GENERIC VS
DOMAIN SPECIFIC
STRICTLY INTERLINKED

FAIR Implementation profiles



FIP wizard

Welcome to the FIP Wizard!



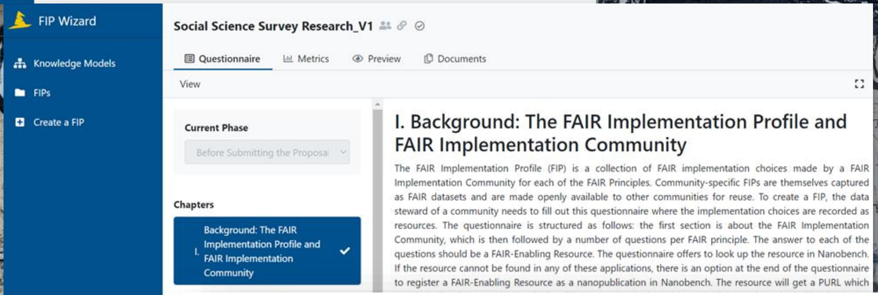
International Conference on Conceptual Modeling

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

2020

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence

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



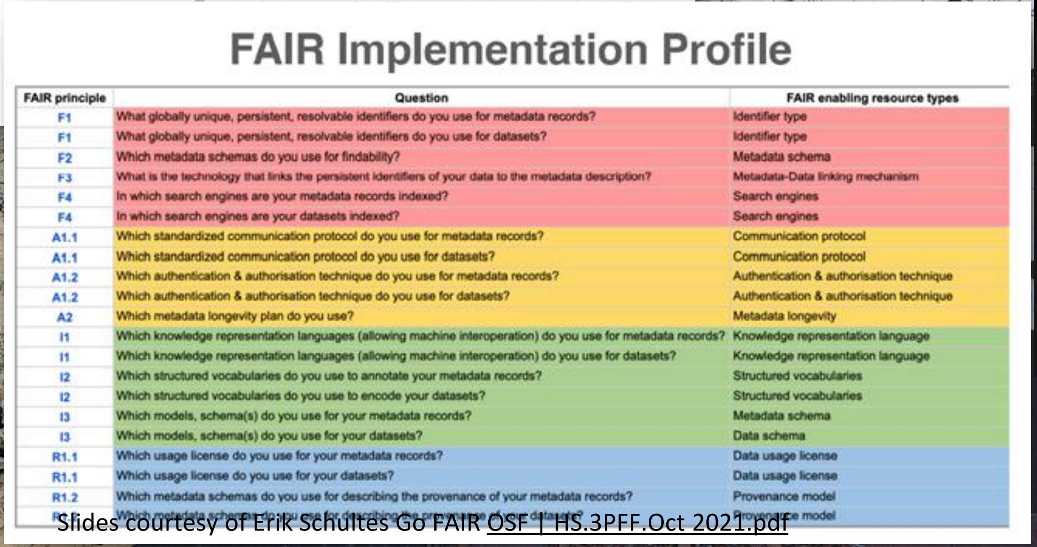
Current Phase: Before Submitting the Proposal

Chapters

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

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

The FAIR Implementation Profile (FIP) is a collection of FAIR implementation choices made by a FAIR Implementation Community for each of the FAIR Principles. Community-specific FIPs are themselves captured as FAIR datasets and are made openly available to other communities for reuse. To create a FIP, the data steward of a community needs to fill out this questionnaire where the implementation choices are recorded as resources. The questionnaire is structured as follows: the first section is about the FAIR Implementation Community, which is then followed by a number of questions per FAIR principle. The answer to each of the questions should be a FAIR-Enabling Resource. The questionnaire offers to look up the resource in Nanobench. If the resource cannot be found in any of these applications, there is an option at the end of the questionnaire to register a FAIR-Enabling Resource as a nanopublication in Nanobench. The resource will get a PURL which



FAIR Implementation Profile

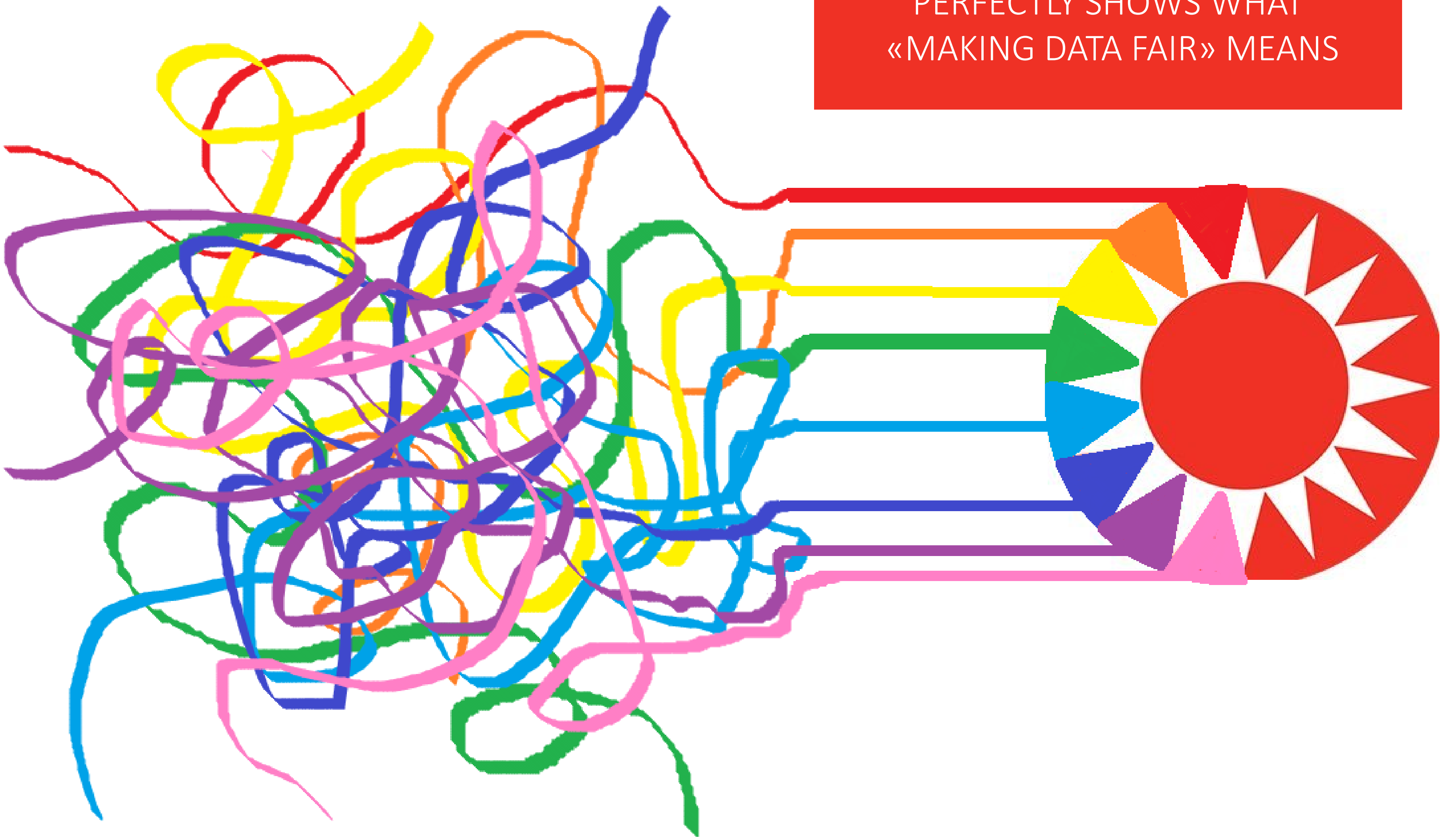
FAIR principle	Question	FAIR enabling resource types
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type
F2	Which metadata schemas do you use for 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

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

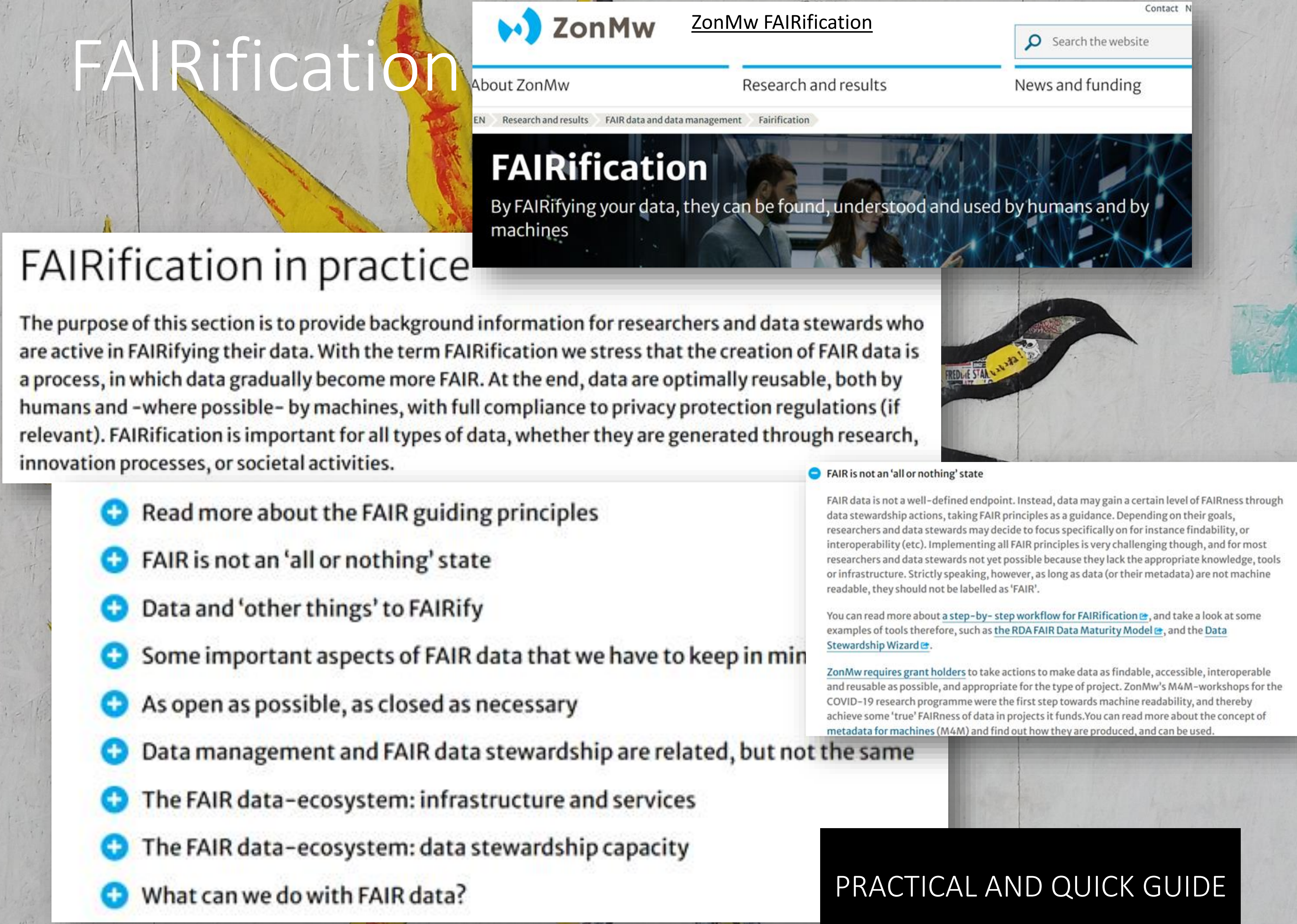
CREATE FAIR IMPLEMENTATION PROFILES REUSBALE BY YOUR COMMUNITY - KEYWORD: CONVERGENCE

How-to

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



FAIRification



ZonMw FAIRification

FAIRification

By FAIRifying your data, they can be found, understood and used by humans and by machines

FAIRification in practice

The purpose of this section is to provide background information for researchers and data stewards who are active in FAIRifying their data. With the term FAIRification we stress that the creation of FAIR data is a process, in which data gradually become more FAIR. At the end, data are optimally reusable, both by humans and -where possible- by machines, with full compliance to privacy protection regulations (if relevant). FAIRification is important for all types of data, whether they are generated through research, innovation processes, or societal activities.

- Read more about the FAIR guiding principles
- FAIR is not an 'all or nothing' state
- Data and 'other things' to FAIRify
- Some important aspects of FAIR data that we have to keep in min
- As open as possible, as closed as necessary
- Data management and FAIR data stewardship are related, but not the same
- The FAIR data-ecosystem: infrastructure and services
- The FAIR data-ecosystem: data stewardship capacity
- What can we do with FAIR data?

FAIR is not an 'all or nothing' state

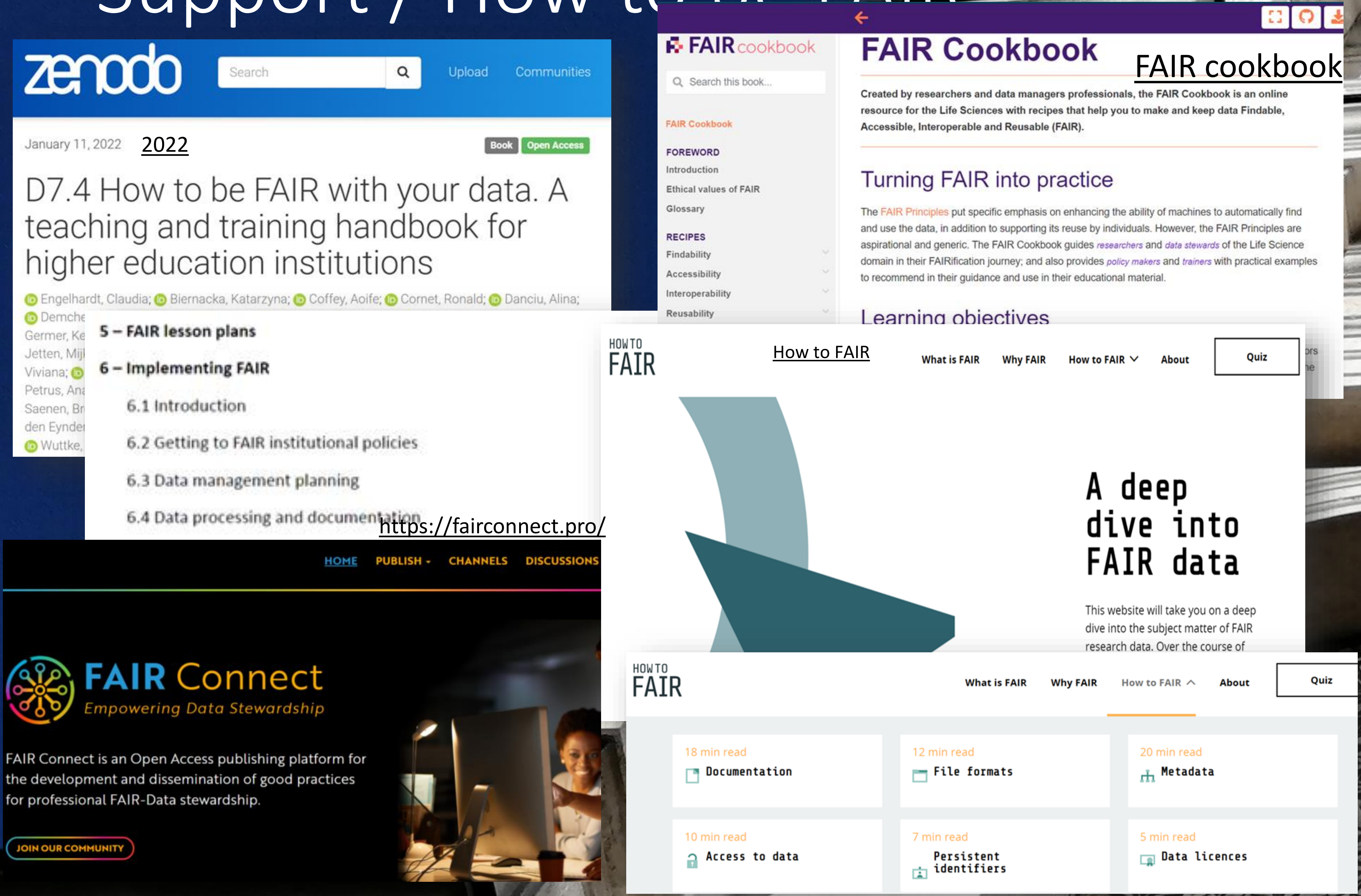
FAIR data is not a well-defined endpoint. Instead, data may gain a certain level of FAIRness through data stewardship actions, taking FAIR principles as a guidance. Depending on their goals, researchers and data stewards may decide to focus specifically on for instance findability, or interoperability (etc). Implementing all FAIR principles is very challenging though, and for most researchers and data stewards not yet possible because they lack the appropriate knowledge, tools or infrastructure. Strictly speaking, however, as long as data (or their metadata) are not machine readable, they should not be labelled as 'FAIR'.

You can read more about a step-by- step workflow for FAIRification, and take a look at some examples of tools therefore, such as the RDA FAIR Data Maturity Model, and the Data Stewardship Wizard.

ZonMw requires grant holders to take actions to make data as findable, accessible, interoperable and reusable as possible, and appropriate for the type of project. ZonMw's M4M-workshops for the COVID-19 research programme were the first step towards machine readability, and thereby achieve some 'true' FAIRness of data in projects it funds.You can read more about the concept of metadata for machines (M4M) and find out how they are produced, and can be used.

PRACTICAL AND QUICK GUIDE

Support / How to be FAIR



zenodo

Search | Upload | Communities

January 11, 2022 — 2022

Book | Open Access

D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions

Engelhardt, Claudia; Biernacka, Katarzyna; Coffey, Aoife; Cornet, Ronald; Danciu, Alina; Demche… Germer, Ke… Jetten, Mij… Viviana; … Petrus, Ana… Saenen, Br… den Eynde… Wuttke, …

5 – FAIR lesson plans

6 – Implementing FAIR

6.1 Introduction

6.2 Getting to FAIR institutional policies

6.3 Data management planning

6.4 Data processing and documentation

https://fairconnect.pro/

HOME | PUBLISH | CHANNELS | DISCUSSIONS

FAIR Connect

Empowering Data Stewardship

FAIR Connect is an Open Access publishing platform for the development and dissemination of good practices for professional FAIR-Data stewardship.

JOIN OUR COMMUNITY

FAIR cookbook

Search this book...

FAIR Cookbook

FOREWORD

Introduction

Ethical values of FAIR

Glossary

RECIPES

Findability

Accessibility

Interoperability

Reusability

FAIR Cookbook

Created by researchers and data managers professionals, the FAIR Cookbook is an online resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable (FAIR).

Turning FAIR into practice

The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. However, the FAIR Principles are aspirational and generic. The FAIR Cookbook guides *researchers* and *data stewards* of the Life Science domain in their FAIRification journey; and also provides *policy makers* and *trainers* with practical examples to recommend in their guidance and use in their educational material.

Learning objectives

HOW TO FAIR — How to FAIR

What is FAIR | Why FAIR | How to FAIR | About | Quiz

A deep dive into FAIR data

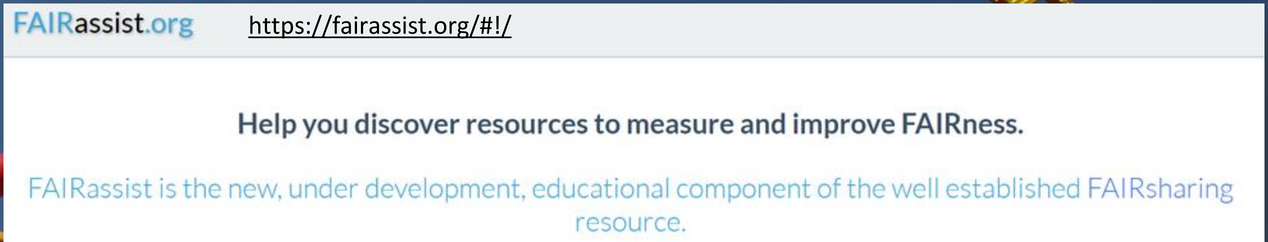
This website will take you on a deep dive into the subject matter of FAIR research data. Over the course of

HOW TO FAIR

What is FAIR | Why FAIR | How to FAIR | About | Quiz

18 min read Documentation	12 min read File formats	20 min read Metadata
10 min read Access to data	7 min read Persistent identifiers	5 min read Data licences

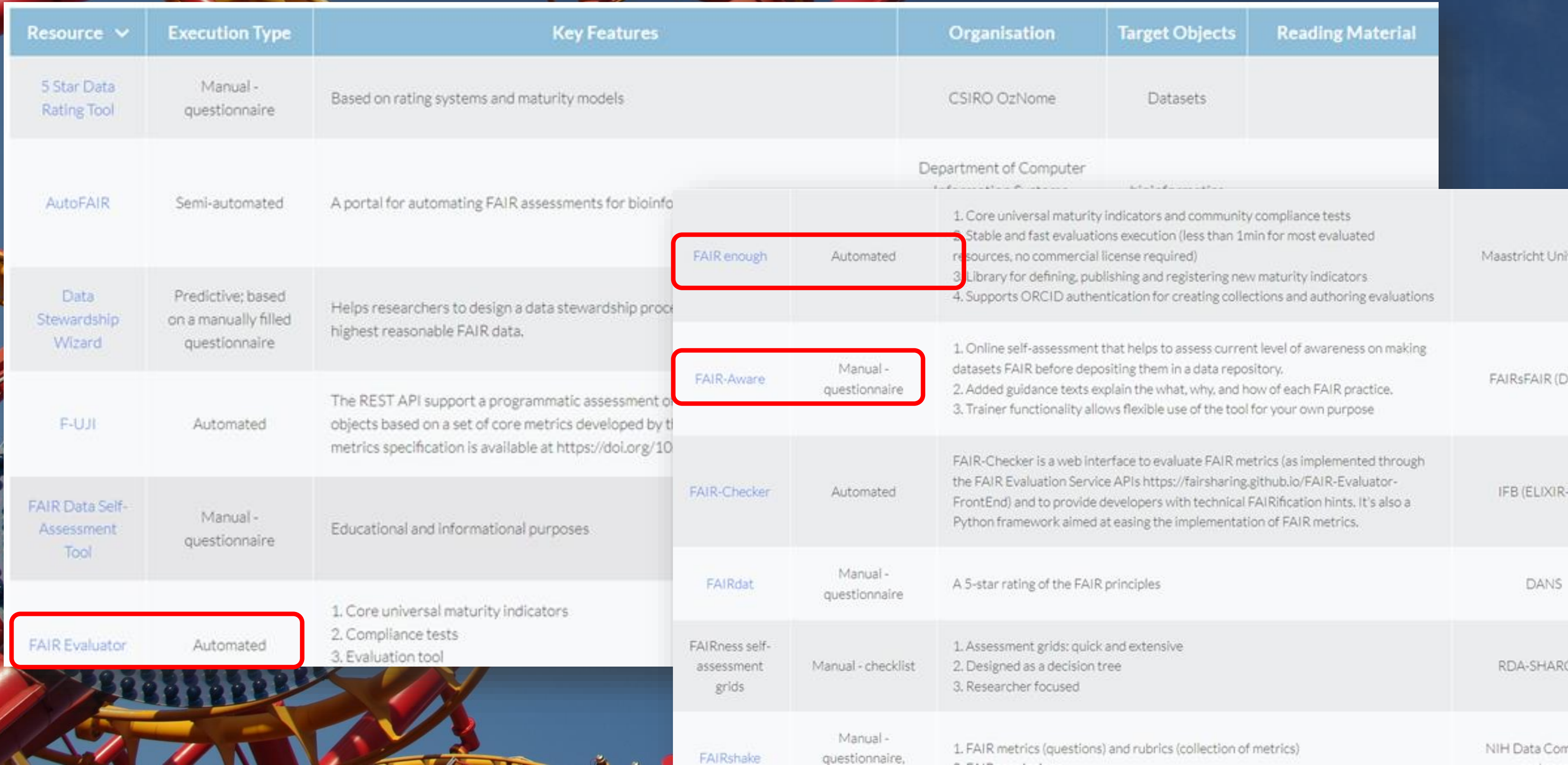
To check your FAIRness



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.

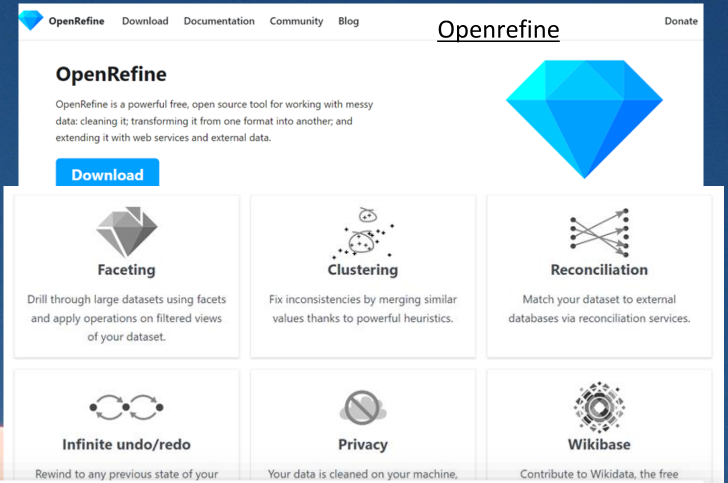


Resource	Execution Type	Key Features	Organisation	Target Objects	Reading Material
5 Star Data Rating Tool	Manual - questionnaire	Based on rating systems and maturity models	CSIRO OzNome	Datasets	
AutoFAIR	Semi-automated	A portal for automating FAIR assessments for bioinfo	Department of Computer		
Data 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-SHAR
FAIRshake	Manual - questionnaire,	1. FAIR metrics (questions) and rubrics (collection of metrics)	NIH Data Con

F=Findable – Metadata

Openrefine



OpenRefine Download Documentation Community Blog Donate

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,

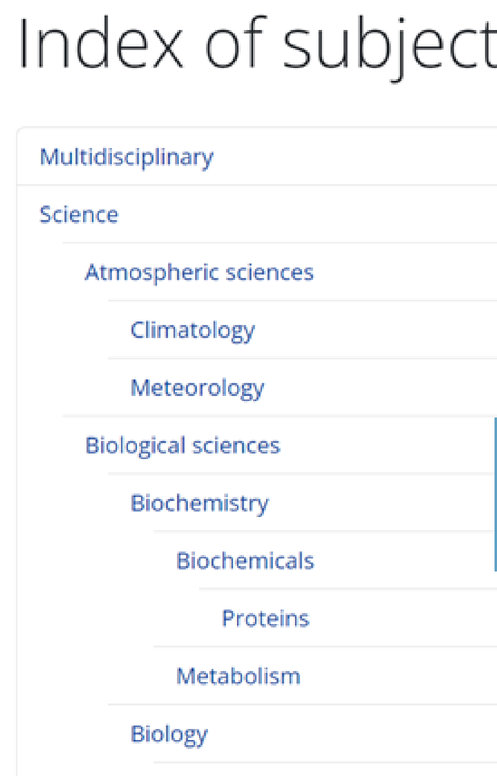
Wikibase
Contribute to Wikidata, the free

Metadata standars catalog

Metadata Standards Catalog Search Sign in

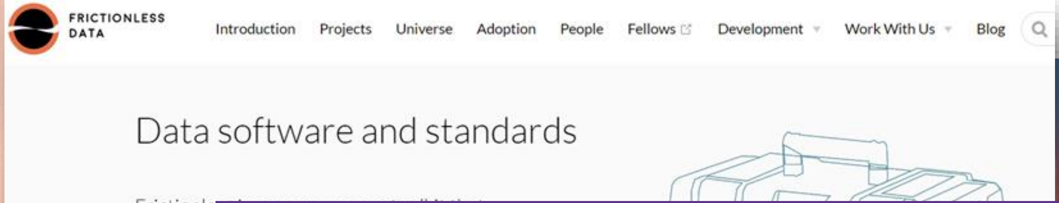
Metadata Standards Catalog

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



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

Data software and standards

Frictionle
brings sim
whether y
engineeri

Why Fric

FAIRcookbook

FAIRCOOKBOOK GITH

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

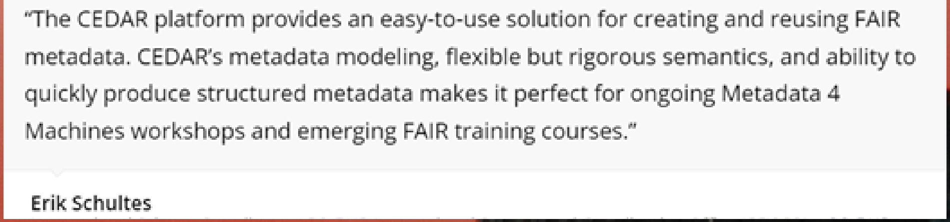
Recipe Overview
Reading Time: 20 minutes
Executable Code: No
Difficulty

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

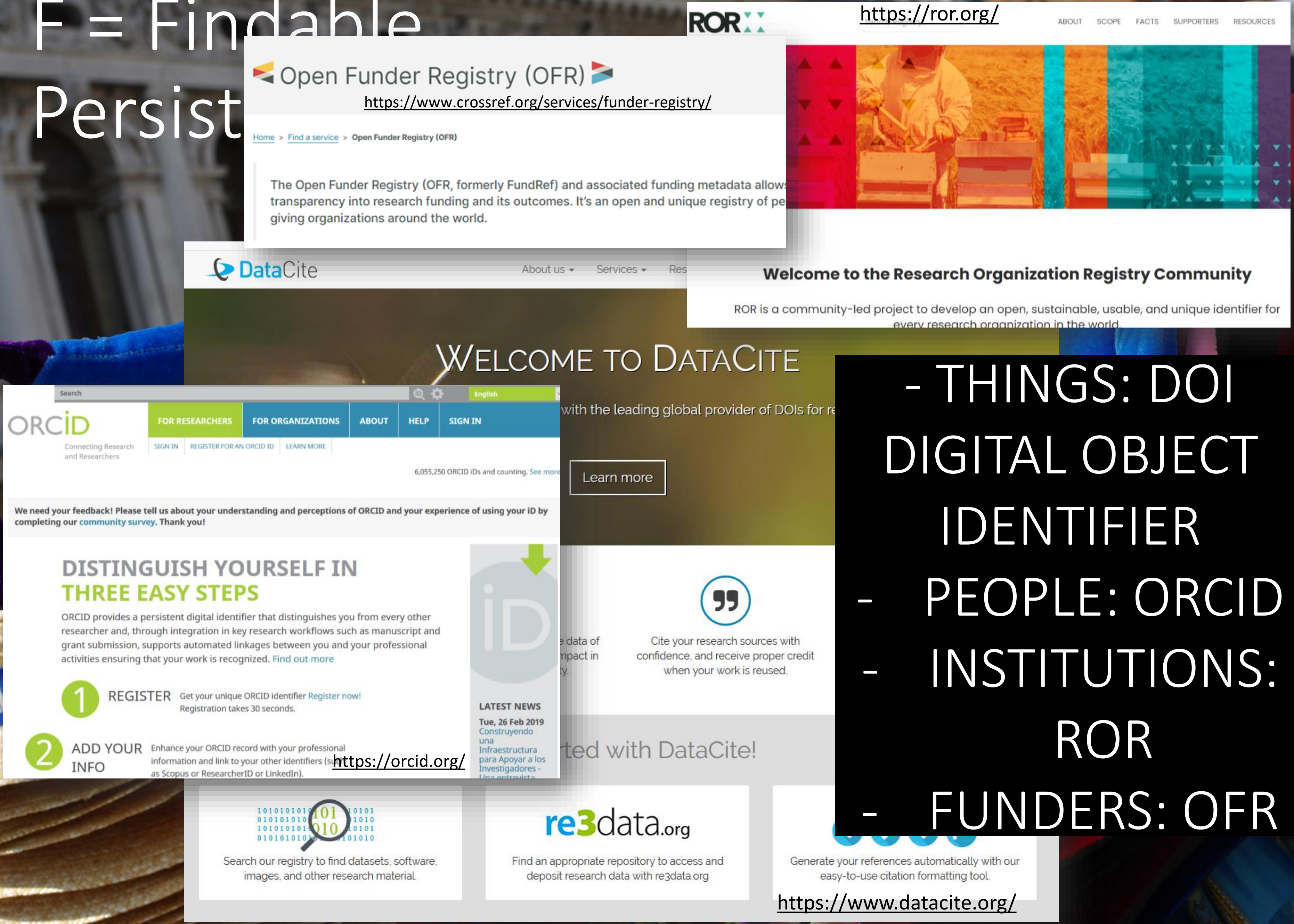
Better metadata means better science



"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



https://ror.org/

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.

Open Funder Registry (OFR)

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

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.

WELCOME TO DATACITE

with the leading global provider of DOIs for re

data of pact in y.

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.

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/

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 (su as Scopus or ResearcherID or LinkedIn).

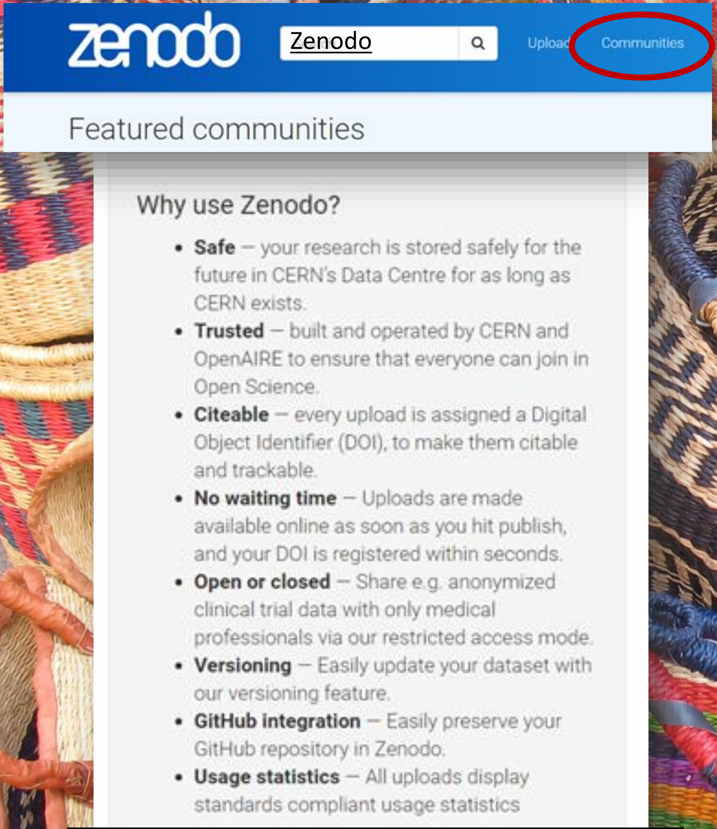
LATEST NEWS

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

https://orcid.org/

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

A = Accessible – Data s



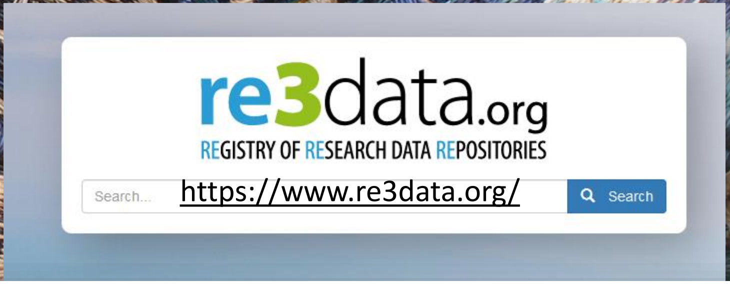
Featured communities

Why use Zenodo?

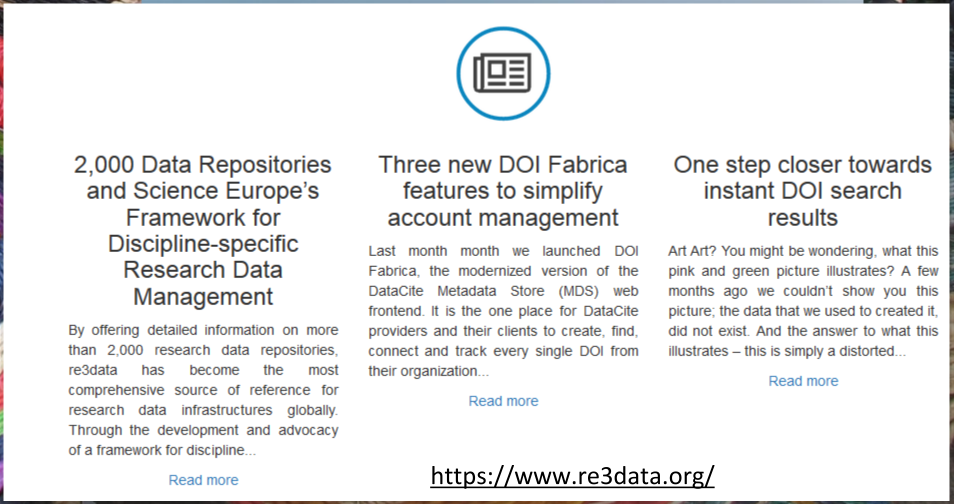
- **Safe** – your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** – built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** – every upload is assigned a Digital Object Identifier (DOI), to make them 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?]

Humanities
Humanities and Social Sciences
Social and Behavioural
Engineering Sciences
Natural Sciences
Life Sciences
Biology
Medicine
Chemistry
Physics
Geosciences
Agriculture



https://www.re3data.org/



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

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

Read more

Three new DOI Fabrica features to simplify account management

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

Read more

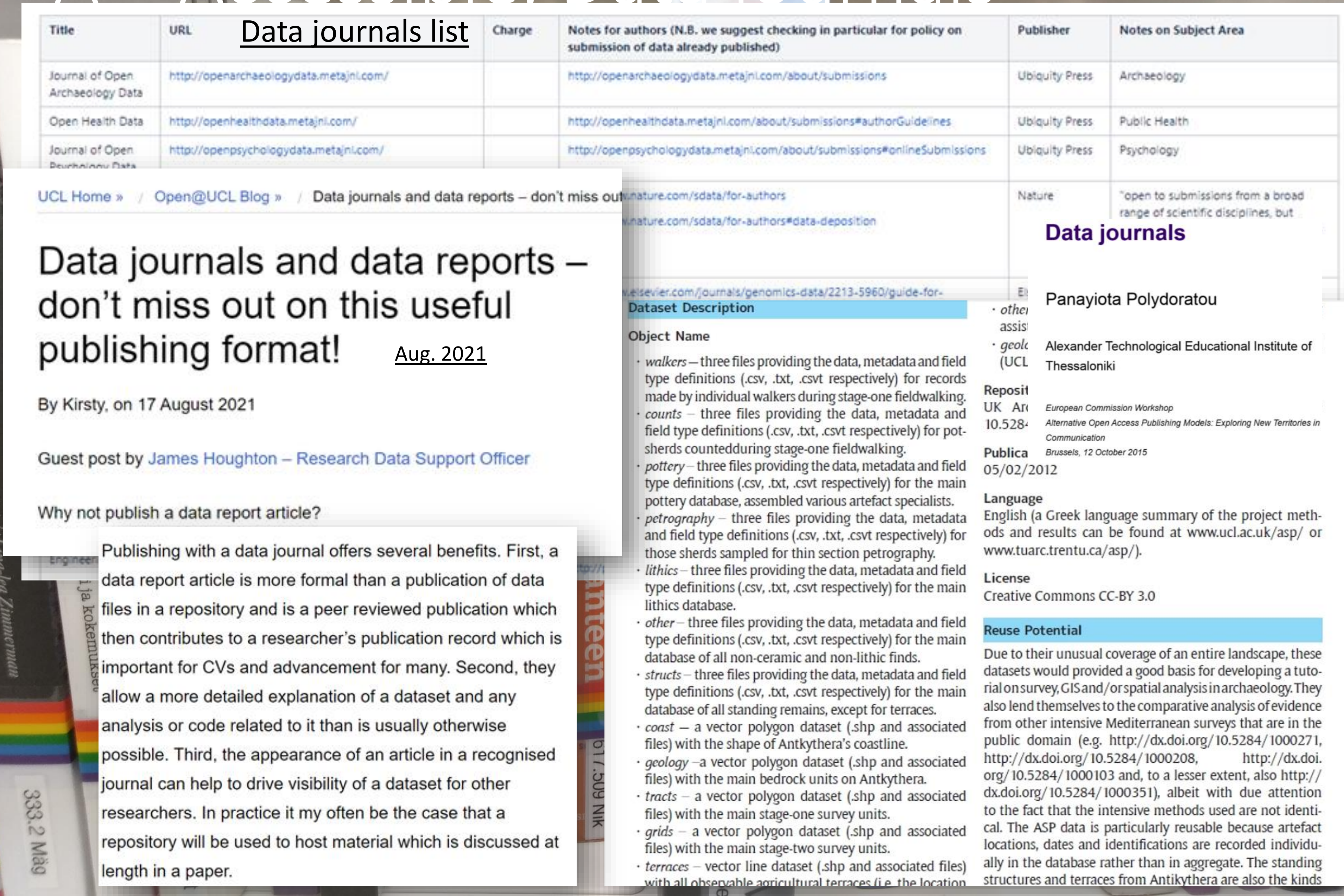
One step closer towards instant DOI search results

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

Read more

https://www.re3data.org/

A = Accessible. Data journals



Data journals list

Title	URL	Charge	Notes for authors (N.B. we suggest checking in particular for policy on submission of data already published)	Publisher	Notes on Subject Area
Journal of Open Archaeology Data	http://openarchaeologydata.metajnl.com/		http://openarchaeologydata.metajnl.com/about/submissions	Ubiquity Press	Archaeology
Open Health Data	http://openhealthdata.metajnl.com/		http://openhealthdata.metajnl.com/about/submissions#authorGuidelines	Ubiquity Press	Public Health
Journal of Open Psychology Data	http://openpsychologydata.metajnl.com/		http://openpsychologydata.metajnl.com/about/submissions#onlineSubmissions	Ubiquity Press	Psychology
			.nature.com/sdata/for-authors .nature.com/sdata/for-authors#data-deposition	Nature	"open to submissions from a broad range of scientific disciplines, but
			.elsevier.com/journals/genomics-data/2213-5960/guide-for-	E	
Engineer					

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

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

Aug. 2021

By Kirsty, on 17 August 2021

Guest post by James Houghton – Research Data Support Officer

Why not publish a data report article?

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

Dataset Description

Object Name

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

- *other*
assis
- *geolo*
(UCL

Reposit
UK Ar
10.528

Publica
05/02/2012

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

License
Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. http://dx.doi.org/10.5284/1000271, http://dx.doi.org/10.5284/1000208, http://dx.doi.org/10.5284/1000103 and, to a lesser extent, also http://dx.doi.org/10.5284/1000351), albeit with due attention to the fact that the intensive methods used are not identical. The ASP data is particularly reusable because artefact locations, dates and identifications are recorded individually in the database rather than in aggregate. The standing structures and terraces from Antikythera are also the kinds

Data journals

Panayiota Polydoratou

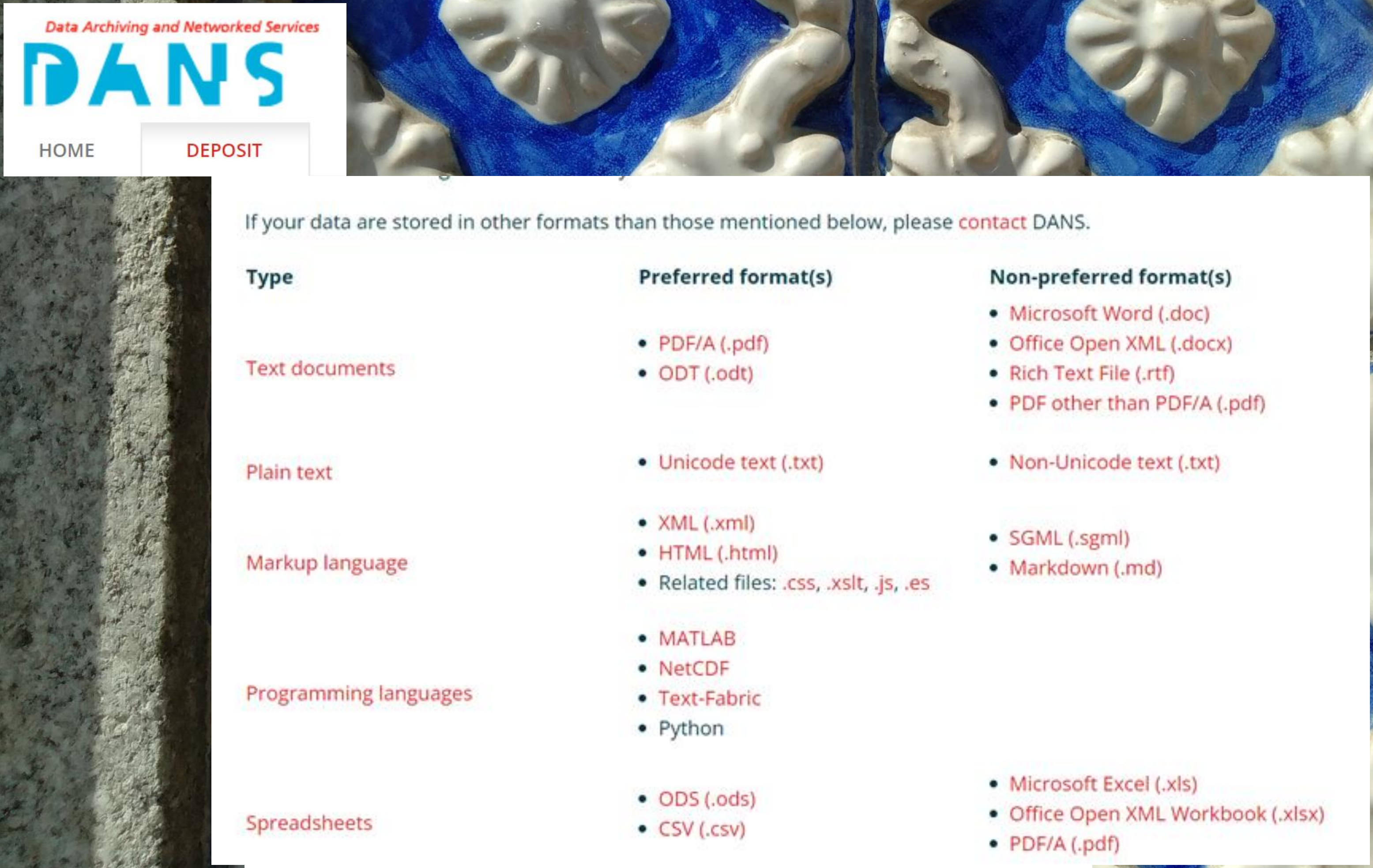
Alexander Technological Educational Institute of Thessaloniki

European Commission Workshop

Alternative Open Access Publishing Models: Exploring New Territories in Communication

Brussels, 12 October 2015

A = Accessible: Formats

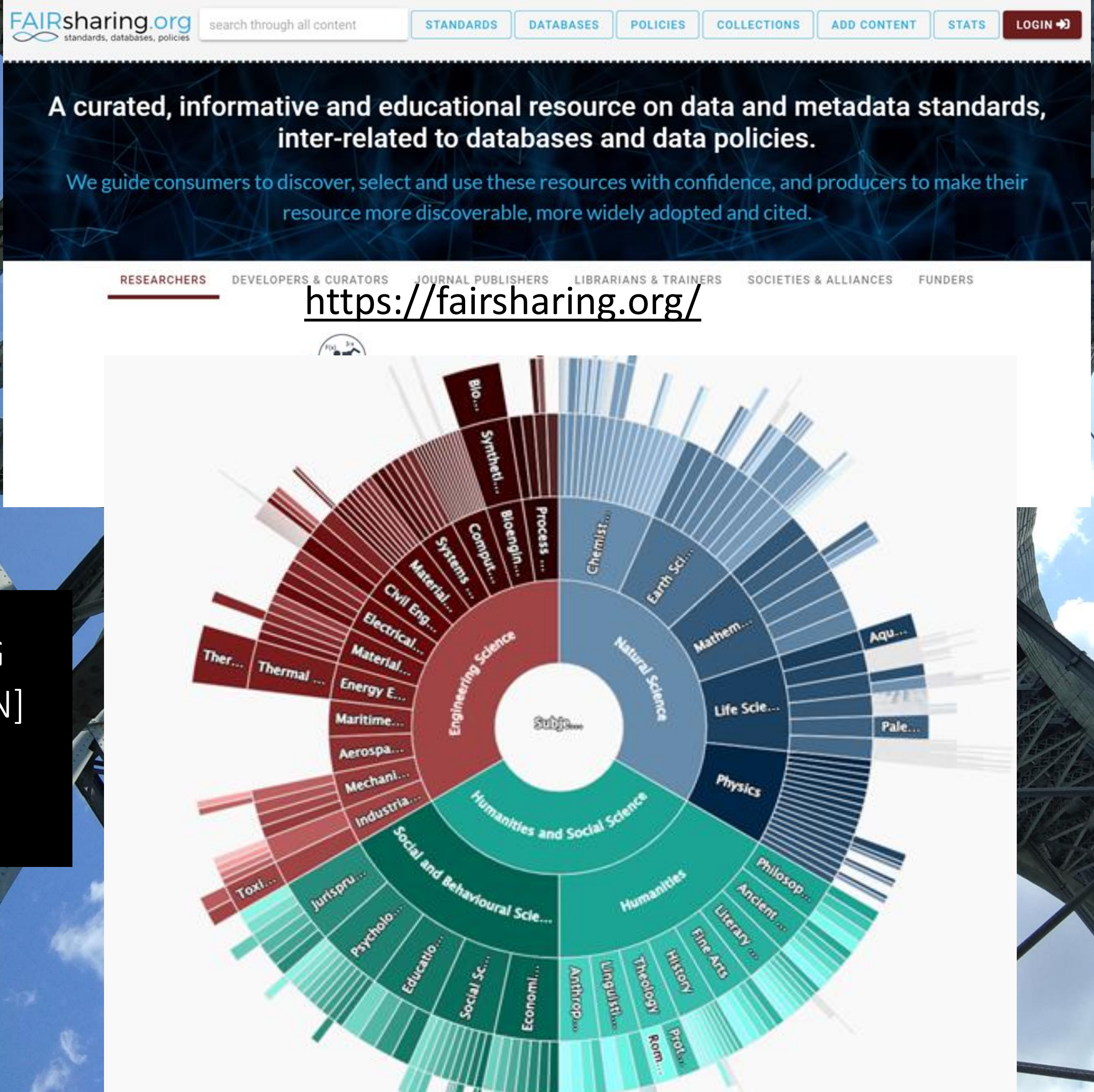


If your data are stored in other formats than those mentioned below, please contact DANS.

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



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

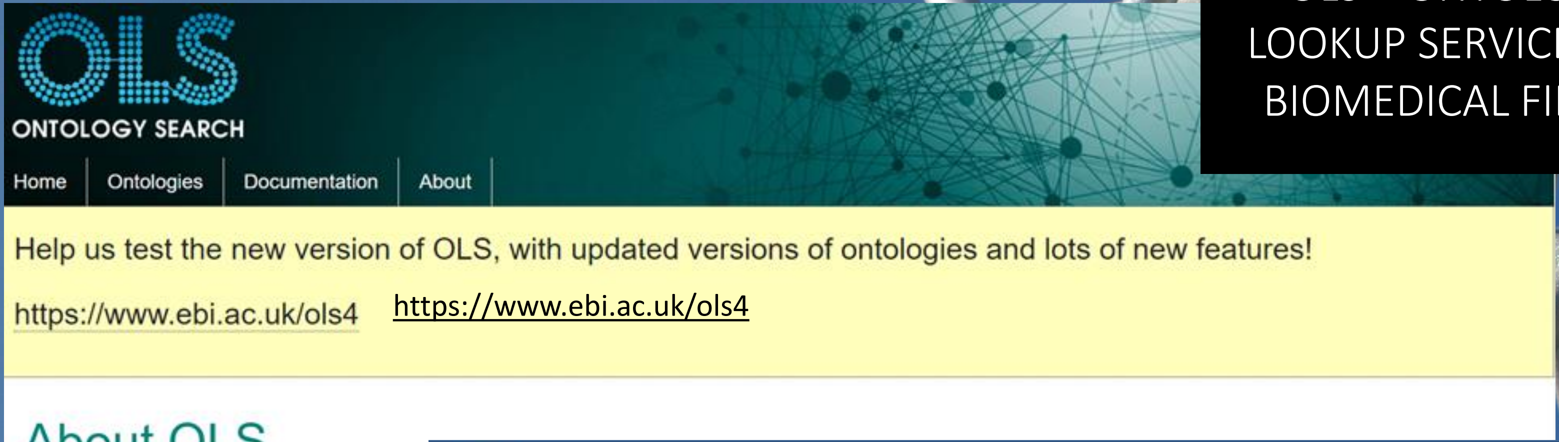
We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

https://fairsharing.org/

FAIRSHARING
[NEW VERSION]
STANDARD
REGISTRY

I = Inteoperable. Ontologies

OLS – ONTOLOGY LOOKUP SERVICE FOR BIOMEDICAL FIELDS

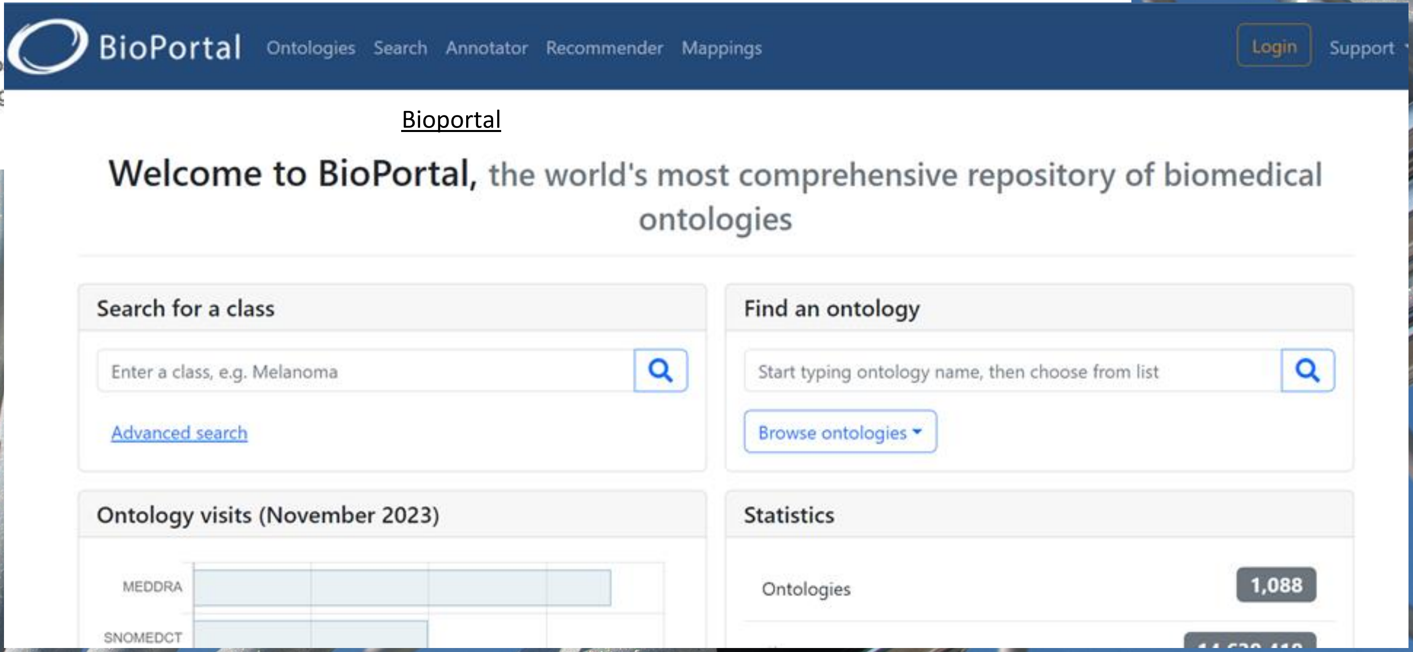


Help us test the new version of OLS, with updated versions of ontologies and lots of new features!

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

About OLS

The Ontology Lookup Service (OLS) is a repo
ontologies through the website as well as prog
EBI.



[Bioportal]()

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

Search for a class

Enter a class, e.g. Melanoma

Advanced search

Find an ontology

Start typing ontology name, then choose from list

Browse ontologies

Ontology visits (November 2023)

Statistics

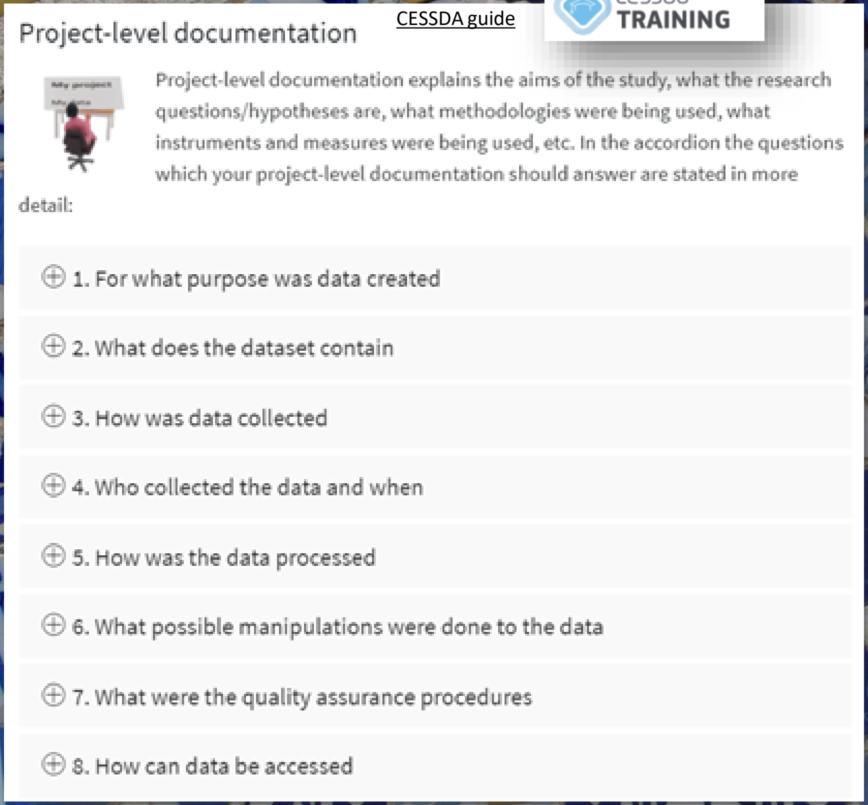
Ontologies 1,088

R = Reusable. Documentation

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

CESSDA guide

cessda TRAINING



Project-level documentation

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

⊕ 1. For what purpose was data created

⊕ 2. What does the dataset contain

⊕ 3. How was data collected

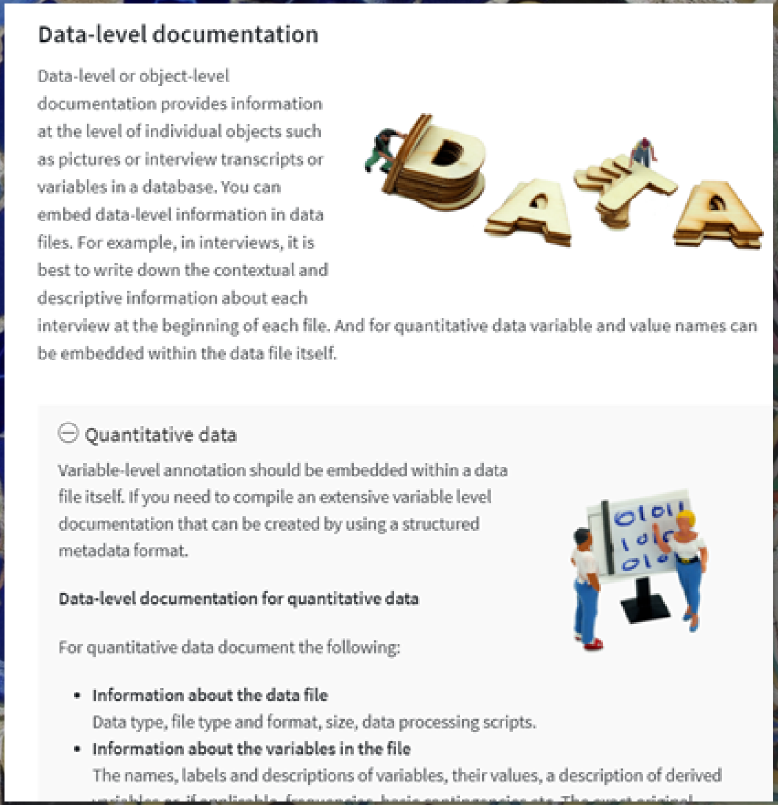
⊕ 4. Who collected the data and when

⊕ 5. How was the data processed

⊕ 6. What possible manipulations were done to the data

⊕ 7. What were the quality assurance procedures

⊕ 8. How can data be accessed



Data-level documentation

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

⊖ Quantitative data

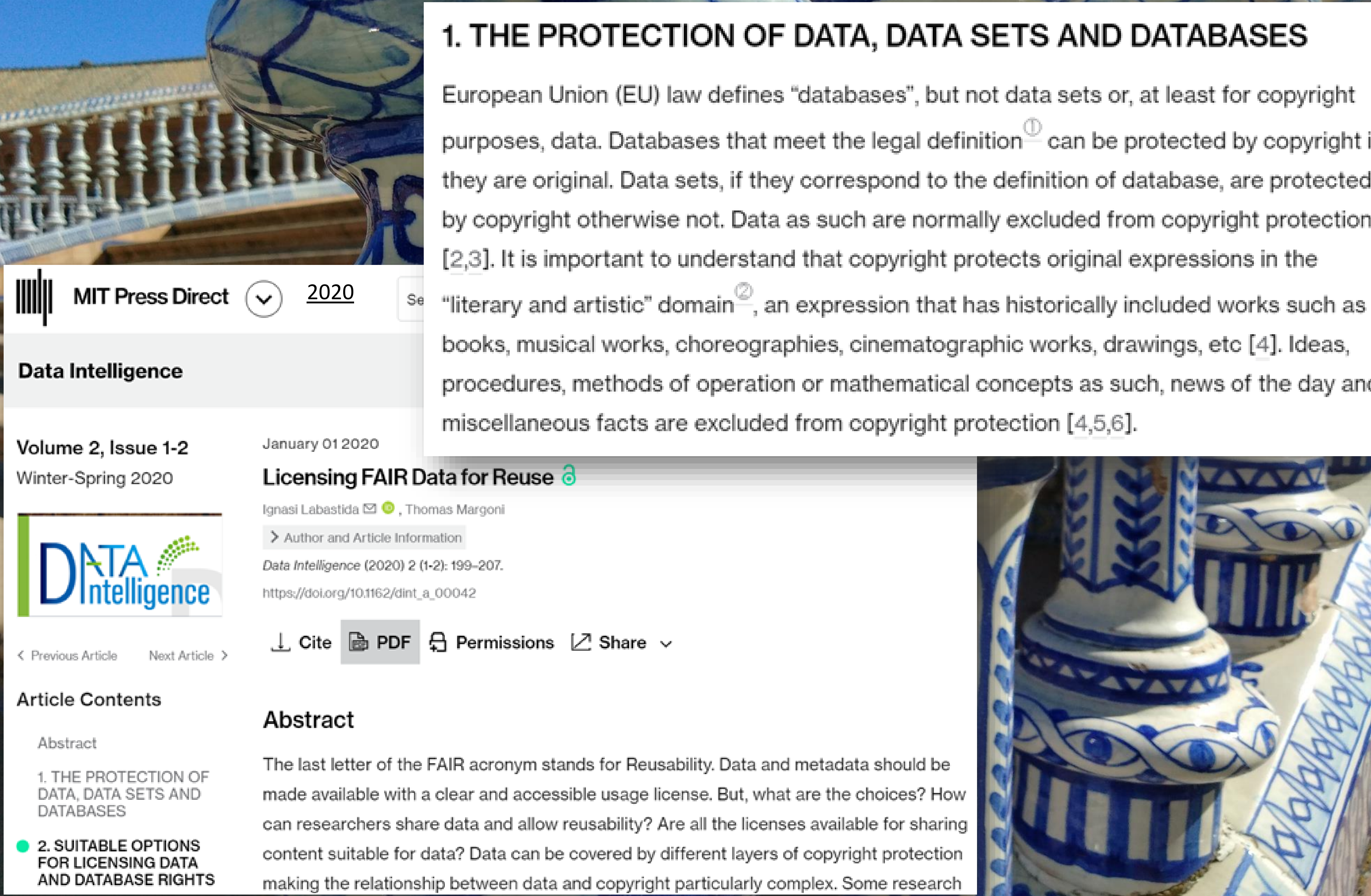
Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.

Data-level documentation for quantitative data

For quantitative data document the following:

- **Information about the data file**
  Data type, file type and format, size, data processing scripts.
- **Information about the variables in the file**
  The names, labels and descriptions of variables, their values, a description of derived variables or, if applicable, frequencies, basic contingencies etc. The exact original

R = Reusable – Licenses



1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

European Union (EU) law defines "databases", but not data sets or, at least for copyright purposes, data. Databases that meet the legal definition[1] can be protected by copyright i they are original. Data sets, if they correspond to the definition of database, are protected by copyright otherwise not. Data as such are normally excluded from copyright protection [2,3]. It is important to understand that copyright protects original expressions in the "literary and artistic" domain[2], an expression that has historically included works such as books, musical works, choreographies, cinematographic works, drawings, etc [4]. Ideas, procedures, methods of operation or mathematical concepts as such, news of the day an miscellaneous facts are excluded from copyright protection [4,5,6].

Data Intelligence

Volume 2, Issue 1-2
Winter-Spring 2020

Article Contents

January 01 2020

Licensing FAIR Data for Reuse

Ignasi Labastida, Thomas Margoni

› Author and Article Information

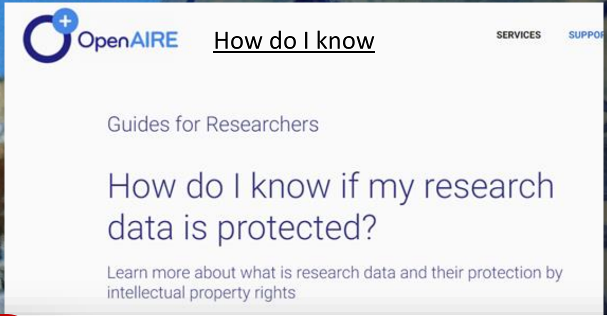
Data Intelligence (2020) 2 (1-2): 199–207.

https://doi.org/10.1162/dint_a_00042

Abstract

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

R = Reusable – Legal aspe

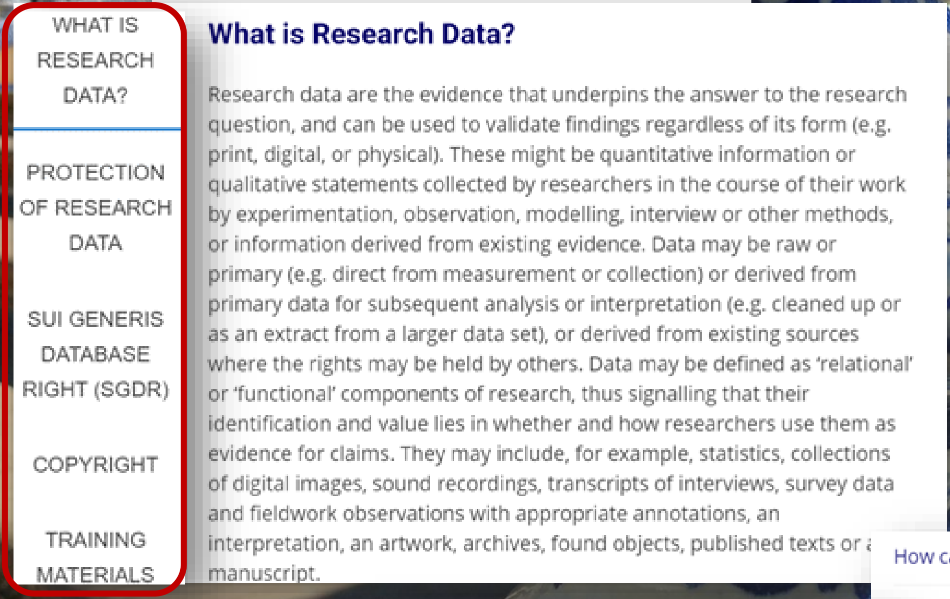


How do I know

Guides for Researchers

How do I know if my research data is protected?

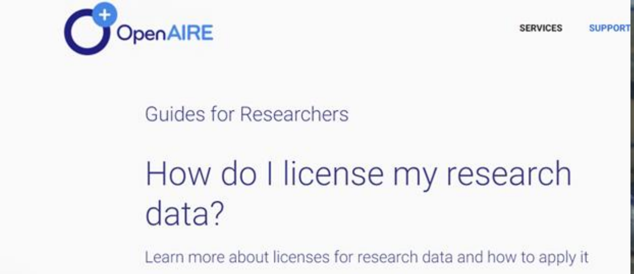
Learn more about what is research data and their protection by intellectual property rights



WHAT IS RESEARCH DATA?
PROTECTION OF RESEARCH DATA
SUI GENERIS DATABASE RIGHT (SGDR)
COPYRIGHT
TRAINING MATERIALS

What is Research Data?

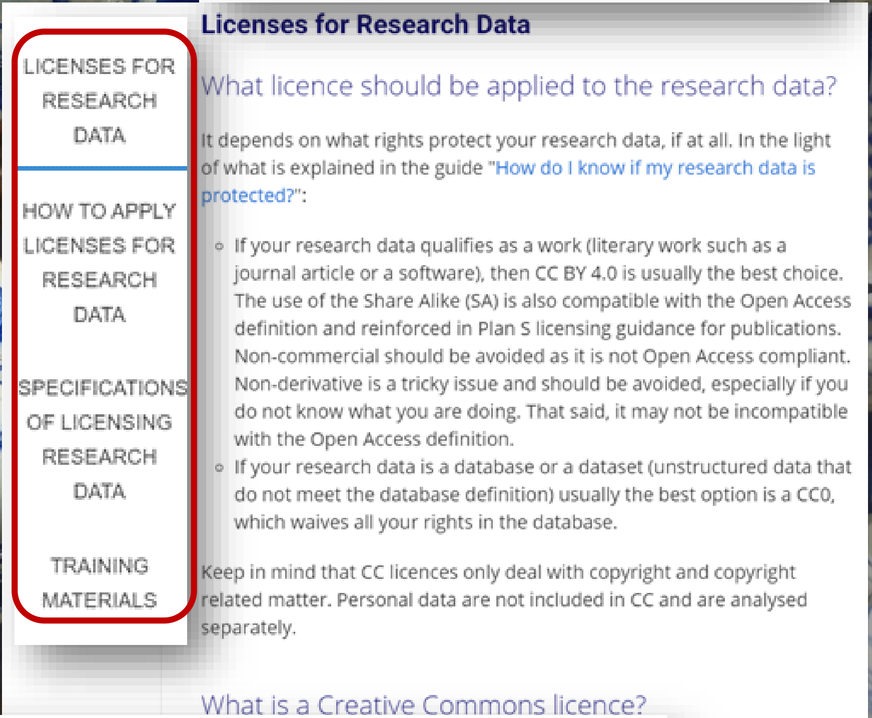
Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by 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.



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

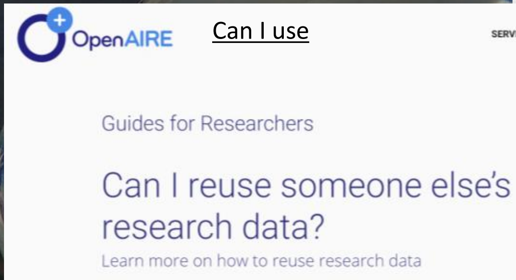
What licence should be applied to the research data?

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?

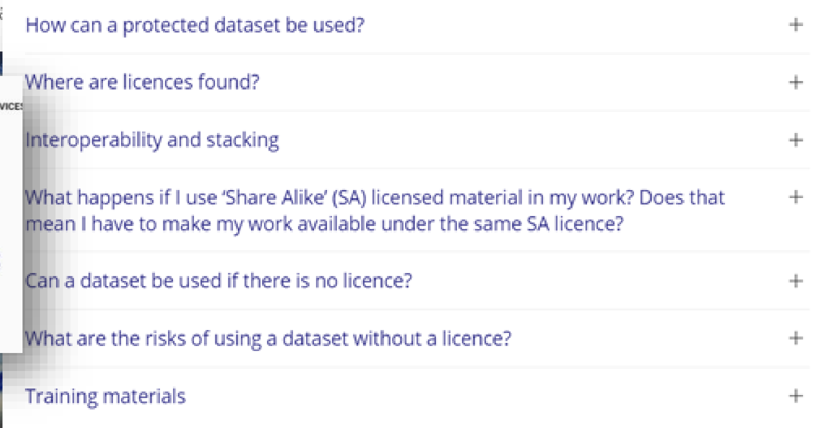


Can I use

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 +

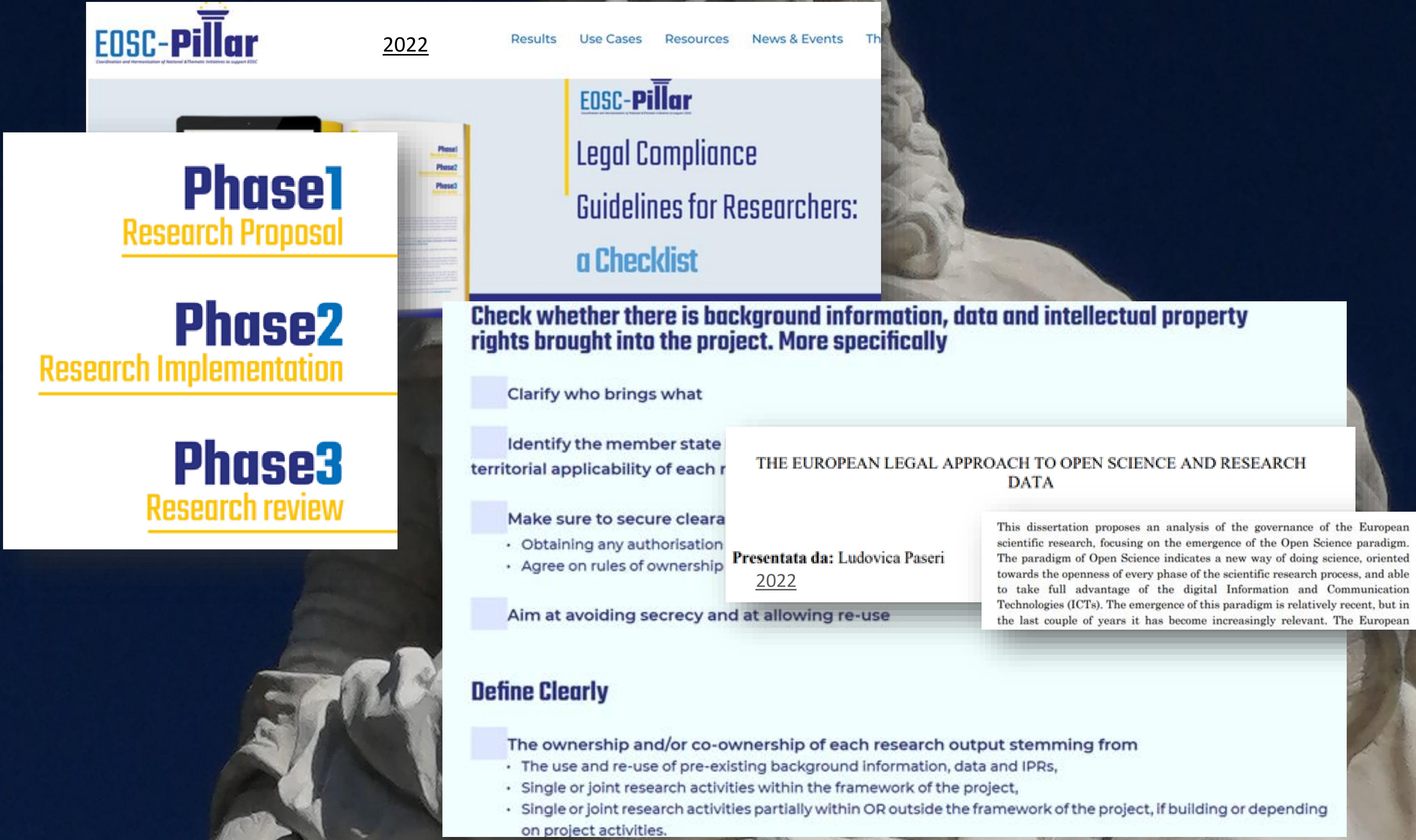
What happens if I use 'Share Alike' (SA) licensed material in my work? Does that mean I have to make my work available under the same SA licence? +

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

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

Training materials +

R = Reusable – Legal aspects



EOSC-Pillar

2022

EOSC-Pillar

Legal Compliance Guidelines for Researchers: a Checklist

Phase1 Research Proposal

Phase2 Research Implementation

Phase3 Research review

Check whether there is background information, data and intellectual property rights brought into the project. More specifically

□ Clarify who brings what

□ Identify the member state

territorial applicability of each r

□ Make sure to secure cleara

- Obtaining any authorisation
- Agree on rules of ownership

□ Aim at avoiding secrecy and at allowing re-use

Define Clearly

□ The ownership and/or co-ownership of each research output stemming from

- The use and re-use of pre-existing background information, data and IPRs,
- Single or joint research activities within the framework of the project,
- Single or joint research activities partially within OR outside the framework of the project, if building or depending on project activities.

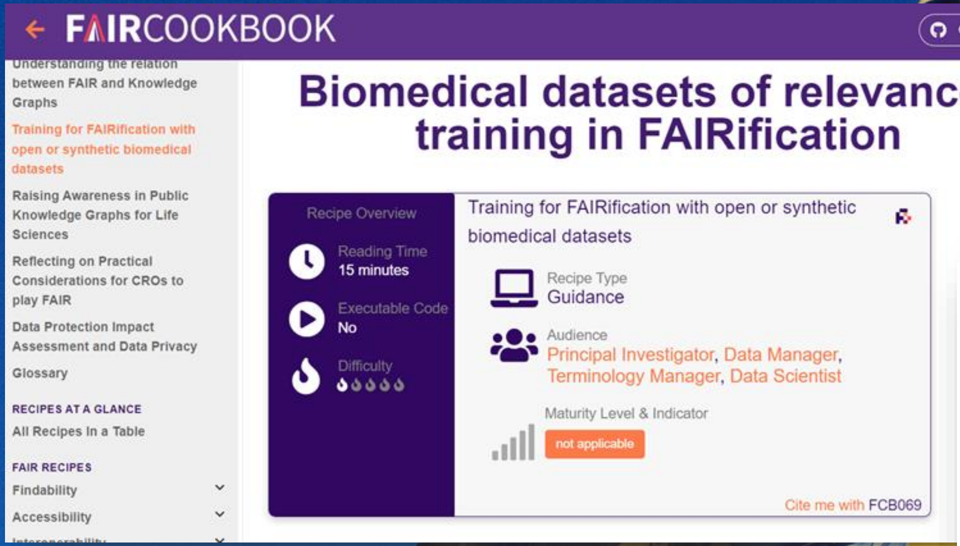
THE EUROPEAN LEGAL APPROACH TO OPEN SCIENCE AND RESEARCH DATA

Presentata da: Ludovica Paseri

2022

This dissertation proposes an analysis of the governance of the European scientific research, focusing on the emergence of the Open Science paradigm. The paradigm of Open Science indicates a new way of doing science, oriented towards the openness of every phase of the scientific research process, and able to take full advantage of the digital Information and Communication Technologies (ICTs). The emergence of this paradigm is relatively recent, but in the last couple of years it has become increasingly relevant. The European

FAIR in health sciences



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 Considerations for CROs to play FAIR

Data Protection Impact Assessment and Data Privacy

Glossary

RECIPES AT A GLANCE

All Recipes in a Table

FAIR RECIPES

Findability

Accessibility

Interoperability

Biomedical datasets of relevance for training in FAIRification

Recipe Overview

Reading Time: 15 minutes

Executable Code: No

Difficulty

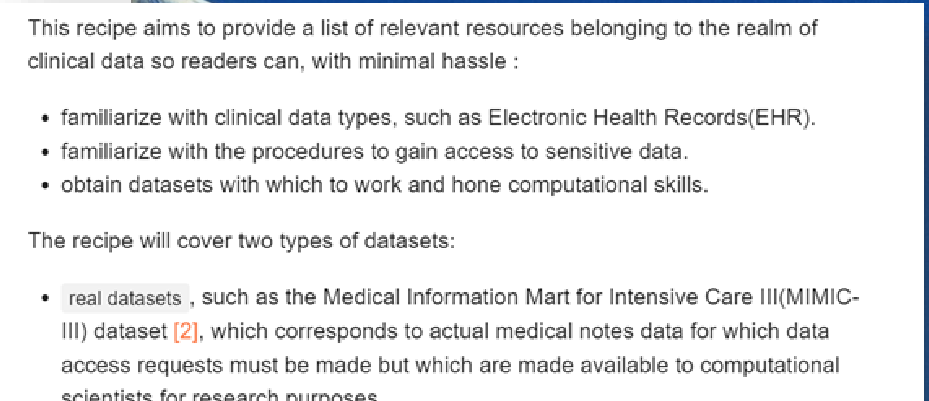
Training for FAIRification with open or synthetic biomedical datasets

Recipe Type: Guidance

Audience: Principal Investigator, Data Manager, Terminology Manager, Data Scientist

Maturity Level & Indicator: not applicable

Cite me with FCB069

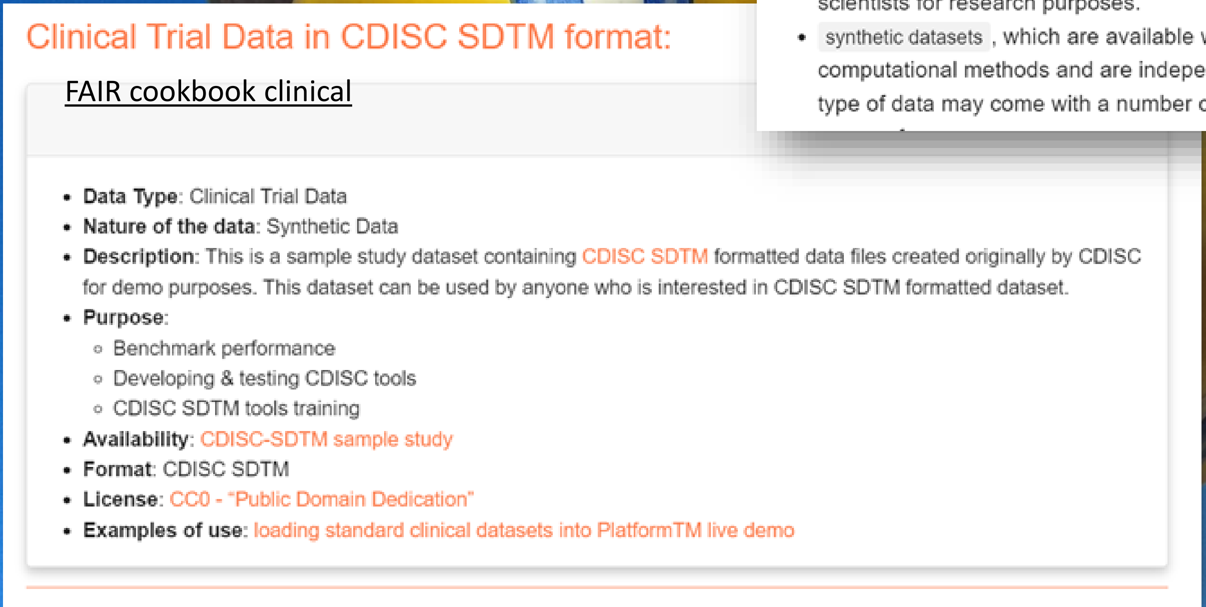


This recipe aims to provide a list of relevant resources belonging to the realm of clinical data so readers can, with minimal hassle :

- familiarize with clinical data types, such as Electronic Health Records(EHR).
- familiarize with the procedures to gain access to sensitive data.
- obtain datasets with which to work and hone computational skills.

The recipe will cover two types of datasets:

- real datasets , such as the Medical Information Mart for Intensive Care III(MIMIC-III) dataset [2], which corresponds to actual medical notes data for which data access requests must be made but which are made available to computational scientists for research purposes.
- synthetic datasets , which are available without restrictions since produced by computational methods and are independent of any real patient. While handy, this type of data may come with a number of limitations prospective users need to be

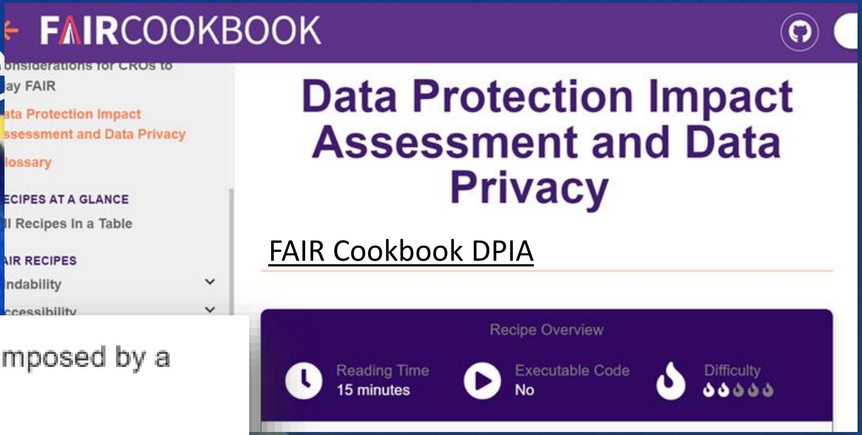


Clinical Trial Data in CDISC SDTM format:

FAIR cookbook clinical

- **Data Type**: Clinical Trial Data
- **Nature of the data**: Synthetic Data
- **Description**: This is a sample study dataset containing CDISC SDTM formatted data files created originally by CDISC for demo purposes. This dataset can be used by anyone who is interested in CDISC SDTM formatted dataset.
- **Purpose**:
  - Benchmark performance
  - Developing & testing CDISC tools
  - CDISC SDTM tools training
- **Availability**: CDISC-SDTM sample study
- **Format**: CDISC SDTM
- **License**: CC0 - "Public Domain Dedication"
- **Examples of use**: loading standard clinical datasets into PlatformTM live demo

FAIR in health scie



Data Protection Impact Assessment and Data Privacy

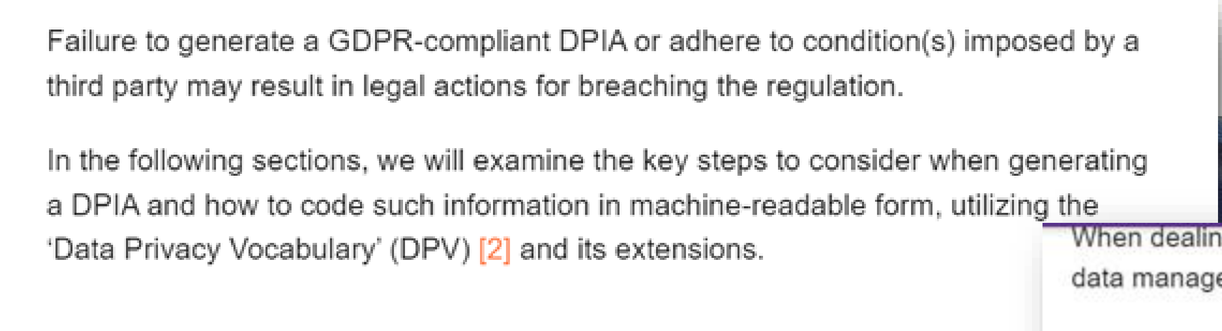
FAIR Cookbook DPIA

Recipe Overview

Reading Time: 15 minutes

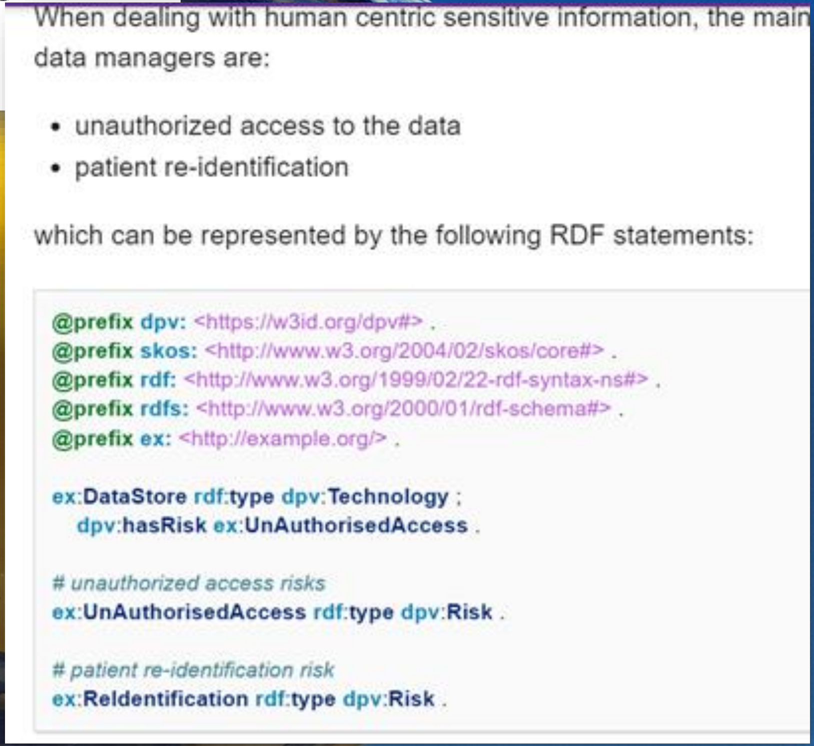
Executable Code: No

Difficulty



Failure to generate a GDPR-compliant DPIA or adhere to condition(s) imposed by a third party may result in legal actions for breaching the regulation.

In the following sections, we will examine the key steps to consider when generating a DPIA and how to code such information in machine-readable form, utilizing the 'Data Privacy Vocabulary' (DPV) [2] and its extensions.



When dealing with human centric sensitive information, the main data managers are:

- unauthorized access to the data
- patient re-identification

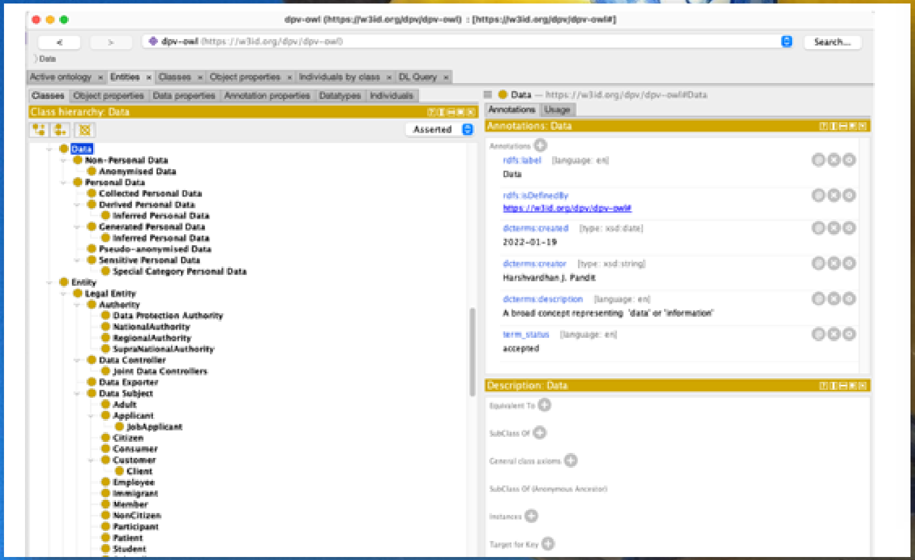
which can be represented by the following RDF statements:

```
@prefix dpv: <https://w3id.org/dpv#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ex: <http://example.org/> .

ex:DataStore rdf:type dpv:Technology ;
   dpv:hasRisk ex:UnAuthorisedAccess .

# unauthorized access risks
ex:UnAuthorisedAccess rdf:type dpv:Risk .

# patient re-identification risk
ex:ReIdentification rdf:type dpv:Risk .
```



- Data
  - Non-Personal Data
    - Anonymised Data
  - Personal Data
    - Collected Personal Data
    - Derived Personal Data
      - Inferred Personal Data
    - Generated Personal Data
      - Inferred Personal Data
    - Pseudo-anonymised Data
    - Sensitive Personal Data
      - Special Category Personal Data
- Entity
  - Legal Entity
    - Authority
      - Data Protection Authority
      - NationalAuthority
      - RegionalAuthority
      - SupraNationalAuthority
    - Data Controller
      - Joint Data Controllers
    - Data Exporter
    - Data Subject
      - Adult
      - Applicant
        - JobApplicant
      - Citizen
      - Consumer
      - Customer
        - Client
      - Employee
      - Immigrant
      - Member
      - NonCitizen
      - Participant
      - Patient
      - Student

rdfs:label [language: en]
Data

rdfs:isDefinedBy
https://w3id.org/dpv/dpv-owl#

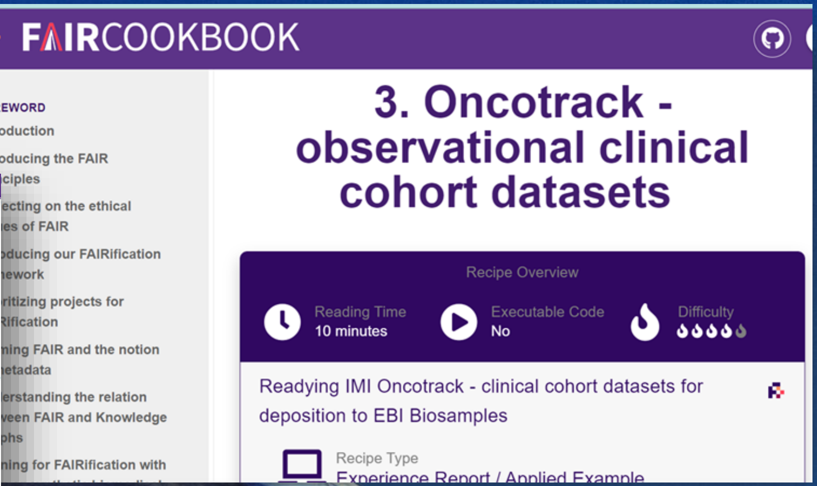
dcterms:created [type: xsd:date]
2022-01-19

dcterms:creator [type: xsd:string]
Harshvardhan J. Pandit

dcterms:description [language: en]
A broad concept representing 'data' or 'information'

term_status [language: en]
accepted

FAIR in health scien



3. Oncotrack - observational clinical cohort datasets

Recipe Overview

Reading Time 10 minutes

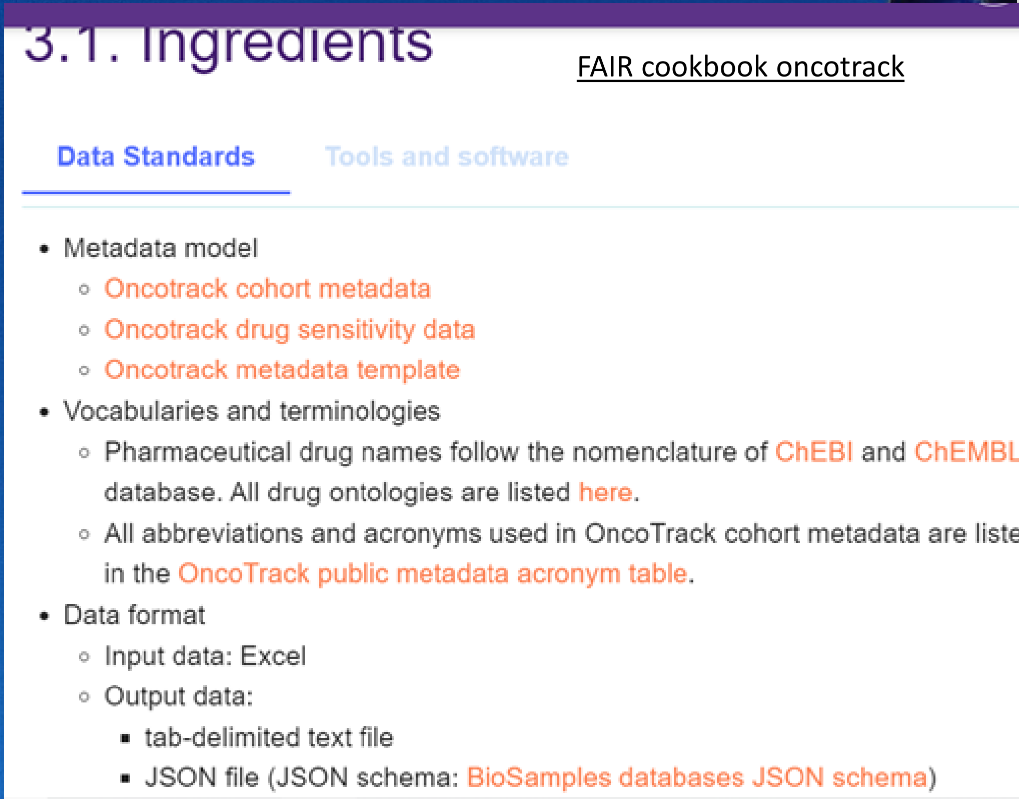
Executable Code No

Difficulty

Readying IMI Oncotrack - clinical cohort datasets for deposition to EBI Biosamples

Recipe Type

Experience Report / Applied Example



3.1. Ingredients

FAIR cookbook oncotrack

Data Standards | Tools and software

- Metadata model
  - Oncotrack cohort metadata
  - Oncotrack drug sensitivity data
  - Oncotrack metadata template
- Vocabularies and terminologies
  - Pharmaceutical drug names follow the nomenclature of ChEBI and ChEMBL database. All drug ontologies are listed here.
  - All abbreviations and acronyms used in OncoTrack cohort metadata are listed in the OncoTrack public metadata acronym table.
- Data format
  - Input data: Excel
  - Output data:
    - tab-delimited text file
    - JSON file (JSON schema: BioSamples databases JSON schema)

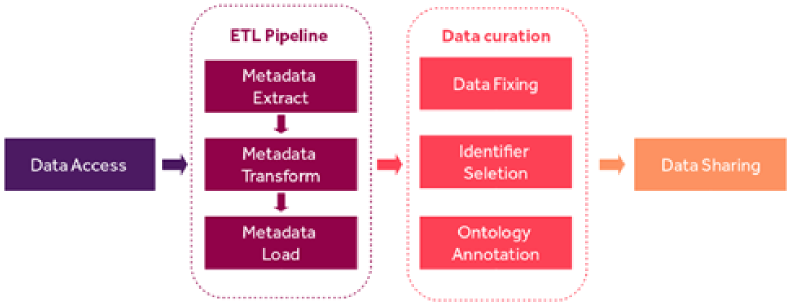
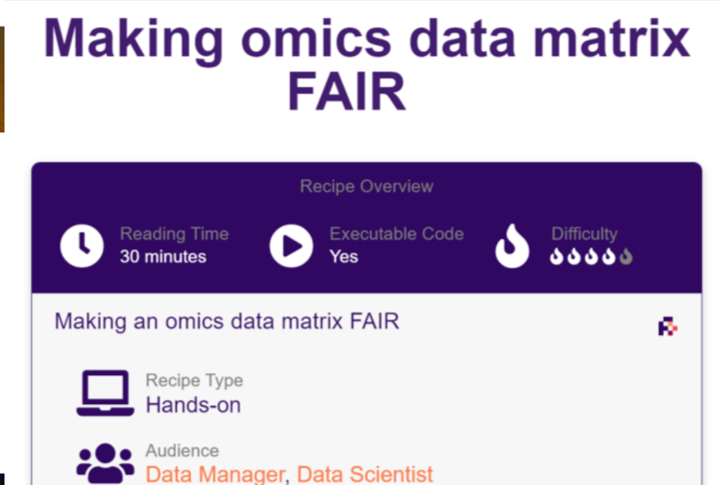


Fig. 3.5 OncoTrack metadata FAIRification pipeline.



Making omics data matrix FAIR

Recipe Overview

Reading Time 30 minutes

Executable Code Yes

Difficulty

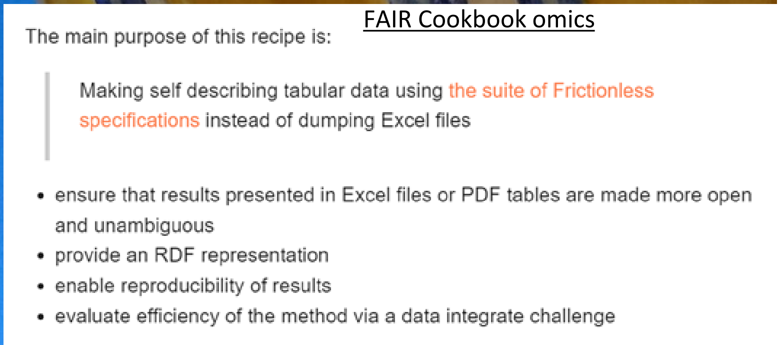
Making an omics data matrix FAIR

Recipe Type

Hands-on

Audience

Data Manager, Data Scientist



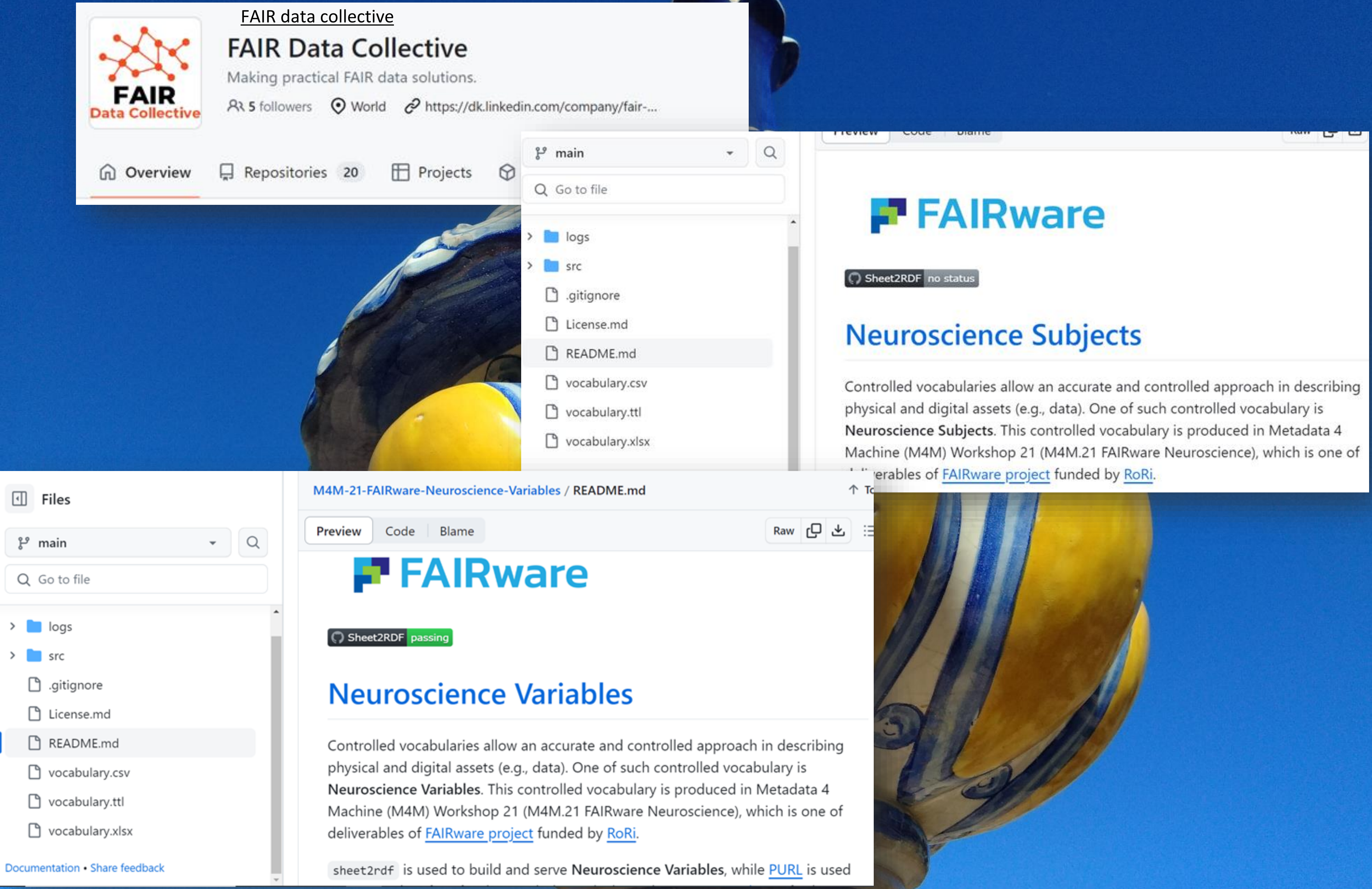
FAIR Cookbook omics

The main purpose of this recipe is:

> Making self describing tabular data using the suite of Frictionless specifications instead of dumping Excel files

- ensure that results presented in Excel files or PDF tables are made more open and unambiguous
- provide an RDF representation
- enable reproducibility of results
- evaluate efficiency of the method via a data integrate challenge

FAIR in health sciences



FAIR data collective

FAIR Data Collective

Making practical FAIR data solutions.

5 followers · World · https://dk.linkedin.com/company/fair-...

Neuroscience Subjects

Controlled vocabularies allow an accurate and controlled approach in describing physical and digital assets (e.g., data). One of such controlled vocabulary is **Neuroscience Subjects**. This controlled vocabulary is produced in Metadata 4 Machine (M4M) Workshop 21 (M4M.21 FAIRware Neuroscience), which is one of erables of FAIRware project funded by RoRi.

M4M-21-FAIRware-Neuroscience-Variables / README.md

Neuroscience Variables

Controlled vocabularies allow an accurate and controlled approach in describing physical and digital assets (e.g., data). One of such controlled vocabulary is **Neuroscience Variables**. This controlled vocabulary is produced in Metadata 4 Machine (M4M) Workshop 21 (M4M.21 FAIRware Neuroscience), which is one of deliverables of FAIRware project funded by RoRi.

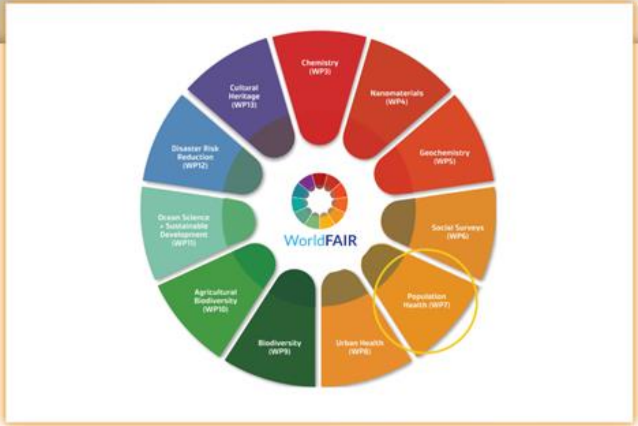
`sheet2rdf` is used to build and serve **Neuroscience Variables**, while PURL is used

FAIR in health sciences

WorldFAIR

World FAIR

Population Health



The Implementation Network for Sharing Population Information from Research Entities (INSPIRE) project is assembling technologies and standards in support of a data hub that facilitates federated and/or shared research capable of interoperating across often-neglected low-resource settings: it aims to provide a platform-as-a-service, which can make data of disparate types available to many different styles of analysis, among which AI systems are increasingly prominent.

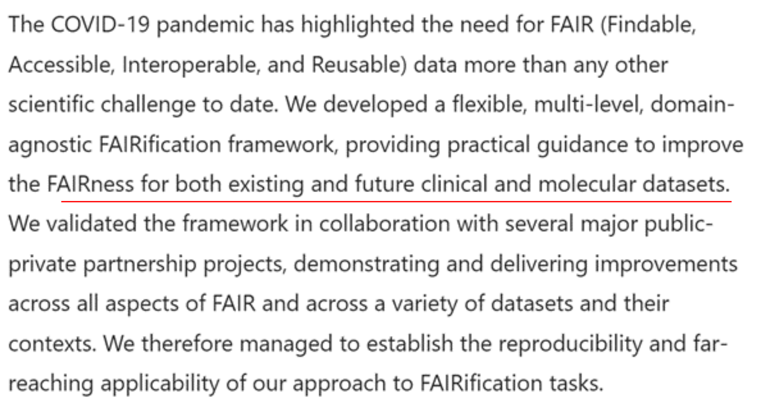
INSPIRE uses OMOP, a common data model that is becoming the gold standard for systematically integrating health data from disparate sources and conducting observational research at scale using routine clinical care data. However, OMOP is not completely FAIR29 and further work is needed to improve the ability to integrate diverse sources of data.

POPULATION HEALTH DATA IMPLEMENTATION GUIDE

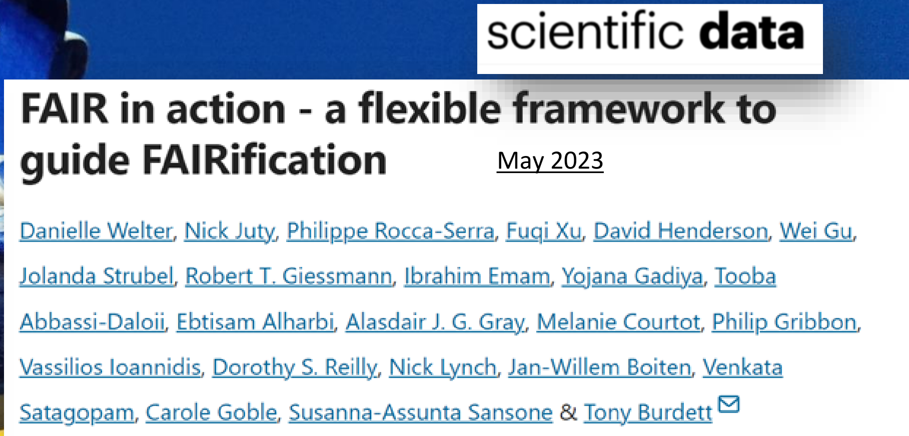
This implementation guide describes the way all aspects of the data are made available for use, both within and from outside the INSPIRE Network community, using standard metadata to describe the data. This is an exploration of how generic standards can be used to express the agreed community metadata set.

Read more.

FAIR in health sciences



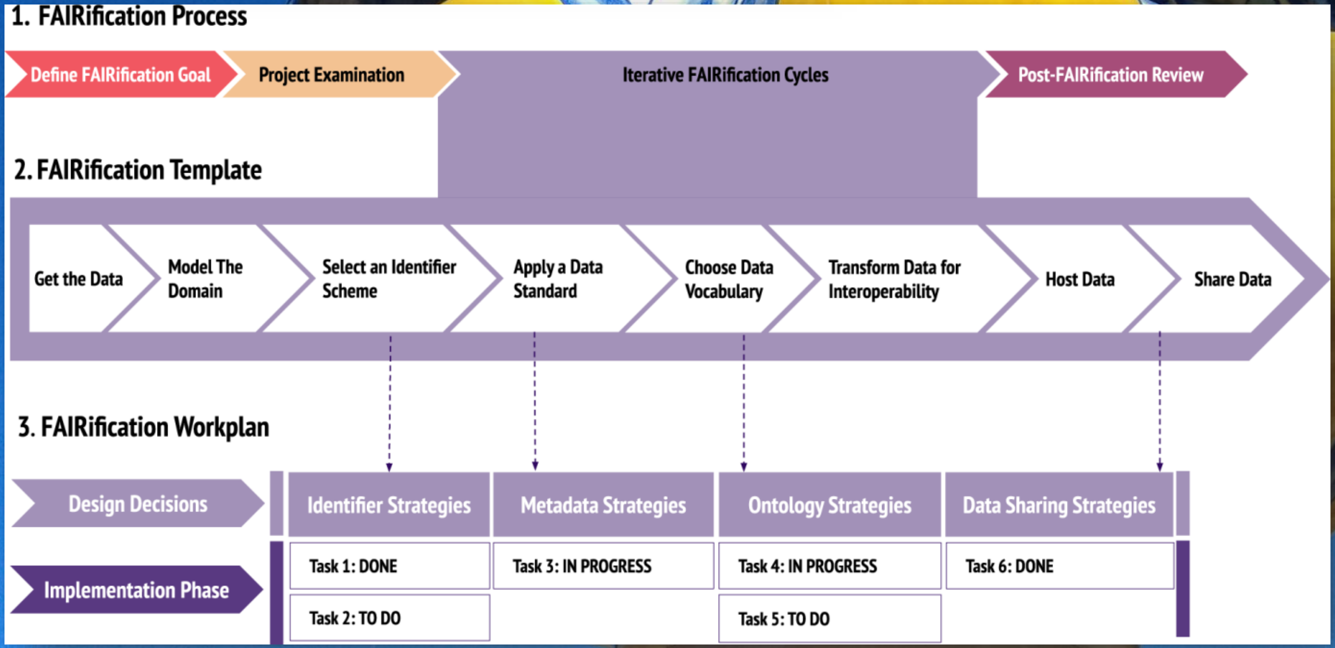
The COVID-19 pandemic has highlighted the need for FAIR (Findable, Accessible, Interoperable, and Reusable) data more than any other scientific challenge to date. We developed a flexible, multi-level, domain-agnostic FAIRification framework, providing practical guidance to improve the FAIRness for both existing and future clinical and molecular datasets. We validated the framework in collaboration with several major public-private partnership projects, demonstrating and delivering improvements across all aspects of FAIR and across a variety of datasets and their contexts. We therefore managed to establish the reproducibility and far-reaching applicability of our approach to FAIRification tasks.



scientific **data**

FAIR in action - a flexible framework to guide FAIRification May 2023

Danielle Welter, Nick Juty, Philippe Rocca-Serra, Fuqi Xu, David Henderson, Wei Gu, Jolanda Strubel, Robert T. Giessmann, Ibrahim Emam, Yojana Gadiya, Tooba Abbassi-Daloii, Ebtisam Alharbi, Alasdair J. G. Gray, Melanie Courtot, Philip Gribbon, Vassilios Ioannidis, Dorothy S. Reilly, Nick Lynch, Jan-Willem Boiten, Venkata Satagopam, Carole Goble, Susanna-Assunta Sansone & Tony Burdett



1. FAIRification Process

Define FAIRification Goal → Project Examination → Iterative FAIRification Cycles → Post-FAIRification Review

2. FAIRification Template

Get the Data → Model The Domain → Select an Identifier Scheme → Apply a Data Standard → Choose Data Vocabulary → Transform Data for Interoperability → Host Data → Share Data

3. FAIRification Workplan

Design Decisions: Identifier Strategies | Metadata Strategies | Ontology Strategies | Data Sharing Strategies

Implementation Phase:

Task 1: DONE

Task 2: TO DO

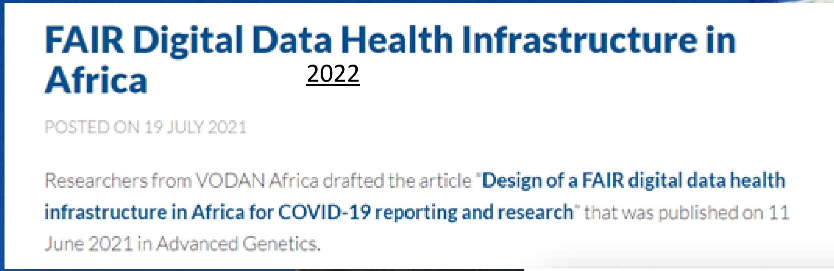
Task 3: IN PROGRESS

Task 4: IN PROGRESS

Task 5: TO DO

Task 6: DONE

FAIR in health sciences

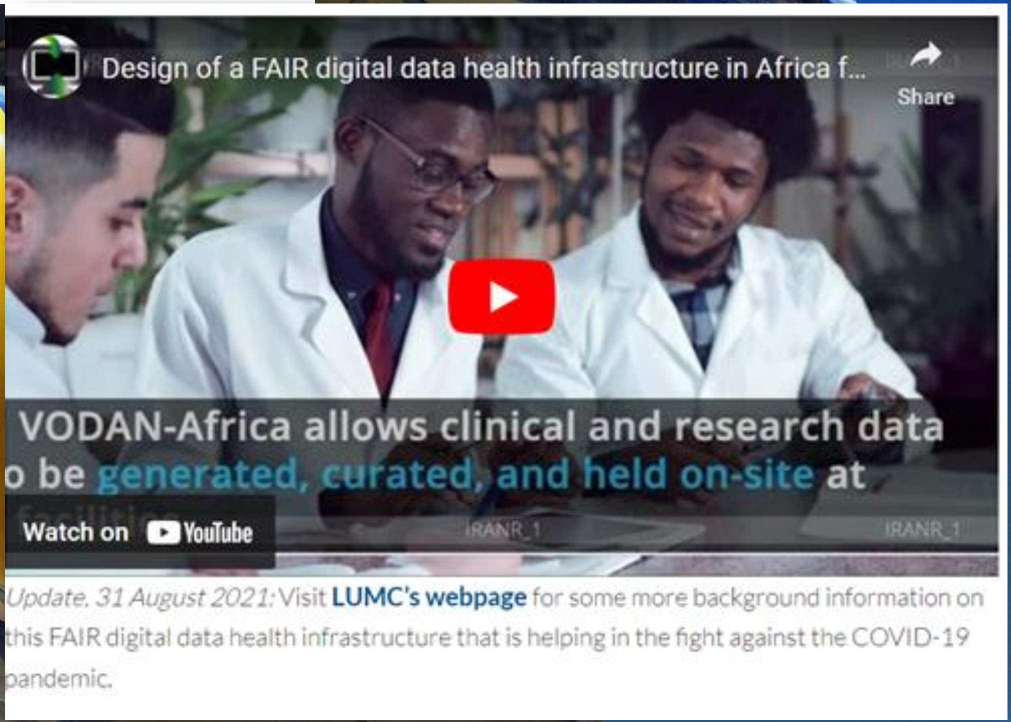


FAIR Digital Data Health Infrastructure in Africa

2022

POSTED ON 19 JULY 2021

Researchers from VODAN Africa drafted the article "**Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research**" that was published on 11 June 2021 in Advanced Genetics.



Design of a FAIR digital data health infrastructure in Africa f...

VODAN-Africa allows clinical and research data o be generated, curated, and held on-site at

Update, 31 August 2021: Visit **LUMC's webpage** for some more background information on this FAIR digital data health infrastructure that is helping in the fight against the COVID-19 pandemic.

STEP #3
OPEN DATA

...concerns

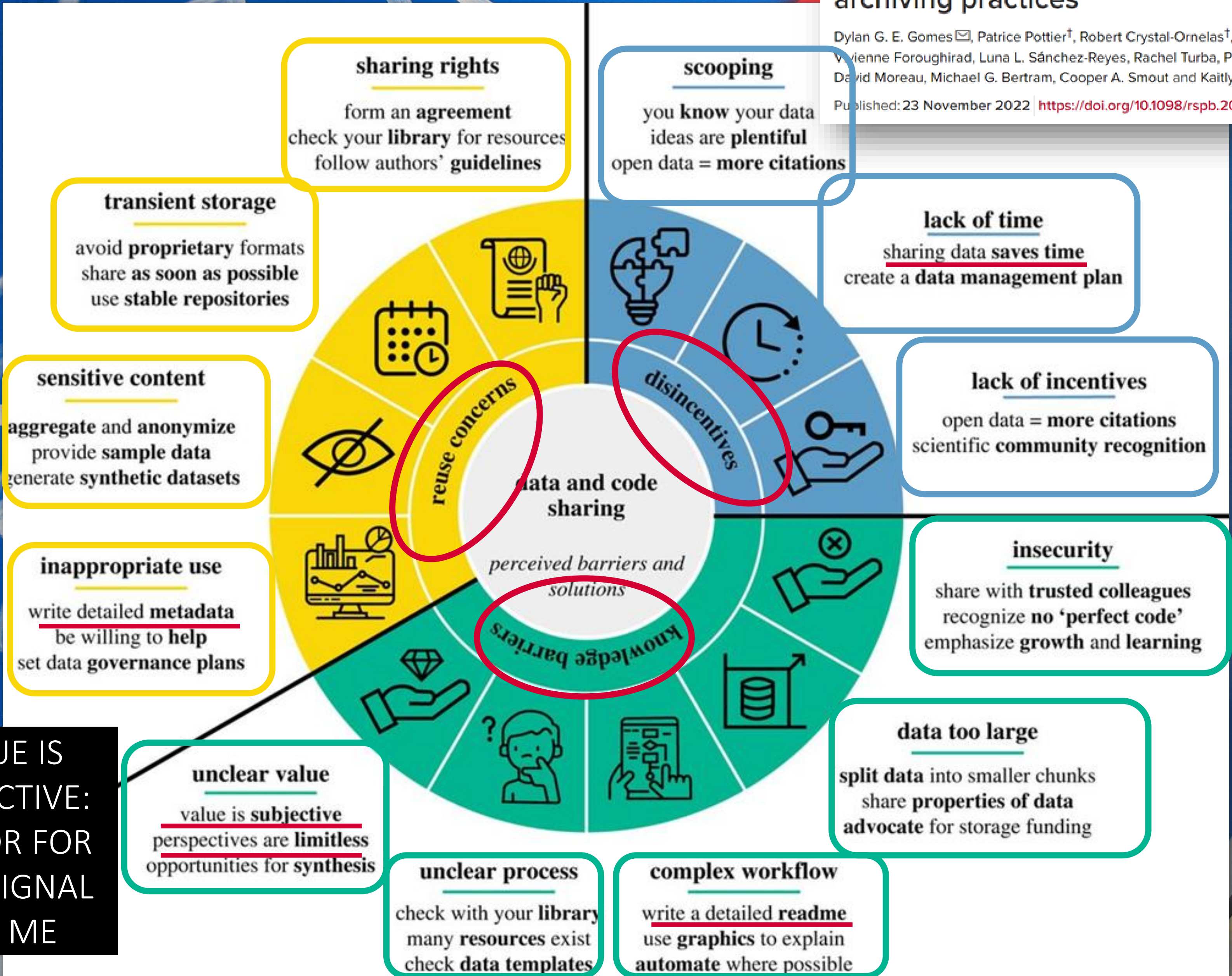
Biological science practices

2022

Why don't we share data and code? Perceived barriers and benefits to public archiving practices

Dylan G. E. Gomes, Patrice Pottier†, Robert Crystal-Ornelas†, Emma J. Hudgins, Vivienne Foroughirad, Luna L. Sánchez-Reyes, Rachel Turba, Paula Andrea Martine David Moreau, Michael G. Bertram, Cooper A. Smout and Kaitlyn M. Gaynor

Published: 23 November 2022 https://doi.org/10.1098/rspb.2022.1113



data and code sharing

perceived barriers and solutions

reuse concerns

disincentives

knowledge barriers

sharing rights
form an **agreement**
check your **library** for resources
follow authors' **guidelines**

scooping
you **know** your data
ideas are **plentiful**
open data = **more citations**

transient storage
avoid **proprietary** formats
share **as soon as possible**
use **stable repositories**

lack of time
sharing data **saves time**
create a **data management** plan

sensitive content
aggregate and **anonymize**
provide **sample data**
generate **synthetic datasets**

lack of incentives
open data = **more citations**
scientific **community recognition**

inappropriate use
write detailed **metadata**
be willing to **help**
set data **governance plans**

insecurity
share with **trusted colleagues**
recognize **no 'perfect code'**
emphasize **growth** and **learning**

data too large
split data into smaller chunks
share **properties of data**
advocate for storage funding

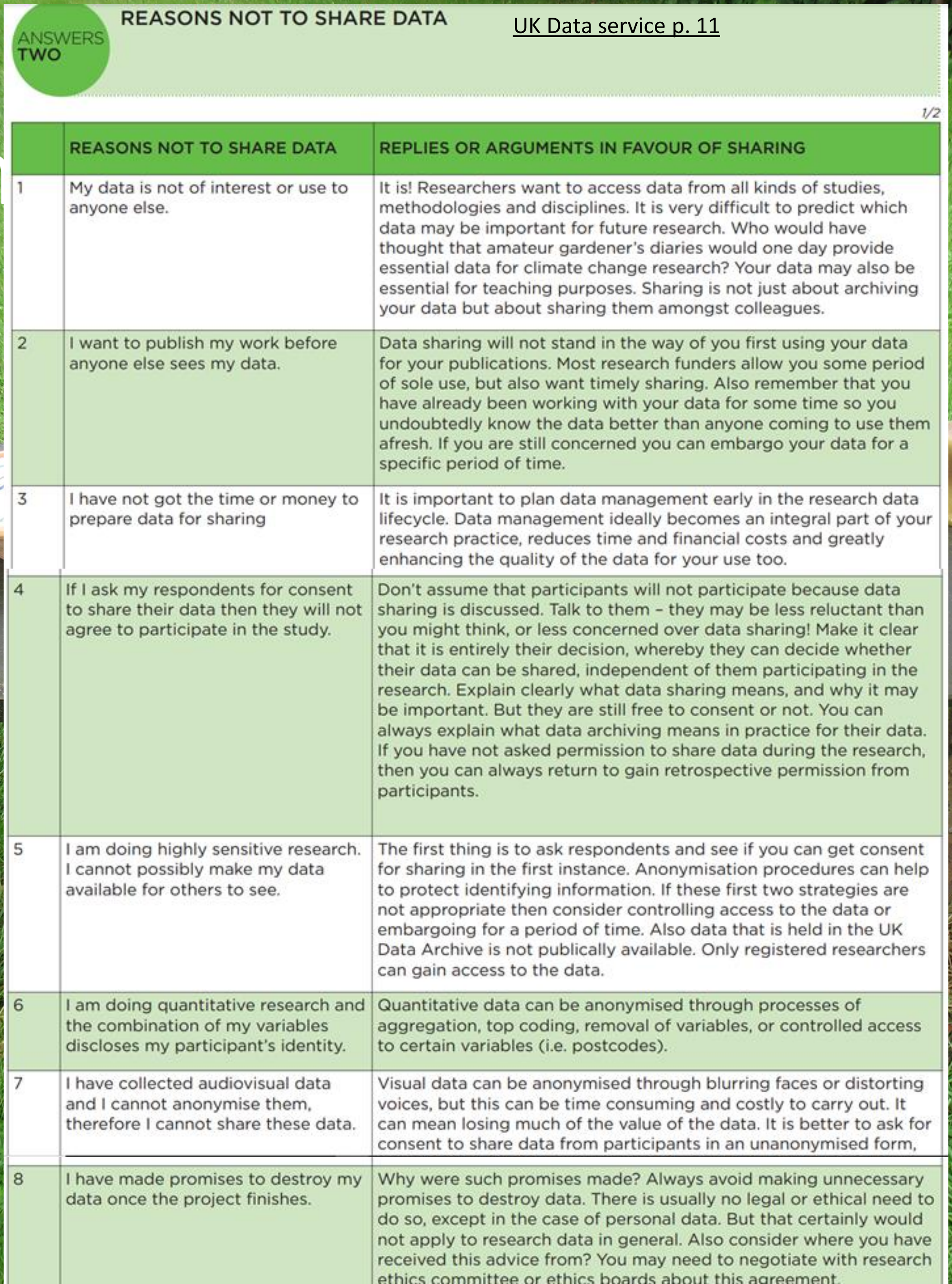
unclear value
value is **subjective**
perspectives are **limitless**
opportunities for **synthesis**

unclear process
check with your **library**
many **resources** exist
check **data templates**

complex workflow
write a detailed **readme**
use **graphics** to explain
automate where possible

VALUE IS SUBJECTIVE: RUMOR FOR YOU, SIGNAL FOR ME

Pro an



ANSWERS TWO

REASONS NOT TO SHARE DATA

UK Data service p. 11

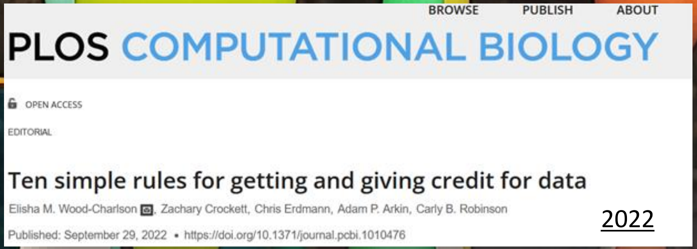
1/2

	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.

ARGUMENTS IN FAVOUR OF SHARING

Data Management Plans: the pillars of your research

DATA MANAGEMENT PLANS ARE YOUR FIRST RESEARCH «PRODUCT» IF YOU WANT YOUR DATA TO BE AVAILABLE AND REUSABLE (EVEN BY YOURSELVES!!!)



BROWSE PUBLISH ABOUT

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

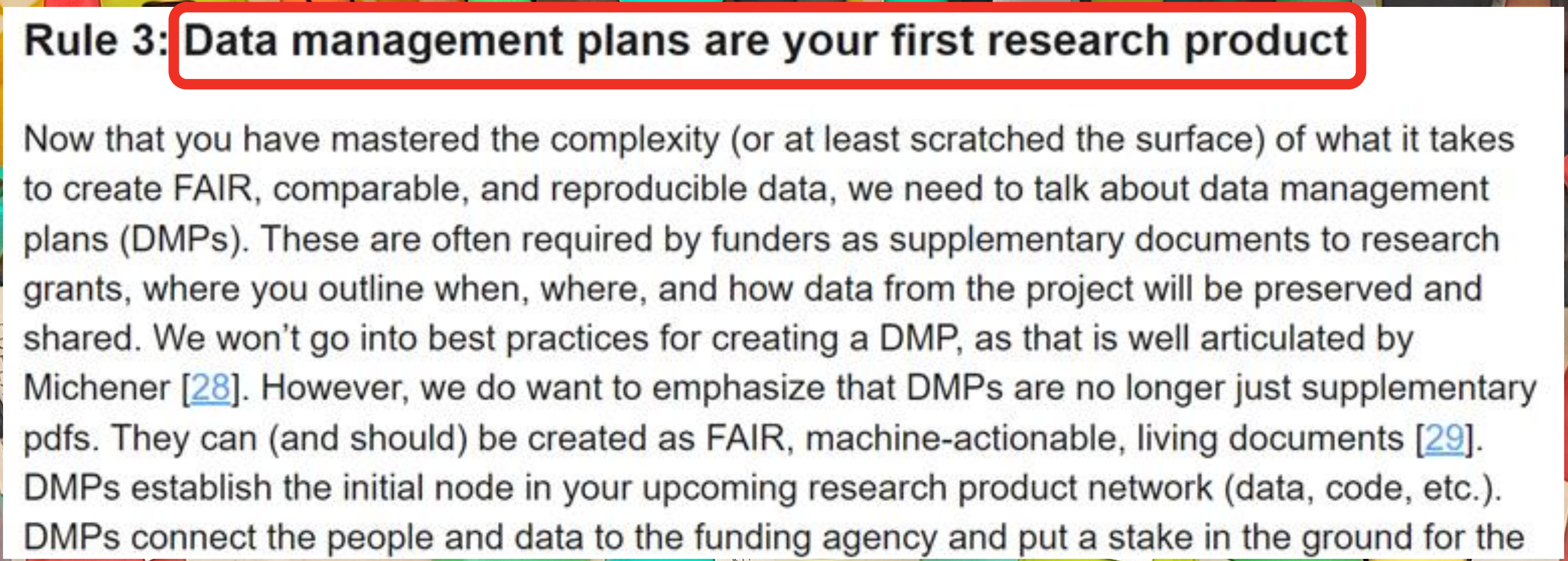
EDITORIAL

Ten simple rules for getting and giving credit for data

Elisha M. Wood-Charlson, Zachary Crockett, Chris Erdmann, Adam P. Arkin, Carly B. Robinson

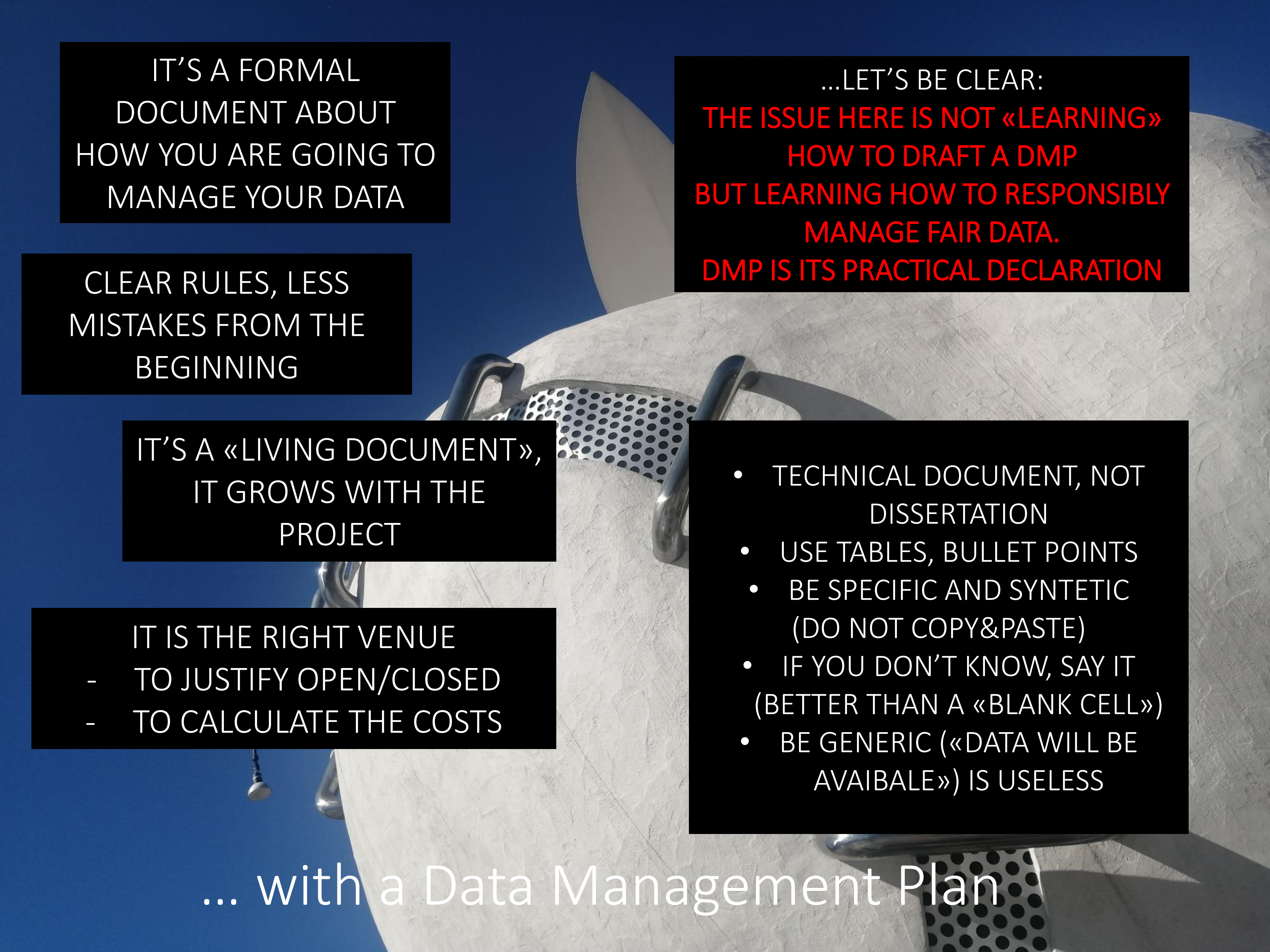
Published: September 29, 2022 • https://doi.org/10.1371/journal.pcbi.1010476

2022



Rule 3: Data management plans are your first research product

Now that you have mastered the complexity (or at least scratched the surface) of what it takes to create FAIR, comparable, and reproducible data, we need to talk about data management plans (DMPs). These are often required by funders as supplementary documents to research grants, where you outline when, where, and how data from the project will be preserved and shared. We won't go into best practices for creating a DMP, as that is well articulated by Michener [28]. However, we do want to emphasize that DMPs are no longer just supplementary pdfs. They can (and should) be created as FAIR, machine-actionable, living documents [29]. DMPs establish the initial node in your upcoming research product network (data, code, etc.). DMPs connect the people and data to the funding agency and put a stake in the ground for the



IT'S A FORMAL DOCUMENT ABOUT HOW YOU ARE GOING TO MANAGE YOUR DATA

CLEAR RULES, LESS MISTAKES FROM THE BEGINNING

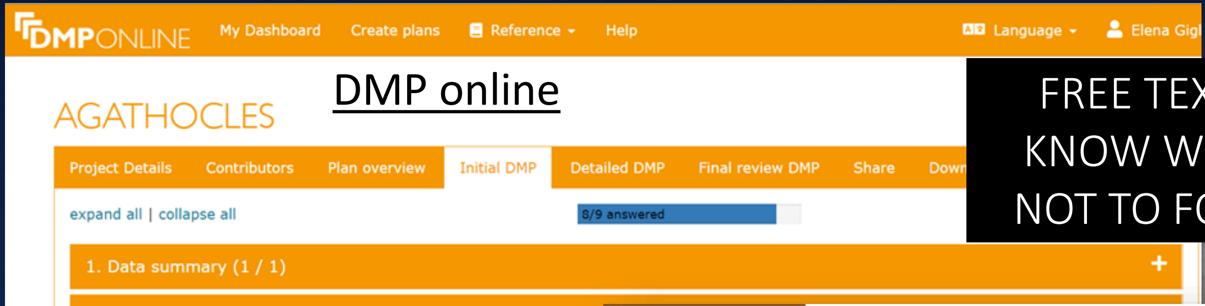
IT'S A «LIVING DOCUMENT», IT GROWS WITH THE PROJECT

IT IS THE RIGHT VENUE
- TO JUSTIFY OPEN/CLOSED
- TO CALCULATE THE COSTS

...LET'S BE CLEAR:
THE ISSUE HERE IS NOT «LEARNING» HOW TO DRAFT A DMP BUT LEARNING HOW TO RESPONSIBLY MANAGE FAIR DATA.
DMP IS ITS PRACTICAL DECLARATION

- TECHNICAL DOCUMENT, NOT DISSERTATION
- USE TABLES, BULLET POINTS
- BE SPECIFIC AND SYNTETIC (DO NOT COPY&PASTE)
- IF YOU DON'T KNOW, SAY IT (BETTER THAN A «BLANK CELL»)
- BE GENERIC («DATA WILL BE AVAIBALE») IS USELESS

... with a Data Management Plan



DMPONLINE My Dashboard Create plans Reference Help Language Elena Gigl

AGATHOCLES

Project Details | Contributors | Plan overview | Initial DMP | Detailed DMP | Final review DMP | Share | Down

expand all | collapse all 8/9 answered

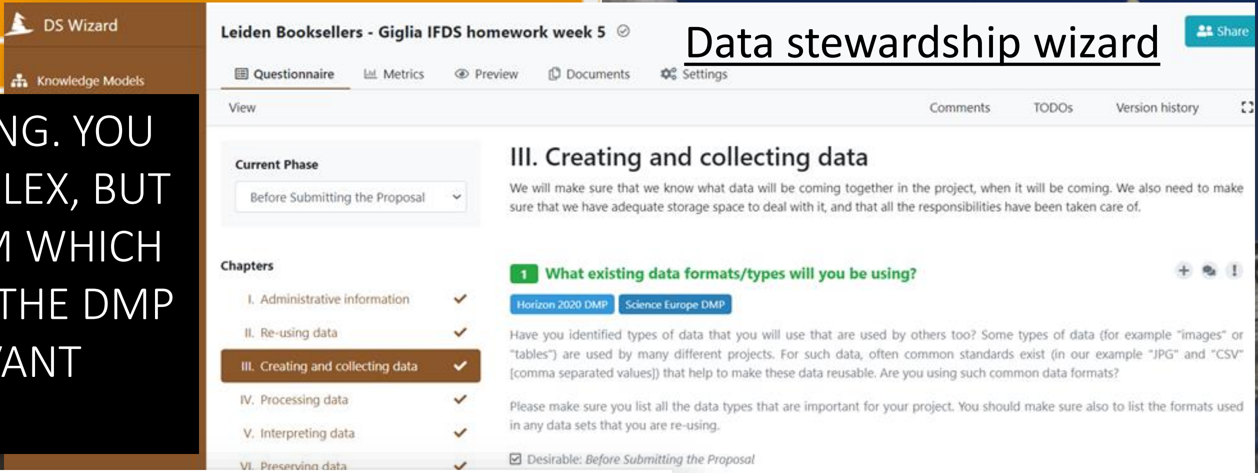
1. Data summary (1 / 1) +

2. FAIR data (3 / 4)

3. Allocation of resources (1 / 1)

DMP online

FREE TEXT. YOU HAVE TO KNOW WHAT TO ADDRESS NOT TO FORGET ANYTHING



DS Wizard

Knowledge Models

Leiden Booksellers - Giglia IFDS homework week 5

Questionnaire | Metrics | Preview | Documents | Settings | Share

View | Comments | TODOs | Version history

Current Phase

Before Submitting the Proposal

Chapters

I. Administrative information ✓

II. Re-using data ✓

III. Creating and collecting data ✓

IV. Processing data ✓

V. Interpreting data ✓

VI. Preserving data ✓

III. Creating and collecting data

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

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

Horizon 2020 DMP | Science Europe DMP

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

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

☑ Desirable: *Before Submitting the Proposal*

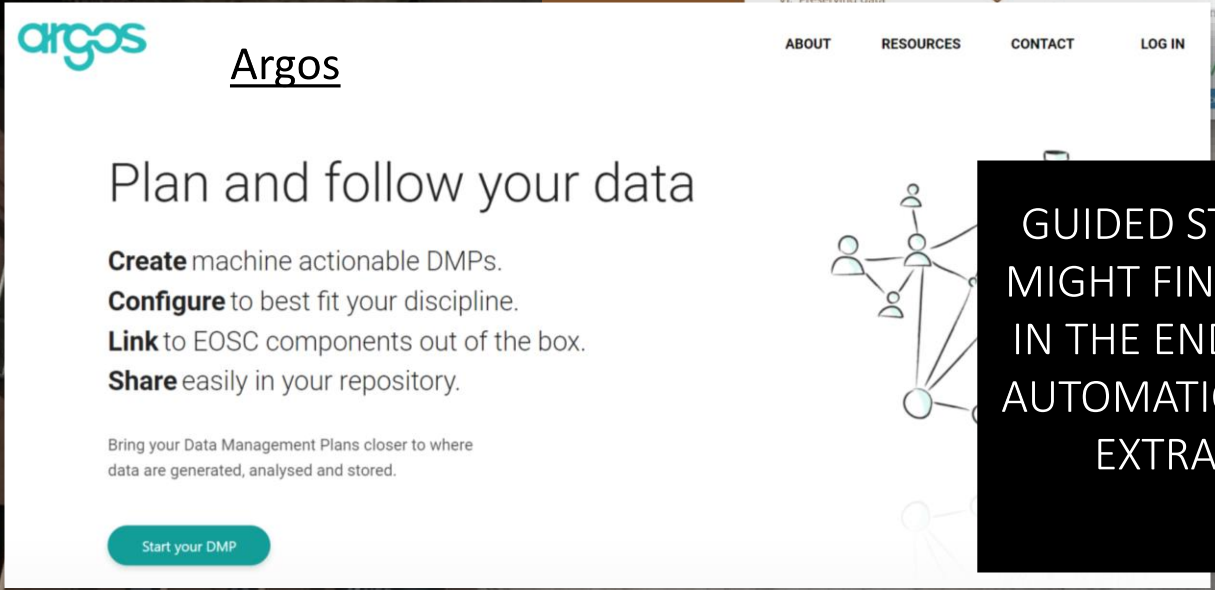
n Science: *njy*

/type

e Europe DMP

Data stewardship wizard

GUIDED STEP TO STEP FILLING. YOU MIGHT FIND IT MORE COMPLEX, BUT IN THE END IT'S THE SYSTEM WHICH AUTOMATICALLY GENERATE THE DMP EXTRACTING THE RELEVANT INFORMATION



argos ABOUT RESOURCES CONTACT LOG IN

Plan and follow your data

Create machine actionable DMPs.
Configure to best fit your discipline.
Link to EOSC components out of the box.
Share easily in your repository.

Bring your Data Management Plans closer to where data are generated, analysed and stored.

Start your DMP

Argos

GUIDED STEP TO STEP FILLING. YOU MIGHT FIND IT MORE COMPLEX, BUT IN THE END IT'S THE SYSTEM WHICH AUTOMATICALLY GENERATE THE DMP EXTRACTING THE RELEVANT INFORMATION

THANK YOU!