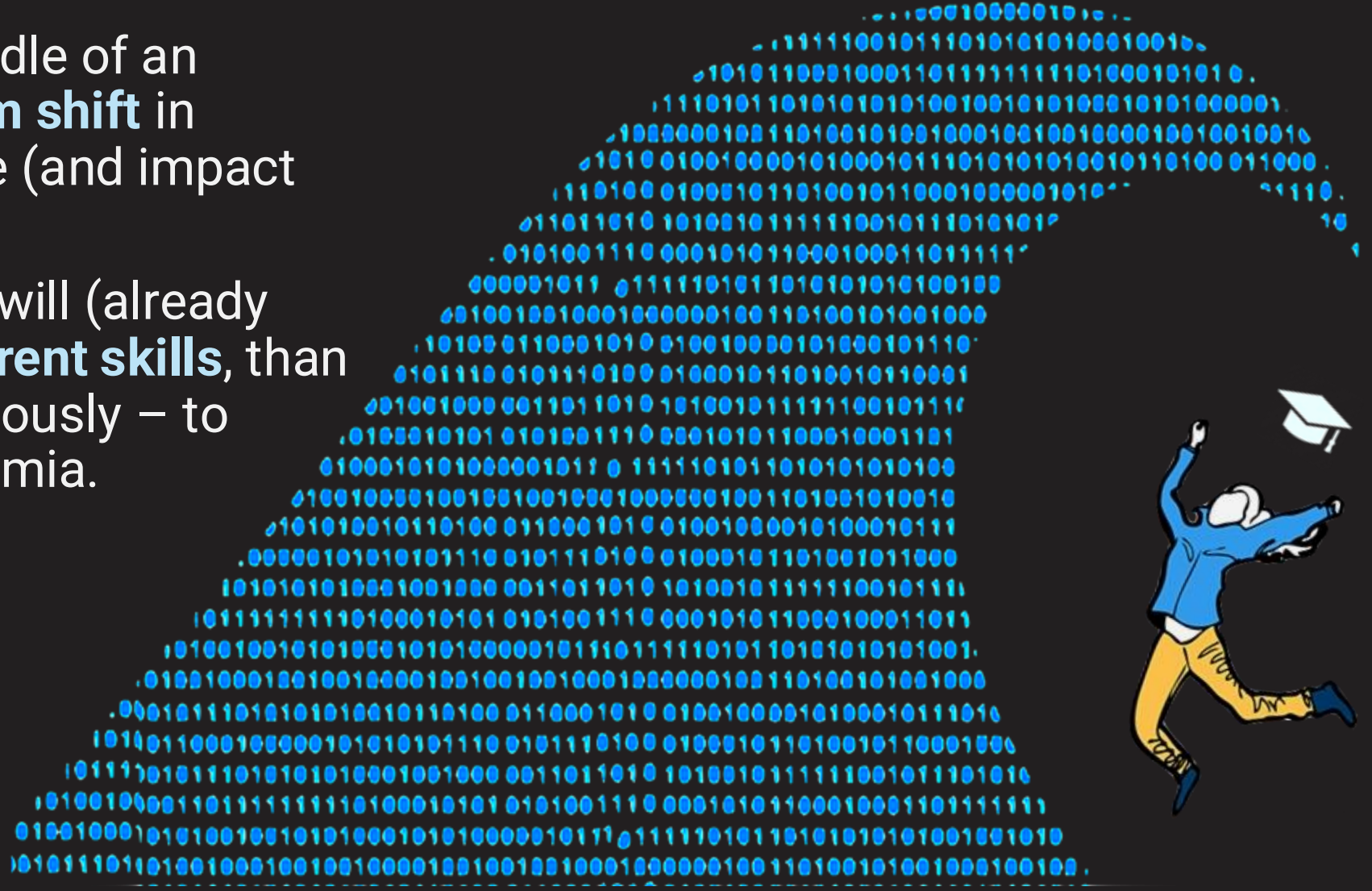


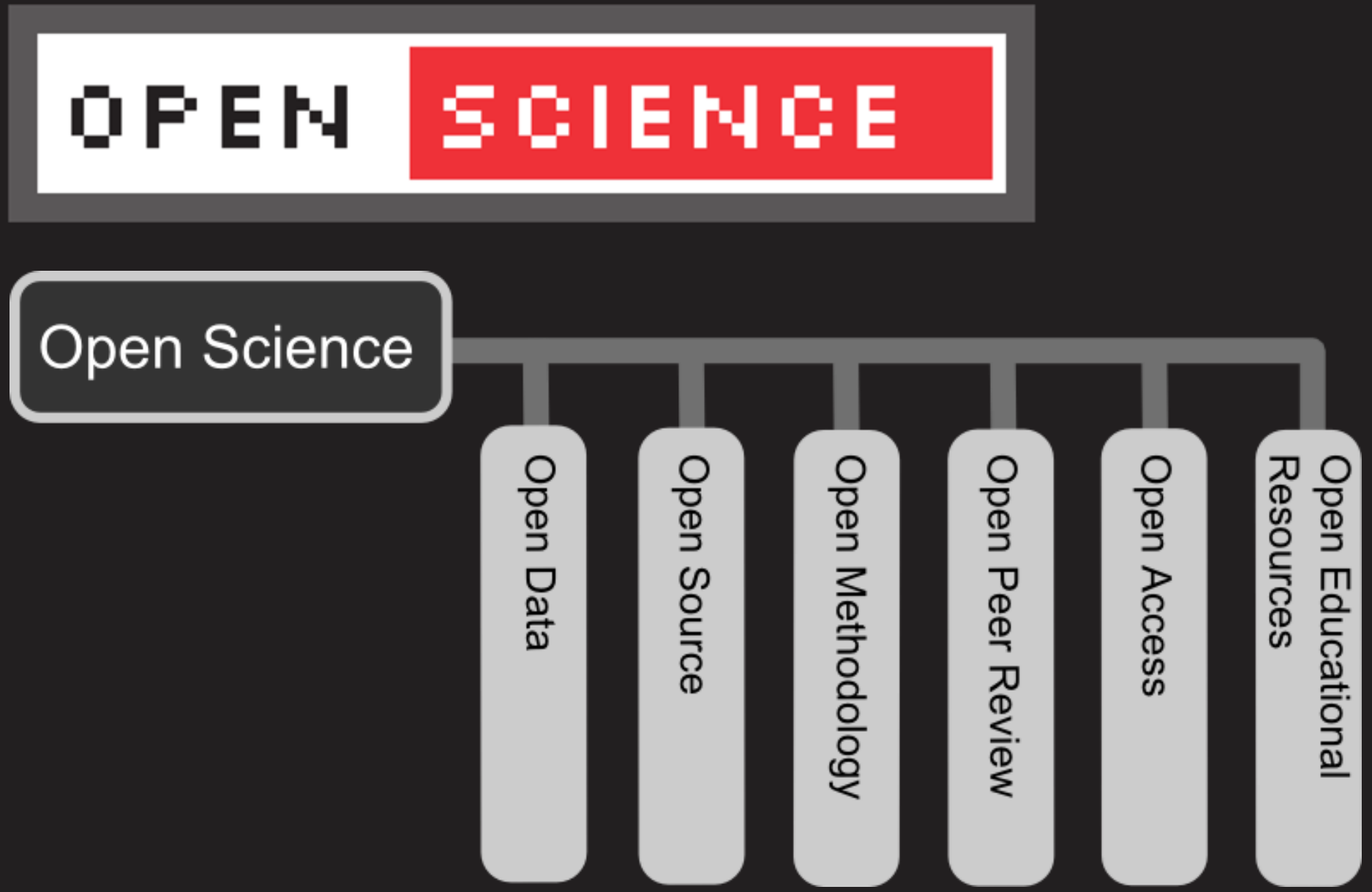
What is Open Science and why is it important for students?

GBIF.no

WHY TEACH STUDENTS OPEN SCIENCE ?

- We are in the middle of an **ongoing paradigm shift** in scientific practice (and impact metrics).
- Young scientists will (already today) **need different skills**, than was needed previously – to succeed in academia.




OPEN SCIENCE
Open Science
Open Data
Open Source
Open Methodology
Open Peer Review
Open Access
Open Educational Resources

WHAT IS OPEN SCIENCE?

- **Open science** is the movement to make scientific research (*including publications, **data**, physical samples, and software*) and its dissemination accessible to all levels of an inquiring society, amateur or professional (Woelfle *et al*. 2011).
- **Open science** is transparent and accessible knowledge that is shared and developed through ***collaborative networks*** (Vicente-Saez *et al*. 2018).
- **Open Science** can be seen as a ***continuation of***, rather than a revolution in, practices begun in the 17th century with the advent of the ***academic journal*** (David 2004).

- Woelfle, M.; Olliaro, P.; Todd, M. H. (2011). "Open science is a research accelerator". Nature Chemistry. 3 (10): 745–748. doi:10.1038/nchem.1149
- Vicente-Saez, Ruben; Martinez-Fuentes, Clara (2018). "Open Science now: A systematic literature review for an integrated definition". Journal of Business Research. 88: 428–436. doi:10.1016/j.jbusres.2017.12.043
- David, P. A. (2004). "Understanding the emergence of 'open science' institutions: Functionalist economics in historical context". Industrial and Corporate Change. 13 (4): 571–589. doi:10.1093/icc/dth023
- https://en.wikipedia.org/wiki/Open_science

Open Access (OA): Research results distributed online and free of costs or other barriers – often meaning free access to research articles.

Open Science: Researchers to share their methods, computer code and research data in central data repositories.

Open Data: is freely available to everyone to use and re-publish as they wish, without restrictions from copyright, patents or other mechanisms of control.

FAIR data principles: **findable**, **accessible**, **interoperable** and **reusable.**



WHY OPEN RESEARCH DATA?



WHY OPEN RESEARCH DATA?

- We are in the middle of an ongoing **paradigm shift** in scientific practice (*and impact metrics*).
- Researchers will need to develop **different approaches**, than they needed in the past – *to remain relevant*.
- Society is gaining Big Data maturity and will **expect new services** from marine sciences.
- The open science wave is moving **fast**!





OPPORTUNITIES

- Enables new research methodologies that were not possible before.
- Skills for open research and open data are in increasing demand!
- Funding opportunities.
- GBIF brings new benefits and opportunities for our museums.

Open science is expanding your possibilities…



INCENTIVES FOR DATA REUSE

To incentivize the sharing of useful data, the scientific enterprise needs a well-defined ***system that links individuals with reuse of data sets they generate***

Pierce *et al*. Credit data generators for data reuse, *Nature* 6 June 2019



DATA CITATION - A NEW CURRENCY OF SCIENCE

- **Peer-reviewed scholarly papers** in high impact journals still maintain considerable weight for scientific careers.
- A **movement** is under way to **build similar status for open data**, open metadata, open material samples, and other open access scientific research products…





DECLARATION ON RESEARCH ASSESSMENT



- DORA recognizes the need to improve the ways in which the outputs of scholarly research are evaluated.
- Worldwide movement covering all scholarly disciplines and all key stakeholders including funders, publishers, professional societies, institutions (universities), and researchers.

- Developed in 2012 in San Francisco
- To date (2025-11-20), 23 057 individuals and 3 516 organizations in 168 countries have signed DORA.
- The Research Council of Norway (RCN) signed DORA in May 2018 [link]

SCIENTIFIC JOURNALS REQUIREMENTS



Open research data policies

Eg. Springer Nature:

- All original articles must include a data availability statement.
- Strongly encourage that all datasets supporting the analysis and conclusions of the paper are made publicly available at the time of publication.
- Peer reviewers are entitled to request access to underlying data (and code) when needed to perform their evaluation of a manuscript.
- Recognize it is not always possible to share research data publicly, for instance when privacy of research participants could be compromised. (Statement of conditions for access required).

SPRINGER NATURE

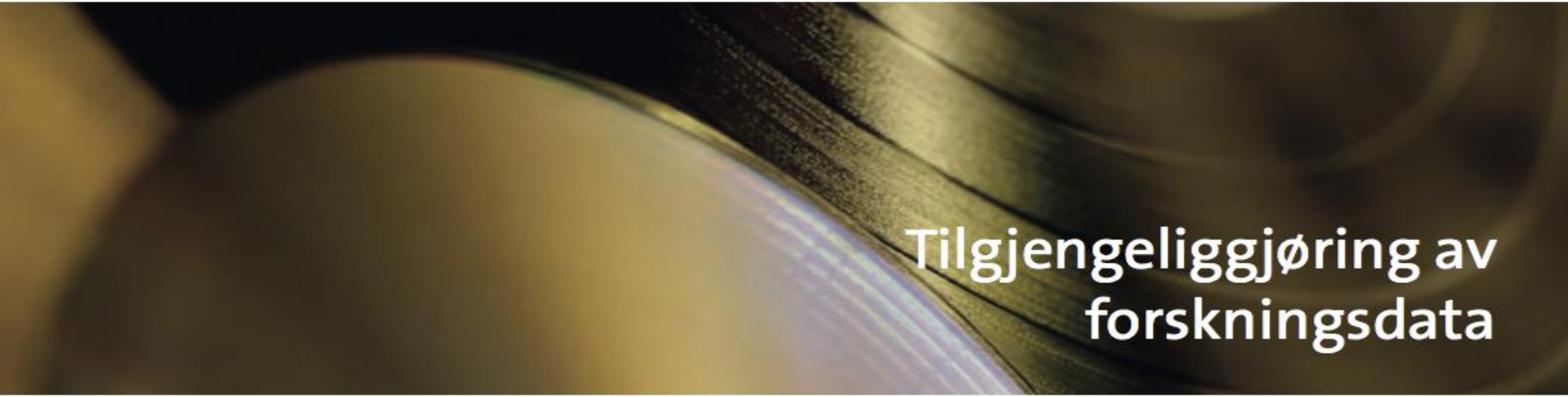
https://www.springernature.com/gp/authors/research-data-policy

RESEARCH FUNDING REQUIREMENTS



*2014: The Research Council of Norway expects **all research data** from projects funded by the Research Council **to be made freely available as open data**.*
In some situations, there can be valid and justified reasons for exceptions.

Forskningsrådet



Forskningsrådet (**2014**). ISBN: 978-82-12-03361-0

Policy for
Norges forskningsråd

https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/

UNIVERSITY REQUIREMENTS



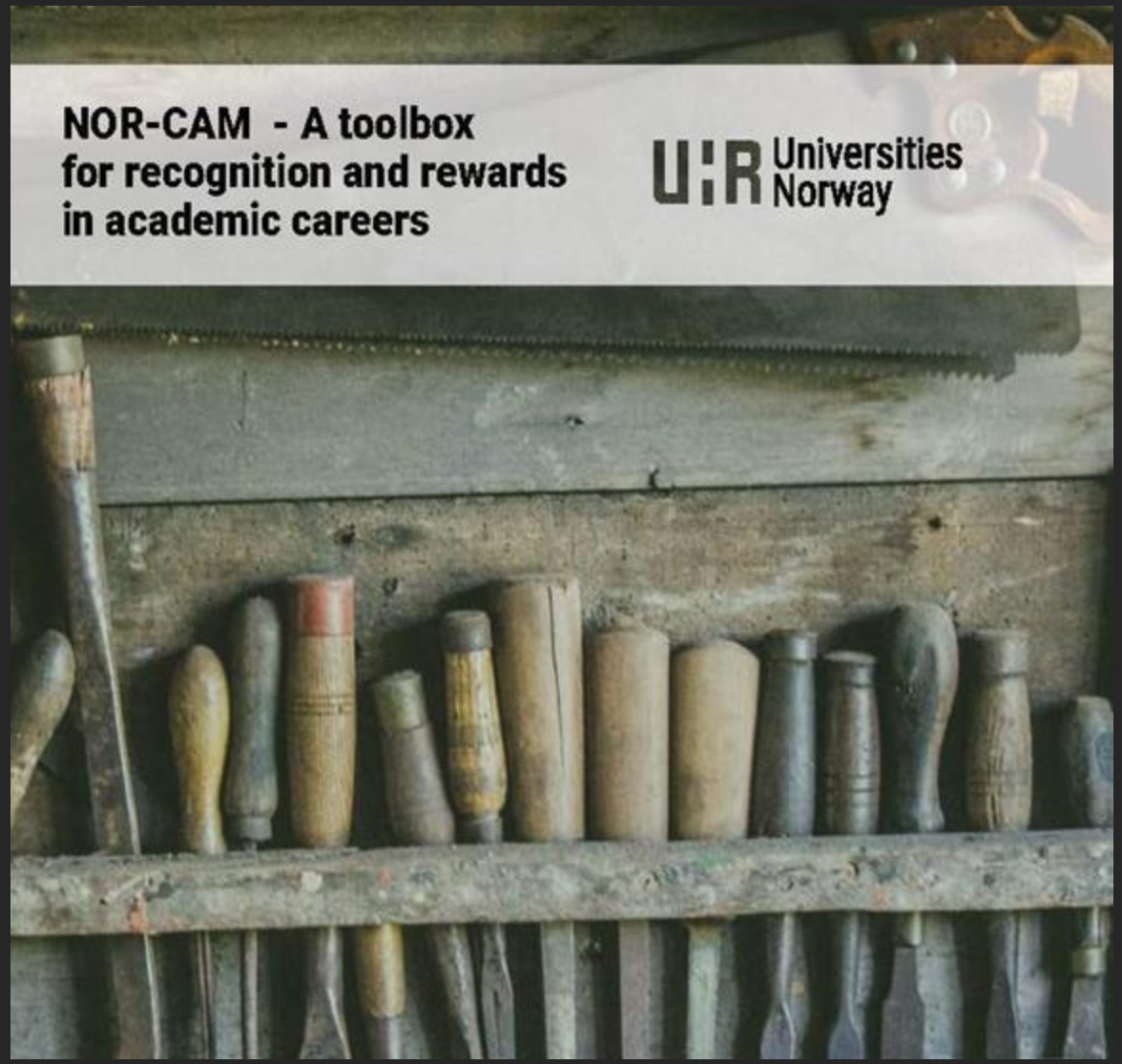
NOR-CAM (2021)

Guiding principles for the assessment and evaluation of research(ers) in light of the transition to Open Science. Proposes a more holistic framework for recognition and rewards in academic research assessment.

Academic institutions when recruiting academic and scientific staff; in processes related to promotion from one scientific career stage to the next; career guidance of employees by department and research leaders

Funders when assessing project managers and participants in connection with research applications.

National authorities when evaluating of Norwegian research and education.



https://www.uhr.no/en/news-from-uhr/nor-cam-a-toolbox-for-recognition-and-rewards-in-academic-careers.5780.aspx

GBIF

GOVERNMENT POLICIES

New national strategy of digitization:

Norway shall become the most digitized country in the world

Press release | Date: 2024-09-26 | Ministry of Digitization and Public Governance | Office of the Prime Minister



https://www.regjeringen.no/no/aktuelt/norge-skal-bli-verdens-mest-digitaliserte-lan/id3055039/

GBIF

Nasjonale retningslinjer for åpen tilgang til forskningsresultater

Norway is to be a careful pioneer in open access to research results.
*Norway to follow the ambition of EU on **full open access to publicly funded research**.*

Results of research supported by public and public-private funds freely available to and reusable by anyone.

Kunnskapsdepartementet (2016)

EU (2016) Competitiveness Council, 26-27/05/2016

EU (2007) INSPIRE Directive

REPRODUCIBILITY CRISIS



REPRODUCIBILITY CRISIS

- "***Scientific irreproducibility*** *— the inability to repeat others' experiments and reach the same conclusion*" (Nature 2016)

Baker (2016) 1,500 scientists lift the lid on reproducibility. Nature. doi:10.1038/533452a



IS THERE A REPRODUCIBILITY CRISIS?



"***Scientific irreproducibility*** *— the inability to repeat others' experiments and reach the same conclusion — is a growing concern*".

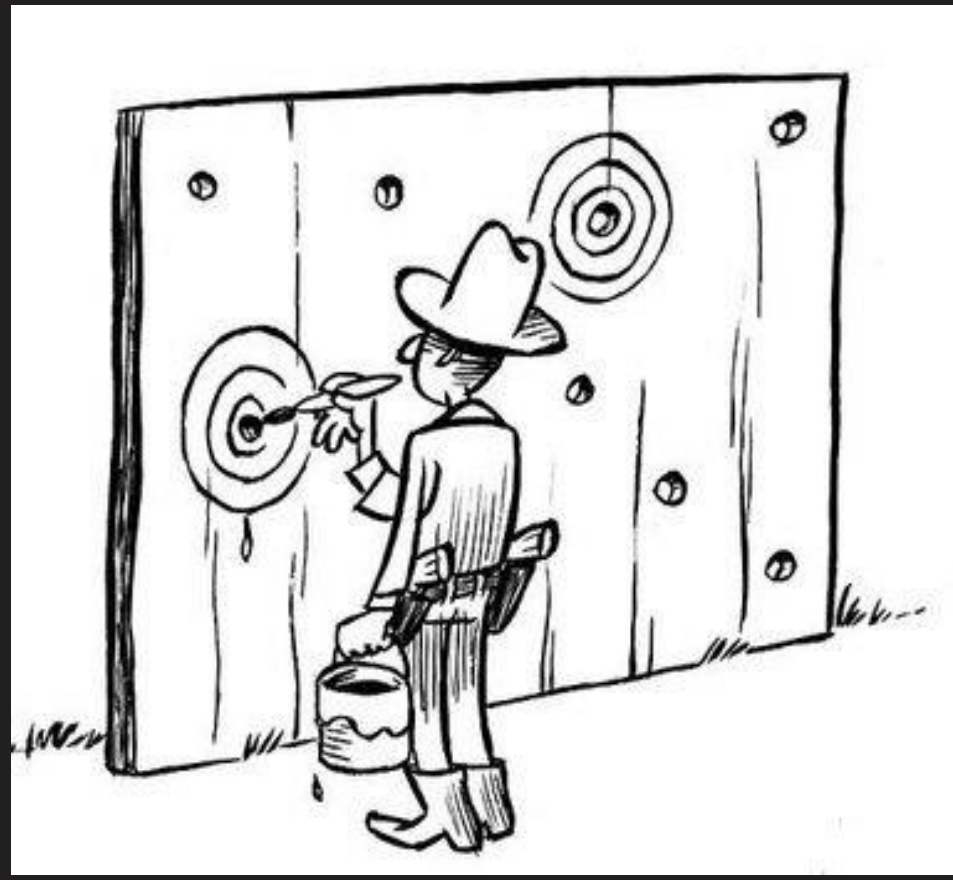
Baker (2016) **Nature** doi:10.1038/533452a

Open Science solution*:* researchers to share their **methods**, **data**, computer **code** and results in central data repositories.

Note that we also need herbarium ***specimen*** *and bio-repositories (eg. museums).*

WILL ANYBODY TRUST CLOSED SCIENCE AGAIN?

- Studies indicates that **p-hacking** is a significant problem - sometimes even without the scientist even being aware of it.
- **Pre-registered (open) data** provides a good insurance against suspicion of both data dredging (and plain data falsification).
- **"p-hacking"** (data dredging, data fishing, …) occurs when researchers collect or select data or statistical analyses until non-significant results become significant.



Ioannidis (2005). "Why Most Published Research Findings Are False". PLoS Medicine. doi:10.1371/journal.pmed.0020124.
Head *et al*. (2015) The Extent and Consequences of P-Hacking in Science. PLoS Biol. doi:10.1371/journal.pbio.1002106

Reuse of research data and other research products



Data Citation Principles

FORCE11
The Future of Research Communications and e-Scholarship

1. Data to be legitimate **citable** products of research.
2. Data citations giving **scholarly credit** and attribution.
3. In scholarly literature, whenever claims are based on data, **data should always be cited**.
4. Persistent method for **identification of data**, that is machine actionable, globally unique, universal.
5. Data citation facilitate **access to data** or at least to **metadata**.
6. **Unique identifiers** that persist even beyond the lifespan of the data.
7. Data citation identify and access the specific data that **support verification** of the claim (provenance, time-slice, version).
8. Flexible, but attention to **interoperability** of practices across communities.

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014

Findable Accessible Interoperable Reusable



FAIR data is about ***machine-readable*** data

*researchers need to do more than simply post their data on the web for it to be **re-usable**.*

What is FAIR Data?

FINDABLE

- Data and supplementary materials have sufficiently rich **metadata** and a unique and **persistent identifier**.

ACCESSIBLE

- Metadata and data are **understandable** to humans and **machines**. Data is deposited in a trusted **repository**.

INTEROPERABLE

- Metadata use a formal, accessible, shared, and broadly applicable language for **knowledge representation**.

REUSABLE

- Data and collections have a **clear usage licenses** and provide accurate information on **provenance**.

https://www.go-fair.org/



FAIR DATA PRINCIPLES

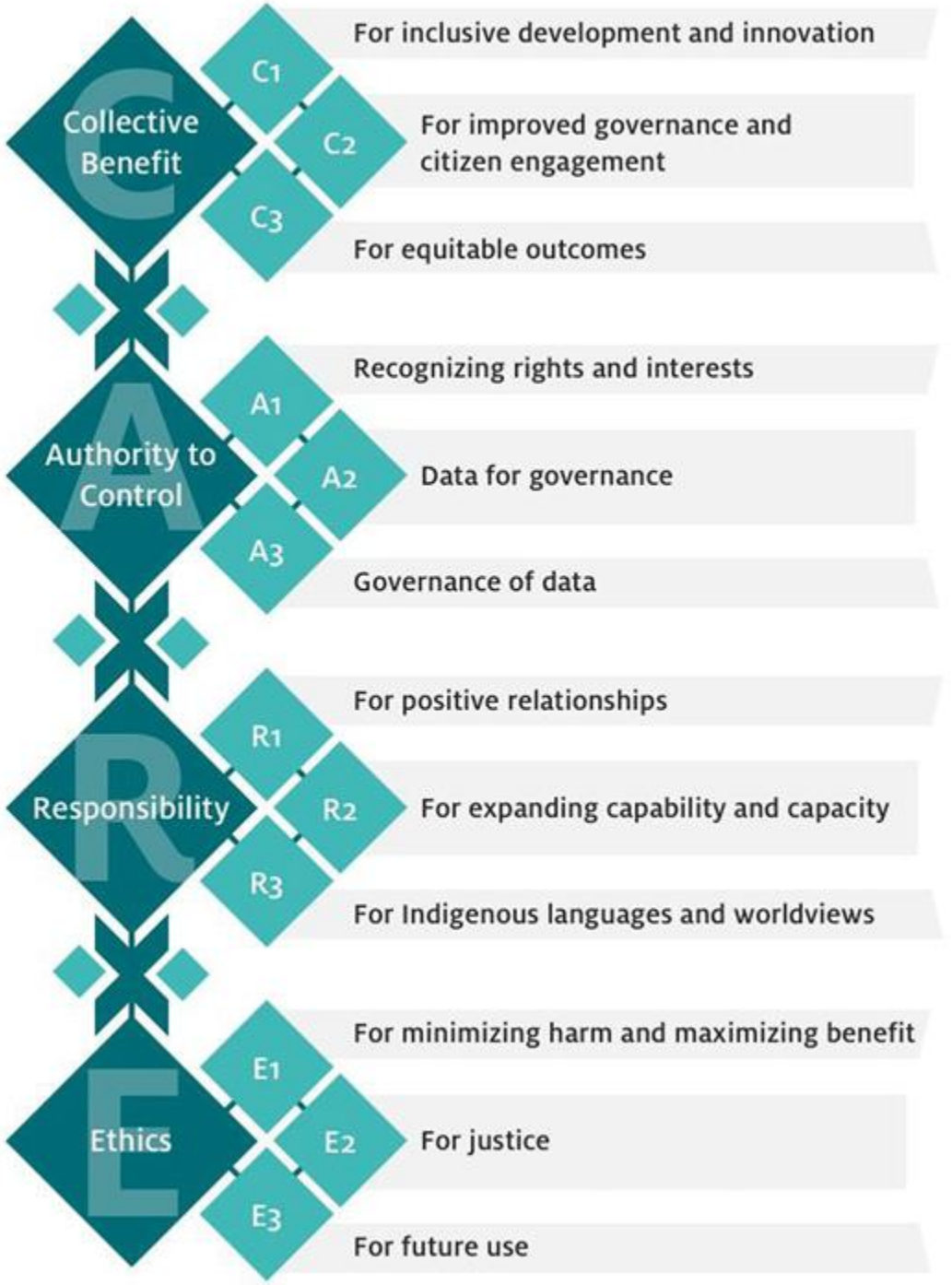



Be
FAIR
Findable
Accessible
Interoperable
Reusable
and
CARE
Collective Benefit
Authority to Control
Responsibility
Ethics
https://www.gida-global.org/care

CARE DATA PRINCIPLES

- CARE is an acronym which stands for:
  - **Collective** Benefit,
  - **Authority** to Control,
  - **Responsibility**,
  - **Ethics**
- CARE was created by the International Indigenous Data Sovereignty Interest Group, a group that is a part of the Research Data Alliance.
- Resolve **Indigenous Peoples' rights** to and interests in their data across the data lifecycle.
- Building on **FAIR** data principles while ensuring indigenous interests.

https://www.gida-global.org/care




META-
DATA

Why publish open data?

- Data produced using **public funds** should be regarded as a common good, and should be made available for inspection, interpretation and re-use by third parties.
- Needless **duplication** of data-collecting **efforts and costs** will be reduced.
- Open data increases transparency and overall **quality of research**.
- Published data can be re-analysed, verified, and **improved by others**.
- Data publication increase **recognition and opportunities** for collaboration.
- Published data can be **cited and re-used**, alone or in combination with other data.
- Data owners and collection managers can **trace data use and citation**.
- Data creators, their institutions and funding agencies can **be credited**.
- Data can be **integrated with other datasets** across space and time.
- Open data increases potential for interdisciplinary research and **re-use in new contexts not envisioned by the data creator**.

Penev *et al*. (2017) https://doi.org/10.3897/rio.3.e12431

Why metadata?

In general, metadata should allow a prospective end user of data to:

1. identify/discover its existence,
2. learn how to access or acquire the data,
3. understand its fitness-for-use,
4. learn how to transfer (obtain a copy of) the data, and
5. learn how the data should be used.



Photo CC BY-SA Jennifer Fagan-Fry (NOAA)

What is Metadata?

Metadata, literally “***data about data***”

An essential component of a data management system, describing such aspects as the “what, where, when, who and how” pertaining to a resource.

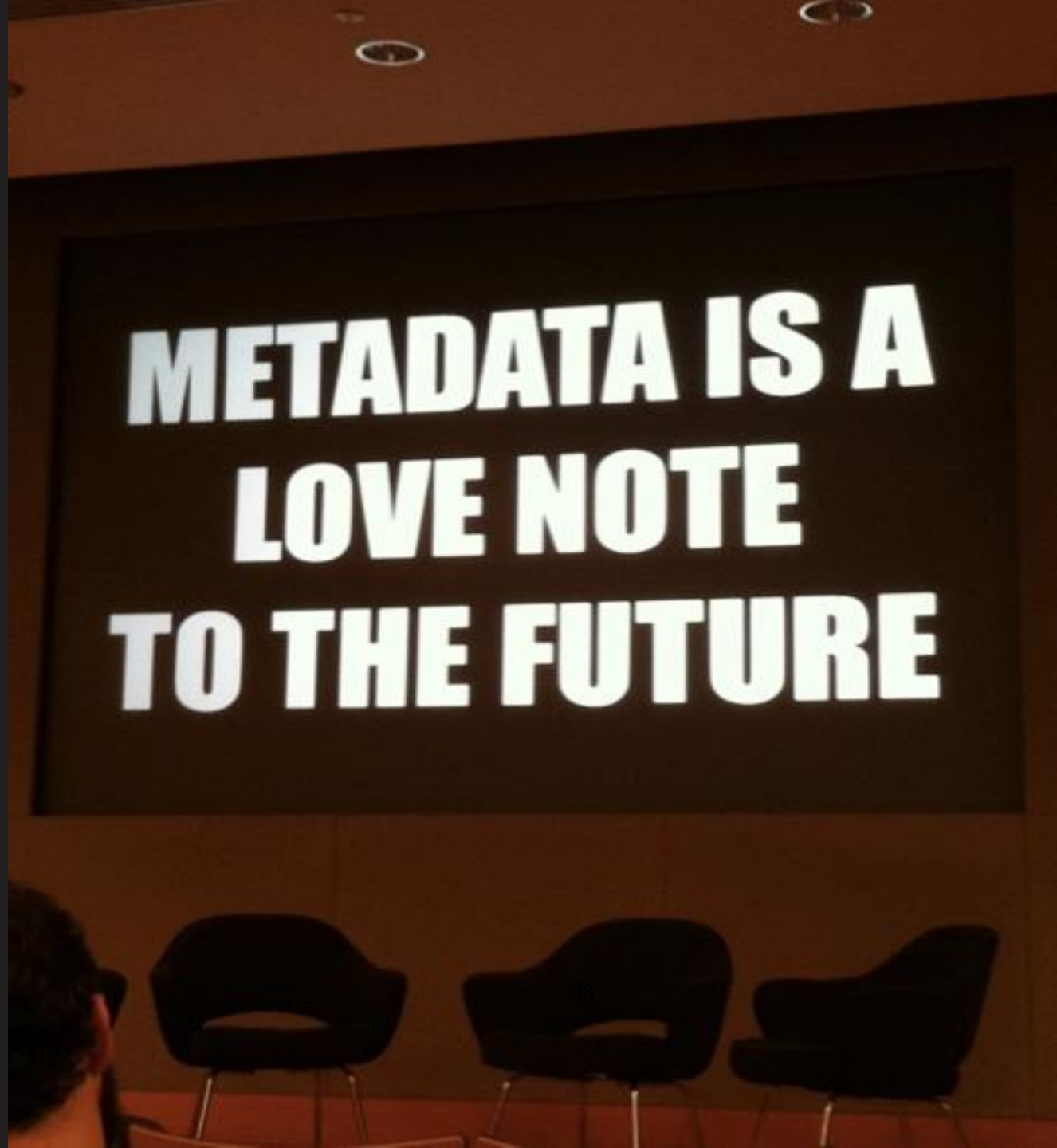


Photo: CC-BY by Cea+ http://www.flickr.com/photos/centralasian/8071729256
Slide source CC BY EUDAT (2016) |

Data entropy

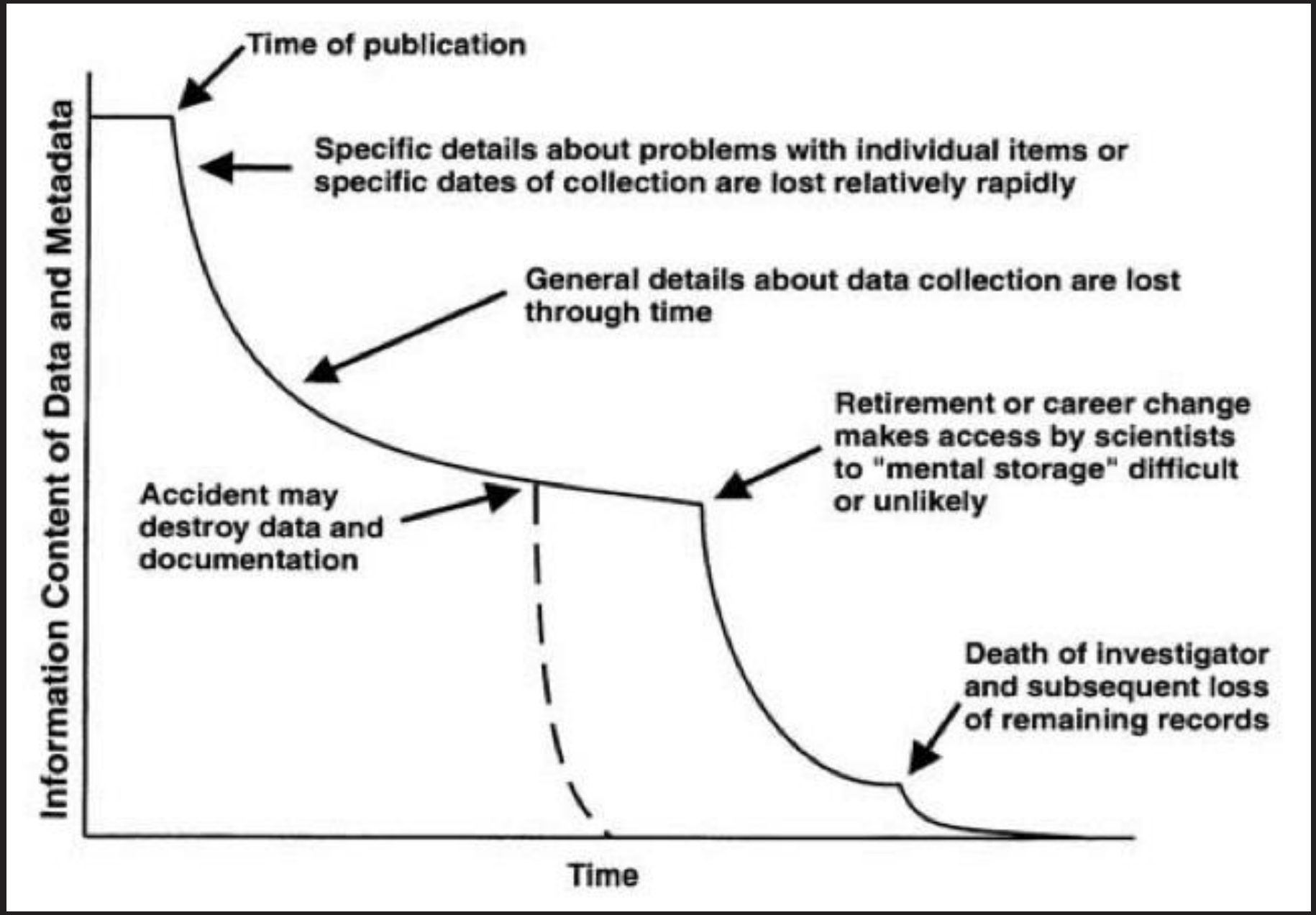


Illustration from: The Loss of Information about Data (Metadata) Over Time, Michener *et al*, 1997

DATA MANAGEMENT PLAN



Illustration CC BY Jørgen Stamp

Data Management Plan (DMP)

A formal document that outlines **how data are to be handled** during a research project, and after the project is completed.

The goal is to **plan** data management **before** the project begins.

Including a plan for the **costs** of data management and archiving.

This **saves time** in the long run and promotes **data fitness for reuse**.

Reduce **duplication** of existing scientific studies.

Reduce the **loss of data**.



Illustration CC BY Jørgen Stamp

Why write Data Management Plans?

A data management plan is a tool for making your research **reproducible** and thus **trustworthy**.

Good data curation saves you research time, because you, your collaborators, and others, will **find**, **understand**, and get **access** to your (own) research data.

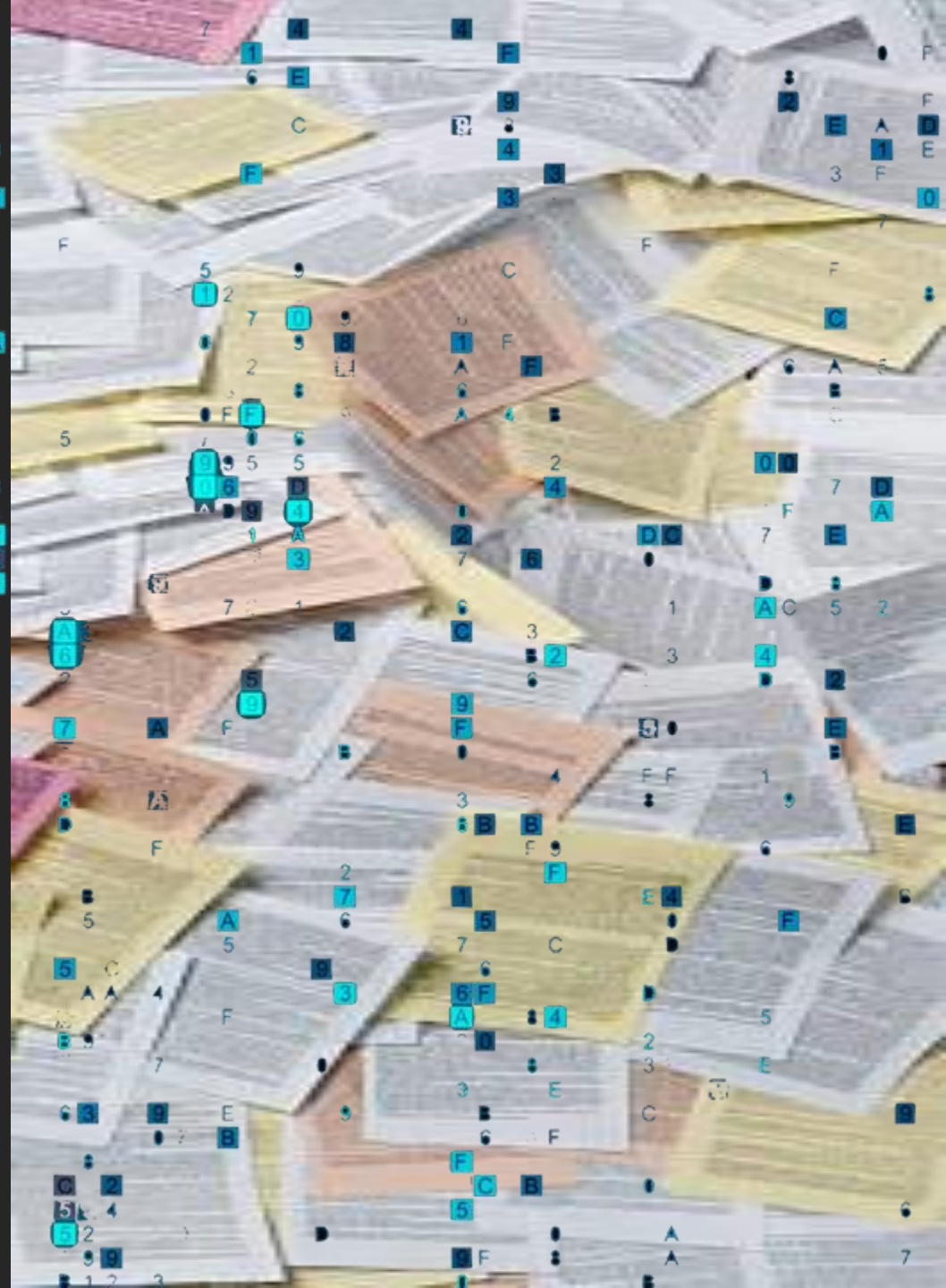
Efficient data sharing provides broader distribution and **impact** for your research results.

Open research data, available for reuse, strengthens open and **curiosity-driven research**, and scientific breakthrough not originally foreseen by the original data producer.



Illustration CC BY Jørgen Stamp

DATA PAPER



What is a «*data paper*»?

A data paper is a peer reviewed document **describing a dataset**, published in a peer reviewed journal. It takes effort to prepare, curate and describe data. Data papers **provide recognition** for this effort by means of a scholarly article.

- Getting **scholarly recognition** for your datasets.
- Promote and improve the **fitness for reuse** of your research data.

https://www.gbif.org/data-papers



Data papers explained

A data paper is a metadata document, describing a particular dataset or a group of datasets, published in the form of a peer-reviewed article in a scholarly journal.

Unlike a conventional research article, the primary purpose of a data paper is to **describe data** and the circumstances of their collection, rather than to report hypotheses and conclusions.

GBIF has been working with partners in academic publishing to promote the data paper as a means of bringing **credit and recognition** to all those involved in data publication.

https://www.gbif.org/data-papers



Why publish data papers?

- Authoring clear, informative **metadata** is an essential step if biodiversity data are going to be discovered and used to inform research and decisions. **This involves extra work, and data publishers need incentives to do it**. In the absence of such incentives, too many datasets are published with poorly-documented metadata or, worse still, no metadata at all.
- Data papers help to overcome barriers to **authoring of metadata** by **providing clear acknowledgement of all those involved** in the collection, management, curation and publishing of biodiversity data.

https://www.gbif.org/data-papers



By publishing a data paper

- **Receive credit** through indexing and **citation of the published paper**, in the same way as with any conventional scholarly publication, offering benefits to authors in terms of recognition and career building.
- **Increase the visibility, usability and credibility** of the data resources you publish.
- **Track** more effectively the **usage** and **citations** of the data you publish.

https://www.gbif.org/data-papers



Machine-readable data …



MACHINE-READABILITY REQUIRES PERSISTENT IDENTIFIERS

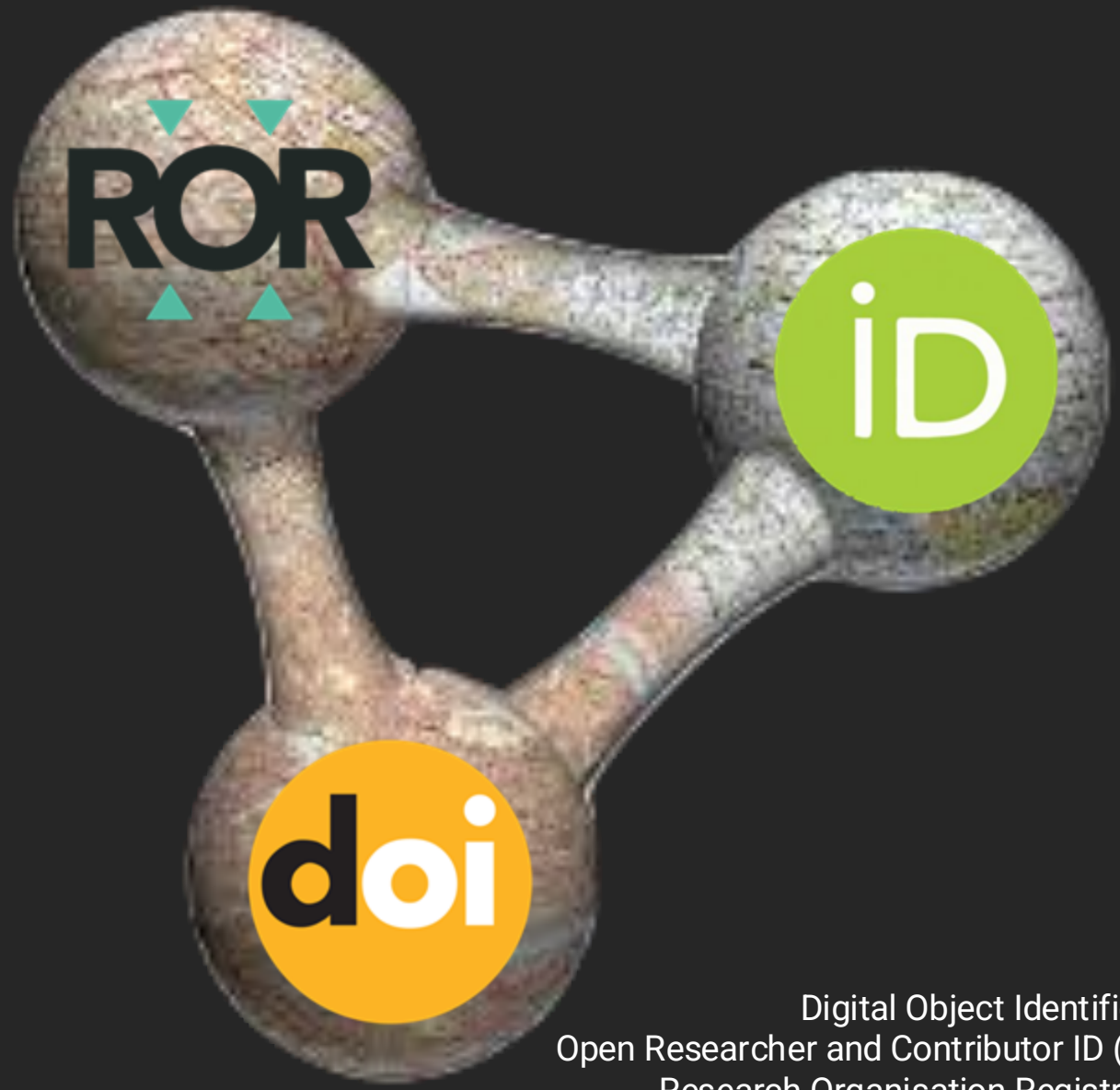
The purpose of identifiers is
… to name things
… making it possible to refer to them



- To uniquely identify something, it needs a **persistent identifier**, a PID.
- A Persistent Identifier is **globally unique**, **persistent**, and **resolvable**“.
- A PID is resolvable when it allows both **human and machine** users to access an object or its representation, and its *Kernel* Information.
- **Kernel Information** is a structured record that contains information (metadata) about the referred object, such as a pointer to the location where the data for the object can be found.

ROR for institutions
ORCID for people
DOI for datasets
GRSciColl UUID for collections

will enable the linking of museum collection specimens to scientific litterature and scientific actors (authors, curators, etc)



Digital Object Identifier (DOI)
Open Researcher and Contributor ID (ORCID)
Research Organisation Registry (ROR)
Universally Unique Identifier (UUID)

THANK YOU

GBIF Norway
Dag Endresen | Node Manager
helpdesk@gbif.no





EXERCISES

EXERCISES FOR THE OPEN SCIENCE SESSION

- Identify 5 data journals and describe why you would choose one journal over the other for documenting a GBIF dataset with a data paper.
  - How is a data paper different from a normal research paper?
  - Do you know of other relatively new types of research papers related to open science?
- Identify 5 tools for writing a data management plan and describe the pros and cons of using each of them (for projects generating biodiversity data).
  - What are some of the main topics to cover in a data management plan?
- What data license would you choose for your biodiversity datasets – why?
  - Which data licenses are supported by GBIF? For data? – for auxiliary data types such as photos?
- The FAIR data principles states “as open as possible, as closed as necessary”. When should FAIR data be open and when should it be closed access? (In your opinion).