

## Global Biodiversity Information Facility

Dag Endresen - GBIF Norway University of Oslo, Natural History Museum



## WHY OPEN RESEARCH DATA?

a (111110010111010101010100100100 ~1989600198119166181981660186188186601981661981& 1110100 0100010 11010010110001000001010\*\* J11011010 1010010 11111100101110101010 . 0101001110 0001010 110001000110111111 d00001011 a1111101011010101010100100 4010010010001000000100110100101001000 ,10105 011050 101 0 5100 100 50 101000 101110. 0101110010111010001001011010010110001 J01001000 001101 1010 1010010111111100101110 40100010101 0101001111 0 000101011100010001101 **#10010930100193100193010093001001101001010010** 2101010010110100 011000101 0 0100100001010001011 )81811191016610661661061660016916001881000189000016011610010106100010010010.





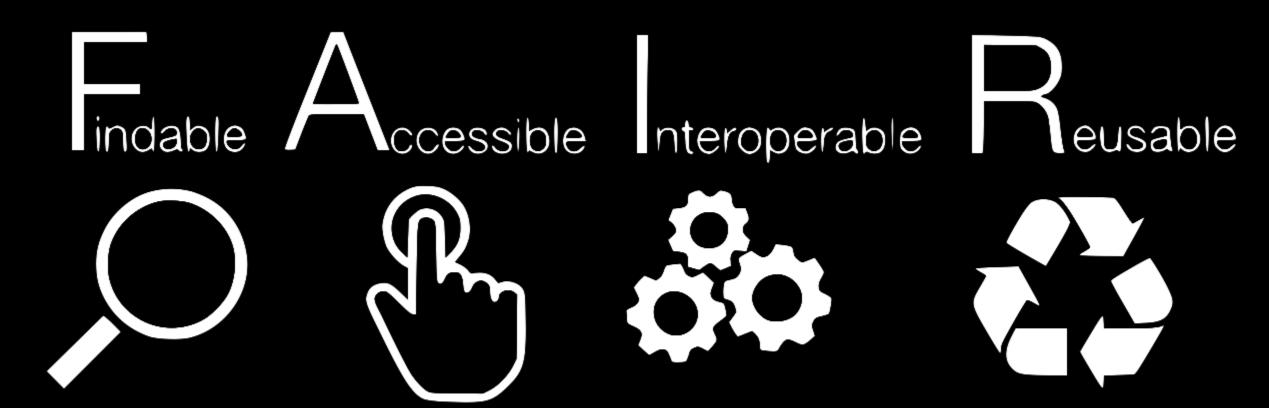
## OPPORTUNITIES

- 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!





## 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







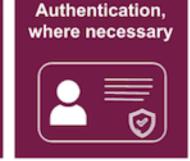






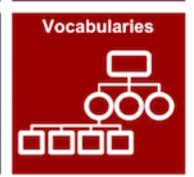




























# Be Findable Accessible Interoperable Reusable

and

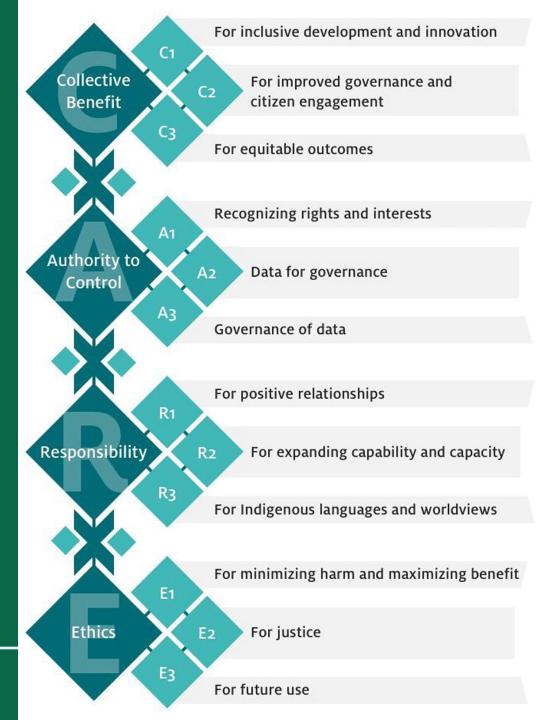
Collective Benefit

Authority to Control Responsibility

**Ethics** 

#### 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.



#### 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

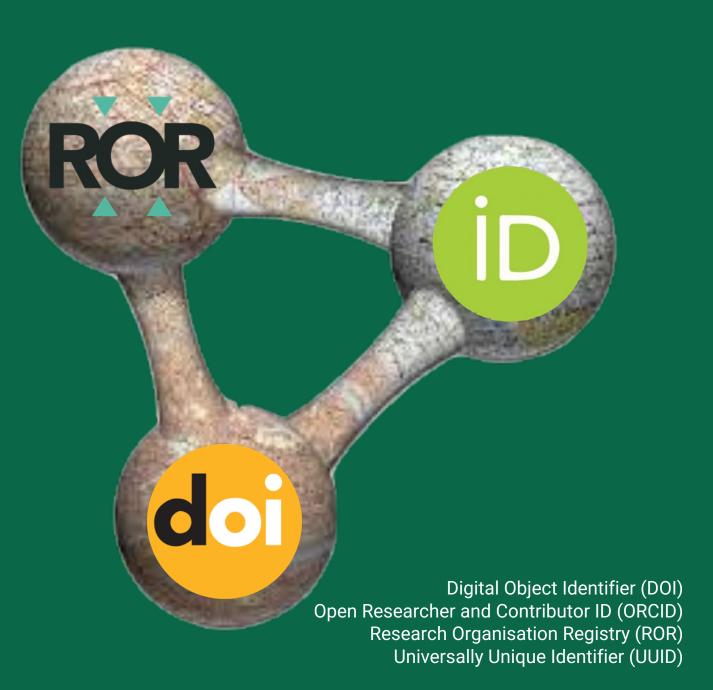
The purpose of identifiers is ... to name things ... making it possible to refer to them





## 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)



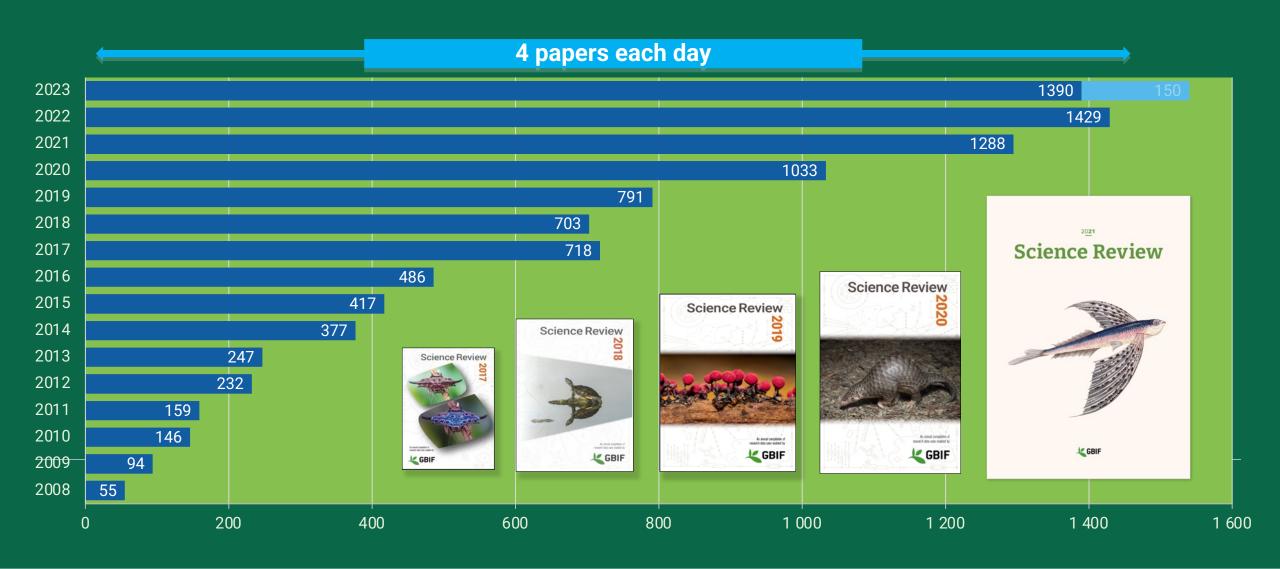


#CiteTheDOI

GBIF started issuing DOIs on 3 February 2015



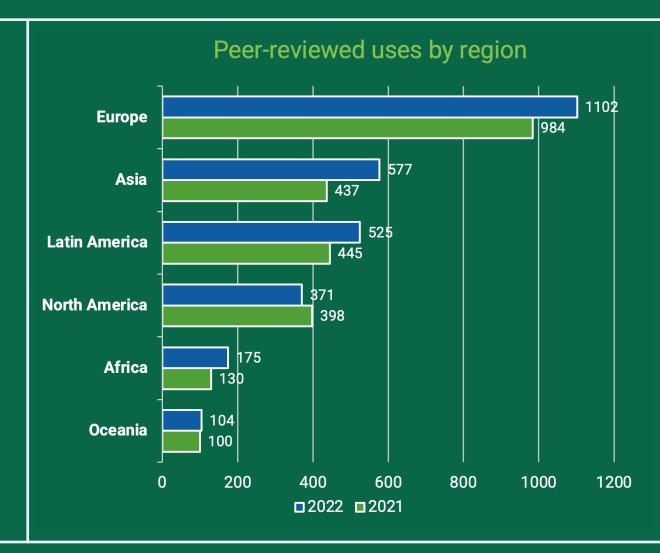
#### PEER-REVIEWED PUBLICATIONS USING GBIF-MEDIATED DATA





#### 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	





#### 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

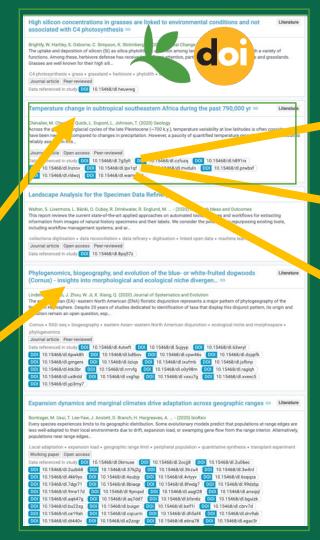


#### DOI BASED DATA CITATION AT GBIF.ORG -- #CITETHEDOI

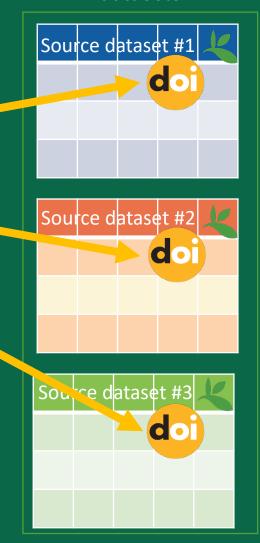
#### paper



#### data citations



#### datasets







Publish

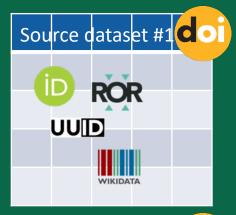
datasets

in GBIF





#### **Dataset DOIs**



Source dataset #2

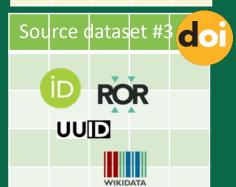
ROR

UUID

WIKIDATA

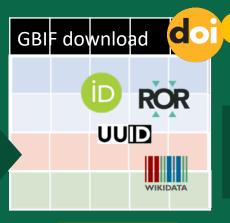
Filter &

download



**Download DOI** 

Bibliographic DOI



Analyze & publish

Final state of data

UUID



Process & archive





Publish

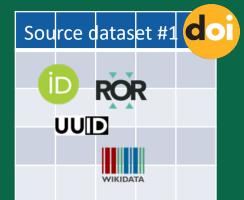
datasets

in GBIF





#### **Dataset DOIs**



Source dataset #2 **ROR** 

UUD

do Source dataset #3

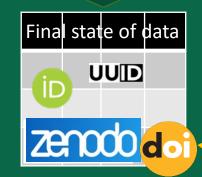


**Download DOI** 

GBIF downlo

Filter & download

> **Process &** archive



Bibliographic DOI











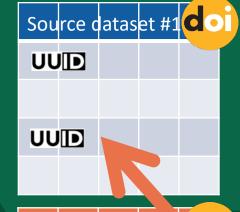


#### **Dataset DOIs**

Download DOI

Bibliographic DOI

Possible with persistent object identifiers



Source dataset #3

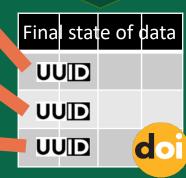


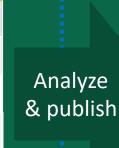
UUD

Filter & download



Process & archive







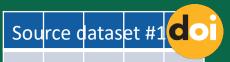








#### **Dataset DOIs**



Source data et #2 Filter & download

Cource dataset #3



blish

datase

in GBIF

#### Download DOI

Bibliographic DOI

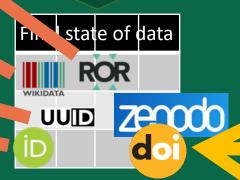
#### Possible with persistent **external** identifiers

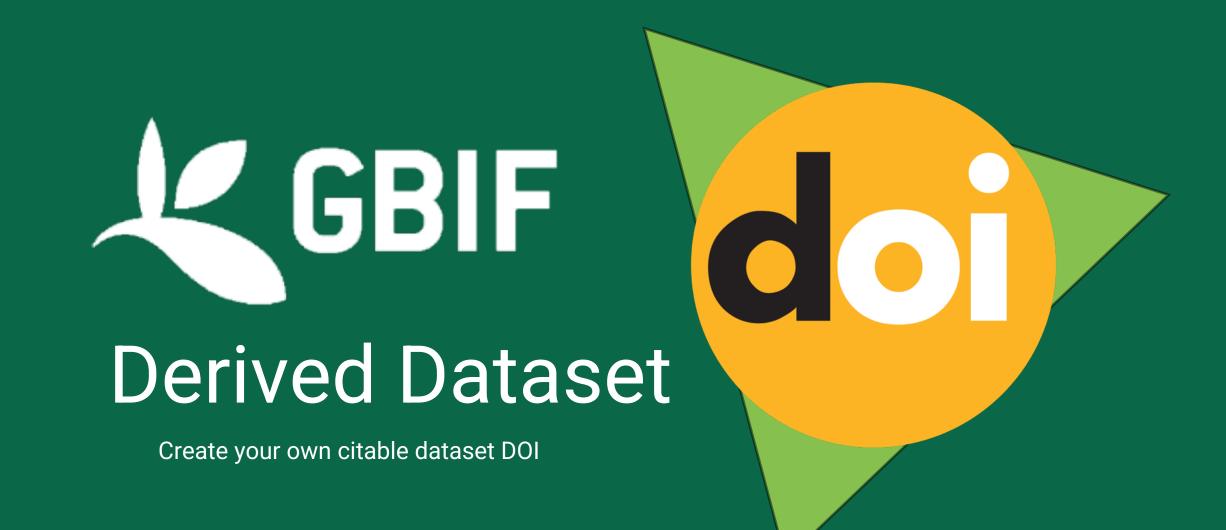


& publish



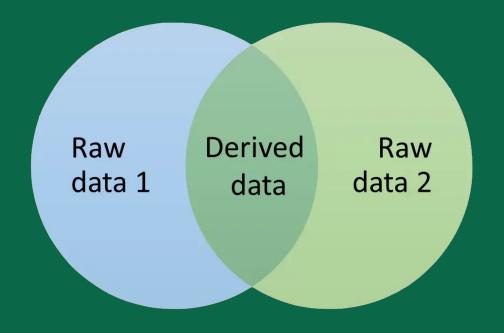
**Process &** archive





#### 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 .





## Global Registry of Scientific Collections

GRSciColl was established at Smithsonian in 2013 Hosting of GRSciColl was transferred to GBIF in 2019

#### **GRSCICOLL**

- The Global Registry of Scientific Collections (GRSciColl) was a community-curated clearing house of colletions information developed by the Consortium of the Barcode of Life (CBOL) launched in 2013.
- Hosting the GRSciColl was transferred to GBIF in 2018 and the upgraded portal came back online in 2019.





w-to

nmunity

About







#### **GBIF** Registry of Scientific Collections

ABOUT

NSTITUTIONS

COLLECTIONS

STAFF MEMBER

#### Institutions

**Institution** refers to any institution or organization that owns and manages scientific collections. This includes herbaria, museums, zoos, botanical gardens, biobanks, among others. An institution may contain multiple collections, in which case those should be entered individually under **Institutional/Project Collections** and linked back to the parent institution.

Norway					Q
10 RESULTS			SUG	GEST A NEW II	NSTITUTION
Code	Name	City/Town	State/Province	Country	Status
UiT	UiT The Arctic University of Norway	Tromsø		Norway	Active
NMBU	Norwegian University of Life Sciences (NMBU)	Ås		Norway	Active
NLH	Agricultural University of Norway	N-1432 Aas-NLH		Norway	Active
MUST	Museum Stavanger	Stavanger		Norway	Active
NINA	Norwegian Institute for Nature Research	Trondheim		Norway	Active
NIBIO	Norwegian Institute of Bioeconomy Research	Ås		Norway	Active
UiO	University of Oslo	Oslo		Norway	Active
MTMU "Balsfjord"	Balsfjord Fjord Museum and Wetland Centre, Midt-Troms Museum	Storsteinnes		Norway	Inactive
IMR	Norwegian Institute of Marine Research	Bergen		Norway	Inactive
NTNU-VM	NTNU Museum of Natural History and Archaeology	Trondheim		Norway	Active

#### **GRSCICOLL**

- 8 117 collection-holding institutions
- 7 010 scientific collections
- 17 000 collection staff members
- https://www.gbif.org/grscicoll













#### dag.endresen

#### **GBIF Registry of Scientific Collections**

#### Collections

Institutional Collections are those that have been formally accessioned into an institution and receive some level of institutional support.

Project Collections are those collected by a researcher associated with an institution that may have not been formally accessioned into its collections. Project Collections may eventually become part of the Institutional Collections or remain under the management and control of the individual researchers.

Personal Collections are under the control of an individual researcher and are not formally associated with or accessioned into an institution. They may be the personal property of a private collector or otherwise orphaned.

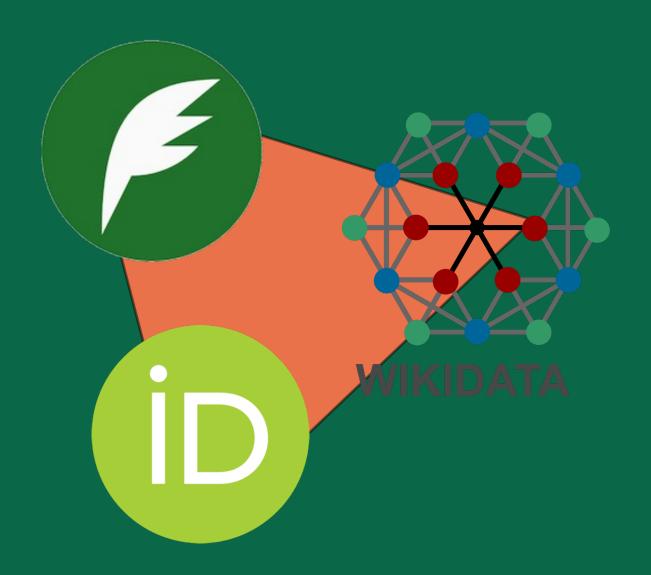
Norway			Q
14 RESULTS			SUGGEST A NEW COLLECTION
Code	Name	Accession status	Status
MA	Mammal collection NTNU University Museum	Institutional	Active
BG-L	Lichen herbarium	Institutional	Active
O-L	Lichen herbarium	Institutional	Active
0-V	Vascular Plant Herbarium		Active
0-F	Mycology Herbarium		Active
O-A	Grethe Hasle Diatoms		Active
DP	DNA Bank Vascular plants		Active
DFL	DNA Bank Fungi and Lichens		Active
0-В	Bryophyte Herbarium		Active
O-A	Algae		Active
DMA	Mammals		Active
ENT	Entomology		Inactive
DOT	DNA Bank Other groups		Inactive
DFH	DNA Bank Fish and Herptiles		Inactive

#### INSTITUTION IDENTIFIERS FOR SOME OF THE MUSEUMS IN NORWAY

Museums	ROR ID	Wikidata	GRSciColl	GBIF publisher institution
Universitetet I Agder (UiA)	03x297z98	Q3375341	UiA	826d1920-7f5c-4091-a84e-668aa2e35b61
Agder naturmuseum og botaniske hage		Q3375341	KMN	
Universitetet I Bergen	03zga2b32	Q204457	UiB	3f3967bf-ecc7-4455-ba89-4e0ab6d6fd3c
University Museum of Bergen		Q301787		
Institutt for naturhistorie		Q11990981		
Helgeland Museum	02gyhy076	Q11057676	Helgeland	a030b53d-7ab0-41dc-8ca7-77f65d5c8157
Rana Museum		Q11997066		
Midt-Troms Museum		Q12327078		
Balsfjord Fjordmuseum og Våtmarkssenter		Q105533121	MTMU Bjalsfjord	9e8e7946-cd17-4c58-81c1-dc8bef359360
Museum Stavanger (MUST)	03bq5ar94	Q19382034	MUST	ecc5cd9e-2d25-4b8d-89c8-a0711eee813b
Universitetet I Oslo (UiO) Naturhistorisk museum i Oslo (NHMO) Botanisk museum (Oslo herbarium - O)	01xtthb56 	Q486156 Q1840963 Q2036576	UiO NHMO O	f314b0b0-e3dc-11d9-8d81-b8a03c50a862  
Randsfjordmuseene AS	-	Q11997108		
Universitetet I Tromsø (UiT)	00wge5k78	Q279724	UiT	689b40c4-ff31-4cd0-83a5-a7a828f1cd92
Tromsø Museum, Universitetsmuseet		Q1686510	TROM	
Varanger Museum	-	Q12009007	-	
Norges Teknisk Naturvitenskapelige Universitet (NTNU)	05xg72x27	Q314536	TRH	a8144f37-5ff7-4137-9400-94b5b2ea4ec4
Vitenskapsmuseet		Q1770886	NTNU-VM	

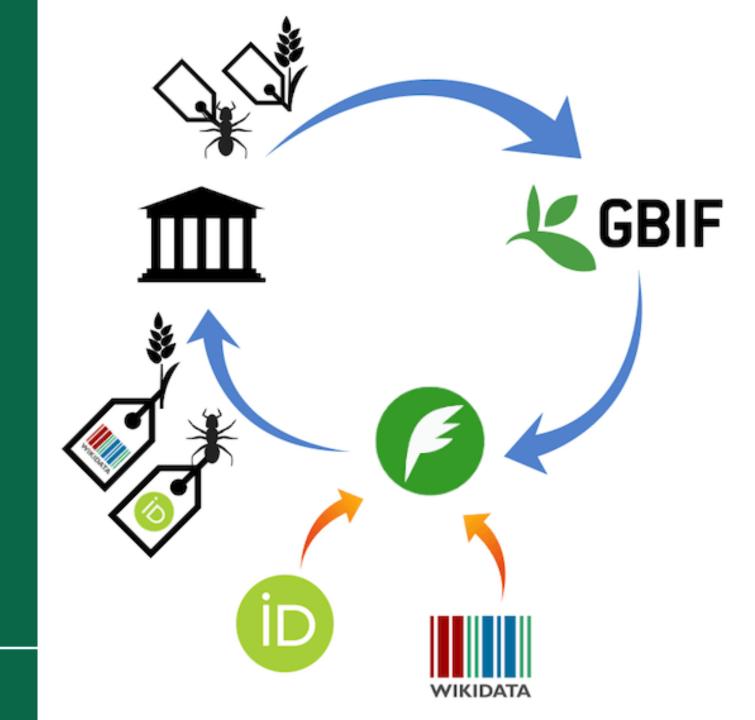


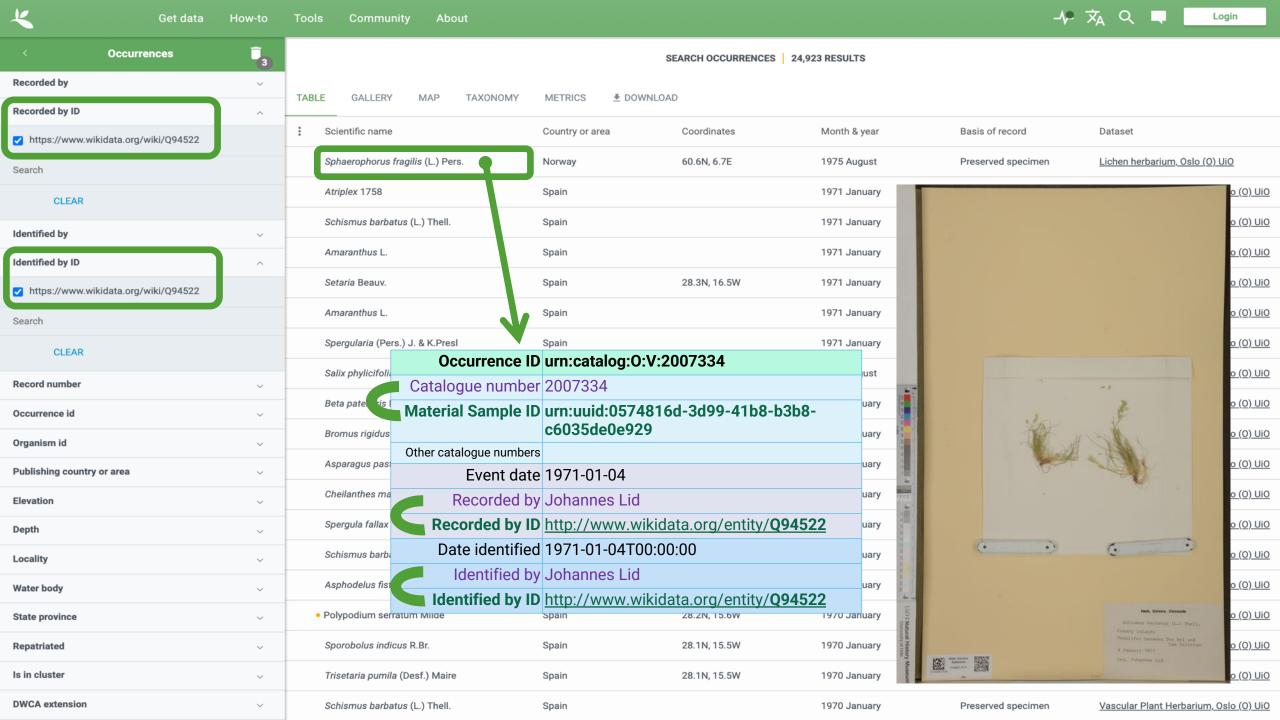
## new tools



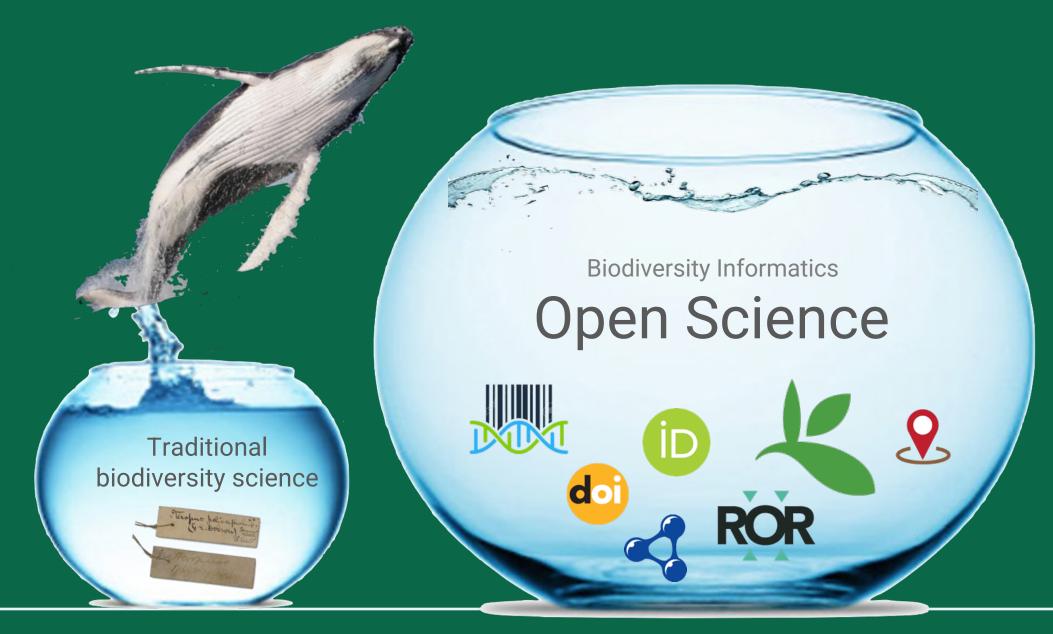
## BIONOMIA

Link natural history specimens to the world's collectors.











## THANK YOU

GBIF Norway

Dag Endresen | Node Manager

helpdesk@gbif.no





