Global Biodiversity Information Facility

Dag Endresen - GBIF Norway

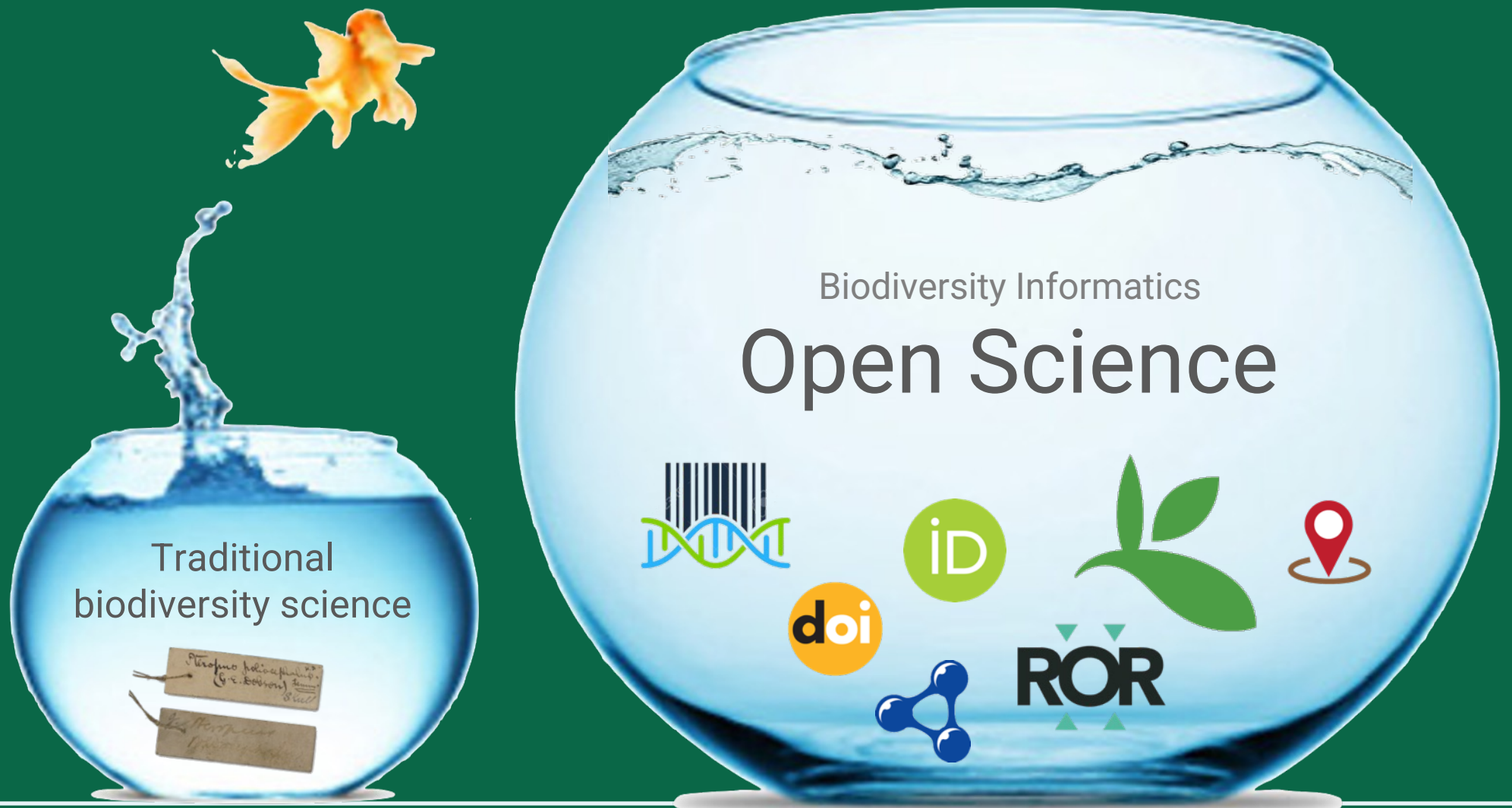
University of Oslo, Natural History Museum

GBIF Global Biodiversity Information Facility

Illustration: GBIF data portal

Open Science course, Hjerkinn | 17th November 2023

WHY OPEN RESEARCH DATA?



new possibilities for novel curiosity-driven research

GBIF

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.

WHY OPEN RESEARCH DATA?

- We are in the middle of an ongoing **paradigm shift** in scientific practice (*and impact metrics*).
- Marine science will also 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**!

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**.*

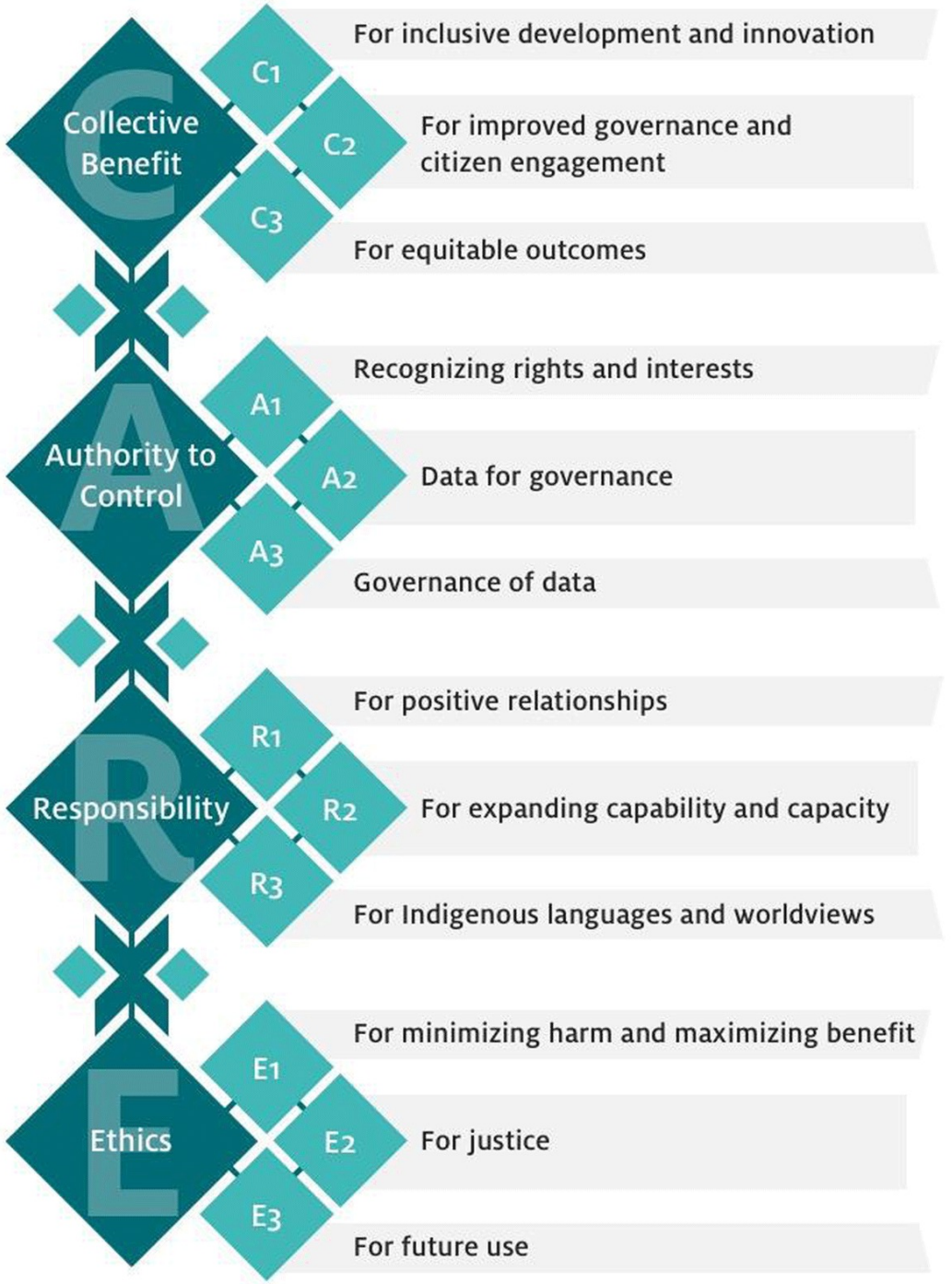
FAIR DATA PRINCIPLES


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



GBIF IS A FAIR & OPEN BIODIVERSITY DATA INFRASTRUCTURE

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 impact metrics.
- 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.

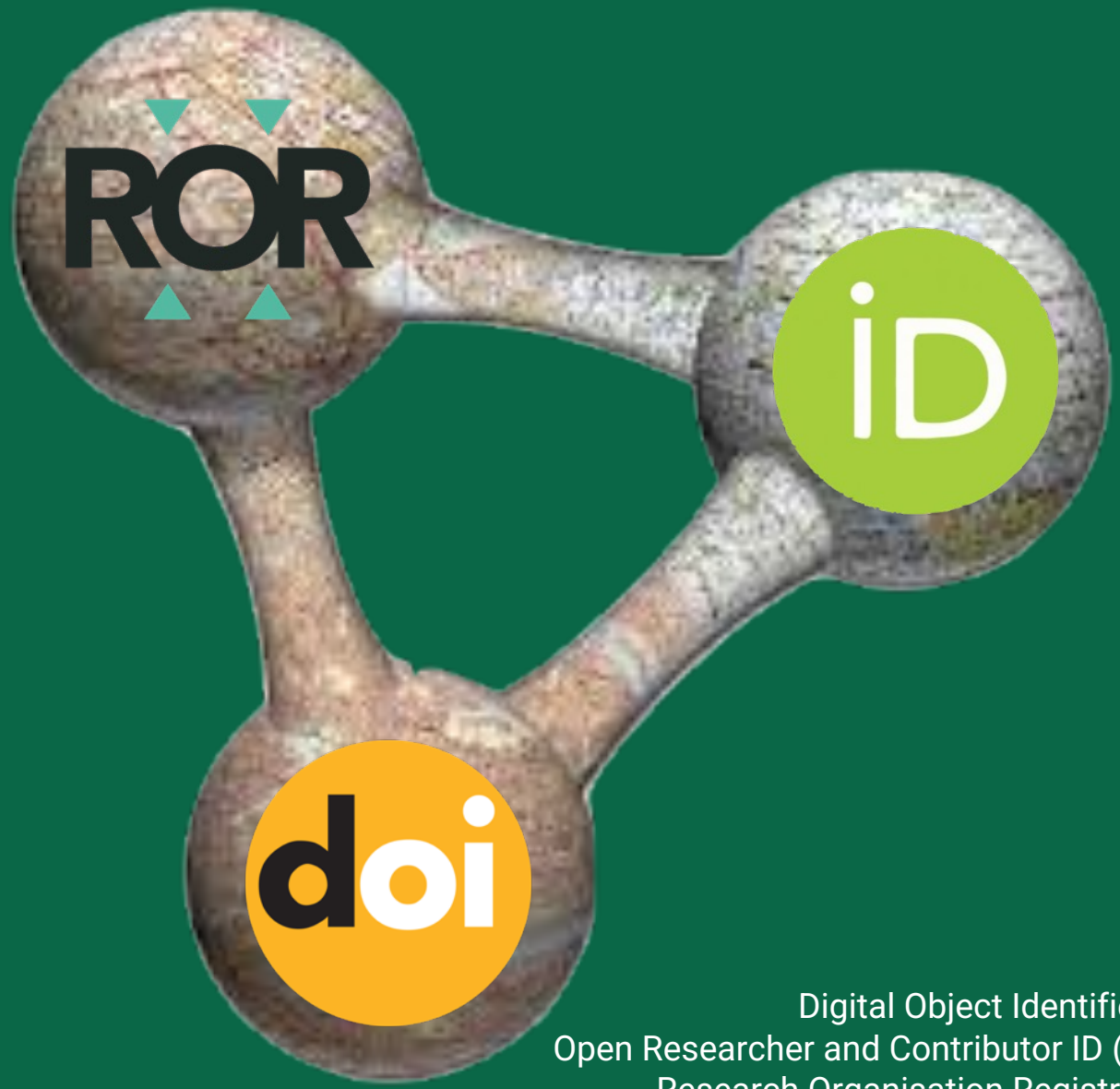
- Developed in 2012 in San Francisco
- To date (2023-11-17), 21 157 individuals and 3 031 organizations in 164 countries have signed DORA.
- The Research Council of Norway (RCN) signed DORA in May 2018 [link]

MACHINE-READABILITY REQUIRES PERSISTENT IDENTIFIERS

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

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)

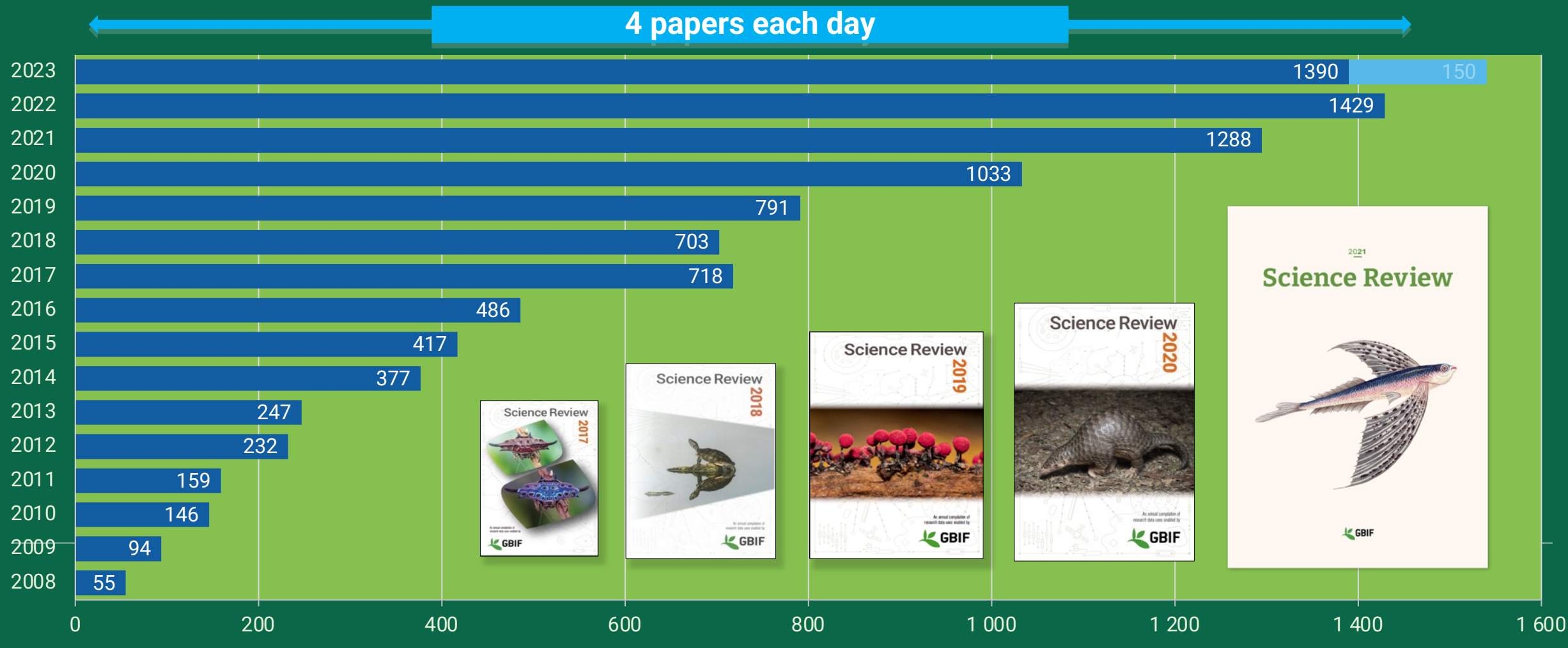


Digital Object Identifier (DOI)
Open Researcher and Contributor ID (ORCID)
Research Organisation Registry (ROR)
Universally Unique Identifier (UUID)

#CiteTheDOI

GBIF started issuing DOIs on 3 February 2015

PEER-REVIEWED PUBLICATIONS USING GBIF-MEDIATED DATA



https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true

#CiteTheDOI

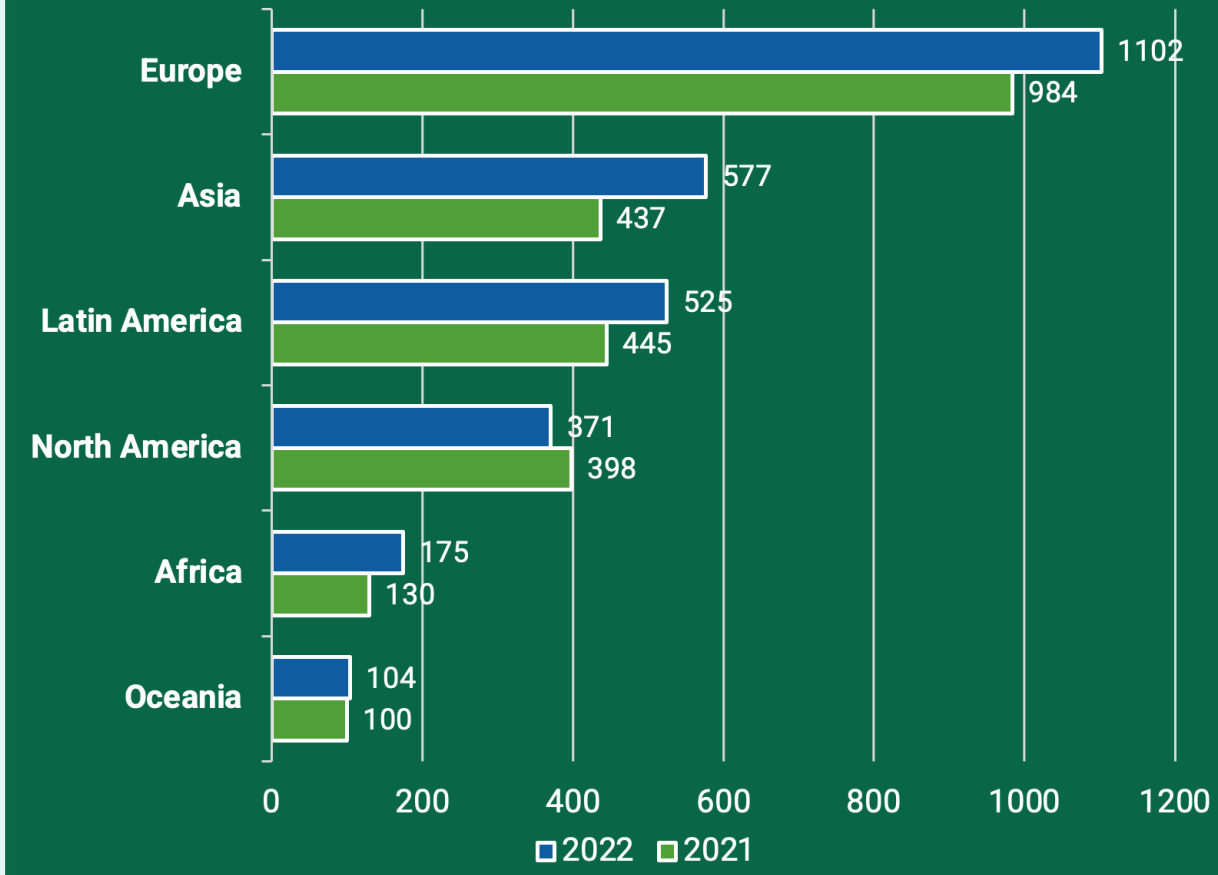
GBIF

DATA USE IN PEER-REVIEWED JOURNALS

Peer-reviewed uses by country

End of Year		*2023*	*2022*	*2021*
1	United States	306	292	328
2	China	300	300	228
3	Brazil	145	152	147
4	United Kingdom	144	146	137
5	Germany	126	120	131
6	Mexico	105	131	134
7	Spain	95	100	103
8	France	88	75	75
9	Australia	80	78	69
10	India	61	62	35
-	Norway	30	41	37

Peer-reviewed uses by region



https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true

GBIF

HOW TO CITE DATA MEDIATED BY GBIF?

1. **Download data** from GBIF.org
2. and receive recommended citation with a **download DOI**
3. **Cite the DOI** in published research or other work

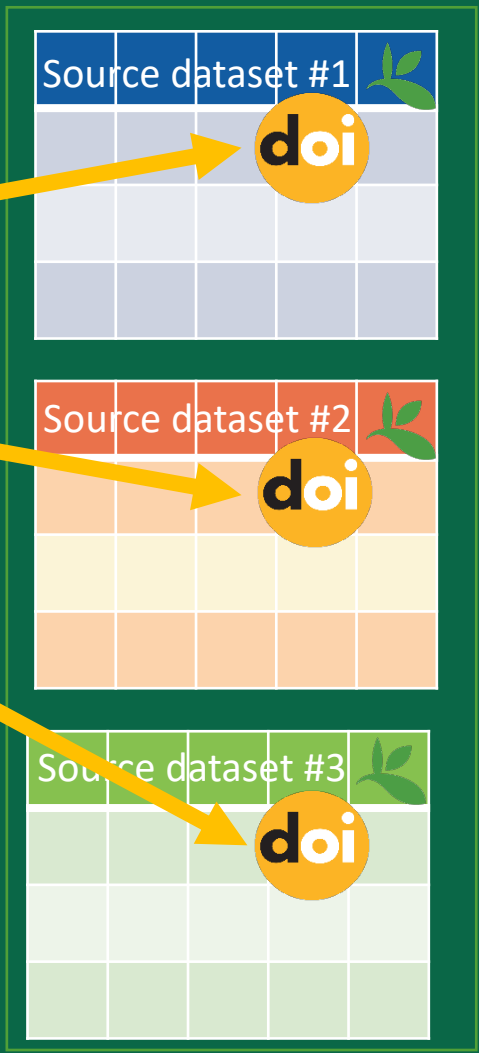
Example: GBIF.org (9 November 2021) GBIF Occurrence Download **https//doi.org/10.15468/dl.xxxxxx**

DOI BASED DATA CITATION AT GBIF.ORG -- #CITETHEDOI

paper

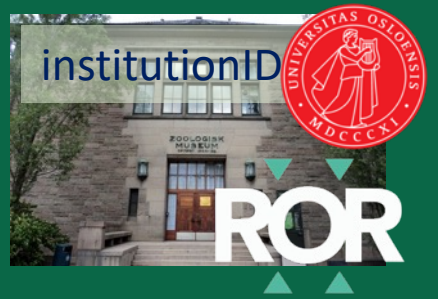
data citations

datasets



#CiteTheDOI

GBIF



Publish datasets in GBIF

Dataset DOIs

Source dataset #1

Source dataset #2

Source dataset #3

Filter & download

Download DOI

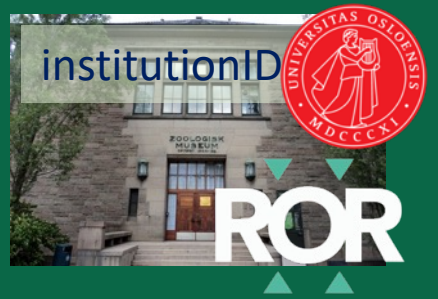
GBIF download

Process & archive

Final state of data

Analyze & publish

Bibliographic DOI



Publish datasets in GBIF

Dataset DOIs

Source dataset #1

Source dataset #2

Source dataset #3

Filter & download

Download DOI

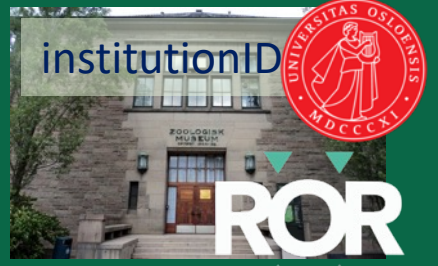
GBIF download

Process & archive

Final state of data

Analyze & publish

Bibliographic DOI



Publish datasets in GBIF

Dataset DOIs

Source dataset #1 doi
UUID
UUID

Source dataset #2 doi
UUID

Source dataset #3 doi
UUID

Filter & download

Download DOI

GBIF download doi
UUID
UUID
UUID
UUID

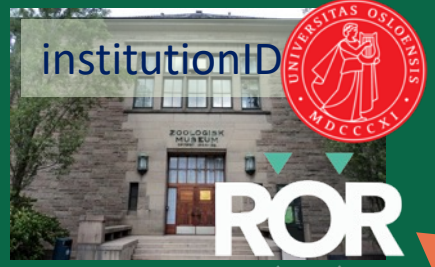
Analyze & publish

Process & archive

Final state of data
UUID
UUID
UUID doi

Bibliographic DOI

Possible with persistent object identifiers



Dataset DOIs

Download DOI

Bibliographic DOI

*Possible with persistent **external** identifiers*

Source dataset #1

Source dataset #2

Source dataset #3

Publish datasets in GBIF

Filter & download

GBIF download

Analyze & publish

Process & archive

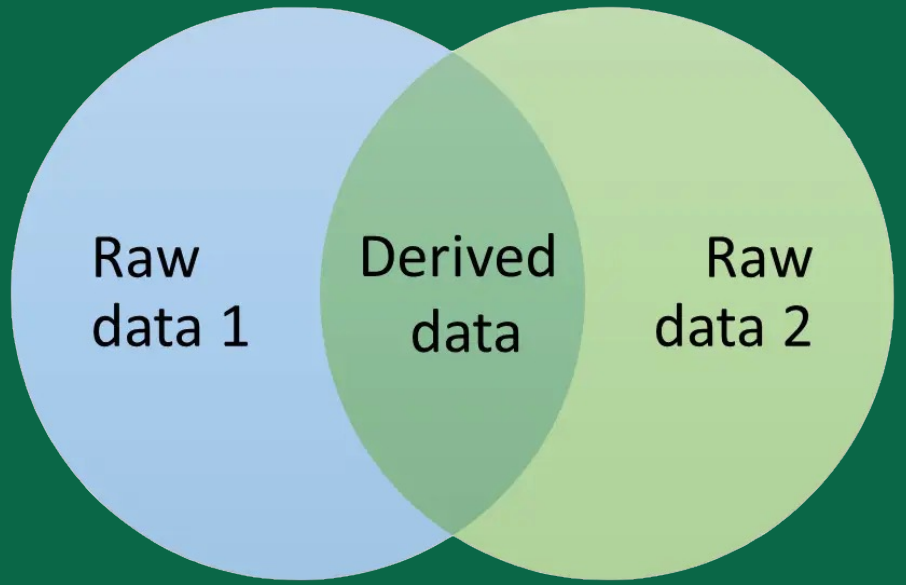
Final state of data

Derived Dataset

Create your own citable dataset DOI

DERIVED DATASET

- a GBIF.org download that has been filtered/reduced significantly, or
- data accessed through a cloud service, e.g. Microsoft AI for Earth (Azure), or
- data obtained by any means for which no DOI was assigned, but one is required (e.g. third-party tools accessing the GBIF search API)
  - *… including rgbif*



HOW TO CREATE A DERIVED DATASET

When created, a derived dataset is assigned a unique DOI that can be used to cite the data.

To create a derived dataset, you will need to authenticate using a GBIF.org account and provide:

- a title of the dataset,
- a list of the GBIF datasets (by DOI or datasetKey) from which the data originated, ideally with counts of how many records each dataset contributed,
- a persistent URL of where the extracted dataset can be accessed,
  - [e.g. Zenodoo, Dataverse, Dryad, …]
- a description of how the dataset was prepared,

- (***optional***) the GBIF download DOI, if the dataset is derived from an existing download , and
- (***optional***) a date for when the derived dataset should be registered if not immediately .

THANK YOU

GBIF Norway

Dag Endresen | Node Manager

helpdesk@gbif.no