



**biodiversa+**

European Biodiversity Partnership

# Data Management Workshop

For the 2022 – 2023 BiodivMon  
funded projects

---

6 June 2024, 12:45-16:45 CEST  
(break @ 15:00 CEST)



# Welcome words & presentation of the objective of the workshop

*By Rainer Sodtke, Biodiversa+ Co-Chair & Harri Hautala, Science Adviser at AKA*

# Objectives of the workshop

## Projects participating in the workshop:

- 2022 - 2023 2<sup>nd</sup> Biodiversa+ Call (2022 – 2023 [BiodivMon](#)) 33 projects

## Main objective:

- **Exchange on best practices** and come up with updated Data Management Plans (DMPs), in line with Biodiversa+ data policies.

## Specific objectives:

- **Information** - Make sure the projects understand well the principles and concepts around open data & the advantages of having a well-structured and complete DMPs.
- **Support/capacity-building** - Provide support and advice to the projects, hints to develop their DMPs and answer to their identified issues

⇒ For BiodivMonprojects: Please send your updated DMP after the workshop (**deadline: 31st of July**) to Harri Hautala ([harri.hautala@aka.fi](mailto:harri.hautala@aka.fi))

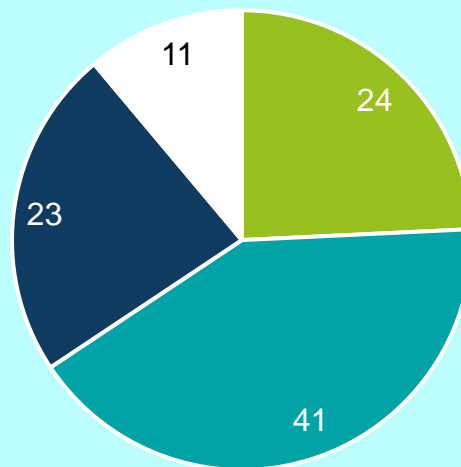
We welcome

# YOU

to the  
Biodiversa+  
Data Management Workshop  
(~ 100 registered people for both days)

June 6, 2024  
12:45 – 16:45 CEST

Current knowledge level on data management



■ Not very experienced ■ Moderate experience ■ Good experience ■ Expert

## Presenters and moderators

- Rainer Sodtke, Biodiversa+ Co-Chair
- Gaby Rerig, Programme Officer, DFG
- Harri Hautala, Science Adviser, AKA
- Céline Billiere, Scientific project officer at ANR
- Sophie Germann, Biodiversa+ operational manager for biodiversity monitoring and research at ANR
- Jennifer Anderson, Consortium Coordinator of project FUNACTION (BiodivProtect)
- Dimitri Brosens, Biodiversity data liaison officer, INBO, Belgium.
- Birgit Gemeinholzer, GBIF Science Committee Chair
- Melissa Liu (GBIF, Asia Support Team)

## Ice-breaker (from 13:10 to 13:30 CEST)



Go to: [www.menti.com](https://www.menti.com)

**Enter:6394 9460**

Data Sharing and Management - A video looking back in 2012



# Data Management Plans: Why scientists should create & sustain them

Jennifer Anderson, Swedish University of Agricultural Sciences  
FUNACTION Coordinator (BiodivProtect)



**Funaction**  
Aquatic fungal  
biodiversity



Funaction  
Aquatic  
fungal  
biodiversity

# FUNACTION — Aquatic FUNgal biodiversity: developing knowledge and strAtegies to inform ConservaTION priorities and measures

[www.funaction.eu](http://www.funaction.eu)



@FUNACTION\_EU



FUNACTION

**Jennifer Anderson (SE)**

Michael Bruun-Nielsen

Olga Vinnere Pettersson

Ziming Wang

**Isabel Fernandes (PT)**

Ronaldo Sousa

Diana Graça

Sérgio Costa

**Andreas Bruder (CH)**

Daniel Romero-Mujalli

Red Calore

**Monika Böhm (US)**

Cátia Canteiro

**Veljo Kisand (EE)**

**Leho Tedersoo (EE)**

Kristel Panksep

Victoria Prins

**Hans-Peter Grossart (DE)**

Alice Retter

Lars Ganzert

Solvig Pinnow

**Laura Garzoli (IT)**

Ester Eckert

Emanuele Ferrari

Diego Fontaneto



REPUBLIC OF ESTONIA  
MINISTRY OF RURAL AFFAIRS



**FORMAS**







FUNaction  
Aquatic  
Fungal  
biodiversity

# FUNACTION — Aquatic FUNgal biodiversity: developing knowledge and strAtegies to inform ConservaTION priorities and measures





FUNACTION  
Aquatic  
Fungal  
biodiversity

# FUNACTION — Aquatic FUNgal biodiversity: developing knowledge and strAtegies to inform ConservaTION priorities and measures

## Foundation:

Produce the knowledge needed for the basis of AF-aware conservation actions.

## Conservation:

Translate that knowledge into conservation planning and monitoring tools.

## Communication:

Build support for AF-aware conservation and produce a useful scientific basis for implementation.



Funaction  
Aquatic  
fungal  
biodiversity

# FUNACTION — Aquatic FUNgal biodiversity: developing knowledge and strAtegies to inform ConservaTION priorities and measures

## Foundation:

Produce the knowledge needed for the basis of AF-aware conservation actions.

## Conservation:

Translate that knowledge into conservation planning and monitoring tools.

## Communication:

Build support for AF-aware conservation and produce a useful scientific basis for implementation.



**WP1**  
**Continental patterns in aquatic fungal biodiversity**



**WP2**  
**Multinational study to test the effectiveness of existing protected areas (PAs)**



**WP3**  
**Systematic conservation planning using aquatic fungi**



**WP4**  
**Developing monitoring of aquatic fungal biodiversity and related ecosystem functions**



**WP5**  
**Effective engagement, communication, and dissemination of information**





Funaction  
Aquatic  
fungal  
biodiversity

# Data

## Samples:

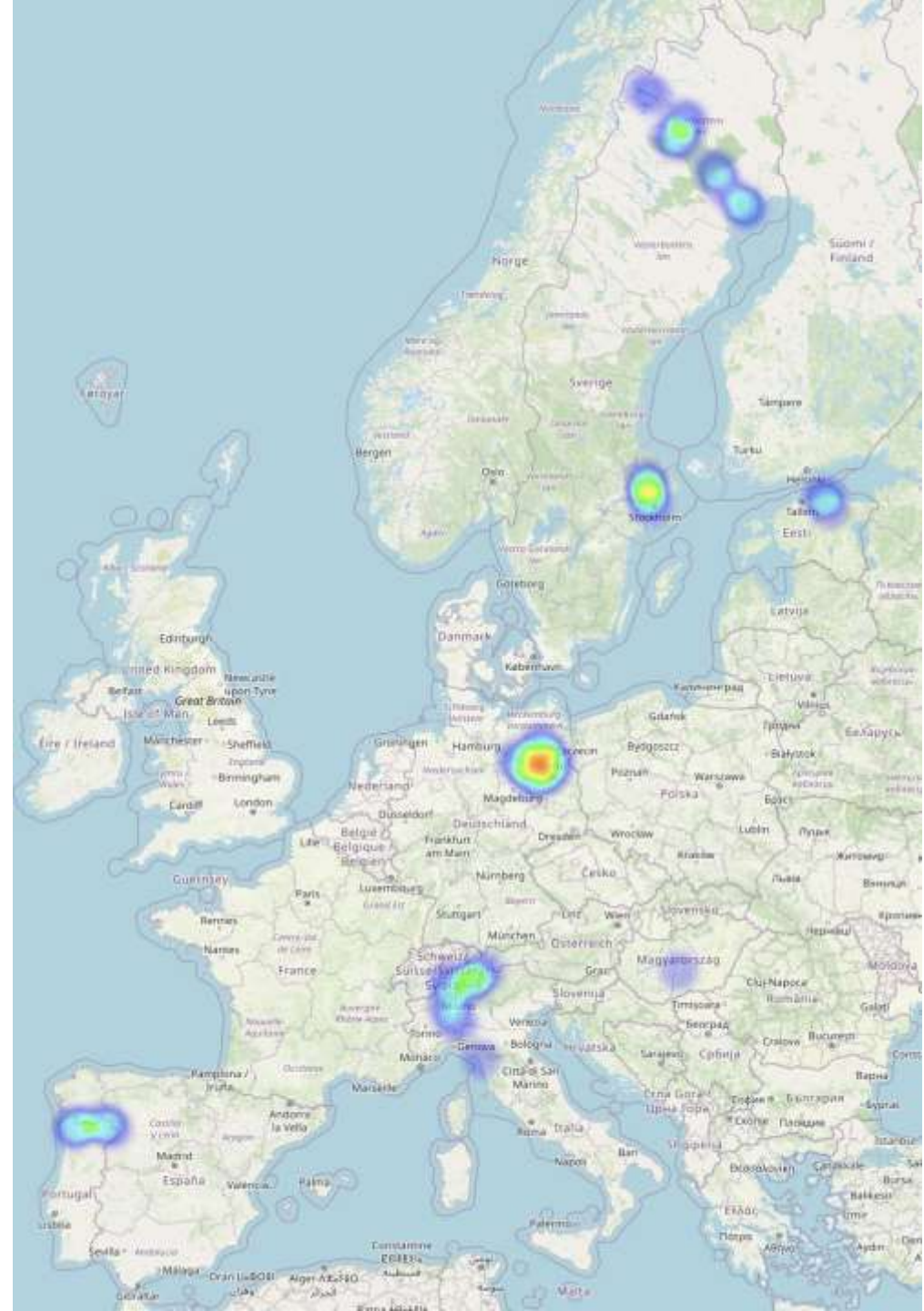
- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

## Primary Scientific Data:

- DNA sequences-  
metabarcodes &  
metagenome
- Water chemistry
- Site metadata

## Secondary Scientific Data:

- Maps
- Models
- Other derivative outputs





Funaaction  
Aquatic  
Fungal  
biodiversity

# Data

## Samples:

- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

## Primary Scientific Data:

- DNA sequences-metabarcodes & metagenome
- Water chemistry
- Site metadata

## Secondary Scientific Data:

- Maps
- Models
- Other derivative outputs

Independent datasets—public

Independent datasets—unpublished

Photos – with people, w/out people

Consortium Meeting Notes

Consortium Documents--other

Consortium Participants--names, contact info, photos, social media accounts, titles, positions, work addresses.

National Funders—contacts, logos

Stakeholder information

Survey responses

Webinar registrations

Event outcomes (who, what, etc...)

Videos

Outreach materials

Media reach

Outputs tracking

Scripts

Readme

Analysis documentation



Funaaction  
Aquatic  
Fungal  
biodiversity

# Data

## Samples:

- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

## Primary Scientific Data:

- DNA sequences-  
metabarcoding &  
metagenome
- Water chemistry
- Site metadata

## Secondary Scientific Data:

- Maps
- Models
- Other derivative outputs

Independent datasets—public

Independent datasets—unpublished

Photos – with people w/out people

Consortium Meeting Notes

Consortium Documents--other

Consortium Participants--names,  
contact info, photos, social media  
accounts, titles, positions, work  
addresses.

National Funders—contacts, logos

Stakeholder information

Survey responses

Webinar registrations

Event outcomes (who, what, etc...)

Videos

Outreach materials

Media reach

Outputs tracking

Scripts

Readme

Analysis documentation





Funaaction  
Aquatic  
Fungal  
biodiversity

# Data Management Plans: Why scientists should create & sustain them

1	Project
8	Partners/teams
7	Countries
6	National funding agencies
25	Team members
5	Work Packages (2 co-leads each)
10	Case study watersheds (WP2)
8	Sampling times
448	Unique collections (WP2)
~2000	Total samples (WP2)

Personal data -- GDPR

Unique raw data

Internal/external documents

Shared/cooperative tasks

Open Science



Funaction  
Aquatic  
fungal  
biodiversity

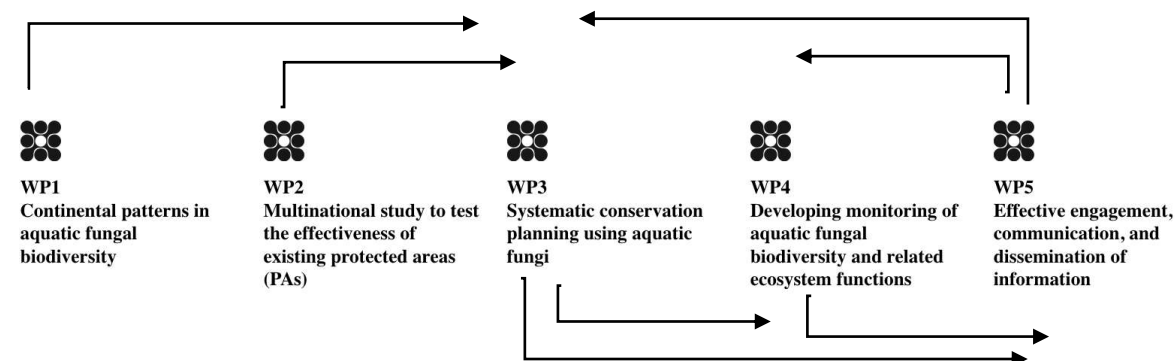
# Data Management Plans: Why scientists should create & sustain them

1	Project
8	Partners/teams
7	Countries
6	National funding agencies
25	Team members
5	Work Packages (2 co-leads each)
10	Case study watersheds (WP2)
8	Sampling times
448	Unique collections (WP2)
~2000	Total samples (WP2)

Personal data -- GDPR  
Unique raw data  
Internal/external documents

Shared/cooperative tasks

Open Science



Spain (water chemistry)

Estonia (extractions, library prep)

Sweden (sequencing)



Funaction  
Aquatic  
Fungal  
Biodiversity

# Data Management Plans: Why scientists should create & sustain them

Avoidance of chaos



Softulka | Shutterstock



Funaction  
Aquatic  
Fungal  
Biodiversity

# Data Management Plans: Why scientists should create & sustain them

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases

(adapted from: Goudeseune L. et al 2023 DOI : 10.5281/zenodo.3448251)







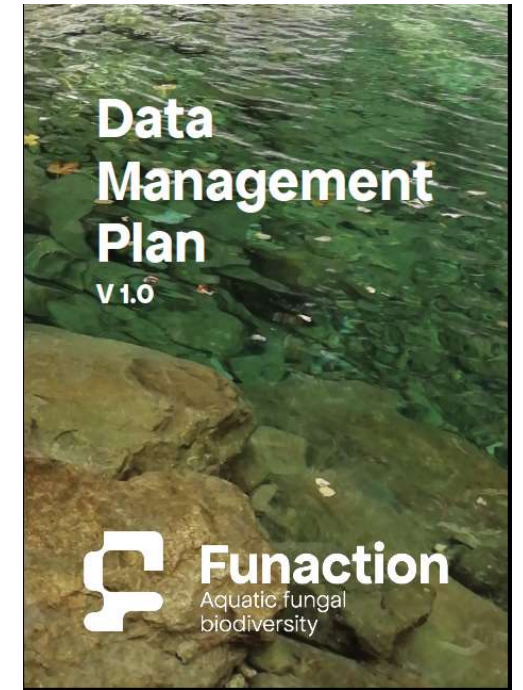
Funaction  
Aquatic fungal  
biodiversity

# Data Management Plan: FUNACTION

## The DMP

- Supports common understanding and expectations among Partners and associated teams for the management of data associated with FUNACTION.
- Serves as a tool to ensure FUNACTION meets research community, partner institution, national and EU standards for data access (inclusive of FAIR principles and deposition of data into appropriate public repositories).
- Provides a mutually agreed framework defining the use, sharing, and storage of data.

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases



Canteiro, C., Fernandes, I., Ferrari, E.,  
Kisand, V., Retter, A., Romero Mujalli, D., &  
Anderson, J. L. (2024). FUNACTION Data  
Management Plan. Zenodo.  
<https://doi.org/10.5281/zenodo.11073579>



Funaction  
Aquatic  
Fungal  
biodiversity

# Data Management Plan: FUNACTION

## Samples:

- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

## Primary Scientific Data:

- DNA sequences-  
metabarcoding &  
metagenome
- Water chemistry
- Site metadata

## Secondary Scientific Data:

- Maps
- Models
- Other derivative outputs

Independent datasets—public

Independent datasets—unpublished

Photos – with people w/out people

Consortium Meeting Notes

Consortium Documents--other

Consortium Participants--names,  
contact info, photos, social media  
accounts, titles, positions, work  
addresses.

National Funders—contacts, logos

Stakeholder information

Survey responses

Webinar registrations

Event outcomes (who, what, etc...)

Videos

Outreach materials

Media reach

Outputs tracking

Scripts

Readme

Analysis documentation





Funaction  
Aquatic fungal  
biodiversity

# Data Management Plan: FUNACTION

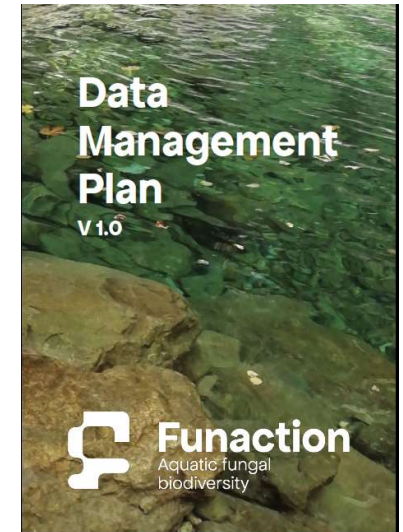
## Sample naming conventions

= persistent unique identifiers

**Table 5.2** Components of the USID and ST and their accepted values.

	Unique Site ID (USID)			Specific Sample types (ST)			
	Broad geographic ID	Site ID (SID)	Sampling occasion (SO)	Sterivex filter	Chemistry (water for chemical analysis)	Sediment	Litter
FUNACTION (WP2)	WC Table 5.3	## 01-30	T1-T8	W1, W2	C1, C2	S	L
FunAqua (WP1)	CC Table 5.4	## 01-30	T1-T8	W1	C1, C2	S	L

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases





Funaction  
Aquatic fungal  
biodiversity

# Data Management Plan: FUNACTION

## Sample naming conventions

= persistent unique identifiers

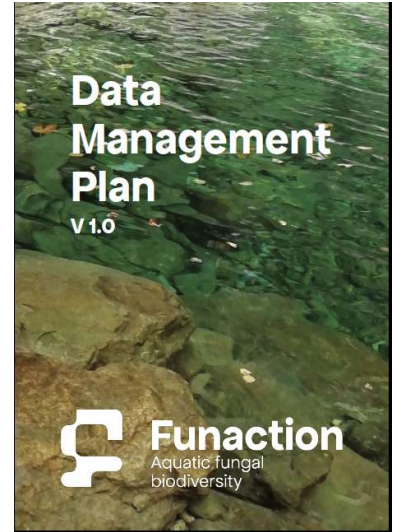
+ flexible for use throughout project

LU09T1\_W1\_MB1 = DNA sample relating to sample from above for barcode “1”

LU09T1\_W1\_ MB1\_raw = Raw sequence data from sample

LU09T1\_1-wWWW-xy = USID + number to identify unique photo file + w[initials to identify people pictured in photo if applicable] + initials of photographer for photo credit

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases





Funaction  
Aquatic  
Fungal  
biodiversity

# Data Management Plan: FUNACTION

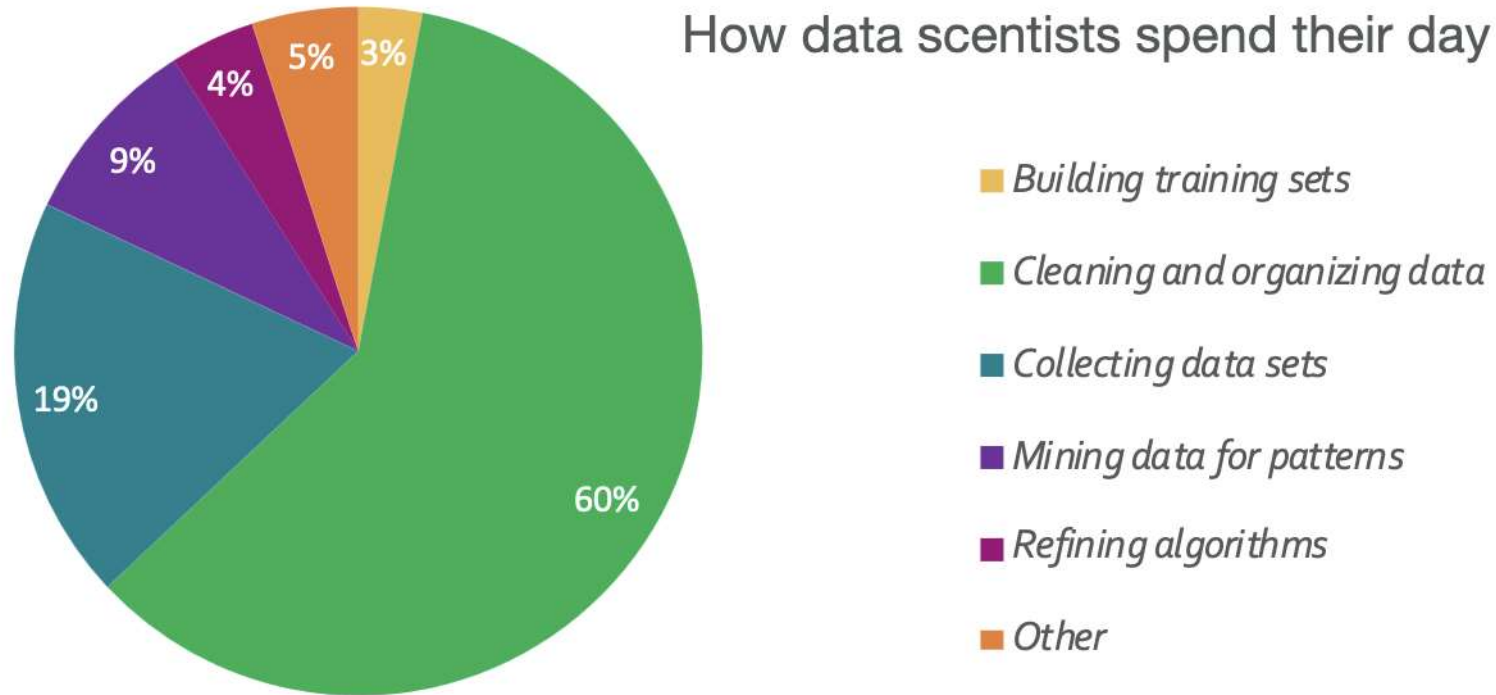


Figure 2: How data scientists spend their day (redrawn after [CrowdFlower, 2016](#))

Goudeseune L. et al 2023 DOI : 10.5281/zenodo.3448251

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases



Funaction  
Aquatic  
Fungal  
biodiversity

# Data Management Plan: FUNACTION

## Standardized terminologies

## Centralized data collection

- = Uniform, correctly spelled/typed/etc, terminology
- = Standard formats
- = Unified dataset from the start, across the countries
- = Transparency
- = Availability

- Efficiency
- Structured data collection & storage
- Reduce risk of loss
- Consistency within/among teams and continuity over time
- Promotes share/re-use of data and reproducibility
- Longevity – supports deposits to public databases

**FUNACTION\_v1**

▼ **Determine the Unique Site ID (USID) and declare the samples taken**

\* **Project**

☐ FUNACTION

☐ FunAqua

\* **Country**

\* **Watershed**

\* **Unique Site ID (USID)**

FUNACTION: WC+SID+SO; FunAqua: CC+SID+SO. WC & CC find from previous questions. SID = number assigned to a specific site and always re-used for that specific site. SO = Sampling occasion. If a site is visited for the first time and/or the only time, it gets T1. Example L009T1

\* **Sample types taken**

Select all that apply. NOTE: this will be used to trace samples—so double check you are accurate.

☐ Sterivex (label as USID-W1)

☐ Sterivex (label as USID-W2)

☐ Water for chemical analysis (label as USID-C1) — DO NOT filter

☐ Water for chemical analysis (label as USID-C2) — Filter this sample through 0.45 µm

☐ Sediment (label as USID-S)

☐ Litter/Organic matter (label as USID-L)

**Volume filtered for Sterivex filter (ml) -W1**

**Volume filtered for Sterivex filter (ml) -W2**

Kobotoolbox



Funaction  
Aquatic fungal  
biodiversity

# Data Management Plan: FUNACTION

Switzerland	Daniel Romero Mujalli (Andreas Bruder)
Germany	Alice Retter
Estonia	Veljo Kisand
Italy	Emanuele Ferrari
Portugal	Isabel Fernandes
USA	Catia Canteiro







SCIENCE AND  
EDUCATION **FOR**  
**SUSTAINABLE**  
**LIFE**



# Menti for the attendants

(from 14:00 to 14:15 CEST)



Go to: [www.menti.com](https://www.menti.com)

**Enter: 6394 9460**

# *Thought experiment: What do people do with stuff they don't need and why?*

*No one else can or shall benefit from it.*



**Type A:** Throw it away / destroy it

*Maybe it will be useful for me later, better keep it.*



**Type B:** Store it in their cellar for long

*Fine, if someone takes it, but I don't invest (much) effort.*



**Type C:** Give it to the bulky waste

*I like the idea that my old things will still be of use.*



**Type D:** Bring it to the social store

*It depends on my current situation and the stuff.*



**Type E:** No fixed type

***Transfer to data (beyond requirements for publication): Is it like clearing out your flat? – Researchers' willingness to provide their data & to make their data FAIR (Findable, Accessible, Interoperable, Reusable):***

**A:** No open data at all

**B:** Open data, but after a long embargo

**C:** Fine with open data, but data is not (very) FAIR

**D:** All data (perfectly) FAIR

**E:** Behavior strongly depends on data

# Open Science and Open Data

main principles and concepts

RESEARCH INSTITUTE  
NATURE AND FOREST



Dimitri Brosens  
Coördinator Flemish Biodiversity Portal  
GBIF Belgium



[0000-0002-0846-9116](tel:0000-0002-0846-9116)  
[Dimitri.brosens@inbo.be](mailto:Dimitri.brosens@inbo.be)

2024/06/06

# Open science lab for biodiversity

We make biodiversity research open and reproducible



[ABOUT](#) • [OUR SERVICES](#) • [RESOURCES](#) • [NEWS](#) • [EVENTS](#) • [GENERAL INFO](#) • [CONTACT](#)



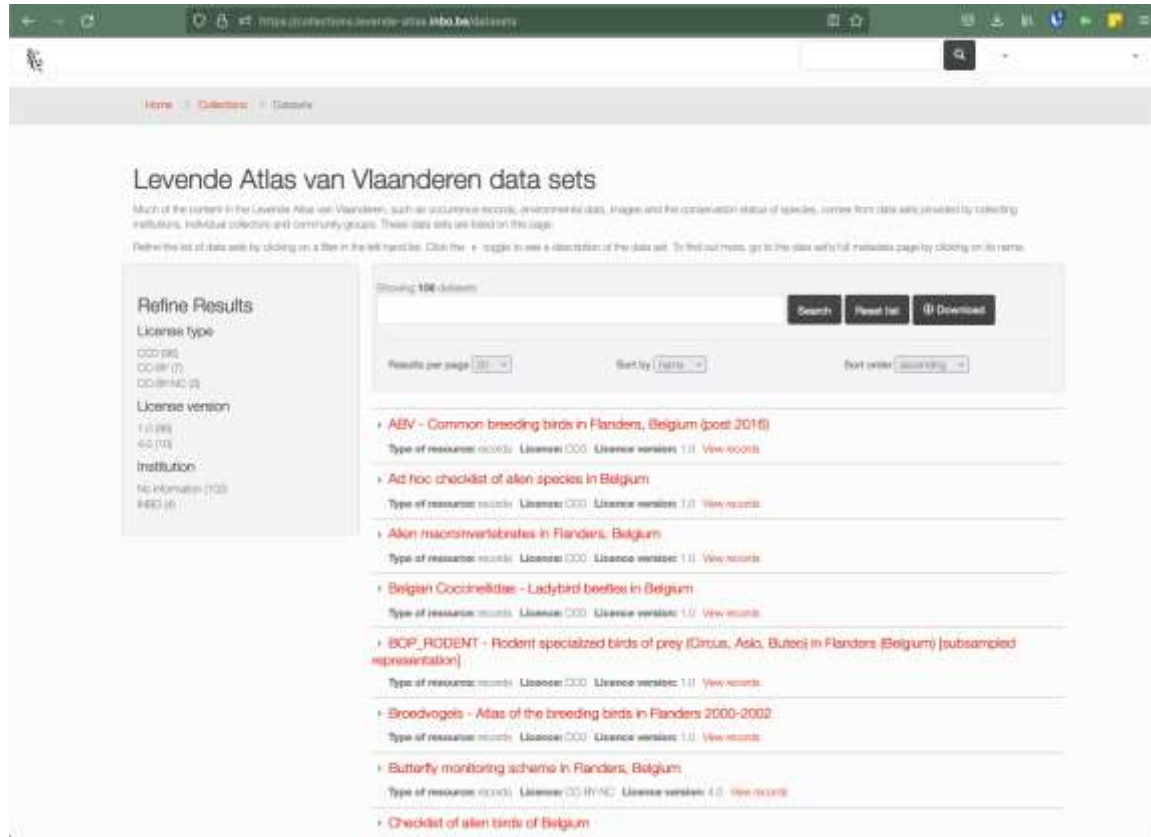
Biodiversity.be

**BELGIAN BIODIVERSITY PLATFORM**  
FOR SCIENCE, POLICY, AND PRACTICE





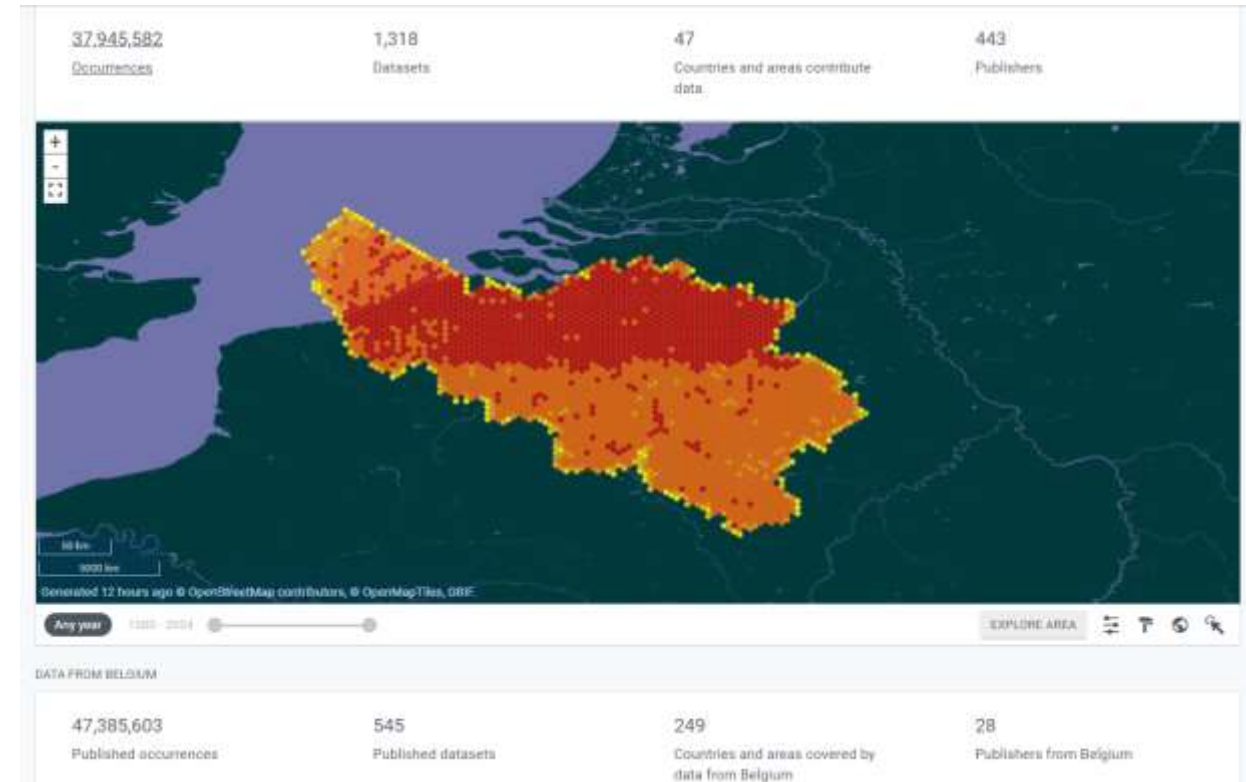
# Flemish Living Atlas project: Vlaams Biodiversiteitsportaal



The screenshot shows the 'Levende Atlas van Vlaanderen data sets' page. It features a search bar at the top, a 'Refine Results' sidebar on the left, and a list of data sets on the right. The data sets include:

- ABV - Common breeding birds in Flanders, Belgium (post 2016)
- Ad hoc checklist of alien species in Belgium
- Alien macroinvertebrates in Flanders, Belgium
- Belgian Coccinellids - Ladybird beetles in Belgium
- BOP\_RODENT - Rodent specialized birds of prey (Circus, Asio, Buteo) in Flanders (Belgium) [subsampling representation]
- Broodvogels - Atlas of the breeding birds in Flanders 2000-2002
- Butterfly monitoring scheme in Flanders, Belgium
- Checklist of alien birds of Belgium

# Belgian Biodiversity Platform: Belgian GBIF Node



Theme 1: Innovation and harmonisation of methods and tools for collection and management of biodiversity monitoring data

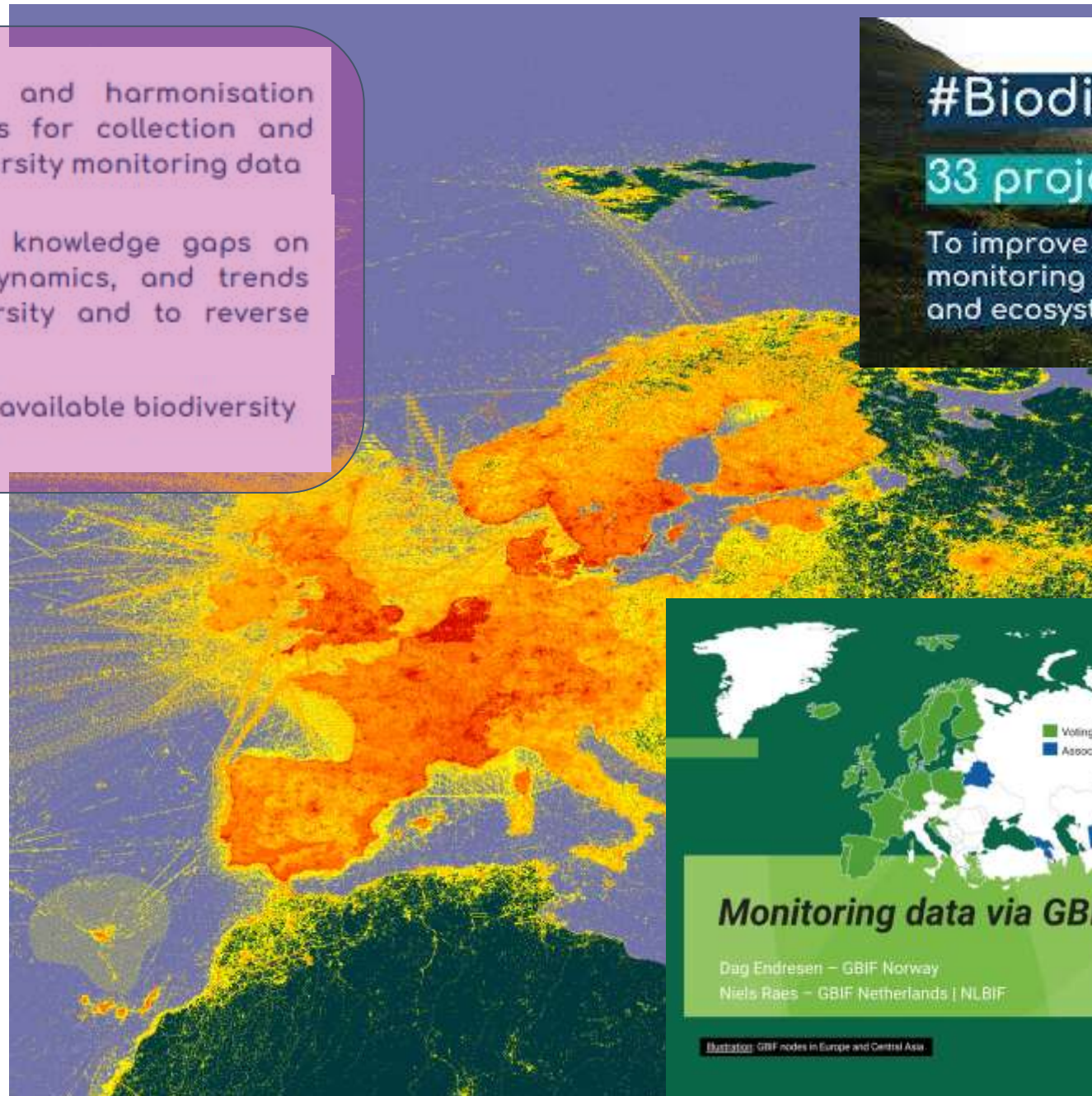
Theme 2: Addressing knowledge gaps on biodiversity status, dynamics, and trends to safeguard biodiversity and to reverse biodiversity loss

Theme 3: Making use of available biodiversity monitoring data

#BiodivMon

33 projects funded

To improve transnational monitoring of biodiversity and ecosystem change



Monitoring data via GBIF

Dag Endresen – GBIF Norway

Niels Raes – GBIF Netherlands | NLBIF

Voting Participants  
Associate Participants



Illustration: GBIF nodes in Europe and Central Asia

Biodiversa monitoring hub, Helsinki | 16<sup>th</sup> April 2024





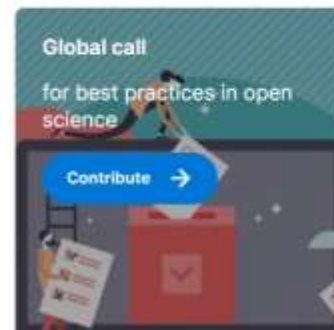
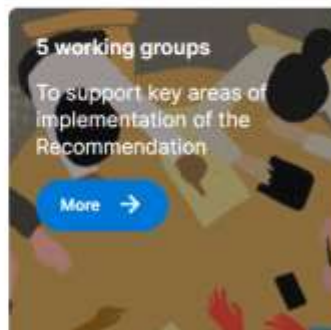






# UNESCO Recommendation on Open Science

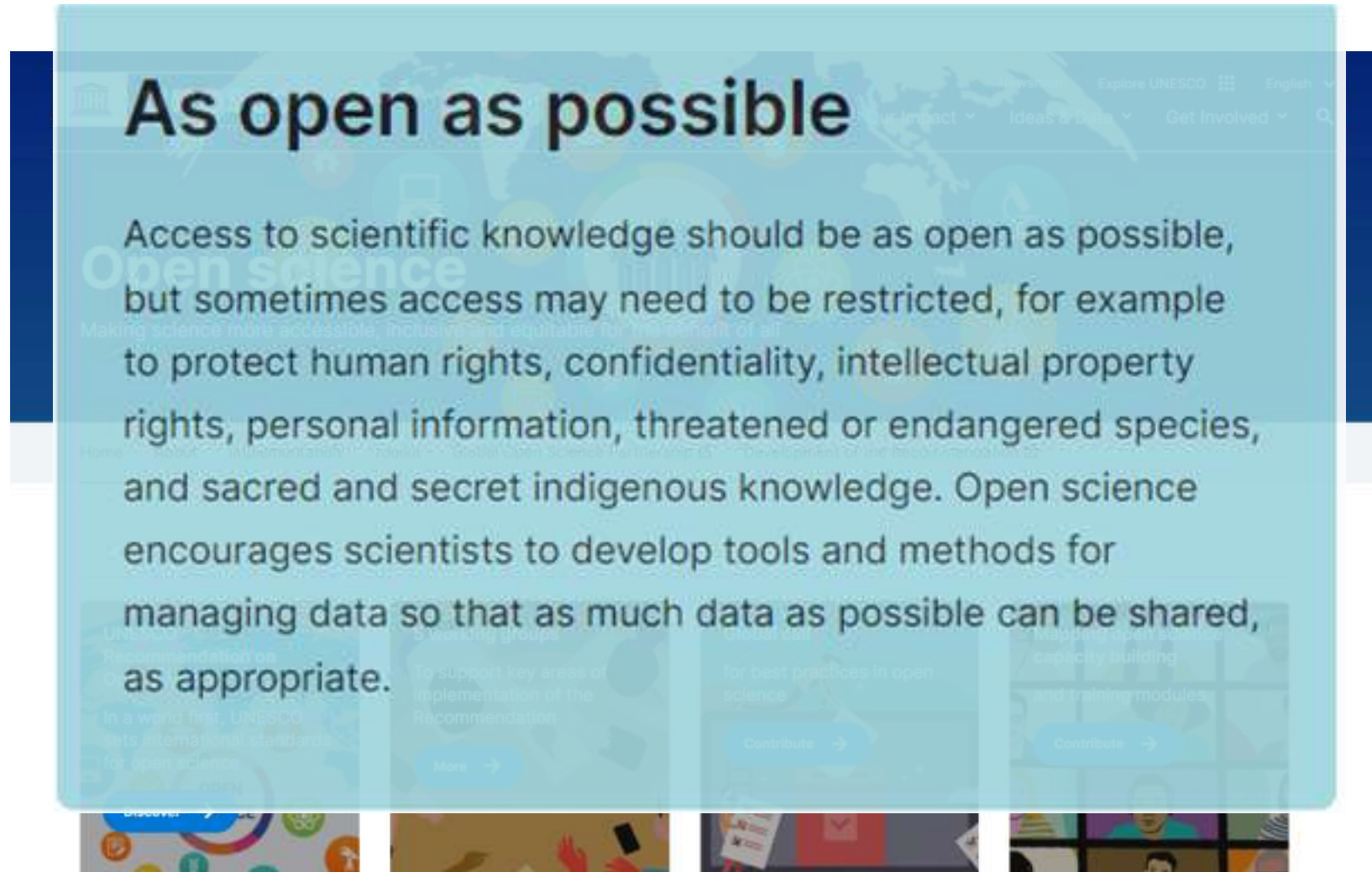
The UNESCO Recommendation on Open Science is the first international standard setting instrument on open science.





# UNESCO Recommendation on Open Science

The UNESCO Recommendation on Open Science is the first international standard setting instrument on open science.







## &gt; Science, technology and innovation policy

## &gt; Industry and globalisation

## &gt; Emerging technologies

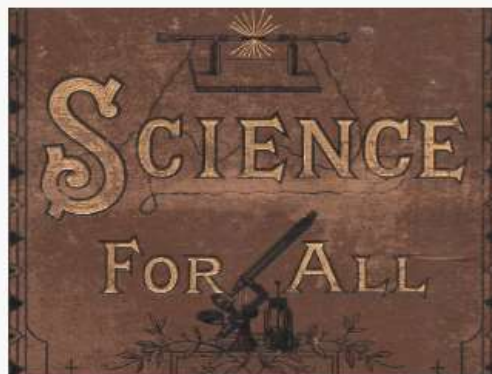
## &gt; Digital economy

## &gt; Broadband and telecom

## &gt; Consumer policy

## Open Science

Open science encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives. [Broadening access to scientific publications and data is at the heart of open science](#), so that research outputs are in the hands of as many as possible, and potential benefits are spread as widely as possible:



- Open science promotes a more accurate verification of scientific results. By combining the tools of science and information technologies, scientific enquiry and discovery can be sped up for the benefit of society.
- Open science reduces duplication in collecting, creating, transferring and re-using scientific material.
- Open science increases productivity in an era of tight budgets.
- Open science results in great innovation potential and increased consumer choice from public research.
- Open science promotes citizens' trust in science. Greater citizen engagement leads to active participation in scientific experiments and data collection.

The OECD is working with member and non-member economies to review policies to promote open science and to assess their impact on research and innovation.

---

### RECOMMENDATION OF THE OECD COUNCIL CONCERNING ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

The OECD Council adopted a [revised Recommendation on Access to Research Data from Public Funding](#) in January 2021. This legal instrument, in force since 2006, has been updated to address new technologies and policy developments.

---

### REPORTS AND POLICY NOTES



OECD Policy Responses to Coronavirus (COVID-19)

# Why open science is critical to combatting COVID-19

Updated 12 May 2020

[Open PDF](#)

## Disclaimer

Search



## Key messages

### Achievements of open science initiatives and commitments

### Remaining challenges

### The way forward: Resilient open science policies

### Further reading

## Key messages

- In global emergencies like the coronavirus (COVID-19) pandemic, open science policies can remove obstacles to the free flow of research data and ideas, and thus accelerate the pace of research critical to combating the disease.
- While global sharing and collaboration of research data has reached unprecedented levels, challenges remain. Trust in at least some of the data is relatively low, and outstanding issues include the lack of specific standards, co-ordination and interoperability, as well as data quality and interpretation.
- To strengthen the contribution of open science to the COVID-19 response, policy makers need to ensure adequate data governance models, interoperable standards, sustainable data sharing agreements involving public sector, private sector and civil society, incentives for researchers, sustainable infrastructures, human and institutional capabilities and mechanisms for access to data across borders.

## Research and innovation

[Home](#) > [Strategy on research and innovation](#) > [Strategy 2020-2024](#) > [Our digital future](#) > [Open Science](#)

### Open Science

An approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Expert groups, publications, news and events.

#### PAGE CONTENTS

[The EU's open science policy](#)

[8 ambitions of the EU's open science policy](#)

[Future of open science under Horizon Europe](#)

[Tracking open research trends - Open Science Monitor](#)

[Documents](#)

[Latest](#)

[Events](#)

### The EU's open science policy

Open science is a policy priority for the European Commission and the standard method of working under its research and innovation funding programmes as it improves the quality, efficiency and responsiveness of research.

When researchers share knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.

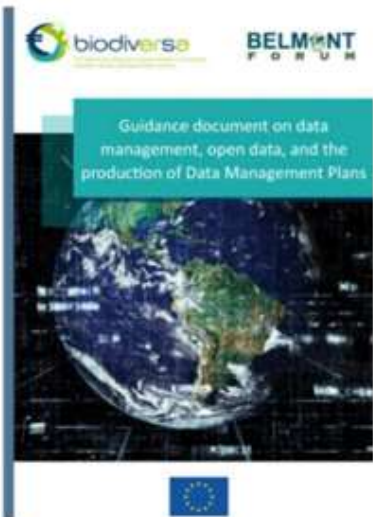
And when partners from across academia, industry, public authorities and citizen groups are invited to participate in the research and innovation process, creativity and trust in science increases.

That is why the Commission requires beneficiaries of research and innovation funding to make their publications available in open access and make their data as open as possible and as closed as necessary. It recognises and rewards the participation of citizens and end users.

Furthermore, the [European Open Science Cloud](#) <sup>(EN|...)</sup> will enable researchers across disciplines and countries to store, curate and share data.



# Guidance document on data management, open data, and the production of Data Management Plans



Biodiversa+ encourages open sharing of research data and digital outputs to stimulate new approaches to the collection, reuse, analysis, validation and management of data and information, thus increasing the transparency of the research process and robustness of the results.

- Applicants are thus requested to submit
  - Preliminary data management information in their pre-proposals
  - A first **Data Management Plan** in their full proposals

To help you, consult our guidance document & learn more about

- Main principles and policies for data management
- Existing tools and resources
- & Get advice on the writing of your data management plans

# Biodiversa+ and Open Science

## 1. Promotes Accessibility and Sharing of Data

- Biodiversa+ advocates for open access to research data.
- Ensures data from biodiversity studies are freely available.
- Fosters further research and innovation.

## 2. Encourages Collaborative Research

- Open Science emphasizes cross-disciplinary collaboration.
- Biodiversa+ supports international, interdisciplinary projects.
- Addresses complex biodiversity challenges.





# Biodiversa+ and Open Science

## 3. Transparency in Research Processes

- Both Biodiversa+ and Open Science value transparency.
- Biodiversa+ promotes publication of research protocols, methodologies, and data sets.
- Enhances credibility and reliability of scientific findings.

## 4. Public Engagement and Citizen Science

- Biodiversa+ involves citizen science.
- Public participates in data collection and monitoring.
- Aligns with Open Science's goal of involving non-scientists.



# Biodiversa+ and Open Science

## 5. Funding and Policy Support

- Biodiversa+ funds projects adhering to Open Science principles.
- Encourages researchers to adopt Open Science practices.
- Promotes openness and accessibility.

## 6. Utilization of Open Access Journals and Repositories

- Biodiversa+ research often published in open access journals.
- Ensures findings are freely available.
- Reduces barriers to information dissemination.



# 89 ambitions of the EU's open science policy

- Open Data
  - FAIR and Open Data
- European Open Science Cloud (EOSC)
  - Store, share, process and reuse research digital objects
- New generation metrics
  - New indicators
- Alternative metrics and rewards
  - Research quality and impact
- Future of scholarly communication
  - Open access policy

The EU's open science policy

8 ambitions of the EU's open science policy

Future of open science under Horizon Europe

Tracking open research trends  
- Open Science Monitor

Documents

Latest

Events

## 9 ambitions of the EU's open science policy

- Rewards
  - Recognise the diverse outputs, practices and activities
- Research integrity and reproducibility of scientific results
  - results of research and innovation activities should be reproducible
- Education and skills
  - All scientists in Europe should have the necessary skills and support to apply open science research routines and practices.
- Citizen science
  - The general public should be able to make significant contributions

The EU's open science policy

8 ambitions of the EU's open science policy

Future of open science under Horizon Europe

Tracking open research trends  
- Open Science Monitor

Documents

Latest

Events





## Transform to Open Science

Transform to Open Science (TOPS) is a \$40 million, 5-year mission, led by NASA's Science Mission Directorate's Open-Source Science initiative.



### Open science lab for biodiversity

The [Open science lab for biodiversity](#) is an externally funded team at the Research Institute for Nature and Forest (INBO). It offers technical support to researchers of projects in which it participates (such as the Belgian Biodiversity Platform, LifeWatch, TriAS and GloBAM). This support is mainly focused on **open data publication** and **research software development**. Its approach is open by default, international, and community-oriented, with the goal of making biodiversity research more efficient and reproducible.

Indexed in

OpenAIRE

### Publication date:

February 24, 2023

### DOI:

DOI [10.5281/zenodo.7672934](https://doi.org/10.5281/zenodo.7672934)

### Keyword(s):

[open science](#)

### License (for files):

[Creative Commons Attribution 4.0 International](#)



eosc

European Open Science Cloud



*Potential spearhead: INBO: **integrated data management and retrieval***

- *How can we do more with long-term data and repurpose historical data for research?*
- *Can we develop data networks?*
- *How do we put more effort into European cooperation to align biodiversity monitoring?*
- *How do we encourage the application of open science objectives within INBO?*





## What is open science?

- Open Science is not different to traditional science.
- It just means that you carry out your research in a more **transparent** and **collaborative** way.
- Open Science applies to **all research disciplines**.

- understand what Open Science means and why **you should care** about it
- be aware of some of **the different ways** to go about making your own research more open over the research lifecycle
- understand why **funding bodies** are in support of Open Science and what their basic requirements are
- be aware of the **potential benefits of practicing open science**

### Introductory Course

FOSTER consortium. (2018, November). What is Open Science? (Version 1.0). Zenodo.  
<http://doi.org/10.5281/zenodo.2629946>

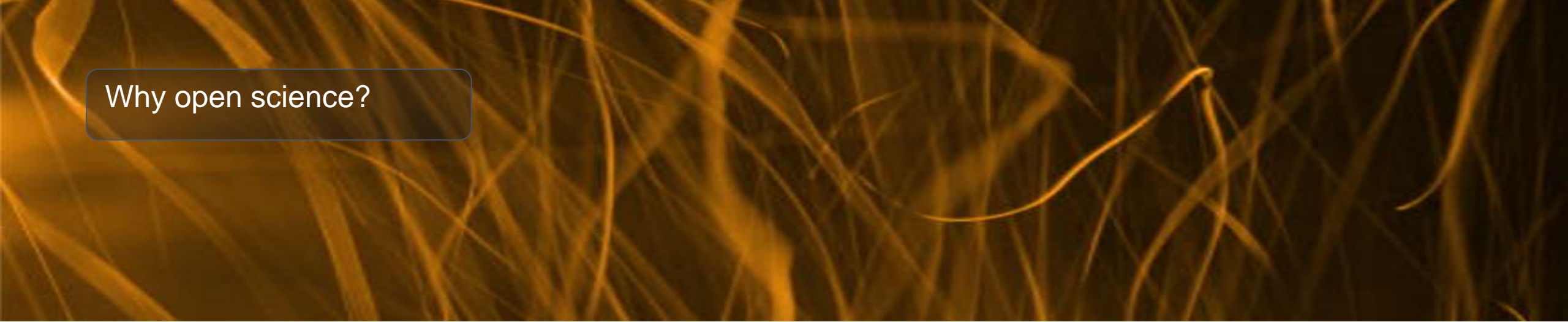




What is open science?

...the practice of science in such a way that others can **collaborate** and **contribute**...

where research data, lab notes and other research processes are **freely available**, under terms that **enable** reuse, redistribution and reproduction of **the research and its underlying data and methods**.

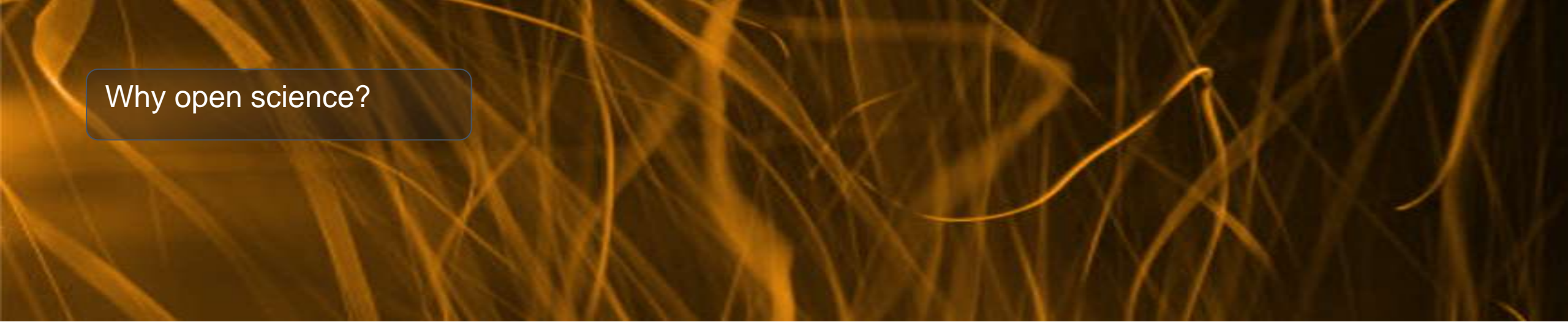


## Why open science?

....Opening the research process supports **validation**,  
**reproducibility** and reduces cases of academic misconduct....

It helps to maximise the **impact** of your research and provides  
the foundations for others to build upon.

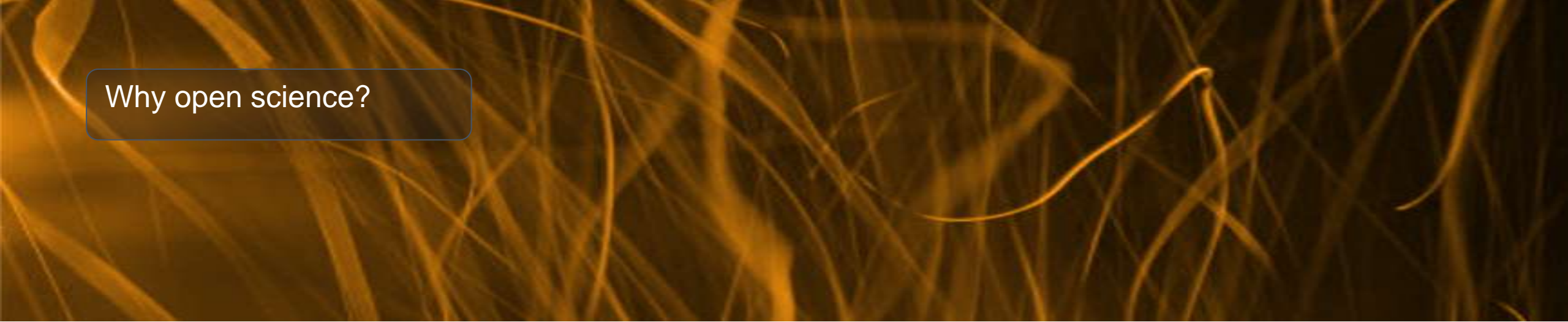
applying open science in your daily workflows is just part of good research practice!



## Why open science?

In a nutshell, Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks (Vicente-Sáez & Martínez-Fuentes 2018)

**applying open science in your daily workflows is just part of good research practice!**



Why open science?

**Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics**

**<https://doi.org/10.1093/joc/jqac017>**

You do not currently have access to this article.

**applying open science in your daily workflows is just part of good research practice!**



## Why open science?



## Good for Research

Practicing open science means that research **outputs are accessible to all** – not stuck behind pay walls. This helps to ensure that all researchers, and other stakeholders, have access to information regardless of their location or economic situation. It means that the **research process can be accelerated** and new **knowledge can be more quickly generated and built upon** to help **solve grand challenges**.

## Why open science?

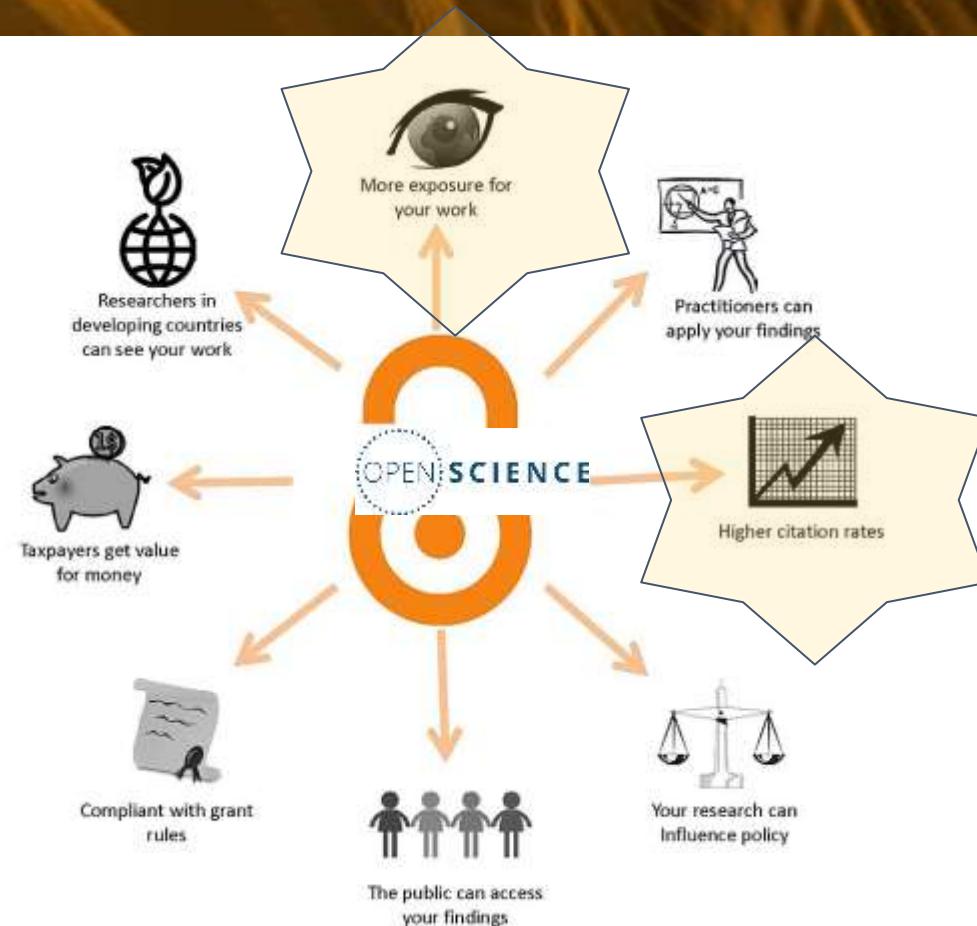


Good for Society

Open science offers a **better return on investment** from research funded by public money and contributes to better economic growth.



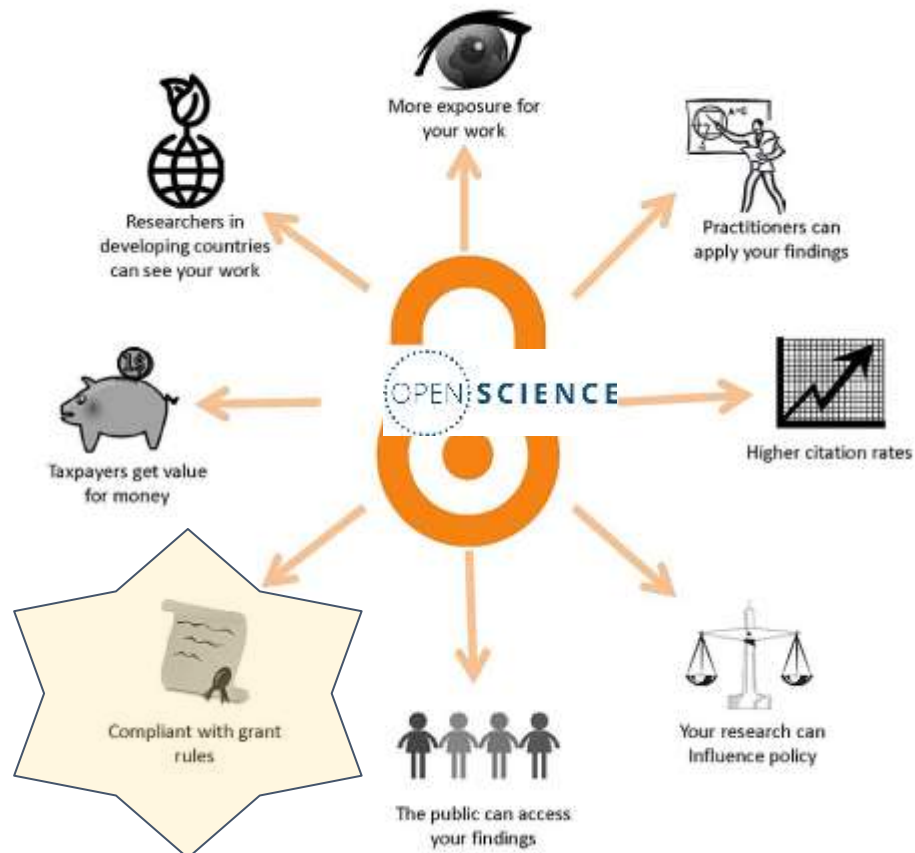
## Why open science?



Good for You

By sharing your articles, data, code and methods, you are **multiplying the number of citable outputs**. Your research will be more visible and understandable to others, which may mean that you might see your citation rate increase. If people can find and access your research, **the potential impact increases**. In addition, practicing open science can foster new collaborations and research partnerships.

## Why open science?

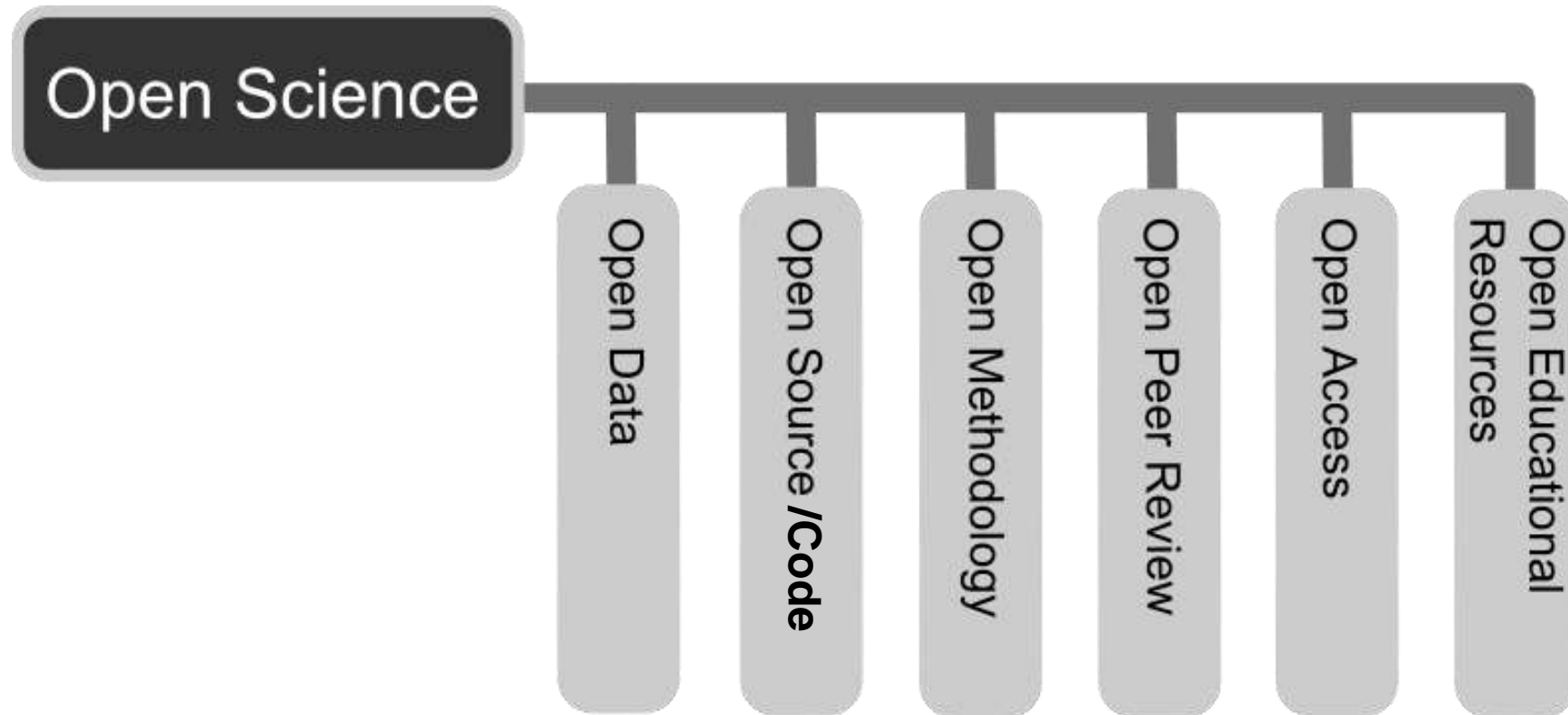


Compliant with grant rule....

**Biodiversa+** strongly supports open science, including open sharing of research data and digital outputs to stimulate novel approaches to the collection, reuse, analysis, validation, and management of data and information, thus increasing the transparency of the research process and robustness of the results. Therefore, submitted projects are expected to make produced data, digital outputs, and supporting material (including metadata) **publicly available**, possibly after a short period of exclusivity, unless there are legitimate reasons to constrain access. In particular, raw data should be made **accessible** to allow for integrated data analysis across different datasets. Data and digital outputs must be **discoverable** through machine readable catalogues, information systems and search engines. Projects should generate FAIR4 data and knowledge products, particularly in the context of real-time data feeds, exploring workflows that can provide “**FAIR-by-design**” data, i.e., data that is FAIR from its generation, and building on and widening data availability in European Research Infrastructures federated under the European Open Science Cloud (**EOSC**). To this end, project proposals will need to develop and implement a **Data and Digital Outputs Management Plan**, which will also ensure ethical approaches and compliance with the Data Policy of this call.



## Open Science - The pillars of Open Science



Open Science - more than just open access to publications!



...Share your **data**

the research data that underpins publications should also be **accessible** to support **validation** and **facilitate reuse**. In cases where data sensitivities won't allow open access, be sure to provide details on how someone could request authorised access.





Open Science - more than just open access to publications!

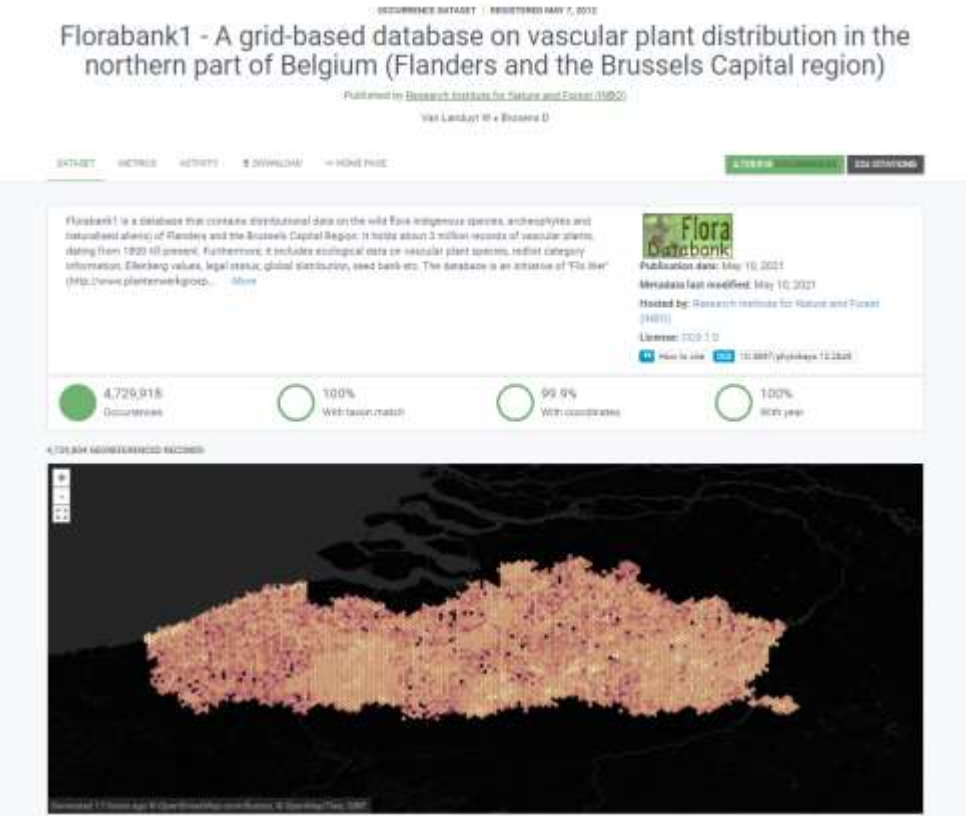
...Share your **data**

**BOLD**SYSTEMS

“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).”



Open Science - more than just open access to publications!





# Open Science - more than just open access to publications!



Studies	
View Download Env-DATA	
Study Details	
Study Name	White Stork Denmark
Contact Person	JessEred (Jess Frederiksen)
Principal Investigator	Anders P. Tøttrup/Kasper Thorup
Citation	
Acknowledgements	
Grants used	Storkene.dk
License Type	Custom
License Terms	not set
Study Summary	Migration routes of the small Danish population of white storks.
Study Reference Location	
Longitude	12.578
Latitude	55.687
Movebank ID	1578840976
Study Statistics	
Number of Animals	11
Number of Tags	8
Number of Deployments	11
Time of First Deployed Location	2021-05-19 09:00:23.000
Time of Last Deployed Location	2023-05-24 10:00:19.000
Taxa	Ciconia ciconia
Number of Deployed Locations	676823
Number of Records	Deployed (outliers) / Total (outliers)
Acceleration	299462 (0) / 349655 (0)
GPS	676823 (0) / 699524 (0)
About study details	
Processing Status	Up-to-date

## Marine Biological Sample Database, JAMSTEC

URL	<a href="https://www.godac.jamstec.go.jp/lp/resource/?jamstec_sampledb">https://www.godac.jamstec.go.jp/lp/resource/?jamstec_sampledb</a>
Repository URL	<a href="https://www.godac.jamstec.go.jp/lp/">https://www.godac.jamstec.go.jp/lp/</a>
Node	OBIS Japan
Published	2023-05-18 12:15
Abstract	This dataset contains data of biological samples which were collected during scientific missions of JAMSTEC ships (NATSUSHIMA, KAIYO, YOKOSUKA, KAIREI and MIRAI) and <a href="#">submersibles</a> . Data of this dataset is derived from the Marine Biological Sample Database of JAMSTEC. At the original database, you may search sample information such as number of individuals, preservation methods, sex, life stages, identification, collecting information and related literatures.
Citation	Japan Agency for Marine-Earth Science and Technology (2016 onwards). JAMSTEC Marine Biological Samples Database. <a href="https://doi.org/10.48518/00001">https://doi.org/10.48518/00001</a> . Accessed on yyyy-mm-dd.
Rights	This work is licensed under a Creative Commons Attribution Non-Commercial (CC-BY-NC) 4.0 License
Keywords	Occurrence, Occurrence
Contacts	Creditor Data Management Group, JAMSTEC Contact Data Management Group, JAMSTEC Metadata Provider Data Management Group, JAMSTEC

[View this issue](#) [Source Data](#) [Download](#)

Open Science - more than just open access to publications!



## Gesäuse-Johnsbachtal - Austria, Soil base saturation

### Basic information

**Related Site:** [Gesäuse-Johnsbachtal - Austria](#)  
**Abstract:** Soil base saturation data of the Gesäuse-Johnsbachtal site  
**Keywords:** [soil parameter](#)  
**Contact:** [Referat Landesforstdirektion Steiermark](#)  
**Owner/Creator:** [Referat Landesforstdirektion Steiermark](#)  
**DOI:** <https://doi.org/10.23728/b2share-61c61c3d3100c46f09cc67c5669685428>  
**UUID:** [225324a3-9c8a-ef8a-a3f4-f0fb19e31fa7](#)

### Dates

**Date Range:** Sun, 01/01/1989 - 12:00  
**Post date:** Wednesday, May 17, 2023 - 10:40  
**Last modified:** Monday, May 22, 2023 - 11:04

### Downloads

**Online Locator**  
**Distribution Function:** Web address (URL)  
**Distribution URL:** [B2Share Landing Page](#)  
**Distribution Function:** File for download  
**Distribution URL:** [B2Share Download Link](#)

### Geographic



## Dataset: Chesapeake Bay Nutrients 2021

View Data: Data not available yet

[Data This Dataset](#)

Temporal Span: 2021-03-04 - 2021-05-10

**Project:** [NODD Database in Oregon Watershed Zones \(NODD-CMZ\)](#)  
**Principal Investigator:** [Alex B. Zedler](#) (Portland University)  
**BCO-DMO Data Manager:** [Brendan R. Smith](#) (Woods Hole Oceanographic Institution, WHOI BCO-DMO)  
**Version:** 1  
**Version Date:** 2023-05-10  
**Restricted:** No  
**Validated:** No  
**Current State:** Data not available

Chesapeake Bay Nutrients 2021



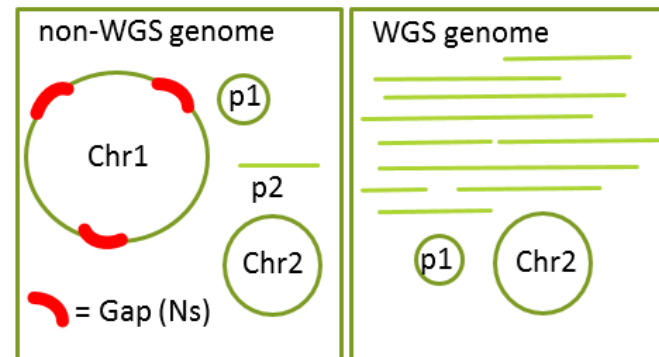
Open Science - more than just open access to publications!

BOLD SYSTEMS



### Prokaryotic and Eukaryotic Genomes Submission Guide

Both WGS and non-WGS genomes, including gapless complete bacterial chromosomes, can be submitted via the Submission Portal. You will be asked to choose whether the genome being submitted is considered WGS or not. The differences for GenBank purposes are:



#### non-WGS

- Each chromosome is in a single sequence and there are no extra sequences
- Each sequence in the genome must be assigned to a chromosome or plasmid or organelle
- Plasmids and organelles can still be in multiple pieces.

#### WGS

- One or more chromosomes are in multiple pieces and/or some sequences are not assembled into chromosomes

#### In both cases

Specimens:

Sequences:

Combined:

Map:

### Results Summary

Found **412** published records,  
with **412** records with sequences,  
forming **8** BINs (clusters),  
with specimens from **30** countries,  
deposited in **18** institutions.

Of these records, **412** have species names, and represent **1** species.

### Specimen Distribution



Open Science - more than just open access to publications!



[Upload](#) [Communities](#)

[Log in](#) [Sign up](#)

May 30, 2021

[Dataset](#) [Open Access](#)

## A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

Bando, Juan M. Yokumaki, Remya, Wang, Guanyu, Yu, Jingyuan, Liu, Tao, Ding, Yuting, Artemova, Katya, Tutubalina, Elena Crowell, Gerardo

Version 64 of the dataset

Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter Stream related to COVID-19 chatter. Since our first release we have received additional data from our new collaborators, allowing this resource to grow to its current size. Dedicated data gathering started from March 11th yielding over 4 million tweets a day. We have added additional data provided by our new collaborators from January 27th to March 27th, to provide extra longitudinal coverage. Version 10 added ~1.5 million tweets in the Russian language collected between January 1st and May 8th, graciously provided to us by: Katya Artemova (NRU HSE) and Elena Tutubalina (KFU). From version 12 we have included daily hashtags, mentions and emojis and their frequencies the respective zip files. From version 14 we have included the tweet identifiers and their respective language for the clean version of the dataset. Since version 20 we have included language and place location for all tweets.

The data collected from the stream captures all languages, but the higher prevalence are: English, Spanish, and French. We release all tweets and retweets on the full\_dataset.csv file (1,098,118,400 unique tweets), and a cleaned version with no retweets on the full\_dataset-clean.csv file (278,205,848 unique tweets). There are several practical reasons for us to leave the retweets, tracing important tweets and their dissemination is one of them. For NLP tasks we provide the top 1000 frequent terms in frequent\_terms.csv, the top 1000 bigrams in frequent\_bigrams.csv, and the top 1000 trigrams in frequent\_trigrams.csv. Some general statistics per day are included for both datasets in the full\_dataset-statistics.csv and full\_dataset-clean-statistics.csv files. For more statistics and some visualizations visit: <http://www.panaceslab.org/covid19/>

More details can be found (and will be updated faster at: [https://github.com/thepanaceslab/covid19\\_twitter](https://github.com/thepanaceslab/covid19_twitter)) and our pre-print about the dataset (<https://arxiv.org/abs/2004.03668>)

As always, the tweets distributed here are only tweet identifiers (with date and time added) due to the terms and conditions of Twitter to re-distribute Twitter data ONLY for research purposes. They need to be hydrated to be used.

This dataset will be updated bi-weekly at least with additional tweets, look at the github repo for these updates.  
Release: We have standardized the name of the resource to match our pre-print manuscript and to not have to update it every week.

[Preview](#)

89,594  
views

106,721  
downloads

[See more details...](#)

linked to

Publication date:  
May 30, 2021

DOI:  
[10.5281/zenodo.4816034](https://doi.org/10.5281/zenodo.4816034)

Keyword(s):  
[world maps](#) [twitter](#) [nlp](#) [covid-19](#) [covid19](#)

Related identifiers:  
Continued by:  
<http://www.panaceslab.org/covid19/> (Other)  
Supplement to:  
<https://arxiv.org/abs/2004.03668> (Preprint)

Alternate identifiers:  
[https://github.com/thepanaceslab/covid19\\_twitter](https://github.com/thepanaceslab/covid19_twitter) (Software)

Communities:  
[BioRxiv](#)  
[COVID-19 Dataset Research Community](#)  
[COVID-19](#)  
[Zenodo](#)

License (for files):  
[CC BY](#) (Public Domain)



# Open Science - more than just open access to publications!

The screenshot displays the Zenodo website interface. At the top, the Zenodo logo is on the left, followed by a search bar and navigation links for 'Communities' and 'My dashboard'. On the right, there are 'Log in' and 'Sign up' buttons. Below the header, the 'RESEARCH INSTITUTE NATURE AND FOREST' (INBO) profile is featured, including its name, a brief description, a URL, and a 'New upload' button. The main content area shows a list of research records. On the left, there are filters for 'Versions', 'Access status', 'Resource types', and 'Subjects'. The records are sorted by 'Newest'. The first record is titled 'R package n2khabmon: prepare and manage N2KHAB monitoring schemes' by Vanderhaeghe, Floris, dated May 6, 2024. The second record is titled 'Data from a cross-sectional study of fifth grade children in a sample of primary schools in Belgium that differ in amount of greenness at school and landscape level' by Van Calster, Hans, et al., dated February 6, 2024. The third record is titled 'March 26, 2024 (P1610026)' and is dated March 26, 2024.

zenodo

RESEARCH INSTITUTE NATURE AND FOREST

Research Institute for Nature and Forest (INBO)

https://www.vlaanderen.be/natuur-gebiedsmanagement

Organization

Research Institute for Nature and Forest INBO

132 results found

Sort by Newest

May 6, 2024 (v2.0)

Software

Open

R package n2khabmon: prepare and manage N2KHAB monitoring schemes

Vanderhaeghe, Floris

n2khabmon is an R package with utilities to prepare and manage Flemish monitoring schemes regarding Natura 2000 habitats and regionally important biotopes (RIBs). URL: <https://inbo.github.io/n2khabmon/>

Part of Research Institute for Nature and Forest (INBO)

Uploaded on May 6, 2024

4 more versions exist for this record

132

February 6, 2024 (v1)

Dataset

Open

Data from a cross-sectional study of fifth grade children in a sample of primary schools in Belgium that differ in amount of greenness at school and landscape level

Van Calster, Hans, Lommelen, Els, Groenewald, Antoine, and 17 others

The data in this deposit were collected as part of the B@SEBALL project (Biodiversity at School Environments - Benefits for All). The project investigated how biodiversity in the school environment can positively affect children's health and mental well-being. B@SEBALL also investigated the opportunities for reducing health inequalities among...

Part of Research Institute for Nature and Forest (INBO)

Uploaded on April 13, 2024

101

March 26, 2024 (P1610026)

Dataset

Open

Open Science - more than just open access to publications!



Search

Explore Data | About ▼ | Help ▼ | Login

## Data from: Latitudinal patterns of phenology and age-specific thermal performance across six *Coenagrion* damselfly species

Nilsson-Örtman, Viktor, Umeå University  
Stoks, Robby, KU Leuven  
Block, Marjan De, KU Leuven  
Johansson, Frank, Umeå University  
Publication date: October 15, 2014  
Publisher: Dryad  
<https://doi.org/10.5061/dryad.1q389>

### Citation

Nilsson-Örtman, Viktor; Stoks, Robby; Block, Marjan De; Johansson, Frank (2014). Data from: Latitudinal patterns of phenology and age-specific thermal performance across six *Coenagrion* damselfly species, Dryad, Dataset, <https://doi.org/10.5061/dryad.1q389>

### Abstract

Using a combination of computer simulations and laboratory experiments we test if the thermal sensitivity of growth rates change during ontogeny in damselfly larvae and if these changes can be predicted based on the natural progression of average temperature or thermal variability in the field. The laboratory experiment included replicated species from Southern, Central and Northern Europe. Although annual fluctuations in temperature represent a key characteristic of temperate environments, few studies of thermal performance have considered the ecological importance of the studied traits within a seasonal context. Instead, thermal performance is assumed to remain constant throughout ontogeny and reflect selection acting over the whole life cycle. The laboratory experiment revealed considerable variation among species in the strength and direction of ontogenetic performance shifts. In four species from Southern and Central Europe, reaction norms were steepest during early ontogeny, becoming less steep during later ontogenetic stages.

### Data Files

[Download dataset](#)

October 15, 2014

### Related Works

#### Article

<https://doi.org/10.1890/12-1383.1>

### Metrics

146 views

93 downloads

1 citations

### Keywords

[Acclimatization](#)

[Coenagrion armatum](#)

[Coenagrion caeruleum](#)



Open Science - more than just open access to publications!



## STORM tropical cyclone wind speed return periods

Cite

Download all (77.99 MB)

Share

Embed

+ Collect

**Version 2** Dataset posted on 30.09.2020, 14:44 by [Nadia Bloemendaal](#), H. (Hans) de Moel, [S. Mus](#), I.D. (Ivan) Haigh, J.C.J.H. (Jeroen) Aerts

Datasets containing tropical cyclone maximum wind speed (in m/s) return periods, generated using the STORM datasets (see <https://www.nature.com/articles/s41597-020-0381-2>). Return periods were empirically calculated using Weibull's plotting formula. The STORM\_FIXED\_RETURN\_PERIOD dataset contains maximum wind speeds for a fixed set of return periods at 10 km resolution in every ocean basin. The STORM\_FIXED\_WIND\_SPEED dataset contains return periods for a fixed set of maximum wind speeds at 10 km resolution in every ocean basin. The STORM\_CITIES dataset contains return periods at fixed wind speeds and wind speeds at fixed return periods (on two separate sheets), occurring within 100 km from a selection of 18 coastal cities. The STORM\_ISLANDS dataset contains return periods at fixed wind speeds and wind speeds at fixed return periods (on two separate sheets), occurring within 100 km from the capital city of an island. We included the Small Island Developing States and a set of other islands.

### HISTORY

- 13.05.2020 - First online date, Publication date
- 30.09.2020 - Submission date, Posted date
- 22.02.2021 - Revision date

### PUBLISHER

4TU Centre for Research Data

### USAGE METRICS

711  
views

165  
downloads

3  
citations



Read the peer-reviewed publication

Generation of a global synthetic tropical cyclone hazard dataset using STORM

4TU ResearchData

### CATEGORIES

- Atmospheric Sciences
- Natural Hazards
- Atmosphere and Weather
- Climate and Climate Change

### KEYWORDS

Natural Hazards, Data, Research

## Open Science - more than just open access to publications!

...Open data should be FAIR data



### Findable

- Persistent identifiers (DOI)
- Metadata
- Naming conventions
- Keywords
- Versioning



### Accessible

- Choice of datasets
- Data repository
- Software, documentation
- Access status
- Retrievable data
- Metadata access



### Interoperable

- Standards
- Vocabulary
- Methodology
- References



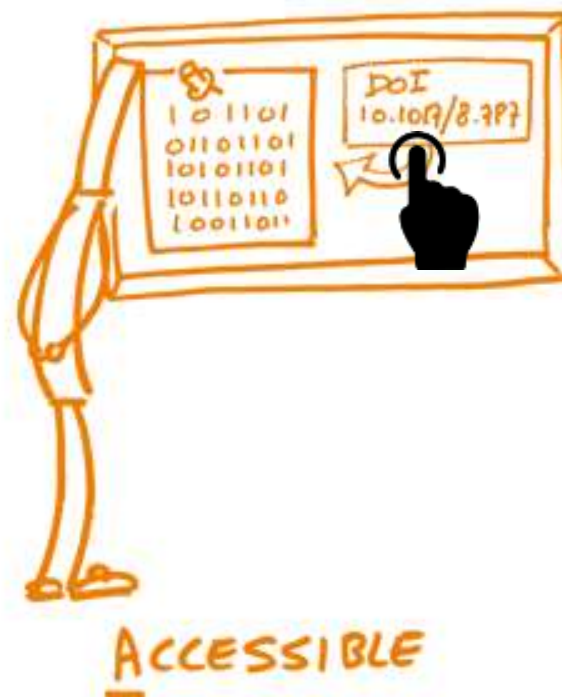
### Reusable

- Licensing
- Provenance
- Community standards



Open Science - more than just open access to publications!

## FAIR DATA PRINCIPLES



# Open Science - more than just open access to publications!

Total across F.A.I.R

## Findable



Does the dataset have any identifiers assigned?

Globally Unique, citable and persistent (e.g. DOI) ▼

Is the dataset identifier included in all metadata records/files describing the data?

Yes ▼

How is the data described with metadata?

Comprehensively (see suggestion) using a record ▼

What type of repository or registry is the metadata record in?

Generalist public repository ▼

## Accessible



How accessible is the data?

Publicly accessible ▼

Is the data available online without requiring specialised protocols or tools once access has been approved?

Standard web service API (e.g. OGC) ▼

Will the metadata record be available even if the data is no longer available?

No ▼

## Interoperable



What (file) format(s) is the data available in?

In a structured, open standard, machine-readable format ▼

What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?

Standardised open and universal using resolved URIs ▼

How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?

The metadata record includes URI links to related data ▼

## Reusable



Which of the following best describes the license/usage rights attached to the data?

Standard machine-readable license (e.g. Creative Commons) ▼

How much provenance information has been captured to facilitate data reuse?

Fully recorded in a machine readable format ▼

Total across F.A.I.R

FAIR	Open data	Similarity	Difference
Findable: Data can be <b>easily found</b> by machines and humans.	<b>Data must be available</b> in its entirety and at a low reproduction cost.	Findability and accessibility in FAIR is quite similar to availability in open data, because it provides accessibility to the user.	Availability in open data refers to data integrity and does not mention other conditions. Open data focuses on no barriers to data accessibility, while accessibility in FAIR principles highlights the need for data protection and the conditions for access to be formulated to meet the specific circumstances that relate to the data.
Accessible: <b>Authentication</b> and verification is <b>possible</b> after the user accesses the data.			
Interoperable: <b>Data can be integrated with other data</b> and systems or workflows for analysis, storage and processing are interoperable.			Interoperability is promoted by the creation of machine-readable instances of ontologies that the data represent, linked to metadata in languages such as JASON or RDF, widely used for the Semantic Web.
Reusable: Metadata and data should be <b>defined for reuse</b> and can be replicated and/or mixed in different environments.	• The data should be made available under the condition of reuse and redistribution.	Both <b>principles have the purpose of making data reusable.</b>	Open data does not mention metadata and focuses on redistribution neutrality.
	• <b>Anyone should be able to use, reuse and redistribute the data</b> — there is no discrimination based on the purpose for which the data is to be used or the individuals/groups wishing to use it.		



Open Science - more than just open access to publications!

...Share your **code**

- many researchers now develop bespoke bits of code to help them analyse and/or visualise the data they have collected. Having access to this code is essential for supporting the validation of your findings and to help others to build upon your work.



Open Science - more than just open access to publications!

[master](#)
[2 branches](#)
[0 tags](#)
[Go to file](#)
[Add file](#)
[Code](#)
[Use this template](#)

[peterdasmet Add VMM dataset](#)
10a1c1e on 11 Mar
221 commits

datasets	Merge branch 'master' of <a href="https://github.com/nbp/mica-occurrences">https://github.com/nbp/mica-occurrences</a>	3 months ago
.gitignore	Ignore all interim data	3 months ago
LICENSE	Initial commit	14 months ago
README.md	Add VMM dataset	3 months ago
mica-occurrences.Rproj	Rename repo	6 months ago

[README.md](#)

## MICA occurrence datasets

### Rationale

This repository contains the functionality to standardize several datasets of **Muskrat** and **Coypu** occurrences to **Darwin Core Occurrence** datasets that can be harvested by **GBIF**. These datasets are published in the framework of the project **Management of Invasive Coypu and muskrat in Europe (MICA)**.


### Datasets


Title (and GitHub directory)	IPT	GBIF
Muskrat captures in Flanders, Belgium	<a href="#">mica-legacy-occurrences</a>	<a href="https://doi.org/10.15468/pequ4z">https://doi.org/10.15468/pequ4z</a>
MICA - Muskrat occurrences collected by RATO in East Flanders, Belgium	<a href="#">mica-rato-occurrences</a>	<a href="https://doi.org/10.15468/5f9s96">https://doi.org/10.15468/5f9s96</a>
MICA - Muskrat and coypu occurrences collected by	<a href="#">mica-uwv-</a>	<a href="https://doi.org/10.15468/qids4c">https://doi.org/10.15468/qids4c</a>

**About**

Darwin Core mapping for the aggregated Muskrat dataset

[dataset](#) [github](#) [accessories](#)

 [Readme](#)

 [MIT License](#)

---

**Releases**

No releases published

[Create a new release](#)


---


**Packages**


No packages published

[Publish your first package](#)



---

**Contributors** 




 **peterdesmet** Peter Desmet

 **DimEvil** Dimi Brocens

master [mice-occurrences / datasets / mice-agouti-occurrences / src / dwt\\_mapping.Rmd](#) [Go to file](#) [...](#)

 peterdemeut Correctly assign vernacular name and rank for dogs Latest commit 16/06/20 on 3 May 

As 1 contributor

361 lines (258 xloc) 7.6 KB [Raw](#) [Blame](#)   

```

1  ---
2  title: "Dandel Core Mapping"
3  subtitle: "MICA : Muekrat and Corega (amers trap occurrences collected in Belgium, the Netherlands and Germany)"
4  author: "Dimitri Brounse, Peter Demeut"
5  date: "r Sys.Date()"
6  output: html_document
7  ---
8
9  ""[r setup, include = FALSE]
10 knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = TRUE)
11 ""
12
13 load libraries:
14
15 ""[r message = FALSE]
16 library(tidyverse) # To do data science
17 library(tidylog) # To provide feedback on knitr functions
18 library(magrittr) # To use %>% pipes
19 library(readr) # To read files
20 library(janitor) # To clean input data
21 ""
22
23 # Quick timezone conversion
24
25 Most of the Dandel Core mapping is done by running queries in sqlite. But timezones get ignored when reading data into sqlite. R converts them nicely to UTC. So the following
26
27 ""[r eval = FALSE]
28 deployments <- read_csv(here::here("datasets", "mice-agouti-occurrences", "data", "raw", "deployments.csv"))
29 multimedia <- read_csv(here::here("datasets", "mice-agouti-occurrences", "data", "raw", "multimedia.csv"))
30 observations <- read_csv(here::here("datasets", "mice-agouti-occurrences", "data", "raw", "observations.csv"))
31 ""

```

Open Science - more than just open access to publications!

...Share your **workflows & methodology**

- without knowing what steps were taken to **capture, process and analyse** the data - and in what order - it can be virtually impossible to validate published findings. This has led to what some are calling the [Reproducibility Crisis](#).





Open Science - more than just open access to publications!



OSFHOME ▾ My Quick Files My Projects Search Support Donate Dimitri Brosens •

Tracking Invasive Alien Species (TrIAS) Files Wiki Analytics Registrations Contributors Add-ons Settings

ergo maintenance between May 26, 2021 2:00 AM and May 26, 2021 4:00 AM (+0200 UTC). Thank you for your patience.

## Tracking Invasive Alien Species (TrIAS)

104.0MB Public P 0

Contributors: Quentin Groom, Tim Adriaens, Peter Desmet, Diederik Strubbe, Thierry Backeljau, Etienne Branquart, Dimitri Brosens, Maxime Couprenmanne, Arny JS Davis, Rozemien De Troch, Valérie De Waele, Sander Devisscher, Hilde Eggermont, Andre Heughebaert, Kris Hostens, Pierre Huybrechts, Jacquemart Anne-Laure, Arnaud Monty, Damiano Oldoni, Thierry Onkelinx, Jean-Yves Pasquet, Lierl Reyserhove, Toon Van Daele, Ruben Van De Kerckhove, Gert Van Hovey, Wouter Vansausel, Bert Van Schaeybroeck

Date created: 2017-01-06 03:02 PM | Last Updated: 2021-05-21 11:03 AM  
Identifier: DOI 10.17605/OSF.IO/7DPGR  
Category: Project  
Description:  
Building an open data-driven framework to support policy on invasive species  
License: CC-BY Attribution 4.0 International ⓘ

Wiki

TrIAS ▶▶▶

Citation

Components Add Component Link Projects

Private documents



Open Science - more than just open access to publications!

### ...Open **peer review**

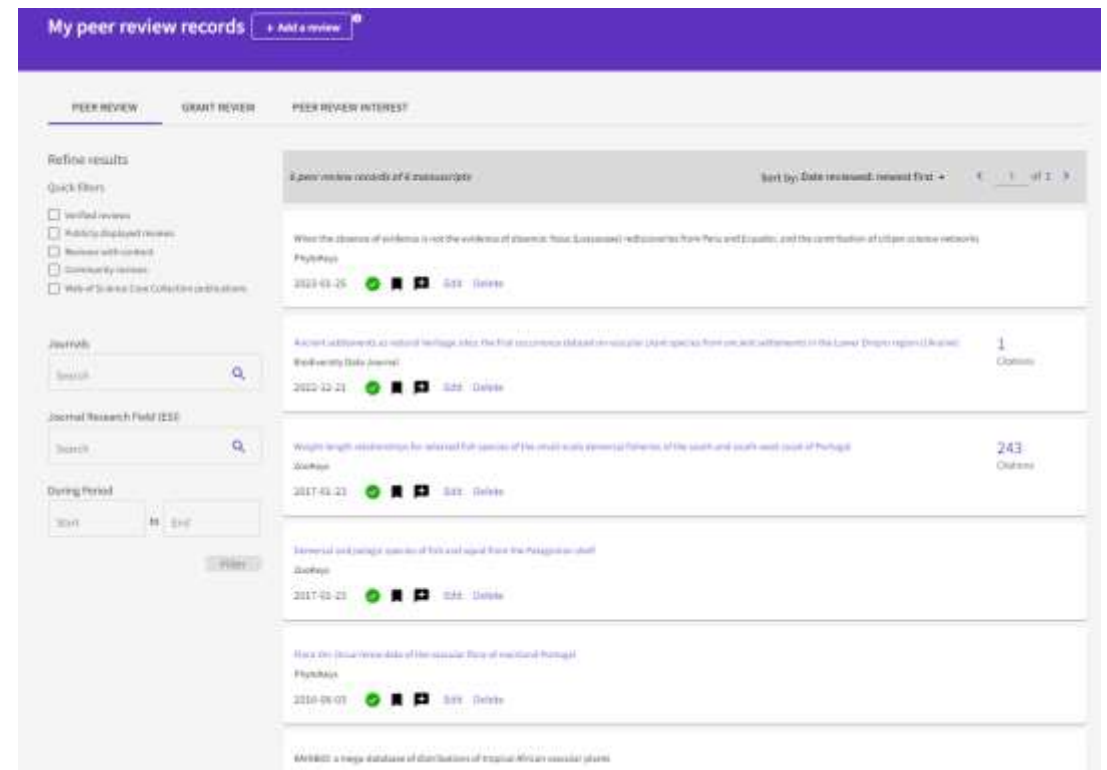
Open peer review (also called "public peer review", "transparent peer review") denotes several, closely related forms of scholarly peer review: Open-identity or attributed peer review (as opposed to anonymous peer review)

Open-disclosure or public peer review, where the **peer review contents are publicly available**.

# Open Science - more than just open access to publications!

## ...Open peer review

making the peer review process **more transparent**, researchers have better access to **peer feedback** at an earlier stage in the lifecycle and consumers of research outputs can have greater confidence in their quality.



The screenshot displays the 'My peer review records' page. At the top, there's a purple header with the title and an 'Add a review' button. Below the header, there are three tabs: 'PEER REVIEW', 'GRANT REVIEW', and 'PEER REVIEW INTEREST'. The 'PEER REVIEW' tab is active. On the left side, there's a 'Refine results' section with 'Quick filters' (verified reviews, publicly displayed reviews, reviews with content, community reviews, Web of Science Core Collection publications) and search boxes for 'Journals' and 'Journal Research Field (ES)'. There's also a 'During Period' section with 'Start' and 'End' date pickers. The main content area shows a list of peer review records. Each record includes a title, a date, a status indicator (green circle), and a 'Details' link. The records are sorted by 'Date received: newest first'. The first record is titled 'When the species of ...' and has a date of '2019-03-25'. The second record is titled 'Ancient ...' and has a date of '2019-12-21'. The third record is titled 'Weight ...' and has a date of '2017-03-23'. The fourth record is titled 'Diversified ...' and has a date of '2017-02-23'. The fifth record is titled 'How ...' and has a date of '2018-06-07'. The sixth record is titled 'ENRICH ...' and has a date of '2018-06-07'.



Open Science - more than just open access to publications!

### ...Open **Educational Resources**

Open Educational Resources (OER) are teaching, learning and research **materials** in any medium – digital or otherwise – **that reside in the public domain** or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions.



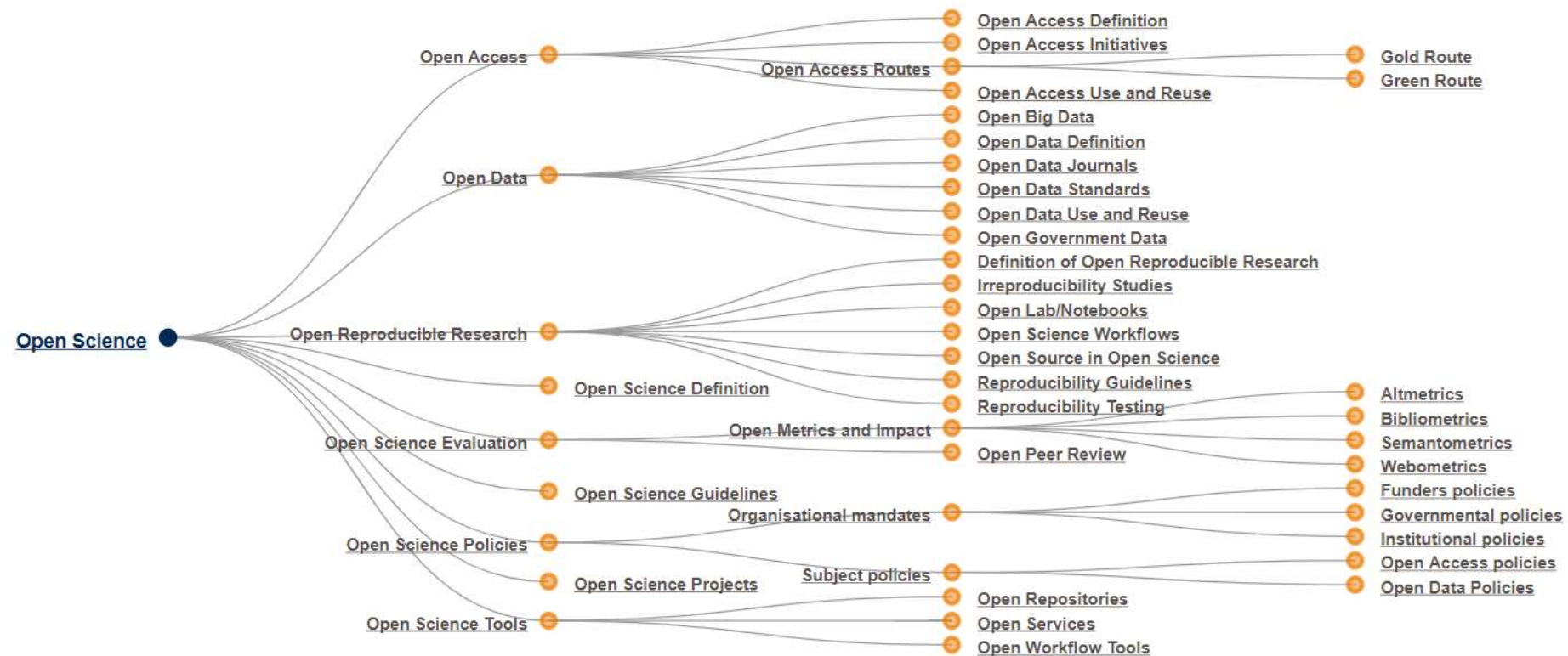
Home

## INBO Tutorials

### Introduction

Welcome to the tutorials website of the Research Institute for Nature and Forest (INBO). It contains a collection of guides, tutorials and further reading material on the installation, use and development of **(research) software** at our institute.

# Open Science - Taxonomy!





Open Science - Ideas for opening up!

## During the planning stage

Consider writing a **blog post/ research idea paper** outlining your ideas and approaches early on to get community feedback.

- *Find the relevant research question*

**Seek involvement** from other researchers, be sure to consider involving other stakeholders too.

- *Make the research Question is sharp and clear*
- *Research is relevant with the research question*

**Check data repositories** to see if there are existing data that you can reuse or build upon.



The screenshot shows the top navigation bar of the RIO platform with links for Home, Articles, About, About Pensoft, Books, Journals, and Blog. Below the navigation bar, there is a teal 'Grant Proposal' button and a red 'Reviewable' button. The article title is 'Tracking Invasive Alien Species (TrIAS): Building a data-driven framework to inform policy'. The authors listed are Sonia Vanderhoeven, Tim Adriaens, Peter Desmet, Diederik Strubbe, Thierry Backeljau, Yvan Barbier, Dimitri Brosens, Julien Cigar, Maxime Coupremagne, Rozemien De Troch, Hilde Eggermont, André Heughebaert, Kris Hostens, Pierre Huybrechts, Anne-Laure Jacquemart, Luc Lens, Arnaud Monty, Jean-Yves Paquet, Céline Prévot, Tim Robertson, Piet Termonia, Ruben Van De Kerchove, Gert Van Hoey, Bert Van Schaeybroeck, Diemer Vercayie, Thomas Jethro Verleye, Sarah Welby, and Quentin John Groom. The abstract section is partially visible with the word 'Abstract' followed by a small upward arrow.

**RIO** Home Articles About About Pensoft Books Journals Blog

Grant Proposal Research Ideas and Outcomes 3: e13414  
<https://doi.org/10.3897/rio.3.e13414> (02 May 2017)

Reviewable v1

### Tracking Invasive Alien Species (TrIAS): Building a data-driven framework to inform policy

▼ Sonia Vanderhoeven, Tim Adriaens, Peter Desmet, Diederik Strubbe, Thierry Backeljau, Yvan Barbier, Dimitri Brosens, Julien Cigar, Maxime Coupremagne, Rozemien De Troch, Hilde Eggermont, André Heughebaert, Kris Hostens, Pierre Huybrechts, Anne-Laure Jacquemart, Luc Lens, Arnaud Monty, Jean-Yves Paquet, Céline Prévot, Tim Robertson, Piet Termonia, Ruben Van De Kerchove, Gert Van Hoey, Bert Van Schaeybroeck, Diemer Vercayie, Thomas Jethro Verleye, Sarah Welby, Quentin John Groom

Abstract ▲



## Open Science - Ideas for opening up!

### During the planning stage

- Write a **Data Management Plan**
- Create a **Communication Plan**
- Foresee revision

**Population genetic study prior to the translocation of graylings within the framework of the Flemish Species Protection Plan (EVINBO)**

7280, n.n. n.n. <https://orcid.org/0000-0002-0000-0000>

#### My Dashboard

The table below lists all the plans associated with the **current user account**. This includes the plans you have created with this account and the plans that have been shared with you for this account.

Project Title	Template	Edited	Date	Owner	Test	Visibility	Shared
Biodiversity+ BIRU+ Biodiversity Indicators for Ruling the Wor...	BELSPD DMP	24-05-2023	Owner	You	<input checked="" type="checkbox"/>	N/A	No
Checking Belgis DMP	BELSPD DMP	01-09-2020	Owner	You	<input type="checkbox"/>	Private	No
My project (FWO DMP)	FWO DMP	02-05-2018	Owner	You	<input type="checkbox"/>	Private	No

[Create plan](#)

#### Research Institute for Nature and Forest 's Plans

The table below lists the plans that users at your organisation have created and shared within your organisation. This allows you to download a PDF and view their plans as samples or to discover new research data.

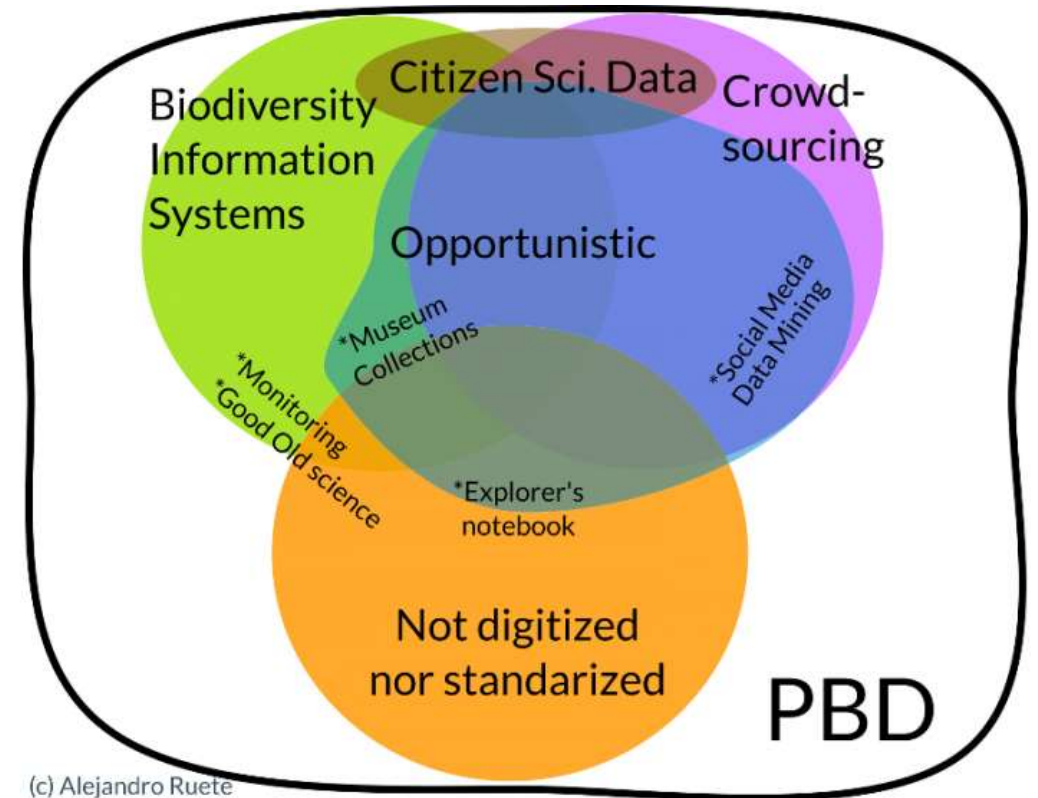
Project Title	Template	Owner	Updated	Download
Population genetic study prior to the ...	INBO Project focus template	an.vandenbroeck@inbo.be	30-06-2022	<a href="#">Download PDF</a>
Onderzoek naar metapopulaties en tran...	INBO Project focus template	karen.cox@inbo.be	17-05-2023	<a href="#">Open in new window</a>
MOVE2GEM - Modelling animal GPS tra...	INBO Project focus template	peter.demot@inbo.be	21-01-2022	<a href="#">Download PDF</a>

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.

## Open Science - Ideas for opening up!

### During the active stage\_ data collection

- Data collection through **well defined protocols**
- Automatic data collection when possible
- **Existing data is (re)used** when possible
- Validation procedures are in play
- Measurement biases are understood and documented

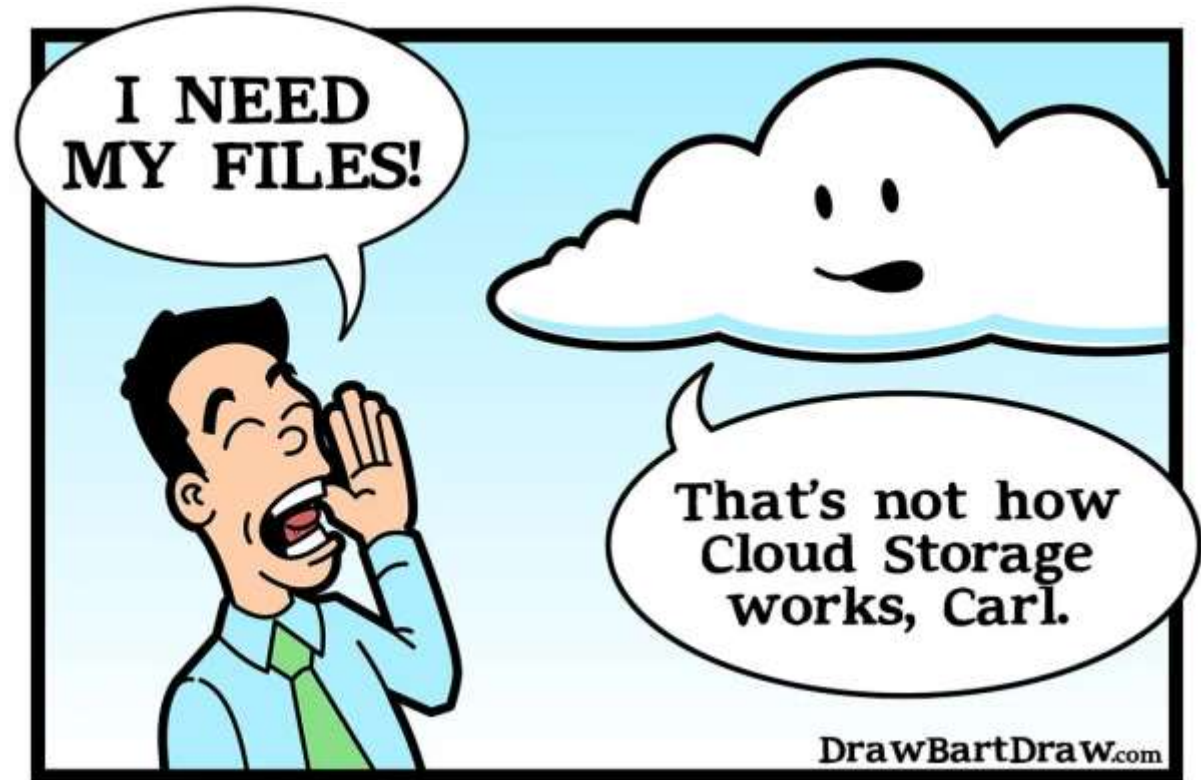




## Open Science - Ideas for opening up!

### During the active stage\_ data storage

- Data is well organised, documented and stored
  - Use **DMP**
  - **Metadata**
  - Be aware of sensitivities (GDPR...), sensitive data





## Open Science - Ideas for opening up!

### During the active stage\_ data analysis

- Data validation is reproducible
  - fixed validation rules
- Data transformation is reproducible
  - scripted data workflow
- Data analysis is adequate and correct
- Analyse workflow is open and reproducible
  - A versioning system is used (git)
  - Code and scripts are kept

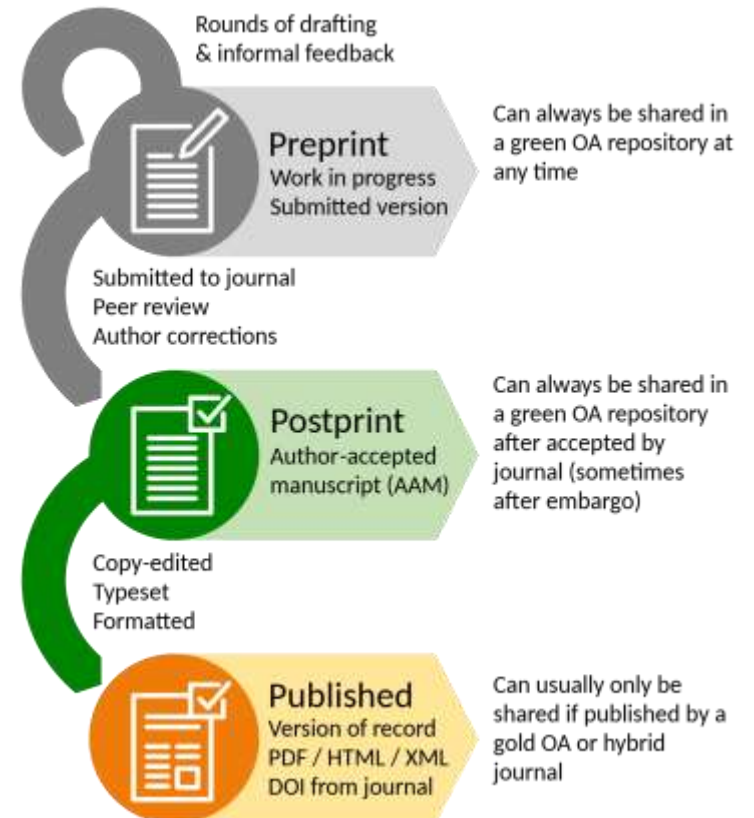


## Open Science - Ideas for opening up!

### During the active stage

You might share your methodologies and early findings via **preprints**.

Worried about getting scooped if you share early? Pre-registering your study gives you **time-stamped evidence of your ideas**. In addition, any peers that review your early work can vouch for you. The Open Science Framework (**OSF**) guide offers great advice on pre-registering your project.





## Open Science - Ideas for opening up!

### Towards the end of your research\_data sharing

- Data is published as open data
- FAIR principles are followed
- Open source code (scripts)
- Open protocols





## Open Science - Ideas for opening up!

### Towards the end of your research\_publications

- Published as open access
- Correct citations of sources (DOI...)
- Use of Orcid ID's

The screenshot shows the ORCID iD profile page for Dimitri Brosens. The page is titled "ORCID Connecting Research and Researchers" and includes a search bar and navigation links: ABOUT, FOR RESEARCHERS, MEMBERSHIP, DOCUMENTATION, RESOURCES, and NEWS & EVENTS. The profile section for Dimitri Brosens includes a biography, employment history, and education. The employment history lists three positions: Research Institute for Nature and Forest: Brussel, BE (2000 to present), Belgian Biodiversity Platform: Brussels, BE (2000 to present), and RBINS: Brussels, BE (2001 to 2008 - Education Department). The education section lists one entry: Ugent: Ghent, BE (1996 to 2001 - Bio (Biology Department)).

**ORCID iD**  
https://orcid.org/0000-0002-0846-9116  
View public version  
Switch account

Display your ID on other sites  
Public record print view  
Get a QR Code for your ID

**Also known as**  
Dimitri Brosens

**Country**  
Belgium

**Keywords**  
Biodiversity

**Websites & Social Links**  
Belgian Biodiversity Platform  
Research Institute for Nature and Forest profile page

**Other IDs**

**Biography**

**Employment (3)**

Research Institute for Nature and Forest: Brussel, BE  
Employment:  
Source: Research Institute for Nature and Forest  
★ Preferred source

Belgian Biodiversity Platform: Brussels, BE  
2000 to present  
Employment:  
Source: Dimitri Brosens  
★ Preferred source

RBINS: Brussels, BE  
2001 to 2008 - Education Department  
Employment:  
Source: Dimitri Brosens  
★ Preferred source

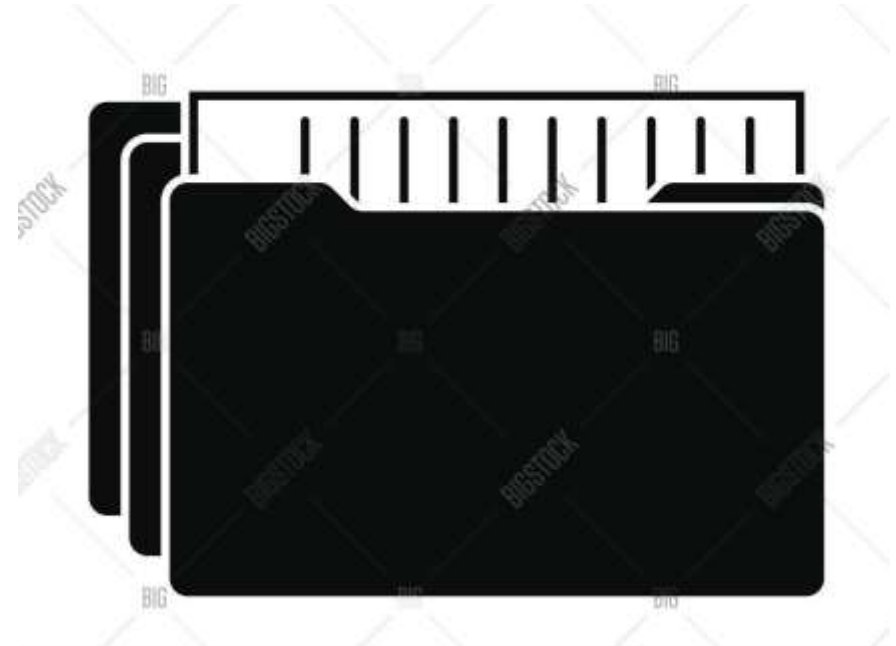
**Education and qualifications (1)**

Ugent: Ghent, BE  
1996 to 2001 - Bio (Biology Department)  
Education:  
Source: Dimitri Brosens  
★ Preferred source

Open Science - Ideas for opening up!

## Towards the end of your research\_archiving

- all documents and achieved sustainable and made open in an efficient way
- all data will be archived
- all samples are archived
- all protocols are archived
- all source code is archived





Open Science - Ideas for opening up!

## **Towards the end of your research\_reuse**

- Create opportunities for internal and external reuse of all resources



Open Science - Ideas for opening up!

## Towards the end of your research

Make sure to publish with an **Open Access** journal and/or to deposit your publications in an **Open Access repository**. This means that anyone can read - and cite - your findings in the short and longer-term.

Be sure to **deposit** any **data** required to validate your findings as well as any **software** you've developed to analyse or visualise them.

Link your papers, data, and code to each other through the assignment of DOIs. Link all of these back to you through your [ORCID](#) !





What is driving funding bodies to embrace Open Science?

The key driver for this is the belief that publicly funded research should be made available to support

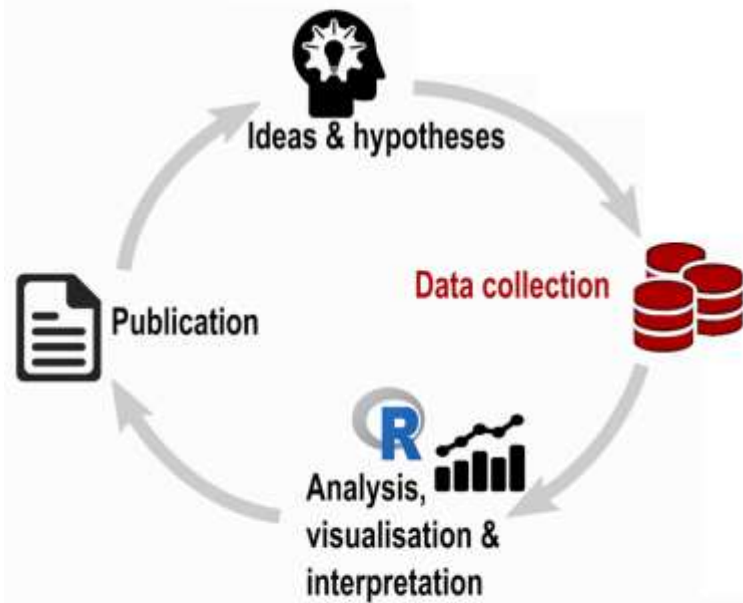
**public trust** in research  
to support **scrutiny** and **validation**  
to enable **reuse**  
and to drive **innovation**.

# Research Life Cycle





# Research Life Cycle



# Research Life Cycle

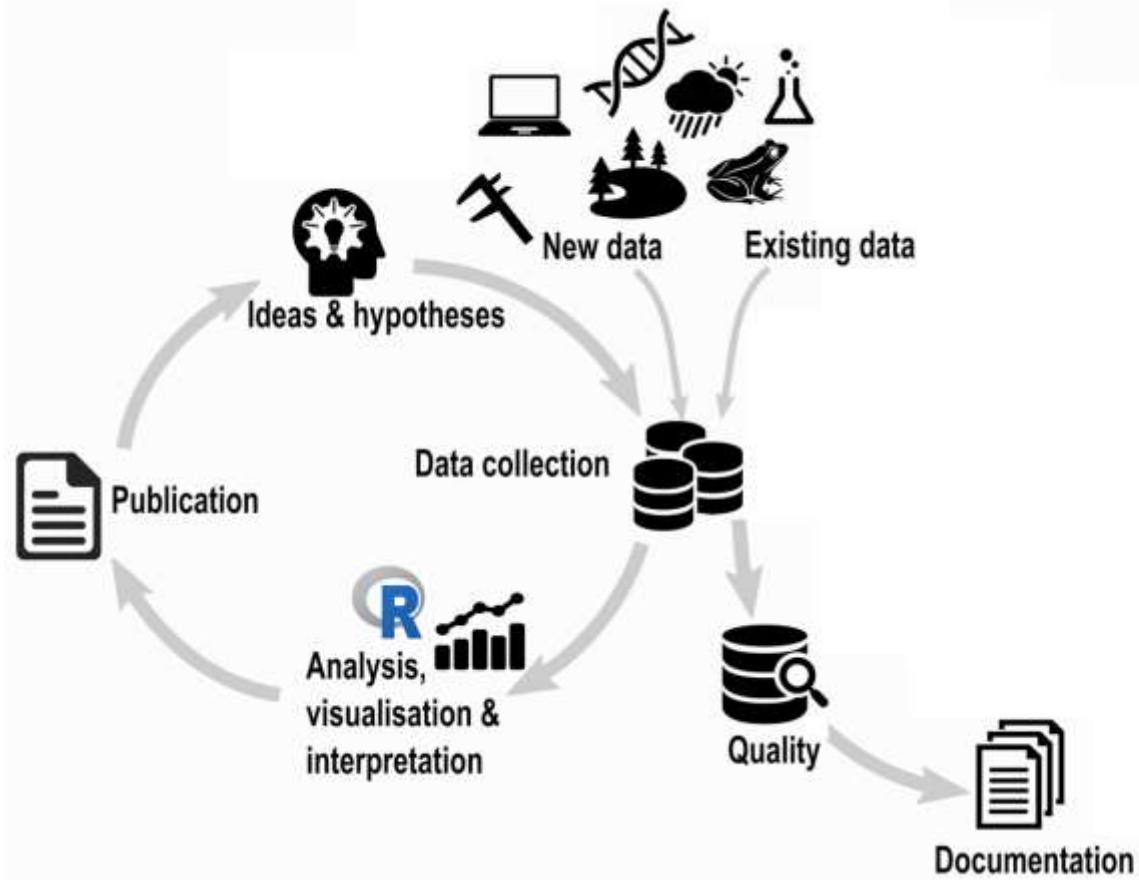


# Research Life Cycle





# Research Life Cycle



# Research Life Cycle



# Research Life Cycle

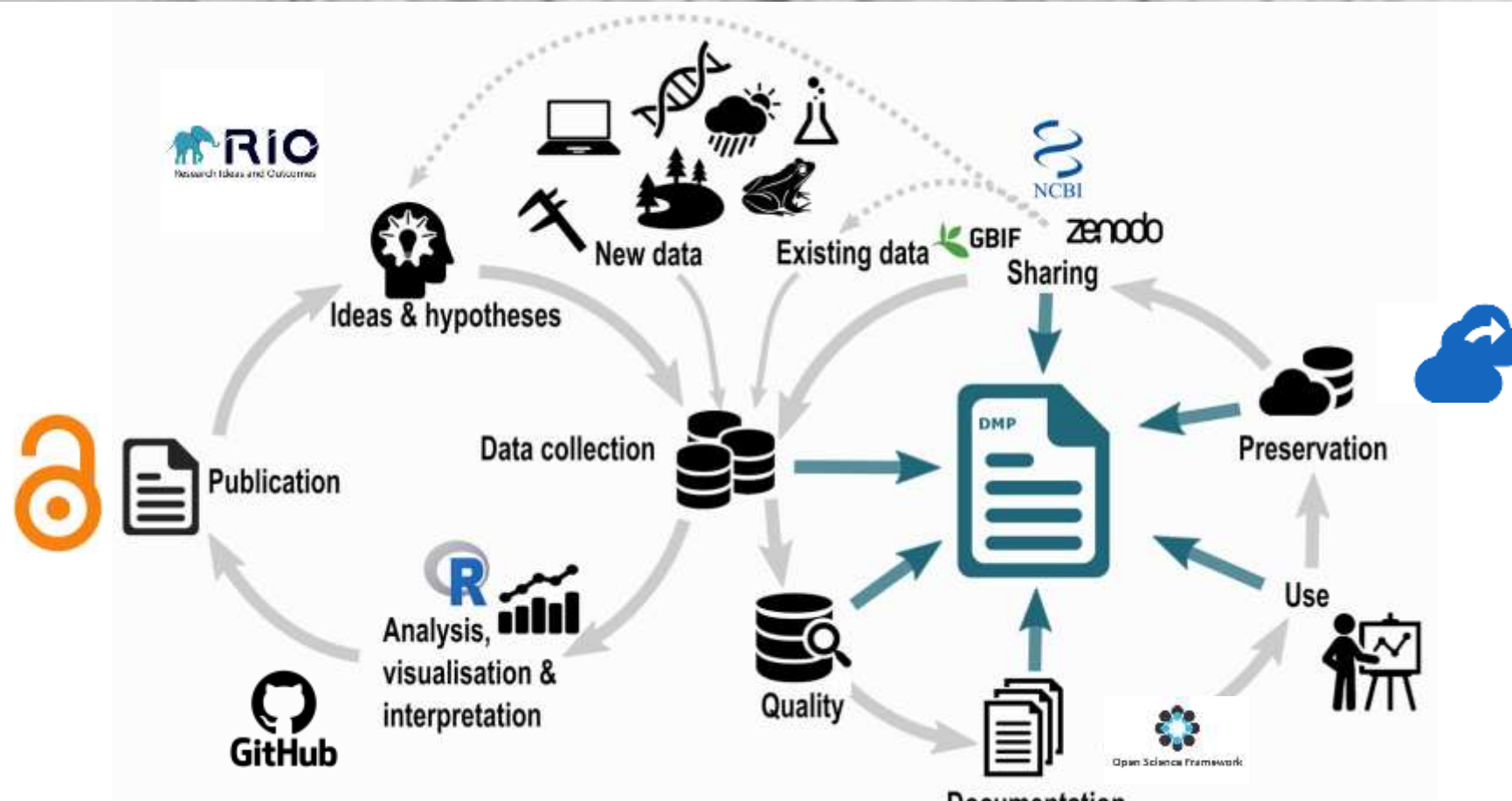




# Research Life Cycle



## Research Life Cycle: Data Management Plan



Thank you....

Some Questions?



# Menti for the attendants

(from 14:45 to 14:55PM)



Go to: [www.menti.com](https://www.menti.com)

**Enter:6394 9460**

**Let's take a break!**

**15h00 – 15h15**

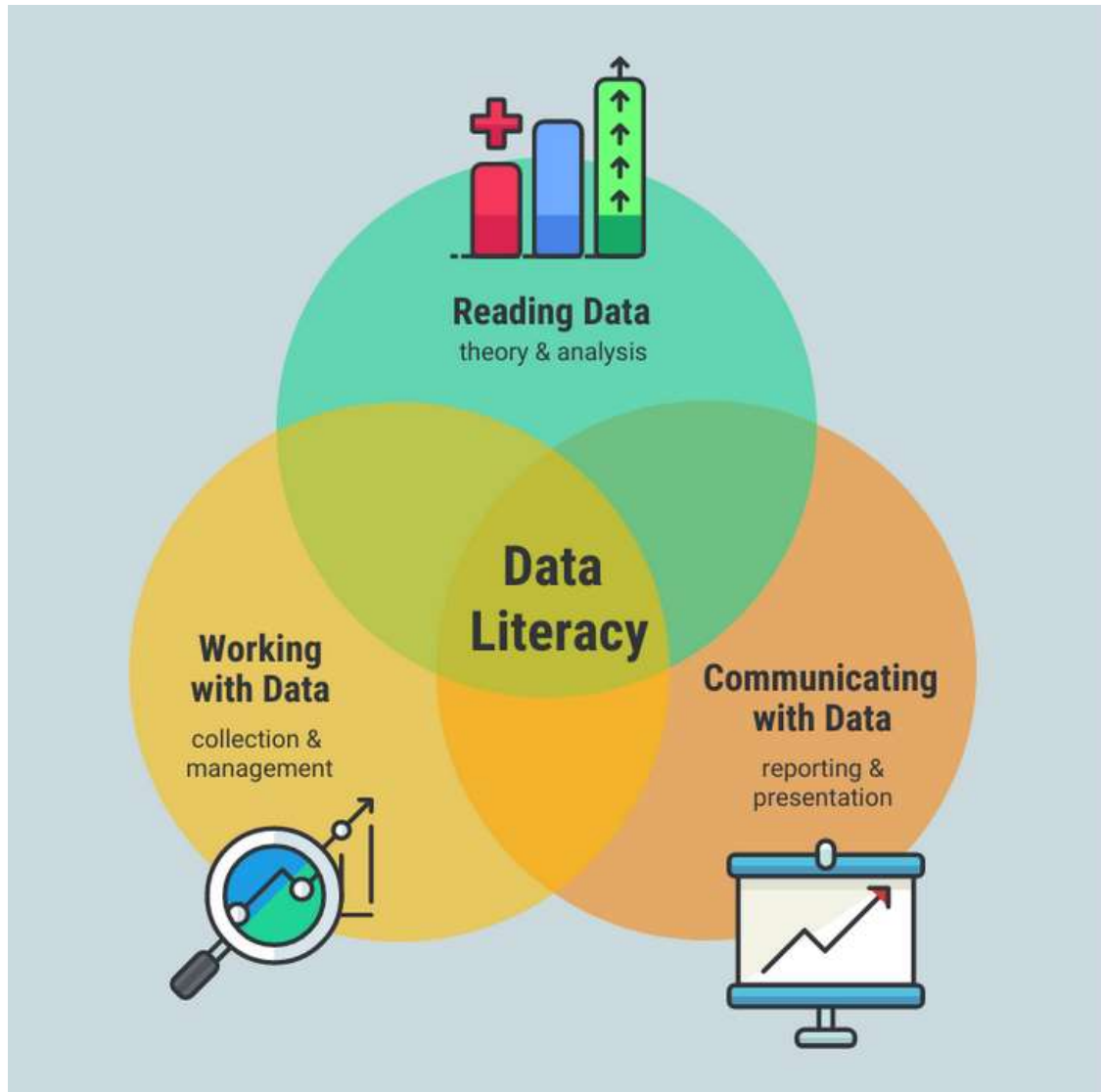


# Data management and open data in GBIF

Prof. Dr. Birgit Gemeinholzer

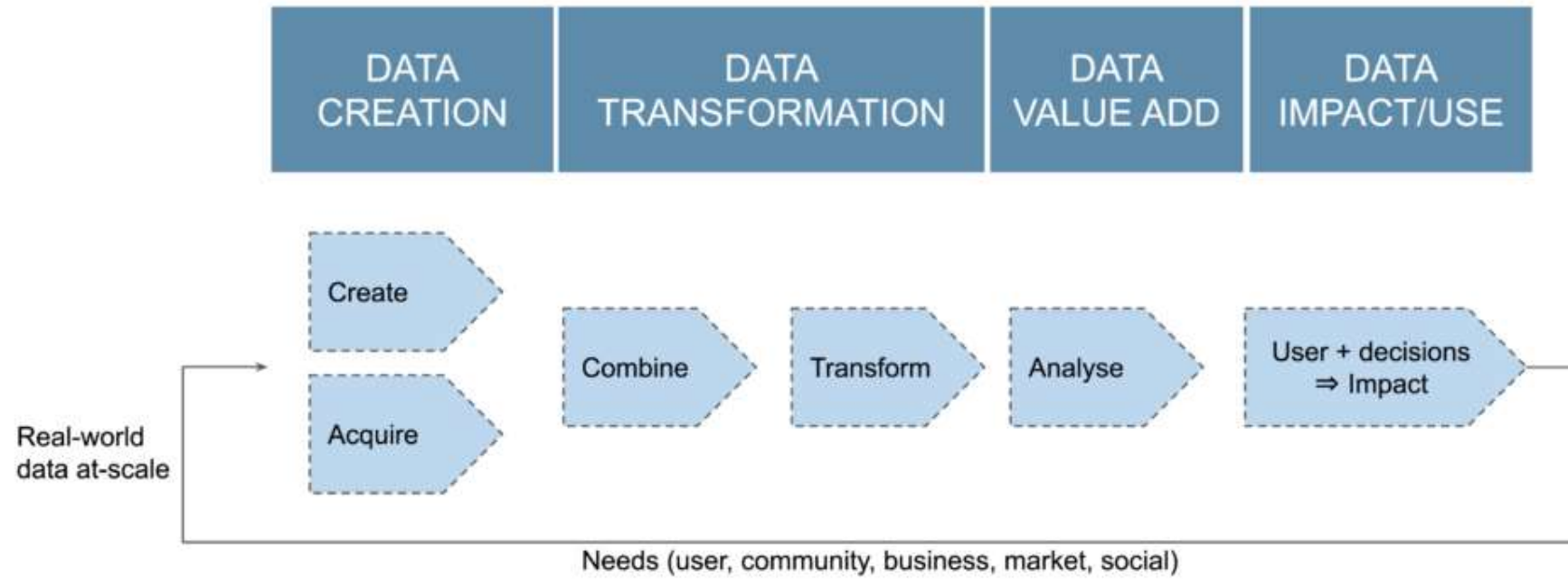
U N I K A S S E L  
V E R S I T Ä T





Integrating research data management into research and teaching.

## The data value chain

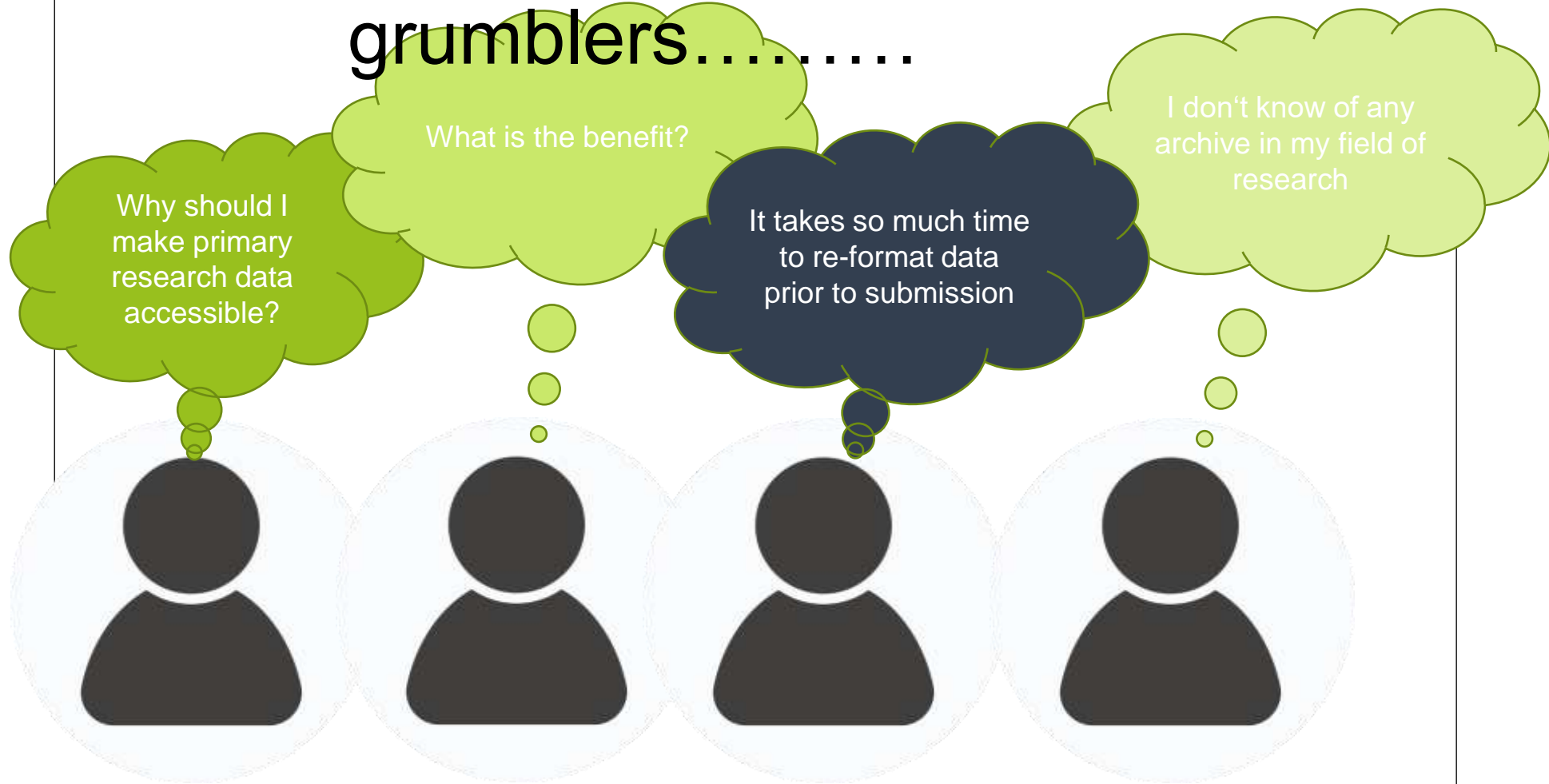


Who is interested  
in my data?



**WE ARE!**

# .....my grumblers.....







<http://www.milkenroar.com/wp-content/uploads/2014/04/us-serbia-eu-illustration.jpg>

The estimated citation counts... 99% higher of all the journal articles

**PLOS ONE**

OPEN ACCESS • PEER-REVIEWED  
RESEARCH ARTICLE

## Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets

Christopher W. Beker 

Published: March 26, 2014 • <https://doi.org/10.1371/journal.pone.0092590>

Article	Authors	Metrics	Comments	Media Coverage
				

**Abstract**

**Introduction**

**Methods**

**Results**

**Discussion**

**Conclusion**

**Acknowledgments**

**Author Contributions**

**References**

**Reader Comments (0)**

**Figures**

**Abstract**

Evaluation of scientific research is becoming increasingly reliant on publication-based bibliometric indicators, which may result in the devaluation of other scientific activities – such as data curation – that do not necessarily result in the production of scientific publications. This issue may undermine the movement to openly share and cite data sets in scientific publications because researchers are unlikely to devote the effort necessary to curate their research data if they are unlikely to receive credit for doing so. This analysis attempts to demonstrate the bibliometric impact of properly curated and openly accessible data sets by attempting to generate citation counts for three data sets archived at the National Oceanographic Data Center. My findings suggest that all three data sets are highly cited, with estimated citation counts in most cases higher than 99% of all the journal articles published in Oceanography during the same years. I also find that methods of citing and referring to these data sets in scientific publications are highly inconsistent, despite the fact that a formal citation format is suggested for each data set. These findings have important implications for developing a data citation format, encouraging researchers to properly curate their research data, and evaluating the bibliometric impact of individuals and institutions.

the average publication tagged with a data-related tag.....achieves a significantly larger citation impact than the average in the field

articles that include statements that link to data in a repository are cited 25.36% ( $\pm 1.07\%$ ) more often on average.

**frontiers in Neuroscience**

published: 10 September 2010  
doi: 10.3389/fnins.2010.00019

**OPEN ACCESS** • **PEER-REVIEWED**  
RESEARCH ARTICLE

## Data Publications Correlate with Citation Impact

Florian Lehner<sup>1,2\*</sup>, Concha Bielza<sup>1,2</sup>, Sean L. Hill<sup>3</sup> and Pedro Larraaga<sup>1</sup>

<sup>1</sup> Computational Intelligence Group, Department for Artificial Intelligence, Universidad Politécnica de Madrid, Madrid, Spain, <sup>2</sup> Data Category S.L., Madrid, Spain, <sup>3</sup> Data Science Project, Campus BioMedic University, Rome, Italy

Neuroscience and molecular biology have been generating large datasets over the past years that are reshaping how research is being conducted. In their wake, open data sharing has been singled out as a major challenge for the future of research. We conducted a comparative study of citations of data publications in both fields, showing that the average publication tagged with a data-related term by the NCI's MeSH (Medical Subject Headings) ontology achieves a significantly larger citation impact than the average in either field. We introduce a new metric, the data article citation index (DAC-index), to identify the most prolific authors among those data-related publications. The study is fully reproducible from an associated R (R Markdown) script together with all the citation datasets. We hope these results can encourage authors to more openly publish their data.


**PLOS ONE**

OPEN ACCESS • PEER-REVIEWED  
RESEARCH ARTICLE

## The citation advantage of linking publications to research data

Eleonora Cordero, Ian Hryniuk, Ina Schab, Kristin Winkler, Barbara M. Olfrey

Published: April 22, 2015 • <https://doi.org/10.1371/journal.pone.0120415>

Article	Authors	Metrics	Comments	Media Coverage	Peer Review
					

**Abstract**

**Introduction**

**Materials and methods**

**Results**

**Discussion**

**Conclusion**

**Data and code availability**

**Acknowledgments**

**References**

**Reader Comments (0)**

**Figures**


Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements at recent literature. Nevertheless, it is still unclear what proportion of these statements actually contain well-formed links to data, for example via a URL, or permanent identifier, and if there is an added value in providing such links. We consider 571, 588 journal articles published by PLOS and BMC, viewing an automatic system for labelling their data availability statements according to four categories based on their content and the type of data availability they display, and finally analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements become very common. In 2016 53.7% of 21,753 PLOS articles and 58.2% of 21,050 BMC articles had data availability statements. Data availability statements containing a link to data in a repository—rather than being available or request or included in supporting information files—are a fraction of the total. In 2017 and 2018, 28.8% of PLOS publications and 12.2% of BMC publications provided DACs containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% ( $\pm 1.07\%$ ) higher citation impact on average, using a citation prediction model. We discuss the potential implications of these results for authors (researchers) and journal publishers who make the effort of sharing their data in repositories. All our data and code are made available in order to reproduce and extend our results.

## Data life cycle





# Data submission .....



SBDI Swedish Biodiversity Data Infrastructure

Search SBDI web site

Support My profile

NEWS AND EVENTS SHARE DATA EXPLORE AND ANALYZE DATA SBDI TOOLS HELP

Open access to Sweden's biodiversity data

Verlässliche Daten sind eine wichtige Grundlage für bessere Beiträge zum Erhalt der weltweiten Artenvielfalt. Dieses Wissen treibt uns an.

Wieder aktivieren

Search About NMIC Content

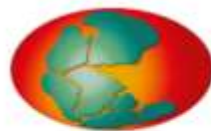
About Norwegian Biodiversity Information Centre

Contact us Publications

Published 10.03.2014 08:36 Started 21.02.2018 10:18

Dutch National Dashboard for Biodiversity

Not logged in



**PANGAEA.**

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT HELP ABOUT CONTACT

Submit

Data

Welcome to PANGAEA® Data Publisher


**INSDC** International Nucleotide Sequence Database Collaboration

ABOUT INSDC
 POLICY
 ADVISORS
 DOCUMENTS





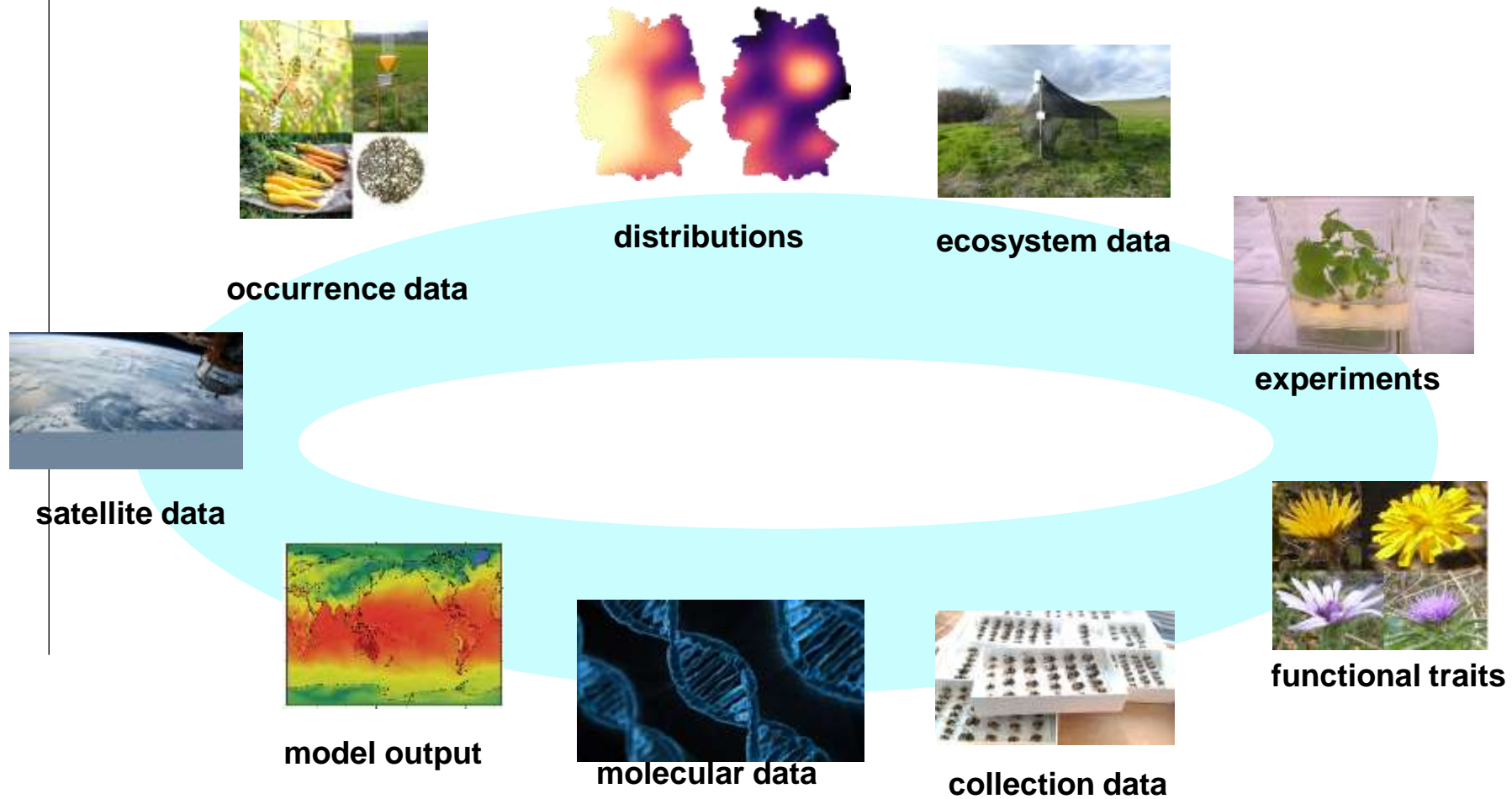
**International Nucleotide Sequence Database Collaboration**

- The International Nucleotide Sequence Database Collaboration (INSDC) is a long-standing foundational initiative that operates between [DDBJ](#), [EMBL-EBI](#) and [NCBI](#). INSDC covers the spectrum of data raw reads, through alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations.

Data type	DDBJ	EMBL-EBI	NCBI
Next generation reads	<a href="#">Sequence Read Archive</a>	European Nucleotide Archive ( <a href="#">ENA</a> )	<a href="#">Sequence Read Archive</a>
Capillary reads	<a href="#">Trace Archive</a>		<a href="#">Trace Archive</a>
Annotated sequences	<a href="#">DDBJ</a>		<a href="#">GenBank</a>
Samples	<a href="#">BioSample</a>		<a href="#">BioSample</a>
Studies	<a href="#">BioProject</a>		<a href="#">BioProject</a>

**no scientific publication without primary data submission**

# Facets of data providers



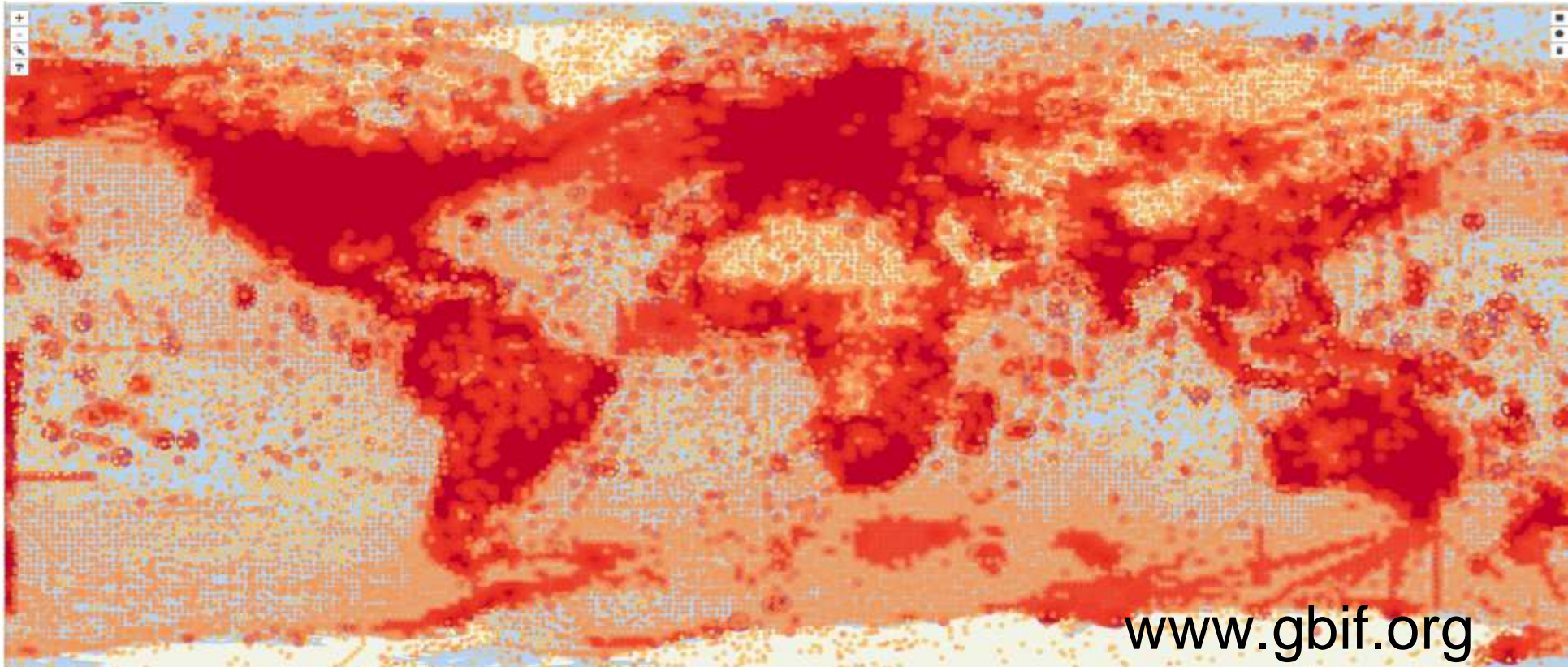


# What is GBIF?

The **G**lobal **B**iodiversity **I**nformation **F**acility is an intergovernmental network and research infrastructure funded by the world's governments.

GBIF provides anyone, anywhere, free and open access to data about all types of life on Earth.

Voluntary collaboration through Memorandum of Understanding, supported by participants nodes and a secretariat in Copenhagen/Denmark.



# Currently GBIF holds

2.951.116.659

Occurrence records

105.561

Datasets

2.233

Publishing institutions

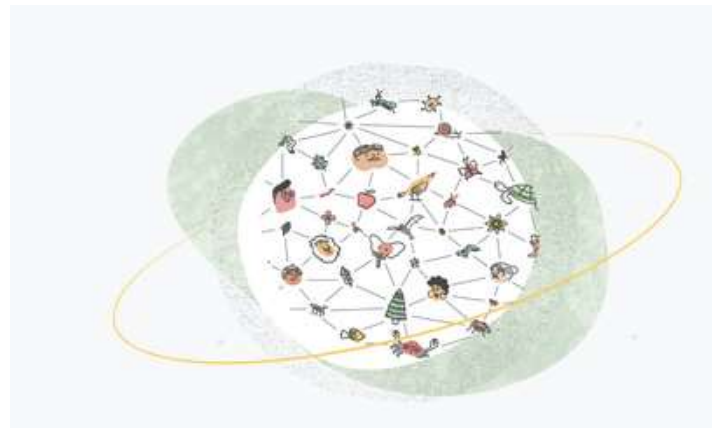
10.599

Peer-reviewed papers  
using data

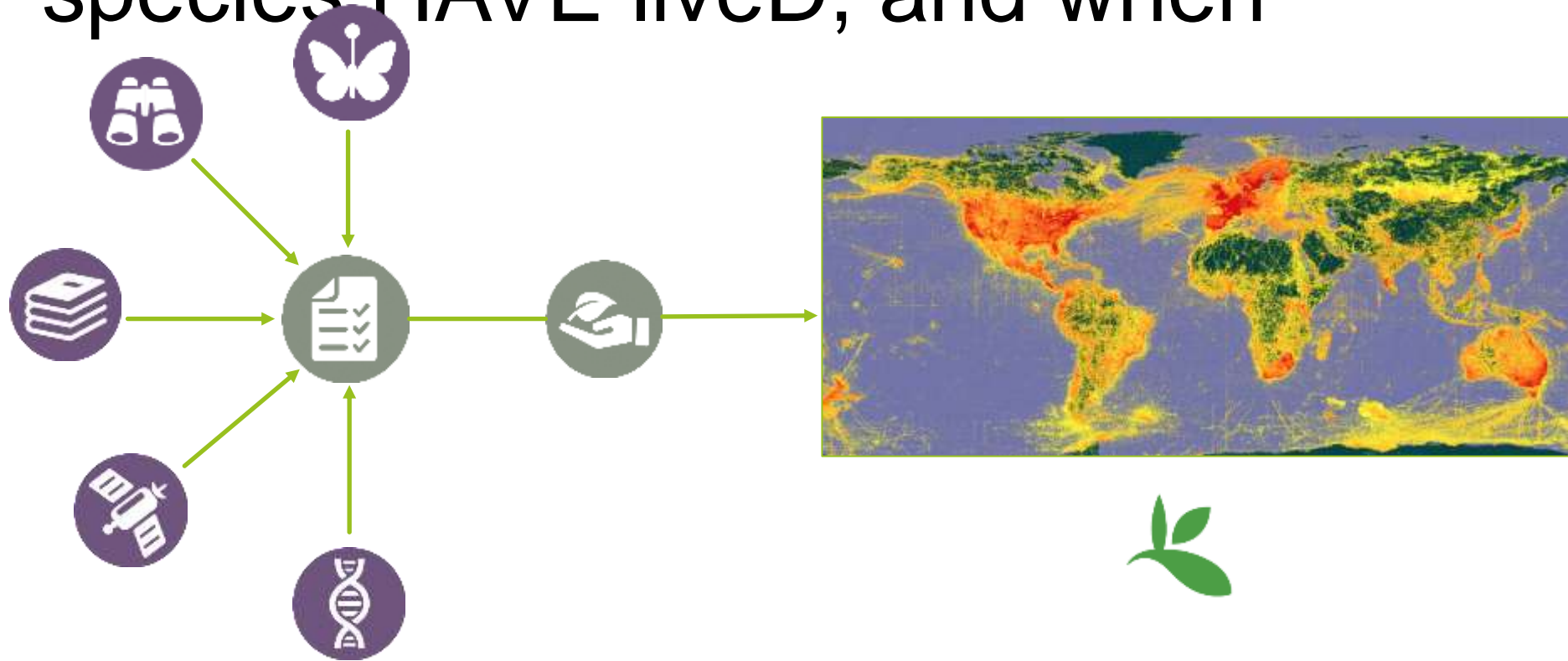


<https://www.gbif.org/analytics/global>

Accessed: 04.06.2024

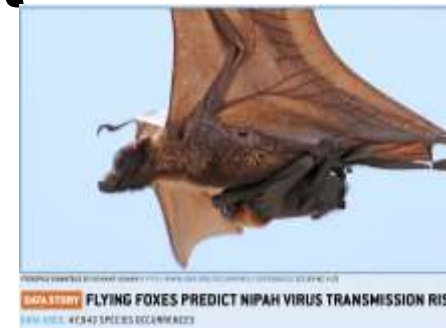
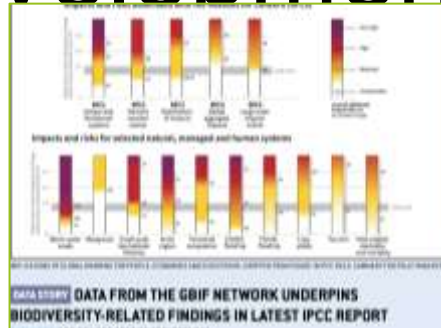


# A window on evidence about where species HAVE lived, and when





# A data resource to support research and sustainable development



## Conservation

- Protected areas
- Threatened species
- Invasive species risk

## Food Security

- Crop wild relatives
- *In situ*, *ex situ* conservation of genetic diversity
- Fisheries planning

## Climate change

- Modelling impacts on species ranges
- Adaptation strategies
- Mitigation benefits, risks

## Human health

- Disease risk based on occurrence of vectors, hosts, reservoirs
- Medicinal plants
- Hazards e.g. snakebite

# A collaborating community of practice

- Open-access tools and guidance
- Training and mentoring
- Nodes staff, partners sharing skills within countries and across continents
- Funded projects for capacity enhancement



Images (top to bottom): Tim Hirsch; Maheva Bagard Laursen; Mélanie Raymond



etaplantcode

Harmonizing plant metabarcoding pipelines in Europe  
to support monitoring activities in the field of plants  
and their functional organismic networks

[Home](#)

[News and Events](#)

[About us](#)

[People](#)



Natural History  
Museum

Hes-500  
Institut de Recherche  
pour le Développement  
FRANCE



INER  
INSTITUT NATIONAL DE RECHERCHE  
EN BIOLOGIE  
Université de la Méditerranée  
Centre National de la Recherche Scientifique

Swiss National  
Science Foundation

NWO  
Netherlands Organisation for Scientific Research



N  
Naturalis  
Biodiversity  
Center

belspo

The Research Council  
of Norway

SIB  
Swedish Institute of  
Biodiversity

uefiscodi

fct

Fundação  
para a Ciência  
e a Tecnologia

KU LEUVEN



UNIKASSEL  
VERSITÄT

DFG  
Deutsche  
Forschungsgemeinschaft

SBDI

RÉPUBLIQUE  
FRANÇAISE  
Ministère  
de la Recherche  
et de l'Enseignement  
Supérieur

anr  
agence nationale  
de la recherche  
sur l'environnement et le climat

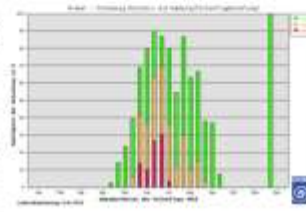
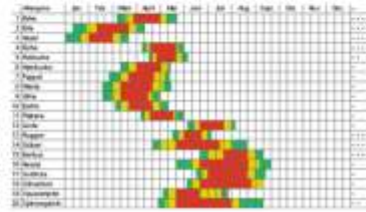
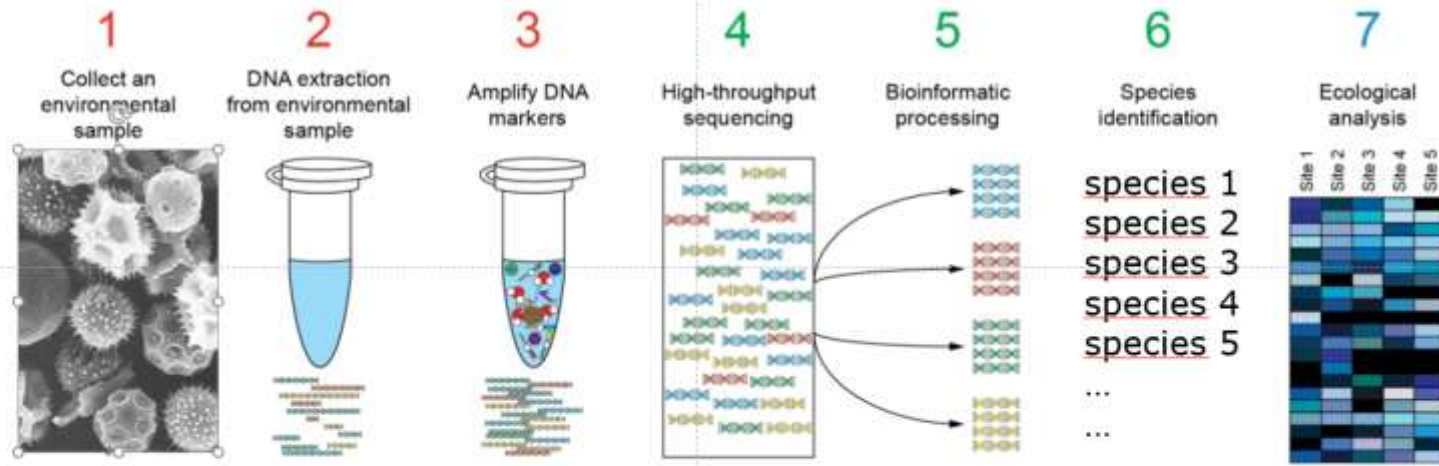
INAB  
Institut National  
de l'Analyse Biologique

CERTH  
Centre for  
Research and  
Technology  
HELAS

Co-funded by  
the European Union

biodiversa+  
European Biodiversity Partnership





## Invitation to share DNA metabarcoding data to test early pilot of data-publishing tool

- Data Publishing ■ Diversifying the GBIF data model

**T**tfroeslev

1 Jun 2023

Jun 2023

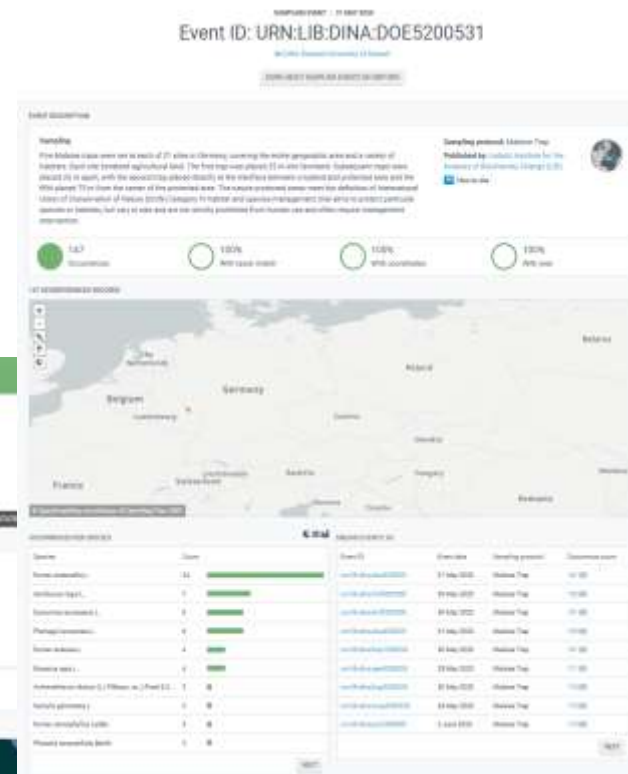
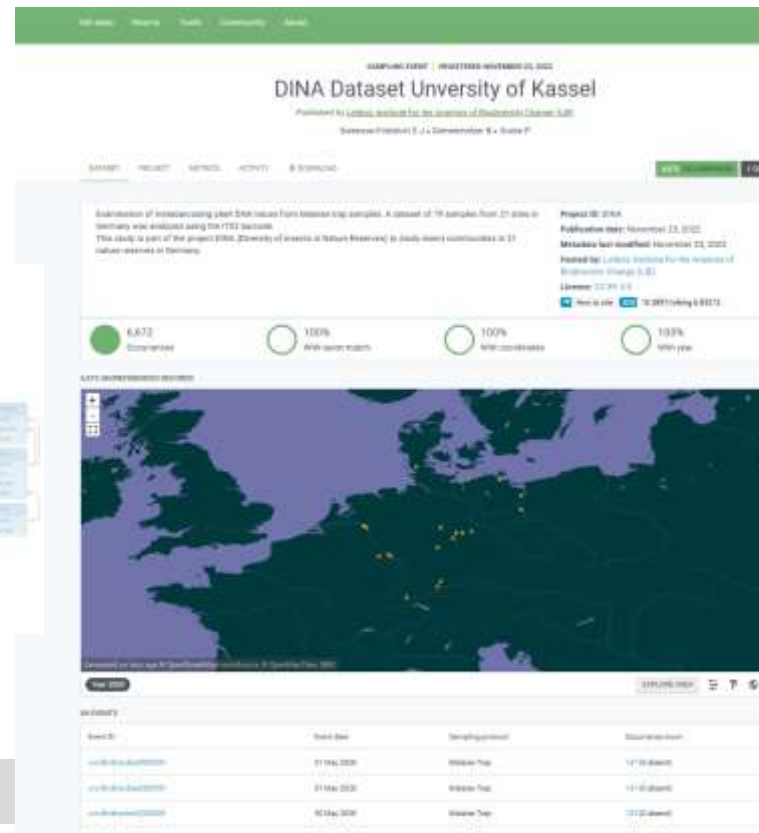
1/1  
Jun 2023

Metabarcoding of environmental DNA samples or bulk samples is one of the major sources of new biodiversity data, and GBIF is exploring ways to expand its support of communities interested in publishing DNA-derived biodiversity data and increase its visibility and reuse beyond molecular repositories and archives.

On the occasion of [updating our guide](#) on sharing such data through biodiversity platforms—which now includes a special section on publishing marine eDNA data—GBIF invites people who hold DNA metabarcoding data to help us pilot an experimental data publishing tool that responds to recent feedback from the omics community.



<https://www.gbif.org/event/33MQ8Zi9xuCS8AekwiMr8/gbif-community-webinar-diversifying-the-gbif-data-model>



www.gbif.org



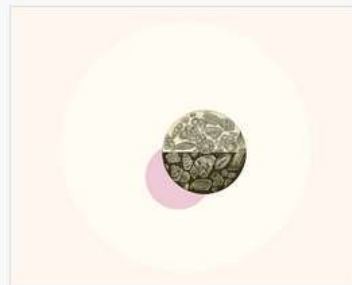
Further information on GBIF  
and DNA



Guidelines for publishing  
DNA-derived data through  
biodiversity data platforms



Invitation to share DNA  
metabarcoding data in  
data-publishing tool pilot



Download webinar  
presentations



New data-clustering feature  
aims to improve data  
quality and reveal cross-  
dataset connections



Five francophone GBIF  
nodes complete first  
translation of guide on  
sharing DNA-derived  
occurrence data



Cross-infrastructure  
collaboration with ENA  
improves processing,  
quality of DNA-derived  
occurrences



Adding sequence-based  
identifiers to backbone  
taxonomy reveals 'dark  
taxa' fungi



## Humboldt Extension

This task group explored concepts and methods of species inventories to fully integrate this type of observational data into existing data exchange schemas. To support proper capture of key information about inventories, current Humboldt extension terms (Guralnick et al. 2018) were reviewed, revised as necessary, and integrated into an extension of the Darwin Core. The TDWG Executive Committee ratified the Humboldt Extension on 2024-02-28. The resulting terms have been incorporated into the Darwin Core and can be found in the Darwin Core, with additional documentation and recommendations on the website linked below. The outcome provides a framework and clear semantics for sharing and integrating biodiversity inventory data.

[Website](#) [GitHub](#)

Last modified: 2024-05-06

### Completion notice

**This Task Group has completed its work and has been dissolved.** Please see the results on:

- Darwin Core, Humboldt Extension Vocabulary List of Terms
- Humboldt Extension documentation

The terms created as by the task group have been incorporated into the Darwin Core and will be maintained by the Darwin Core Maintenance Group. Future task groups may be formed to continue development of the extension if usage experience indicates significant additional work would be beneficial.

This page will be archived under the Humboldt Extension website within a year of ratification date, above.

### Convener

- Yanina Sica - Map of Life, Yale University, USA

On this page

[Completion notice](#)

[Convener](#)

[Core members](#)

[Moderator](#)

[Goals, outputs and outcomes](#)

[Strategy](#)

[Becoming involved](#)

[History and context](#)

[Resources](#)

### Implementation of the vocabulary: testing

Use case: eBird Volunteer–Collected Observations of Birds

Use case: Field Museum Rapid Inventory Data

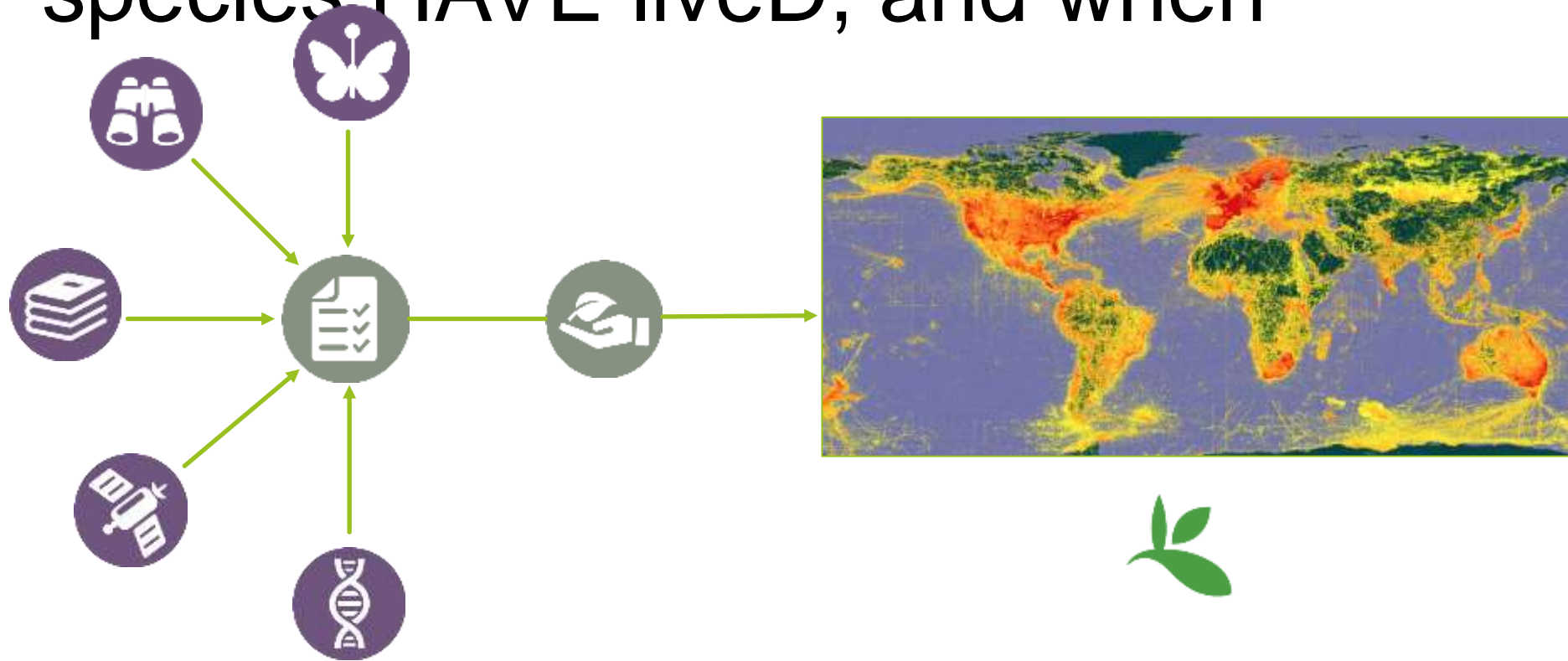
Use case: Distribution of squid and fish in the pelagic zone of the Cosmonaut

Sea and Prydz Bay region during the BROKE-West campaign – data

Use case: Hummingbirds of the Northern Andes

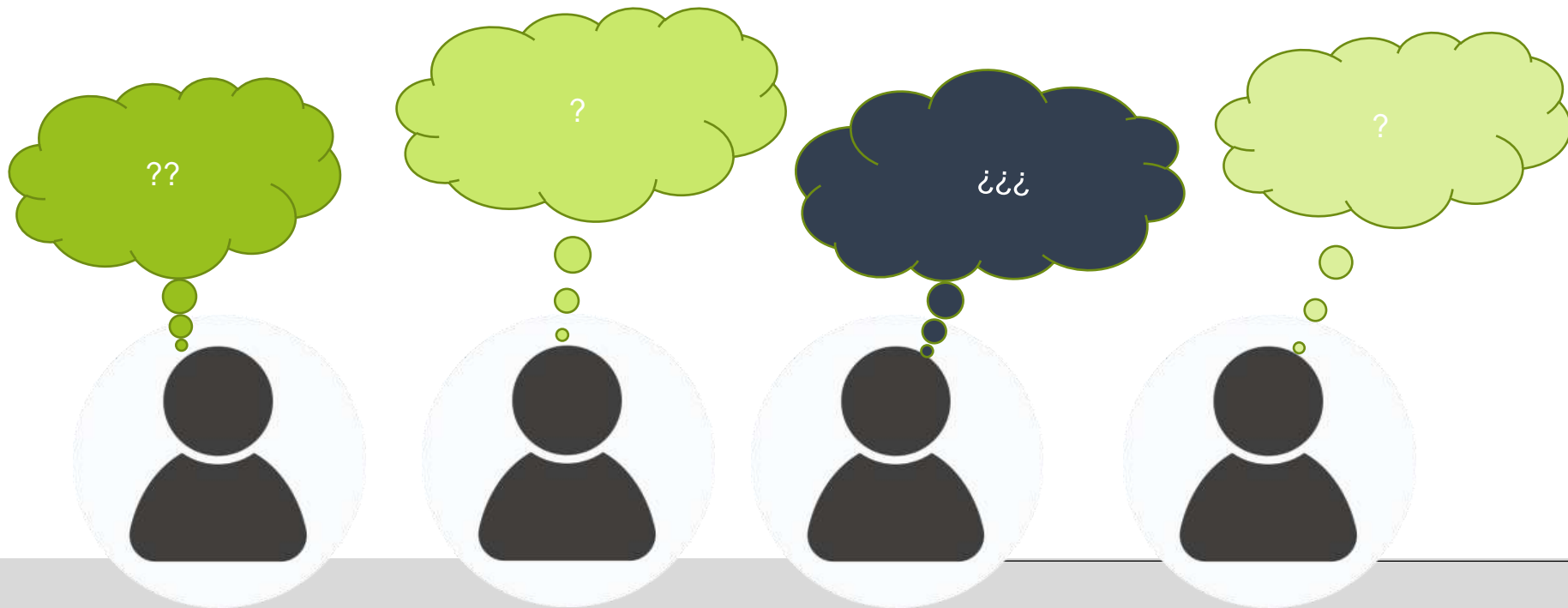
and others.....

# A window on evidence about where species HAVE lived, and when



..... do you still think that your research data is of now value for you and for others and there is nobody interested in it?

**Thank you for your interest!**







**GBIF**

Global Biodiversity  
Information Facility

# Introduction to GBIF training package on data mobilization

Melissa Liu | Asia Regional Support Team



Photo by Melissa Liu. CC BY-NC 4.0

# Gbif strategic plan

01

02

03

04

Develop solutions that increase meaningful participation in the GBIF community by reducing linguistic barriers and providing capacity development and programme support.

<https://www.gbif.org/strategic-plan>



2023 - 2027

# GBIF Strategic Framework

# Capacity Development objectives

**Developing the capacity to mobilize biodiversity data through GBIF**



the ability to manage biodiversity data, standardize and publish it through GBIF

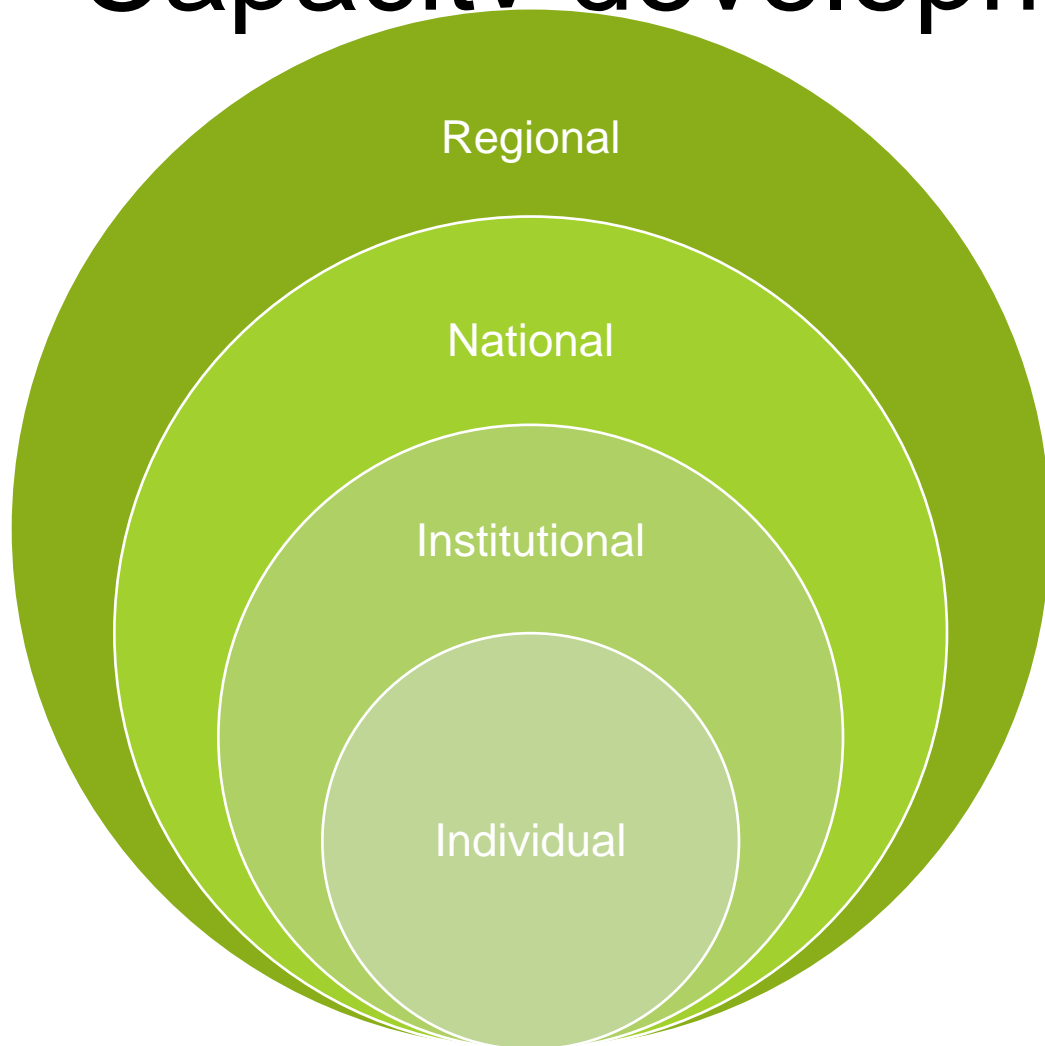
**Developing the capacity to use GBIF-mediated data**



the ability to access, analyze and use biodiversity data accessible through GBIF in scientific research and decision making



# Capacity development levels



Coordinate and raise the visibility of the GBIF community of practice and network



Develop national biodiversity information facilities



Build partnerships that drive the institutionalization of biodiversity data mobilization and use of GBIF-mediated data



Strengthen and grow GBIF's community of practice

Ambassadors

Mentors

# Individual level

Trainers

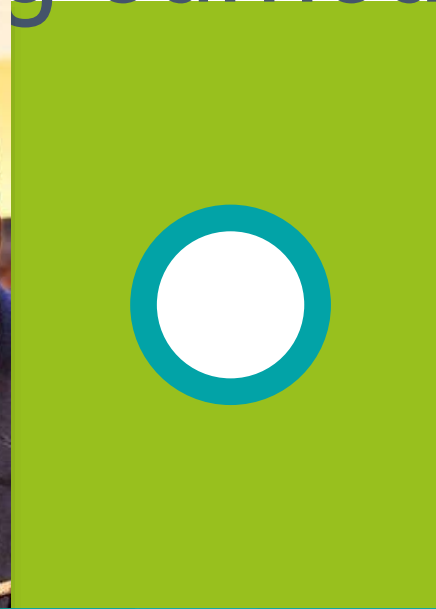
Translators

Strengthen and grow GBIF's community of practice





# Consolidated training curricula



Development of curricula and courses for self-paced use, onsite or online workshops

- Introduction to GBIF
- Biodiversity Data Mobilization
- Biodiversity Data use
- Accelerating biodiversity research through DNA barcodes, collection and observation data
- Establishing a GBIF Participant Node



# Multiplier effect

**15** GBIFs- led training events on **Data Mobilization** and **Data Use** organized through BIFA and BID

Over **370** people trained during these workshops



Data mobilized through BIFA, BID and CESP projects have already been cited in over **1,700** peer-reviewed papers



Over **130** replication workshops using curricula developed by GBIF on Data mobilization and Data use organized by BIFA and BID projects

Over **3,870** professionals across Asia, Africa, the Pacific and the Caribbean during these workshops

GBIF nodes can adapt these curricula to their national and regional context through CESP projects



Photo @arochakenya - <https://pic.twitter.com/kRIulUYdNr>

# METRICS IN 2022-2023 virtual workshop



European Biodiversity Partnership

Need training in data mobilization?

**305** said interested



New Datasets

**14** ~2000 records



Response for application form



**131** applicants

Workshop  
Participants

**60**

Self-Paced  
Participants

**27**



New Publishers

**12**

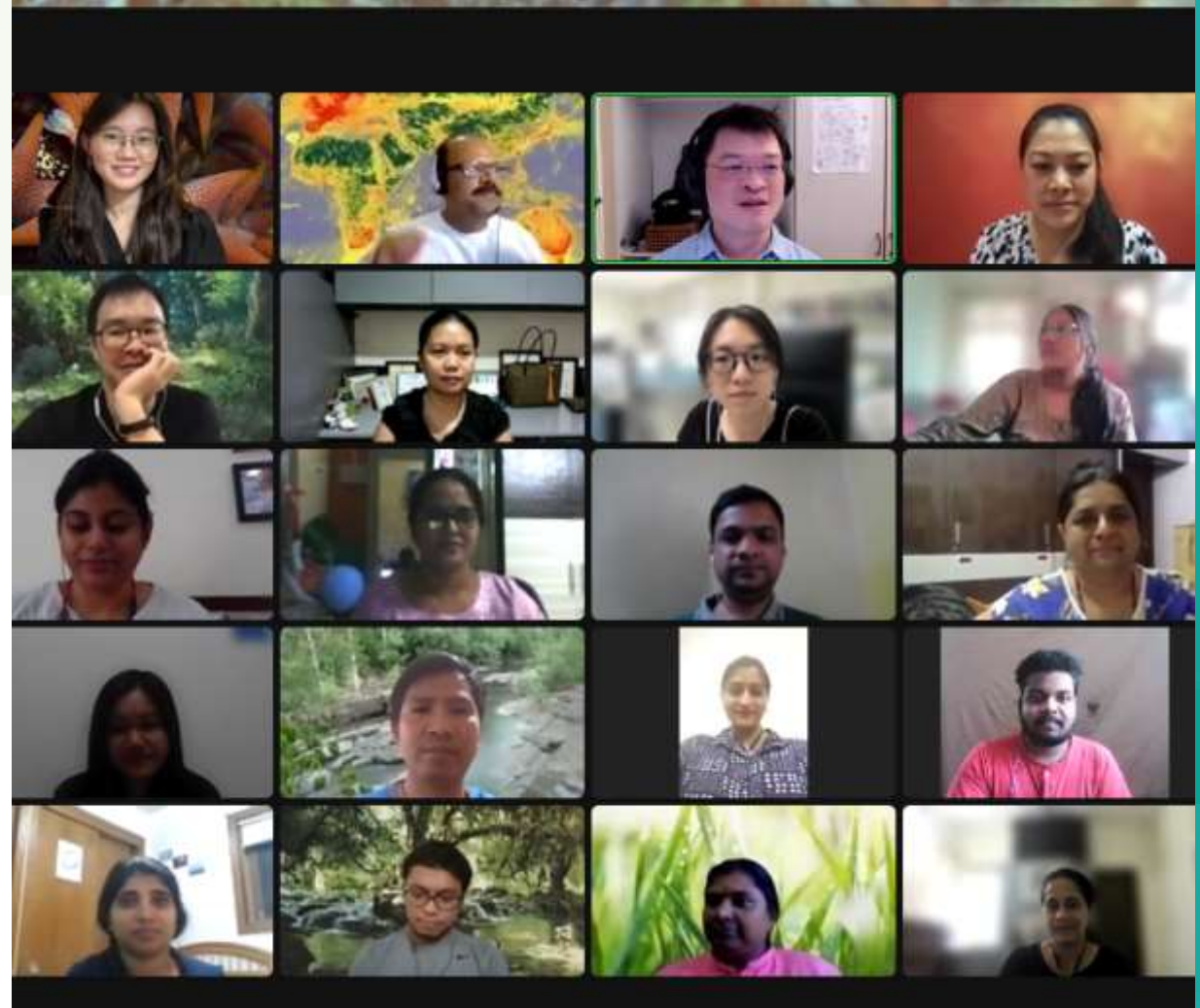
from

**14** countries



# Organizing a virtual regional data mobilization workshop

- ✓ Tasks for organizing a virtual workshop
- ✓ Resources for use
- ✓ Workshop modules
- ✓ Grading & Badges





# organizing workflow



## 01 Interest Survey

Investigate the willingness of participation from Asia community.



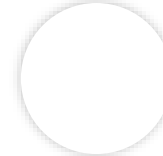
## 02 Decide module & date

Decide how and when to organize the workshop.

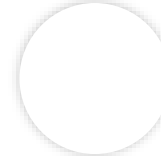


## 03 Announce

Design an application form and spread it out to potential communities.



Set criteria and filter participants when the number of applicants exceeds the limit.



Contact previous workshop mentors, and organize a coordinating group.



# Interest Survey

- ❑ create a google form
- ❑ send emails to communities
- ❑ ask participants in Asia office hour

## 2023 Asia Data Mobilization Workshop Interest Survey

**B** *I* U  

One of the challenges to open access biodiversity data in Asia is the lack of technical capacity among data holders in data mobilization process. GBIF Biodiversity Data Mobilization workshop intends to enhance the capacity of data holders in project management, data capture, data management and data publication, focusing on GBIF data infrastructure. We have received enquiries about the opportunity to participate in a GBIF data mobilization workshop and acquire a badge. This survey is thus, designed to explore interest among the Asian biodiversity community for the workshop.

We are aiming the virtual workshop in 2nd week of August 2023 should enough interests are expressed. It would be conducted in English with some translation support provided by mentors from various Asian countries. Participants would be expected to allocate at least 4 hours per day for 5 days for the workshop in which we arrange self-paced video learning and online zoom sessions for questions, answers and demonstrations.

Please submit your interest as soon as possible, if you intend to participate should the workshop become available. Thank you very much.

# Decide module & date

## 2023 Virtual Data Mobilization Workshop

August 07 - 11, Asia

Date/Time	Jul 31	Aug 01	Aug 02	Aug 03	Aug 04
	Self-paced learning and material review				
Date/Time	Aug 07 (Monday)	Aug 08 (Tuesday)	Aug 09 (Wednesday)	Aug 10 (Thursday)	Aug 11 (Friday)
Hour 1 12:00 Tokyo (GMT+9) 11:00 Manila/Taipei (GMT+8) 10:00 Indochina/Jakarta (GMT+7) 08:45 Nepal (GMT+5:45) 08:30 Mumbai (GMT+5:30) 08:00 Islamabad (GMT+5:00) 05:00 Copenhagen (GMT+2) 03:00 UTC 20:00 Los Angeles (GMT-7)	Zoom Session - 1  Welcome and introductions  Foundations & QA Planning introduction  Leads: Chihjen, Asia supporters / Melissa	Zoom Session - 2  Planning activity, discussion and QA  Lead: Melissa	Zoom Session - 3  Data capture discussion  Lead: Lily	Zoom Session - 4  Data management activity, OpenRefine activity  Lead: Chihjen	Zoom Session - 5  Data publishing practice and Q&A  Lead: Vijay
Hour 2					
Hour 3					Zoom Session - 6
Hour 4	Self-paced: Foundations(Review); Planning	Self-paced: Data capture	Self-paced: Data management, OpenRefine	Self-paced: Data publishing; Review Use Case II & III	Final review, conclusion & assignments & Feedback (google form)  Leads: Asia supporters
Date/Time	Aug 14 - Sep 03	Self-paced certification exercises			
Date/Time	Sep 04 - Sep 29	Grading and awarding of digital badges (trainers)			

- ❑ Abide by the previous training module developed by Laura
- ❑ Decide the length of the workshop
- ❑ Settle the agenda
- ❑ Set up a poll to find the possible dates





# Announcement



GBIF online data mobilization course for Asia - now open for registration [GBIF/GBIF-Asia support work](#)

Mélanie Raymond via Nodes\_lg <nodes\_lg@gmail.com> 2022年8月17日 星期三  
寄件: Asia - Training

Dear Nodes,

The [Asia regional support team](#) is organizing a GBIF data mobilization workshop in September that is from any current or potential new data publishing institutions in the Asia region.

Please see details and registration information on the [event page](#).

Strengthening engagement in Asia remains an important priority for GBIF, and we invite you to share it with any relevant contacts you have from the region.

The course will use GBIF's curriculum for the [biodiversity data mobilization course](#) that has been designed and easily reused by anyone wishing to run a virtual or in-person workshop on this topic (and is available in Portuguese). Please remember that you can find [GBIF's training and learning resources on the website](#) to contact [training@gbif.org](#) if you have any questions about running GBIF-related workshops for your



Send emails to inform the respondents from the interest survey.

- Registered publishers
- Nodes & institutions that have already made contacts
- Past project teams

Inform nodes to spread it out.

- Ask GBIFS to send email to nodes

Announce an event on GBIF.org.

- Suggest an event and provide detailed information
- Example: <https://bit.ly/gibf-2022-virtual-workshop-for-Asia>

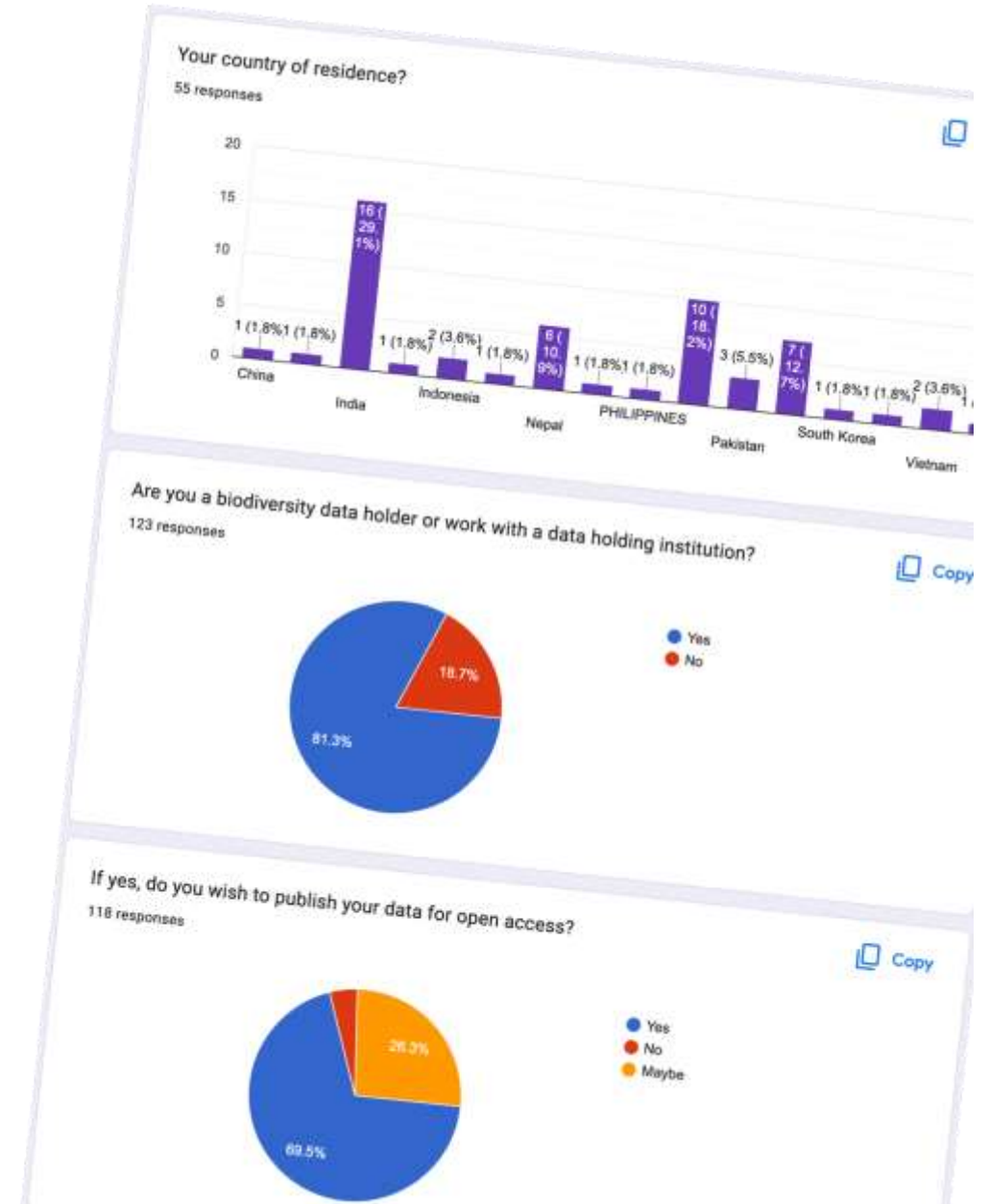
Announce in Asia office hour/ social media.

- Regional virtual office hour
- Facebook, Twitter, etc.
- Mailing lists
- Any other contacts

# Select participants

Select if the applicant:

- ✓ has datasets to publish soon
- ✓ is a data holder or work with data-holding institution
- ✓ has never participated in related workshops
- ✓ plays an important role from nodes or can fill the regional gap
- ✓ plan to hold replication workshops or wants to become a mentor and support future workshops



# Recruit mentors

## Mentors and trainers

*By volunteering their knowledge and expertise, GBIF mentors help others within the community build the skills needed to achieve the GBIF network's common goals*



Volunteer mentors from INBio Costa Rica and GBIF Spain mentor BID Africa 2015 project participants. (From left to right) Maria Auxiliadora Mora, Sara Raquel Figueira Fernandes, Gladys Odey Schwinger, Gilbert Muvunankiko and Katia Cezón Garcia. Photo 2015 GBIF | Mélanie Raymond.



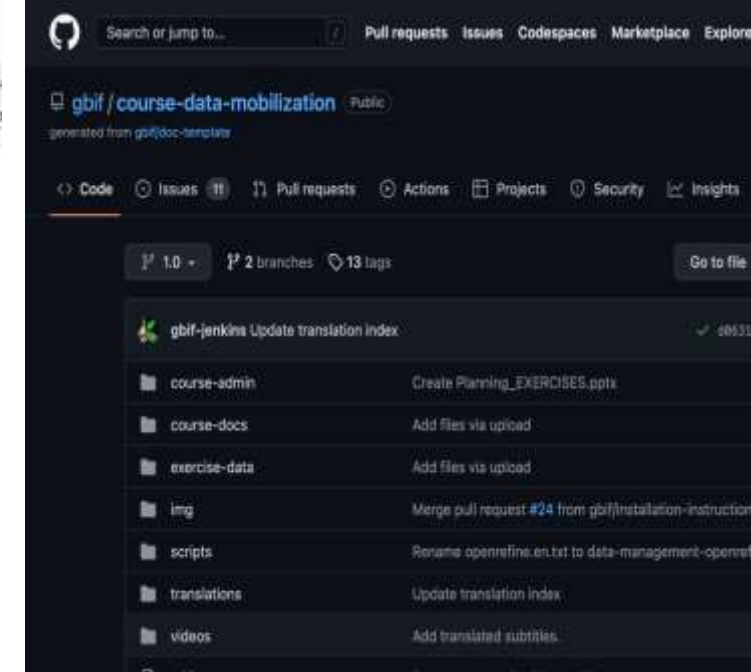
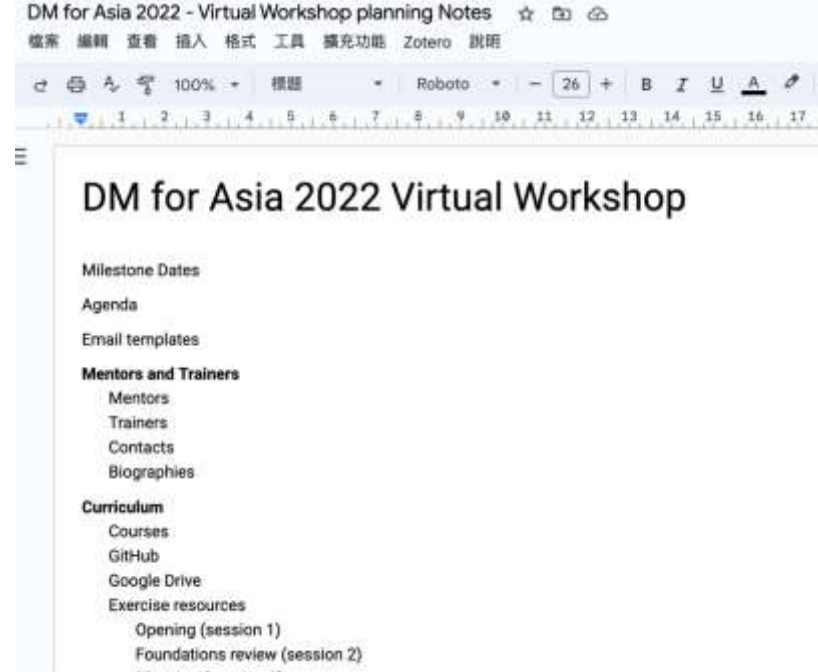
# Biodiversity Data Mobilization Course

GBIF Secretariat – [training@gbif.org](mailto:training@gbif.org) – Version 12, May 2021

This document is also available in [PDF format](#) and in other languages: [español](#), [français](#), [Português](#).



*Zonocerus elegans* (Thunberg, 1815) observed in Zimbabwe by nickypegg (licensed under [CC-BY-NC 4.0](#))



- ❑ [Biodiversity Data Mobilization Course](#): For participants to learn in self-paced and for trainers to refer to the rubrics of each assignment.
- ❑ [Virtual workshop planning notes](#): For organizers to know how to organize a workshop and get all the information needed.
- ❑ [GitHub](#): For trainers to download all the educational resources.
- ❑ Interested in material translation? please contact [Training@gbif.org](mailto:Training@gbif.org)

# Workshop module

The workshop is based on the GBIF Biodiversity Data Mobilization Course and has three phases of activities:

- Week 1: preparatory course activities (online self-paced)
- Week 2: virtual workshop (Zoom sessions and online self-paced, 4 hours per day)
- Week 3: certification exercises (online self-paced)

## 2023 Virtual Data Mobilization Workshop

August 07 - 11, Asia

Date/Time	Jul 31	Aug 01	Aug 02	Aug 03	Aug 04
	Self-paced learning and material review				
Date/Time	Aug 07 (Monday)	Aug 08 (Tuesday)	Aug 09 (Wednesday)	Aug 10 (Thursday)	Aug 11 (Friday)
Hour 1 12:00 Tokyo (GMT+9) 11:00 Manila/Taipei (GMT+8) 10:00 Indochina/Jakarta (GMT+7) 08:45 Nepal (GMT+5:45) 08:30 Mumbai (GMT+5:30) 08:00 Islamabad (GMT+5:00) 05:00 Copenhagen (GMT+2) 03:00 UTC 20:00 Los Angeles (GMT-7)	Zoom Session - 1  Welcome and introductions  Foundations & QA Planning introduction  Leads: Chihjen, Asia supporters / Melissa	Zoom Session - 2  Planning activity, discussion and QA  Lead: Melissa	Zoom Session - 3  Data capture discussion  Lead: Lily	Zoom Session - 4  Data management activity; OpenRefine activity  Lead: Chihjen	Zoom Session - 5  Data publishing practice and Q&A  Lead: Vijay
Hour 2					
Hour 3					Zoom Session - 6
Hour 4	Self-paced: Foundations(Review), Planning	Self-paced: Data capture	Self-paced: Data management, OpenRefine	Self-paced: Data publishing; Review Use Case II & III	Final review, conclusion & assignments & Feedback (google form)  Leads: Asia supporters
Date/Time	Aug 14 - Sep 03	Self-paced certification exercises			
Date/Time	Sep 04 - Sep 29	Grading and awarding of digital badges (trainers)			

# Grading & Badges

- Trainers review all the exercises submitted by participants.
- Refer to the rubrics of each topic and grade the exercises
- Write all the scores in the sheet (template provided by Laura) and see who can get the badge

## Data capture rubric

### Data capture

Skills	Beginning performance 1	Developing performance 2	Accomplished performance 3	Outstanding performance 4
A. Ability to identify the type of digital data that can be extracted from a source of biodiversity data (i.e. that can be published using the GBIF network)	Can identify only the most evident data types from common sources of biodiversity data (e.g. occurrences from natural history collection specimens). Shows little understanding of potential for online publishing using GBIF.	Can frequently identify correctly, at least one digital data type that can be extracted from common sources of data. Has difficulty identifying which ones can be currently published using GBIF.	Can always identify one (or more) types of digital data that can be extracted from common sources of data. Can identify which one of those types can be currently published using GBIF.	Can always identify one or more types of digital data that can be extracted from common and uncommon sources of data. Can identify which one of those types can be currently published using GBIF and which ones are under discussion. Can identify data cores and extensions used for publishing those data types.
B. Capacity to extract relevant information from a source of biodiversity data into simple data structures (e.g. spreadsheets) that follows international standards	Can only extract large pieces of obvious information (e.g. all geographic information as a single unit) which are evident in the data source. Shows little knowledge of current standards for recording biodiversity data.	Can retrieve several information items from the data source (but not all) and can disaggregate them into meaningful pieces. Shows some basic knowledge of the most common standards (e.g. DwC) and the most used data fields in those standards.	Can identify all valuable information in a data source, and extract the mandatory elements in a standard data structure (e.g. a spreadsheet based on Simple DwC). Can identify missing information and infer from existing information (e.g. derive a country name from a province).	Can identify all valuable information in a complex data source, and divide it into meaningful pieces which then translate directly into international standards. Can identify critical information missing in the source and infer it from the existing data or from additional information about the source (metadata).
C. Ability to understand and apply basic principles of data quality to the data capture process	Shows limited understanding of how applying simple data quality principles can have a large impact on the final product, preventing additional required cleaning afterwards.	Knows some of most generic principles of data quality (e.g. avoid misspellings) but shows limited knowledge on how to apply more specific principles to the data capture process.	Knows all the basic principles of data quality and how to apply these in simple ways to the data capture process. Uses formats consistently during the data capture process (e.g. in dates, country names). Documents all procedures and changes connected to data quality in a simple manner.	Shows good knowledge of all common principles of data quality and how to use them to improve the data capture process. Uses data formats consistently and can use gazetteers, reference lists, or software-specific features to improve quality from the original. Documents clearly all changes and decisions taken in connection to data quality.



# Grading & Badges

- Criteria of badges
  - Basic badge: Average score above 2.5/4
  - Advance badge: Average score above 3/4
- Benefits of badges
  - Understandings of skill levels
  - Better quality of support



- Encourage participants to publish datasets to GBIF after the workshop.
  - Data publishing to GBIF.org within a week
  - Get some prizes
- Mail-in GBIF goodies from GBIFS
  - Contact Annie Elkjær Ørum-Kristensen ([akristensen@gbif.org](mailto:akristensen@gbif.org))







Becoming a mentor

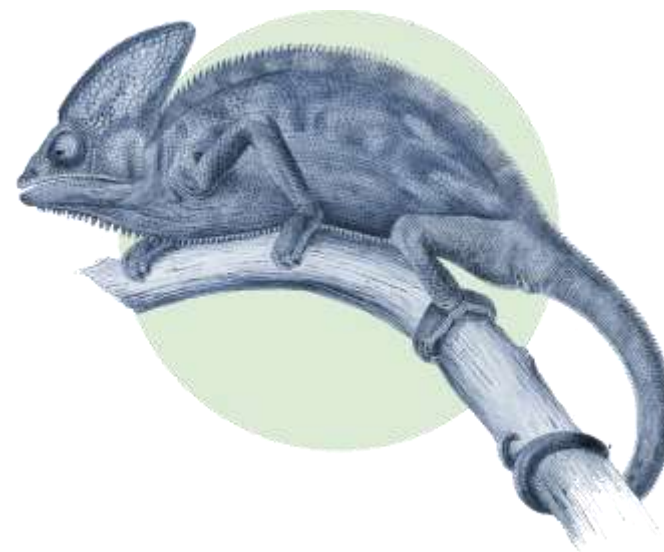


## Course design and instruction

The success of this course depends heavily on the support provided to participants from GBIF's network of capacity enhancement mentors. Visit the GBIF page on [capacity enhancement mentoring](#) to read more about these individuals and their contributions.

The following individuals are recognized for their significant contributions to the course design, materials and instruction:

- Nestor Beltran\*
- David Bloom
- Katia Cezón\*
- Dag Endresen
- Alberto González-Talaván\*
- Sharon Grant\*
- Marie-Elise Lecoq
- Sophie Pamerlon\*
- Nicolas Noé\*
- Mélianie Raymond\*
- Laura Anne Russell\*
- John Wieczorek
- Paula Zermoglio



# REACHING OUT



**Email Us**

*asia\_support@gbif.org*

<https://bit.ly/asia-office-hours>

Every Thursday at 3-4 pm  
(UTC+8)



# Thank you!

Melissa Liu | GBIF Asia Support Team

*[mliu@gbif.org](mailto:mliu@gbif.org)*





# Menti for the attendants

(from 16:15 to 16:30 CEST)



Go to: [www.menti.com](https://www.menti.com)

**Enter:6394 9460**

# CONCLUDING WORDS & NEXT STEPS

*By Harri Hautala, Science Adviser at AKA & Rainer Sodtke, Biodiversa+ Co-Chair*

## NEXT STEPS

⇒ For BiodivProtect projects: Please send your updated DMP after the workshop (**deadline: 31st of July 2024**) to Harri Hautala ([harri.Hautala@aka.fi](mailto:harri.Hautala@aka.fi))

**Format:** no formal template or structure (length, details,... depends on project and data types) – between 4 to 12 pages.

**Assessment:** A short feedback will be sent at the coordinators on their (updated) DMPs.

⇒ A virtual GBIF training workshop on data mobilization under planning for Q2 2025: The projects can attend the course voluntarily. More information coming during autumn!

=> Tomorrow we will continue with Capacity Building workshop on Darwin Core Standard at 9:00 CEST, welcome!



# Resources

- **Guidance document on data management, open data and the Productions of DMP developed by Biodiversa+ and the Belmont Forum (2023 update)**

⇒ Contains the main concepts in Data Management and Open Data, some recommendations on how to write a DMP, a list of practical tools and a selection of publications and documents on the topic.

⇒ The Annex II includes “Links to (and information on) national and funders open data/open access policies”.

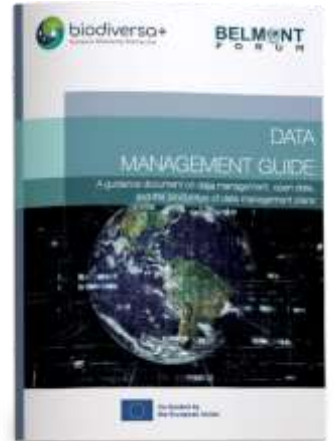
[https://www.biodiversa.eu/wp-content/uploads/2023/05/Biodiversa-Data-Management\\_WEB\\_2023.pdf](https://www.biodiversa.eu/wp-content/uploads/2023/05/Biodiversa-Data-Management_WEB_2023.pdf)

- **Practical Guide to the international alignment of research data management – Science Europe - Extended Edition (January 2021):**

[https://scienceeurope.org/media/4brkxxe5/se\\_rdm\\_practical\\_guide\\_extended\\_final.pdf](https://scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf)

- **Practical Guide to Sustainable Research Data – Science Europe (June 2021):**

<https://scienceeurope.org/media/b3odxx3s/se-practical-guide-sustainable-research-data.pdf>





**biodiversa+**  
European Biodiversity Partnership

EUROPEAN PARTNERSHIP



Co-funded by  
the European Union

**Thank you!**



[www.biodiversa.eu](http://www.biodiversa.eu)



[contact@biodiversa.eu](mailto:contact@biodiversa.eu)



BiodiversaPlus

