FAIR: what's in it for me? [looking ahead to the next 10 years]

Rome, iEntrance Advanced
 school 2024
22-2-2024

Elena Giglia

elena.giglia@unito.it

𝕏 @egiglia

Why are we here today?

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)

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

THANK YOU FOR YOUR UNDIVIDED ATTENTION...I'M DONE

JUST KIDDING! WE ARE GOING TO SEE
- WHY DO WE NEED FAIR OPEN DATA
  (I.E. LESSONS LEARNED FROM COVID)
- THE EUROPEAN/INTERNATIONAL FRAMEWORK
- SOME [HOPEFULLY USEFUL] TOOLS
- PAST AND FUTURE OF FAIR

Why should we care about data? A personal lesson
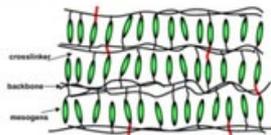

Past scientific interests

Mathematical models for soft-active materials
- Elasticity within large deformation framework (non-linear models)
- Deformation of active-smart materials (swelling materials, nematic elastomers, …)
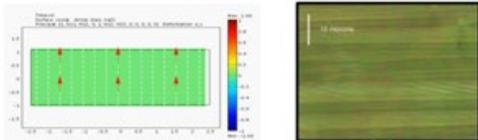
M. de Luca, A. DeSimone. Elastomeric Gels: A Model and First Results. Innovative Numerical Approaches for Multi-Field and Multi-Scale Problems.
Lecture Notes in Applied and Computational Mechanics, vol 81. Springer, Cham. (2016)
https://doi.org/10.1007/978-3-319-39022-2_4

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

Research (FAIR) data management 2023

|Mariarita de Luca|

https://orcid.org/0000-0002-5507-968X
mariarita.deluca@areasciencepark.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
This work © 2023 by Mariarita de Luca is licensed under CC BY 4.0

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

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]

Lessons learned from COVID

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

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...

… …**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
Congratulations to the authors but I am not strong enough for this
Mostra 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

... 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...

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

Retraction Watch
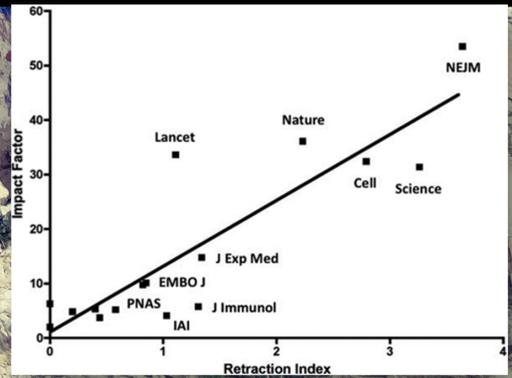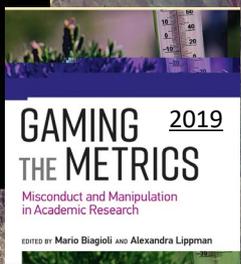Tracking retractions as a window into the scientific process

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

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

GAMING THE METRICS 2019
Misconduct and Manipulation in Academic Research
EDITED BY Mario Biagioli AND Alexandra Lippman

EXTENDED VERSION ON WWW.OA.UNITO.IT

TIME IS UP!!!

- THE REFORM OF RESEARCH EVALUATION HAS STARTED
- COARA LAUNCHED IN 2022, 644 SIGNATORIES
- ITALIAN CHAPTER IS ACTIVE, THENATIONAL AGENCY ANVUR SIGNED
- COMMITTMENT: NO LONGER IMPACT FACTOR OR RANKING

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.

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

Signatories

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

The Commitments

The Agreement, based on 10 commitments, establishes a common direction for research assessment reform, while respecting organisations' autonomy.

The Agreement full text | The Commitments | The Timeframe | COARA | FAQ

The Commitments

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

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

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

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

5. Commit resources to reforming research assessment as is needed to achieve the organisational changes committed to

Open Science
might help?

Open Science – definition

Open Access | Lic. Info | Cite

Qeios

Open Science

'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

[Houston, we have a problem

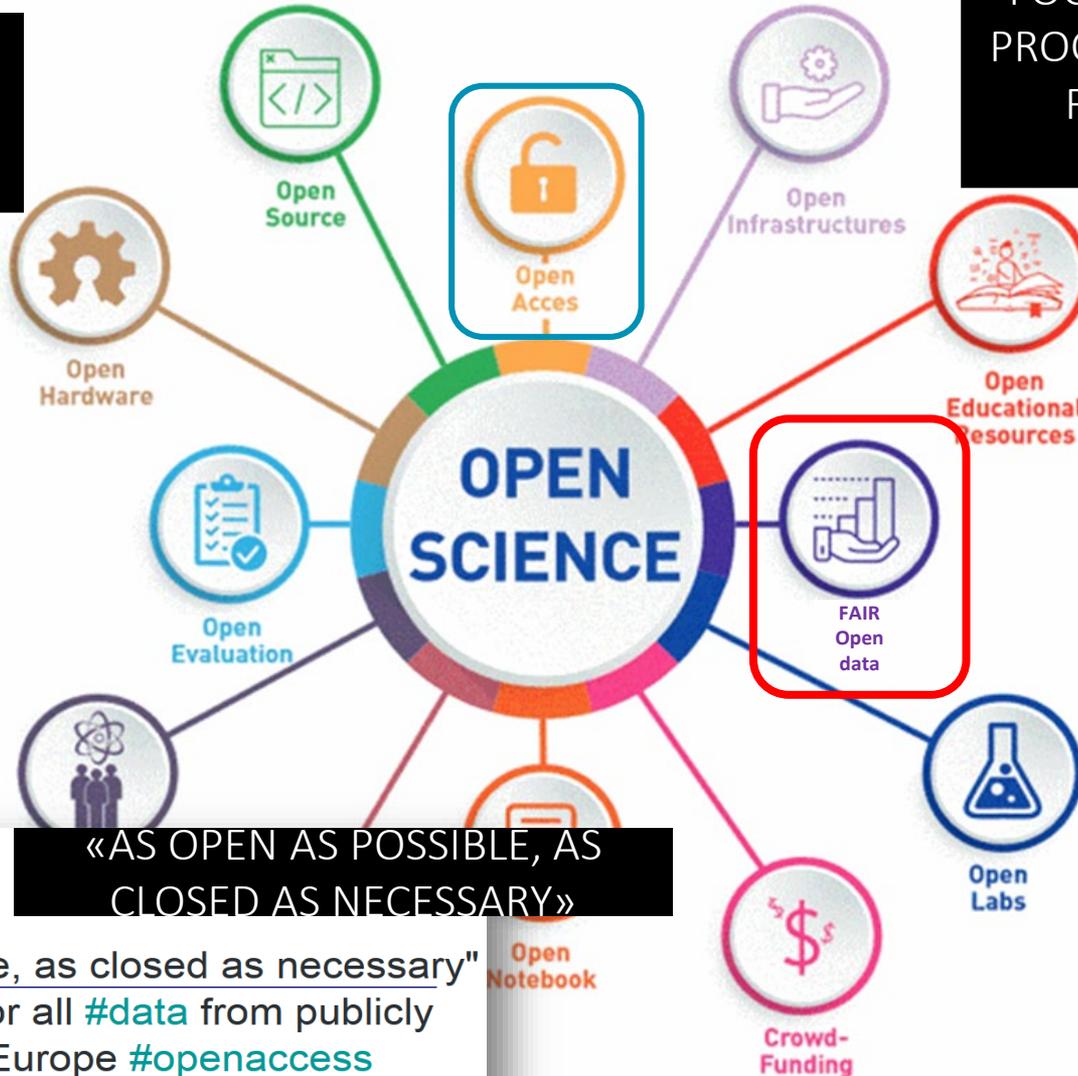## 10 Myths around Open Scholarly Publishing
March 11, 2019

Myth 1
Preprints will get your research 'scooped'

Preprints typically provide a time-stamp and a DOI, therefore establishing priority of discovery

Myth 2
JIF and journal branding are measures of quality for researchers

The JIF is a flawed metrics that was never meant to be used for evaluation of research and researchers

Myth 3
Approval by peer review proves that you can trust a research article

The current peer review system is prone to a number of flaws including corruption, human bias and ghostwriting

Myth 4
Without journal peer review, the quality of science suffers

Researchers are more than responsible and competent enough to ensure their own quality control as part of intrinsic scientific integrity

Myth 5
Open Access has created predatory publishers

Predatory journals have been around for a long time before the recent push towards Open Access publishing

Myth 6
Copyright transfer is required to publish and protect authors

Copyright transfer procedures do not protect authors nor contribute to the advancement of scientific progress

Myth 7
Gold Open Access is synonymous with the APC business model

Most DOAJ-indexed journals do not have APCs and are funded from other sources, such as research institutes and grants

Myth 8
Embargo periods on 'green' OA are needed to sustain publishers

Traditional journals can peacefully coexist with zero-embargo self-archiving policies on author manuscripts

Myth 9
Web of Science and Scopus are global databases of knowledge

Neither represent the sum of current global research knowledge including Africa, Latin America and Southeast Asia

Myth 10
Publishers add no value to the scholarly communication process

Publishers are responsible for quite some key functions, from peer-review management to production and archiving of final version articles

1/12	2/12	3/12	4/12
Open Science is just a gimmick...	Open Science is all about publishing Open Access	Open Science is a plot against publishers	I already deposit my works on ResearchGate
5/12	6/12	7/12	8/12
An open access dissertation has less chances of being published	I'm afraid of plagiarism	There is no open access journal in my discipline	Open Science is for STEM. As a researcher in SSH this is not important to me
9/12	10/12	11/12	12/12
Science is for researchers only. Citizens cannot improve my research	A Data Management Plan is useless	I am not a Data Manager	Open access to research data is not mandatory

Busting myths on Open Science with the YERUN OS Calendar 2021!
Dec. 2021

DIFFUSED MISCONCEPTIONS:
OPEN SCIENCE=OPEN ACCESS, YOU ALWAYS PAY TO PUBLISH,
OA= PREDATORY, I CAN'T OPEN MY DATA...

... Open Science

BMC Research Notes <u>2022</u>

Home About Articles Submission Guidelines Collections | Submit manuscript
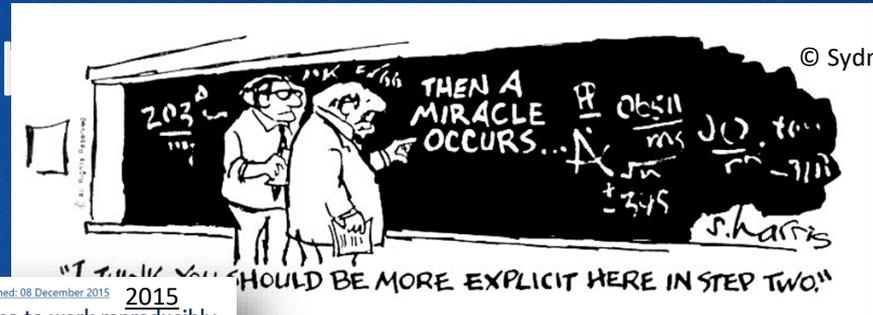
Commentary | Open Access | Published: 20 September 2022
Promoting trust in research and researchers: How open science and research integrity are intertwined
Tamarinde Haven, Gowri Gopalakrishna, Joeri Tijdink, Dorien van der Schot & Lex Bouter

© Sydney Har...

THEN A MIRACLE OCCURS...

S. Harris

"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Comment | Open Access | Published: 08 December 2015 <u>2015</u>
Five selfish reasons to work reproducibly
Florian Markowetz
Genome Biology 16, Article number: 274 (2015) | Cite this article
18k Accesses | 38 Citations | 456 Altmetric | Metrics

Open and Responsible Research
Roles and Responsibilities for Data Stewards
<u>2021</u>
LOUISE BEZUIDENHOUT

Openness as an Extension of Responsibility

Open Lab Books: Transparency in research practices
Sharing and openness: enhance transmission of values

Open Peer Review: Transparency in peer review leads to better dialogue and collegial behaviour

Open Access: Improves availability of research outputs
Open publishing: leads to improved citations, credit and collaboration

RESEARCH MISCONDUCT

MENTOR & TRAINEE RELATIONSHIP

PROTECTION OF HUMAN SUBJECTS

PEER REVIEW

RCR

DATA MANAGEMENT PRACTICES

AUTHORSHIP & PUBLICATIONS

CONFLICTS OF INTERESTS & COMMITMENT

COLLABORATIVE RESEARCH

Open Data and Open Methodologies: Improve transparency and reproducibility of research

KEYWORD=TRANSPARENCY/ RESPONSIBILITY

Open Science Tools: Improve collaboration

FAIR - how

... FAIR principles...

F METADATA, PERSISTENT IDENTIFIERS

A TRUSTED REPOSITORIES, FORMATS

I ONTOLOGIES, STANDARDS

R LICENSES AND DOCUMENTATION

2020

MACHINE ACTIONABLE

TO KNOW MORE

Comment | OPEN

The FAIR Guiding Principles for scientific data management and stewardship

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

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis

FAIR guide, Nature, March 2016

FAIR principles

91 5...

NO MISTAKES!

- **Findability:** Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process [14].

- **Accessibility:** Protocols for retrieving digital resources should be made explicit, for both humans and machines, including well-defined mechanisms to obtain authorization for access to protected data.

- **Interoperability:** When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity. Similarly, when a digital entity is capable of being processed by an online service, a machine should be capable of automatically detecting this compliance and facilitating the interaction between the data and that tool. This requires that the meaning (semantics) of each participating resource – be they data and/or services service – is clear.

- **Reusability:** Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource *should* be reused (i.e., is it relevant to the task at-hand?); if a digital resource *can* be reused, and under what conditions (i.e., do I fulfill the conditions of reuse?); and *who to credit* if it is reused.

GO FAIR Foundation   2022   Home   FAIR   Criteria   Qualification   Fellow...

The FAIR Guiding Principles

Findable

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

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

Interpretation of F1

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

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

Interpretation of A1.2

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

ORIGINAL INTERPRETATION

FAIR research software

FAIR RESEARCH SOFTWARE

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 principles

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 kn

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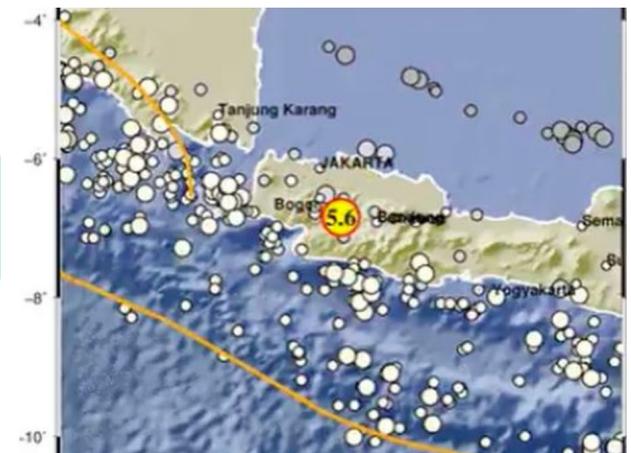DATA CAN BE CLOSED, PROVIDED YOU – AND MACHINES - KNOW WHERE TO FIND THEM AND UNDER WHICH ACCESS CONDITIONS

…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.

IN «FAIR» THE STRESS IS ON «R»

BEWARE…
IF DATA ARE NOT REUSABLE THEY ARE JUST A SELFIE OF DATA
[USELESS]
[Dasapta Erwin Irawan]
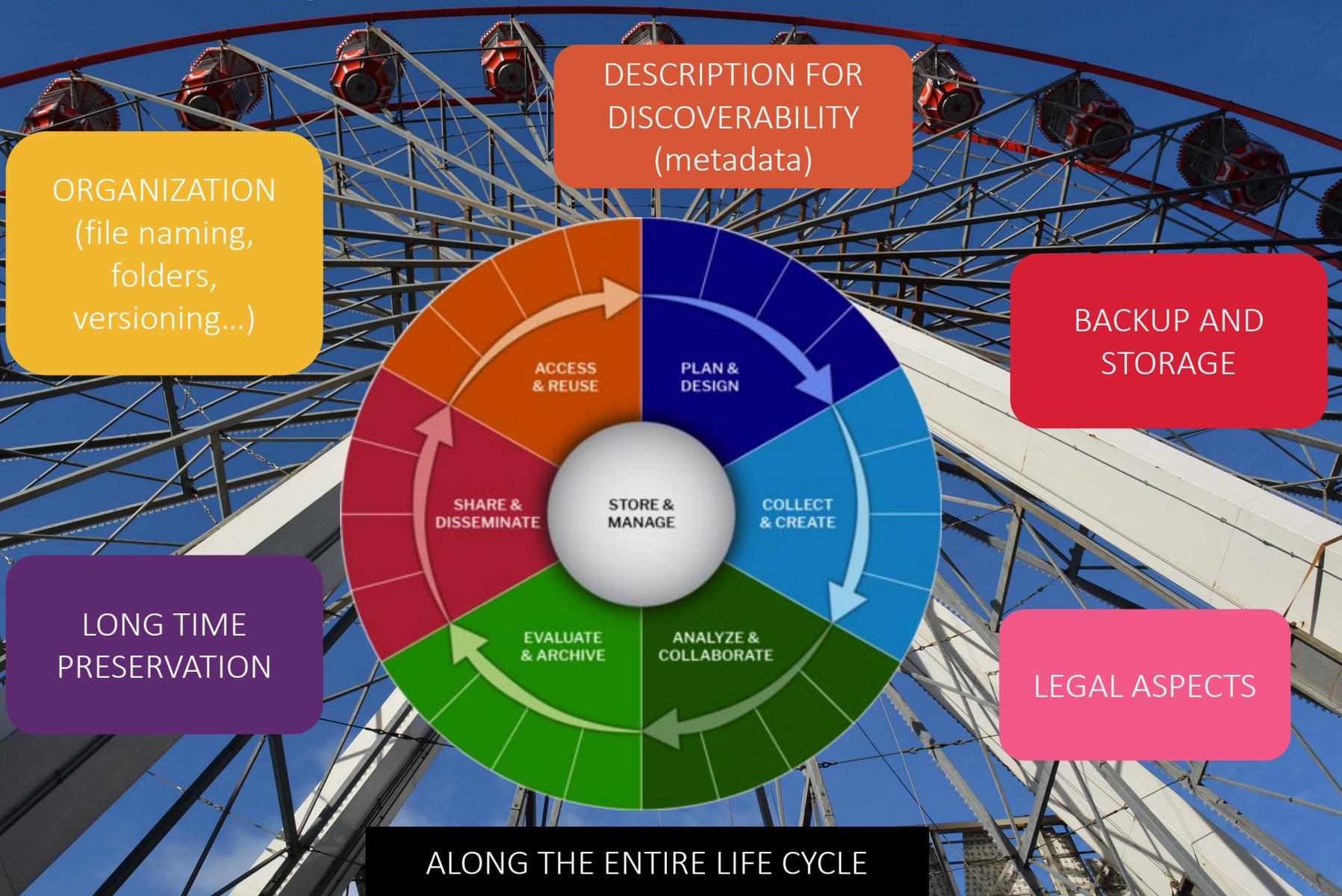
The 3 steps

OPEN **FAIR** **MANAGED**

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»]

1) Manage data

ORGANIZATION (file naming, folders, versioning...)

DESCRIPTION FOR DISCOVERABILITY (metadata)

BACKUP AND STORAGE

LONG TIME PRESERVATION

LEGAL ASPECTS

ACCESS & REUSE

PLAN & DESIGN

SHARE & DISSEMINATE

STORE & MANAGE

COLLECT & CREATE

EVALUATE & ARCHIVE

ANALYZE & COLLABORATE

ALONG THE ENTIRE LIFE CYCLE

2) Make them FAIR

FINDABLE

ACCESSIBLE [≠OPEN]

INTEROPERABLE

REUSABLE

DataCite
FIND, ACCESS, AND REUSE DATA

ORCiD
Connecting Resear
and Researchers

Metadata standards catalog

Metadata Standards Catalog

standards Catalog is a collaborative, open directory of metadata standards
h data. It is offered to the international academic community to help address
nges.

proximus

ROR

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

What are data journals?

Data journals are scholarly journals that publish datasets or data papers. According to
Geoscience Data Journal, "a **data paper describes a dataset**, giving details of its collection,
processing, software, file formats etc, without the requirement of novel analyses or ground
breaking conclusions. It allows the reader to understand the when, how and why data was
t these exist, as this data would

Data journals

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

DANS formats

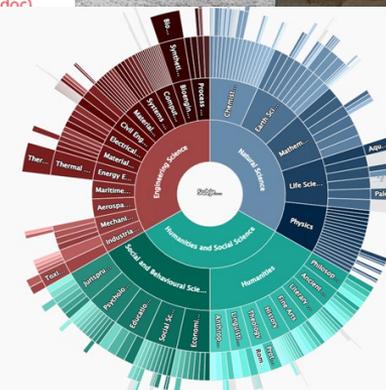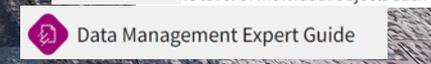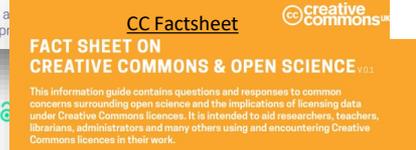| Type	Preferred format(s)	Non-preferred format(s)
Text documents	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open X • Rich Text File (• PDF other than

FAIRsharing.org
standards, databases, policies   search through all content   STANDARDS   DATABASES   POLICIES   COLLECTIONS   ADD CONTENT   STATS
https://fairsharing.org/

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

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

OpenAIRE   SERVICES   SUPP

Guides for Researchers

How do I know if my research
data is protected?

Learn more a
intellectual pr

MIT Press Direct   2020

Data Intelligence

Volume 2, Issue 1-2
Winter-Spring 2020

January 01 2020

Licensing FAIR Data for Reuse
Ignasi Labastida ✉ ● , Thomas Margoni

CC Factsheet

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

Project-level documentation

The project-level documentat
research
instrum
that yo

Data-level documentation

Data-level or object-level
umentation provides information
e level of individual objects such

cessda

Data Management Expert Guide

3) Whenever possible, make ... Open

Sharing Data
Why share data
2. Why share data?

PORT SCHOOL OF ADVANCED STUD... postgraduate online research training UNIVERSITY OF LONDON

Better research

- Demonstrates research integrity, as there is transparency and accountability in the production of the d...
- Encourages research enquiry and debate
- Promotes innovation and potential new...
- Encourages the improvement of researc...
- Prevents research fraud

Better impact

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

Better value

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

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

Digital Science Report
The State of Open Data 2017
f analyses and articles about open data, curated by Figshare
Foreword by Jean-Claude Burgelman
Oct. 2017
OCTOBER 2017

BETTER RESEARCH
- INTEGRITY
- DEBATE
- REUSE

BETTER IMPACT
- VISIBILITY
- CREDIT
- COLLABORATIONS

BETTER VALUE
- AVOID DUPLICATIONS
- MAX RETURN ON INVESTMENTS

FAIR/Open

2018
Final Report and Action Plan from the European Commission Expert Group on FAIR Data

TURNING FAIR INTO REALITY

AS OPEN AS POSSIBILE
AS EARLY AS POSSIBLE
AS FAIR AS POSSIBLE

Increasing degrees ← → Increasing degrees

FAIR data | Open data
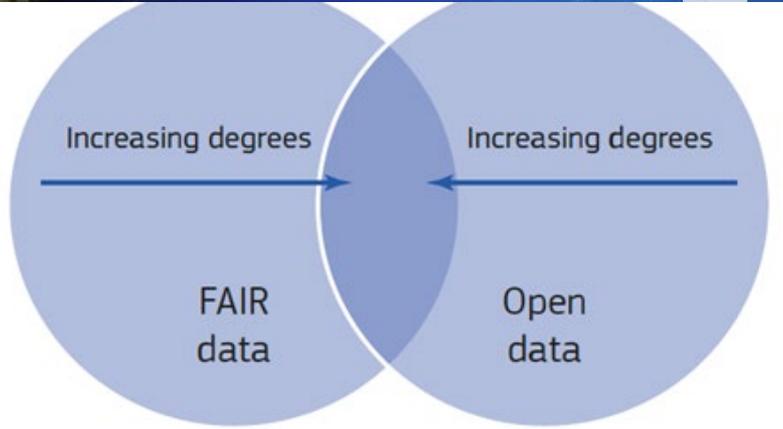
Figure 4. The relationship between FAIR and Open

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

[Open data directive]

CoPER Gruppo Open Science
CONSULTA DEI PRESIDENTI
DEGLI ENTI DI RICERCA

Panoramica sulle politiche istituzionali dei dati negli EPR

Coord. **Mario Locati** (INGV), Stefano Bianco (INFN), Anna Grazia Chiodetti (INGV)

Contributi di Angela Saraò (OGS), Monica Sala e Daniela Palma (ENEA),
Riccardo Scano (CREA), Carlo Cipolloni (ISPRA), Dario Menasce (INFN)

Convegno OpenScience @ INAF
INAF, Sede Centrale, Roma, 14-15 dicembre 2023

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

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

FOR THE ITALIAN FRAMEWORK

FAIR DOES NOT MEAN OPEN ...BUT... WE HAVE THE OPEN DATA DIRECTIVE, SO «OPEN BY DEFAULT»

- Stimulate the publishing of dynamic data and the uptake of Application Programme Interfaces (APIs).
- Limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data.
- Enlarge the scope of the Directive to:
  - data held by public undertakings, under a specific set of rules. In principle, the Directive will only apply to data which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination;
  - research data resulting from public funding – Member States will be asked to develop policies for open access to publicly funded research data. New rules will also facilitate the re-usability of research data that is already contained in open repositories.
- Strengthen the transparency requirements for public–private

«OPEN DATA» DIRECTIVE ENLARGED TO RESEARCH DATA.
THEY MUST BE OPEN [AS OPEN AS POSSIBILE]

…and we have EOSC

EU WEB OF FAIR DATA AND SERVICES TO UNLOCK THE FULL POTENTIAL OF RESEARCH DATA

eosc — EOSC vision in a nutshell

What
EOSC is the European 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
- Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal'
- Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results
- Establish a sustainable and federated infrastructure enabling open sharing of scientific results

→ Strategic Research and Innovation agenda (SRIA) eosc.eu/sria-mar

REPowerEU

EOSC IS NOT A BIG BOX]

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

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

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

EOSC IS NOT A REPOSITORY NOR A «CLOUD»

AND GIVE SEAMLESS ACCESS TO 20 M EU RESEARCHERS

YOU DON'T «UPLOAD» YOUR DATA INTO EOSC

OBJECTIVES

EOSC SRIA 1.0

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

OPEN SCIENCE AS THE «NEW NORMAL»

[and we need data stewards]

nature — Feb. 25, 2020 — Subscribe

WORLD VIEW · 25 FEBRUARY 2020

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons

Change Agents

Learn how change agents, such as data stewards, play an important role

- A network of change agents coordinate data management across the

---

FAIR toolikt

HOW TO

Below are a set of questions designed to build and harness a network of change agents who support the change actively as an important facet of their daily work. They will be able to understand and communicate what is well and what requires attention. These questions are accompanied by example answers to illustrate how change agents relate to the implementation od sustained FAIR data management.

Q1. Who would you identify as key change agents?

- **Data steward:** Appointed to each important group who will be a senior scientist familiar with the concepts and process of data stewardship.
- **Middle managers** must support common data policies which can be reused.
- **Senior managers** must invest appropriate levels of budget for data management training, workshops and data service provision.
- **Support service staff** in Business Technology and Informatics functions are also likely to be important.

Q2. How can the change agents help to drive adoption of the change?

- The **change agents**, especially the **data stewards** supported by **management** need to facilitate new or improved business processes.
- These will foster the attitude that **data sets and corresponding metadata are valuable corporate assets** which must be managed effectively.

Q3. How can the network of change agents help to overcome barriers to change?

- The **network of data stewards** will facilitate implementation of FAIR data management at an optimal level of capability, determined through feasibility studies.
- Iterative application of FAIR maturity indicators will show opportunities for improvement, and the resulting benefits.
- Success will include more reuse of the data, better reproducibility and realisation of value from data and more time for insightful data analytics.

Use cases will show case such benefits. This will be important to communicate the value of the

---

DATA DOMAIN COMPETENCES+ TRANSVERSAL SKILLS

KØBENHAVNS UNIVERSITET

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 data internally and externally in the organisation.

Competence profile examples

- Domain-specific data understanding
- Ability to ensure that structured and unstructured data and meta data is modelled, harvested, stored, and maintained in a documented, and regulated fashion with focus on 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.

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

Why FAIR / Cost

2018

Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

The problem

Discoveries are made using shared data and this requires data that are:

- Retrievable and structured in standard format(s)
- Self-described so that third parties can make sense of it

*Data preparation* accounts for about 80% of the work of data scientists
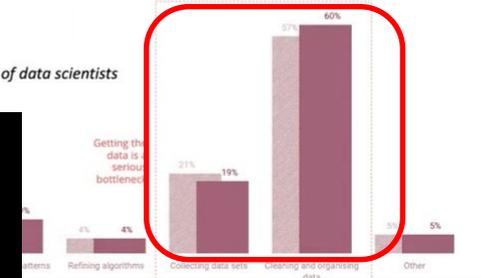
COSTS OF NOT HAVING FAIR DATA

ALMOST 80% TIME «WASTED» IN PREPARING THE DATA

Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

10,2 bn DIRECT
16   bn INDIRECT
26,2 bn TOTAL/YEAR

Why FAIR? / Kissed or missed?
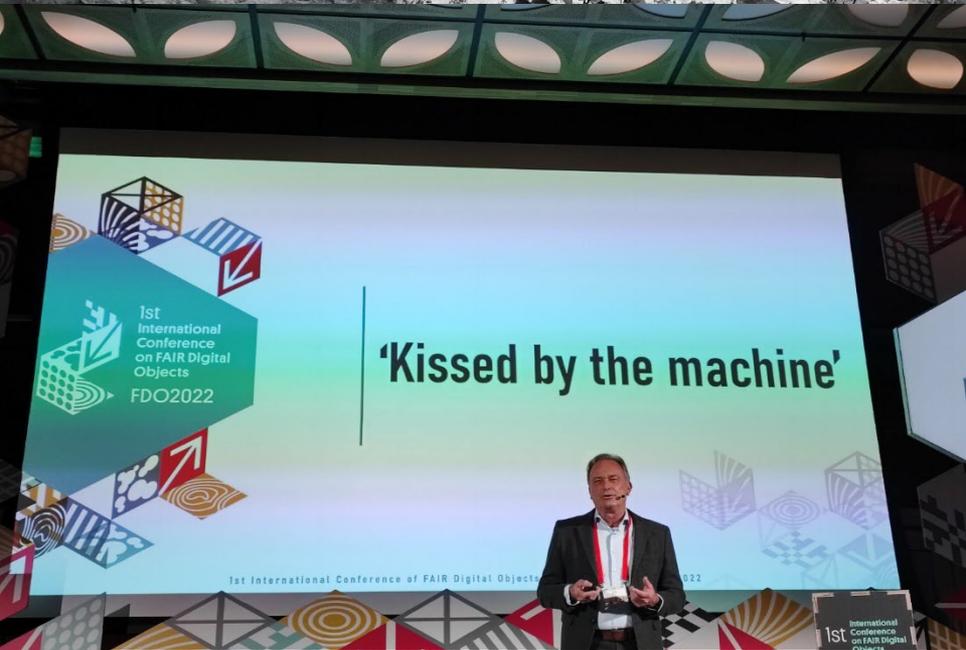
FAIR PRINCIPLES ARE
«MACHINE ACTIONABLE»
(MORE THAN READABLE)
FAIR = FULLY AI READY
IF NOT... YOU'LL BE MISSED (INSTEAD OF KISSED) BY THE MACHINE

Why FAIR? / AI needs good data

RDA 2022

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

Connie Clare, PhD

LEARNING
LEARNING LEARNING LEARNING

RDA **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)*

RDA **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 and whether datasets are fit for purpose.

ARTIFICIAL INTELLIGENCE
- WORKS IF DATA ARE GOOD
- THERE ARE ETHICAL ISSUES

Why FAIR? / to avoid garbage in-garbage out

GARBAGE IN, GARBAGE OUT:
THAT'S WHY WE NEED MACHINE-ACTIONABLE FAIR DATA!

At the end of the day, Artificial Intelligence is a powerful combination of *data* and *algorithms*. These AI algorithms are data-hungry. They require massive amounts of data to train themselves to do their intended job. And if they get bad data, the results are poor, too. *Garbage in, garbage out.*

FAIR MEANS ALSO QUALITY DATA, AS «R» INCLUDES DOCUMENTATION AND PROVENANCE

clearbox^AI   Product_ Pricing_ Use Cases_ About_ Resources_

AI Apocalypse: What you really need to be afraid of

By Shalini Kurapati   2023

Data is Food for AI

	~1% of AI research?	~99% of AI research?
	80%	20%
	PREP	**ACTION**
	Source and prepare high quality ingredients	Cook a meal
	Source and prepare high quality data	Train a model

Slide courtesy of Shalini Kurapati
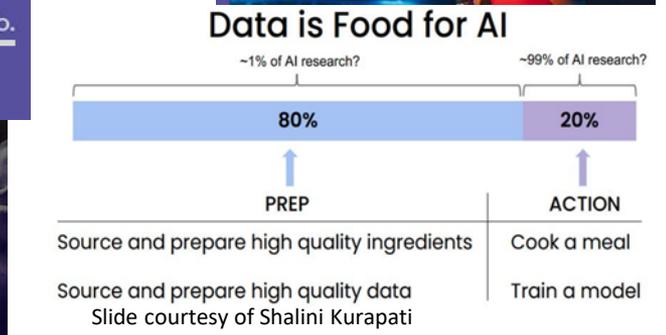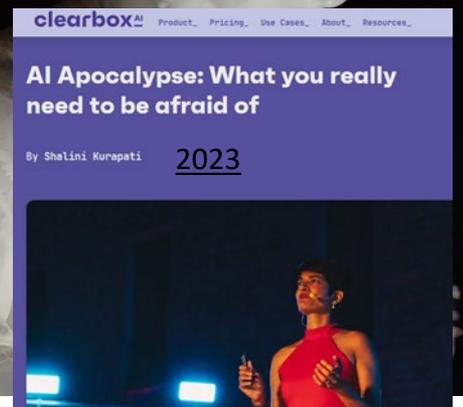
clearbox^AI

I dati tra presente e futuro nell'AI

Shalini Kurapati

How FAIR / 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 proce 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 th 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 Com

How / 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

Authors | Authors and affiliations

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

Social Science Survey Research_V1

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

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

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

How / FAIRification mapping

Nanomaterials

The mapping of existing initiatives to increase the FAIRness of both nanomaterials and mixture toxicity datasets and computational approaches for toxicity and mixture assessment is a critical step towards identifying both the domain- specific features and the general features needed to maximise data and model FAIRness.

Building on this mapping, and the development of a FIP, the case study will foster development and piloting of interoperability standards and guidelines for increasing FAIRness in the interlinked scientific disciplines (chemical toxicity, nanomaterials toxicity and characterisation, risk assessment, advanced materials, environmental science), and across the different domains.

https://worldfair-project.eu/nanomaterials/

FIP
NANOMATERIALS

Nanomaterials domain-specific FAIRification mapping (Deliverable 4.1)   2023

How / FAIR connect

https://fairconnect.pro/

Articles

- Data Stewardship Plan templates designed to support the FAIR principles
- Hourglass-based interoperability through nanopublications in VODAN-A
- A 10 step checklist for starting FAIR discussions in your community: Call for contributions
- FAIR scientific information with the Open Research Knowledge Graph
- FIP2DMP: Linking data management plans with FAIR implementation profiles
- The FAIR hourglass: A framework for FAIR implementation

All articles

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.

Try our new Nanopublication Search engine

See our new Dashboard

JOIN OUR COMMUNITY

FAIR Connect   SEARCH   DASHBOARD

Search FAIR Nanopublications

| FAIR-Enabling Resources ⌄ | Search for FERs, FIPs or FICs... | SEARCH |

FAIR-Enabling Resources
FAIR Implementation Profiles
FAIR Implementation Communities

How / FAIR cookbook

F∧IRCOOKBOOK

GITHUB | FAIRcookbook

The FAIR Cookbook
for FAIR doers

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

F∧IRCOOKBOOK

GITHUB

RECIPES

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

1. Unique, persistent identifiers

Introducing unique, persistent identifiers

Recipe Overview

Reading Time
30 minutes

Executable Code
No

Difficulty

Recipe Type
Background information

Audience
Principal Investigator, Data Manager, Data Scientist

Maturity Level & Indicator
DSM-1-C0

Cite me with FCB006

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

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

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

GITHUB | Search Wizard...

A Accessibility

EXEMPLAR RECIPES
- Transferring data with SFTP
- Downloading data with Aspera

I Interoperability

EXEMPLAR RECIPES
- Selecting terminologies and ontologies
- Creating a metadata profile

R Reusability

EXEMPLAR RECIPES
- Data licenses
- Declaring data's permitted uses

LEARN MORE

LEARN MORE

LEARN MORE

LEARN MORE

Infrastructure

LEARN MORE

Assessments

LEARN MORE

Applied Examples

LEARN MORE

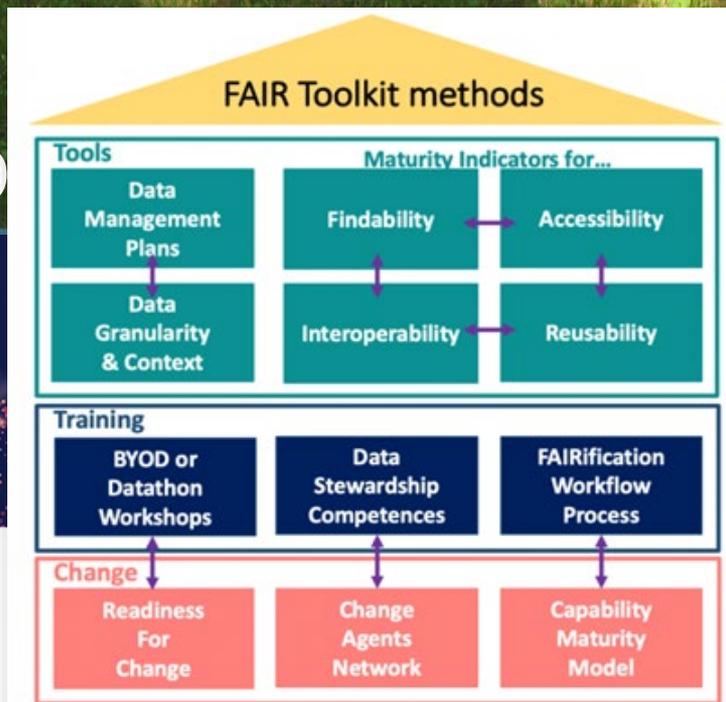Maturity model

LEARN MORE

How / FAIR too

Practical Support for FAIR Data

An overview of how the FAIR Toolkit provides practical support for implementation of FAIR data management through numerous use cases from industry and relevant tools, training and change methods.

FAIR Toolkit

Practical Support

The FAIR Toolkit is designed to provide support for management of the FAIR data life cycle as illustrated in Figure 1 below. It places emphasis on the practical aspects of FAIR data management through the leverage of existing resources that are most relevant to the needs of Life Science industry.

FAIR Toolkit
1. Use Cases
2. Tool methods
3. Training methods
4. Change management

SUPPORTS

FAIR Data Life Cycle Management

Data/metadata Generation
Data Ingestion
Data Processing & Storage
Data Integration
Archive
Finding & Data Application

FAIR Toolkit methods

Tools

| Data Management Plans |
| Data Granularity & Context |

Maturity Indicators for...

| Findability | Accessibility |
| Interoperability | Reusability |

Training

| BYOD or Datathon Workshops | Data Stewardship Competences | FAIRification Workflow Process |

Change

| Readiness For Change | Change Agents Network | Capability Maturity Model |

PISTOIA ALLIANCE FAIR TOOLKIT: TOOLS, CASE STUDIES, TRAINING

TOOLS

Data Management Plans

A Data Management Plan documents the specific attributes expected for your FAIR objectives.

- Prepare the Data Management Plan as early as possible

Find out more >

Findability Maturity Indicators

Find out how to apply the FAIR Maturity Indicators to measure the FINDABILITY of the data and metadata.

- Findability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

Find out more >

Accessibility Maturity Indicators

Learn how to apply the FAIR Maturity Indicators to measure the ACCESSIBILITY of the data and metadata.

- Accessibility of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

Find out more >

Data Granularity and Context

Consider how the granularity and context of data and associated metadata to help to inform your FAIR objectives.

- Understand the granularity and context of the data as early as possible

Find out more >

Interoperability Maturity Indicators

Read how to apply the FAIR Maturity Indicators to measure the INTEROPERABILITY of the data and metadata.

- Interoperability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

Find out more >

Reusability Maturity Indicators

Discover how to apply the FAIR Maturity Indicators to measure the REUSABILITY of the data and metadata.

- Reusability of data is compared with your FAIR objectives to identify and make improvements in an iterative manner

Find out more >

How / CEDAR

https://metadatacenter.org/

CEDAR | CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL

PURPOSE   RESEARCH   TOOLS | TRAINING
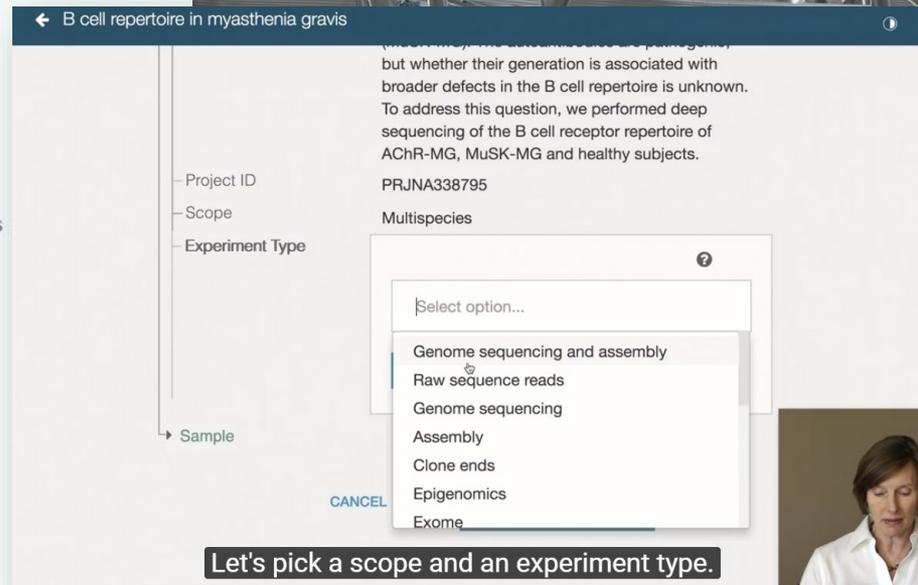
Better data for better science

What CEDAR does

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

What can CEDAR do for me already?

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

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

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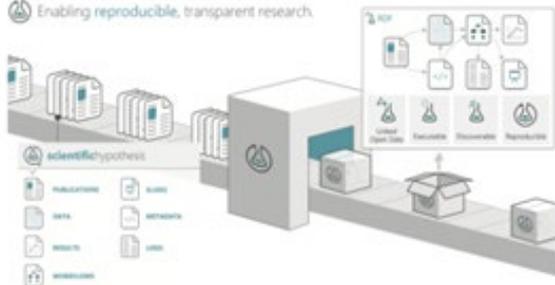
← B cell repertoire in myasthenia gravis

(AChR-MG). The autoantibodies are pathogenic, but whether their generation is associated with broader defects in the B cell repertoire is unknown. To address this question, we performed deep sequencing of the B cell receptor repertoire of AChR-MG, MuSK-MG and healthy subjects.

Project ID   PRJNA338795

Scope   Multispecies

Experiment Type

Select option...

Genome sequencing and assembly
Raw sequence reads
Genome sequencing
Assembly
Clone ends
Epigenomics
Exome

CANCEL

Sample

Let's pick a scope and an experiment type.

How / RO Crate

How / FAIR signposting

How / [FAIRsharing. To be interoperable]

https://fairsharing.org/

FAIRSHARING [NEW VERSION] STANDARD REGISTRY

…FAIR for institutions

DO I-PASS FOR FAIR?

Policy: Does your organization (institute / university (of applied sciences)) have a FAIR research data policy?

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

Services: Does your organization have a (virtual) DCC which provides services (including infrastructure) to allow comply with FAIR?

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

Looking forward to the next 10 years of FAIR

News from Leiden

THE ROAD TO FAIR AND EQUITABLE SCIENCE – LOOKING FORWARD TO THE NEXT 10 YEARS

- **Day 1 (Monday) Workshop introduction: Reflection and looking ahead** - The first day is devoted to reflection (the previous 10 years) and broadly looking ahead (the next 10 years), and will provide participants with a clear agenda and goal for the week. The goal of Day 1 will be to set up the context for the following 3 days where both success and obstacles can be systematically documented. Concluding remarks that summarize the day will be formulated and shared. The day ends with the workshop reception.
- **Day 2 (Tuesday) Focus area 1: Machine actionability** - Day 2 will focus on machine actionability, a central but often dismissed objective of FAIR. The clear role for FAIR Digital Objects and FAIR Digital Twins of real world entities will be highlighted. Concluding remarks that summarize the day will be formulated and shared.
- **Day 3 (Wednesday) Focus area 2: Equitability** - Day 3 will address how FAIR can be used to promote *fair*, in the sense of equal opportunities for (citizen) scientists and innovators around the globe and how we can prevent that FAIR inadvertently contributes to further widening of the digital divide not only between the Global South and North, but also between the public and private sectors. Concluding remarks that summarize the day will be formulated and shared.
- **Day 4 (Thursday) Focus area 3: Fully AI Ready** - The FAIR acronym is now also frequently interpreted in the sense of data and services that are Fully AI Ready, that is, as the substrate for data-intensive science. The recent rise and release of Large Language Models (LLMs) feeding Chatbots has severely aggravated the already difficult problem of misinformation and deliberate disinformation, not only in social media context but also in science. This day will be co-lead by CODATA and the International Science Council. Focus will be on how trustworthy information with full, traceable provenance can play a role. Concluding remarks that summarize the day will be formulated and shared.
- **Day 5 (Friday) Workshop conclusion: Looking forward to the next 10 years** - Based on the successful precedent of the 2014 workshop, we will invite major stakeholders from the public and private sectors to review the successes, obstacles and recommendations logged the previous 4 days. Together, we aim for a final conclusions and recommendations document as well as a concrete decadal roadmap of desired outcomes, potentially published in a leading scientific journal comparable to the first seminal paper. Day 5 will thus contribute to the overall aim to come to a set of commonly agreed upon recommendations and a plan for FAIR and Equitable Science and Innovation in the coming decade.

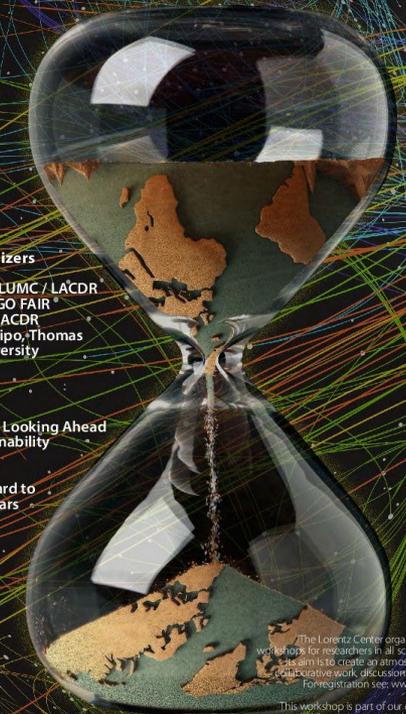**The Road to FAIR and Equitable Science**
22 - 26 January 2024, Leiden, the Netherlands

Scientific Organizers
- Barend Mons, LUMC / LACDR
- Erik Schultes, GO FAIR Foundation / LACDR
- Francisca Oladipo, Thomas Adewumi University

Topics
- Reflection and Looking Ahead
- Machine Actionability
- Equitability
- Fully AI Ready
- Looking Forward to the Next 10 Years

w w w . l o r e n t z c e n t e r . n l

A recap

DAY 1
Monday

Convergence

Video recap on Youtube
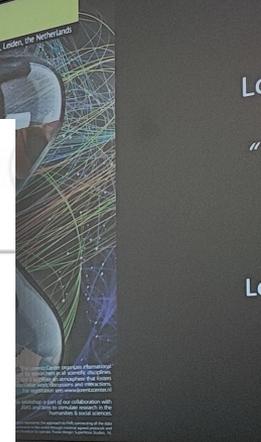
Key thoughts

DATA «VISITING» INSTEAD OF SHARING.
DATA DO NOT MOVE
(VODAN AFRICA)

EMBED «FAIR» INTO A WIDER «RESEARCH CULTURE CHANGE» PROGRAMME

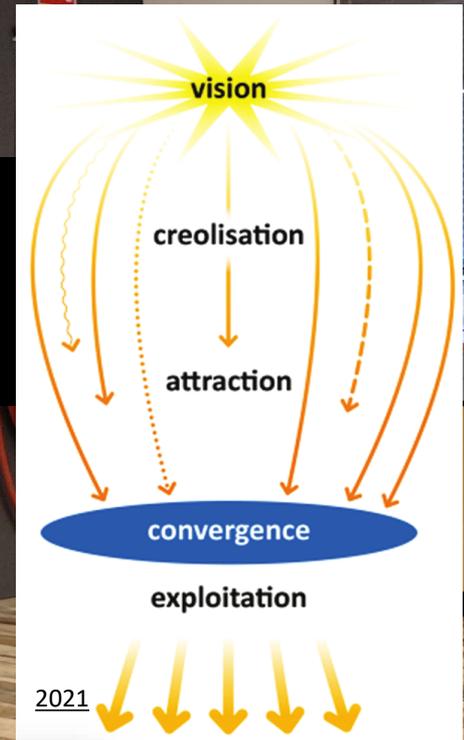WE ARE STILL IN THE «CREOLIZATION» PHASE

GREAT UPTAKE OF FAIR BY INDUSTRY

vision

creolisation

attraction

convergence

exploitation

2021

Equity for FAIR ⟷ Equity through FAIR

The relationship between FAIR and equity is bidirectional.

Two dimensions:

1. **Equity for FAIR:** FAIR is better when it is fair and all stakeholders benefit - for the highest benefit, we need implement it in an equitable way

2. **Equity through FAIR:** The implementation of FAIR can also help us ensure that broader issues of equity are addressed in relation to scientific practice and the use of data

AI to achieve FAIR

Artificial Intelligence can assist with data stewardship, automating some processes to help improve the FAIRness of data and metadata.

- Improve ontology development (text mining, logical constraints).
- Automate metadata capture and maintenance.
- Make recommendations (for metadata, ontologies, knowledge models) based on use.

FAIR to improve AI

1. **Avoiding garbage in, garbage out (afval in, afval uit):** FAIR can help with ensuring the quality and usability of data that goes into AI models; and can help constrain outputs.

2. FAIR can help with **transparency, reproducibility** of the AI processes,

3. FAIR can greatly assist **accuracy**, minimising hallucinations and avoiding inequitable outcomes

4. FAIR for software, including AI models, helps with better **software development** and **management practices.**

5. FAIR can help ensure **proper use of digital objects** to include copyright, intellectual property, and licensing.

Where are we?

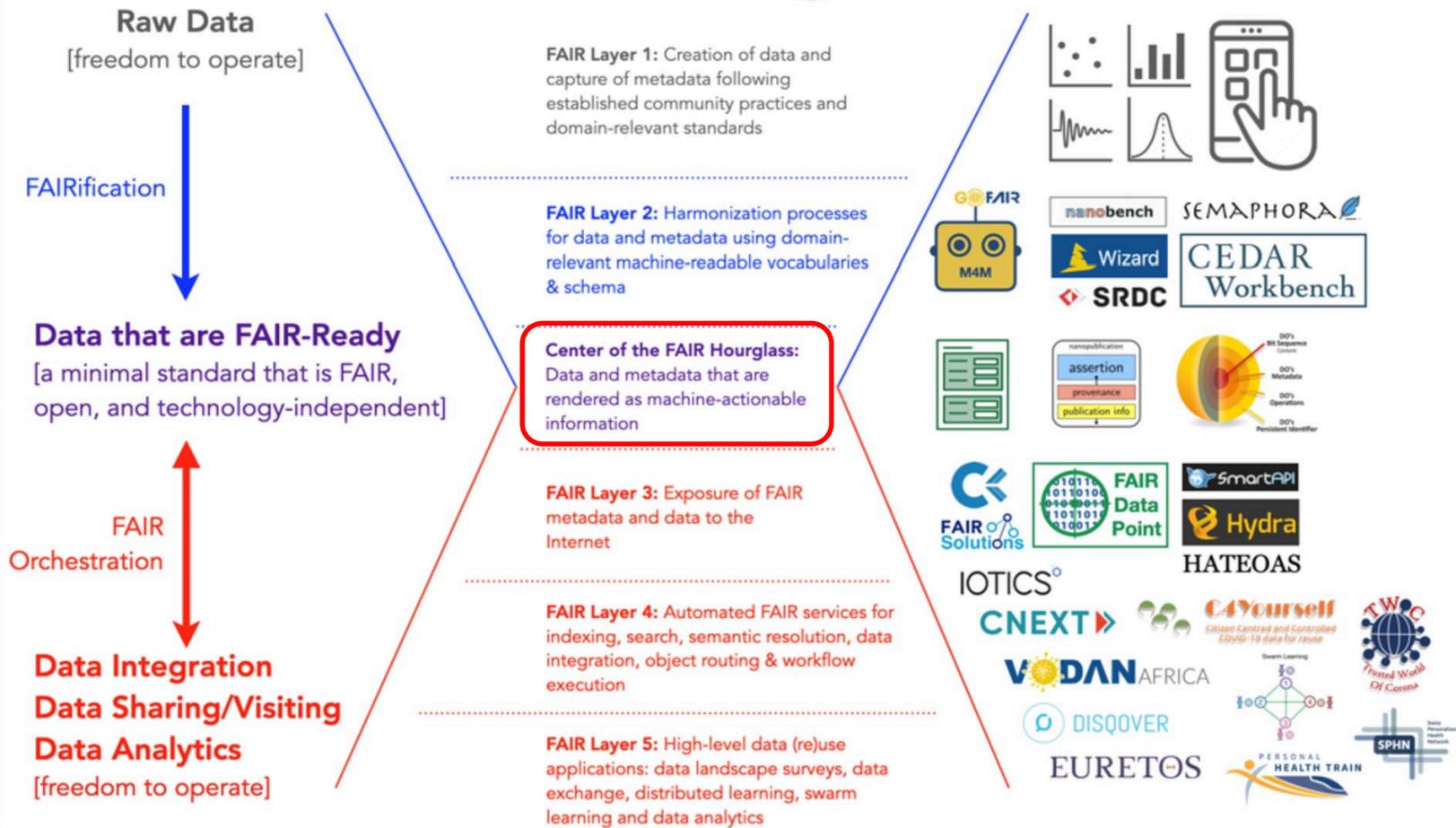The FAIR hourglass: A framework for FAIR implementation

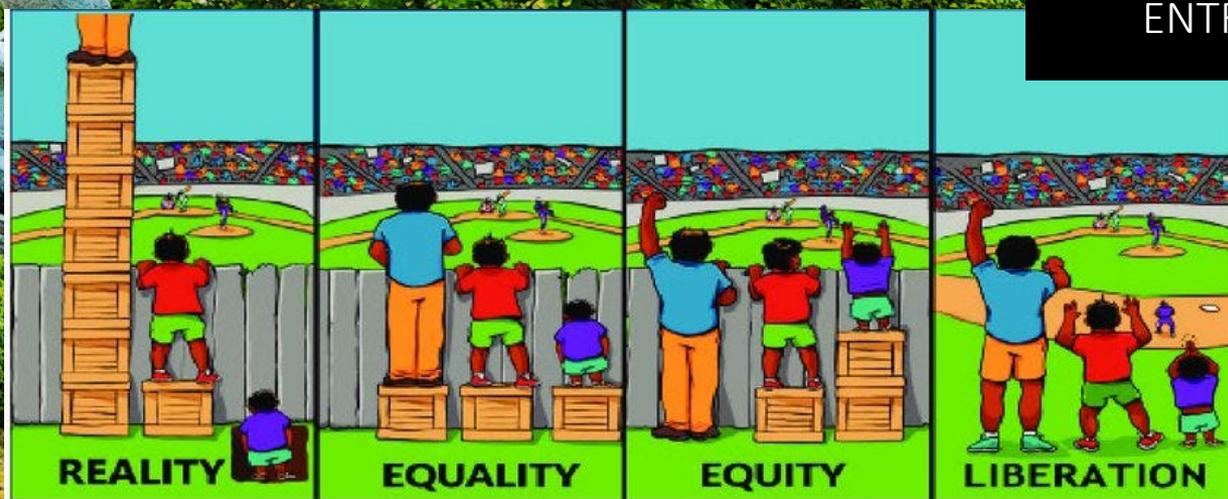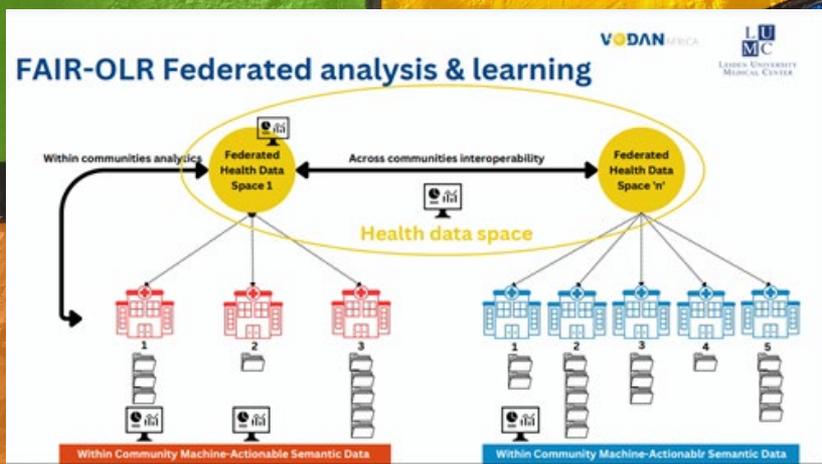Article type: Research Article   2022

Authors: Schultes, Erik

Key thoughts / 2

OUR GOAL SHOULD BE «LOWER THE FENCE», CREATE EQUITY AT THE ENTRY LEVEL

GO FOR EQUITY (AT LEAST!)

[lessons from VODAN Afr

FAIR-OLR Federated analysis & learning

VODAN AFRICA — LUMC LEIDEN UNIVERSITY MEDICAL CENTER

Within communities analytics — Across communities interoperability

Federated Health Data Space 1 ↔ Federated Health Data Space 'n'

Health data space

Within Community Machine-Actionable Semantic Data | Within Community Machine-Actionablr Semantic Data

1.1. Ownership

Ownership is a critical principle from which access and control arrangements are derived [6,10]. In VODAN-Africa the data is stored in residence, governed by data use agreements, and compliant with the regulations identified in each locality [14]. In multiple jurisdictions, data-ownership is federated, and requires data-localisation, meaning the repositioning of machine-actionable meta-data held in the repository where the data is produced. The precise definition of data-ownership may differ in the different regulatory frameworks in place for different situations. The ultimate ownership of personal health record resides with the data subject and/or the health facility where the data is produced who consent for data (re)-use. When the data is aggregated – without inclusion of personal information – into statistical data elements, ownership can be transferred. The main aspect regarding ownership of the data principle is that personal data pertains to the data subject and is stewarded by the data producer with consent of the data subject, and this is a basic non-negotiable principle in VODAN-Africa.

Federated FAIR principles: Ownership, localisation and regulatory compliance (OLR)

Cite

2023

OWNERSHIP, LOCALISATION, REGULATORY COMPLIANCE

Article type: Research Article

Authors: van Reisen, Mirjam[a; b; *] | Amare, Samson Yohannes[a; c] | Nalugala, Reginald[d] | Taye, Getu Tadele[e] | Gebreselassie, Tesfit Gebremeskil[e] | Medhanyie, Araya Abrha[e] | Schultes, Erik[f] | Mpezamihigo, Mohamed[g]
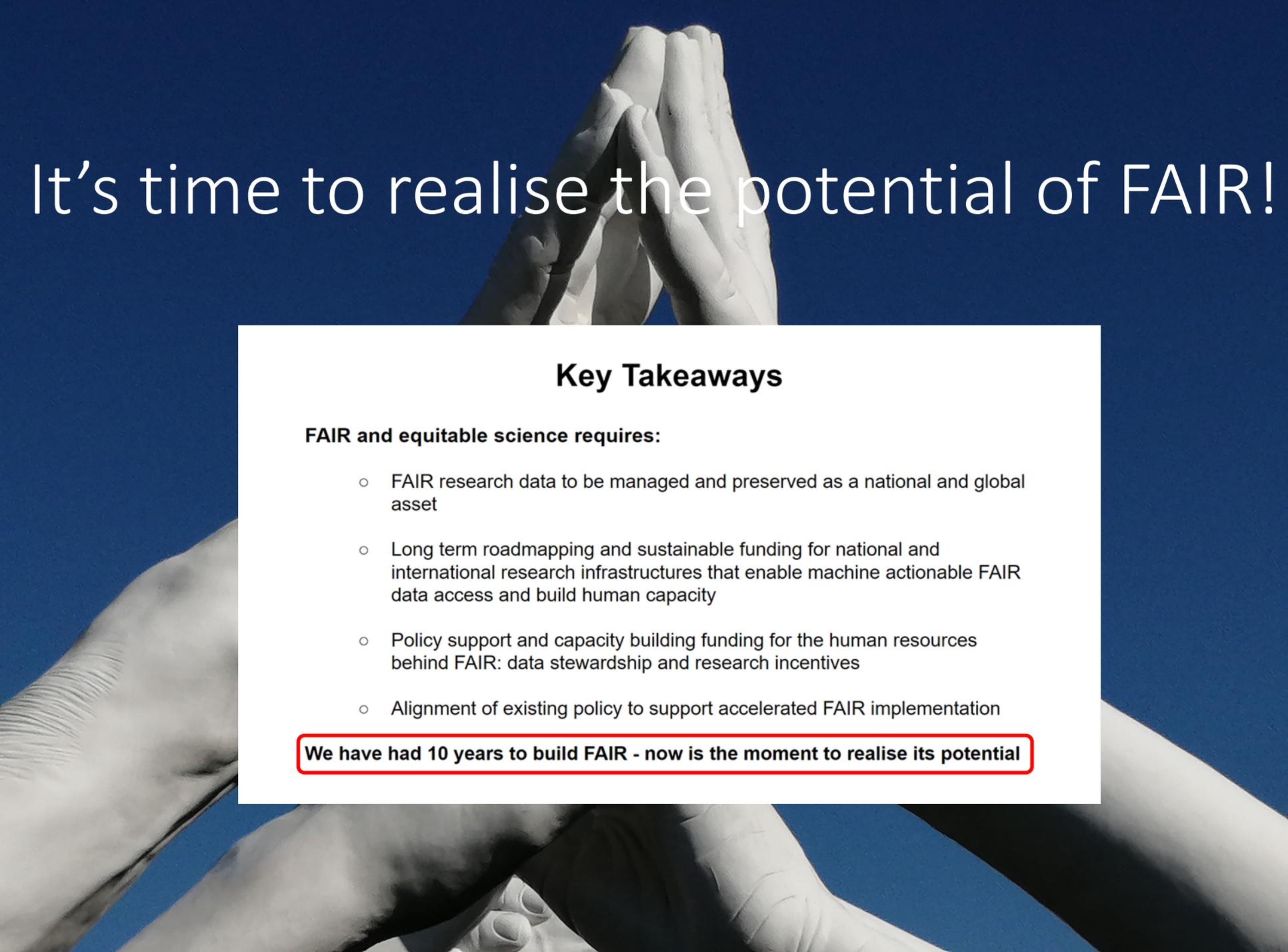
1.2. Localisation

Data Localisation is necessary if the standard of data Ownership is to be met. VODAN-A tested the possibility of repositing data in each of the participating health clinics. The data is entered only once and stored in AllegroGraph, or a triple store for semantic linked data, from where it is available for statistical analysis in the local data dashboard. In addition, horizontal interoperability is arranged across health facilities in the community [17]. The VODAN-A dashboard serves to visualise health population trends. Vertical interoperability with other platforms is supported through interoperability schemas [17]. Specific functions, such as an automated computation in a Health Information Systems (HIS) or reporting into District Health Information System 2 (DHIS2) are arranged within the architecture deployed in residence.

The VODAN-A research found that a 'locale' for data localisation can differ across different geographies. In some locations, depending on availability of resources, a client-server deployment architecture was followed. In a client-server set-up, the MVP is deployed in a

1.3. Regulatory compliance in the jurisdiction

Through a federated data-visiting set up, VODAN-A achieved compliance with data according to the regulatory framework in place in each jurisdiction [1]. Due to the sensitive nature of personal health data, full attention must be given to the rights and obligations under the regulatory frameworks in each jurisdiction in which the data is produced, in

It's time to realise the potential of FAIR!

Key Takeaways

FAIR and equitable science requires:

- FAIR research data to be managed and preserved as a national and global asset

- Long term roadmapping and sustainable funding for national and international research infrastructures that enable machine actionable FAIR data access and build human capacity

- Policy support and capacity building funding for the human resources behind FAIR: data stewardship and research incentives

- Alignment of existing policy to support accelerated FAIR implementation

We have had 10 years to build FAIR - now is the moment to realise its potential

Reasons NOT to go FAIR/Open?

Valid reasons not to participate in open science practices

Casper J. Albers[*]

Abstract

The past years have seen a sharp increase in the attention for open science practices. Such practices include pre-registration and registered reports, sharing of materials, open access publishing and attention to reproducibility of research. Despite the overwhelming amount of evidence highlighting the benefits of open science, some researchers remain reluctant. In this paper, I will outline valid reasons for researchers not to participate in open science practices.

Discussion

There are no valid reasons.

THANKS!

2018   [*]Heymans Institute for Psychological Research, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. c.j.albers@rug.nl