FAIR data: from FAIRytale to FAIR enough

Ingrid Dillo

DANS

Building Bridges across Open Science in the UK

Joint RDA UK and Open AIRE Virtual Workshop

06-05-2020

Building bridges across Open Science in the UK
6 May 2020 - 12:30 UTC

'Just as the Netherlands is a knowledge society, science is a knowledge community. Both benefit from the free exchange of information. For that reason alone you have my fullest support in your quest for Open Science.'

Ingrid van Engelshoven, minister of Education, Culture and Science

Open Science
free exchange of information
for the good of science and society at large

We are now building a European Open Science Cloud. It is a trusted space for researchers to store data and access data from researchers from other disciplines."

Ursula von der Leyen
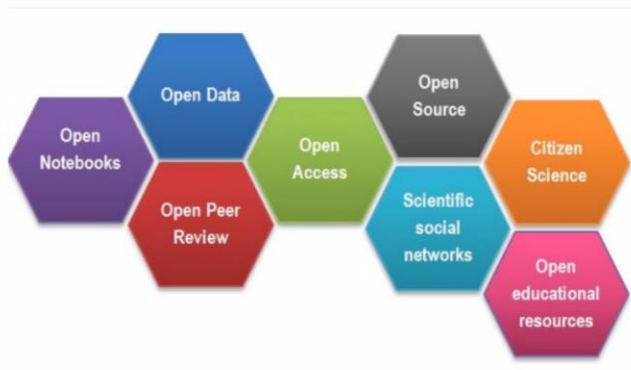European Commission President
World Economic Forum 2020

MAKING OPEN SCIENCE A REALITY

OECD

Components of Open Science

Open science

Open science is an umbrella term for transparent science with ease of access to all products from beginning to end
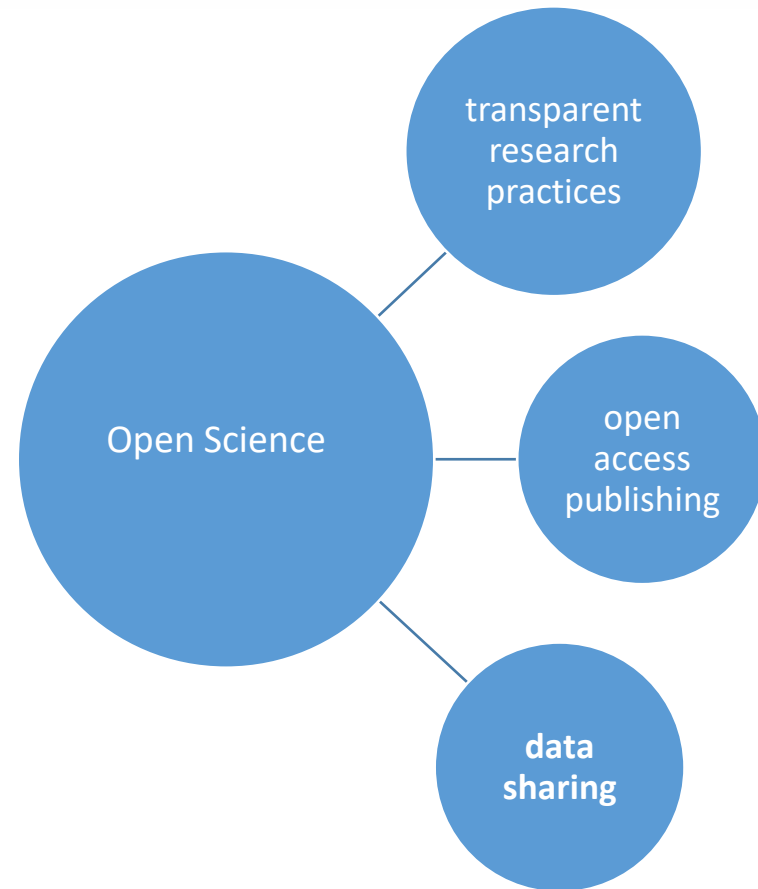
Image credit: Gema Bueno de la Fuente by CC-BY

Open Science

transparent research practices

open access publishing

data sharing

Why data sharing is important

Replication and validation of research outcomes
(scientific integrity and transparency)

nature International weekly journal of science

News

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

The New York Times

Fraud Case Seen as a Red Flag for Psychology Research

By BENEDICT CAREY
Published: November 2, 2011

A well-known psychologist in the Netherlands whose work has been published widely in professional journals falsified data and made up entire experiments, an investigating committee has found. Experts say the case exposes deep flaws in the way science is done in a field, psychology, that has only recently earned a fragile respectability.

ScienceInsider Breaking news and analysis from the world of science policy

Report: Dutch 'Lord of the Data' Forged Dozens of Studies (UPDATE)

by Gretchen Vogel on 31 October 2011, 7:05 PM | 34 Comments

SPIEGEL ONLINE

Niederlande

Renommierter Psychologe gesteht Fälschungen

theguardian

Public sector can combat fraud with data sharing

Outsourcing is not the only thing to blame for procurement fr... says **Graham Kemp**, and the public sector needs to view da... less as a security risk but knowledge to be shared

PLoS one accelerating the publication of peer-reviewed science

Home   Browse Articles   About   For Readers   For Authors and Reviewers

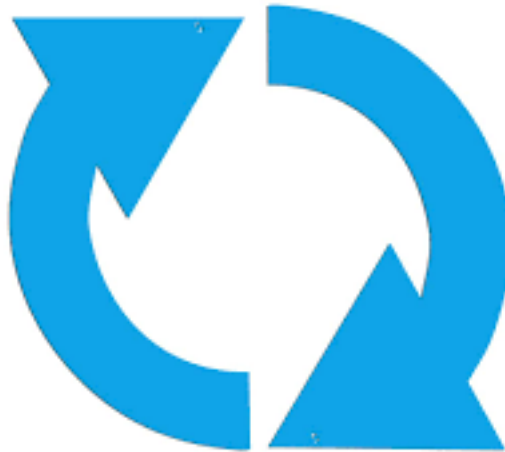RESEARCH ARTICLE   OPEN ACCESS

How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data
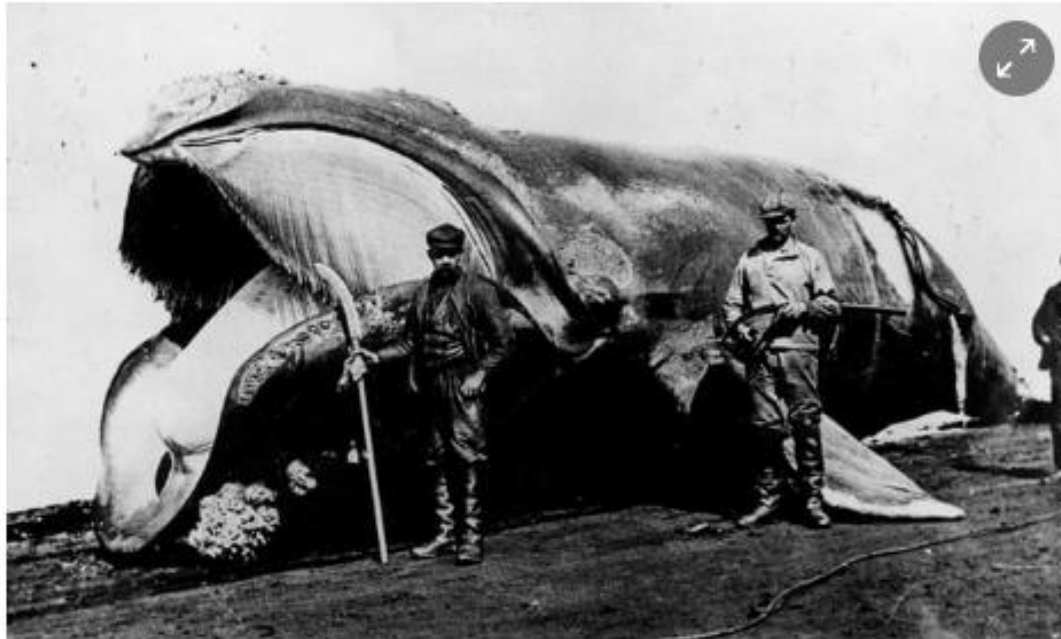
Why data sharing is important

Re-use of data (efficiency, return on investment, standing on the shoulders of others)

The 19th-century whaling logbooks that could help scientists understand climate change

The public are being enlisted to read through detailed logs of whaling ships which include records of ice flows and weather conditions

Hunters with their catch, circa 1900. Whalers' logbooks are providing detailed information about the climate and ice patterns many years ago. Photograph: Hulton Archive/Getty Images

Maritime historians, climate scientists and ordinary citizens are coming together on a project to study the logbooks of 19th-century whaling ships to better understand modern-day climate change and Arctic weather patterns.

https://www.theguardian.com/environment/2015/dec/17/the-19th-century-whaling-logbooks-that-could-help-scientists-understand-climate-change

https://www.covid19dataportal.org
https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0
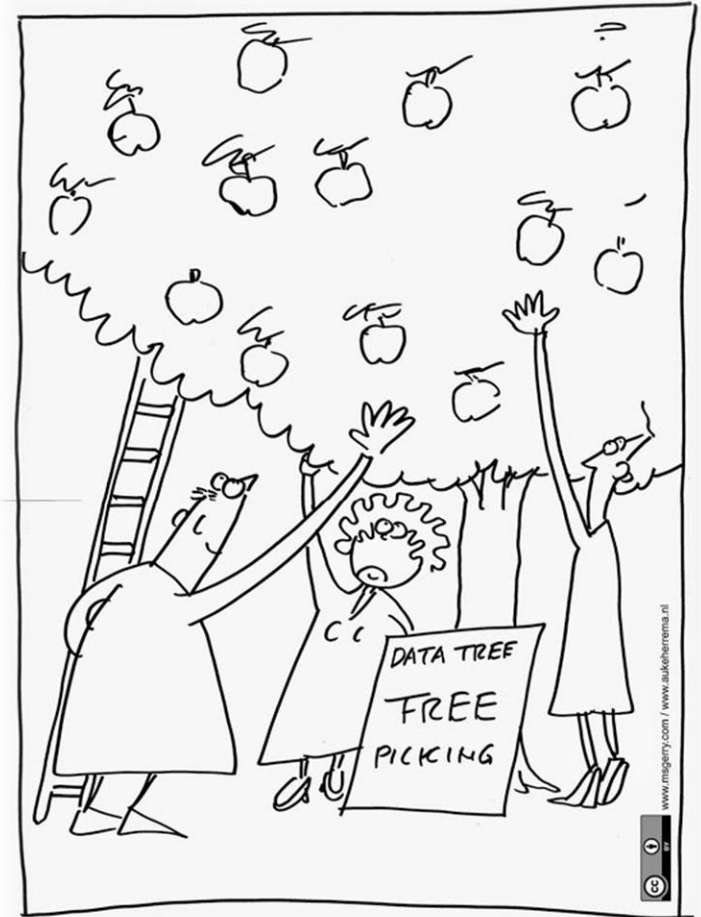
Enabling data sharing

1. Culture change
   - Influence of sharing norms within direct research circle
   - Academic rewards for data sharing

   - External drivers:
     - Funder policies
     - Publishers requirements (DAPs)

2. RDM training and support for researchers

3. Infrastructure: VREs, trustworthy data repositories

FAIR Data Principles (2016)

During the 2014 workshop "Designing a data FAIRport" for the life sciences in Leiden a minimal set of community-agreed guiding principles were formulated.
The FAIR Data Principles:

SCIENTIFIC DATA

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons ✉ - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

110k Accesses | **1458** Citations | **1512** Altmetric | Metrics

- Easy to find by both humans and machines based on metadata

- With well-defined use license and access conditions (Open Access if possible)

- Ready to be linked with other datasets

- Ready to be re-used for future research and to be processed further using computational methods and tools

https://www.nature.com/articles/sdata201618

FAIR metrics

Box 2 | The FAIR Guiding Principles

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

15 Criteria

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

See: http://datafairport.org/fair-principles-living-document-menu and
https://www.force11.org/group/fairgroup/fairprinciples

Different implementations of FAIR data

Creation

Requirements for new data creation (funders)

H2020 Programme

Guidelines on
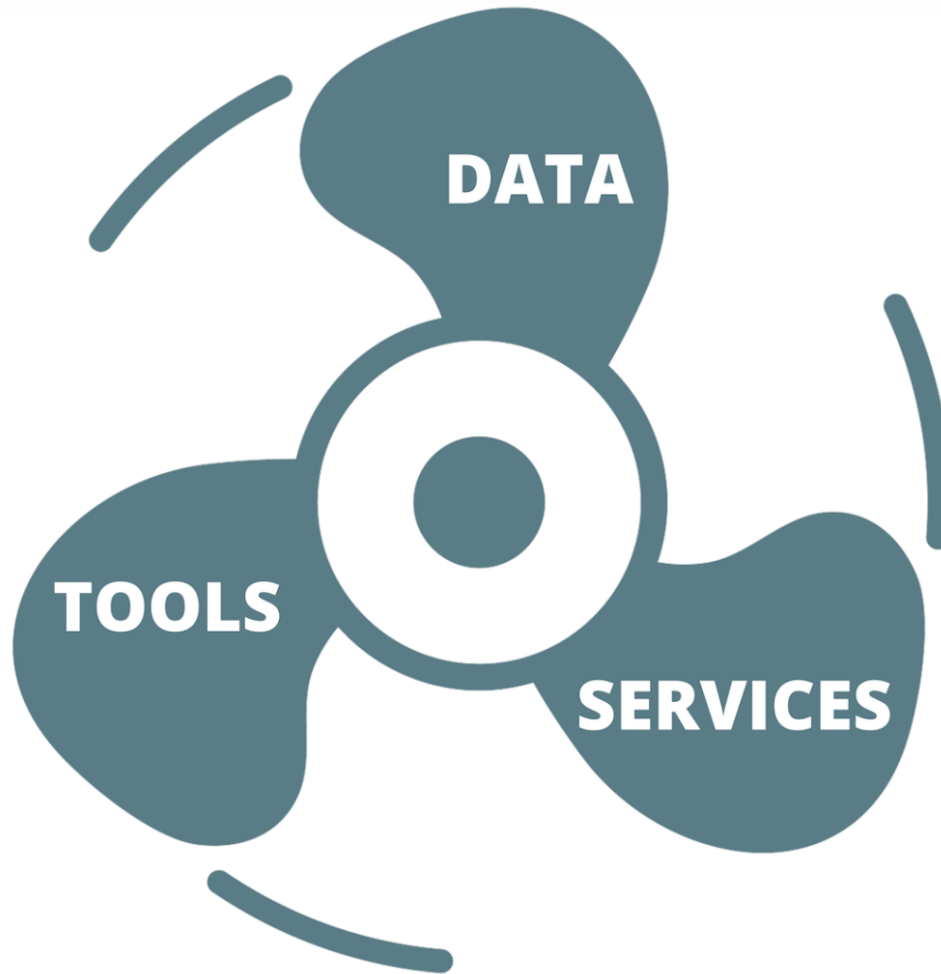FAIR Data Management in Horizon 2020

RISK ALERT

Assessment

FAIR

Establishing the profile for existing data (data repositories)

Transformation

NOT FAIR → FAIR

Transformation tools to make
data FAIR (Go-FAIR initiative)

The internet of FAIR data and services

Europe and FAIR

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

TURNING FAIR INTO REALITY

2018

https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

Key Points: To make FAIR a reality …

- Report takes a holistic approach, not a data centric approach
- Need to address the enabling practices and technologies – not just focus on the data and its attributes
- Need to consider all digital outputs (data, code, metadata etc)
- Objective is to make data and other digital research outputs FAIR for humans and machines.
- Needs: concept of FAIR digital objects, FAIR ecosystem, interoperability frameworks for disciplines and across disciplines, FAIR services including trusted digital repositories, skills, metrics and sustainable funding.

FAIRsFAIR project

Fostering Fair Data Practices in Europe:

- practical solutions for the use of the FAIR data principles throughout the research data life cycle;

- fostering FAIR data culture and the uptake of good practices in making data FAIR;

- Ensuring that research resources such as data, software and services can be rendered and maintained FAIR is FAIRsFAIR's contribution to the EOSC.

FAIRSFAIR IN ACTION

- IMPROVE INTEROPERABILITY OF FAIR RESOURCES
- INCREASE PRODUCTION AND USE OF FAIR DATA
- BUILD A NETWORK OF TRUSTED DIGITAL REPOSITORIES
- SET UP A FAIR COMPETENCE CENTRE FOR ALL COMMUNITIES
- DEVELOP A CAPABILITY MATURITY MODEL TOWARDS FAIR CERTIFICATION
- EMBED FAIR DATA EDUCATION IN UNIVERSITY PROGRAMMES
- ORGANISE AN OPEN CALL FOR REPOSITORIES TO GET SUPPORT FOR CERTIFICATION
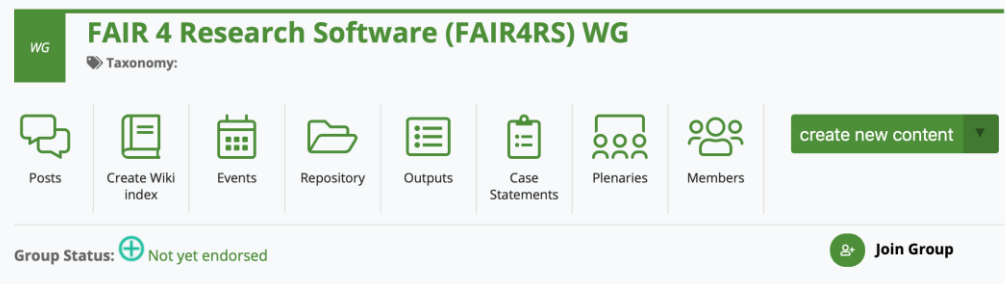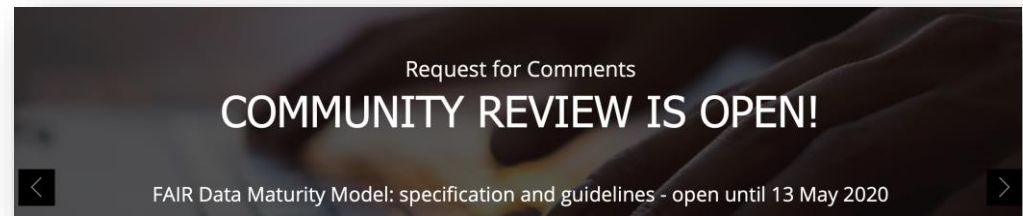
https://www.fairsfair.eu

FAIR in RDA

FAIR Data Maturity Model WG

The indicators developed to answer the question 'What needs to be measured to assess the FAIRness of a digital object' and not 'How to measure the FAIRness of a digital object'.

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines

Request for Comments
COMMUNITY REVIEW IS OPEN!
FAIR Data Maturity Model: specification and guidelines - open until 13 May 2020

FAIR 4 Research Software (FAIR4RS) WG
Taxonomy:

Posts | Create Wiki index | Events | Repository | Outputs | Case Statements | Plenaries | Members

create new content

Group Status: Not yet endorsed

Join Group

https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg

https://fair-software.nl

FIVE RECOMMENDATIONS FOR FAIR SOFTWARE
ENDORSE
LET'S GO! →

Everybody loves FAIR!

Everybody wants to be FAIR, but:
- what does that mean?
- how to put the principles into practice?

The concept of FAIR data: what does it really mean?

Responsible RDM
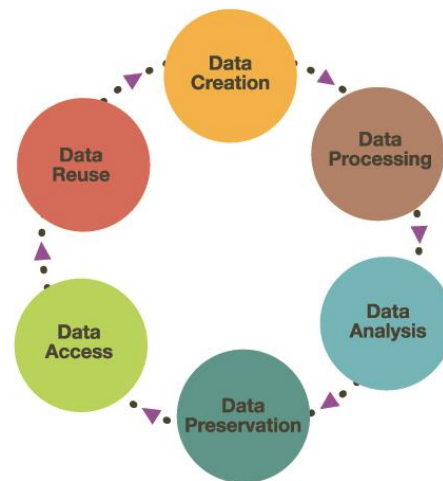
"**Perhaps the biggest challenge in sharing data is trust: how do you create a system robust enough for scientists to trust that, if they share, their data won't be lost, garbled, stolen or misused?**"

The Data Harvest:
How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report
December 2014

Data sharing a FAIRytale?

"Research data will not become nor stay FAIR by magic. We need skilled people, transparent processes, interoperable

technologies and collaboration to build, operate and maintain research data infrastructures."
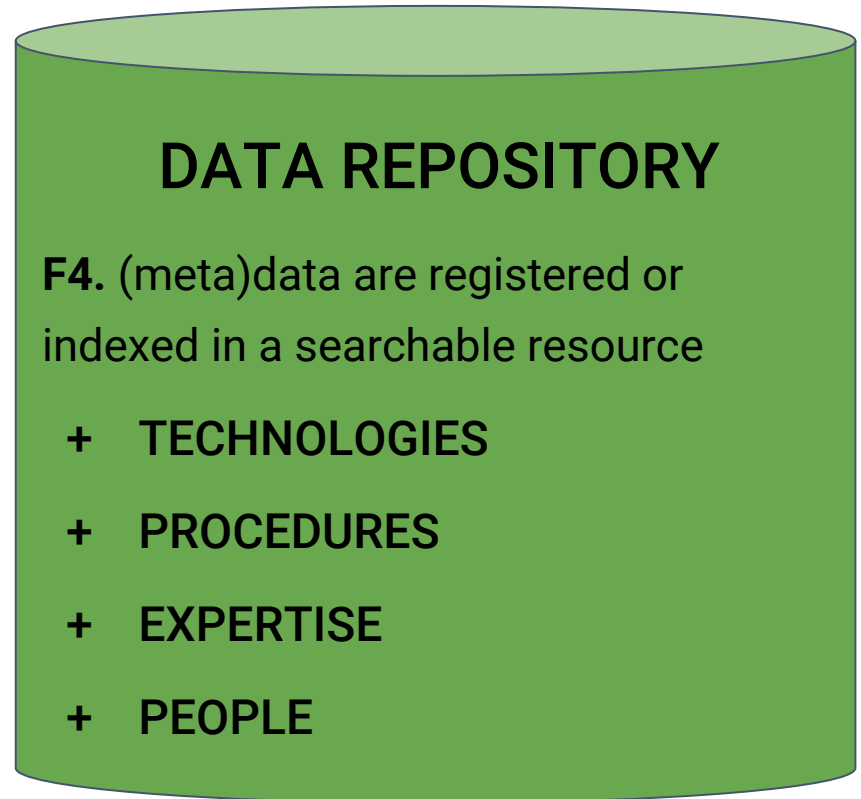
Mari Kleemola,

CoreTrustSeal Board

https://tietoarkistoblogi.blogspot.com/2018/11/being-trustworthy-and-fair.html

FAIR data assessment: levels

(META)DATA

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

F2. data are described with rich metadata

F3. metadata clearly and explicitly include the identifier of the data it describes

DATA REPOSITORY

F4. (meta)data are registered or indexed in a searchable resource

+ TECHNOLOGIES

+ PROCEDURES

+ EXPERTISE

+ PEOPLE

TRUST

Where to store your data?

https://www.re3data.org

Trusting digital repositories

- actions and attributes of the trustee (integrity, transparency, competence, predictability, guarantees, positive intentions)

- external acknowledgements:
  - reputation (researchers)
  - third party endorsements (funders, publishers)

CoreTrustSeal certification

- Community driven repository certification standard

- Developed under the umbrella of RDA

- 16 requirements (organizational infrastructure, digital object management, technology and security)

- Peer review, 3 year cycle, transparent processes

- Global uptake, discipline agnostic

https://www.coretrustseal.org

TDR to guarantee baseline data FAIRness

- Majority of CoreTrustSeal requirements (indirectly) refer to the FAIRness of the repository holdings

- Baseline of data FAIRness, but:

- Some data will be more FAIR than others!

FAIR and TRUST principles

https://www.nature.com/sdata/

www.nature.com/scientificdata

OPEN **The TRUST Principles for digital repositories**

COMMENT

Box 1 The TRUST Principles

Principle	Guidance for repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Takeaway message

- We need to share our data in order to turn open science into a reality;

- The FAIR principles help us to define high quality and transparent research data management practices;

- The TRUST principles and certification mechanisms, like CoreTrustSeal for digital repositories, help us to create trust in the research data infrastructure we need to safeguard the accessibility and assessibility of our (FAIR) data for the future.

Thank you for listening!

ingrid.dillo@dans.knaw.nl