From my data to FAIR data

Benjamin Pfeil
Senior Advisor
Norwegian School of Economics (NHH), & Norwegian Research Centre AS (NORCE)

NARMA Conference
Lillestrøm, 21.03.2023

Professional background in Research Data Management

2001 with various data centres

2007 international committee work (OECD, GEO, IOC UNESCO, IAEA)

2014 established the Bjerknes Climate Data Centre (BCDC)

2016 working with European Research Infrastructures (ICOS, EMSO) and national research infrastructures Norwegian Marine Data Centre and Infrastructure for Norwegian Earth System modelling

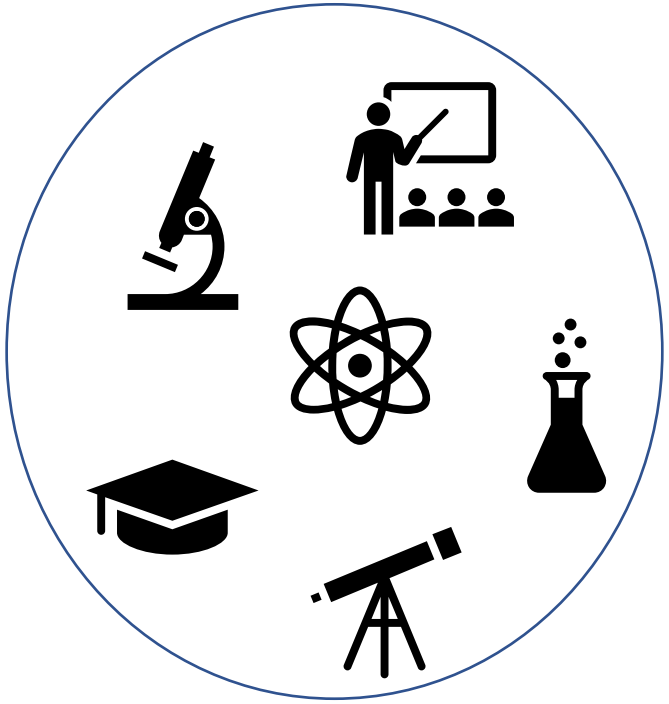
2018 Copernicus Marine Service delivering data towards space agencies

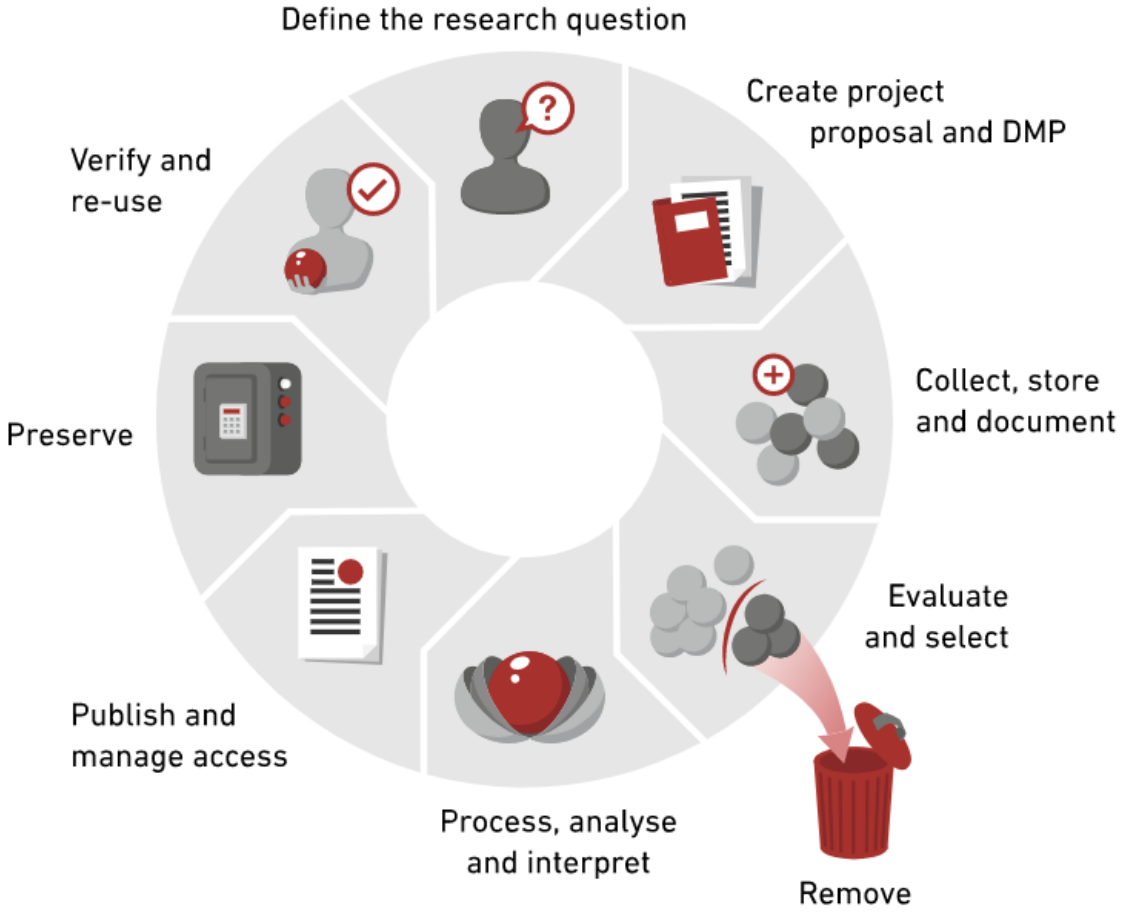
2021 International Oceanographic Data and Information Exchange (IODE) of the International Oceanographic Commission of UNESCO – BCDC become a member

2022 new position at NHH as a Senior Advisor for Open Science and Research Data and Senior Advisor for NORCE

NHH

My Data → **FAIR Data**





Findable

Accessible

Interoperable

Reusable

[Research data life cycle – ETH Library | ETH Zurich]

[Common Data Elements: Standardizing Data Collection (nih.gov)]

SCIENTIFIC DATA

NHH

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

- Urgent need to **improve the infrastructure** supporting the re-use of scholarly data
- A diverse set of stakeholders designed and endorsed a concise and measureable set of principles – the **FAIR Data Principles**
- Guideline for those wishing to enhance the reusability of their data
- 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 individual)

Scientific Data **3**, Article number: 160018 (2016) | Download Citation

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

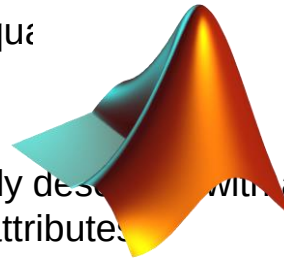
To be Accessible:

- 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 f broadly applicable langua representation.
- I2. (meta)data use vocabu principles
- I3. (meta)data include qua (meta)data

Interoperability = streaming

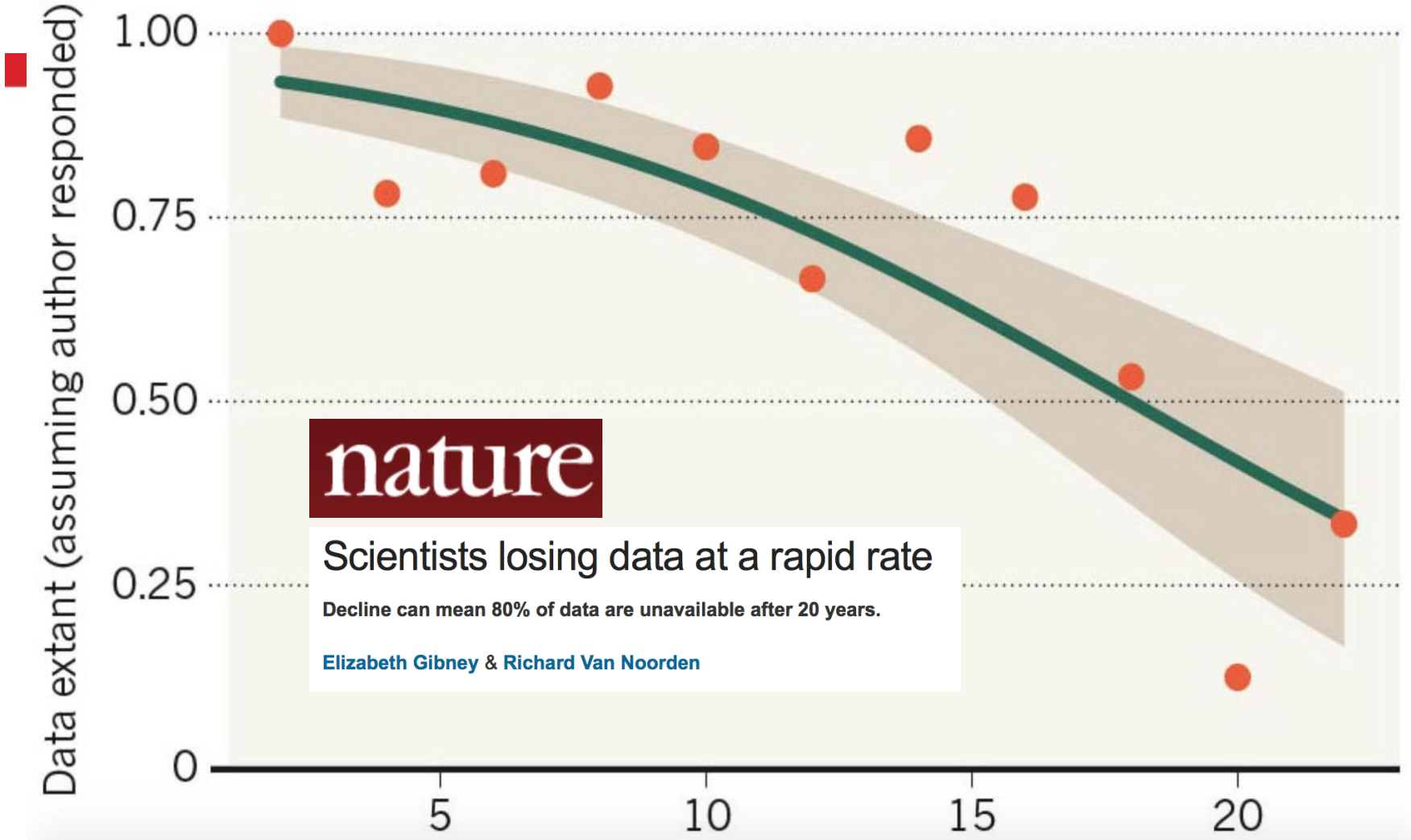


To be Reusable:

- R1. meta(data) are richly des with 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

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



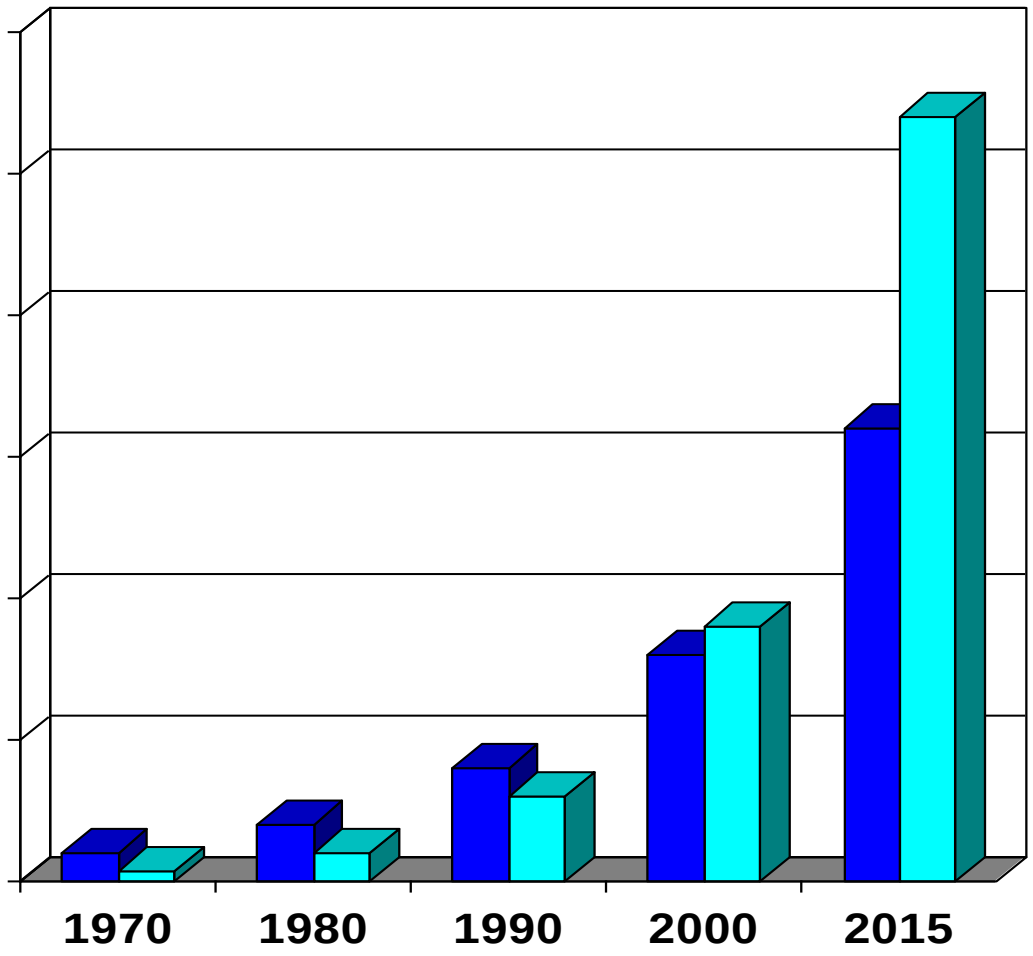
nature

Scientists losing data at a rapid rate

Decline can mean 80% of data are unavailable after 20 years.

Elizabeth Gibney & Richard Van Noorden

Global increase in publications in empirical sciences



newsblog

Nature brings you breaking news from the world of science

News & Comment | News blog Archive | Post

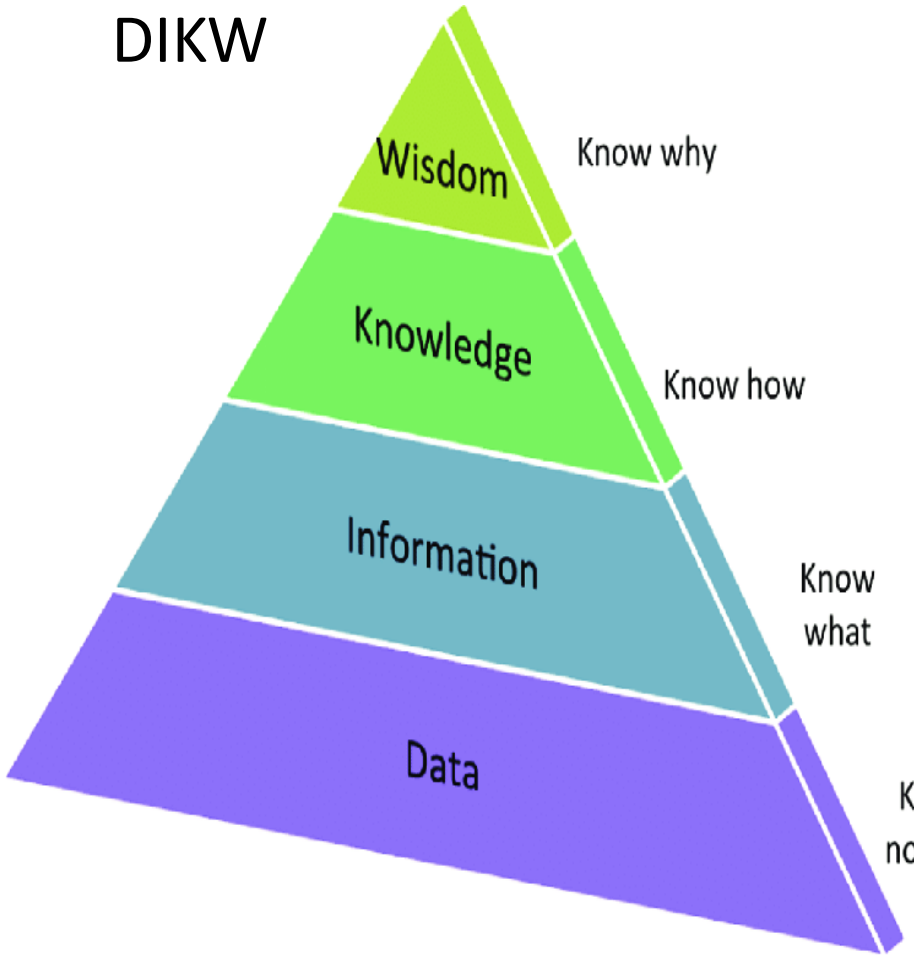
NEWS BLOG

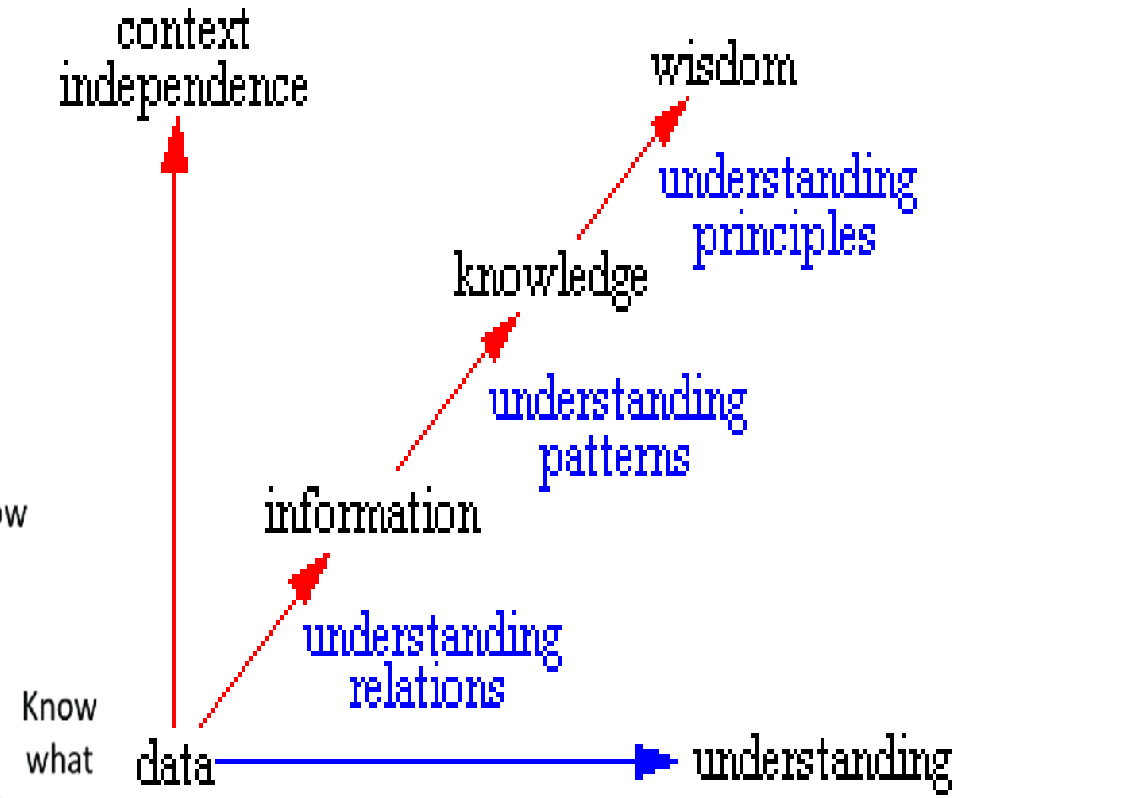
Global scientific output doubles every nine years

07 May 2014 | 16:46 BST | Posted by Richard Van Noorden | Category: Policy, Publishing

NEWS BLOG

DIKW





'Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation'
Wilkinson et al., Sci. Data, 2016

Compiled based on Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3-9. and Zeleny, M. (1987). Management support systems: towards integrated knowledge management. *Human Systems Management*, *7*(1), 59-70.

NHH

SUSTAINABLE DEVELOPMENT GOALS

17 GOALS TO TRA

aws

Contact Sales Support English My Account Create an AWS Account

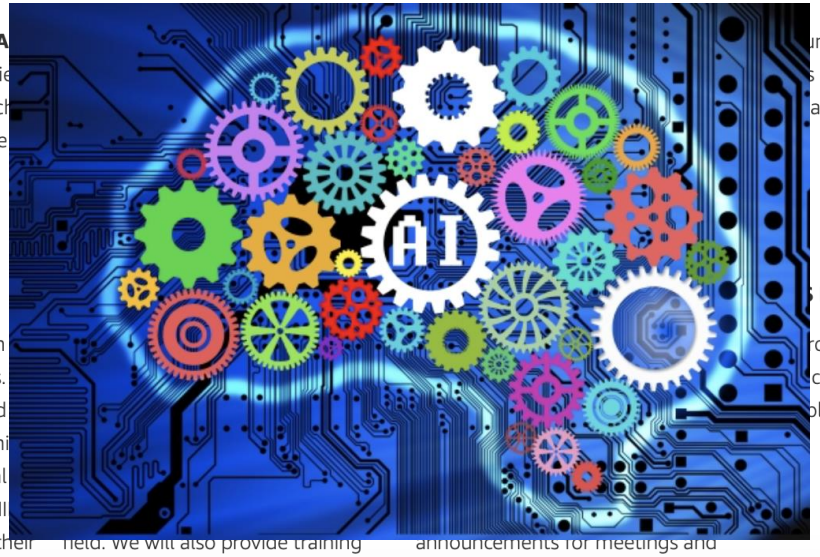
Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

AWS Research Initiatives (ARI)

AWS and federal agencies collaborate to foster innovation

Contact Us

The **AWS Research Initiatives (A** Institutes of Health (NIH) in the fie such as artificial intelligence, mach to help researchers accelerate the



ndation (NSF) and National s of technologies and platforms ance computing, ARI awards aim

Awards

Awards consists of a combination federal funds and AWS resources. Researchers from universities and colleges, nonprofits, non-academi organizations, and state and local government can apply via NSF/NI solicitations, requesting AWS as their cloud for research.

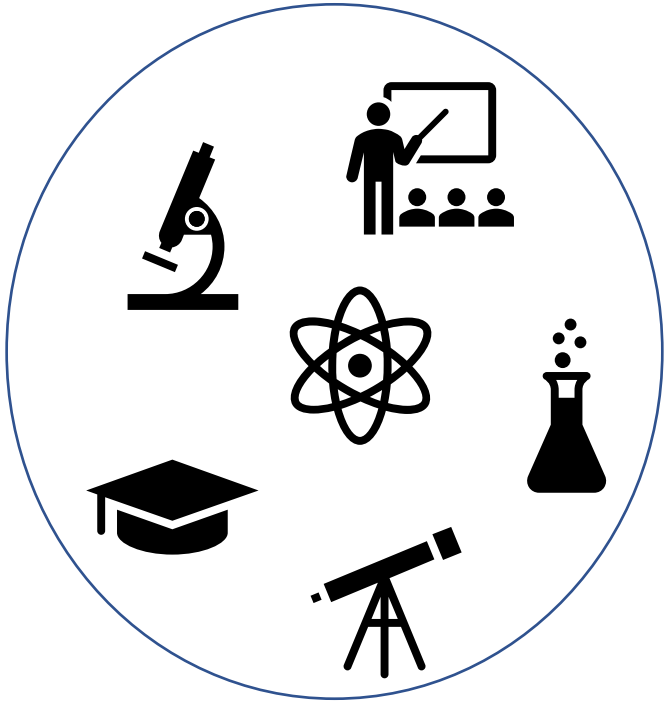
field. We will also provide training resources, expertise and content

announcements for meetings and workshops.

Research Credits

rds include AWS research credits can be redeemed toward le AWS services.

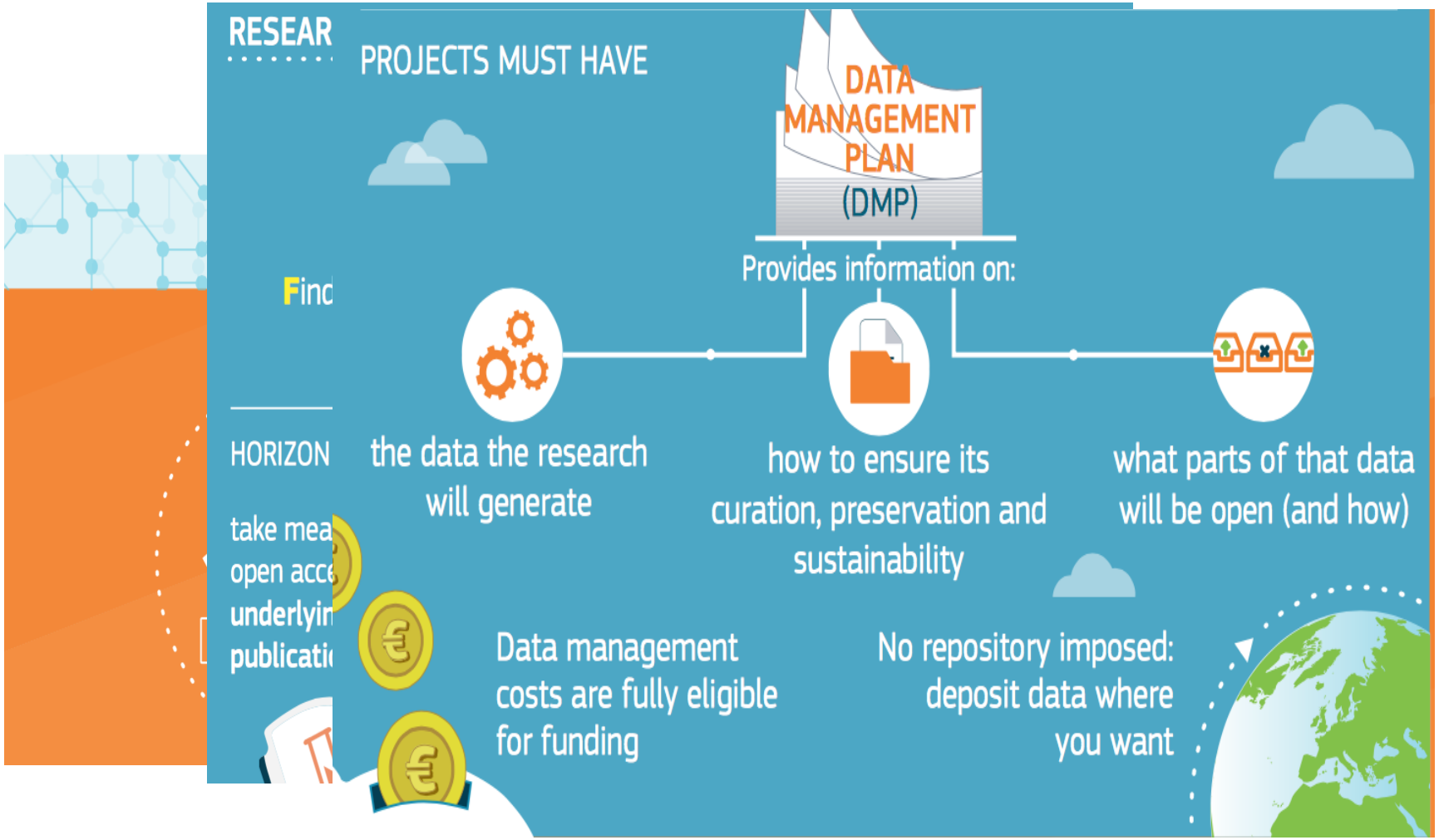
My Data → FAIR Data



Research data life cycle – ETH Library | ETH Zurich

Common Data Elements: Standardizing Data Collection (nih.gov)

Data management is mandatory for all EU and (most) national funded projects



FAIR Guiding Principles

- have rapidly been adopted by publishers, funders, and pan-disciplinary infrastructure programmes and societies
- present guidelines for the publication of digital resources
- are aspirational, in that they do not strictly define how to achieve a state of “FAIRness”
- describe a continuum of features, attributes, and behaviors that will move a digital resource closer to that goal
- this led to a wide range of interpretations of FAIRness
- a number of incompatible methodologies to assess FAIRness have been developed already

Source: Wilkinson, M. D. *et al*. A design framework and exemplar metrics for FAIRness. *Sci. Data* 5:180118 doi: 10.1038/sdata.2018.118 (2018).

SCIENTIFIC DATA

Comment | OPEN | Published: 26 June 2018

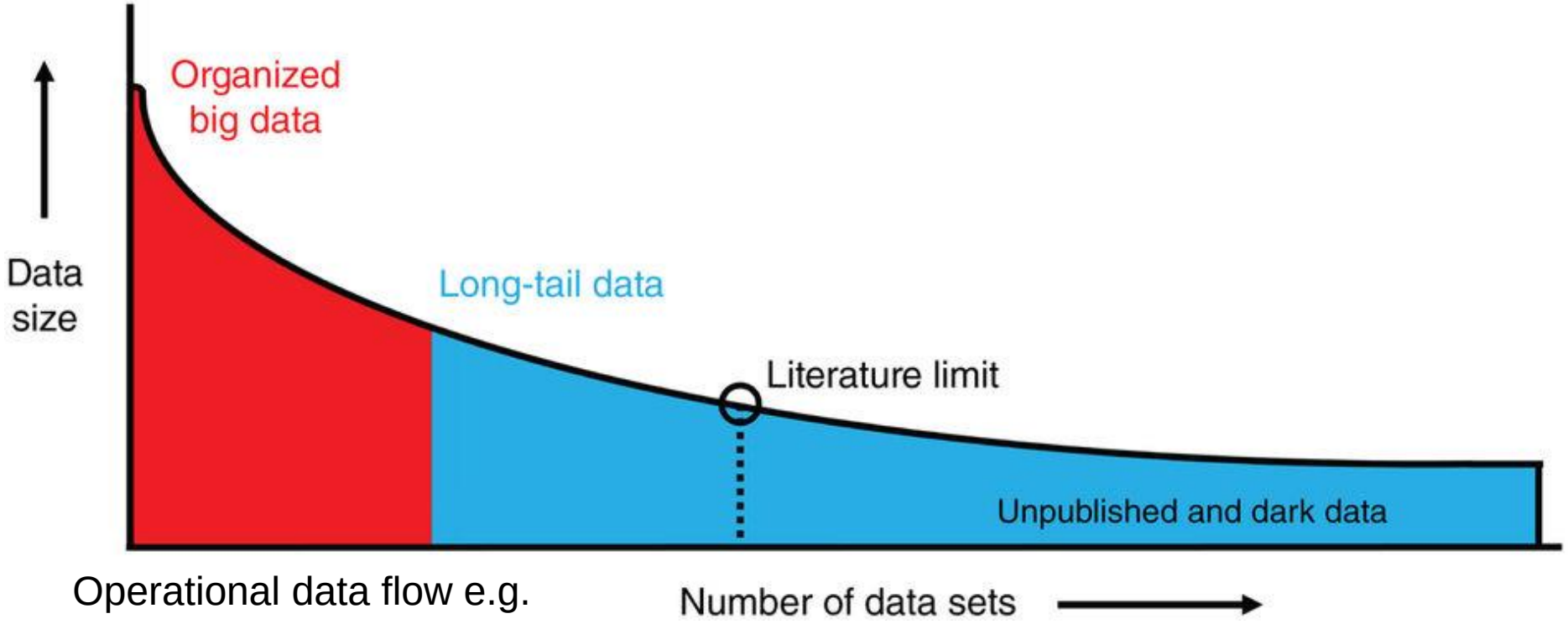
A design framework and exemplar metrics for FAIRness

'It would focus on FAIRness for machines – i.e., the degree to which a digital resource is findable, accessible, interoperable, and re-usable without human intervention'

This was because FAIRness for people would be difficult to measure objectively, as it would often depend on the experience and prior-knowledge of the individual attempting to find and access the data.

Principles have rapidly been adopted by publishers, funders, and pan-disciplinary infrastructure programmes and societies. The Principles are aspirational, in that they do not strictly define how to achieve a state of "FAIRness", but rather they describe a continuum of features,

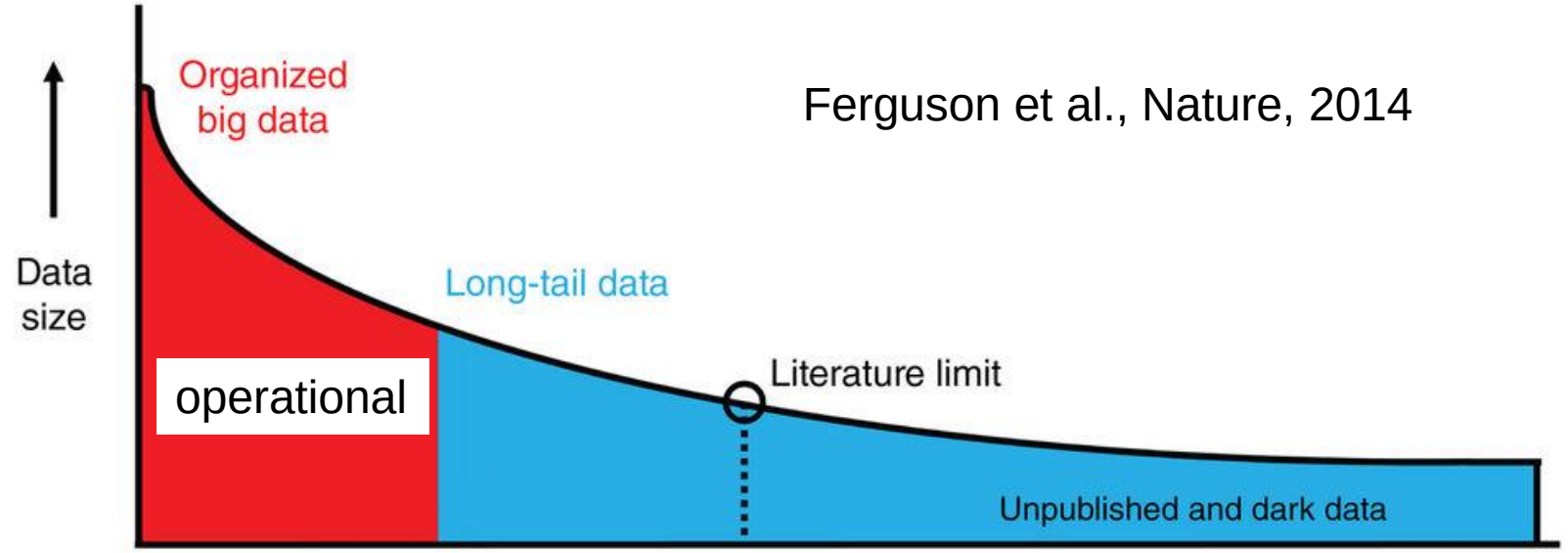
Classification of scientific data



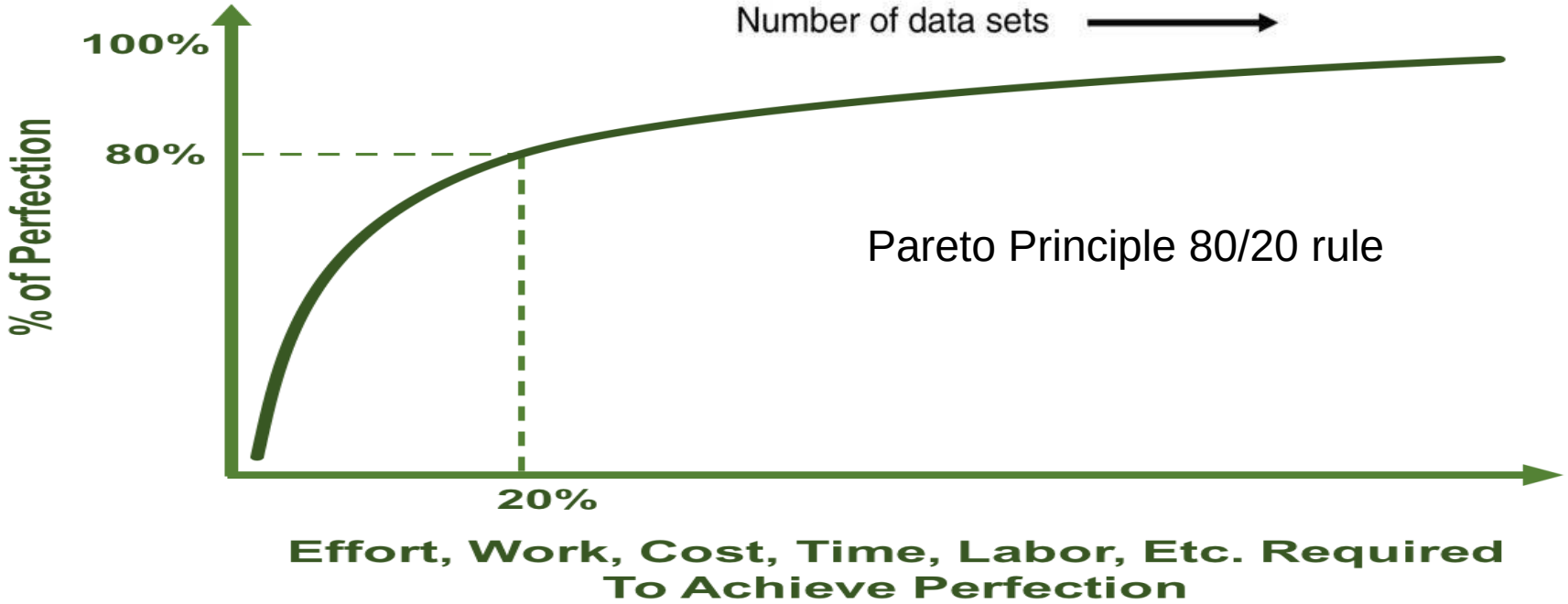
Operational data flow e.g.
Research Infrastructures,
model output, remote sensing

(Ferguson et al., 2014, Nature)

Challenge of handling scientific data:

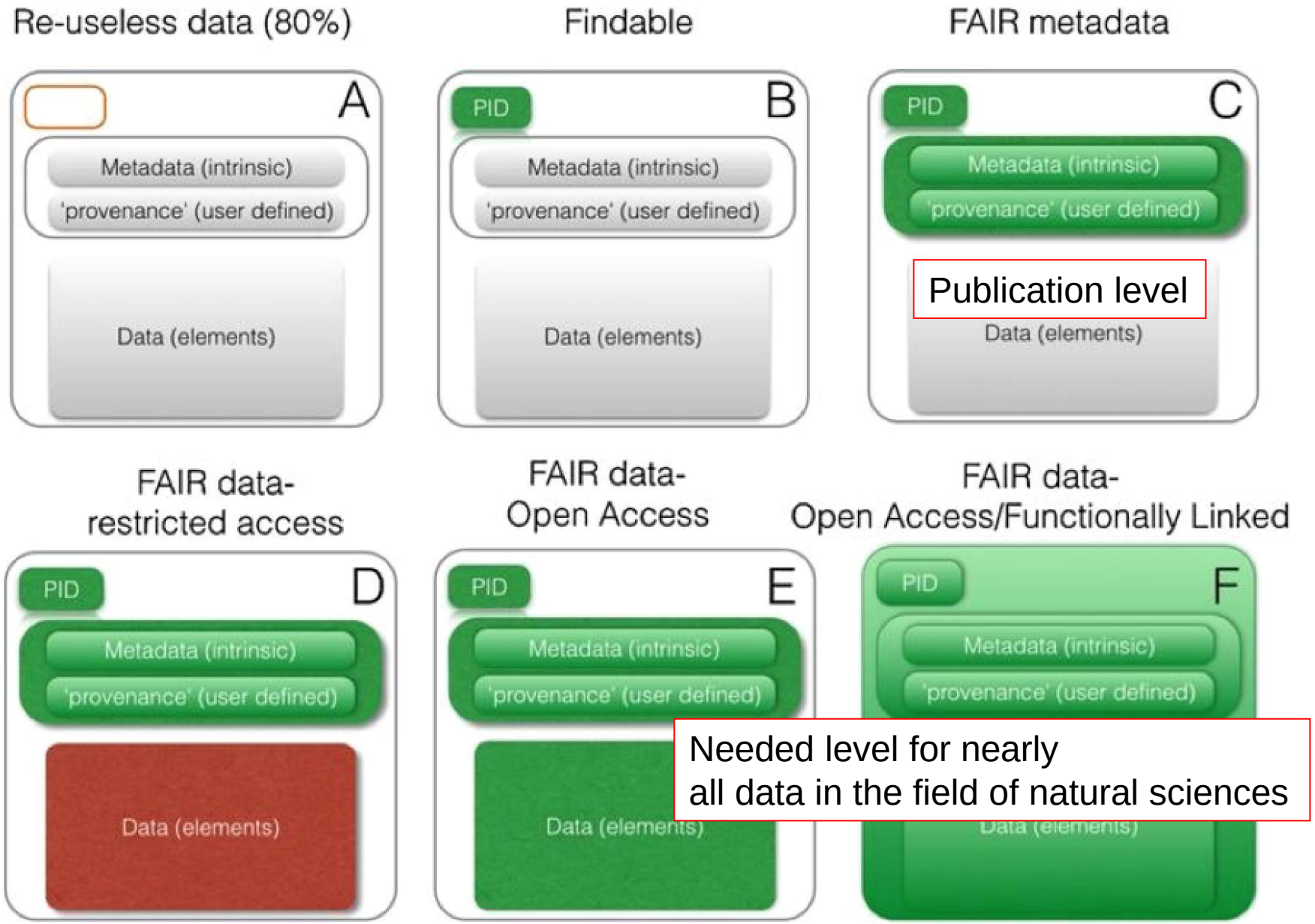


Ferguson et al., Nature, 2014



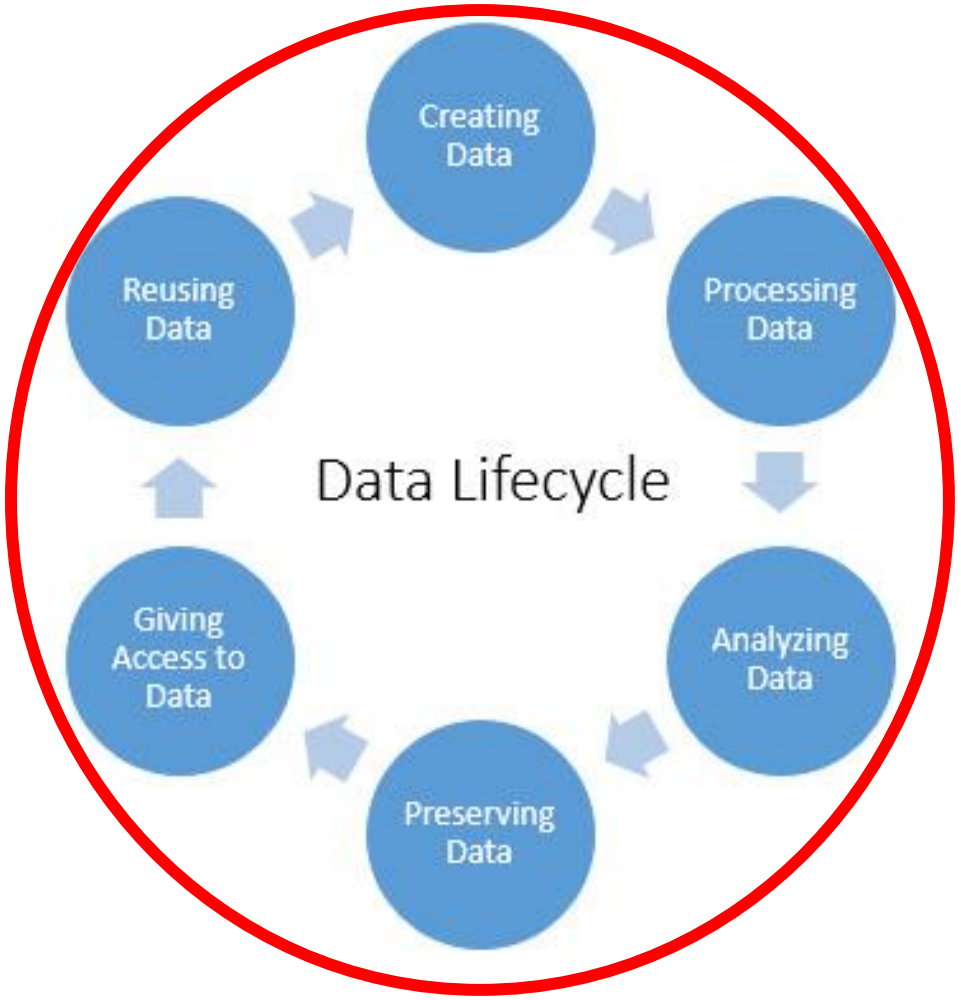
Pareto Principle 80/20 rule

Data as increasingly FAIR Digital Objects



Estimated cost: 5% of the received funding (eligible cost)!

Managing research data according to FAIR principle



The FAIR Guiding Principles

FAIRness in practise: How to achieve it?

- Solely funding either the scientists or long-term repositories infrastructure part will not make data FAIR
- Close communication with the scientists to understand their needs (and fears) and establishing domain specific metrics that can be implemented
- Scientists need an infrastructure that supports them and takes care of the technical details
- Need for professional (scientific) data managers working in close collaboration with the scientific community and long-term repositories

March 19, 2021

Report Open Access

Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship

Mijke Jetten; Marjan Grootveld; Annemie Mordant; Mascha Jansen; Margreet Bloemers; Margriet Miedema; Celia W.G. van Gelder

Other(s)

Sabrina Gunput; Mirjam Brullemans-Spansier; Peter-Bram 't Hoen; Salome Scholtens; Jasmin Böhmer; Iza Witkowska; Kristina Hettne; Joanne Yeomans; Erik Jansen; Inge Slouwerhof; Marta Teperek; Yan Wang; Brett Olivier; Lena

GO FAIR states: “It is irresponsible to support research but not data stewardship” and recommends investing 5% of research funds in ensuring data are reusable.

About the NPOS F project: https://www.openscience.nl/en/projects/project-f-professionalising-data-stewardship-

Publication date:
March 19, 2021

DOI:
DOI 10.5281/zenodo.4623713

The FAIR Guiding Principles - responsibilities

To be Findable:

- F1. (meta)data are assigned a g and persistent identifier
- F2. data are described with rich (defined by R1 below)
- F3. metadata clearly and explici identifier of the data it describes
- F4. (meta)data are registered or searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier us
- A im
- A au
- A ar

mal, accessible, shared, and uage for knowledge

bularies that follow FAIR

ualified references to other

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes

We need critical support staff on permanent positions!
And we have to ensure that science is not detached from technology and that scientists trust the services possible through FAIRness!

Blue: Long-term archive
Green: Scientific community (Scientist and data manager)

Thank you!

NHH