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 2nd 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)

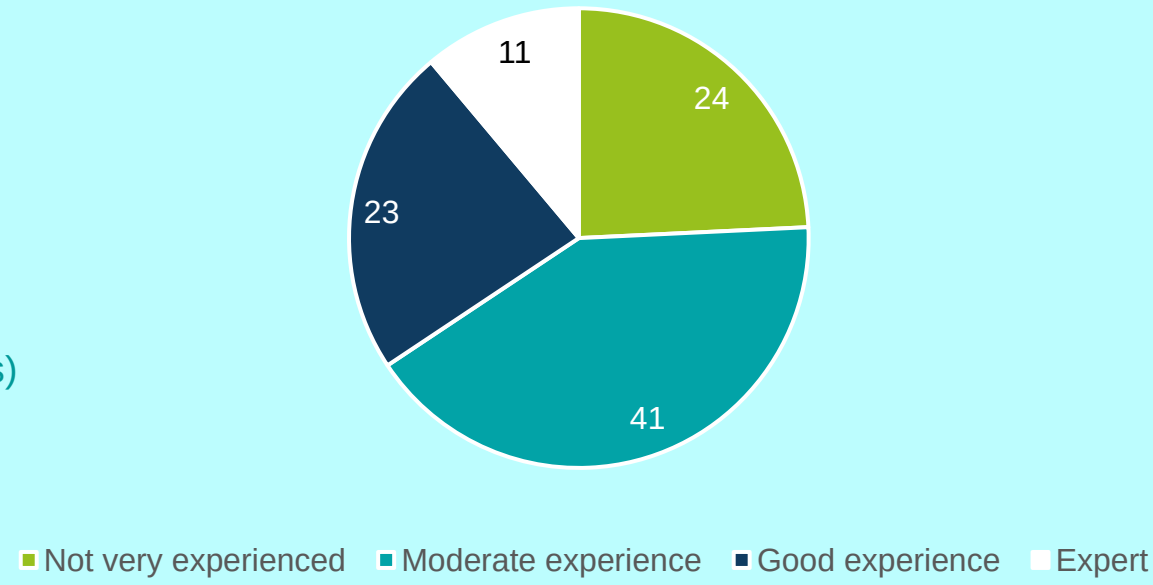
We welcome

YOU

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

June 6, 2024
12:45 – 16:45 CEST

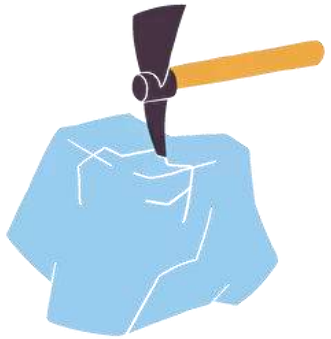
Current knowledge level on data management



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

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: developing knowledge and strAtegies to inform ConservaTION priorities and measures

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

Funaction
Aquatic
fungal
biodiversity

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

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

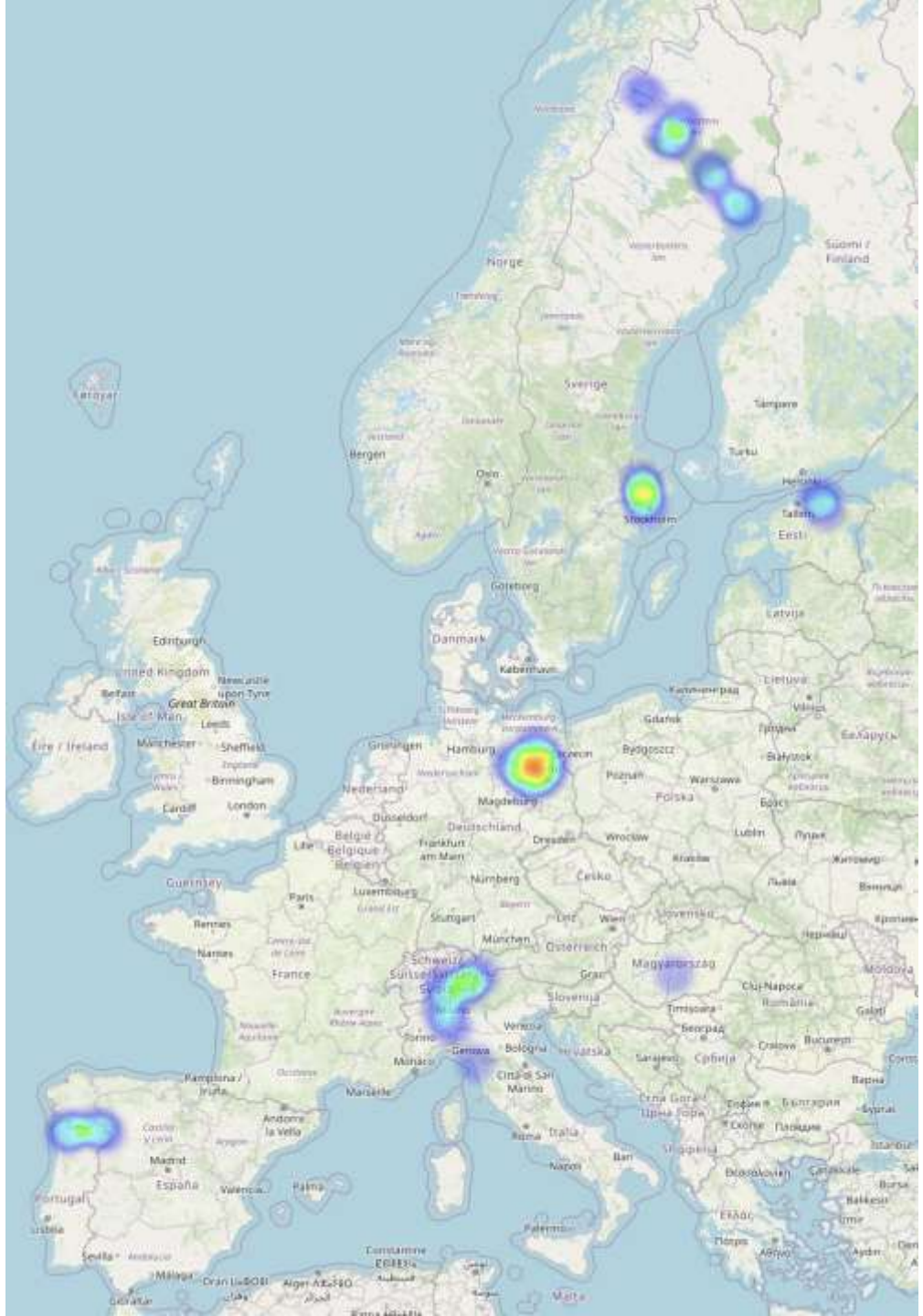
WP5
Effective engagement, communication, and dissemination of information

Data

Samples:
- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

Primary Scientific Data:
- DNA sequences- metabarcode & metagenome
- Water chemistry
- Site metadata

Secondary Scientific Data:
- Maps
- Models
- Other derivative outputs



Data

Samples:
- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

Primary Scientific Data:
- DNA sequences-metabarcode & 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

Data

Samples:
- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

Primary Scientific Data:
- DNA sequences-metabarcode & 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

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

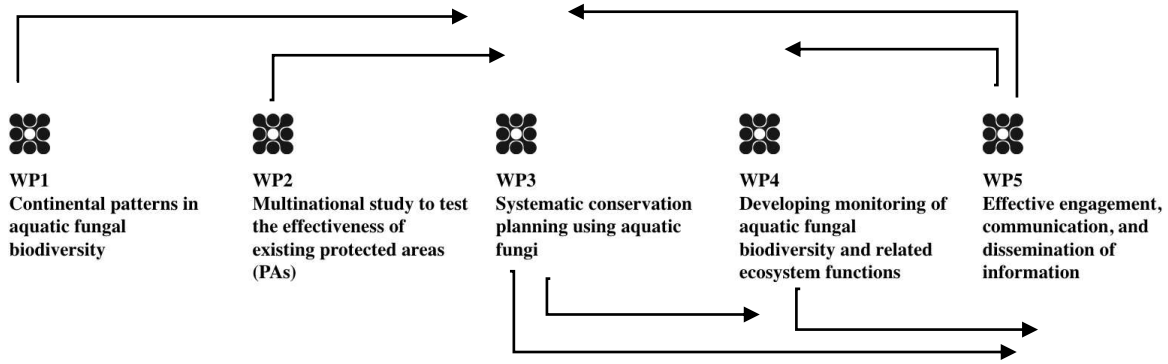
Data Management Plans: Why scientists should create & sustain them

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)

Data Management Plans: Why scientists should create & sustain them

Avoidance of chaos

Softulka | Shutterstock

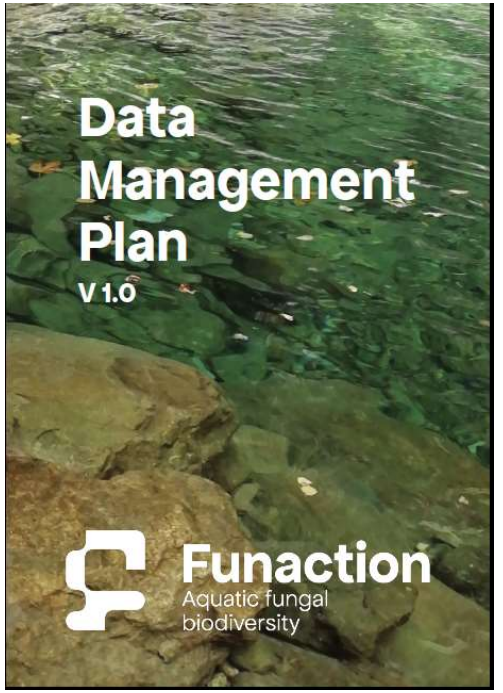
Data Management Plans: Why scientists should create & sustain them

- 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)

Softulka | Shutterstock

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

Data Management Plan: FUNACTION

Samples:
- Water-microbes
- Water-chemistry
- Sediment
- Leaf litter

Primary Scientific Data:
- DNA sequences-metabarcode & 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

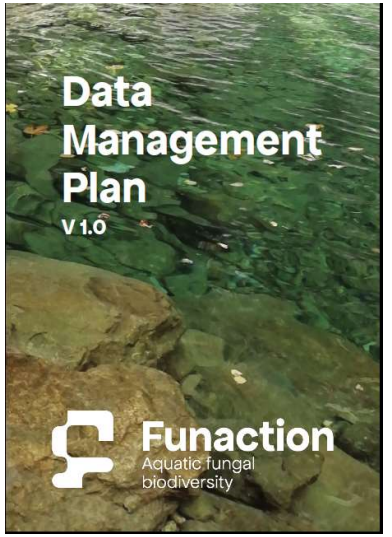
Outputs tracking

Scripts

Readme

Analysis documentation

Data Management Plan: FUNACTION



Sample naming conventions

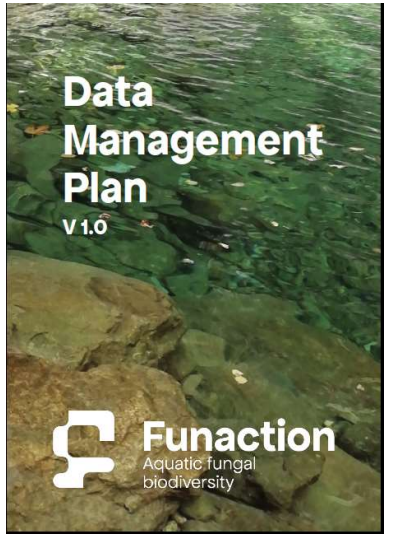
= persistent unique identifiers

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

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

Data Management Plan: FUNACTION

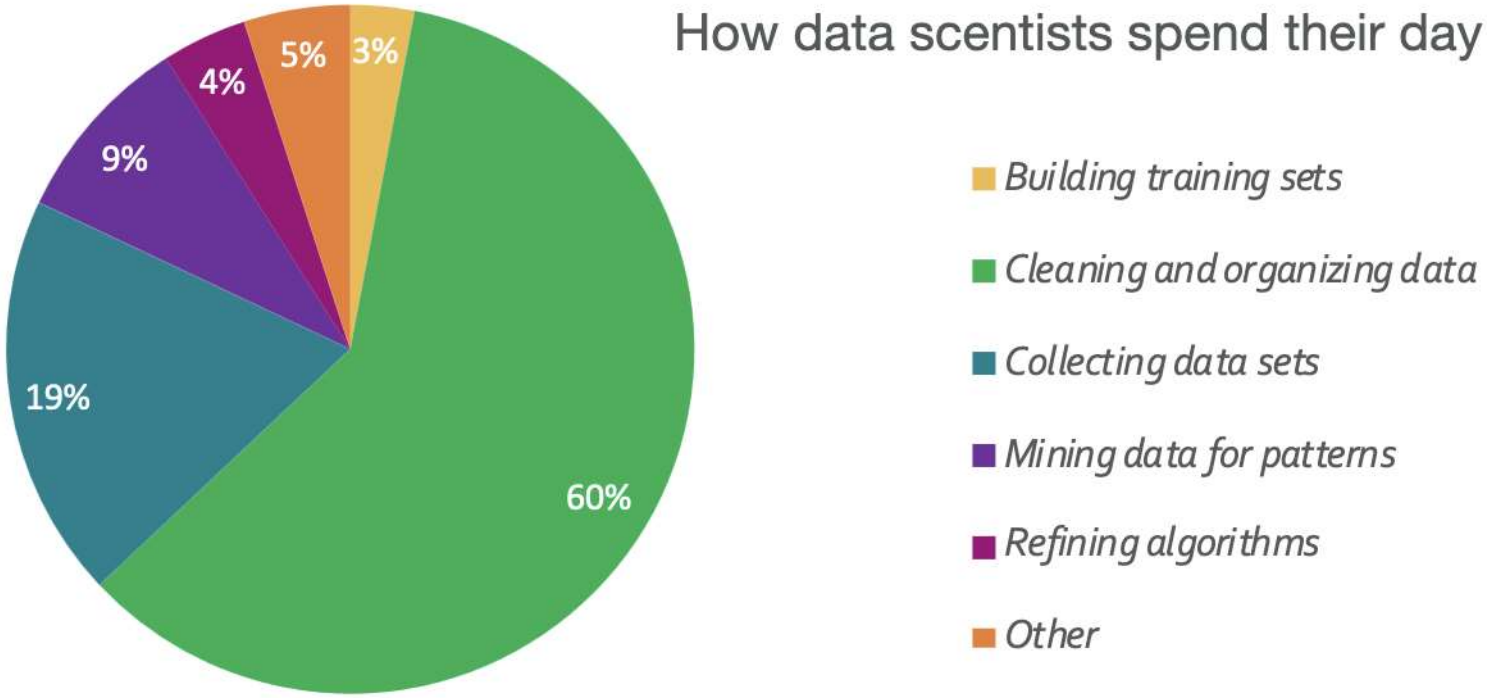


Figure 2: How data scentists spend their day (redrawn after CrowdFlower, 2016)

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

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

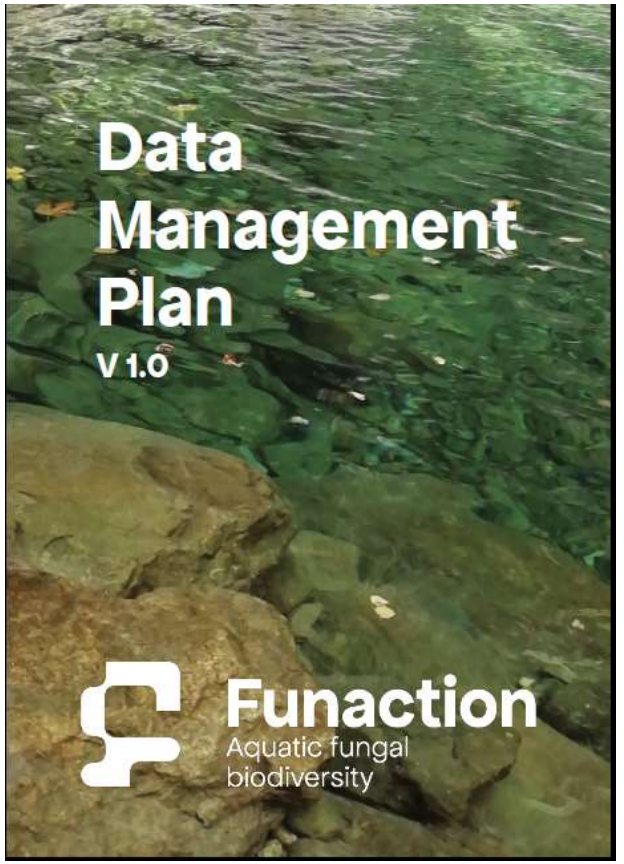
= Availability

Kobotoolbox

- 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

Data Management Plan: FUNACTION

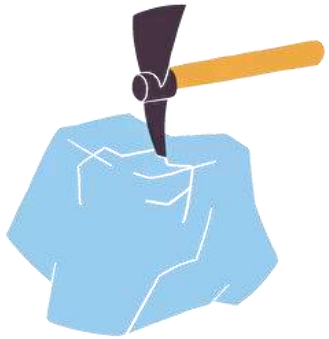
Switzerland	Daniel Romero Mujalli (Andreas Bruder)
Germany	Alice Retter
Estonia	Veljo Kisand
Italy	Emanuele Ferrari
Portugal	Isabel Fernandes
USA	Catia Canteiro




SLU
SCIENCE AND EDUCATION FOR SUSTAINABLE LIFE

Menti for the attendants

(from 14:00 to 14:15 CEST)



Go to: 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

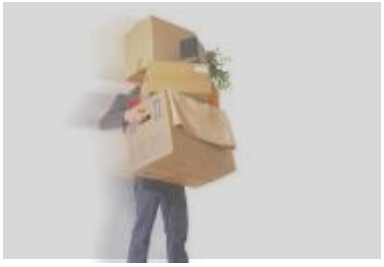
Maybe it will be useful for me later, better keep it.

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

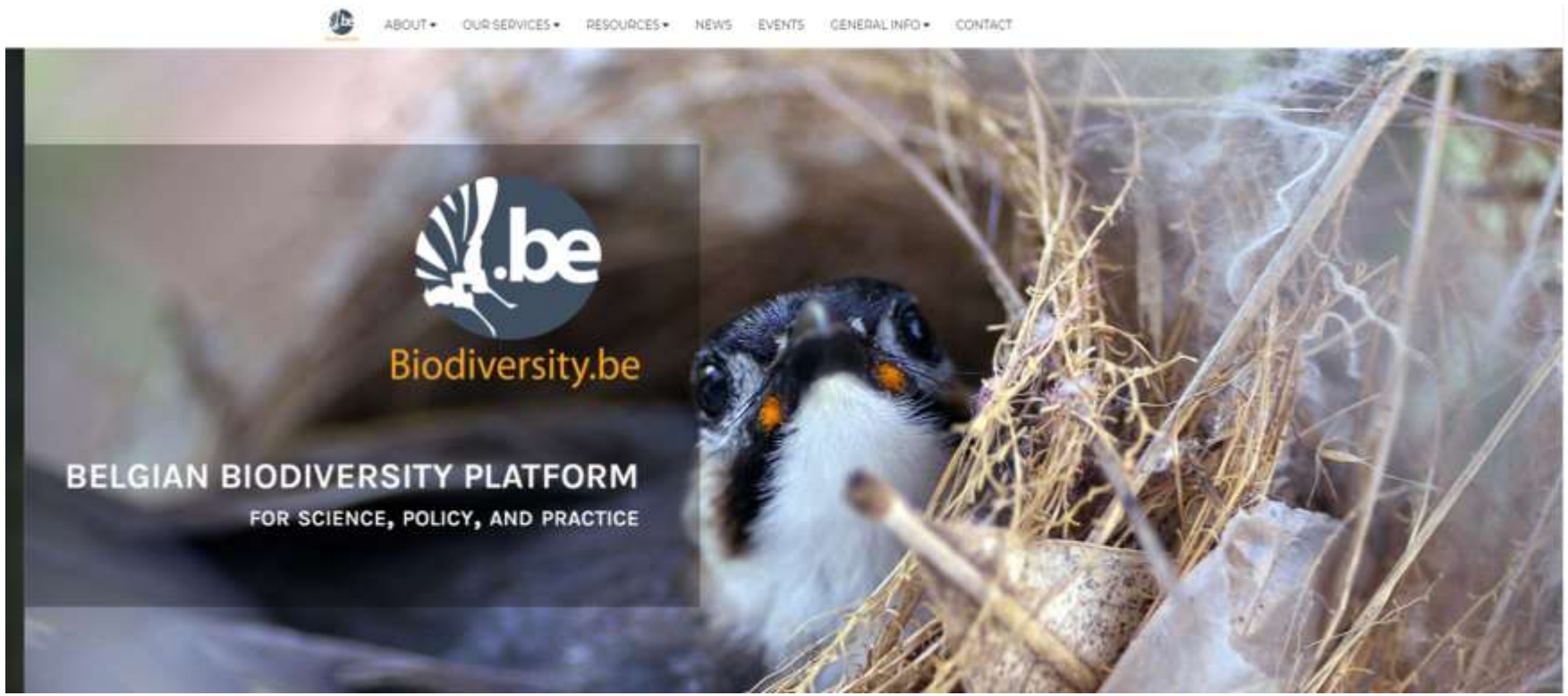
Open Science and Open Data

main principles and concepts

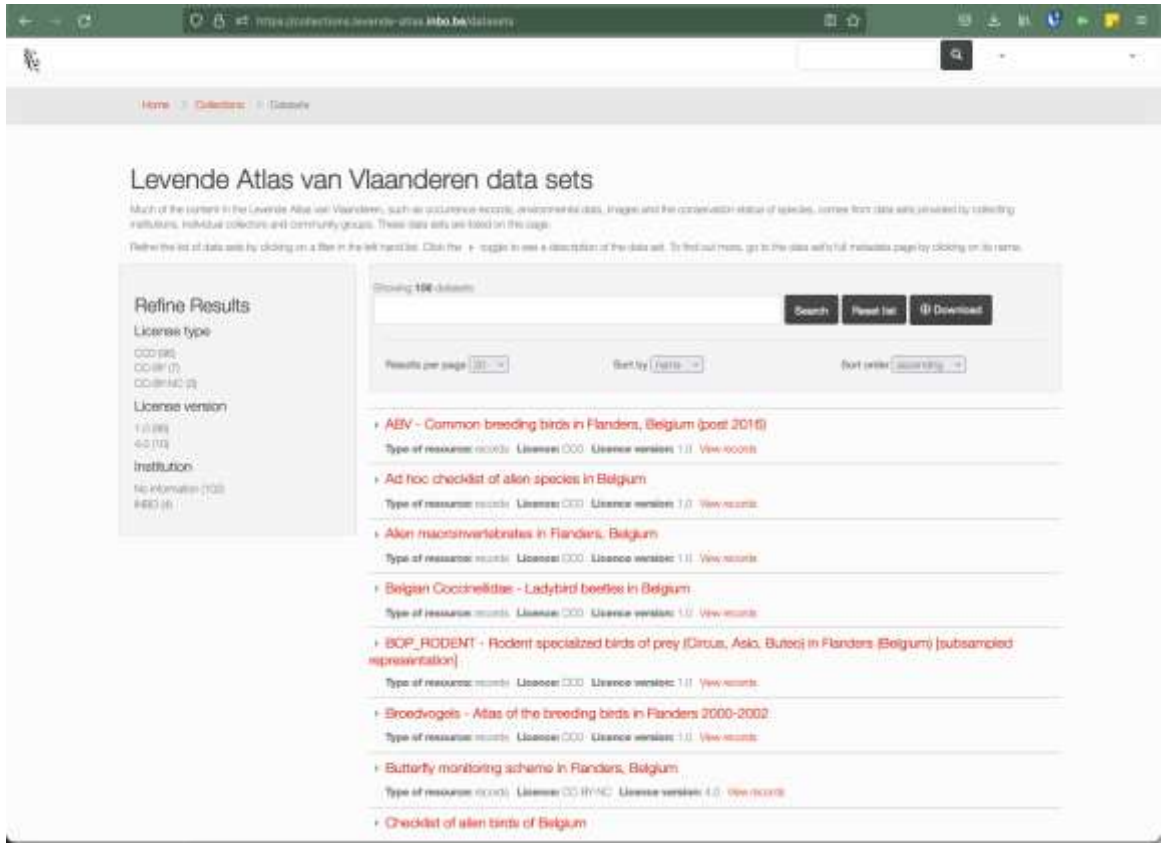
Dimitri Brosens
Coördinator Flemish Biodiversity Portal
GBIF Belgium
0000-0002-0846-9116
Dimitri.brosens@inbo.be
2024/06/06

Open science lab for biodiversity

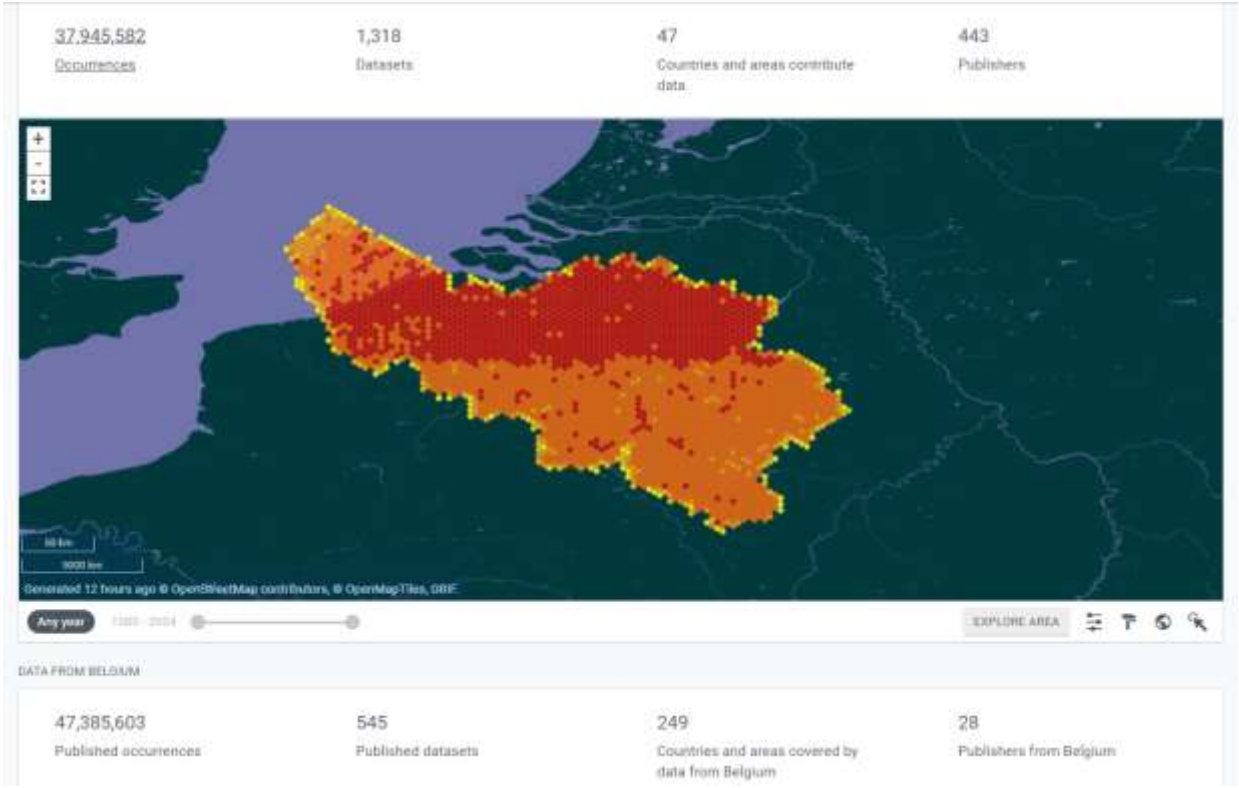
We make biodiversity research open and reproducible



Flemish Living Atlas project:
Vlaams Biodiversiteitsportaal



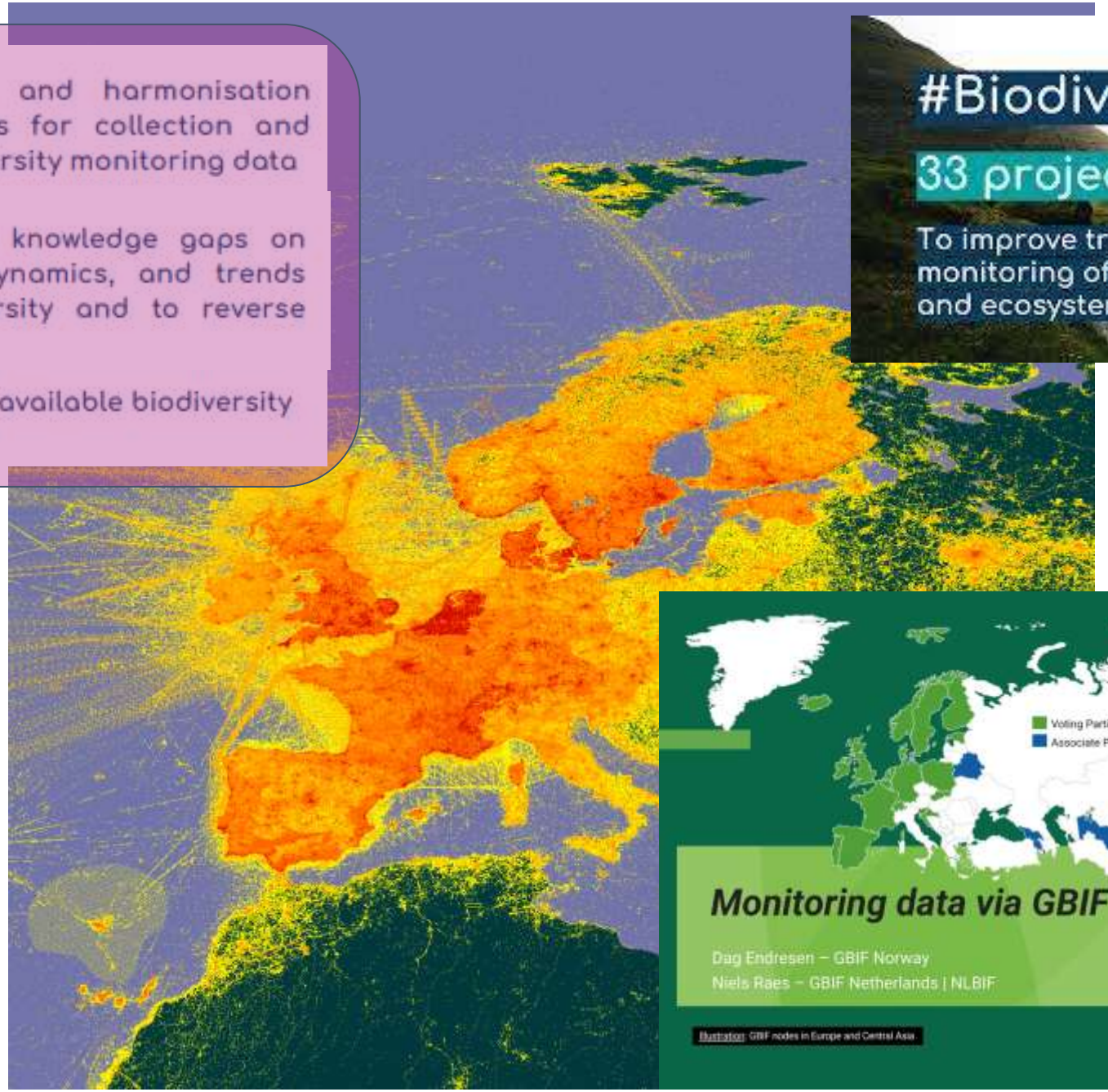
Belgian Biodiversity Platform:
Belgian GBIF Node



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




OPEN
SOFTWARE
OPEN
SOURCE
FREE
SOURCE
COPYRIGHT
PUBLIC
ACCESS
COST
TECHNOLOGY
RESEARCH
CODE
CULTURE
DEVELOPMENT
COMMUNITY
DIGITAL
INTERNET
AVAILABLE
INFORMATION
USERS
CULTURAL
CONTENT
SYSTEM
MEDIA

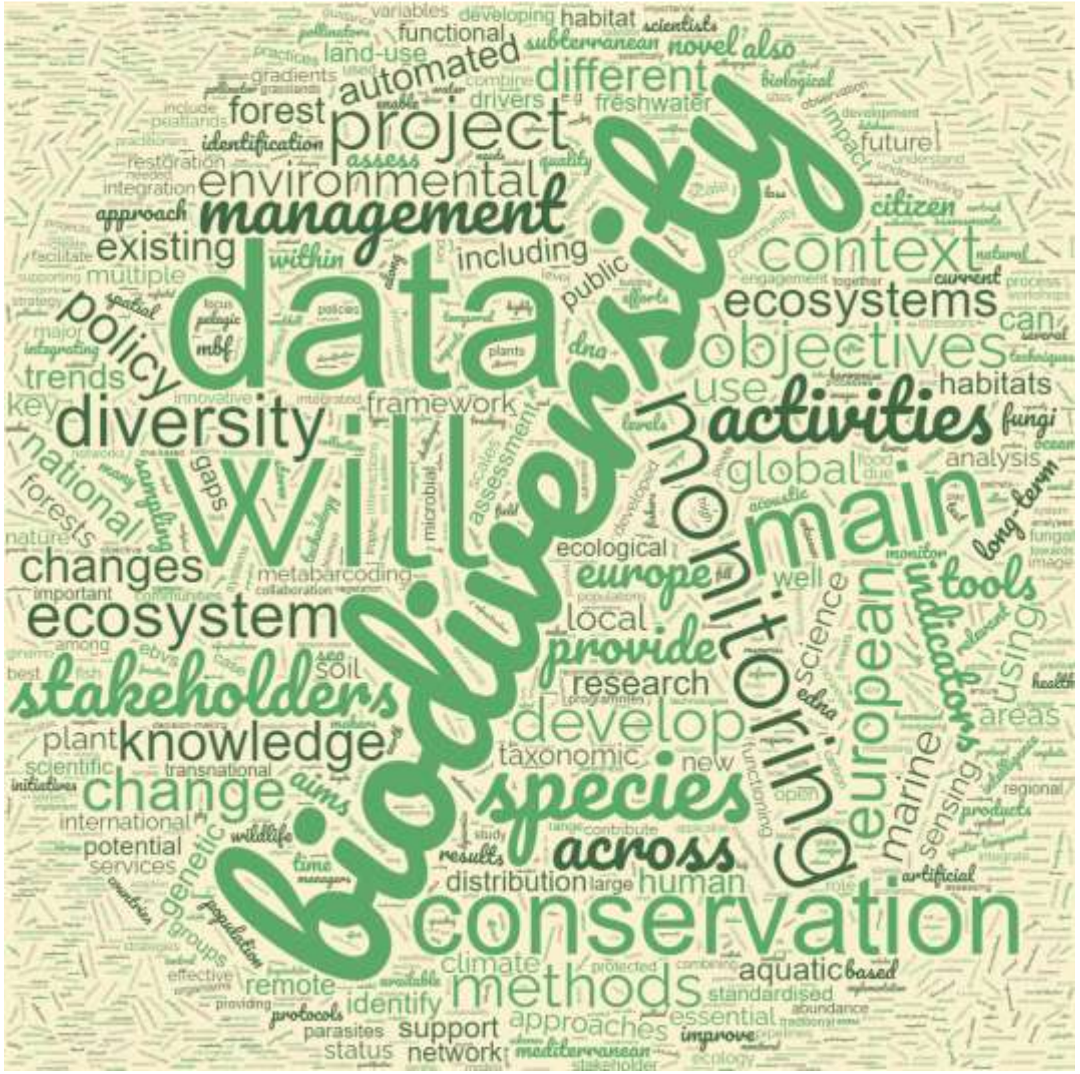
DOI
FAIR
Openess
DataCite
Open
data
Knowledge
Preservation
Repository
Archiving
Publishing
Open
research
data
Researchers
Data management
plans
DataDOI
Open Access
FAIRest
DOI
Open
science
Storage
Reuse
Research data
OpenAIRE

open
science
people
new
way
knowledge
tools
scientific
make
community
better
open-source
world
discovery
just
research
ideas
access
sharing
progress
truth
forward
everyone
accessible
problems
together
approach
first
even
find
one
now
year
code
time
brain
allows
thousands
making
working
publish
movement
environment
formal
source
build
solve
discoveries
take

Open
Access
journals
journal
agreements
publishers
publishing
self-archiving
articles
article
authors
publication
copyright
available
rights
various
based
freely
DiVA
PLoS
subscription
alternative
used
areas

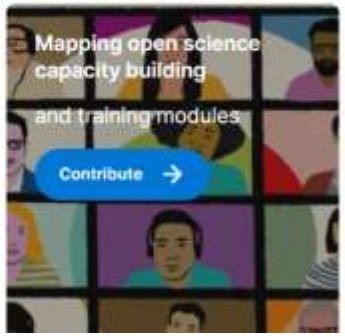
open
data
students
teaching
species
sharing
management
software
education
tools
need
gbif
writing
practices
people
using
learning
integrated
workshops
universities
supervision


research
will
data
biodiversity
monitoring
call
project
university
institute
projects
conservation
management
science
sciences
knowledge
marine
diversity
change
european
stakeholders
across
national
department
germany
provide
ecosystem
ecosystems
species
context
environmental
information
biology
main
europe
grant
partners
funded
figure
centre
activities
develop
changes
belgium
status
global
italy
use
existing
technology
forest
committee
netherlands
policy
ecology
tools
proposals
human
different
trends
france
resources
innovation
sweden
indicators
evaluation
portugal
distribution
objectives
total
framework
countries
republic
nature
methods
finland
key
local
funding
support
habitats
plant
spain
norway
communities
community
theme
network
applied
poland
functional
wildlife
areas
also
sensing
impact
taxonomic
fully
natural
climate
citizen
aims
development
identification
population
mediterranean


biodiversity
data
will
conservation
monitoring
main
species
activities
management
project
diversity
ecosystem
stakeholders
knowledge
methods
develop
european
objectives
context
ecosystems
environmental
change
across
policy
different
automated
indicators
tools
provide
europe
changes
existing
national
global
marine
forest
forests
trends
habitats
research
local
use
using
including
public
science
areas
framework
ecological
metabarcoding
taxonomic
plant
distribution
human
climate
remote
identify
support
network
status
approaches
essential
aquatic
standardised
potential
services
international
transnational
scientific
genetic
groups
effective
parasites
functional
land-use
variables
habitat
subterranean
novel
also
freshwater
drivers
gradients
restoration
integration
approach
identification
assess
citizen
future
impact
multiple
spatial
key
nature
important
among
soil
new
open
functioning
sensing
fungi
fungal
analysis
long-term
regional
artificial
products
mediterranean
improve
protocols
population
wildlife
time
aims
results
collaboration
abundance
based
ecology
stakeholder
scientists
developing
natural
current
can
engagement
dna
edna
acoustic
sampling
gaps
microbial
developed
assessment
integrated
innovative
major
trends
best
fish
well

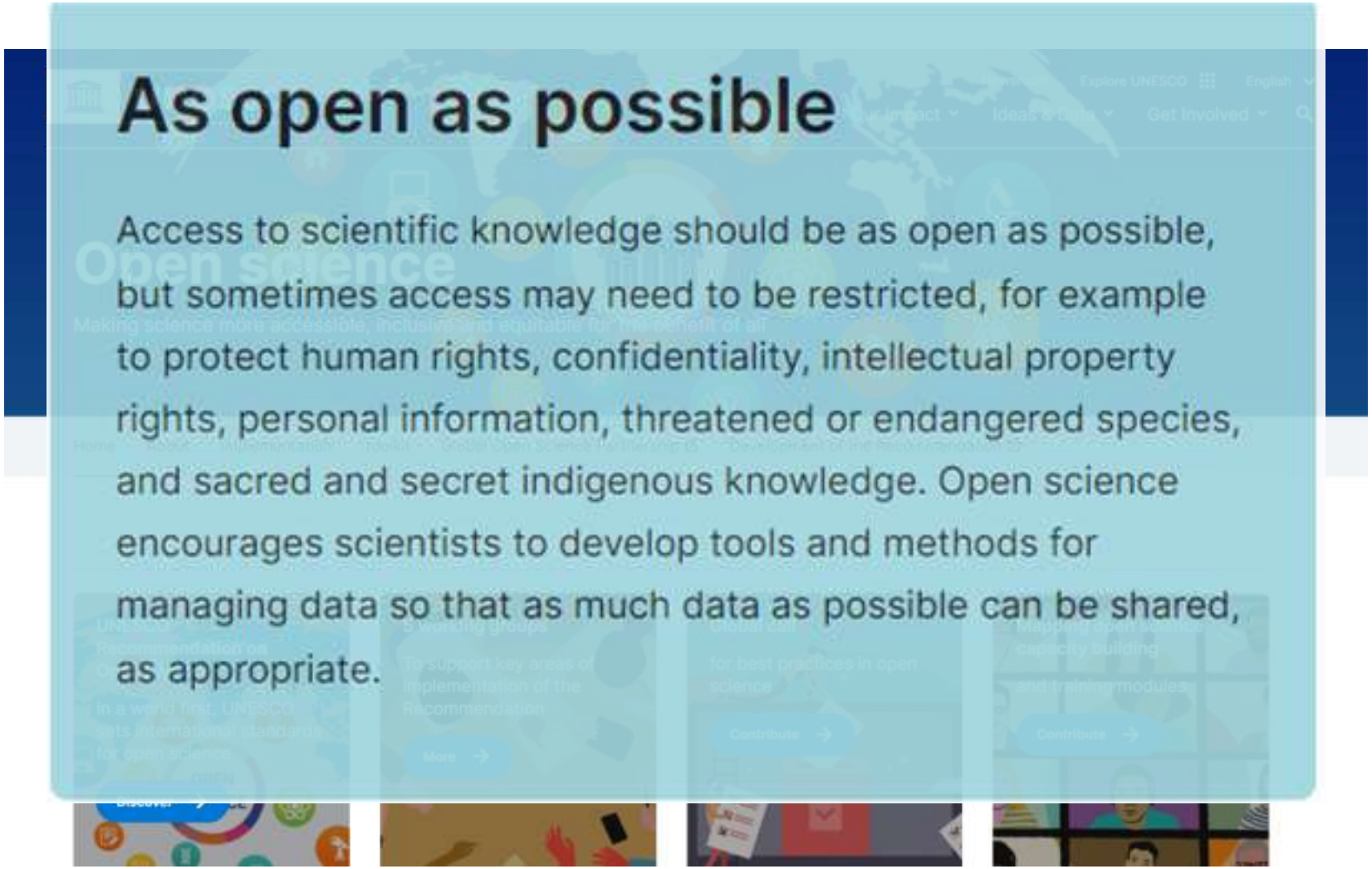
UNESCO Recommendation on Open Science

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



UNESCO Recommendation on Open Science

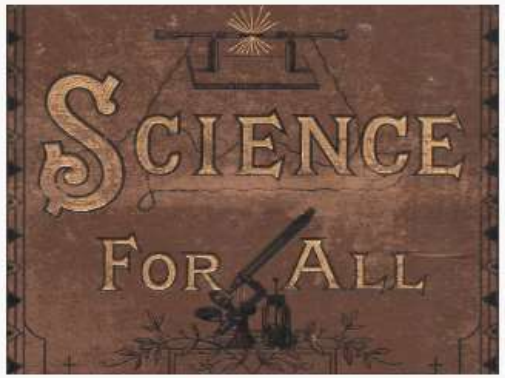
The UNESCO Recommendation on Open Science is the first international standard setting instrument on open science.



> Science, technology and innovation policy
> Industry and globalisation
> Emerging technologies
> Digital economy
> Broadband and telecom
> 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

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.

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

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.

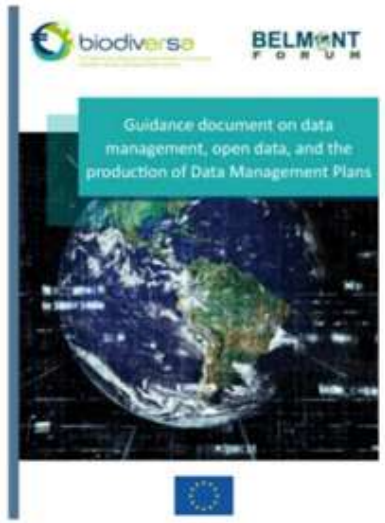
When researchers share knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.

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 EN ••• 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+
European Biodiversity Partnership

LINK: https://www.biodiversa.org/1677/download

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.

~~8~~9 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 objec
- New generation metrics
  - New indicators
- Alternative metrics and rewards
  - Research quality and impact
- Future of scholarly communication
  - Open access policy

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

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

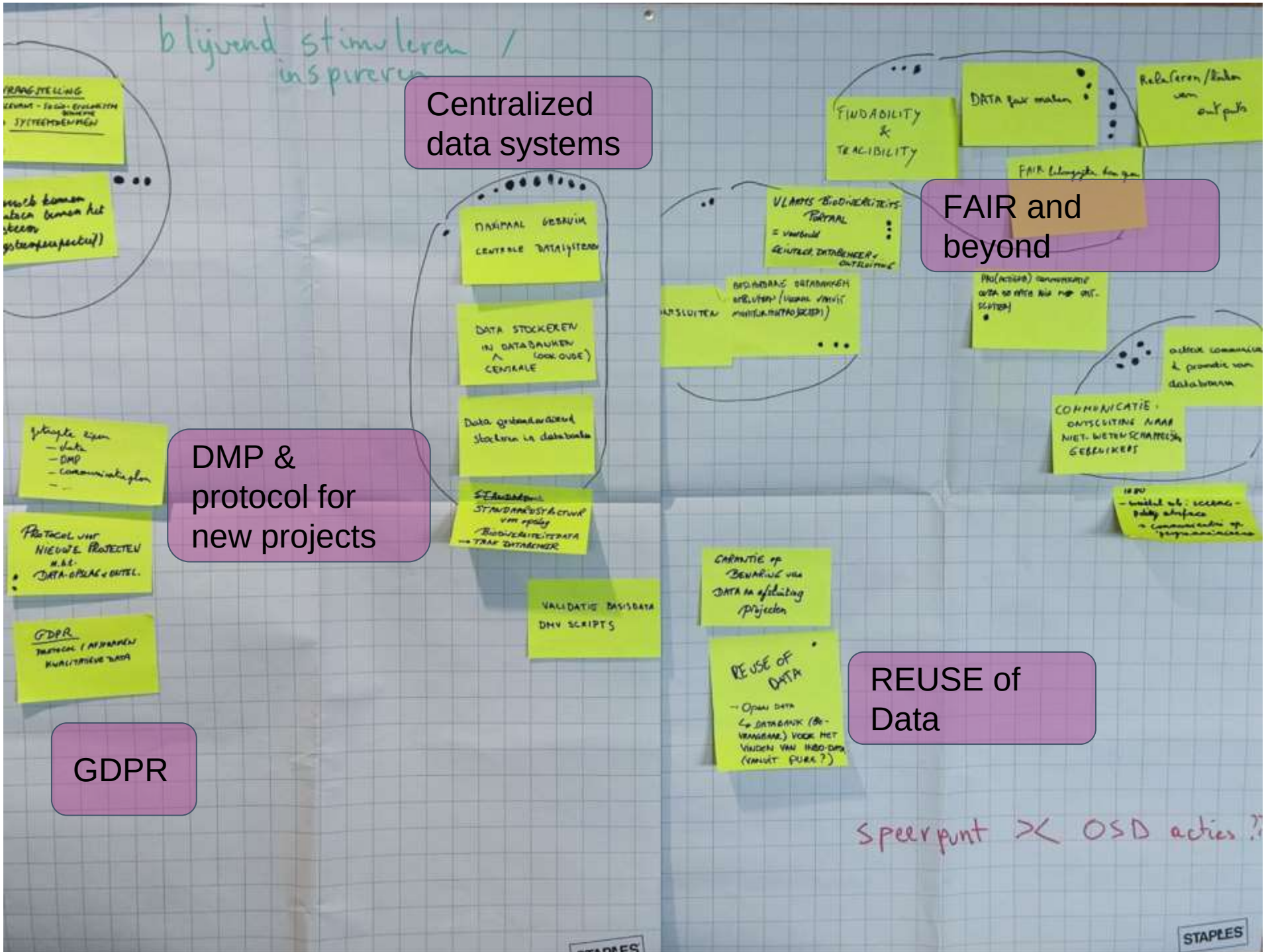
DOI:
DOI 10.5281/zenodo.7672934

Keyword(s):
open science

CERN Open Science

FRDN.

eosc
European Open Science Cloud


blijvend stimuleren / inspireren
Centralized data systems
DMP & protocol for new projects
GDPR
FAIR and beyond
REUSE of Data
FINDABILITY & TRACIBILITY
DATA FAIR maken
DATA STOCKEREN
Data gedecentraliseerd stockeren in databanken
VALIDATIE BASISDATA
DMV SCRIPTS
GDPR
REUSE OF DATA
COMMUNICATIE
Speerpunt >< OSD acties ?
STAPLES

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

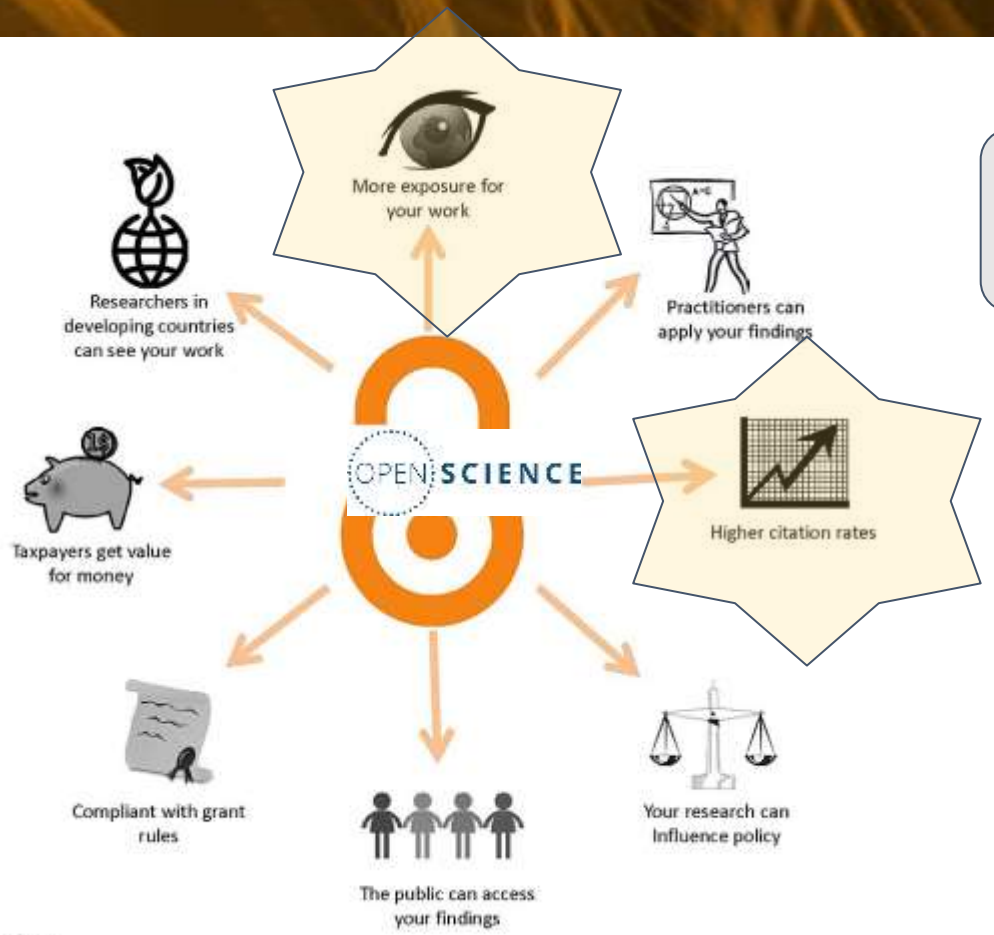
Practicing open science means that research **outputs are accessible to all** – not stuck behind pay walls. This helps to ensure that all researchers, and other stakeholders, have access to information regardless of their location or economic situation. It means that the **research process can be accelerated 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?

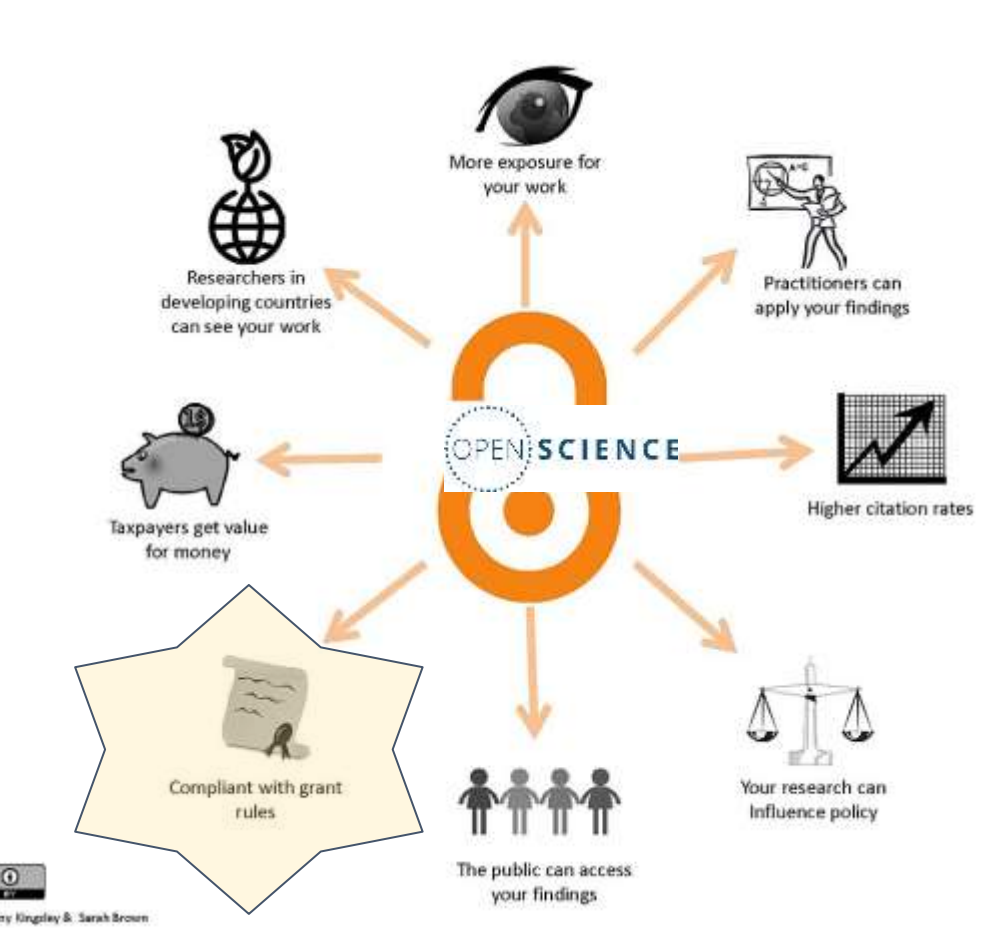


Good for You

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.

https://www.enago.com/academy/open-access-publications-get-cited-more-often/
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040157

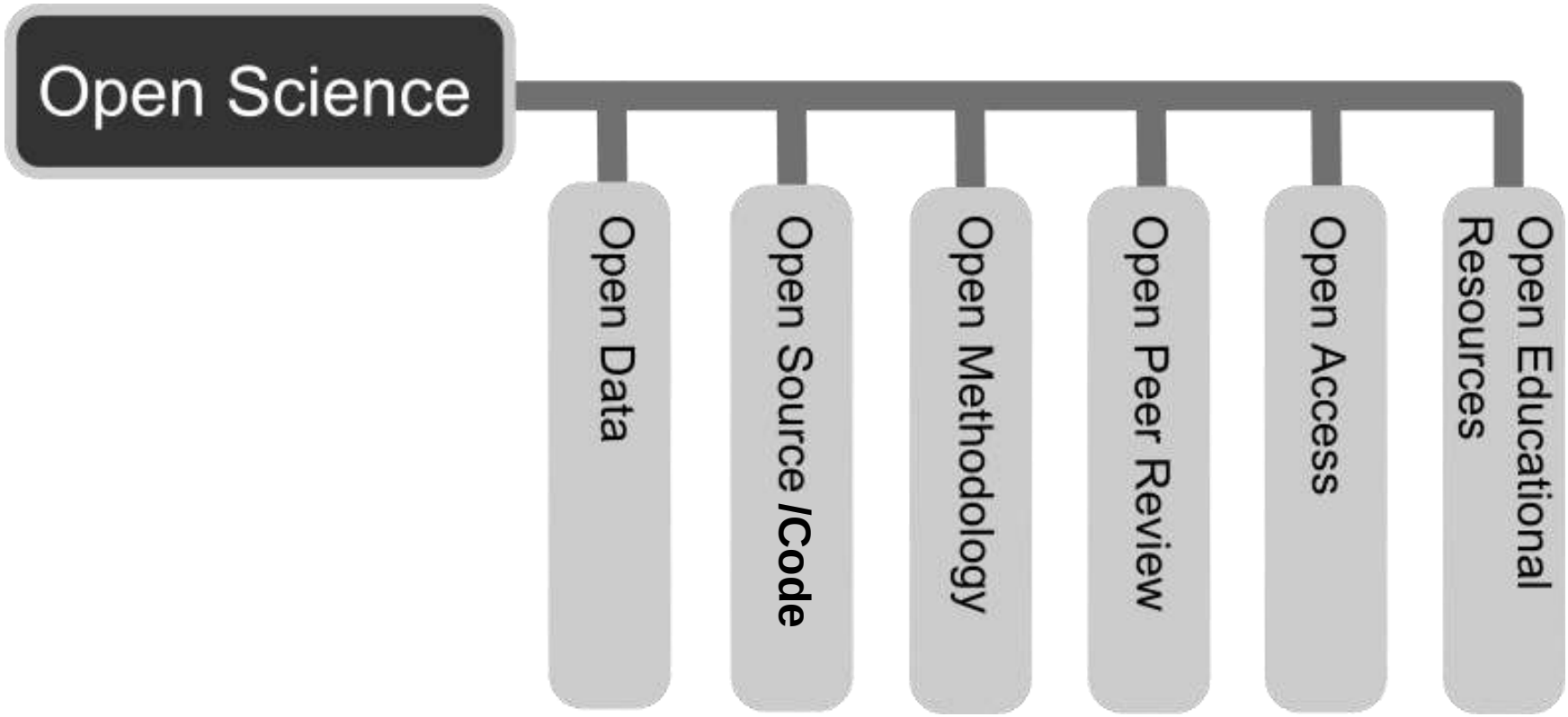
Why open science?



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**-bydesign" 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.

GBIF Global Biodiversity Information Facility

zenodo

DRYAD

figshare

Open Science - more than just open access to publications!

…Share your **data**

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

BOLD SYSTEMS

National Center for Biotechnology Information

MOVEBANK for animal tracking data

OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM

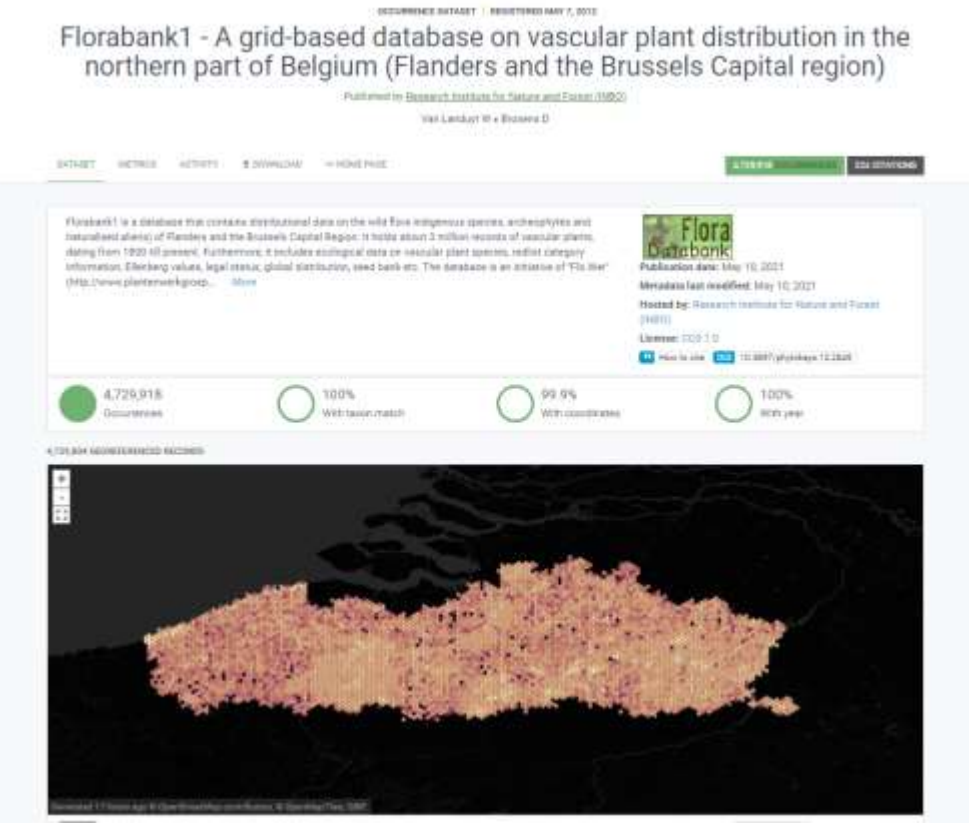
GEO GROUP ON EARTH OBSERVATIONS

BCO-DMO Biological & Chemical Oceanography Data Management Office

eLTER

Open Science - more than just open access to publications!

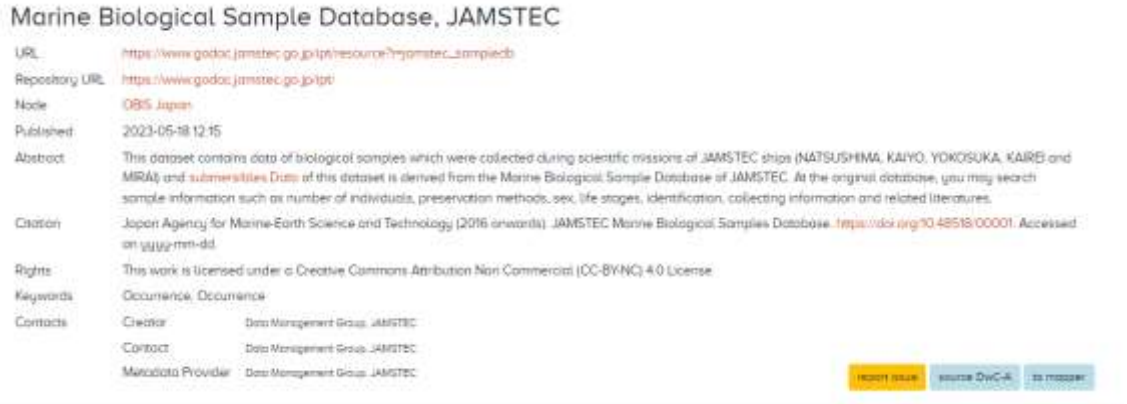
GBIF
Global Biodiversity Information Facility



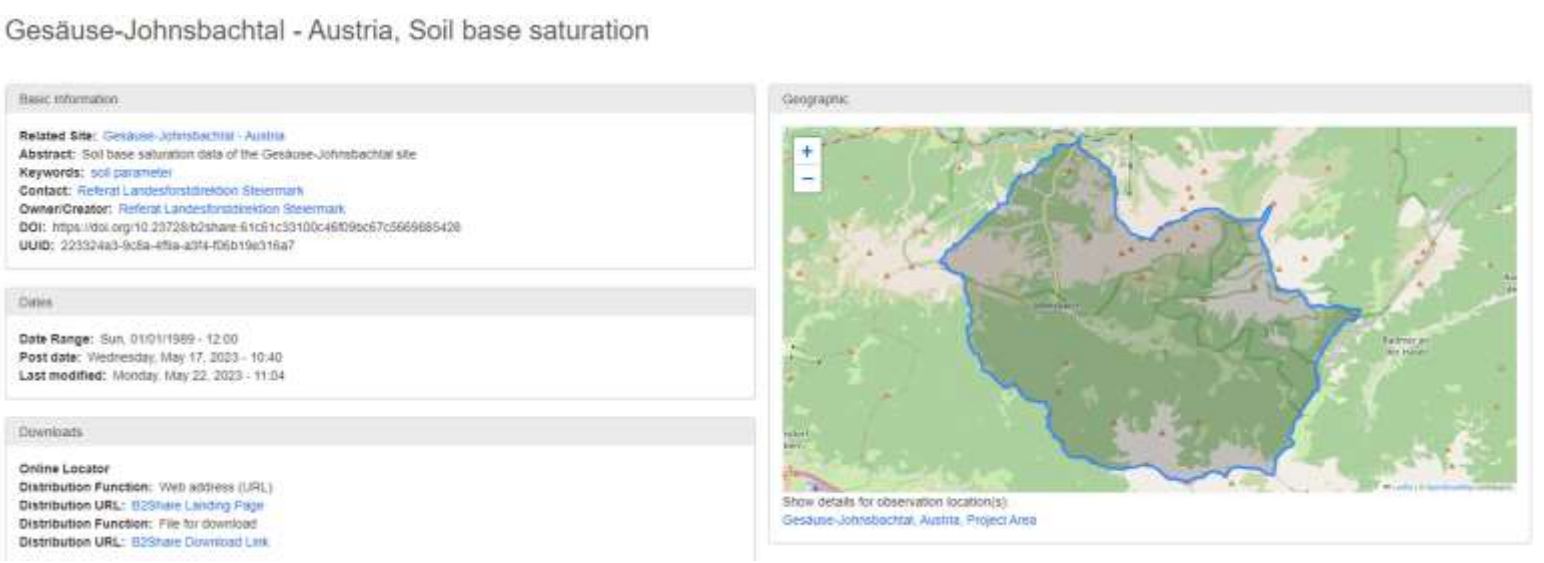
Open Science - more than just open access to publications!

OBIS
OCEAN BIODIVERSITY INFORMATION SYSTEM

MOVEBANK
for animal tracking data



Open Science - more than just open access to publications!



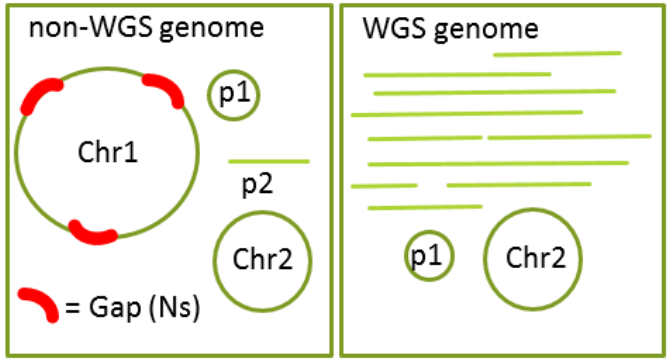
Open Science - more than just open access to publications!

BOLD SYSTEMS

National Center for Biotechnology Information
NCBI

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:



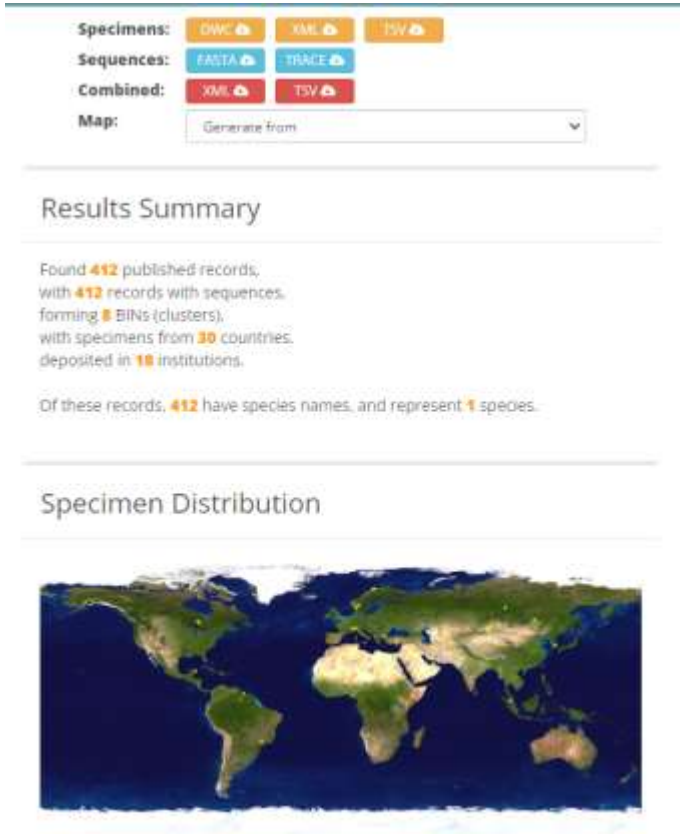
non-WGS

- 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

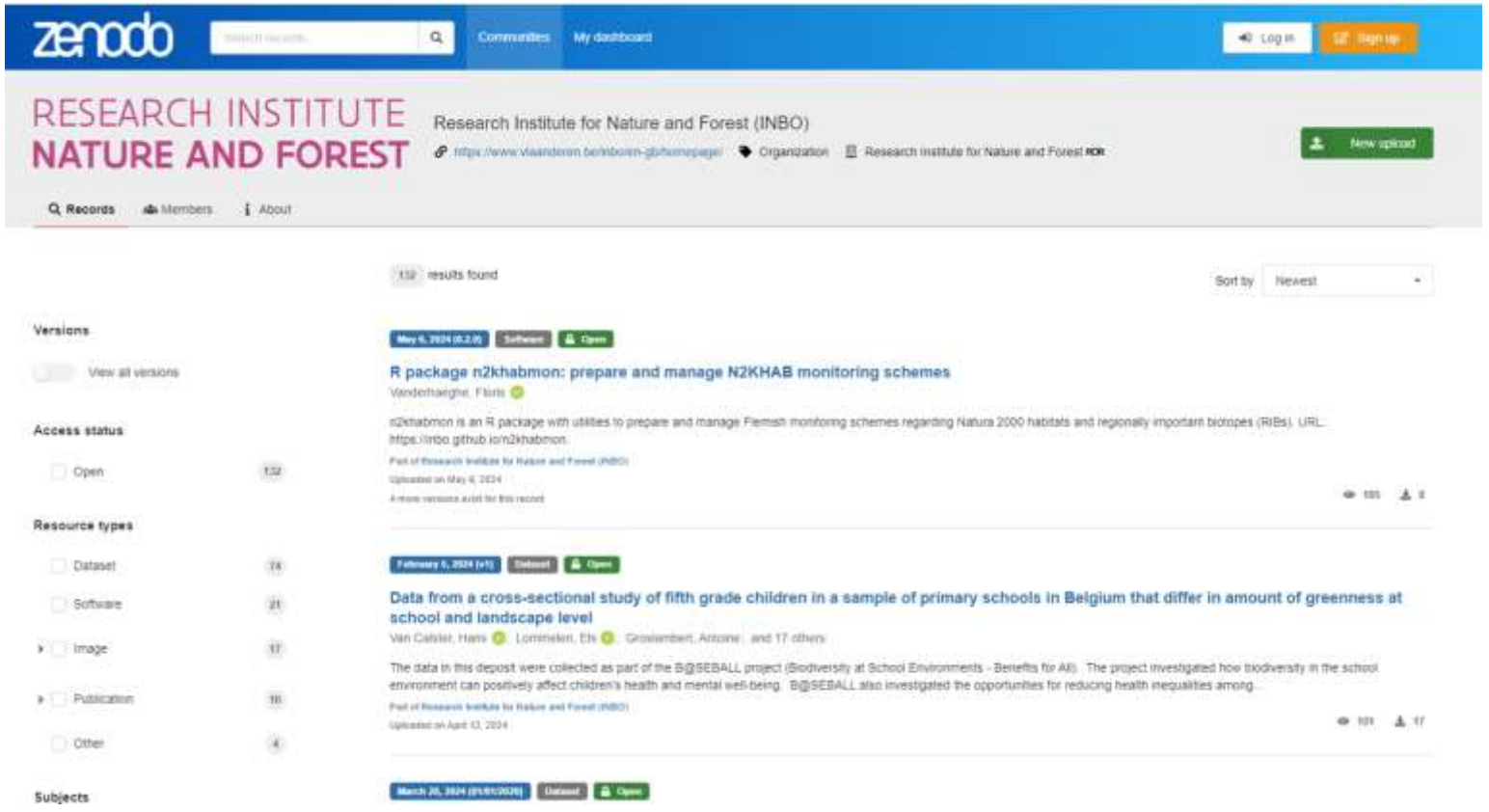
In both cases



Open Science - more than just open access to publications!

zenodo

Open Science - more than just open access to publications!



Open Science - more than just open access to publications!

DRYAD

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

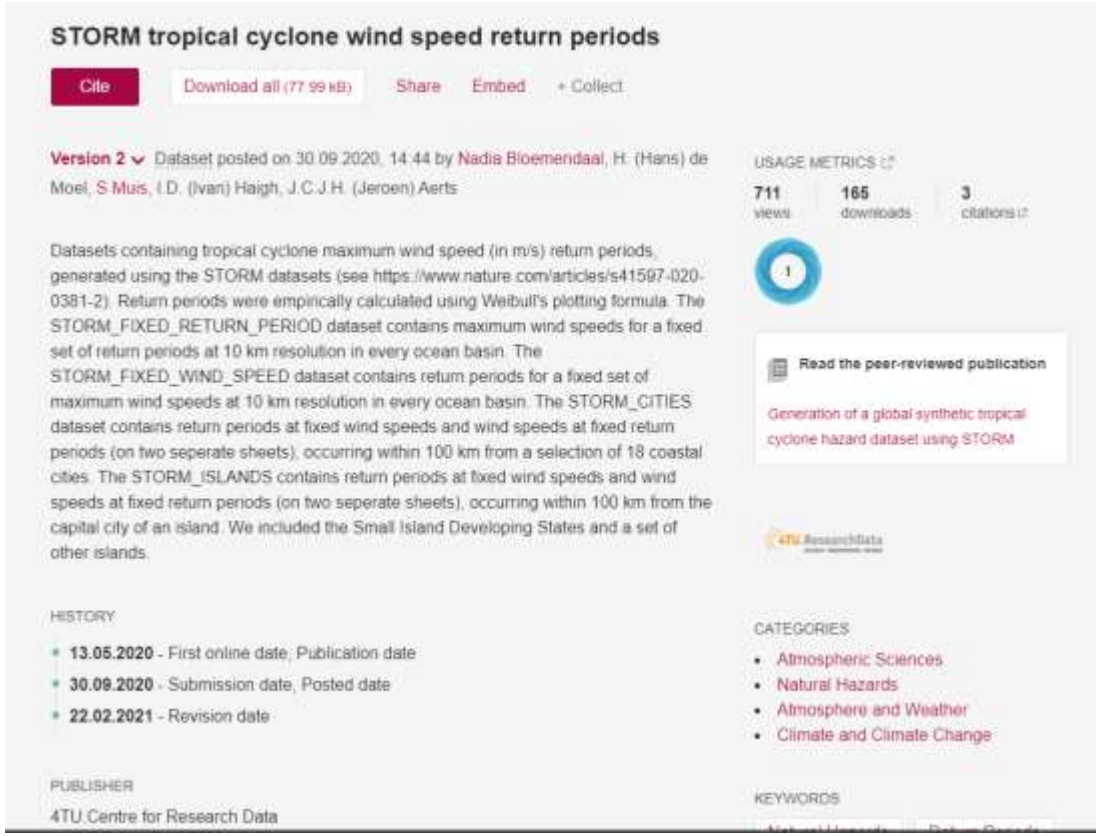
1 citations

Keywords

Acclimatization
Coenagrion armatum
Coenagrion caerulescens

Open Science - more than just open access to publications!

figshare



STORM tropical cyclone wind speed return periods

Cite | Download all (77.99 kB) | Share | Embed | + Collect

Version 2 Dataset posted on 30.09.2020, 14:44 by Nadia Bloemendaal, H. (Hans) de Moel, S Muis, 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 seperate sheets), occurring within 100 km from a selection of 18 coastal cities. The STORM_ISLANDS contains return periods at fixed wind speeds and wind speeds at fixed return periods (on two seperate 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

CATEGORIES

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

KEYWORDS

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

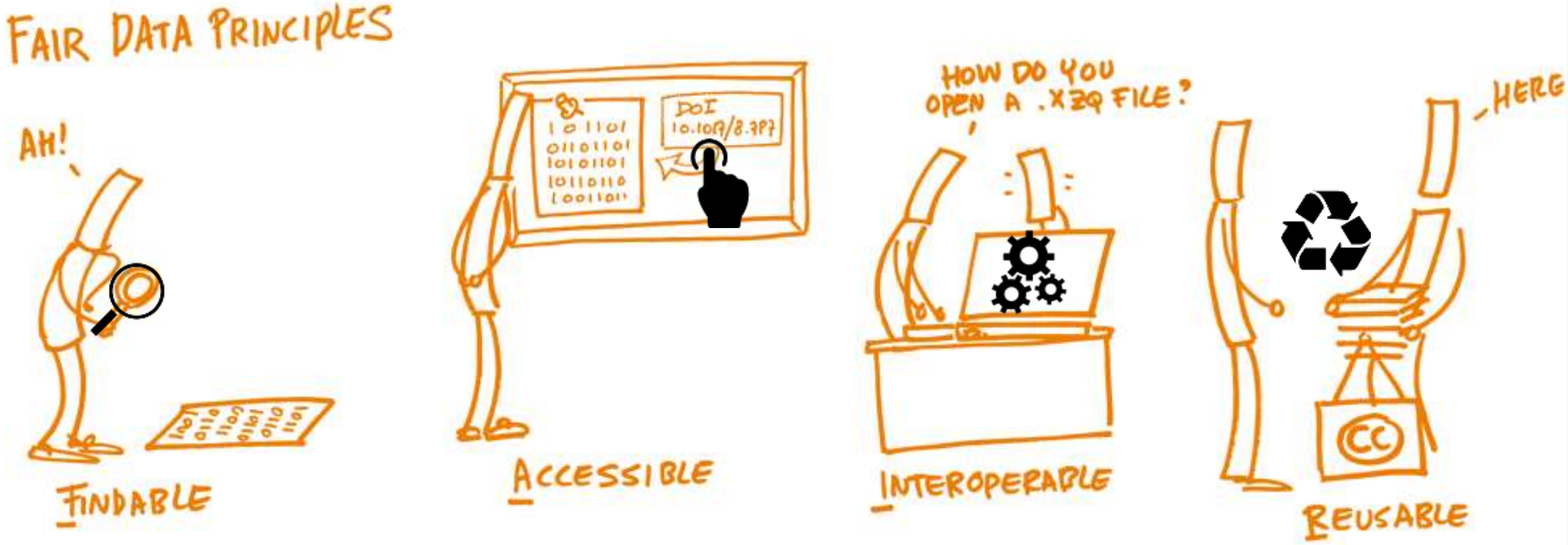
- Standards
- Vocabulary
- Methodology
- References

Reusable

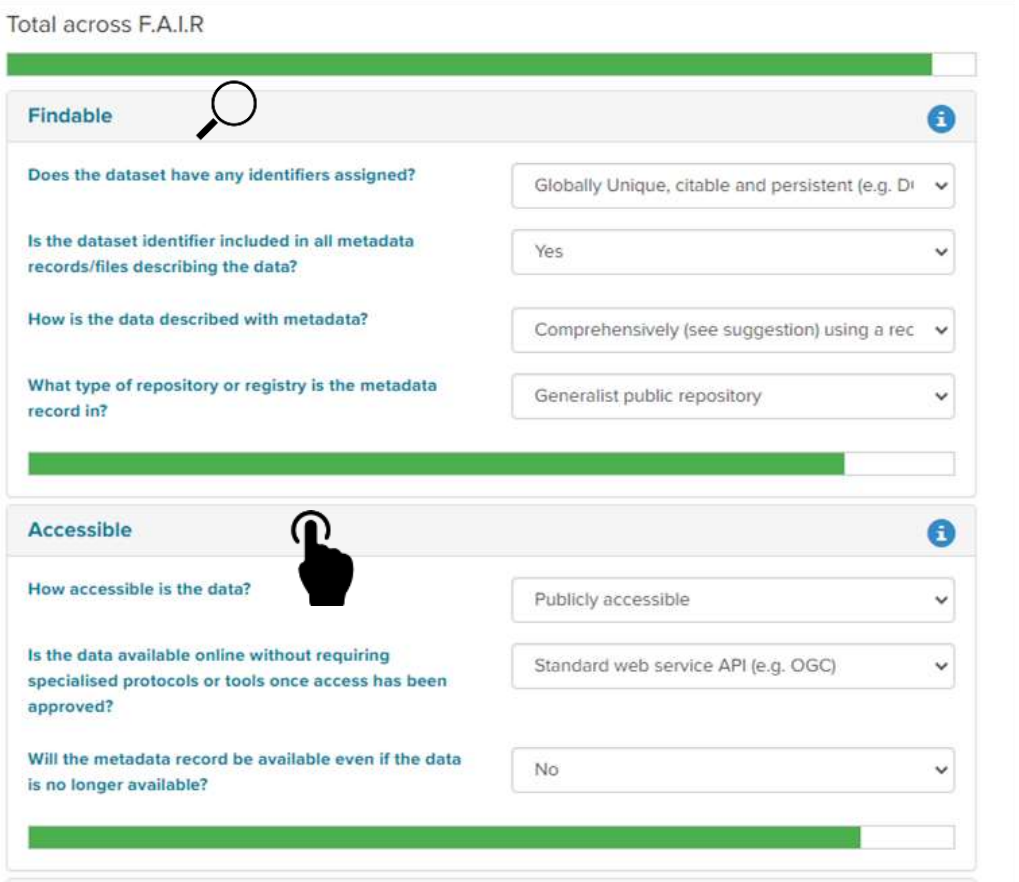
- Licensing
- Provenance
- Community standards

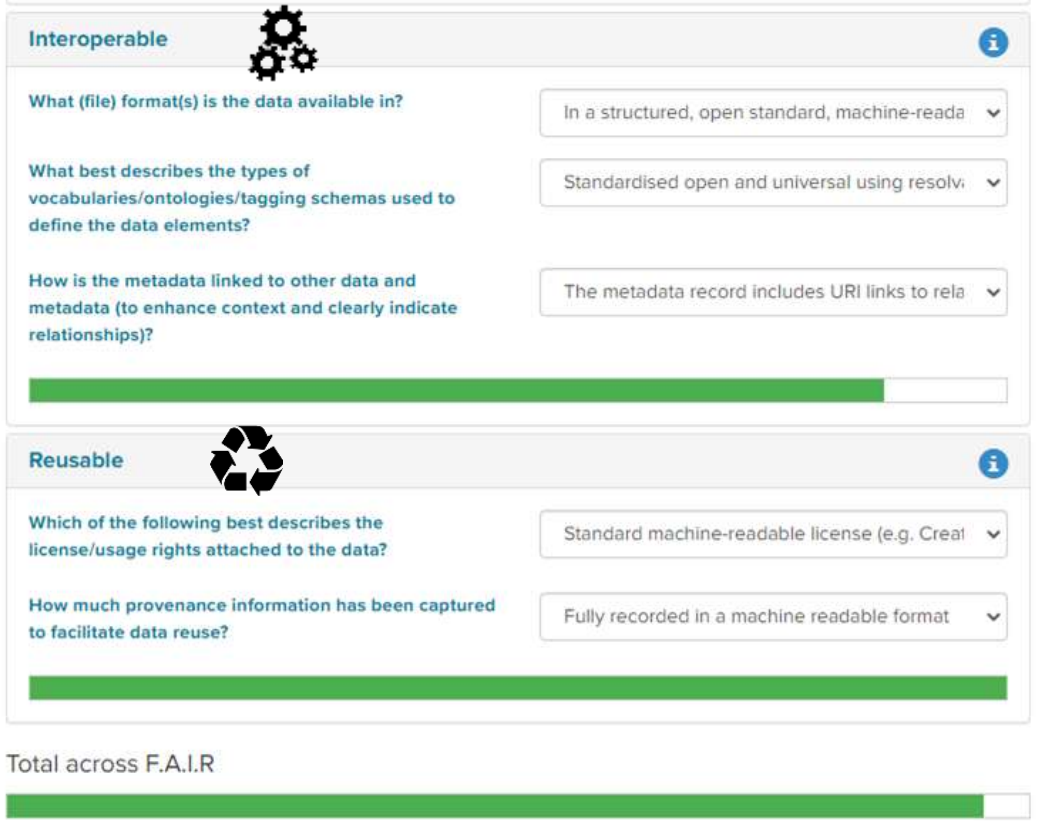
https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/

Open Science - more than just open access to publications!



Open Science - more than just open access to publications!





FAIR	Open data	Similarity	Difference
Findable: Data can be **easily found** by machines and humans.	**Data must be available** 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: **Authentication** and verification is **possible** after the user accesses the data.			
Interoperable: **Data can be integrated with other data** 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 **defined for reuse** and can be replicated and/or mixed in different environments.	• The data should be made available under the condition of reuse and redistribution.	Both **principles have the purpose of making data reusable.**	Open data does not mention metadata and focuses on redistribution neutrality.
	• **Anyone should be able to use, reuse and redistribute the data**— there is no discrimination based on the purpose for which the data is to be used or the individuals/groups wishing to use it.		

https://direct.mit.edu/dint/article/4/4/867/112737/FAIR-Versus-Open-Data-A-Comparison-of-Objectives

Open Science - more than just open access to publications!

…Share your **code**

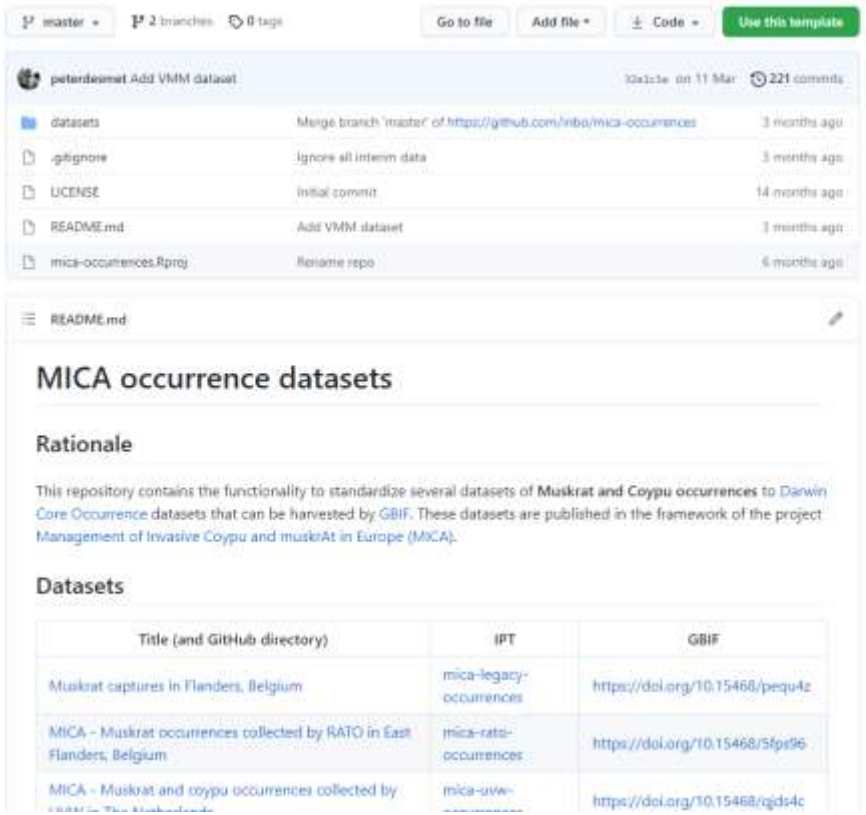
- many researchers now develop bespoke bits of code to help them analyse and/or visualise the data they have collected. Having access to this code is essential for supporting the validation of your findings and to help others to build upon your work.

GitLab

GitHub

git

Open Science - more than just open access to publications!



master · 2 branches · 0 tags · Go to file · Add file · Code · Use this template

peterdesmet Add VMM dataset · 30a1c3e on 11 Mar · 221 commits

datasets	Merge branch 'master' of https://github.com/inbo/mica-occurrences	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	mica-legacy-occurrences	https://doi.org/10.15468/pequ4z
MICA - Muskrat occurrences collected by RATO in East Flanders, Belgium	mica-rato-occurrences	https://doi.org/10.15468/5fps96
MICA - Muskrat and coypu occurrences collected by	mica-uve-	https://doi.org/10.15468/qjds4c

About

Darwin Core mapping for the aggregated Muskrat dataset

dataset · gbif · occurrences

Readme

MIT License

Releases

No releases published

Create a new release

Packages

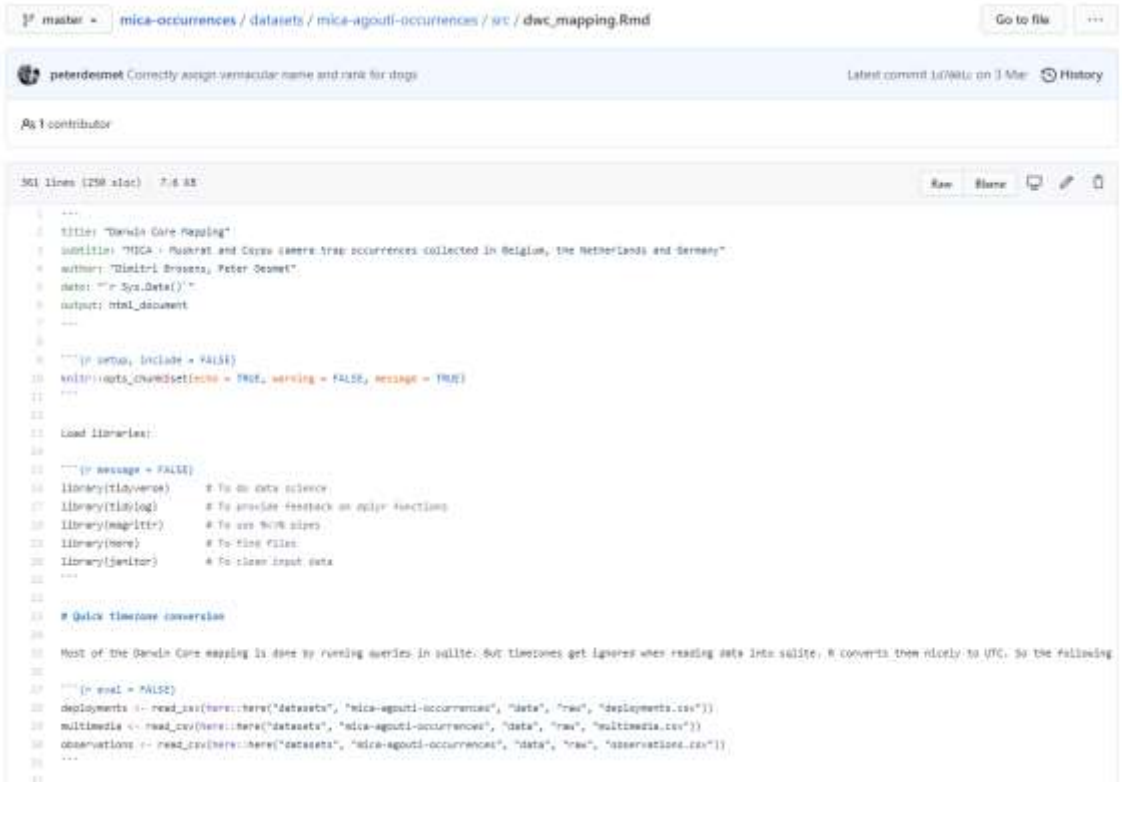
No packages published

Publish your first package

Contributors 2

peterdesmet Peter Desmet

DimEvil Dimi Brosens



master · mica-occurrences / datasets / mica-agouti-occurrences / src / dwc_mapping.Rmd · Go to file

peterdesmet Correctly assign vernacular name and rank for dogs · Latest commit 1d7661c on 3 Mar · History

1 contributor

361 lines (250 sloc) · 7.6 KB · Raw · Blame

```
---
title: "Darwin Core Mapping"
subtitle: "MICA - Muskrat and Coypu camera trap occurrences collected in Belgium, the Netherlands and Germany"
author: "Dimitri Brosens, Peter Desmet"
date: "`r Sys.Date()`"
output: html_document
---

```{r setup, include = FALSE}
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = TRUE)
```

Load libraries:

```{r message = FALSE}
library(tidyverse)      # To do data science
library(tidylog)        # To provide feedback on dplyr functions
library(magrittr)       # To use %<>% pipes
library(here)           # To find files
library(janitor)        # To clean input data
```

Quick timezone conversion

Most of the Darwin 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

```{r eval = FALSE}
deployments <- read_csv(here::here("datasets", "mica-agouti-occurrences", "data", "raw", "deployments.csv"))
multimedia <- read_csv(here::here("datasets", "mica-agouti-occurrences", "data", "raw", "multimedia.csv"))
observations <- read_csv(here::here("datasets", "mica-agouti-occurrences", "data", "raw", "observations.csv"))
```
```

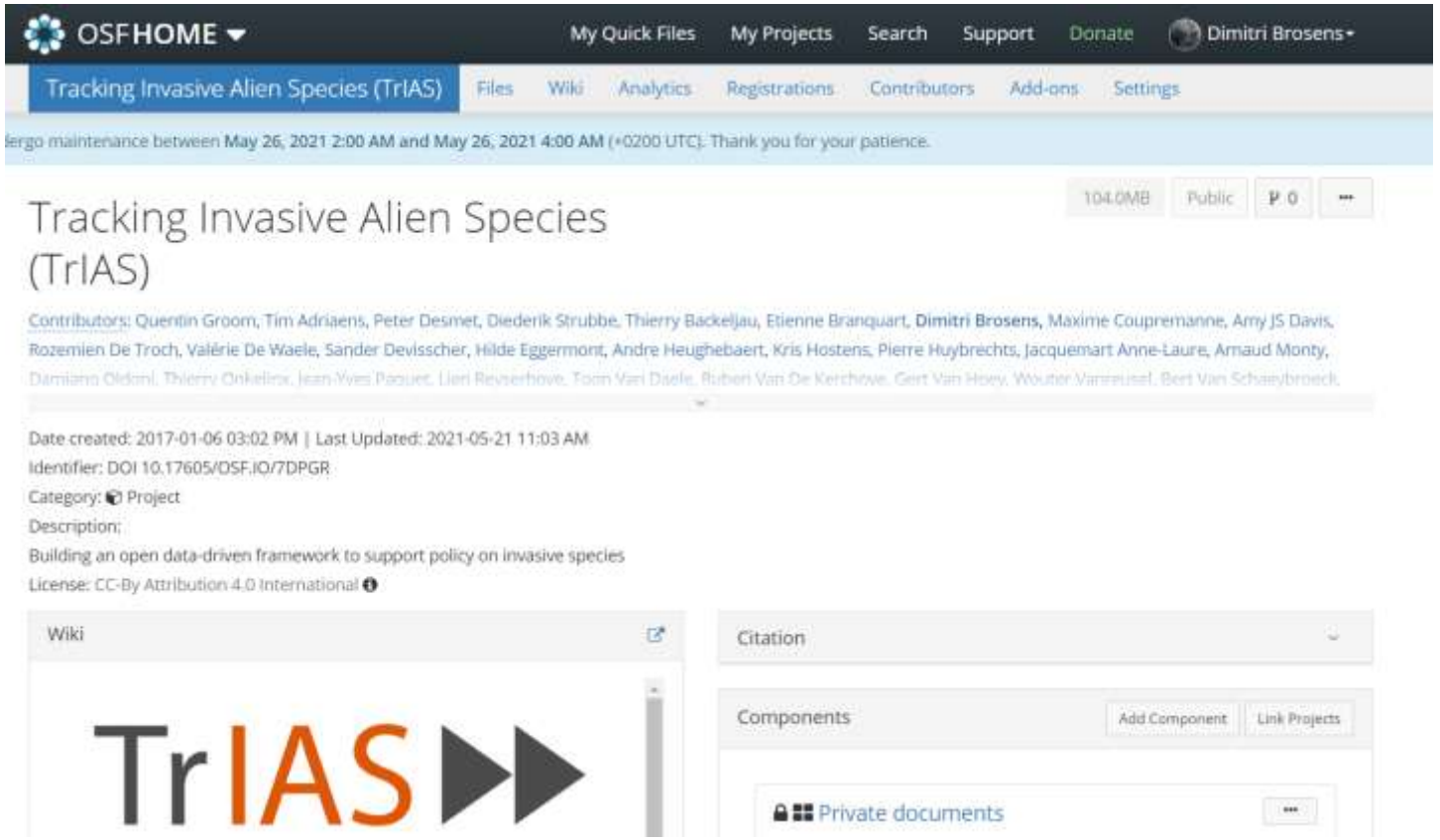
Open Science - more than just open access to publications!

…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!

OSF



Open Science - more than just open access to publications!

…Open **peer review**

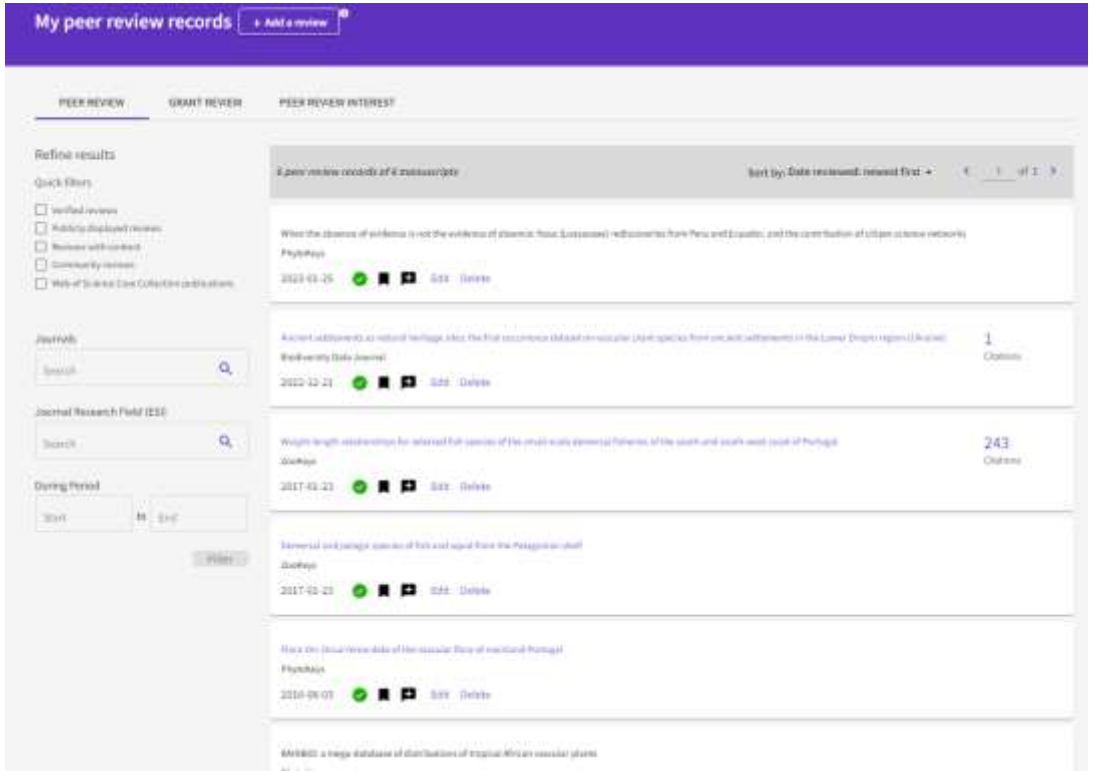
Open peer review (also called "public peer review", "transparent peer review") denotes several, closely related forms of scholarly peer review: Open-identity or attributed peer review (as opposed to anonymous peer review)

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.

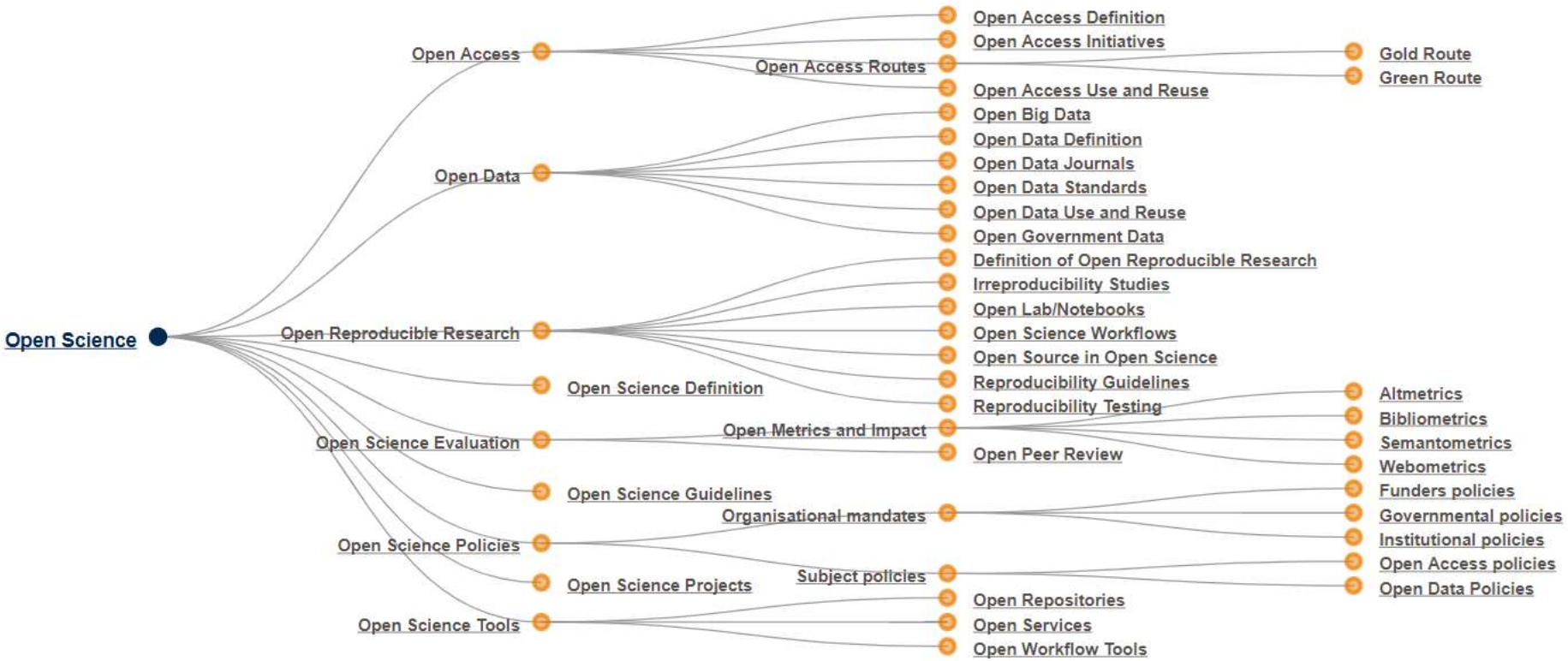


Open Science - more than just open access to publications!

…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.

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.

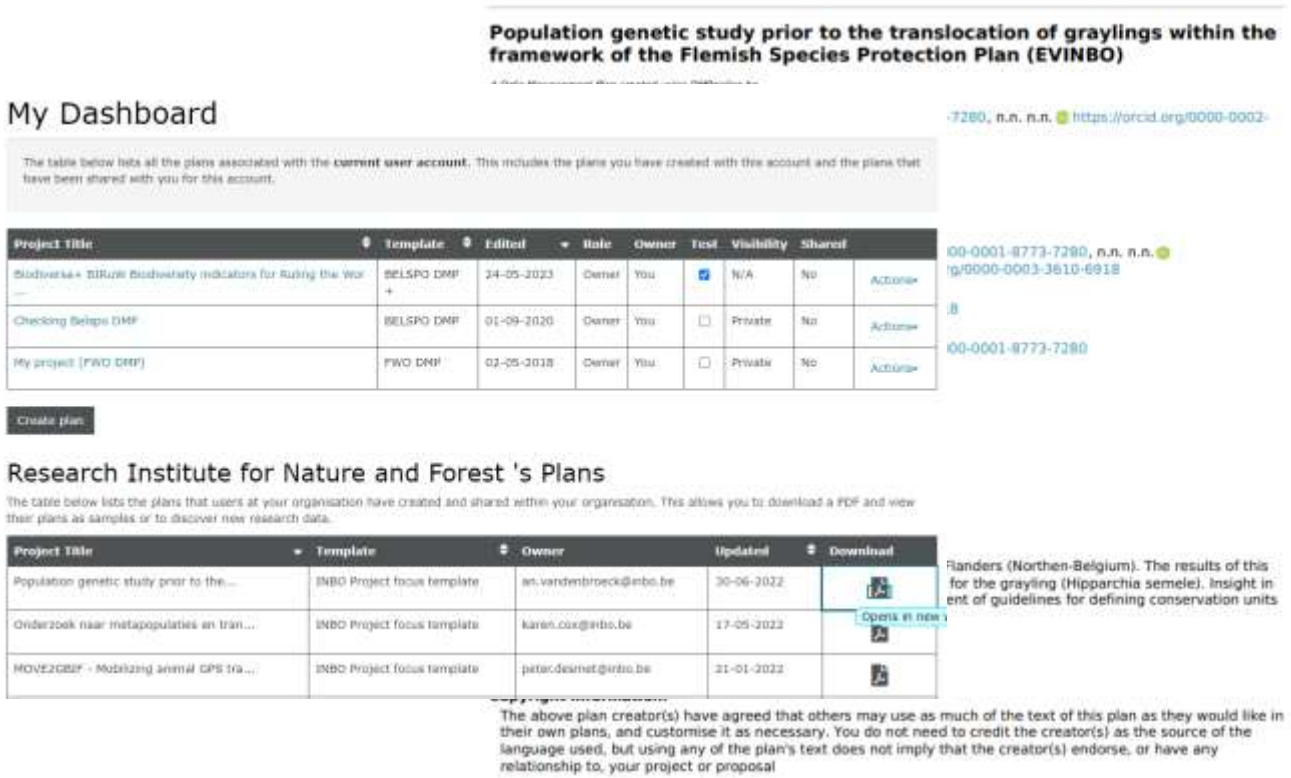
- *Make the research Question is sharp and clear*
- *Research is relevant with the research question*

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

Open Science - Ideas for opening up!

During the planning stage

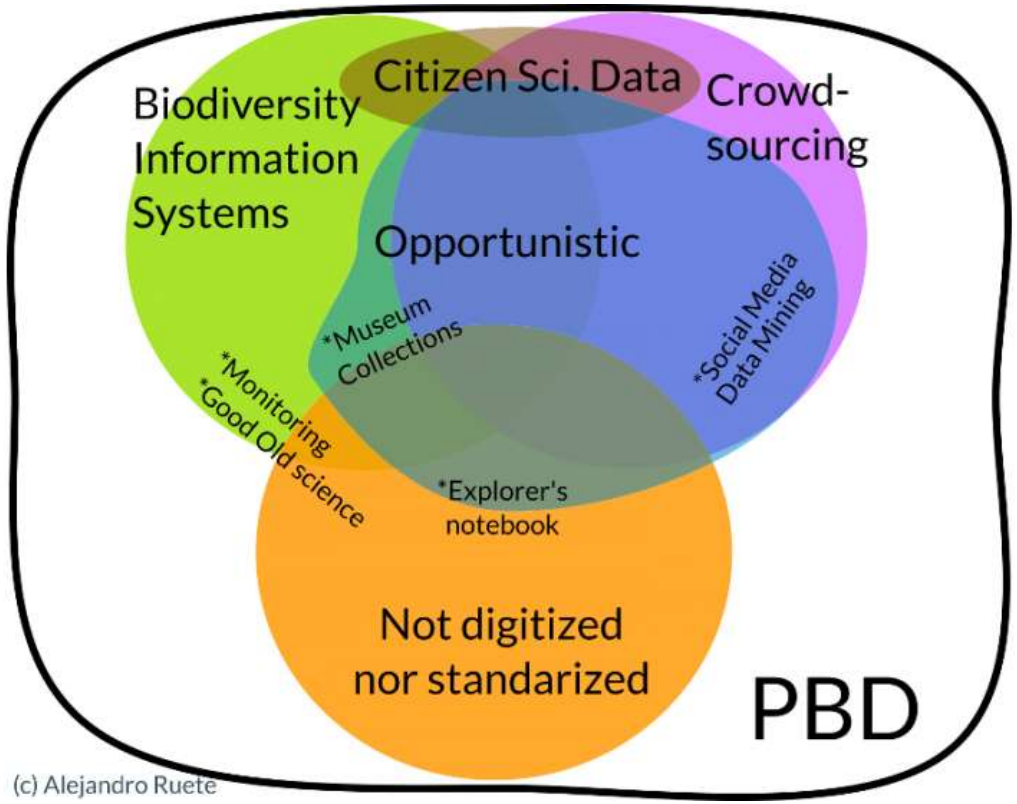
- Write a **D**ata **M**anagement **P**lan
- Create a **C**ommunication **P**lan
- Foresee revision



Open Science - Ideas for opening up!

During the active stage_ data collection

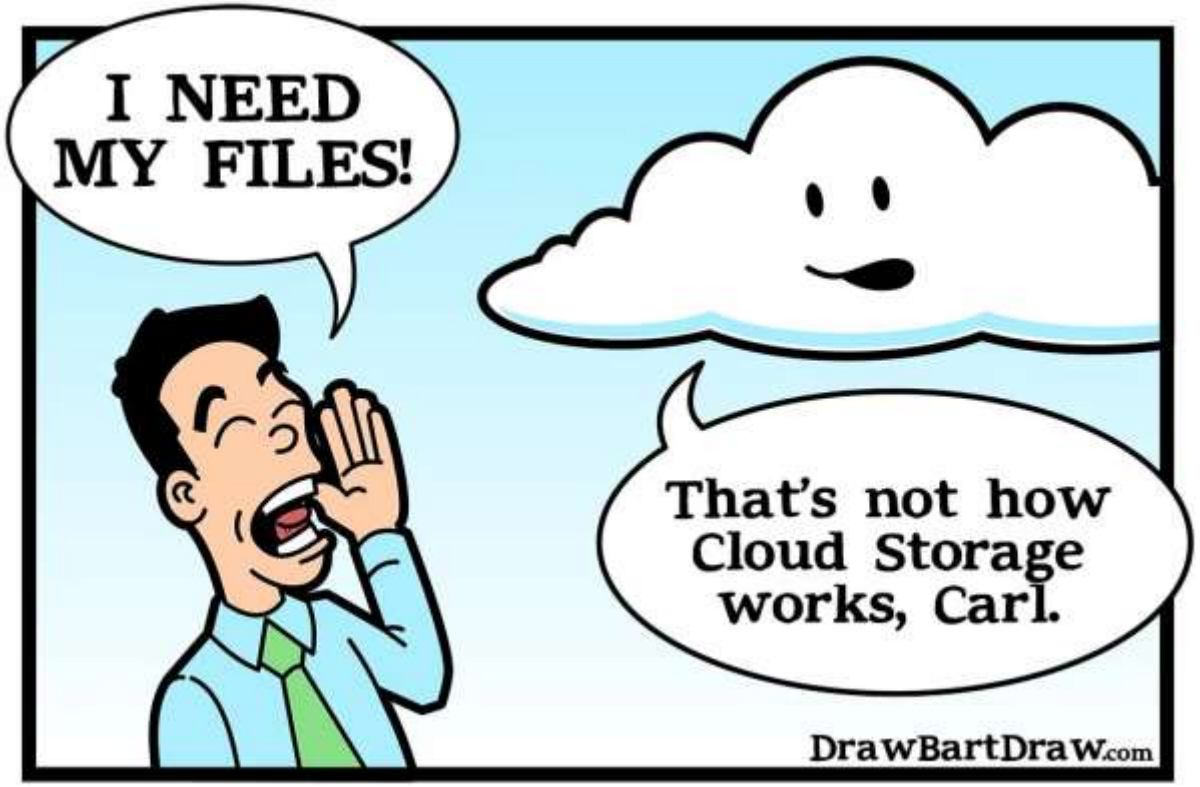
- Data collection through **well defined protocols**
- Automatic data collection when possible
- **Existing data is (re)used** when possible
- Validation procedures are in play
- Measurement biases are understood and documented



Open Science - Ideas for opening up!

During the active stage_ data storage

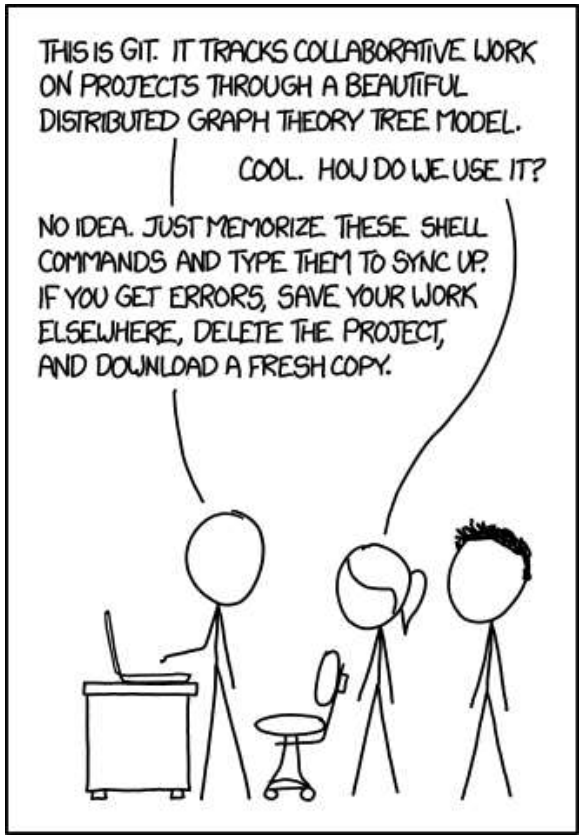
- Data is well organised, documented and stored
  - Use **DMP**
  - **Metadata**
  - Be aware of sensitivities (GDPR…), sensitive data



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

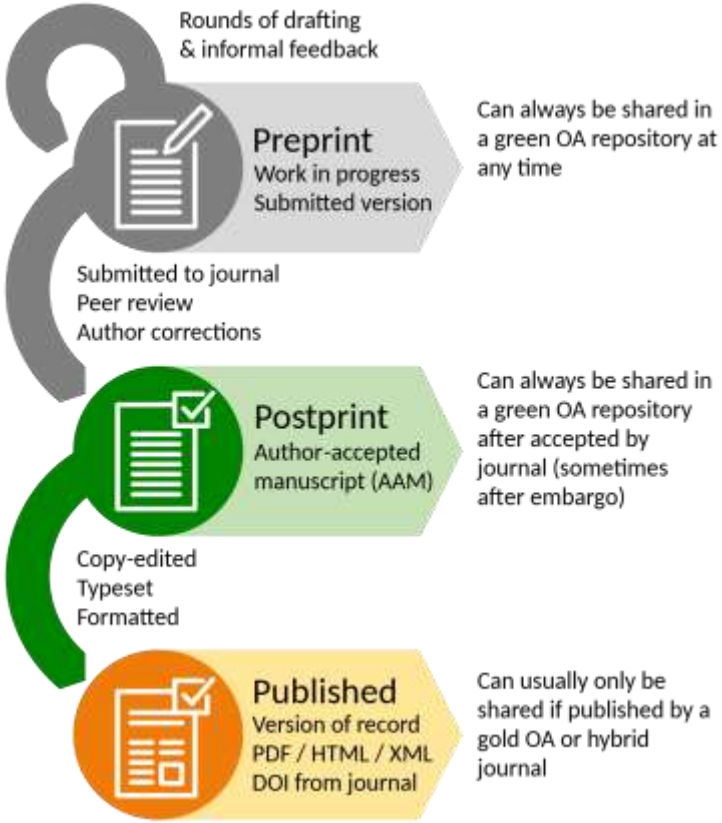


Open Science - Ideas for opening up!

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!

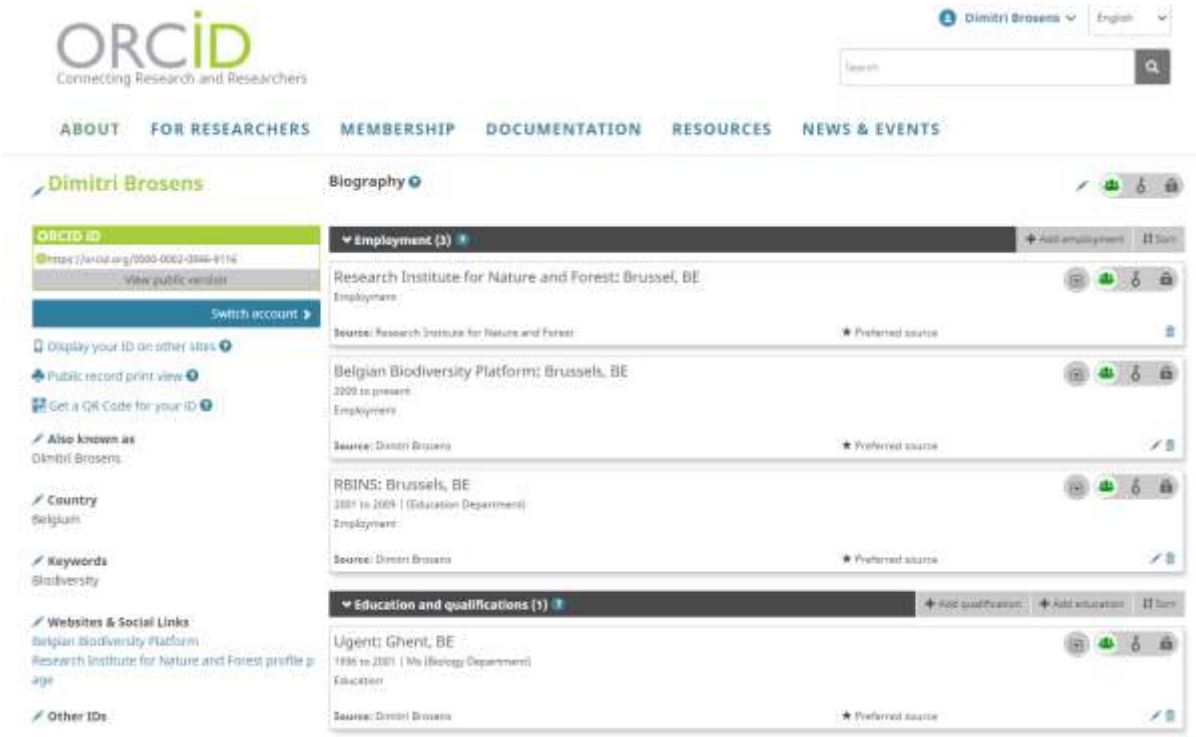
Towards the end of your research_data sharing

- 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

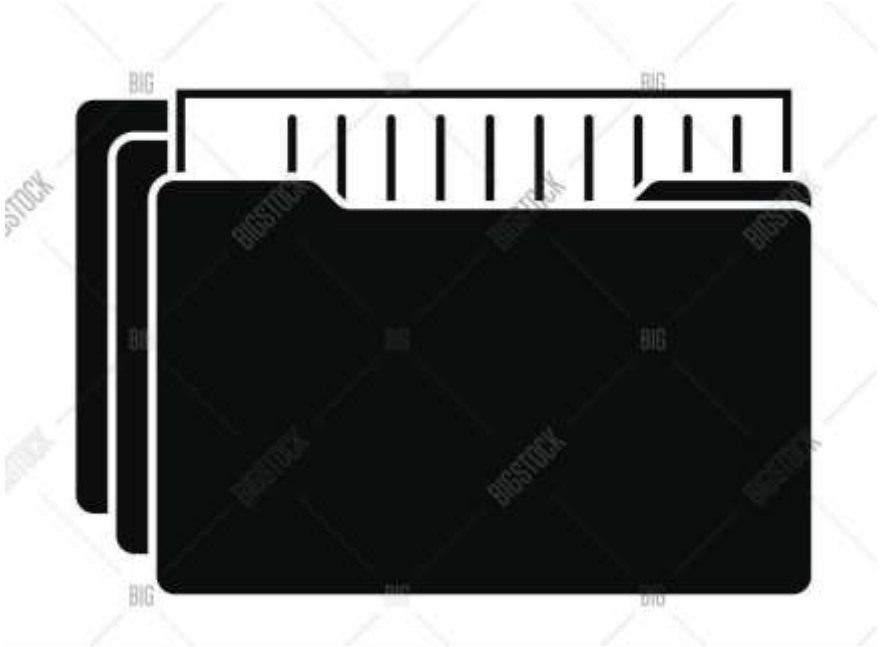
- Published as open access
- Correct citations of sources (DOI…)
- Use of Orcid ID’s



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.

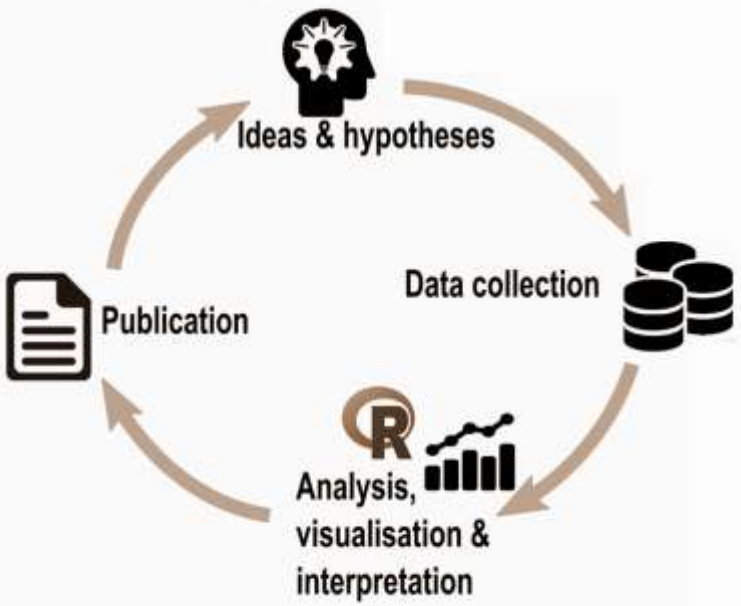
Be sure to **deposit** any **data** required to validate your findings as well as any **software** you've developed to analyse or visualise them.
Link your papers, data, and code to each other through the assignment of DOIs. Link all of these back to you through your ORCID !

What is driving funding bodies to embrace Open Science?

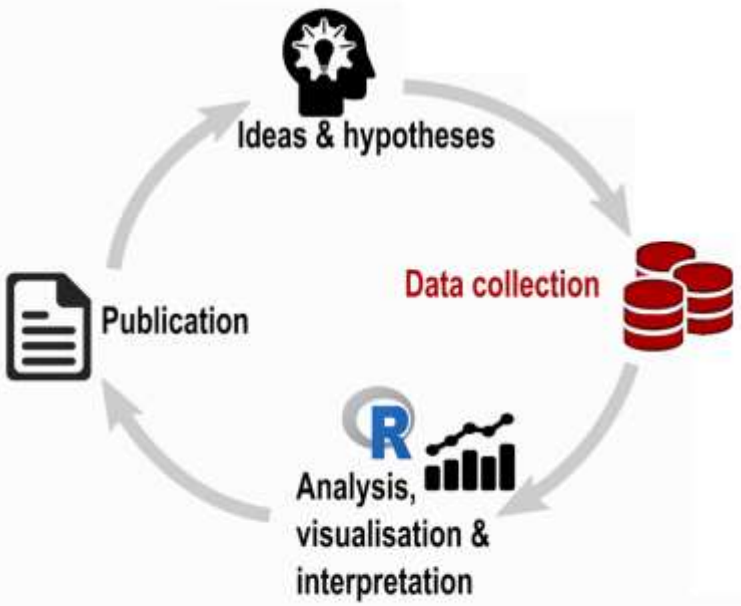
The key driver for this is the belief that publicly funded research should be made available to support

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

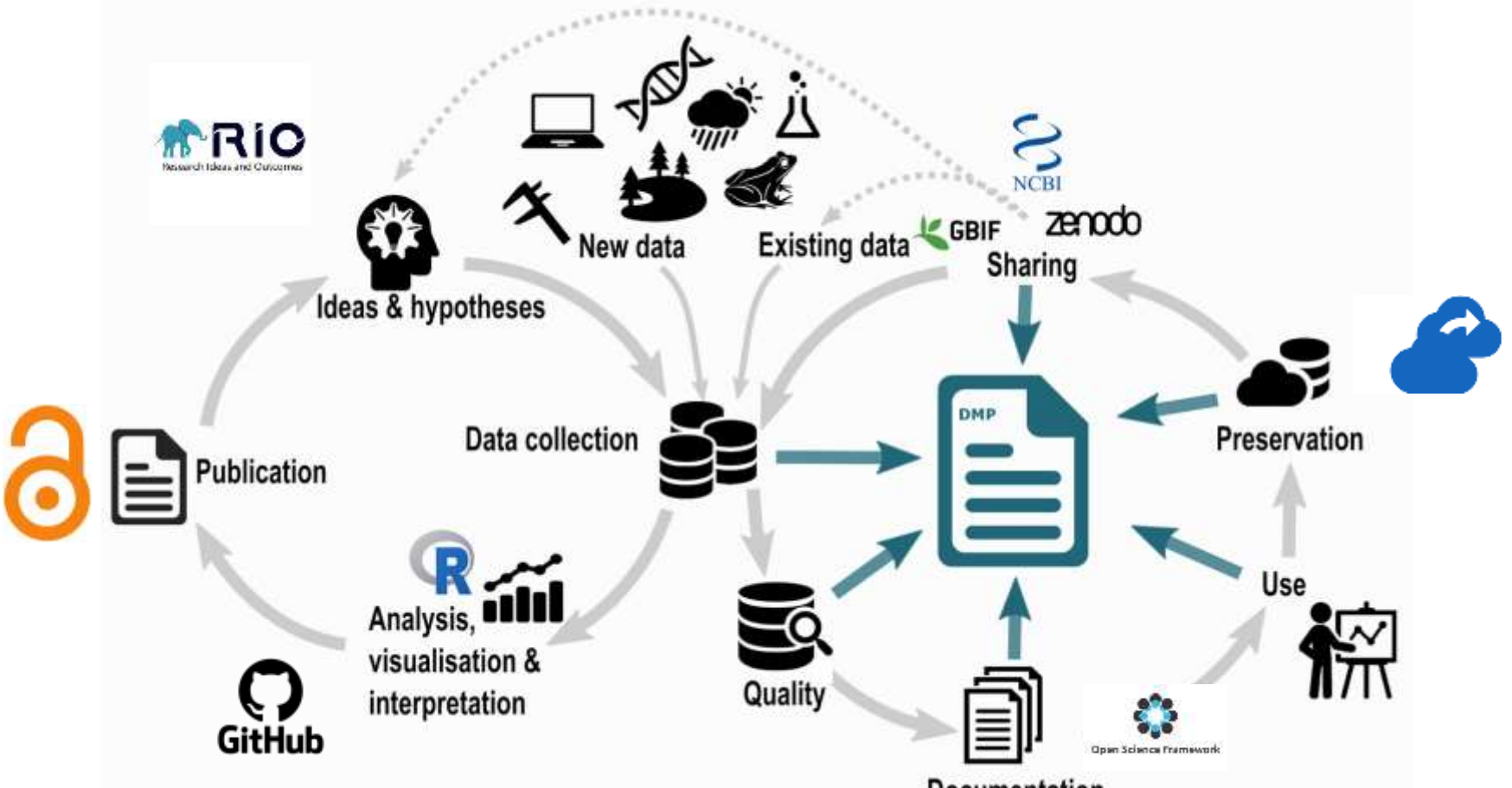
Research Life Cycle

Research Life Cycle

Research Life Cycle

Research Life Cycle

Research Life Cycle: Data Management Plan

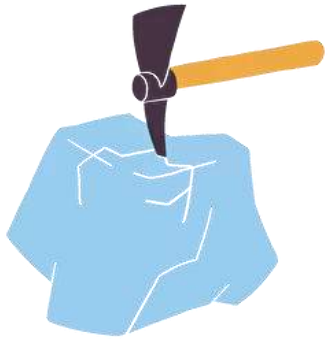


Thank you….

Some Questions?

Menti for the attendants

(from 14:45 to 14:55PM)



Go to: www.menti.com

Enter:6394 9460

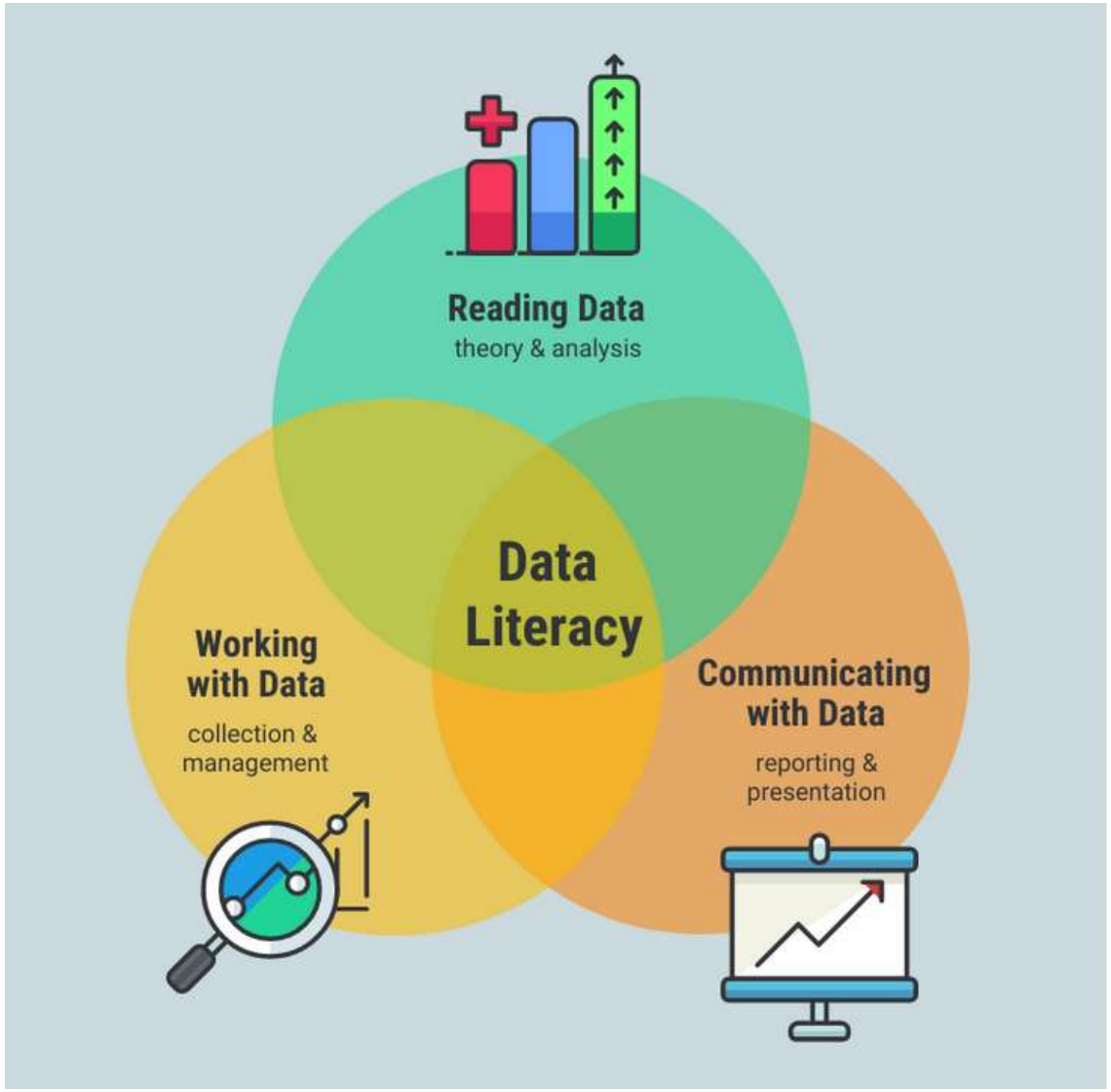
Let’s take a break!

15h00 – 15h15

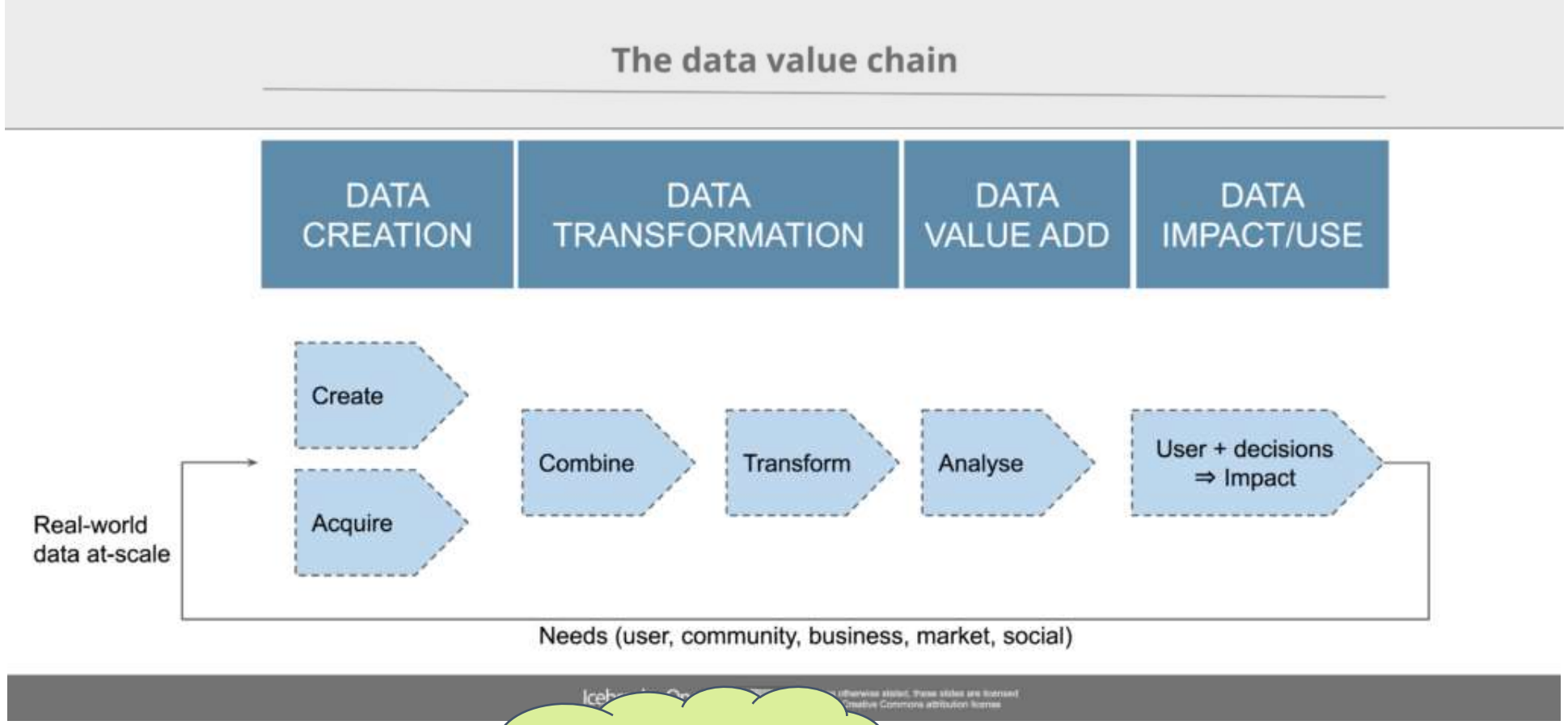
Data management and open data in GBIF

Prof. Dr. Birgit Gemeinholzer

UNIKASSEL
VERSITÄT



Integrating research data management into research and teaching.



Who is interested in my data?

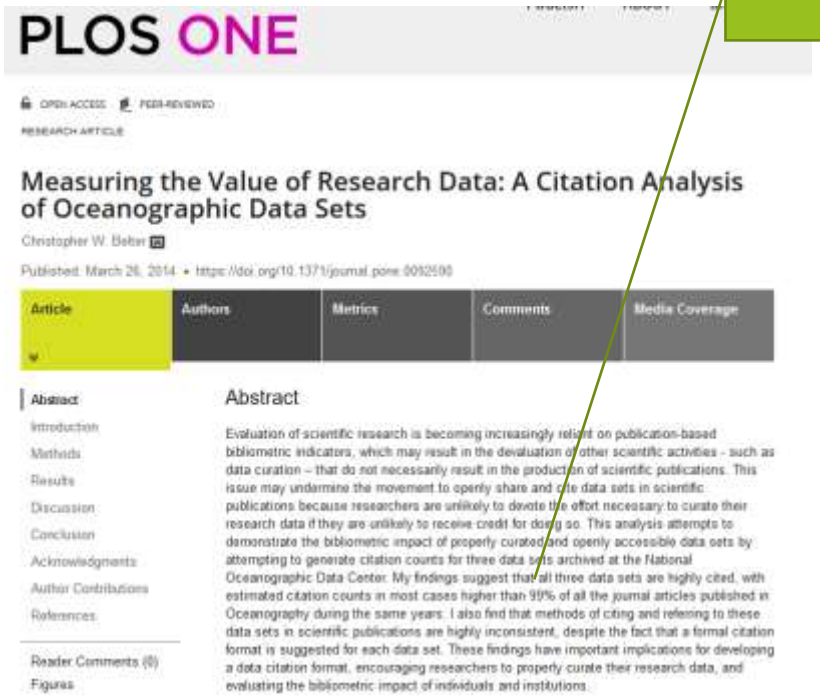
WE ARE!

https://icebreakerone.org/wp-content/uploads/2020/10/IB1-DataValueChain-1024x454.png

……my grumblers……

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

The estimatet 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. Belter

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

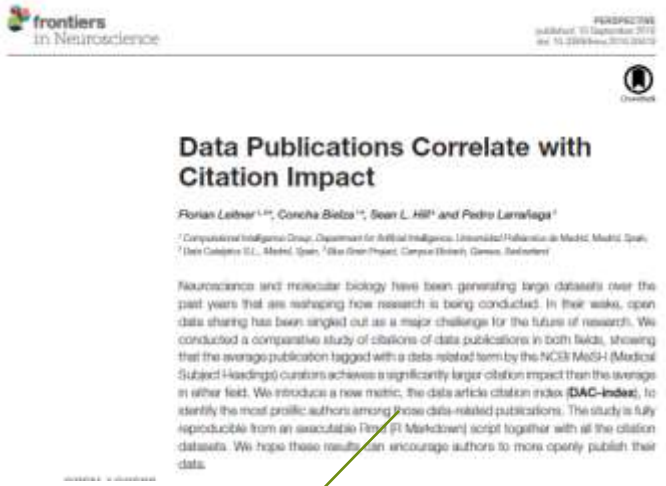
Article | Authors | Metrics | Comments | Media Coverage

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.


frontiers in Neuroscience

PERSPECTIVE

Data Publications Correlate with Citation Impact

Florian Leitner, Concha Bielza, Sean L. Hill and Pedro Larrañaga

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 NCBI MeSH (Medical Subject Headings) curators 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 executable Rmd (R Markdown) script together with all the citation datasets. We hope these results can encourage authors to more openly publish their data.

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


PLOS ONE

PUBLISH ABOUT BROWSE

RESEARCH ARTICLE

The citation advantage of linking publications to research data

Published: April 22, 2020 • https://doi.org/10.1371/journal.pone.0230416

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

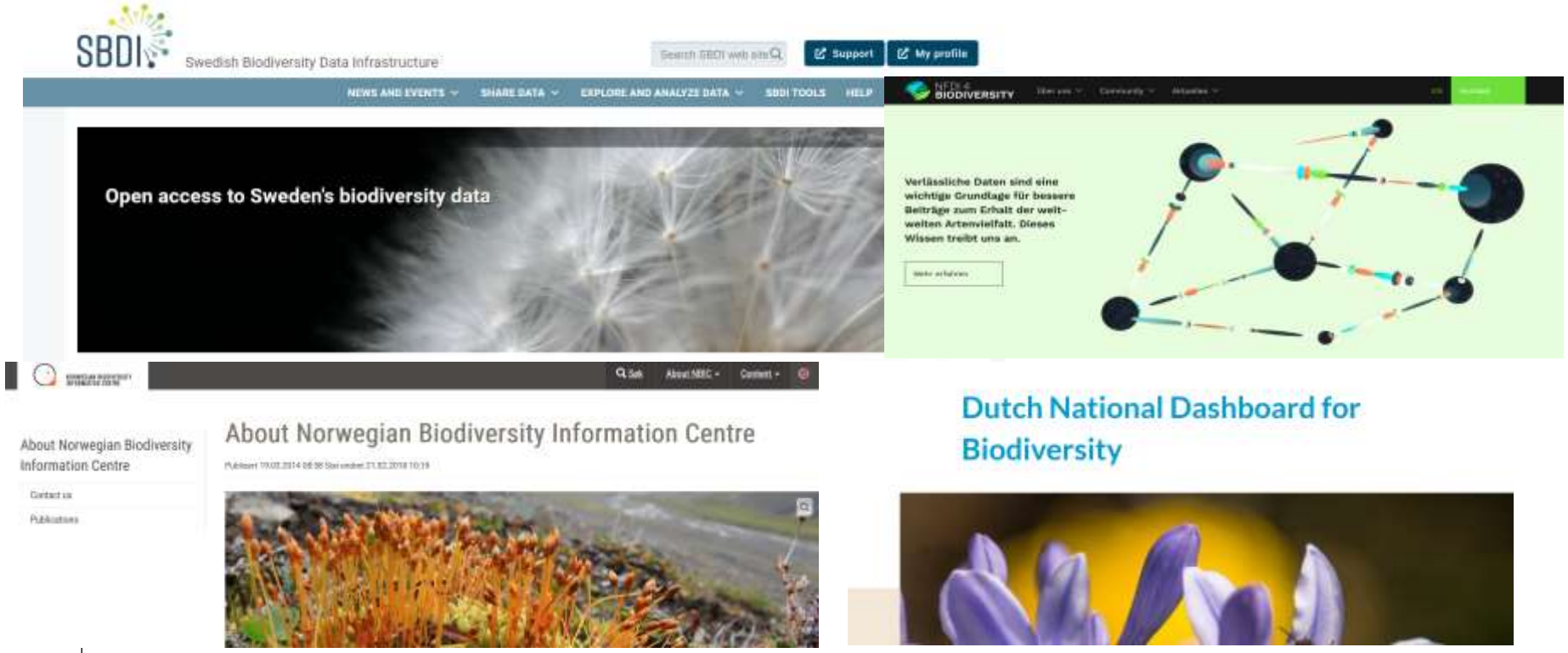
Abstract

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 in 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 531, 808 journal articles published by PLOS and BMC, develop 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 regressions. We find that, following mandated publisher policies, data availability statements become very common. In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository—rather than being available on request or included as supporting information files—are a fraction of the total in 2017 and 2018. 20.8% of PLOS publications and 12.2% of BMC publications provided DAS 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% (± 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.

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

Data life cycle

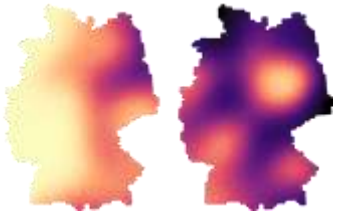
Data submission …….



no scientific publication without primary data submission

Facets of data providers

occurrence data

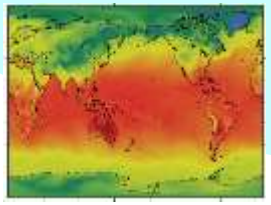


distributions

ecosystem data

experiments

satellite data



model output

molecular data

collection data

functional traits

Adapted from Christian Wirth, iDiv,
https://tse4.mm.bing.net/th?id=OIP.8KcU_c-

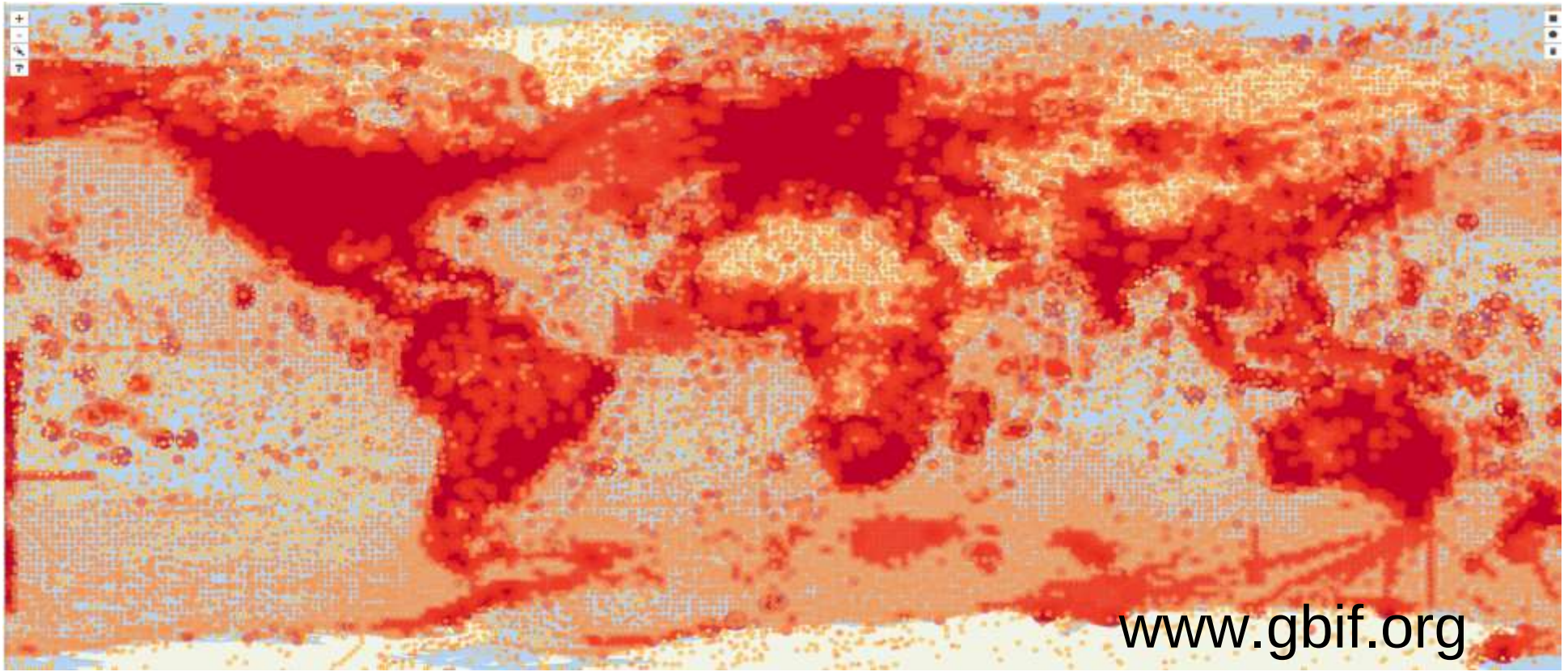
NFDI 4 BIODIVERSITY
BIODIVERSITY, ECOLOGY & ENVIRONMENTAL DATA

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.

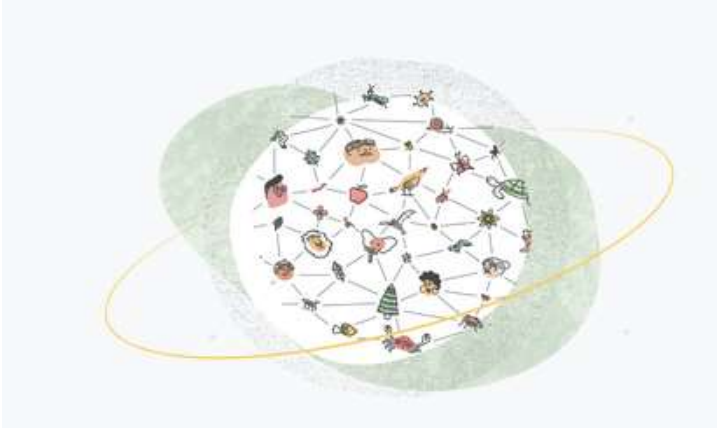


Currently GBIF holds

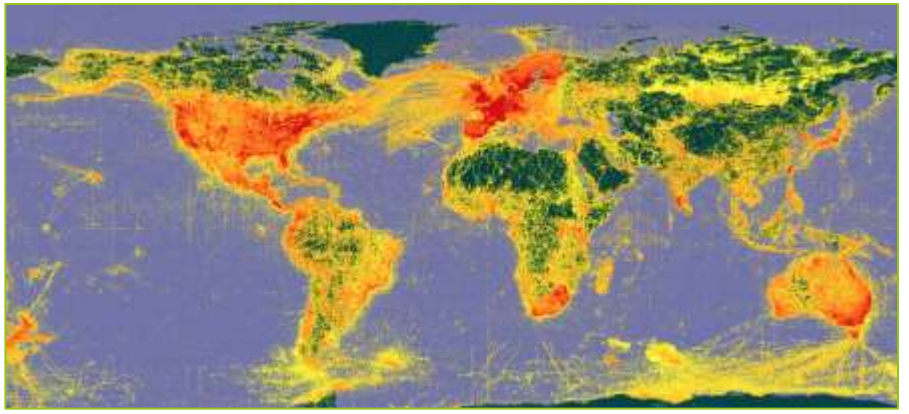
2.951.116.659	105.561	2.233	10.599
Occurrence records	Datasets	Publishing institutions	Peer-reviewed papers using data

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

Accessed: 04.06.2024

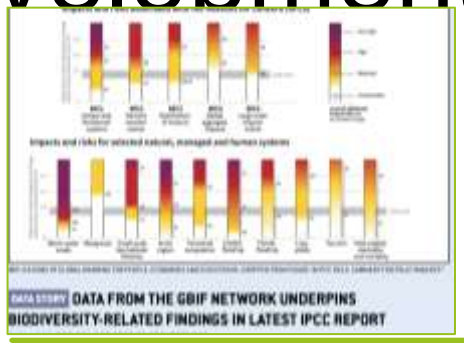


A window on evidence about where species HAVE liveD, and when



https://www.gbif.org/occurrence/search

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

https://www.gbif.org/science-review

GBIF

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; Mélianie Raymond

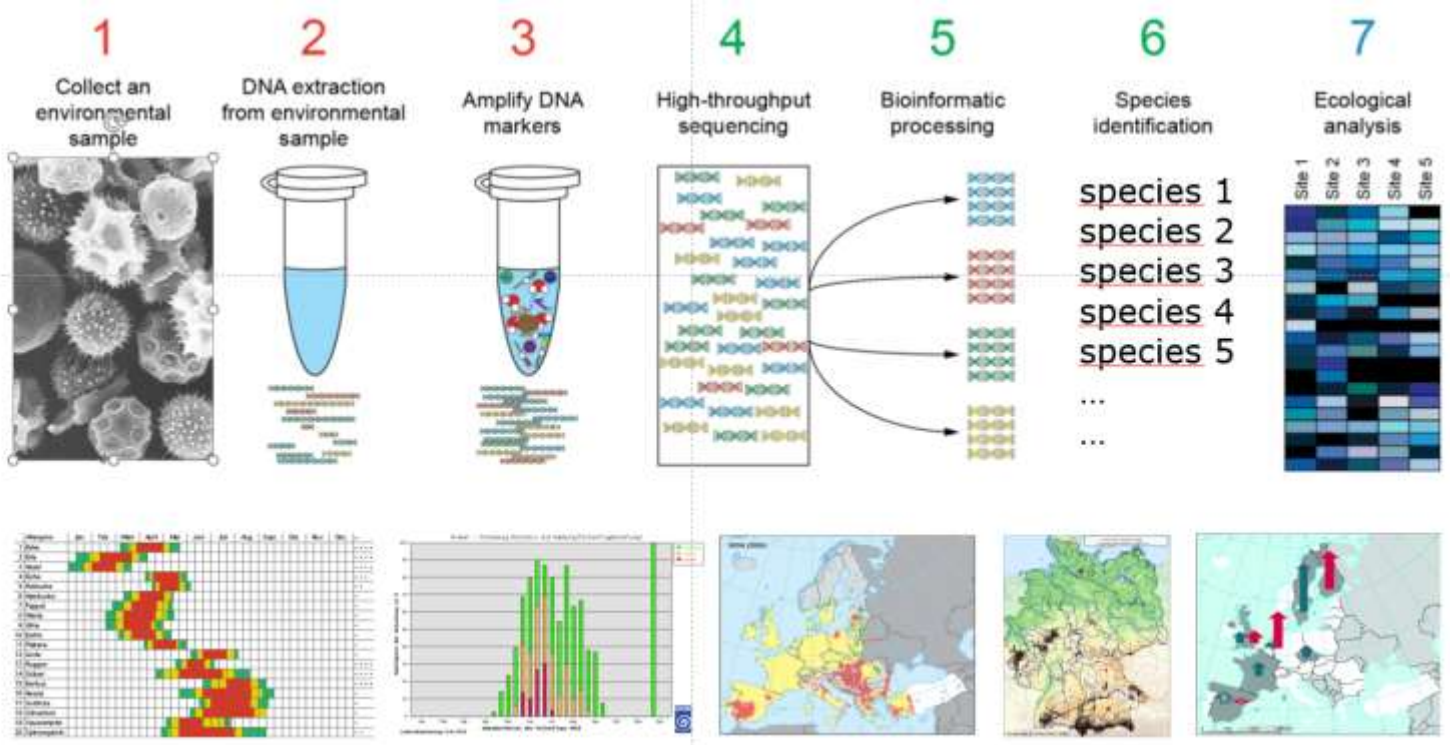
metaplantcode

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

Co-funded by the European Union

biodiversa+
European Biodiversity Partnership


1
Collect an environmental sample
2
DNA extraction from environmental sample
3
Amplify DNA markers
4
High-throughput sequencing
5
Bioinformatic processing
6
Species identification
species 1
species 2
species 3
species 4
species 5
...
...
7
Ecological analysis
Site 1
Site 2
Site 3
Site 4
Site 5

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

Data Publishing · Diversifying the GBIF data model

tfroeslev — 1 · Jun 2023

Jun 2023

1 / 1
Jun 2023

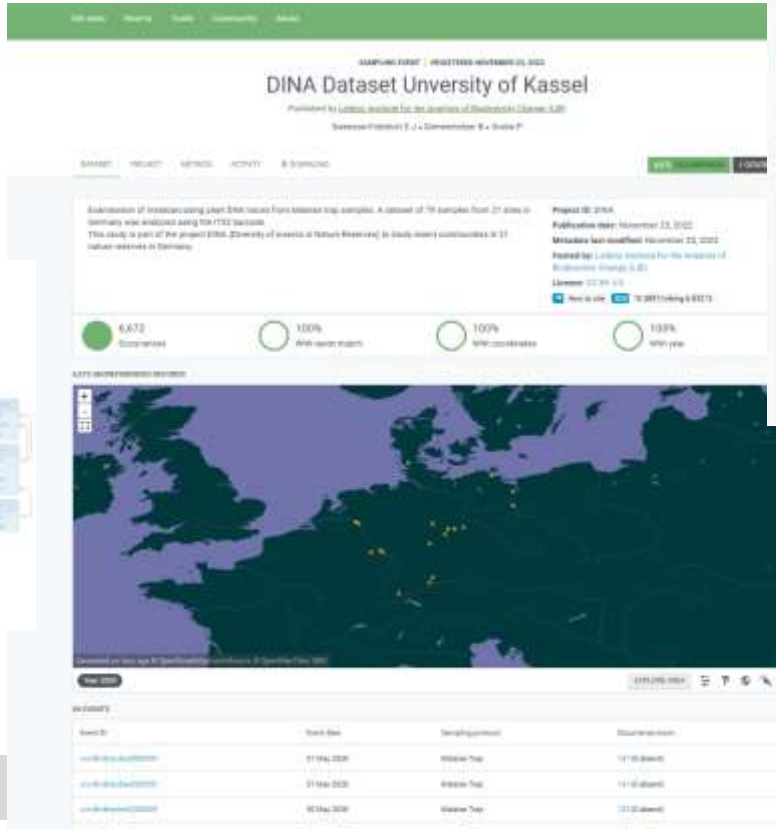
Metabarcoding of environmental DNA samples or bulk samples is one of the major sources of new biodiversity data, and GBIF is exploring ways to expand its support of communities interested in publishing DNA-derived biodiversity data and increase its visibility and reuse beyond molecular repositories and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

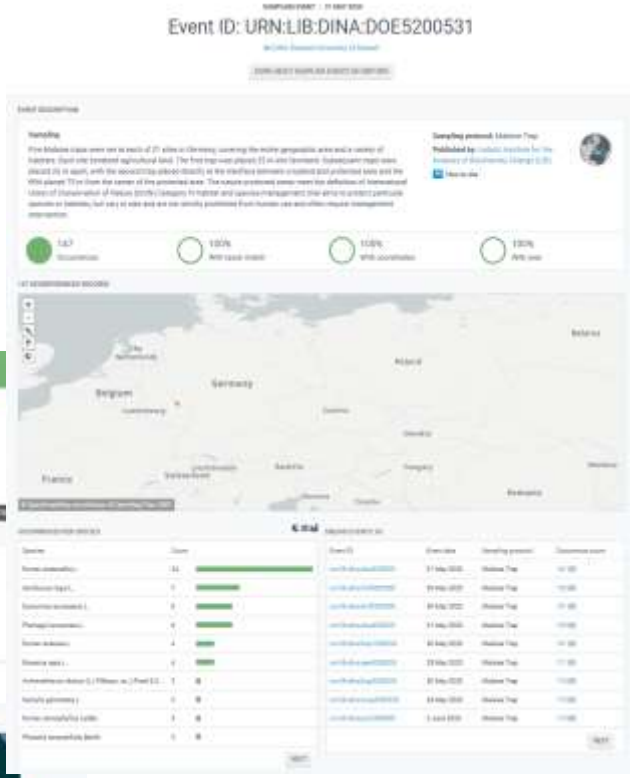
On the occasion of updating our guide 25 on sharing such data through biodiversity platforms—which now includes a special section on publishing marine eDNA data—GBIF invites people who hold DNA metabarcoding data to help us pilot an experimental data publishing tool that responds to recent feedback from the omics community. .

Use Case

Global Malaise Programme in the "Grand Unified Model"

https://www.gbif.org/event/33MQ8ZI9xuC98AekewiMr8/gbif-community-webinar-diversifying-the-gbif-data-model

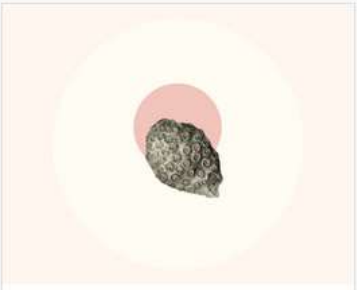




www.gbif.org

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

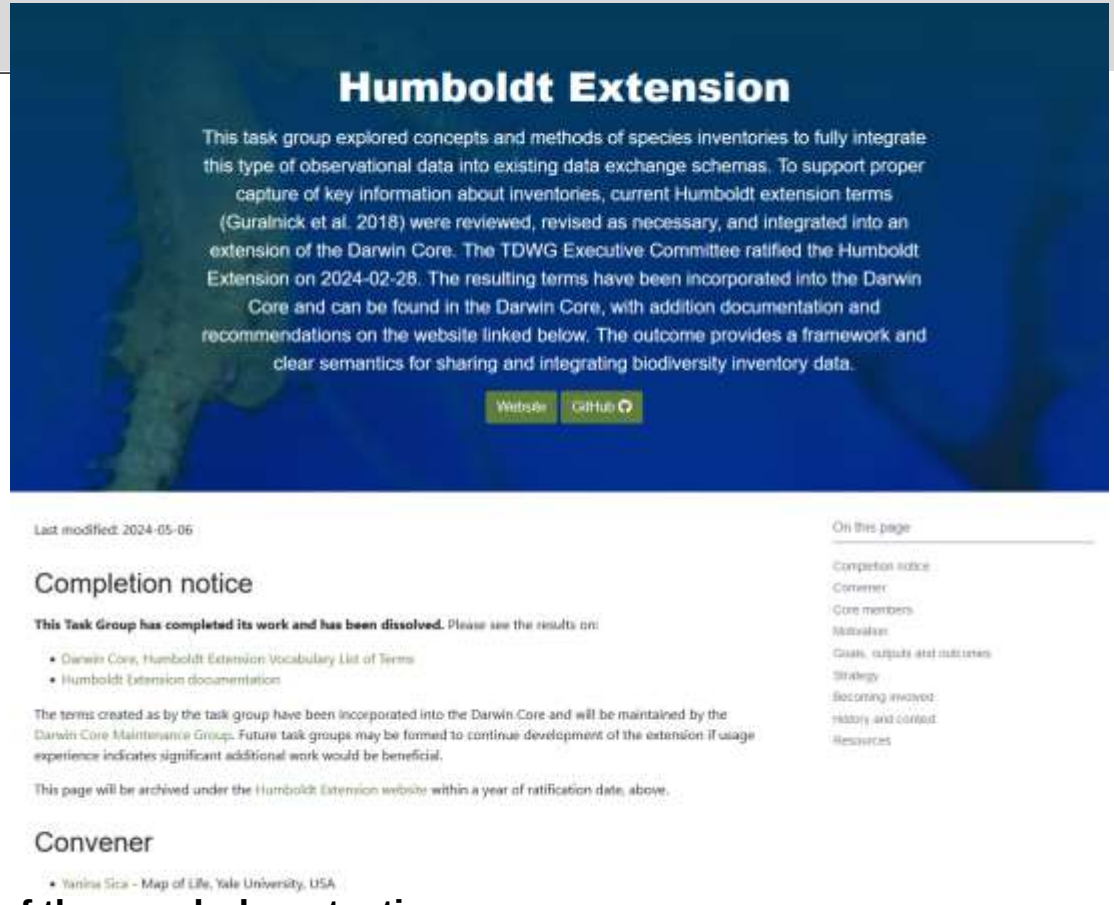
Download webinar presentations

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



Implementation of the vocabulary: testing

Use case: eBird Volunteer–Collected Observations of Birds

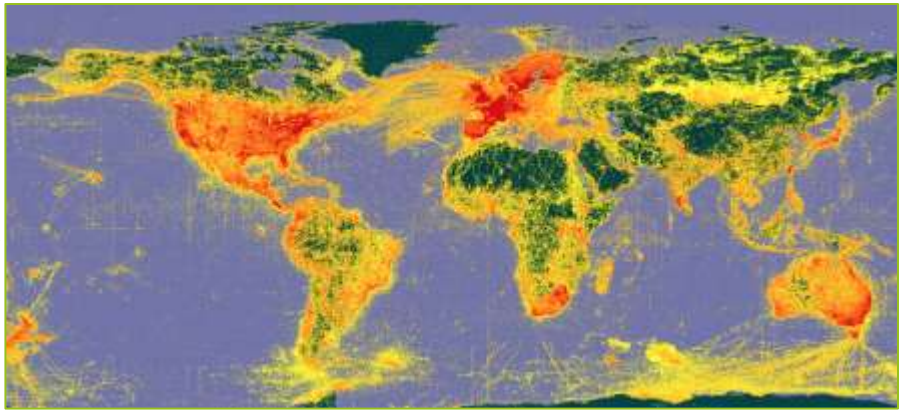
Use case: Field Museum Rapid Inventory Data

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



https://www.gbif.org/occurrence/search

….. 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!

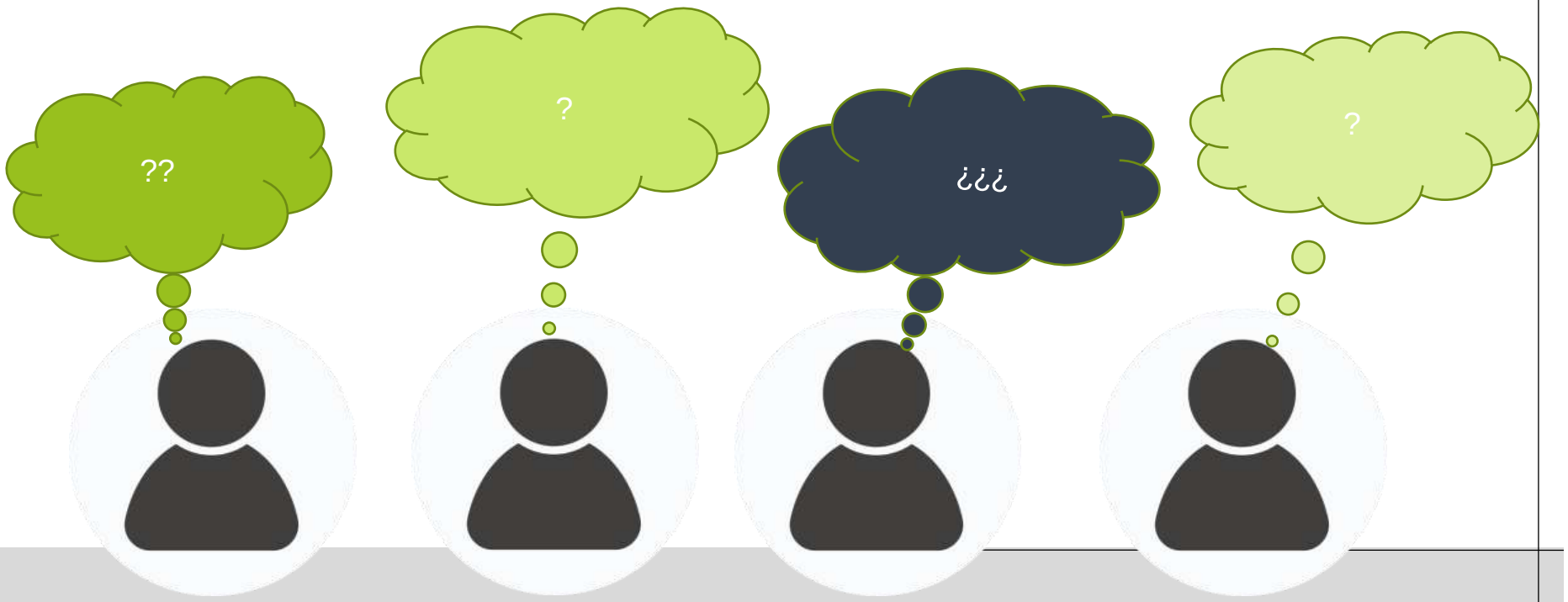


Photo by Melissa Liu. CC BY-NC 4.0

GBIF | Global Biodiversity Information Facility

Introduction to GBIF training package on data mobilization

Melissa Liu | Asia Regional Support Team

Gbif strategic plan

01

02

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

03

04

2023 - 2027

GBIF Strategic Framework

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

GBIF

Capacity Development objectives

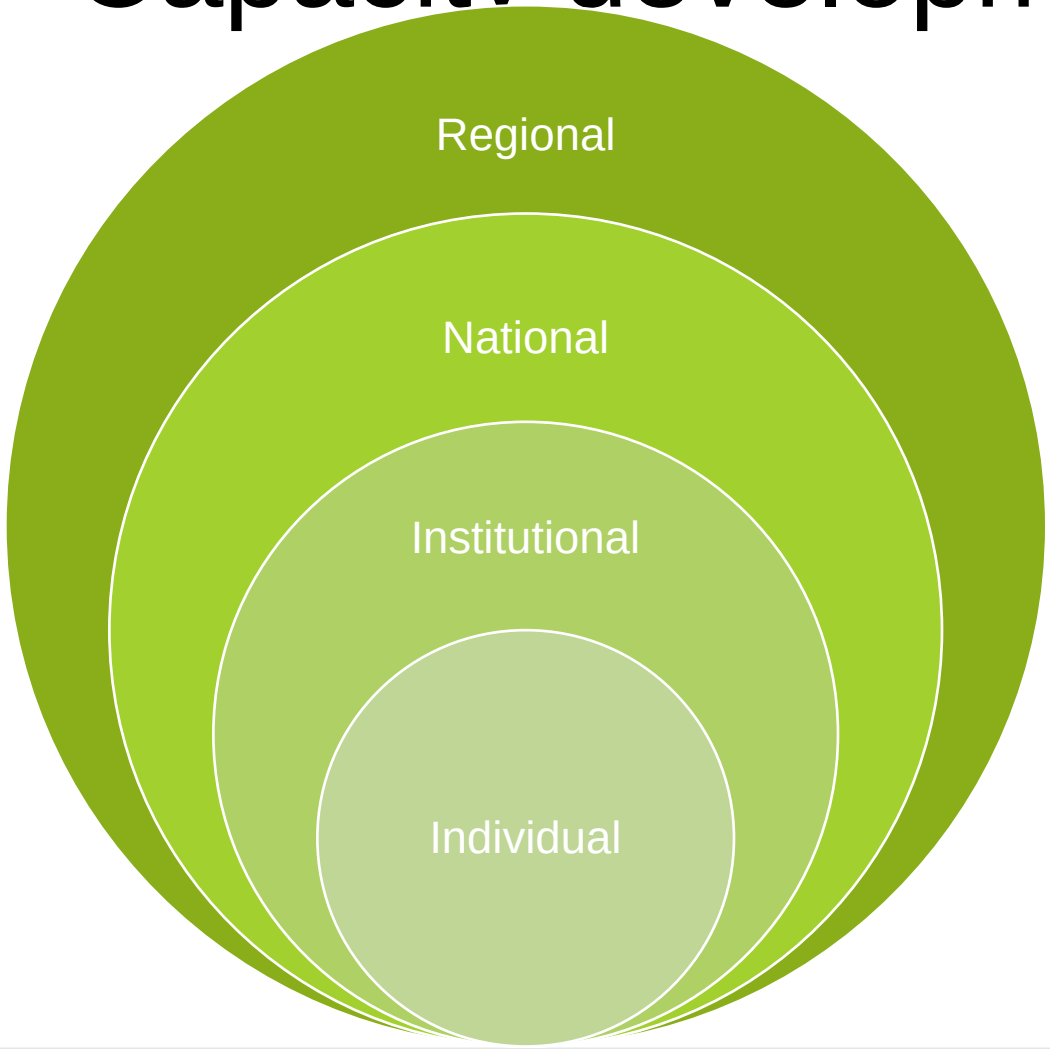
Developing the capacity to mobilize biodiversity data through GBIF

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

https://www.gbif.org/training

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/kRIuIUYdNr

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

New Publishers

12

from

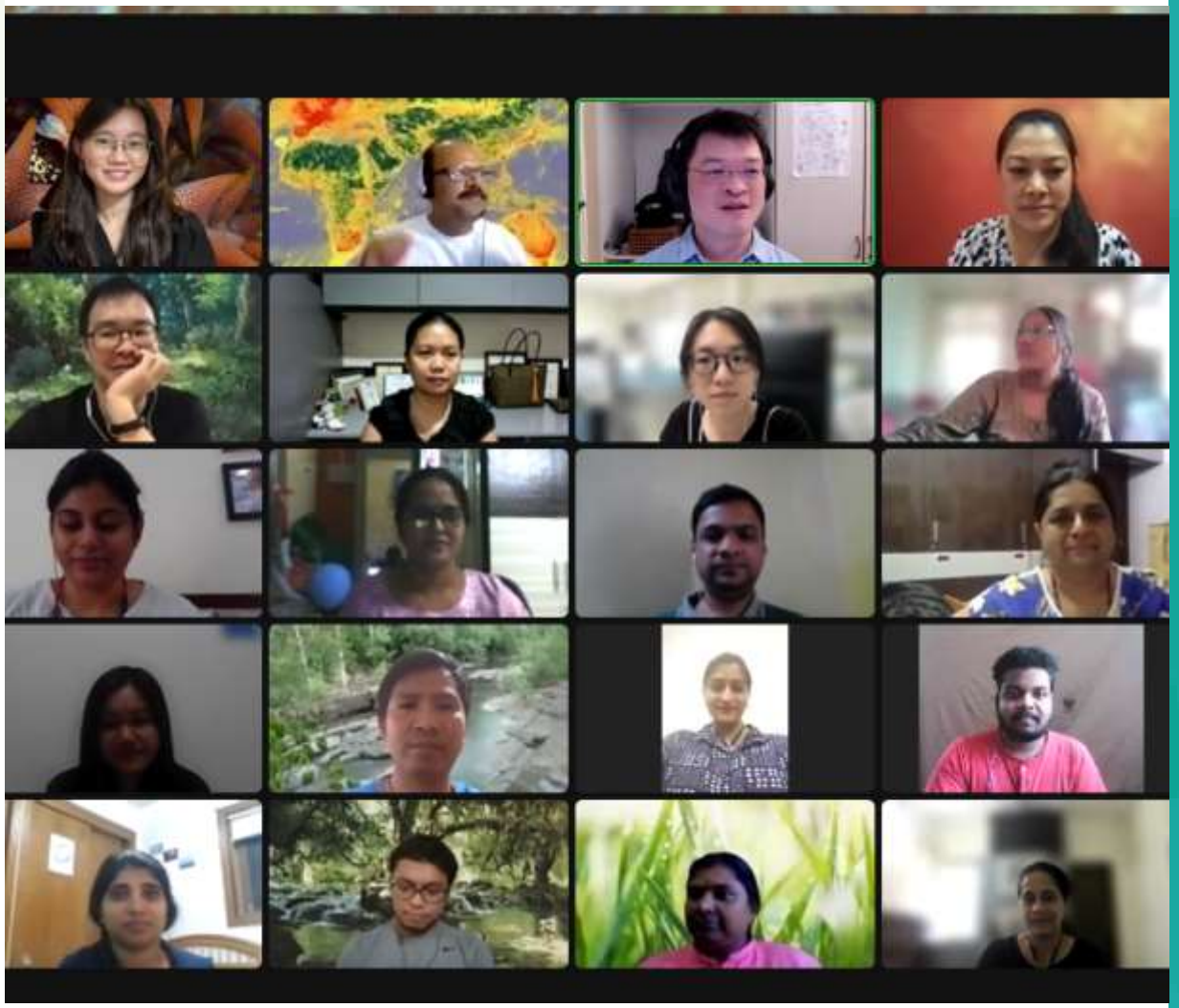
14 countries

Workshop Participants **60**

Self-Paced Participants **27**

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

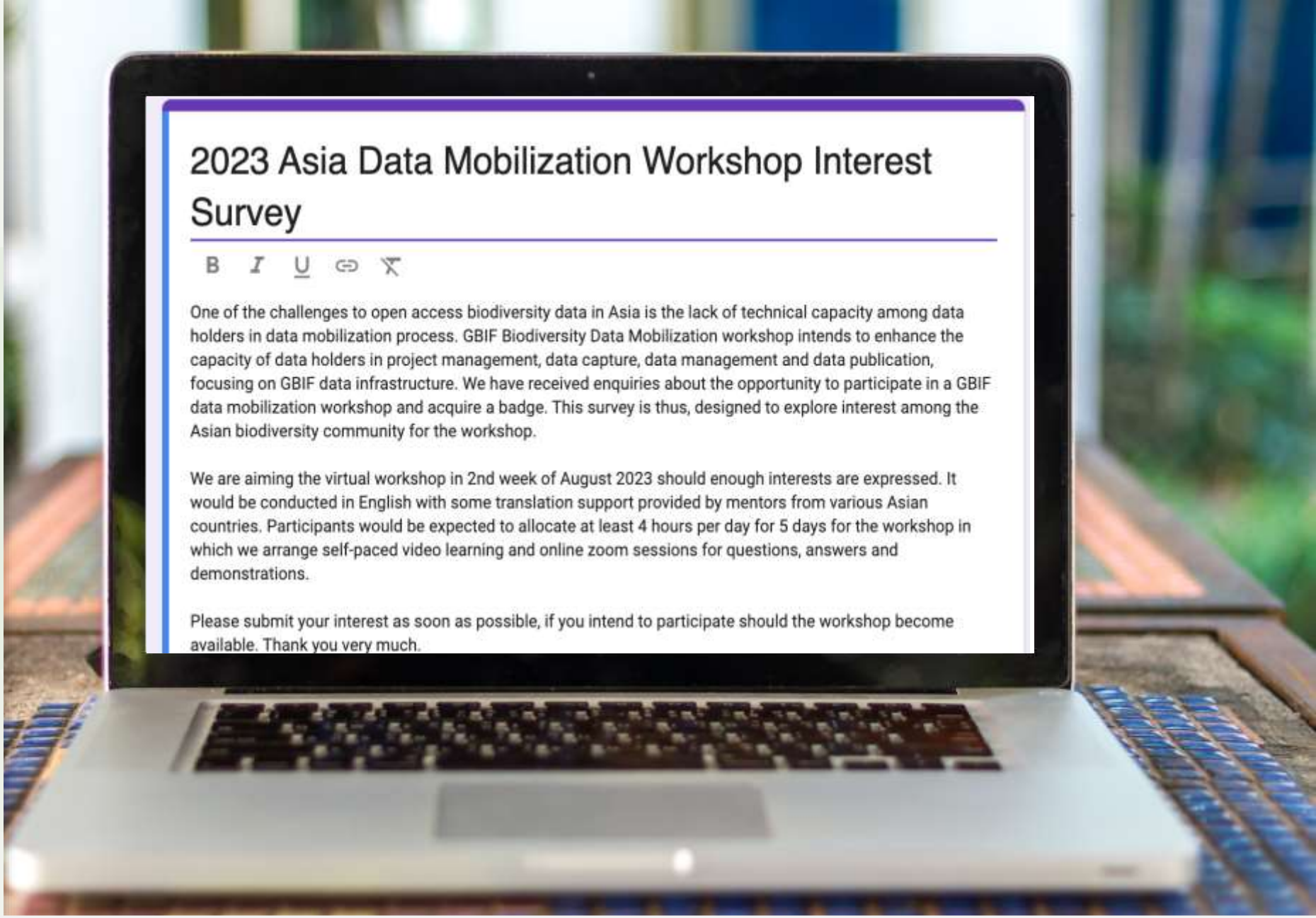
Design an application form and spread it out to potential communities.

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



Decide module & date

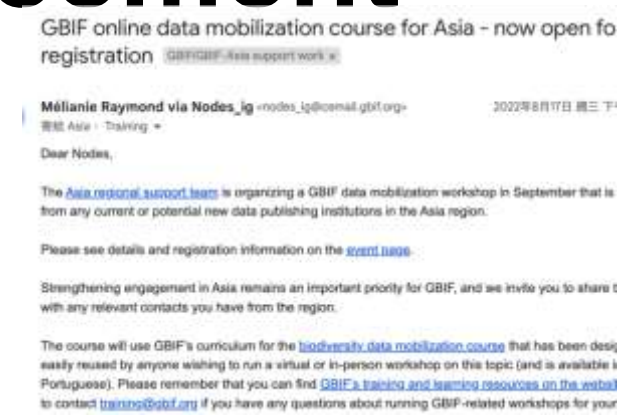
2023 Virtual Data Mobilization Workshop

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 Final review, conclusion & assignments & Feedback (google form) Leads: Asia supporters
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	
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



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.

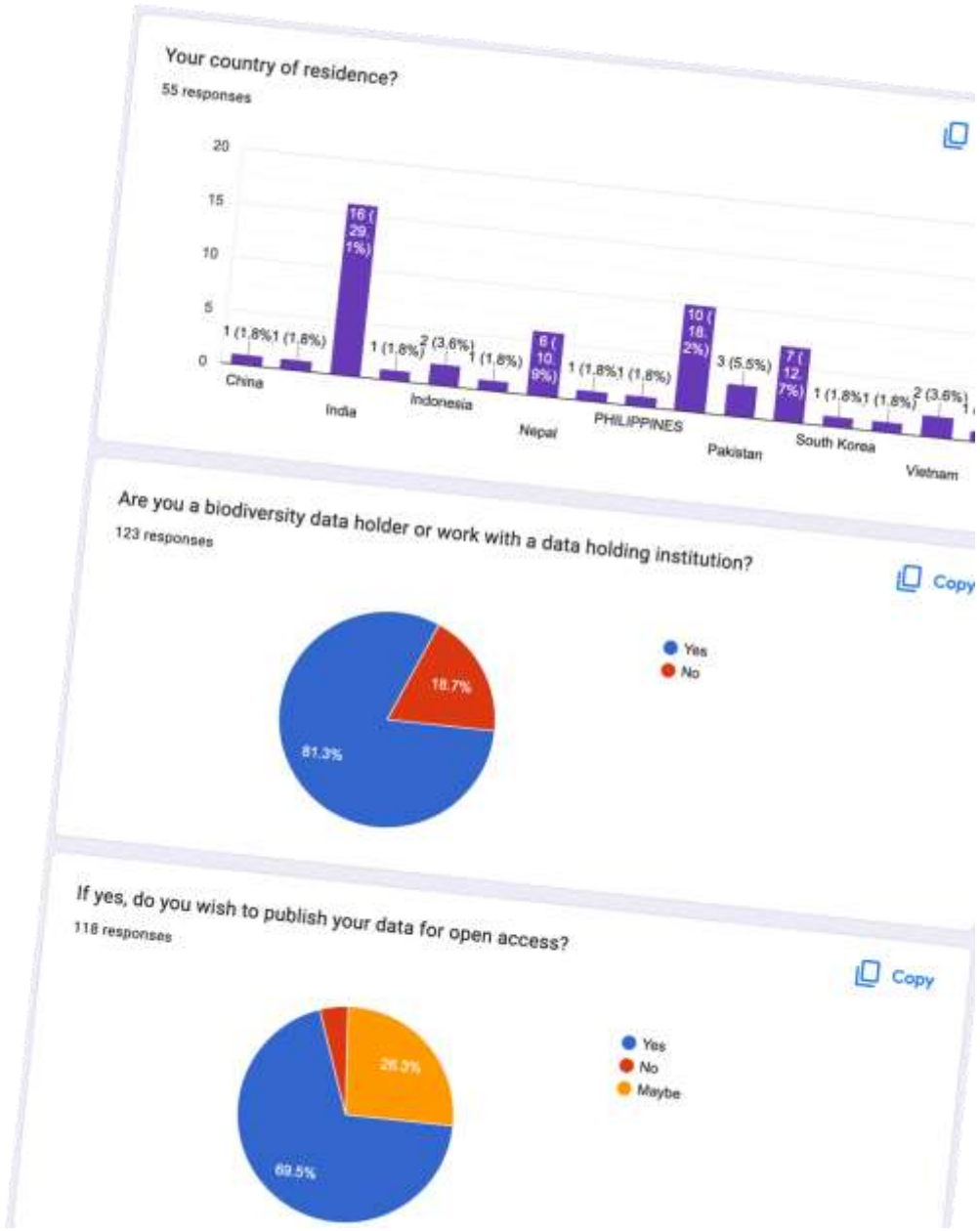
- Regional virtual office hour
- Facebook, Twitter, etc.
- Mailing lists
- Any other contacts

GBIF

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

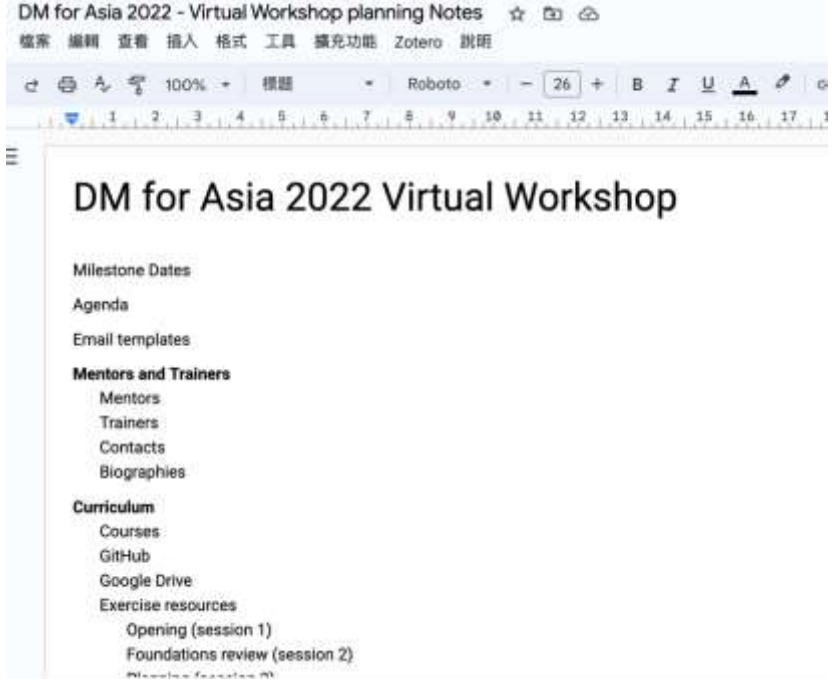


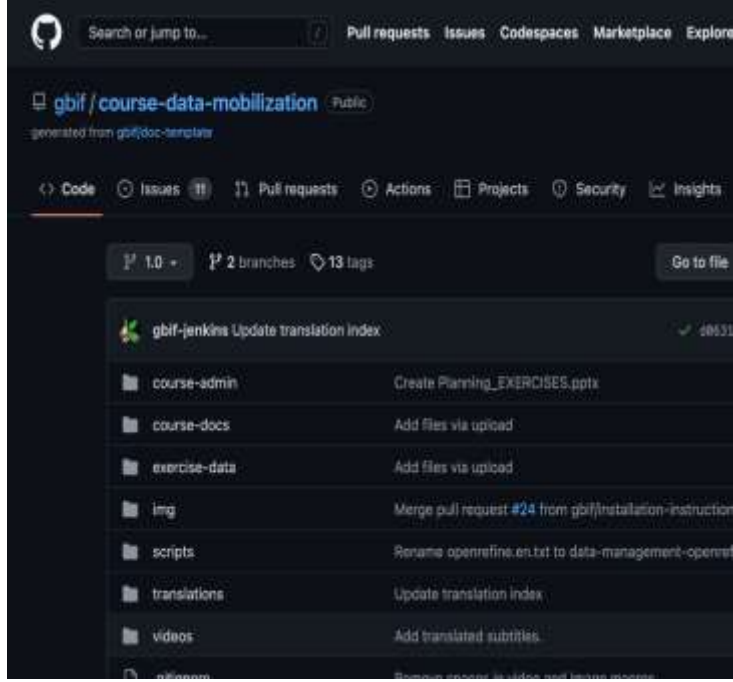
Recruit mentors

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 García. Photo 2015 GBIF | Mélianie Raymond.





- 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

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 Final review, conclusion & assignments & Feedback (google form) Leads: Asia supporters
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	
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 of 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.

← Planning rubric | ↑ Assessment and certification | ⌂ Home | Data management rubric →

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

Incentives

- 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)

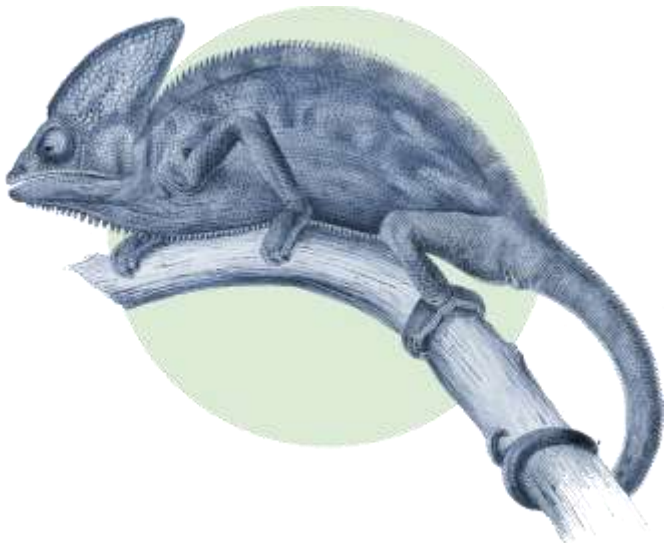
Becoming a mentor

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)

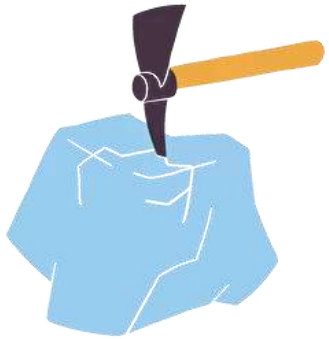
Thank you!

Melissa Liu | GBIF Asia Support Team

mliu@gbif.org

Menti for the attendants

(from 16:15 to 16:30 CEST)



Go to: 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)

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 Cuilding 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

- **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

- **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

contact@biodiversa.eu

BiodiversaPlus