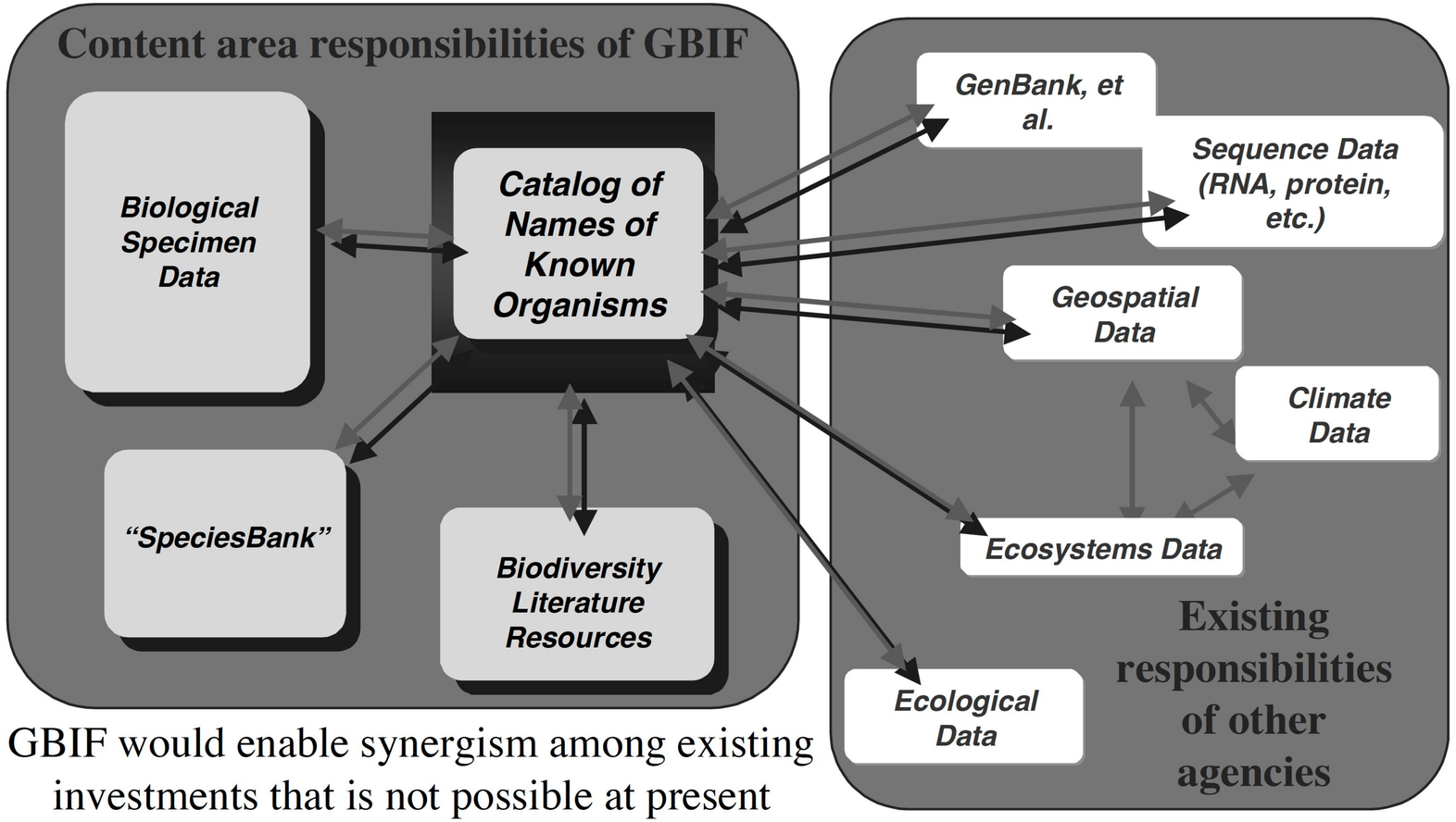
The Global Biodiversity Information Facility (GBIF)

Prof. Dr. Birgit Gemeinholzer/GBIF Science Committee Chair

GBIF Global Biodiversity Information Facility

1999 Recommendation of the OECD Biodiversity Working Group



GBIF would enable synergism among existing investments that is not possible at present

https://web-archive.oecd.org/2012-06-15/164761-2105199.pdf

Vision

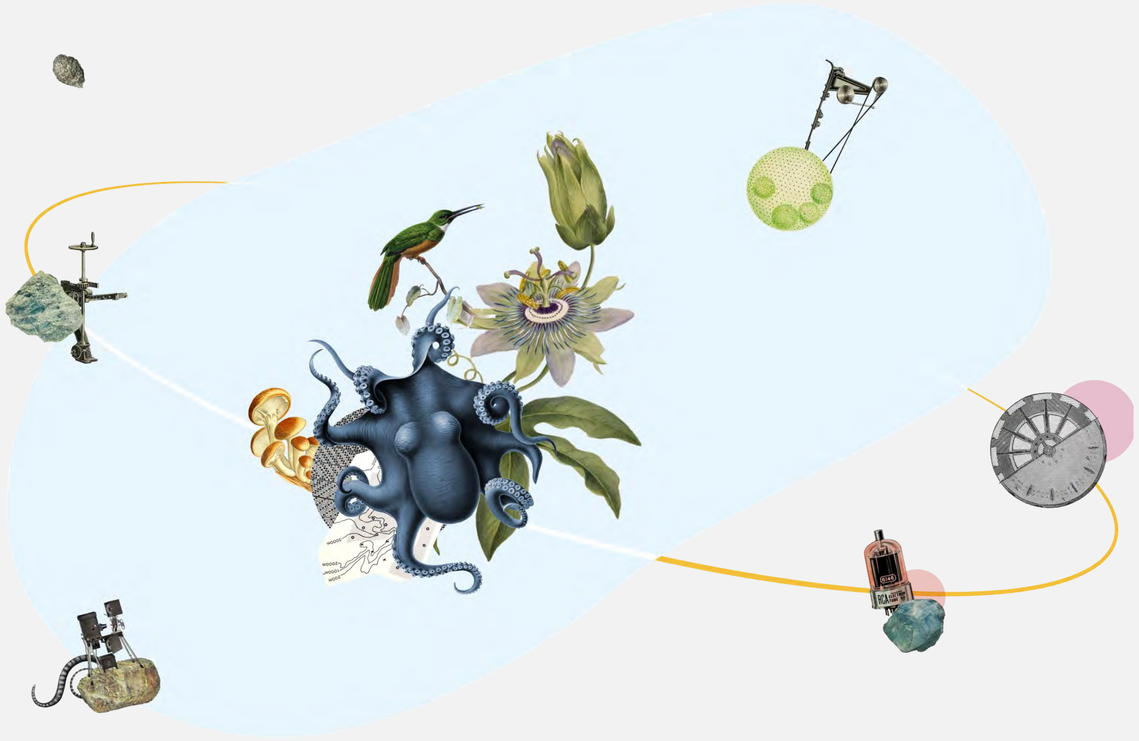
A world in which the best possible biodiversity data underpins research, policy and decisions.

Mission

To mobilize the data, skills and technologies needed to make comprehensive biodiversity information freely available for science and decisions addressing biodiversity loss and sustainable development

Shared infrastructure

By embracing hosted and cloud-based environments, GBIF's infrastructure expands easy access to biodiversity data, promotes collaboration, facilitates data harmonization and fosters innovative research



https://www.gbif.org/value

Datasets

103,568

Peer-review papers using data

10,288

Country Participants

63

Average records downloaded per month (2023)

173.8 billion

Organizational Participants

43

Species occurrence records

2,690,679,082

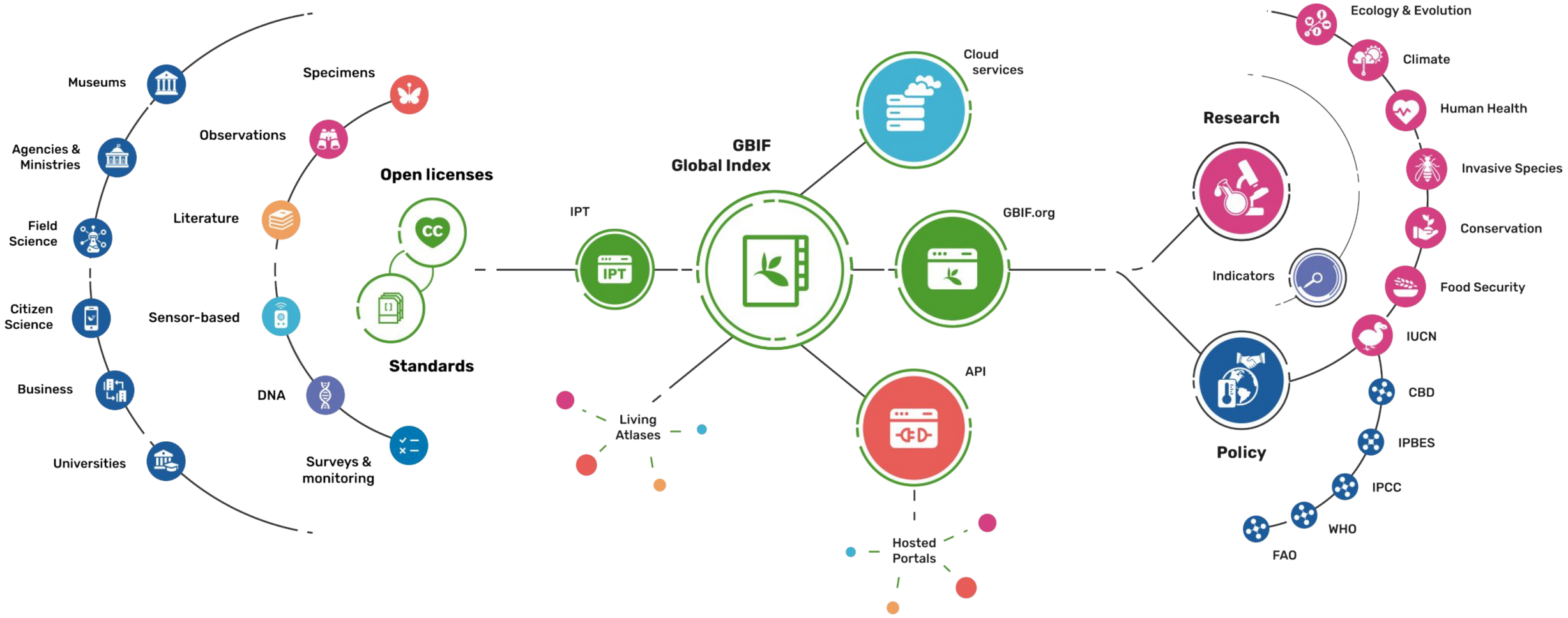
Publishers

2,202

By the numbers | 1 April 2024

www.gbif.org

Providing biodiversity evidence for research and policy



Create

Combine sources of evidence

Share

FAIR and open access

Transform

Apply and use data

Thematic communities

Fostering focused collaboration and knowledge exchange among experts in specific biodiversity domains, these communities serve as dynamic hubs where people come together to address shared challenges.

https://www.gbif.org/value

GBIF’s contribution to the Sustainable Development Goals

Flying foxes (*Pteropus* sp.) predict Nipah virus transmission risk

- Study maps the risks of Nipah infection in South and Southeast Asia using models confirmed by current outbreak in India.
- GBIF-mediated data resources used : 28,418 species occurrences
- Main risk factor is via human consumption of *Pteropus* urine or saliva-contaminated raw sap from date palm (*Phoenix dactylifera*)

Indian flying fox (*Pteropus giganteus*) by Hemant Kumar via iNaturalist. Photo licensed under CC BY-NC 4.0.

Deka M and Morshed N (2018) Mapping Disease Transmission Risk of Nipah Virus in South and Southeast Asia. Tropical Medicine and Infectious Disease. MDPI AG 3(2): 57. Available at: https://doi.org/10.3390/tropicalmed3020057.

Evaluating trends in research uses

PNAS

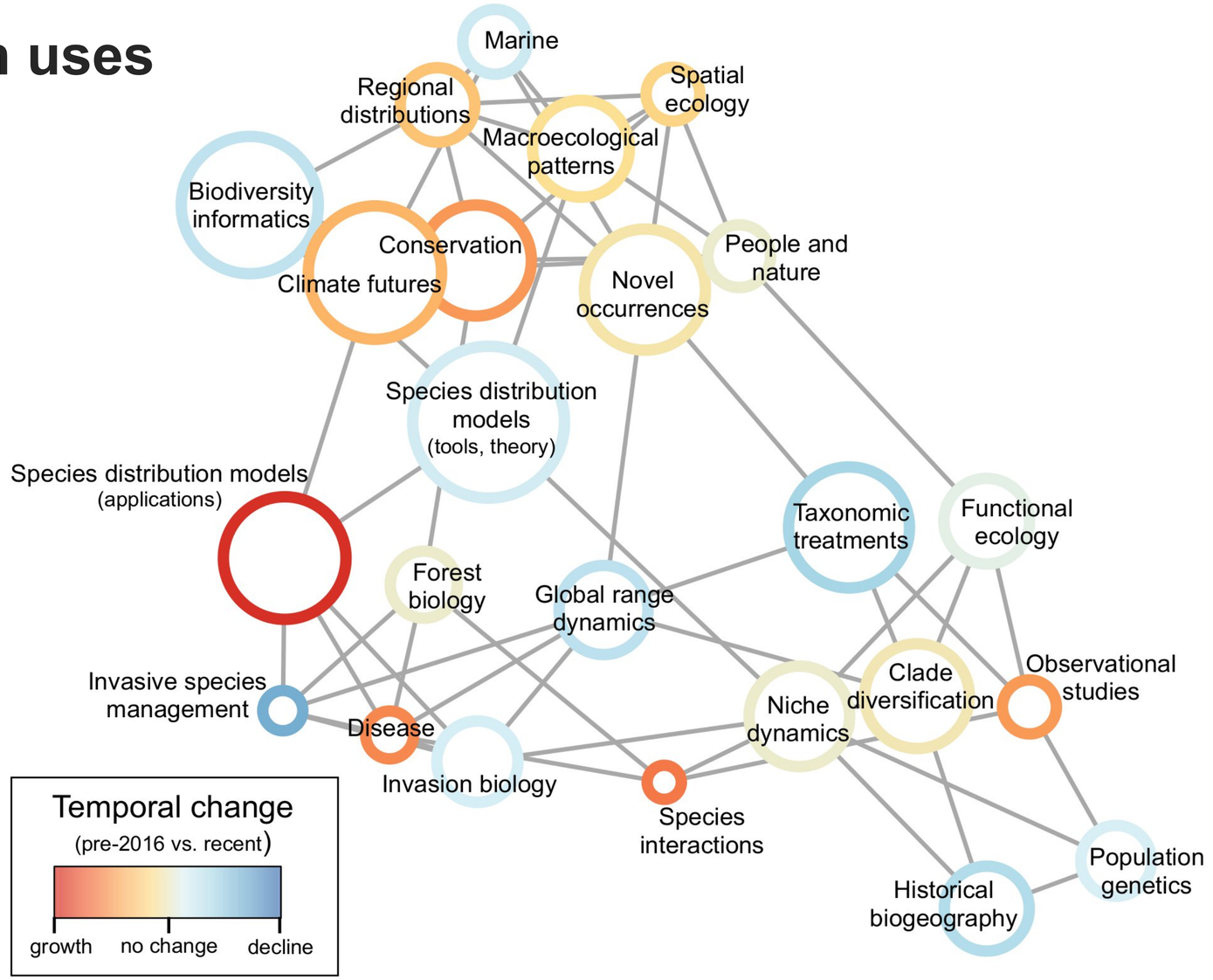
RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Data integration enables global biodiversity synthesis

J. Mason Heberling, Joseph T. Miller, Daniel Noesgaard, +1, and Dmitry Schigel Authors Info & Affiliations

Edited by Douglas E. Soltis, University of Florida, Gainesville, FL, and approved December 8, 2020 (received for review September 1, 2020)

February 1, 2021 | 118 (6) e2018093118 | https://doi.org/10.1073/pnas.2018093118



https://www.gbif.org/news/4NBqq6IfjVYLwX0u252oNM/call-for-proposals-for-a-systematic-review-of-gbif-enabled-research; https://doi.org/10.1073/pnas.2018093118

Evaluating trends in research uses

PNAS

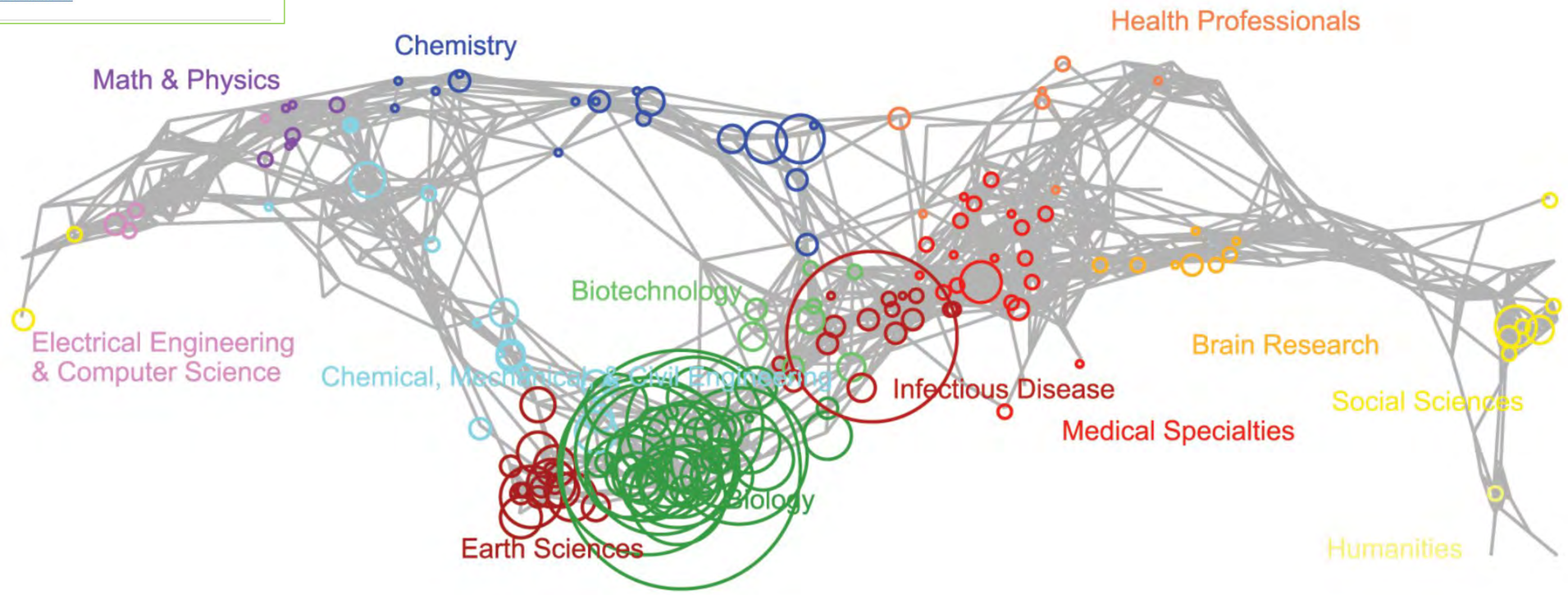
RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Data integration enables global biodiversity synthesis

J. Mason Heberling, Joseph T. Miller, Daniel Noesgaard, +1, and Dmitry Schigel Authors Info & Affiliations

Edited by Douglas E. Soltis, University of Florida, Gainesville, FL, and approved December 8, 2020 (received for review September 1, 2020)

February 1, 2021 | 118 (6) e2018093118 | https://doi.org/10.1073/pnas.2018093118



https://www.gbif.org/news/4NBqq6IfjVYLwX0u252oNM/call-for-proposals-for-a-systematic-review-of-gbif-enabled-research; https://doi.org/10.1073/pnas.2018093118

Geography of GBIF data use and authorship

PNAS

RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Data integration enables global biodiversity synthesis

J. Mason Heberling, Joseph T. Miller, Daniel Noesgaard, +1, and Dmitry Schigel Authors Info & Affiliations

Edited by Douglas E. Soltis, University of Florida, Gainesville, FL, and approved December 8, 2020 (received for review September 1, 2020)

February 1, 2021 | 118 (6) e2018093118 | https://doi.org/10.1073/pnas.2018093118

https://www.gbif.org/news/4NBqq6IfjVYLwX0u252oNM/call-for-proposals-for-a-systematic-review-of-gbif-enabled-research; https://doi.org/10.1073/pnas.2018093118

Capacity enhancement

By focusing on people, GBIF recognises that the success of data sharing and conservation efforts relies on individuals' skills, knowledge, and engagement at various levels.

- providing documentation and learning resources
- capacity enhancement mentoring
- biodiversity open data ambassadors
- regional knowledge networks and a global community of practice

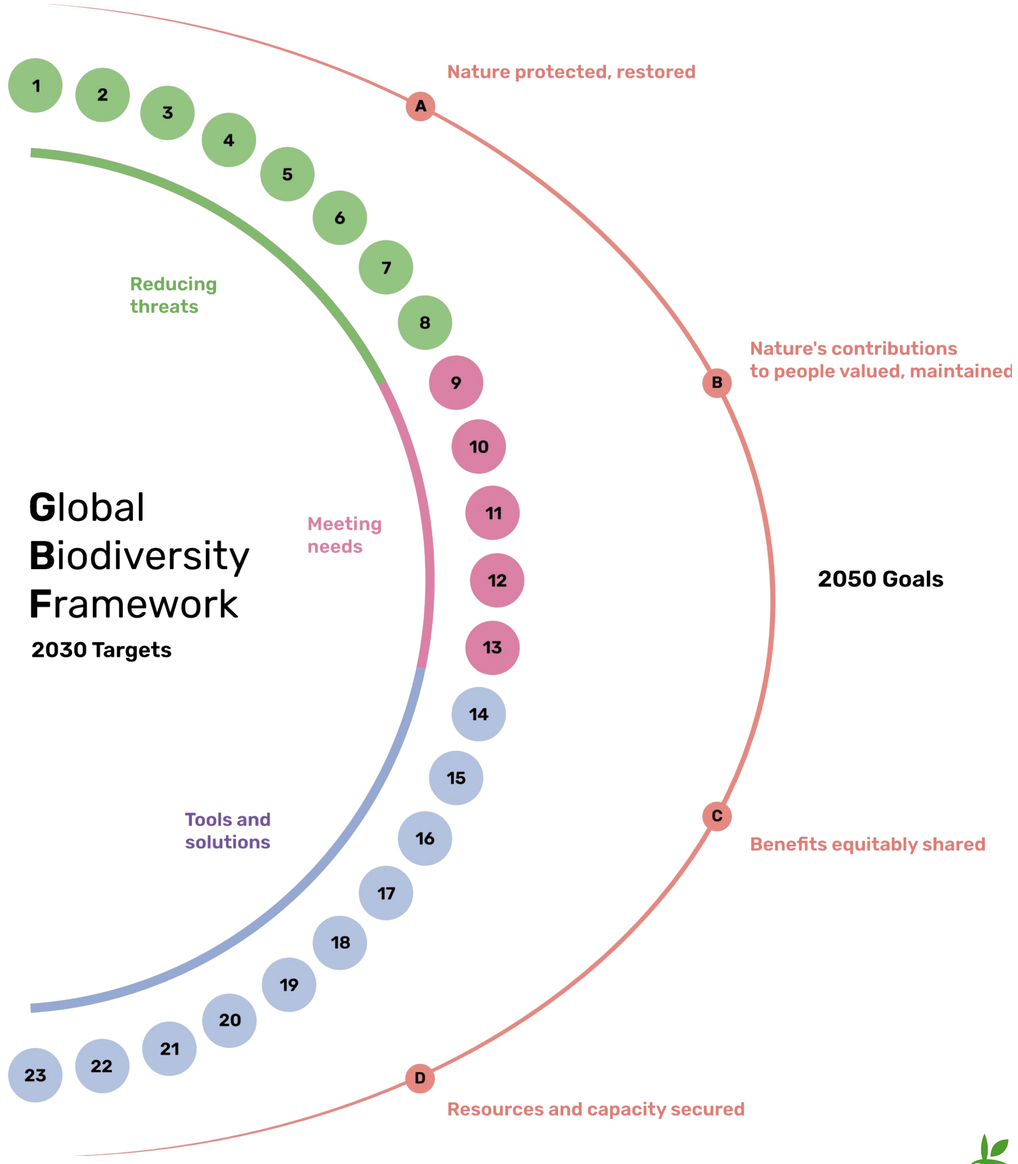
Research and policy

GBIF empowers its community of users to unlock new insights, enabling ground-breaking scientific studies and facilitating evidence-based policy decision-making.

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

GBIF and the Kunming-Montreal global biodiversity framework

GBIF relevance (illustrative)	
Data to identify key biodiversity areas	1
Data to monitor restoration	2
Data to locate, monitor protected areas	3
Data for species conservation	4
Data on invasive species occurrence	6
Data to model climate change impacts	8
Platform for sharing EIA data	15
Capacity programmes for data mobilization and use	20
Making data available for implementation	21



https://www.cbd.int/gbf/

Biodiversity Information for Development (BID)

+1,3 million

Species occurrence records

Data used in

1,141

Peer-reviewed publications*

BID

▶ *As of 31 May 2022*

103

Professionals trained

BID | Biodiversity Information for Development

Funded by the European Union

102

Projects

** As of 21 Feb 2024*

A multi-year programme funded by the European Union led by GBIF with the aim of enhancing capacity for effective mobilization and use of biodiversity data in research and policy

BID 2

Funded by
the European Union

global environment facility
INVESTING IN OUR PLANET

A multilateral family of funds dedicated to confronting biodiversity loss, climate change, and pollution, supporting land and ocean health. The partnership includes 186 member governments as well as civil society, Indigenous Peoples, women, and youth, with a focus on integration and inclusivity.

Economic dividend from GBIF-mediated data

For every €1 invested in GBIF there are €3 of benefits to users and up to €12 to society

https://www.gbif.org/value

GBIF is “the most comprehensive, openly available, application-agnostic (most unbiased), easiest-to-use, and modern access point to known digital species occurrence data”

CoDATA *Twenty-Year Review of GBIF*

https://www.gbif.org/value

Thank you!

Prof. Dr. Birgit Gemeinholzer - GBIF Science Committee Chair

Open Science and Open Data

main principles and concepts

RESEARCH INSTITUTE
NATURE AND FOREST

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

Data & Infrastructure

Open data and Open Science

INBO aims to work as much as possible according to the methods of open and reproducible research - Open Science for short - in order to achieve solid and demonstrable scientific quality.

Open Science Goals as a guide

To concretise Open Science for INBO, we have drawn up Open Science Goals (OSDs) for projects. The Open Science Goals concretise for each phase of the research cycle how we want to apply Open Science: from design, through data collection, storage, management and analysis to publication, archiving and encouraging reuse.

Open Science Goals as a long-term perspective

The Open Science Goals offer an ideal picture for research at INBO. They came about through a participatory process within the organisation and paint a picture of where we want to go in the long term. We focus on incremental learning and growth to gradually apply these goals more broadly.

Open Science bij het INBO: naar een kwaliteitsvolle en transparante onderzoekscyclus

Een overzicht van de INBO Open Science Doelen voor projecten - versie 1.1

Floris Vanderhaeghe, Aaike De Wever, Lien Reyserhove, Peter Desmet, Thierry Onkelinx, Bart Goossens, Saskia Wanner, Gerrit Genouw, Lymke Janssens, Francis Turkelboom, Toon Van Daele

INSTITUUT
NATUUR- EN BOSONDERZOEK

English abstract

Open Science is increasingly preferred in scientific research. It results in high-quality and transparent research, paving the way to increased trustworthiness. Furthermore, Open Science provides the best potential for collaboration and a potentially faster application of the research results. These properties make it the most suitable research approach for public authorities.

By 'Open Science' we mean methods of **open and reproducible research**. At the Research Institute for Nature and Forest (INBO), the Open Science approach historically received much attention already. Through the INBO 'Open Science programme' it now gets even more **aligned with the whole research cycle**.

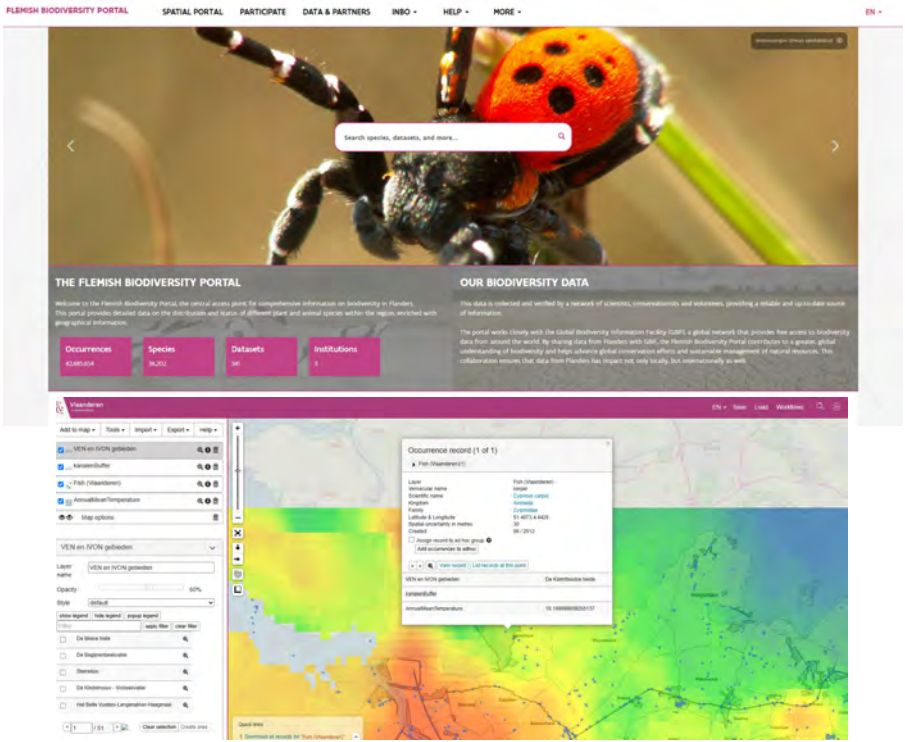
With this report we want to reach out to scientists within and outside INBO and:

- offer tangible leads to engage with Open Science methods in every research step;
- inspire to invest more in research quality and transparency.

More specifically, we list the **INBO Open Science Goals for projects**: in paragraph 2.2 and elaborated in appendix B. They are the idealized picture of how we want to proceed in every stage of the research cycle. The goals have arisen as a **bottom-up** process throughout the organization. We want to create a welcoming and stimulating environment, and establish gradual and long-lasting progress. We encourage everyone to learn by trial and error, and to disseminate knowledge and skills to others. In doing so, not all goals will be implemented at the same speed or extent: they rather serve as a roadmap.

At a higher level the **Flemish Open Science Board (FOSB)** encourages Flemish research organizations to increasingly apply Open Science. The FOSB appointed **data stewards** to accelerate the Open Science ideas within the Flemish Scientific Institutes, including INBO. At INBO the data steward helps in coordinating the Open Science Goals. The FOSB has put forward five objectives (treated as KPIs): ORCID iDs, Data Management Plans, FAIR data, open access and open data.

Flemish Living Atlas project:
Vlaams Biodiversiteitsportaal



Belgian Biodiversity Platform:
Belgian GBIF Node

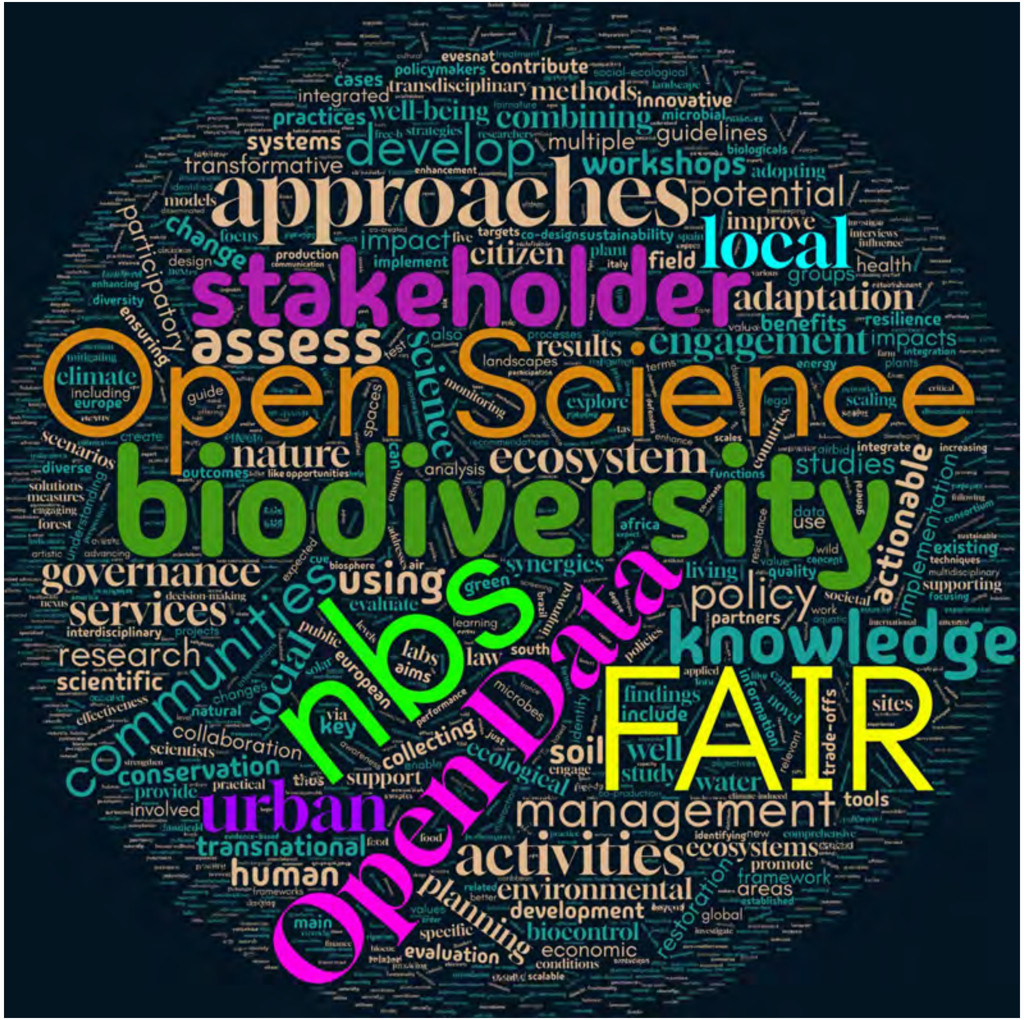
Biodiversity.be

This call, co-funded by the European Commission, focused on three (non-exclusive) key themes:

1. **THEME A:** Synergies and trade-offs of Nature-based Solutions in the context of human well-being
2. **THEME B:** Nature-based Solutions mitigating anthropogenic drivers of biodiversity loss
3. **THEME C:** The contribution of Nature-based Solutions for just transformative change

Open Science?


biodiversity
nbs
stakeholder
local
urban
approaches
communities
knowledge
management
develop
assess
activities
policy
governance
engagement
adaptation
human
science
transnational
environmental
combining
services
studies
workshops
research
ecosystem
using
change
climate
social
potential
nature
actionable
conservation
transdisciplinary
implementation
planning
guidelines
citizen
impact
scientific
european
practices
improve
integrated
restoration
evaluation
outcomes
economic
well-being
framework
biocontrol
sustainability
participatory
development
synergies
resilience
collecting
field
impacts
results
scenarios
carbon
water
methods
findings
multiple
collaboration
benefits
countries
supporting
ensuring
tools
diverse
scientists
labs
key
support
projects
explore
groups
energy
production
design
soil
analysis
include
adopting
ecological
public
partners
well
understanding
contribute
cases
general
influence
implement
enhance
effectiveness
values
systems
law
consortium
mitigating
resistance
opportunities
forest
legal
targets
engage
value
aims
trade-offs
involved
terms
awareness
interactions
sites
plants
living
promote
spaces
specific
south
microbes
innovative
decision-making
focus
models
test
integrate
risks
five
evaluate
microbial
policymakers
monitoring
health
diversity
transformative
ecosystems
strategies
provide
study
green
relevant
co-design
co-production
identify
novel
biologicals
landscapes
areas
global
existing
thus
interdisciplinary
measures
established
including
identified
advancing
interviews
clean


biodiversity
Open Science
stakeholder
approaches
FAIR
nbs
OpenData
knowledge
local
urban
management
activities
ecosystem
policy
assess
engagement
communities
governance
services
nature
develop
potential
adaptation
citizen
impact
science
results
research
scientific
social
using
actionable
studies
planning
environmental
development
human
transnational
conservation
collaboration
soil
water
well
study
include
findings
trade-offs
tools
sites
biocontrol
economic
evaluation
workshops
combining
methods
practices
systems
transformative
guidelines
multiple
well-being
integrated
transdisciplinary
cases
contribute
policymakers
innovative
participatory
climate
scenarios
diverse
solutions
measures
forest
change
models
focus
design
production
implement
improve
sustainability
co-design
field
health
groups
benefits
resilience
impacts
landscapes
explore
scaling
ecological
collecting
labs
law
south
green
synergies
evaluate
living
existing
supporting
partners
restoration
framework
areas
promote
ecosystems
main
specific
support
provide
practical
involved
scientists
natural
key
european
aims
public
monitoring
spaces
analysis
can
outcomes
opportunities
europe
including
guide
ensuring
diversity
interdisciplinary
decision-making
projects
changes
effectiveness
data
use
africa
quality
concept
value
wild
implementation
countries
functions
integrate
increasing
learning
microbes
air
biosphere
improved
policies
information
carbon
novel
international
applied
global
identifying
new
comprehensive
food
evesnat
microbial
biologicals
adopting
targets
production
italy
plant
energy
critical
also
test


biodiversity
data
monitoring
conservation
species
management
stakeholders
diversity
objectives
project
methods
change
context
indicators
europe
european
ecosystem
ecosystems
activities
knowledge
environmental
policy
provide
develop
research
changes
forest
global
national
local
automated
habitats
human
tools
marine
approaches
including
aquatic
genetic
functional
existing
citizen
ecological
potential
taxonomic
fungi
parasites
mediterranean
habitat
forests
novel
long-term
distribution
trends
key
available
framework
network
improve
areas
standardised
scientists
status
subterranean
assess
identify
spatial
transnational
assessment
international
soil
remote
sampling
impact
land-use
identification
integration
intelligence
support
climate
dna
edna
sensing
science
restoration
future
gradients
current
analysis
functioning
population
integrating
scientific
protocols
public
groups
results
freshwater
artificial
pollinators
developing
wildlife
research
metabarcoding
relevant
scales
levels
abundance
community
effective
variables
engagement
ecology
plant
multiple
communities
sea
essential
facilitate
practitioners
innovative
natural
harmonise
collaboration
acoustic
major
biological
peatlands
fungal
habitats
products
techniques
new
gaps
based
practices
trophic
combining


biodiversity
FAIR
Open Source
Open science
monitoring
conservation
stakeholders
data management
diversity
open data
standards
species
european
ecosystem
context
indicators
methods
knowledge
objectives
develop tools
environmental
changes
provide
tools
national
change
policy
automated
ecosystems
europe
existing
research
approaches
science
forest
including
support
citizen
distribution
genetic
approach
global
aquatic
habitats
network
marine
long-term
climate
areas
international
identification
public
ecological
local
trends
gaps
essential
integration
quality
multiple
functional
framework
impact
soil
remote
drivers
novel
dna
based
fungi
plant
mediterranean
parasites
pollinators
restoration
integrating
stakeholder
artificial
microbial
population
sampling
technology
human
taxonomic
forests
current
key
fungal
metabarcoding
sensing
edna
wildlife
subterranean
freshwater
habitat
improve
developed
groups
practices
important
spatial
protected
gradients
storm
scientists
interactions
developing
available
contribute
temporal
natural
countries
analysis
Open access
process
populations
future
identify
acoustic
intelligence
peatlands
pelagic
scientific
programmes
assessments
variables
efforts
relevant
combining
practitioners
observation
engagement
initiatives
collaboration
new
biological
modelling
standardised
providing
however
trophic
role
land-use
transnational
assessment
integrated
monitoring
large
among
results
regional
ecology
applied
rocky
studied
scale
needed
ponds
accessible
stressors
high
collection
used
plants
due
assess
food
facilitate
protocols
abundance
threats
challenges
design
sharing
impacts
nature
increasing
guidelines
authorities
projects


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


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

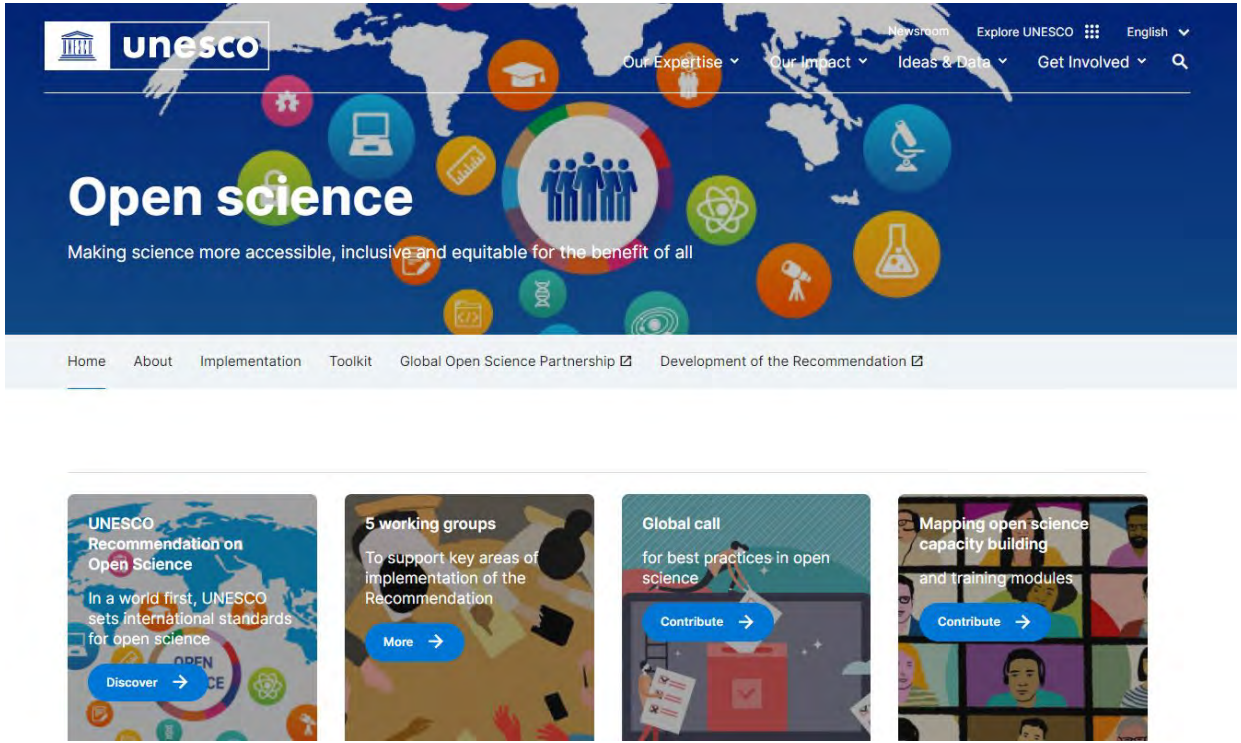

open
science
people
new
community
knowledge
tools
way
make
scientific
ideas
world
discovery
open-source
research
better
access
just
forward
everyone
accessible
sharing
progress
truth
problems
environment
movement
find
take
one
now


Open
Access
journals
journal
publishers
agreements
self-archiving
publishing
articles
authors
article
publication
rights
based
various
freely
DiVA


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

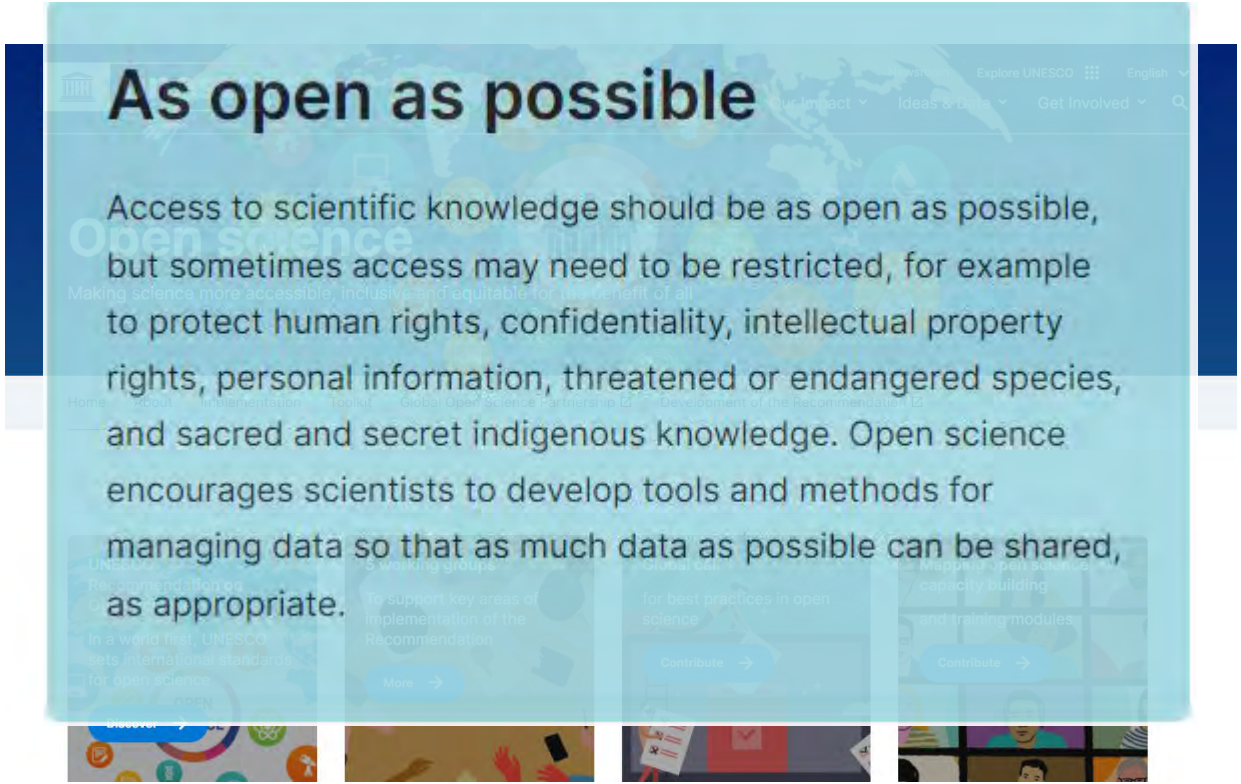
UNESCO Recommendation on Open Science

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.

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

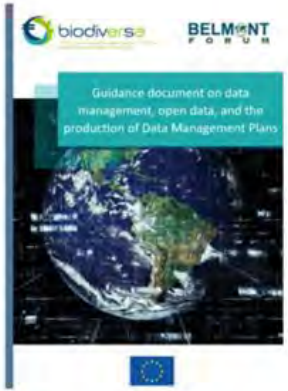
9 ambitions of the EU's open science policy

- 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

Biodiversa+ and Open Science

1. **Promotes Accessibility and Sharing of Data**
2. **Encourages Collaborative Research**
3. **Transparency in Research Processes**
4. **Public Engagement and Citizen Science**
5. **Funding and Policy Support**
6. **Utilization of Open Access Journals and Repositories**

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

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

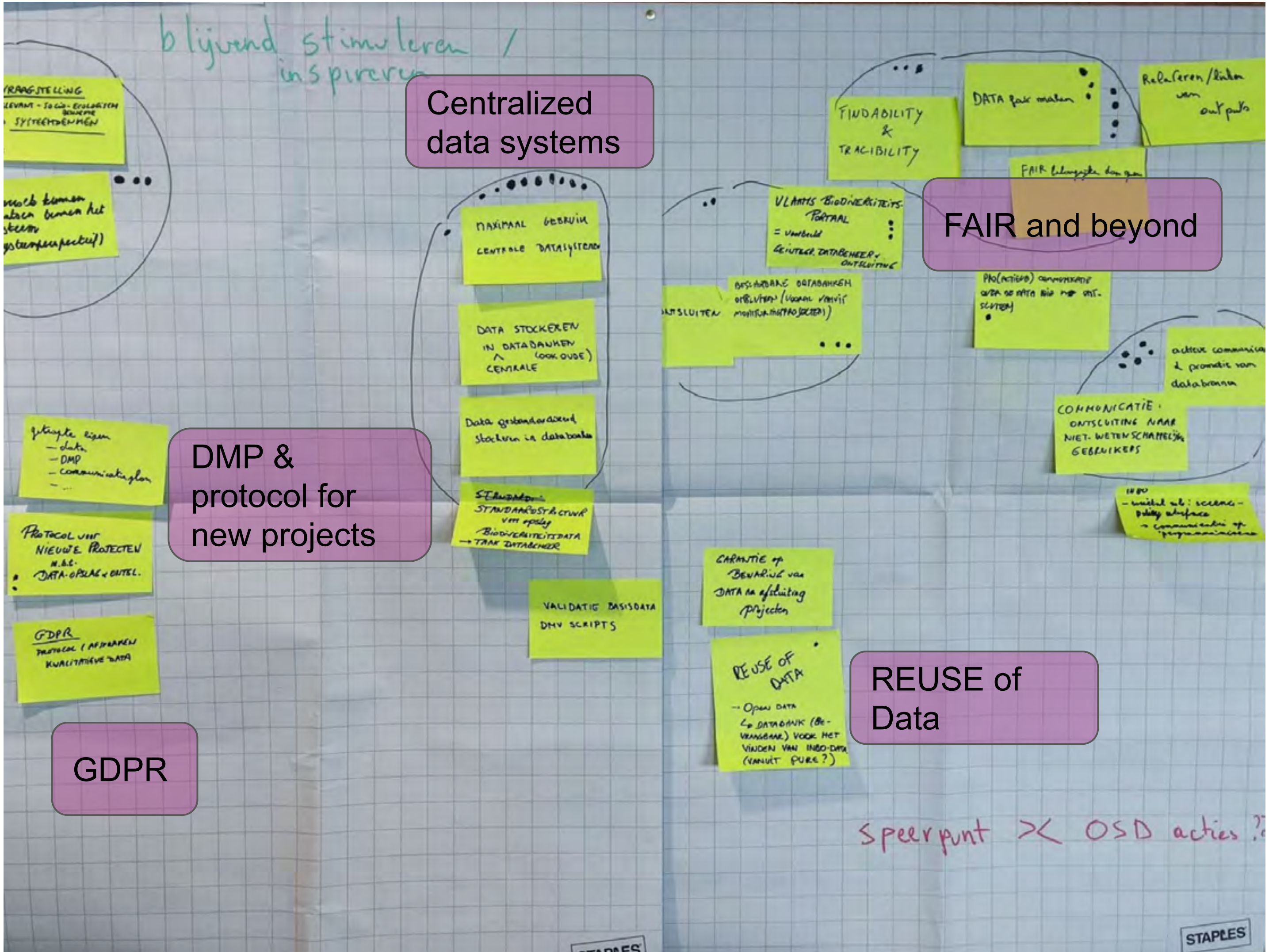
DOI:
DOI 10.5281/zenodo.7672934

Keyword(s):
open science

CERN Open Science

FRDN.

eosc
European Open Science Cloud


blijvend stimuleren / inspireren
Centralized data systems
FAIR and beyond
DMP & protocol for new projects
GDPR
REUSE of Data
FINDABILITY & TRACIBILITY
DATA STOCKEREN IN DATABANKEN (OOK OUDE) CENTRALE
VALIDATIE BASISDATA DMV SCRIPTS
REUSE OF DATA
COMMUNICATIE ONTSLUITING NAAR NIET-WETENSCHAPPELIJKE GEBRUIKERS
Speerpunt >< OSD acties ?

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

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

Why open science?

Good for Research

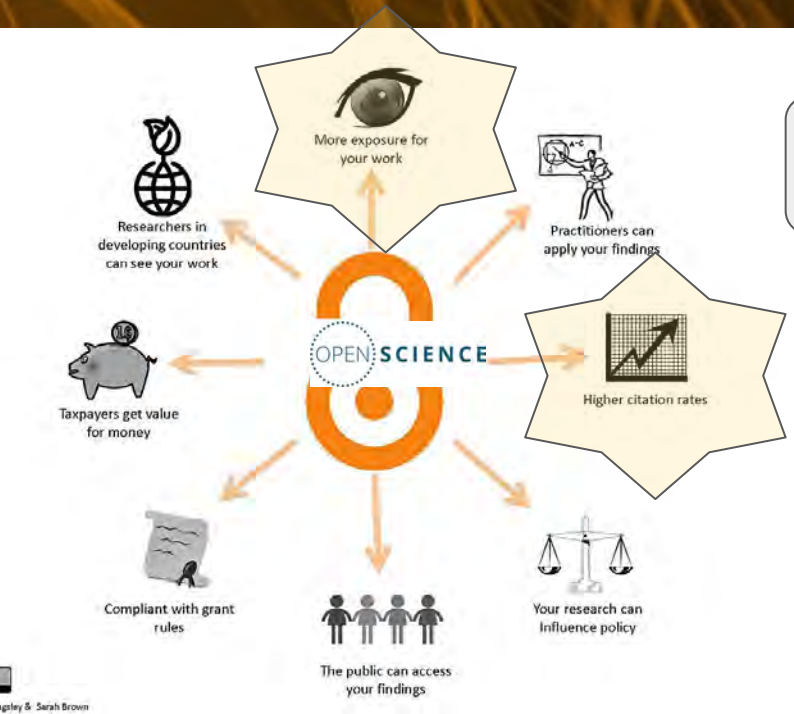
Practicing open science means that research **outputs are accessible to all** – not stuck behind pay walls. This helps to ensure that all researchers, and other stakeholders, have access to information regardless of their location or economic situation. It means that the **research process can be accelerated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

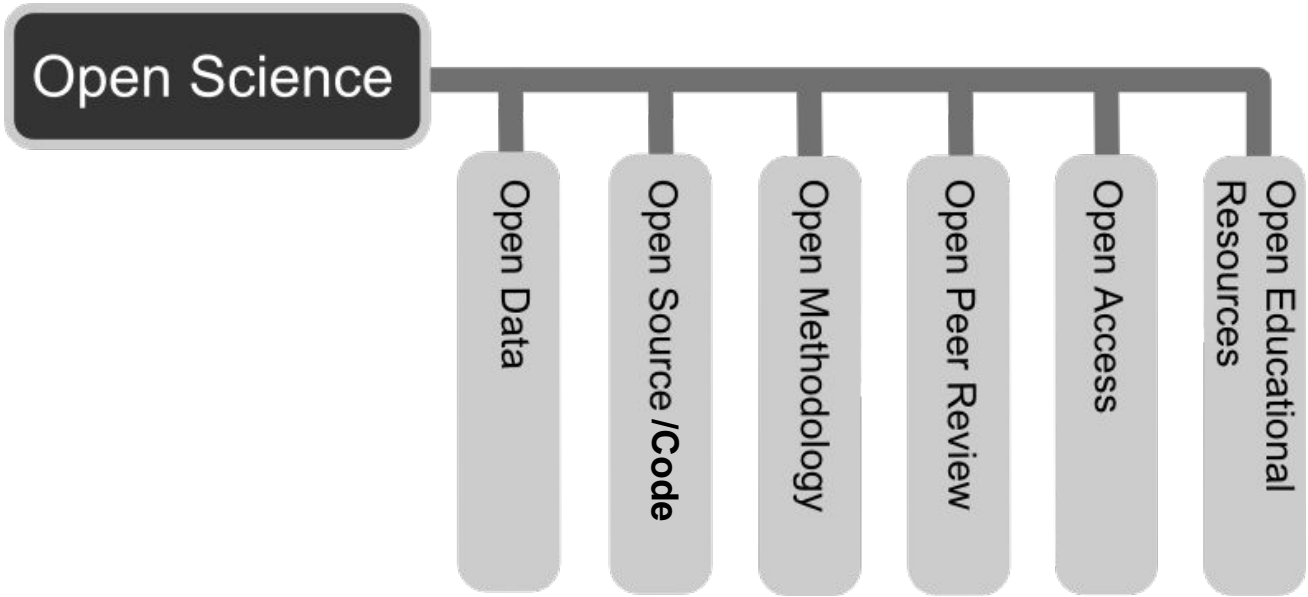
https://www.enago.com/academy/open-access-publications-get-cited-more-often/
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040157

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 .

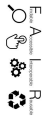
Open Science - The pillars of Open Science



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
NCBI

MOVEBANK
for animal tracking data

OBIS
OCEAN BIODIVERSITY INFORMATION SYSTEM

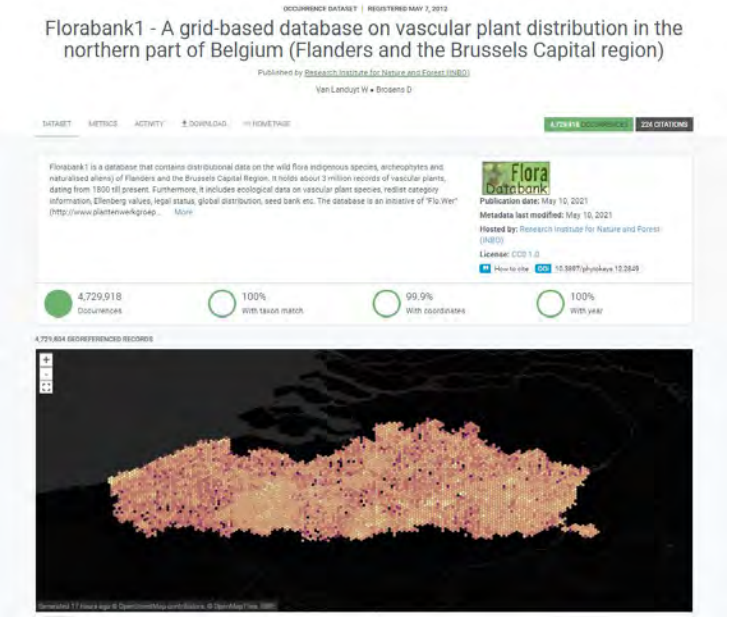
GEO GROUP ON EARTH OBSERVATIONS

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

Studies

View | Download | Env-DATA

Study Details

Study Name	White Stork Denmark
Contact Person	JassFred (Jess Frederiksen)
Principal Investigator	Anders P. Tøttrup/Kasper Thorup
Citation	
Acknowledgements	
Grants used	Storkene.dk
License Type	Custom
License Terms	*not set*
Study Summary	Migration routes of the small Danish population of white storks.
Study Reference Location	
Longitude	12.578
Latitude	55.687
Movebank ID	1578840076

Study Statistics

Number of Animals	11
Number of Tags	8
Number of Deployments	11
Time of First Deployed Location	2021-06-19 09:00:23.000
Time of Last Deployed Location	2023-05-24 10:00:19.000
Taxa	Ciconia ciconia
Number of Deployed Locations	676823
Number of Records	Deployed (outliers) / Total (outliers)
Acceleration	295462 (0) / 349655 (0)
GPS	676823 (0) / 699524 (0)

About study details

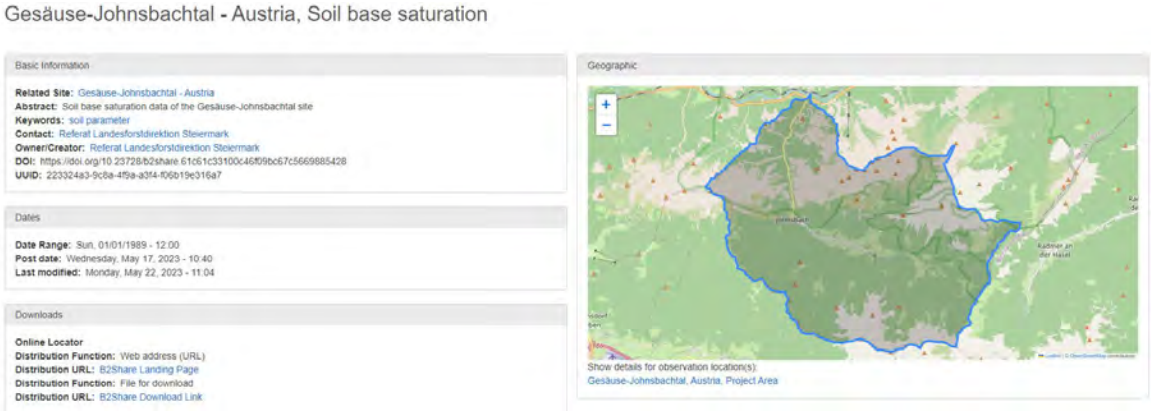
Processing Status	Up-to-date

Marine Biological Sample Database, JAMSTEC

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

report issue | source DwC-A | to mapper

Open Science - more than just open access to publications!



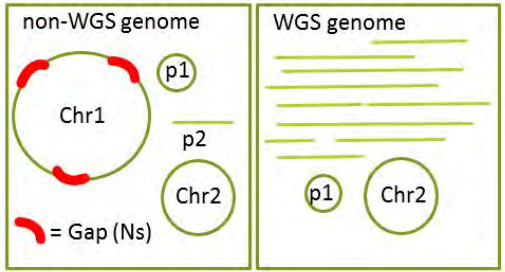
Open Science - more than just open access to publications!

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:



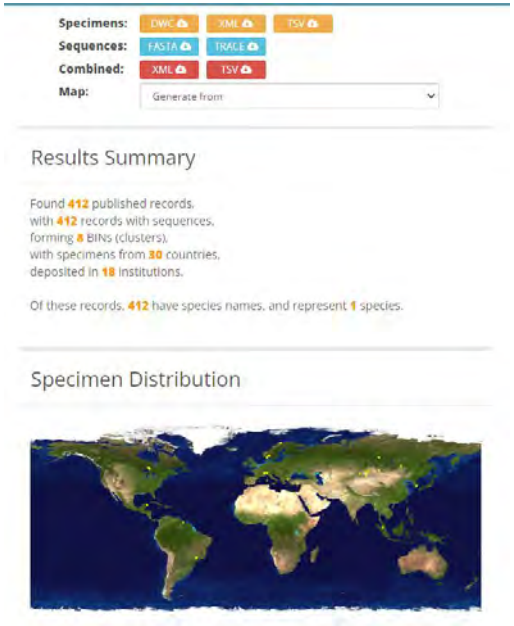
non-WGS

- 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

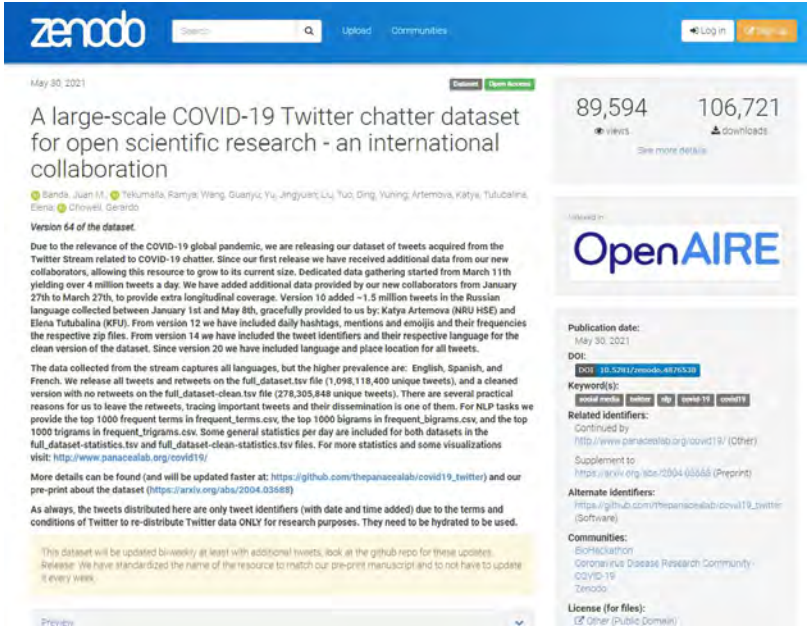
- One or more chromosomes are in multiple pieces and/or some sequences are not assembled into chromosomes

In both cases

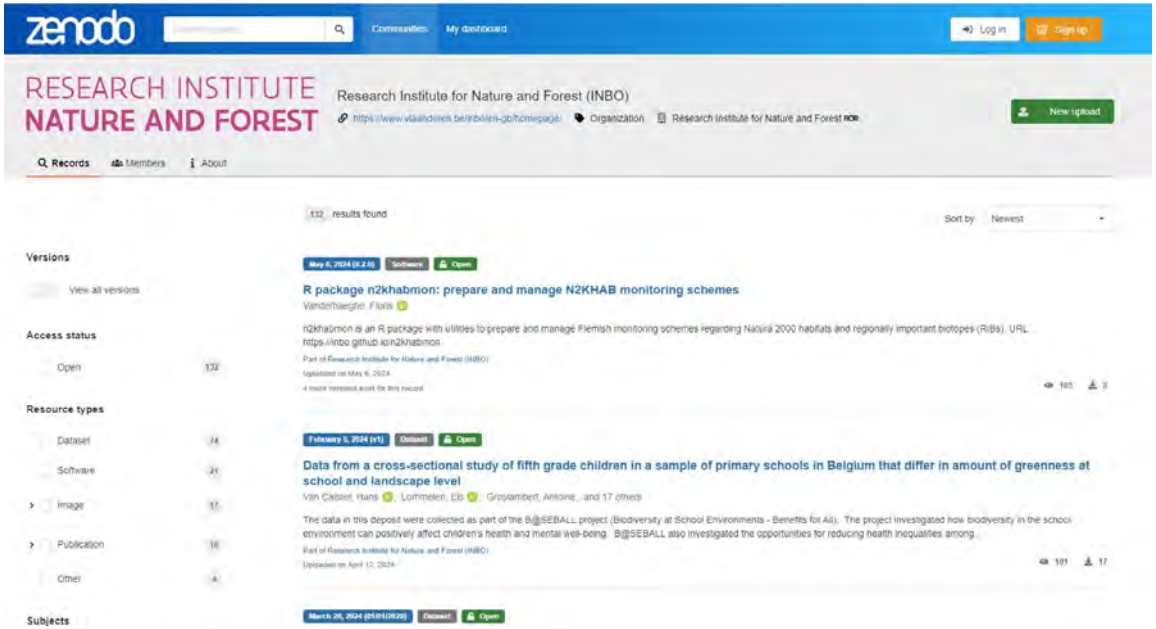


Open Science - more than just open access to publications!

zenodo



Open Science - more than just open access to publications!

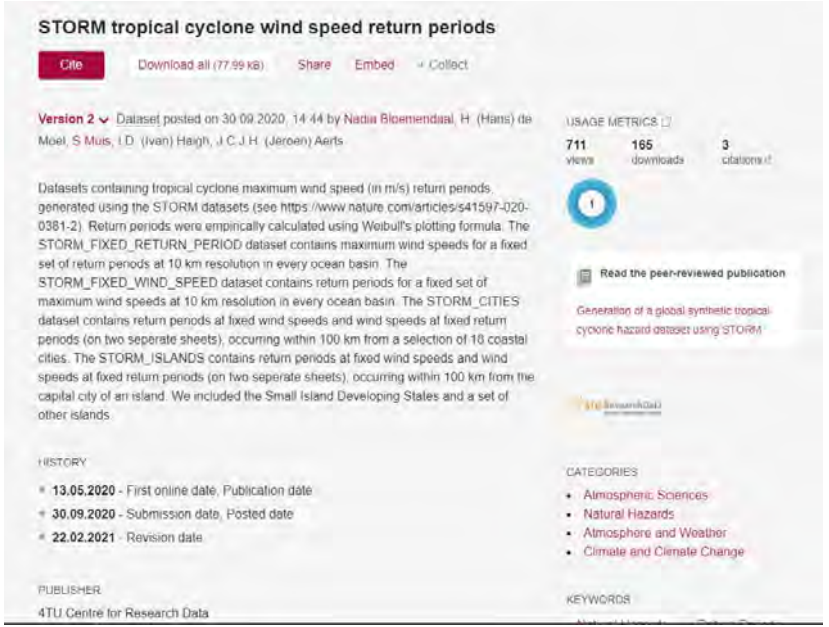


Open Science - more than just open access to publications!

DRYAD

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 10 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

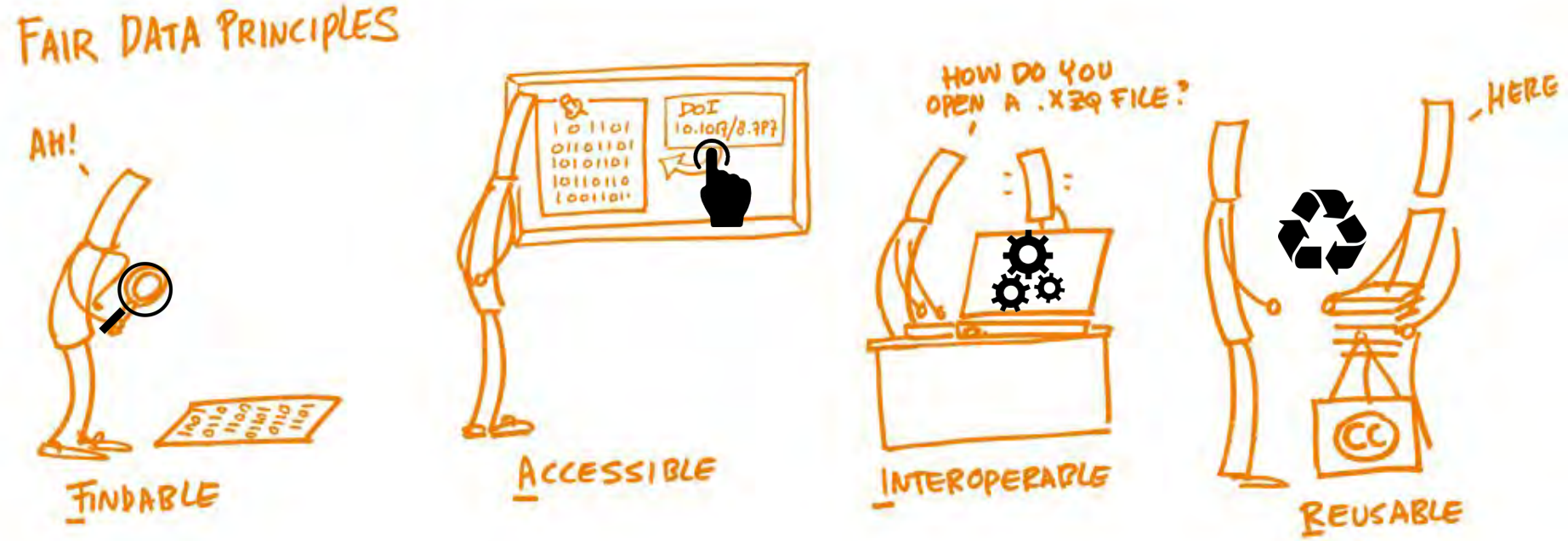
- Standards
- Vocabulary
- Methodology
- References

Reusable

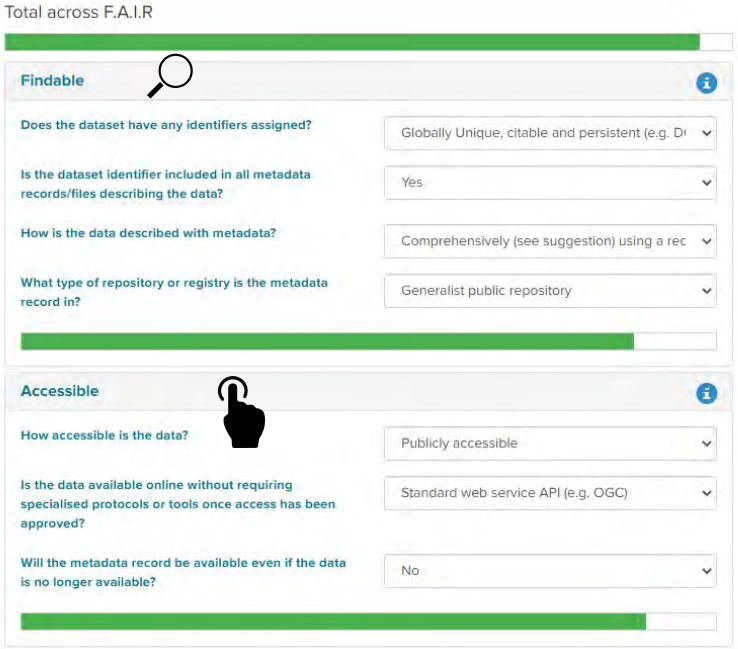
- Licensing
- Provenance
- Community standards

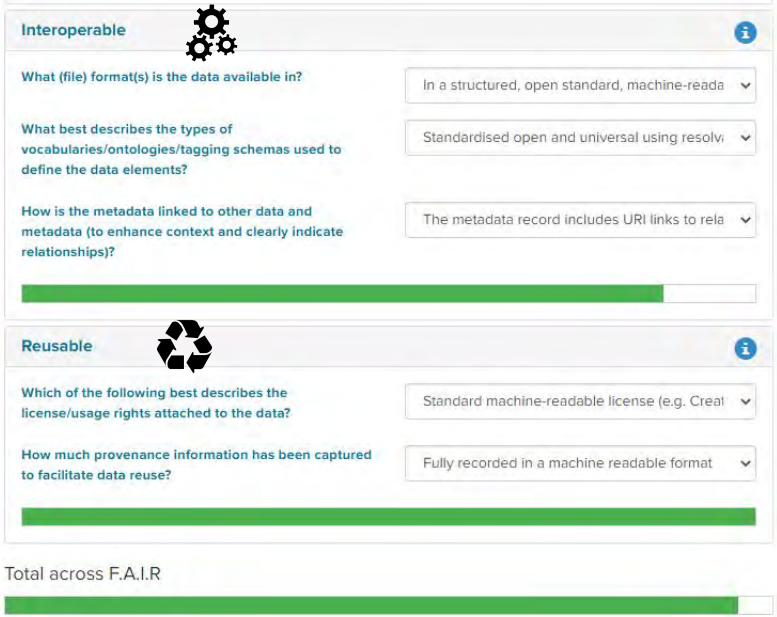
https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/

Open Science - more than just open access to publications!



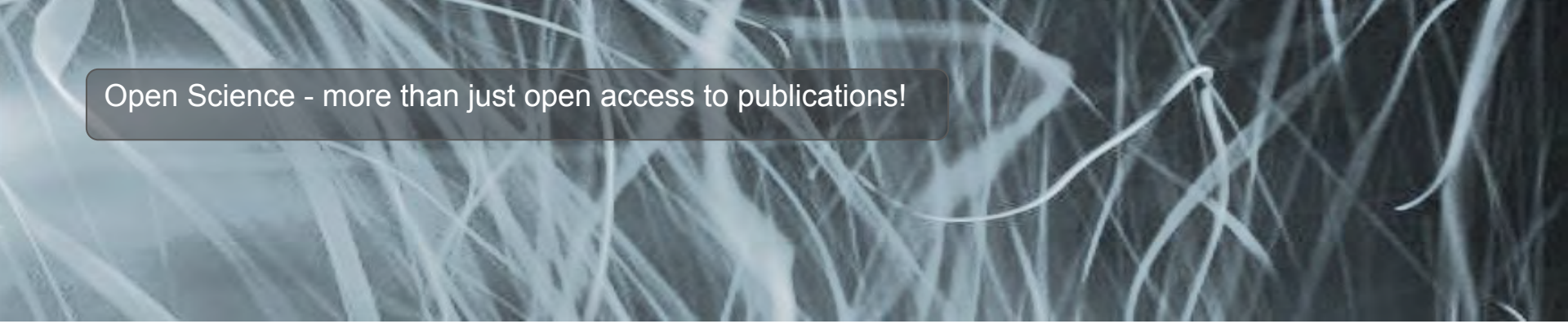
Open Science - more than just open access to publications!





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**

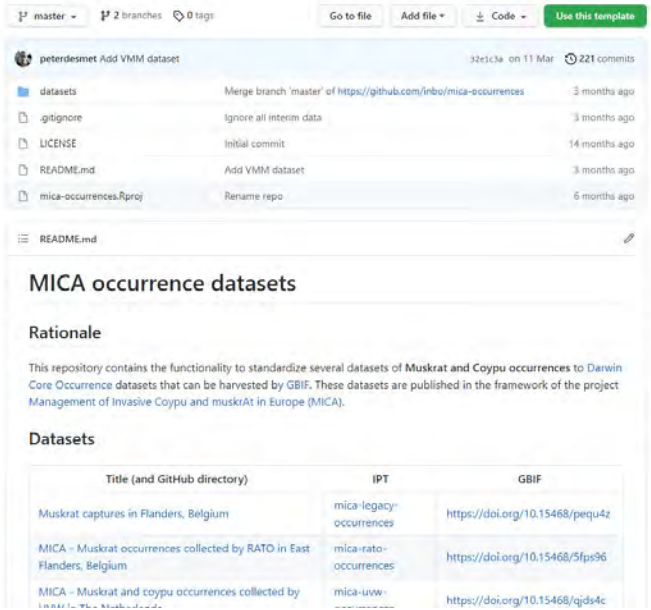
- many researchers now develop bespoke bits of code to help them analyse and/or visualise the data they have collected. Having access to this code is essential for supporting the validation of your findings and to help others to build upon your work.

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 — 32e1c3a 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-uvw-	https://doi.org/10.15468/qjds4c

About

Darwin Core mapping for the aggregated Muskrat dataset

r · dataset · gbif · occurrences

Readme

MIT License

Releases

No releases published
Create a new release

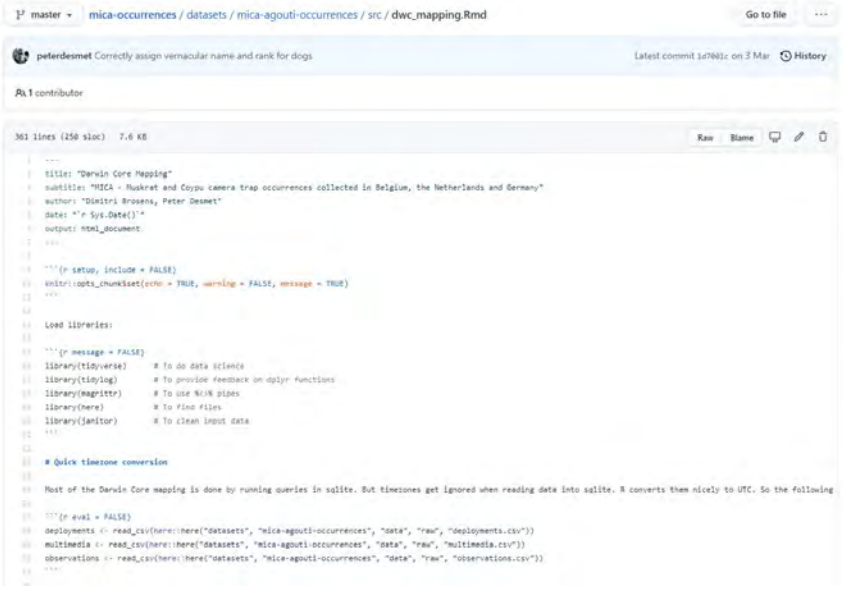
Packages

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 1d7001c 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"))
```
```

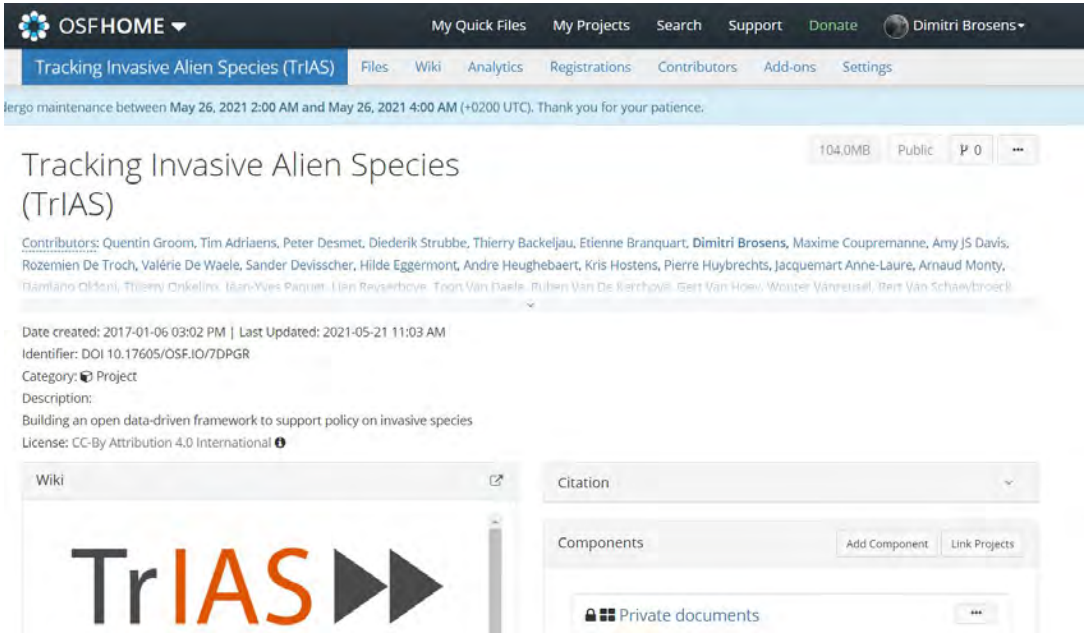
Open Science - more than just open access to publications!

…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**

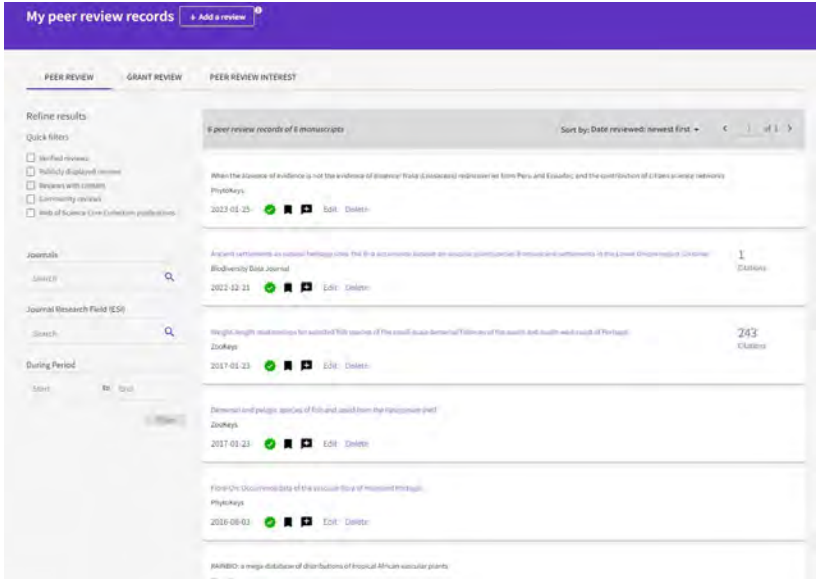
Open peer review (also called "public peer review", "transparent peer review") denotes several, closely related forms of scholarly peer review: Open-identity or attributed peer review (as opposed to anonymous peer review)

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.

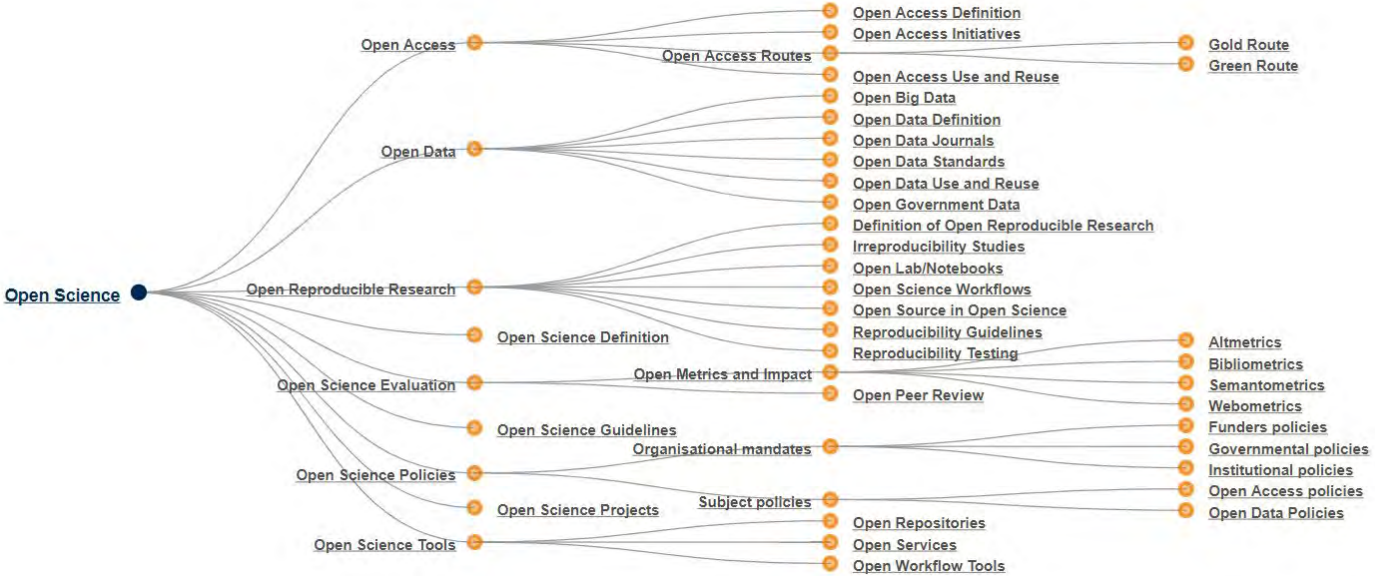


Open Science - more than just open access to publications!

…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.

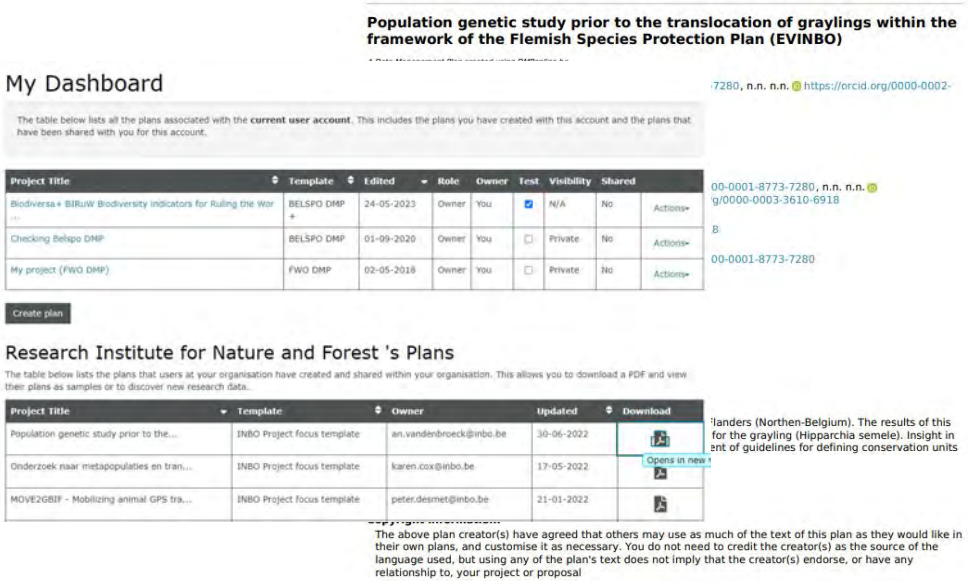
- *Make the research Question is sharp and clear*
- *Research is relevant with the research question*

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

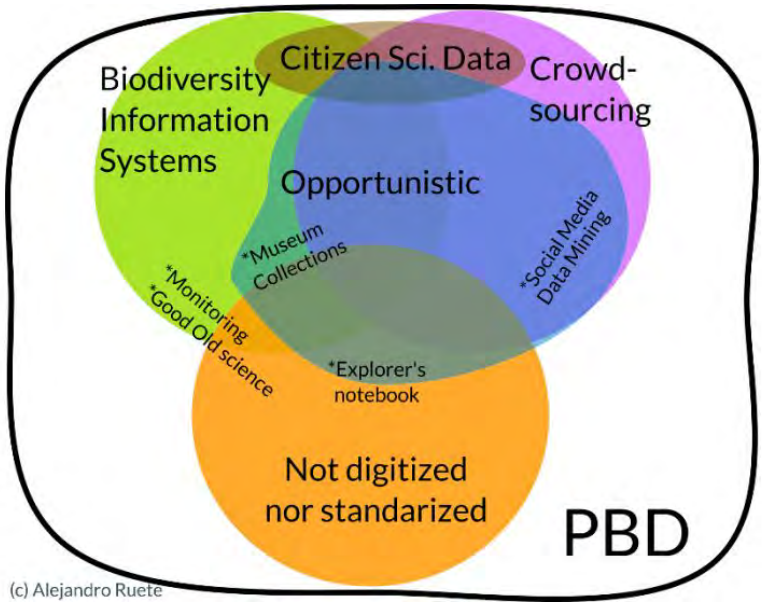
- Write a **D**ata **M**anagement **P**lan
- Create a **C**ommunication **P**lan
- Foresee revision



Open Science - Ideas for opening up!

During the active stage_ data collection

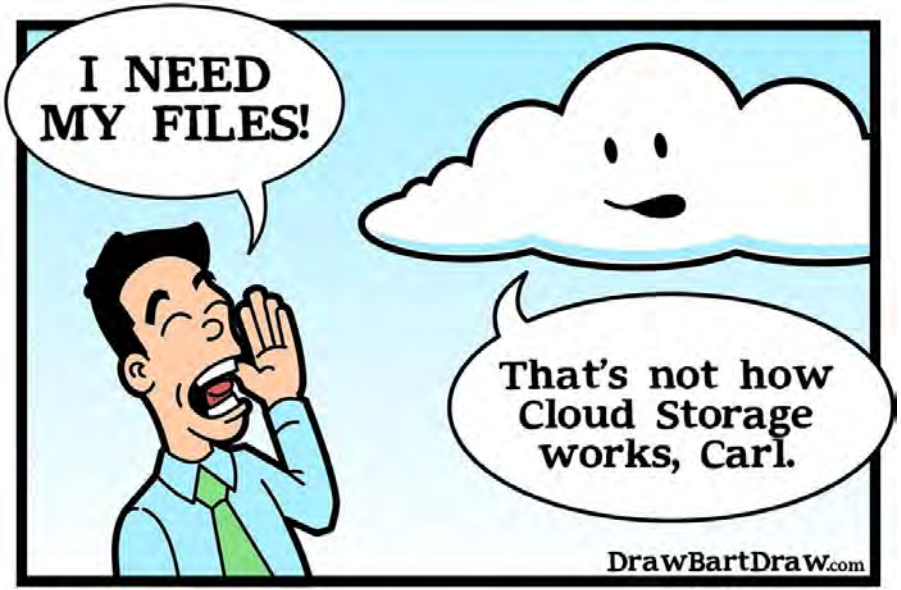
- Data collection through **well defined protocols**
- Automatic data collection when possible
- **Existing data is (re)used** when possible
- Validation procedures are in play
- Measurement biases are understood and documented



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



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!

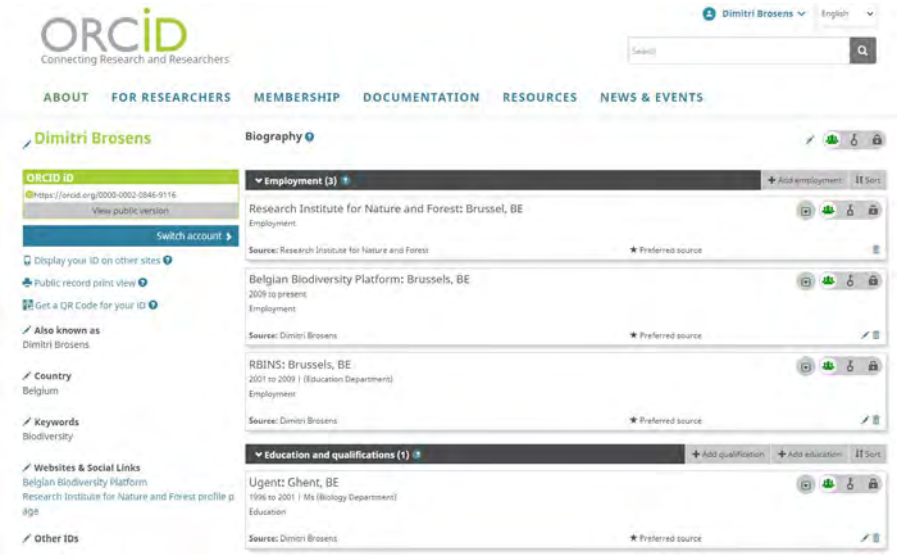
Towards the end of your research_data sharing

- 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

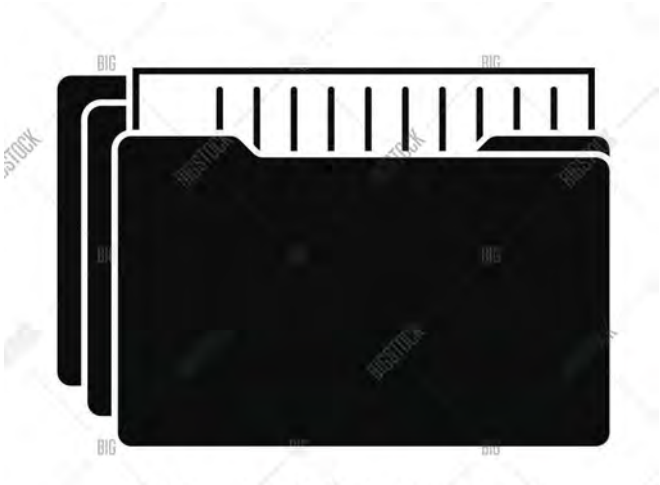
- Published as open access
- Correct citations of sources (DOI…)
- Use of OrcID's



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



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**.

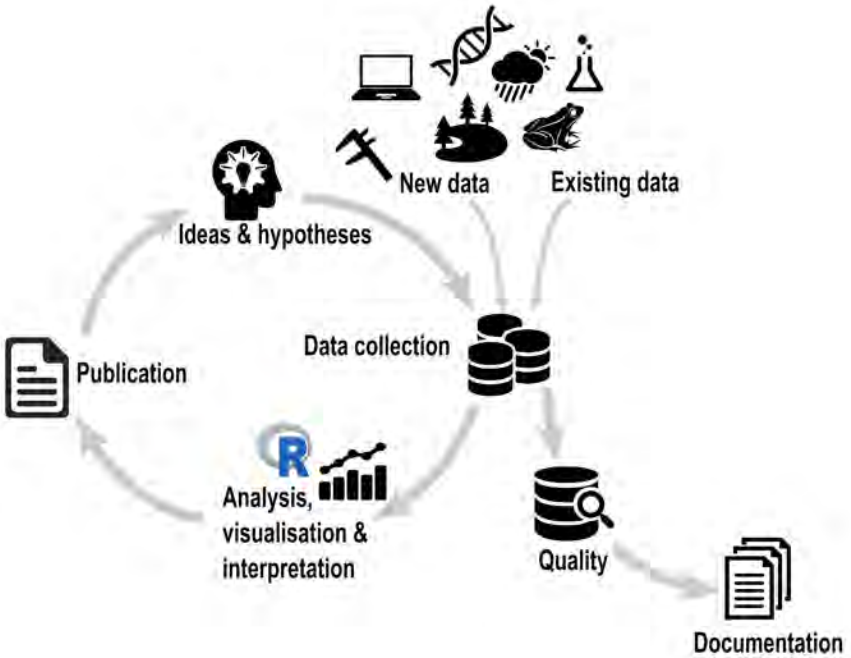
Research Life Cycle

Research Life Cycle

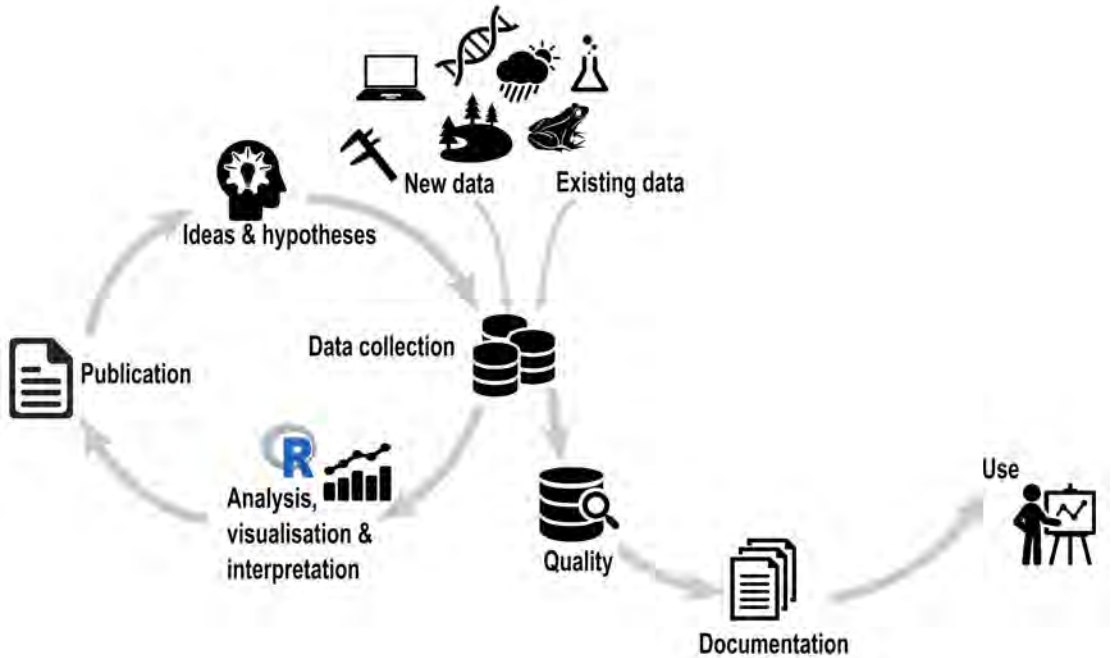
Research Life Cycle

Research Life Cycle

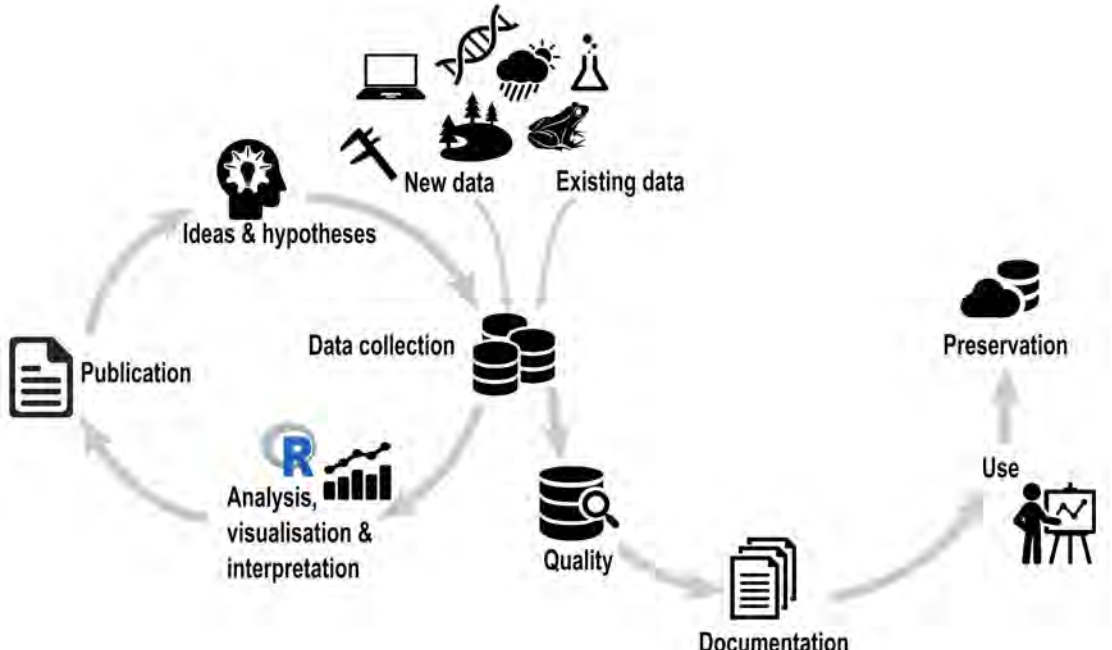
Research Life Cycle



Research Life Cycle



Research Life Cycle



Research Life Cycle

Research Life Cycle: Data Management Plan

Thank you….

Some Questions?

Data Management Plans:
Why scientists should create & sustain them

Jennifer Anderson, Swedish University of Agricultural Sciences

FUNACTION Coordinator (BiodivProtect)
MoSTFun Partner (BiodivMon)

Funaction
Aquatic fungal biodiversity

MOSTFUN
Monitoring strategies for aquatic fungi biodiversity

Funaction
Aquatic
fungal
biodiversity

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

Funaction
Aquatic
fungal
biodiversity

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

Foundation:

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

Conservation:

Translate that knowledge into conservation planning and monitoring tools.

Communication:

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

Funaction
Aquatic
fungal
biodiversity

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

Foundation:

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

Conservation:

Translate that knowledge into conservation planning and monitoring tools.

Communication:

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

WP1
Continental patterns in aquatic fungal biodiversity

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

WP3
Systematic conservation planning using aquatic fungi

WP4
Developing monitoring of aquatic fungal biodiversity and related ecosystem functions

WP5
Effective engagement, communication, and dissemination of information

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

www.funaction.eu @funaction.bsky.social FUNACTION

Jennifer Anderson (SE)
Olga Vinnere Pettersson
Ziming Wang
Isabel Fernandes (PT)
Ronaldo Sousa
Diana Graça
Andreas Bruder (CH)
Daniel Romero-Mujalli
Red Calore
Veljo Kisand (EE)
Leho Tedersoo (EE)
Kristel Panksep
Victoria Prins
Monika Böhm (US)

Hans-Peter Grossart (DE)
Alice Retter
Lars Ganzert
Solvig Pinnow
Laura Garzoli (IT)
Ester Eckert
Emanuele Ferrari
Diego Fontaneto
Selene Chinaglia

Former members
Michael Bruun-Nielsen
Sérgio Costa
Cátia Canteiro
Mercè Montoliu Nerín

Data

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

www.funaction.eu @funaction.bsky.social FUNACTION

biodiversa+ European Biodiversity Partnership

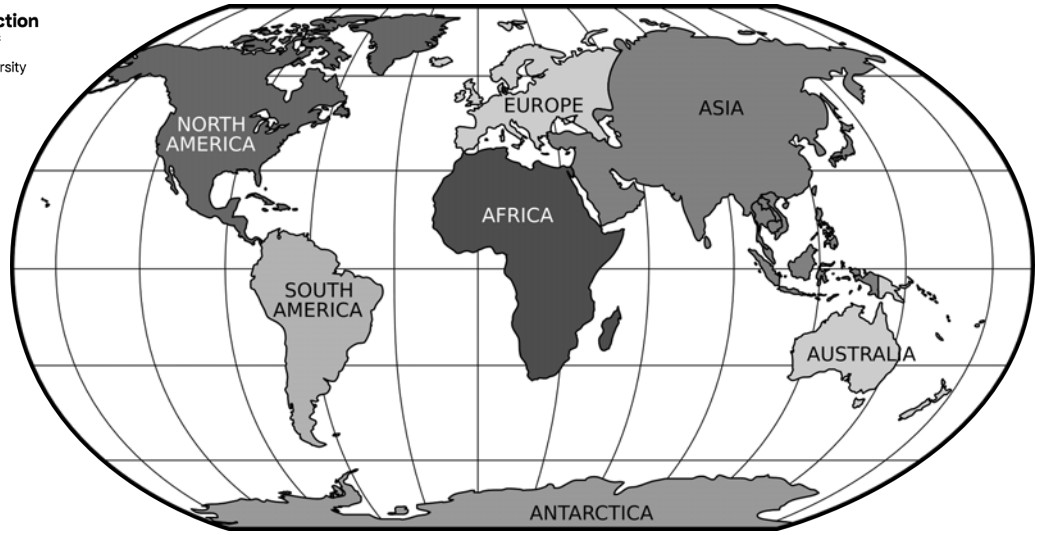
Jennifer Anderson (SE)
Olga Vinnere Pettersson
Ziming Wang
Isabel Fernandes (PT)
Ronaldo Sousa
Diana Graça
Andreas Bruder (CH)
Daniel Romero-Mujalli
Red Calore
Veljo Kisand (EE)
Leho Tedersoo (EE)
Kristel Panksep
Victoria Prins
Monika Böhm (US)
Hans-Peter Grossart (DE)
Alice Retter
Lars Ganzert
Solvig Pinnow
Laura Garzoli (IT)
Ester Eckert
Emanuele Ferrari
Diego Fontaneto
Selene Chinaglia

Former members
Michael Bruun-Nielsen
Sérgio Costa
Cátia Canteiro
Mercè Montoliu Nerín

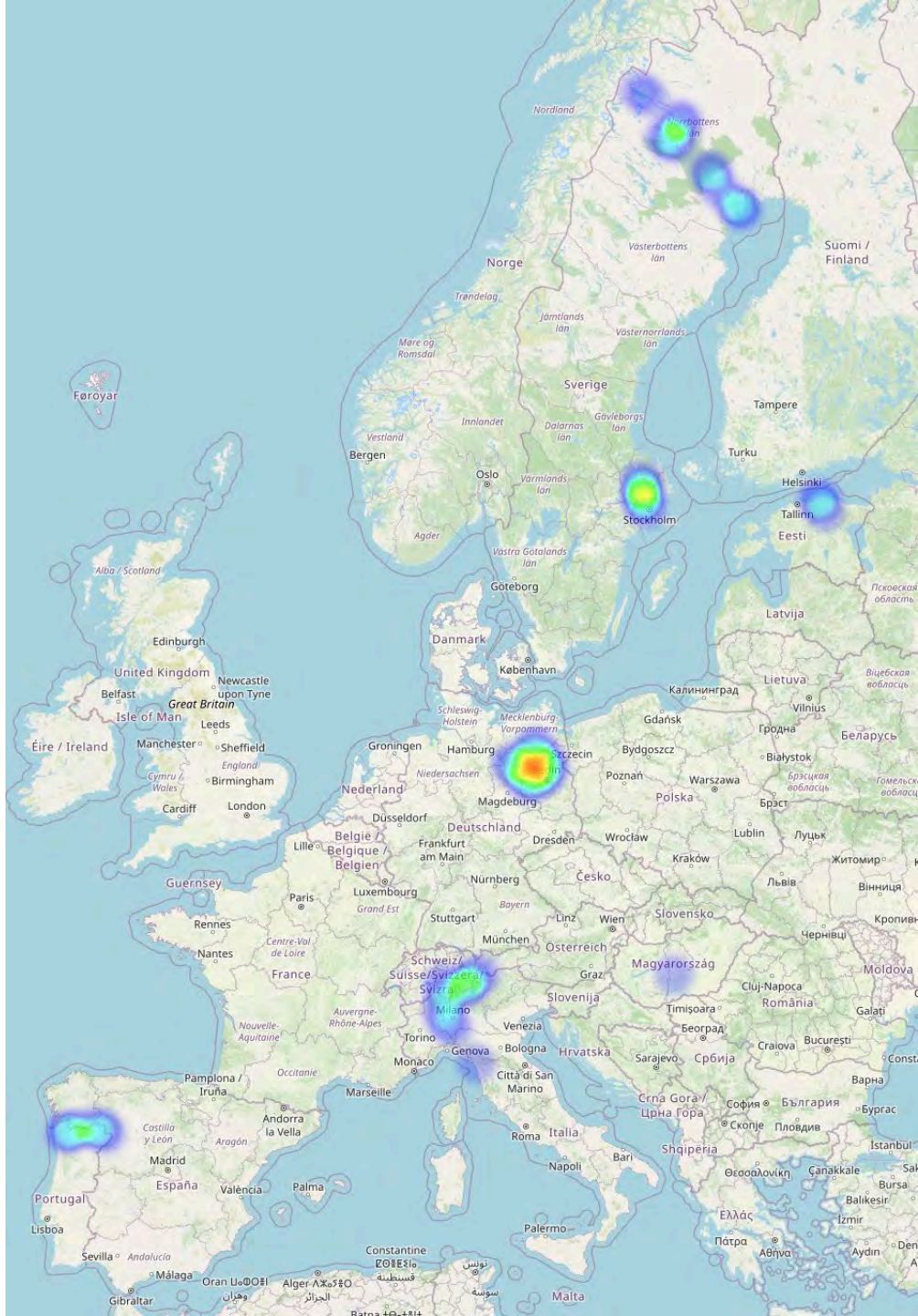
- Consortium members
  - names, contact info, social media accounts,
  - titles, positions, work addresses.
- Partner organizations—logos, rules
- National Funders—contacts, logos
- Consortium Meeting Notes
- Consortium Documents--other

Funaction
Aquatic
fungal
biodiversity

Data



- Consortium members
  - names, contact info, social media accounts,
  - titles, positions, work addresses.
- Partner organizations—logos, rules
- National Funders—contacts, logos
- Consortium Meeting Notes
- Consortium Documents--other



Data

Funaction
Aquatic
fungal
biodiversity

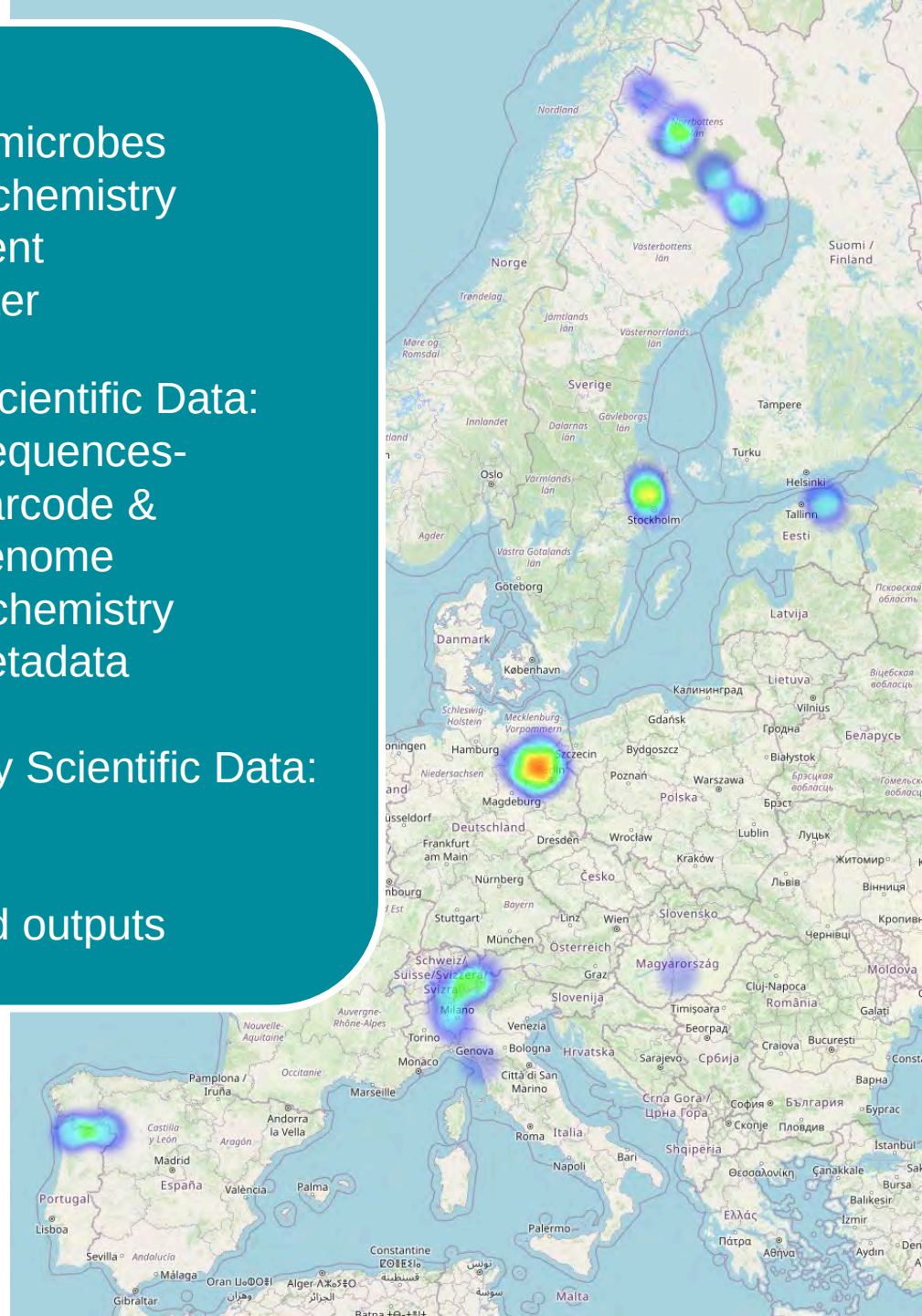
- Stakeholder info
- Surveys & responses
- Webinars & registrations

- Consortium members
  - names, contact info, social media accounts,
  - titles, positions, work addresses.
- Partner organizations—logos, rules
- National Funders—contacts, logos
- Consortium Meeting Notes
- Consortium Documents--other

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

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

Secondary Scientific Data:
- Maps
- Models
- Derived outputs



Data

Funaction
Aquatic
fungal
biodiversity

- Event outcomes
- Videos
- Outreach materials
- Media reach
- Outputs tracking

- Stakeholder info
- Surveys & responses
- Webinars & registrations

- Consortium members
  - names, contact info, social media accounts,
  - titles, positions, work addresses.
- Partner organizations—logos, rules
- National Funders—contacts, logos
- Consortium Meeting Notes
- Consortium Documents--other

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

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

Secondary Scientific Data:
- Maps
- Models
- Derived outputs

- Scripts
- Readme
- Analysis documentation

- Project datasets
- Independent datasets

Funaction
Aquatic fungal biodiversity

Data

- Event outcomes
- Videos
- Outreach materials
- Media reach
- Outputs tracking

- Stakeholder info
- Surveys & responses
- Webinars & registrations

- Permits, permissions, Nagoya, GDPR…..

- Photos – with people, w/out people

- Consortium members
  - names, contact info, social media accounts,
  - titles, positions, work addresses.
- Partner organizations—logos, rules
- National Funders—contacts, logos
- Consortium Meeting Notes
- Consortium Documents--other

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

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

Secondary Scientific Data:
- Maps
- Models
- Derived outputs

- Scripts
- Readme
- Analysis documentation

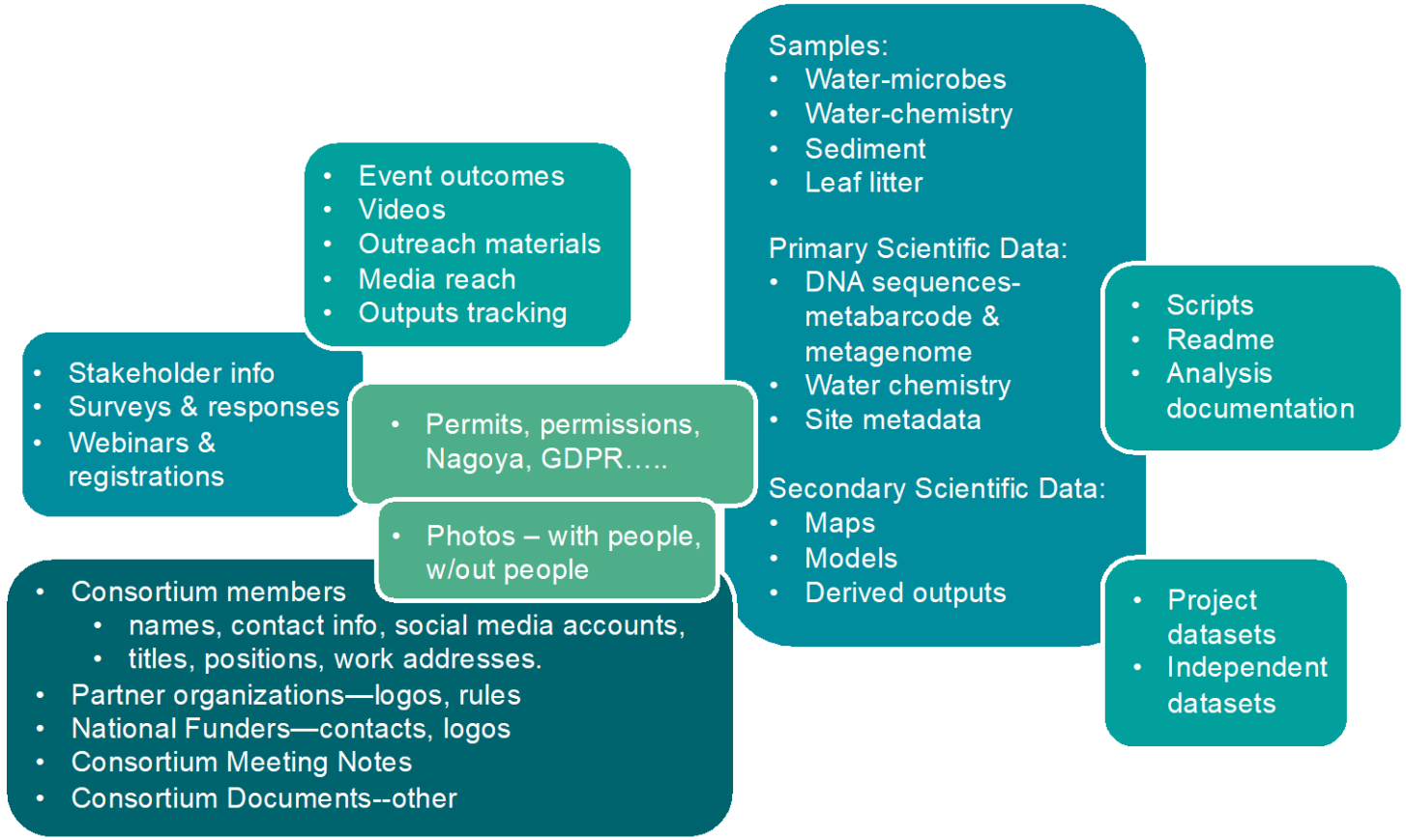
- Project datasets
- Independent datasets

Funaction
Aquatic
fungal
biodiversity

Data

Special considerations

- Large datasets
- Personal data - GDPR
- Public/Private
- who is responsible?
- When?
- Where?
- Backup
- Archiving
- Formats
- Terminologies
- Sharing internal/external



Funaction
Aquatic
fungal
biodiversity

Data Management Plans: Why scientists should create & sustain them

1	Project
8	Partners/teams
7	Countries
6	National funding agencies
23	Team members
4	Former members
5	Work Packages (2 co-leads each)
10	Case study watersheds
10	Timepoints
~800	Unique collections
>3000	Total samples

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

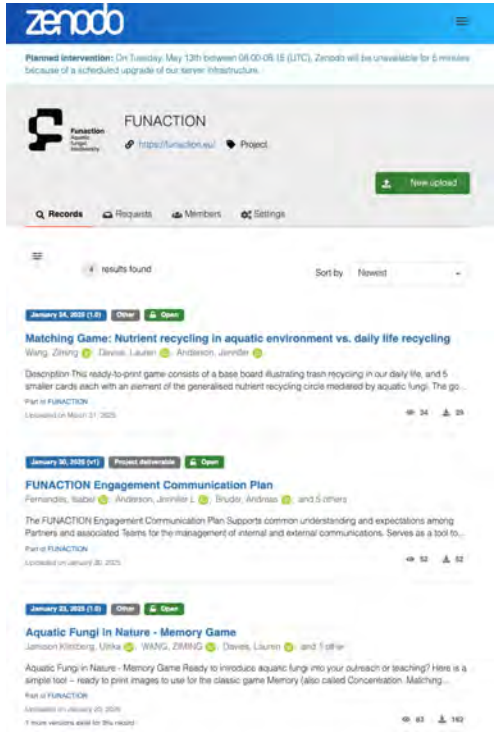
Shared/cooperative tasks

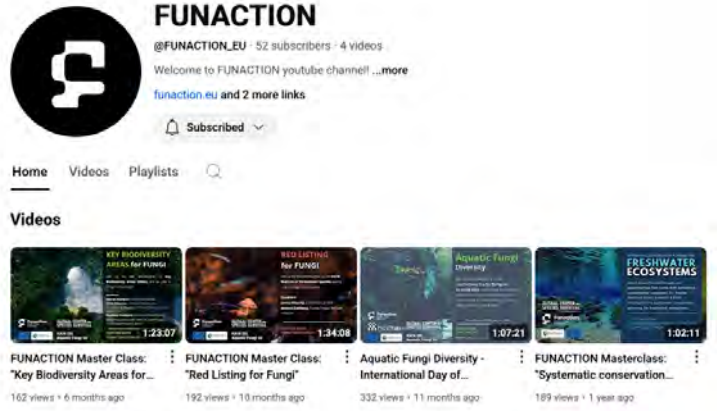
Open Science

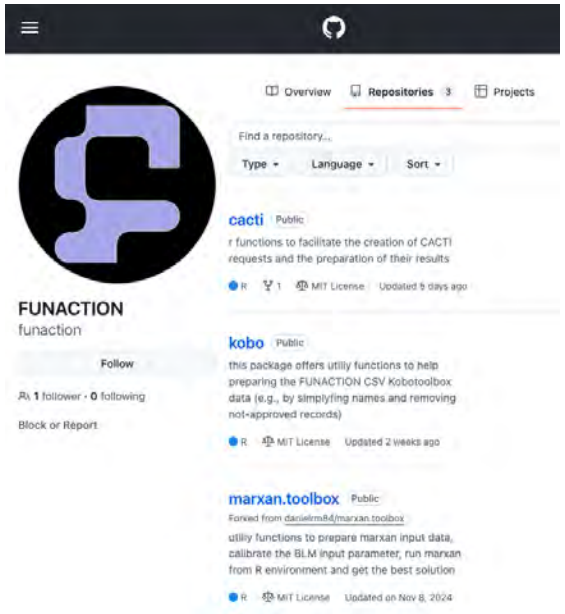
FUNACTION

@funaction.bsky.social

www.funaction.eu



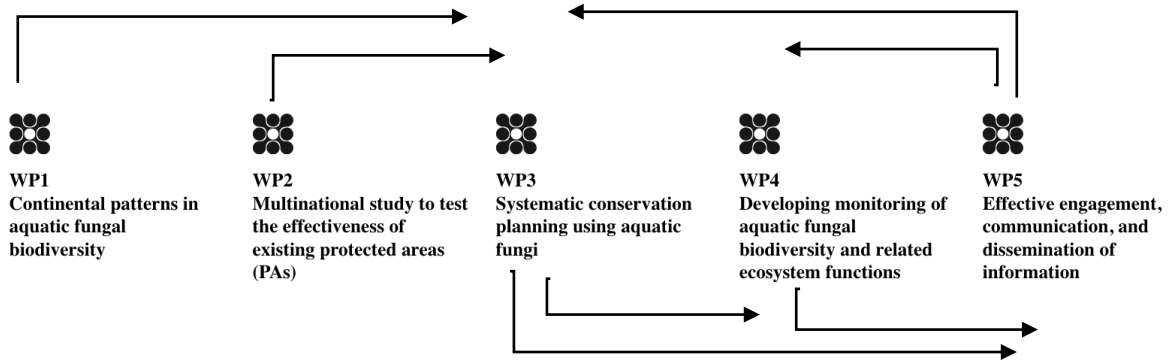




Data Management Plans: Why scientists should create & sustain them

Funaction
Aquatic fungal biodiversity

1	Project
8	Partners/teams
7	Countries
6	National funding agencies
23	Team members
4	Former members
5	Work Packages (2 co-leads each)
10	Case study watersheds
10	Timepoints
~800	Unique collections
>3000	Total samples



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

Shared/cooperative tasks

Spain (water chemistry)

Estonia (extractions, library prep)

Sweden (sequencing)

Open Science

Funaction
Aquatic
fungal
biodiversity

Data Management Plans: Why scientists should create & sustain them

Avoidance of chaos

Softulka | Shutterstock

Funaction
Aquatic
fungal
biodiversity

Data Management Plans: Why scientists should create & sustain them

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

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

Softulka | Shutterstock

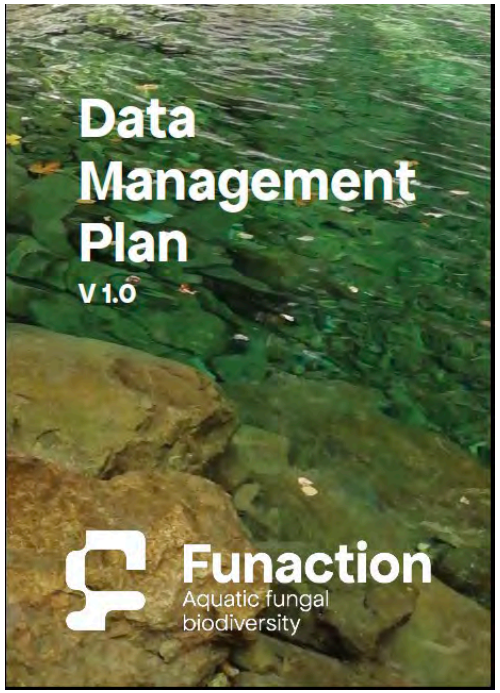
Data Management Plan:

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.



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

- 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

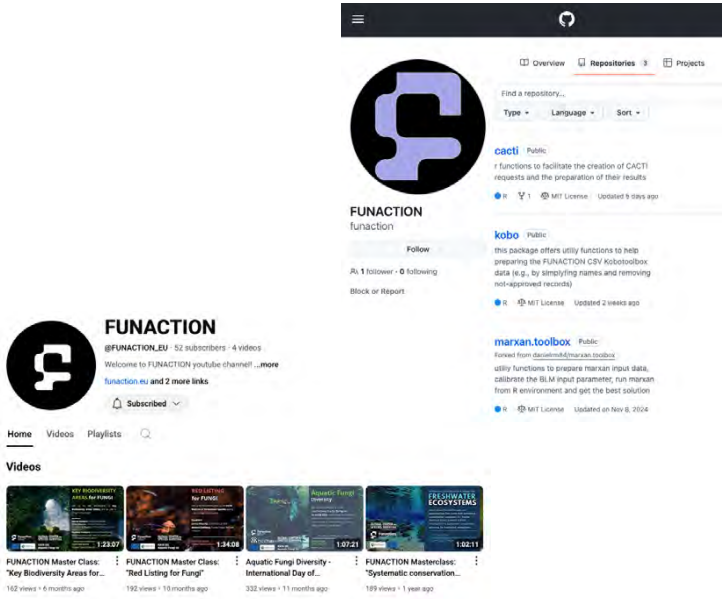
Element

Shared online docs

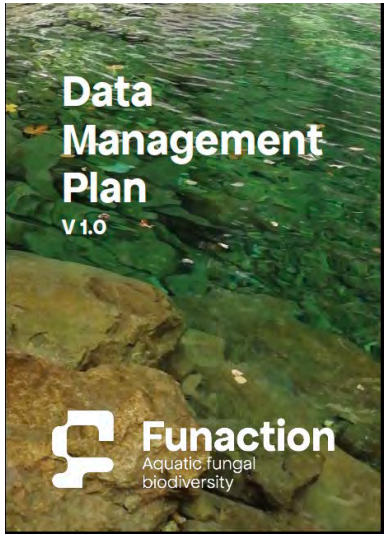
DMC meetings

Reporting at monthly meeting

Data - component	Solution
Sequence data	SLU: Physical server with backup. Redundant copies.
Personal data	SLU: Sharepoint w/ restricted access & physical server
Terminologies for metadata	Enforced through Kobotoolbox collection
IDs	Persistent unique identifiers as per DMP



Data Management Plan: FUNACTION



Sample naming conventions

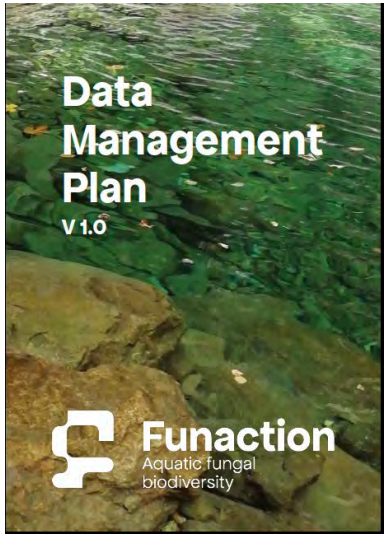
= persistent unique identifiers

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

Standardized terminologies

Centralized data collection

= Uniform, correctly spelled/typed/etc, terminology

= Standard formats

= Unified dataset from the start, across the countries

= Transparency

= Availability

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

Kobotoolbox

Data Management Plan: FUNACTION

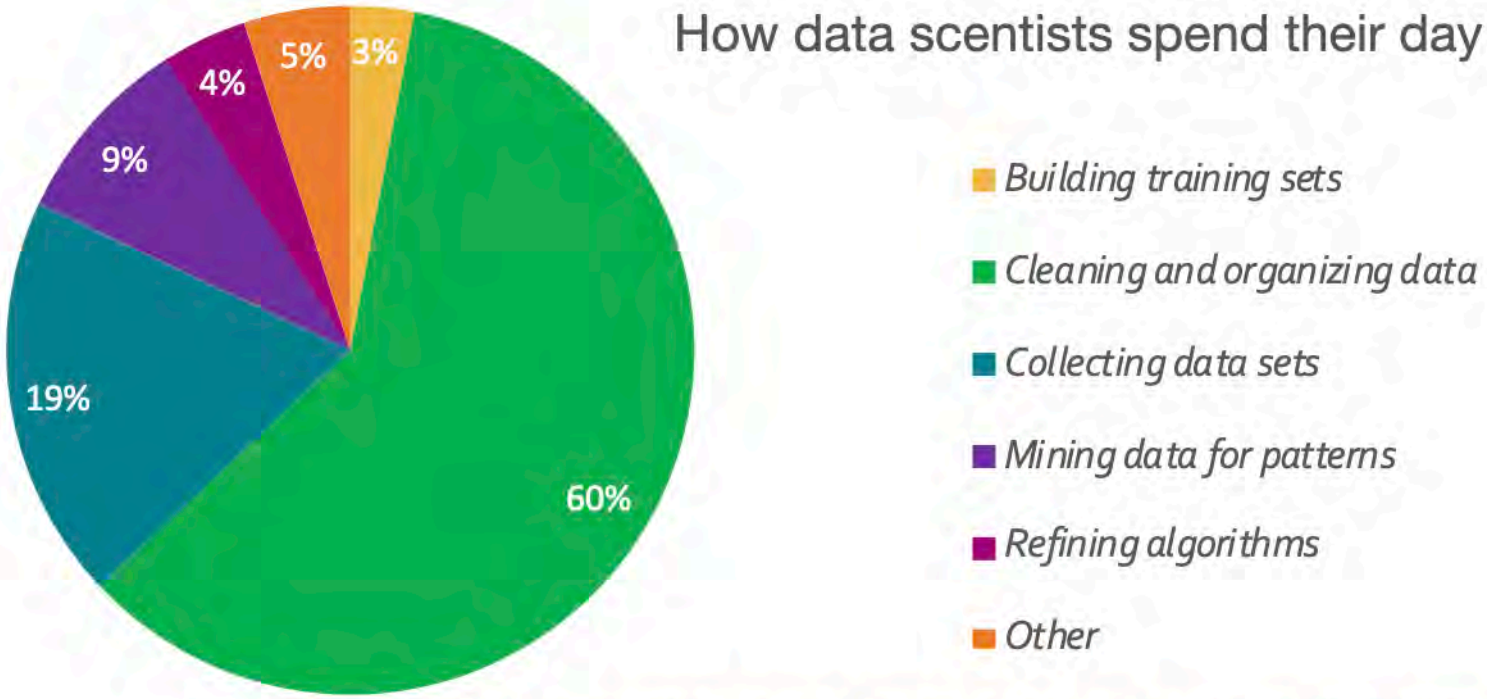


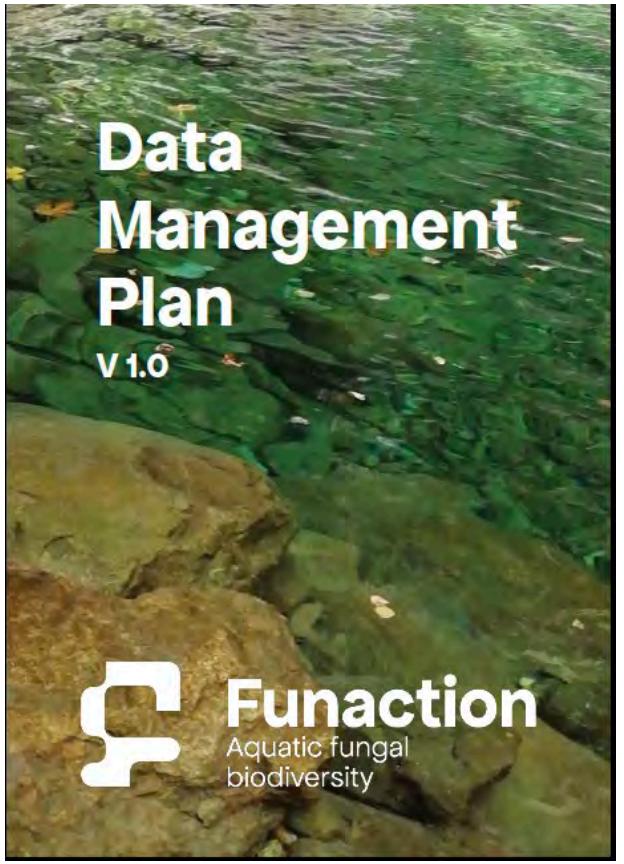
Figure 2: How data scentists spend their day (redrawn after CrowdFlower, 2016)

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

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

Artificial intelligence (AI) in data management

Eero Siivola, Finnish environment institute (Syke)

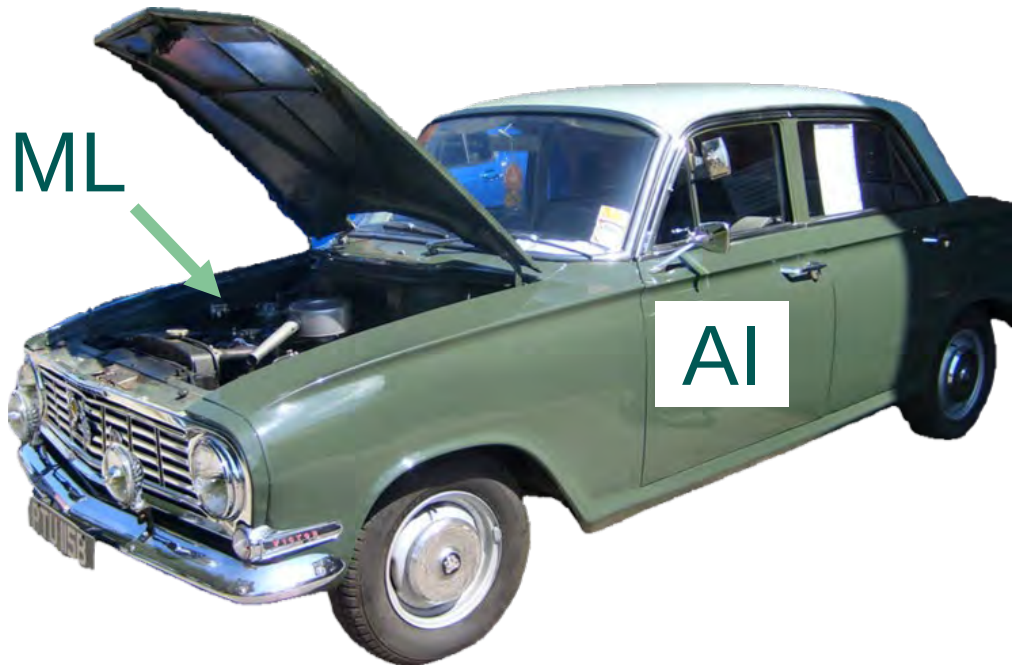
Suomen ympäristökeskus
Finlands miljöcentral
Finnish Environment Institute

Contents of the talk

- What is AI?
  - Beyond misconceptions and hype
- Foundations of the use of AI
- AI in Biodiversity Data Management
- Ethical concerns
- Future aspects of AI research
- Conclusion

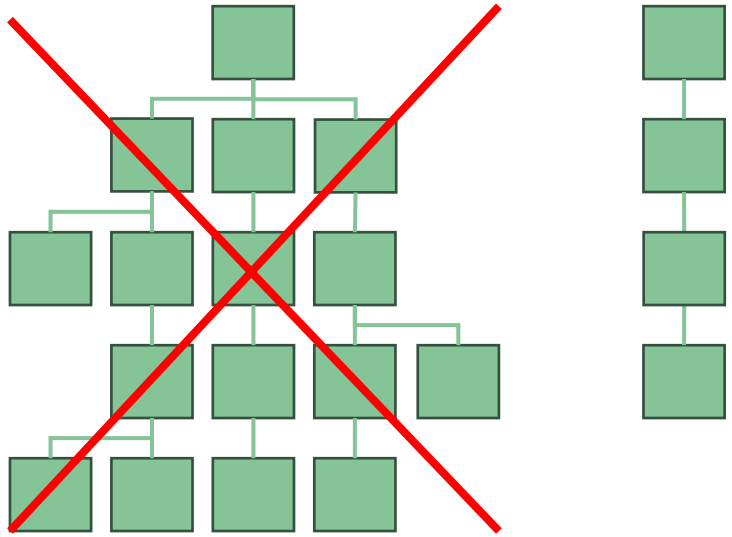
What is AI? (demystifying AI)

- AI = computer systems designed to perform tasks that typically require human intelligence
    - E.g. recognize patterns, make decisions, interpret images/language
- Now-a-days AI is commonly used as a synonym for different advanced tools like chat GPT
    - Most AI is driven by Machine Learning (ML) – methods (=methods that learn from data)
- Disclaimers:
    - AI is not magic – it always makes mistakes
    - AI works reliably on a set of very limited tasks,
    - Setting up AI is often laborous
    - AI is often black box



Foundations of the use of AI

- AI is data driven
  - AI depends on high-quality, well-documented and FAIR (Findable, Accessible, Interoperable, Reusable) data
  - Poor data = poor AI outcomes
  - AI often requires lots of data
- AI requires infrastructure: storage systems, processing power, tools for sharing and standardizing data
- AI requires interdisciplinary collaboration
  - E.g. domain experts (biologists, chemists, GIS experts or ecologists) and AI experts
- Building the foundations for AI is often laborous



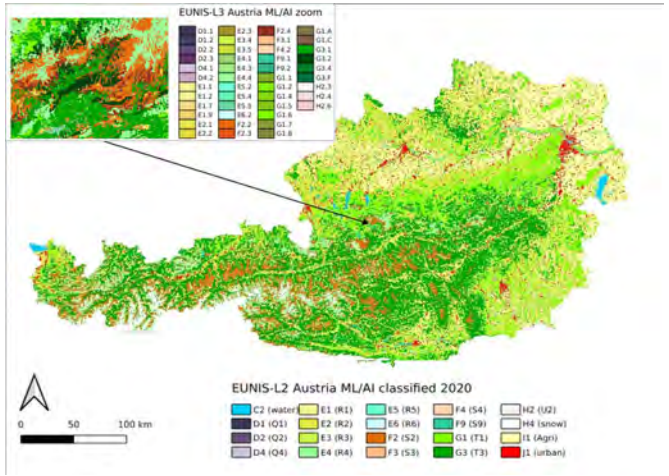
Occam's razor: all else being equal, the simplest explanation that fits the data is usually the best one.

Simpler, purpose-built models often provide better results than overly complex, generalized models.

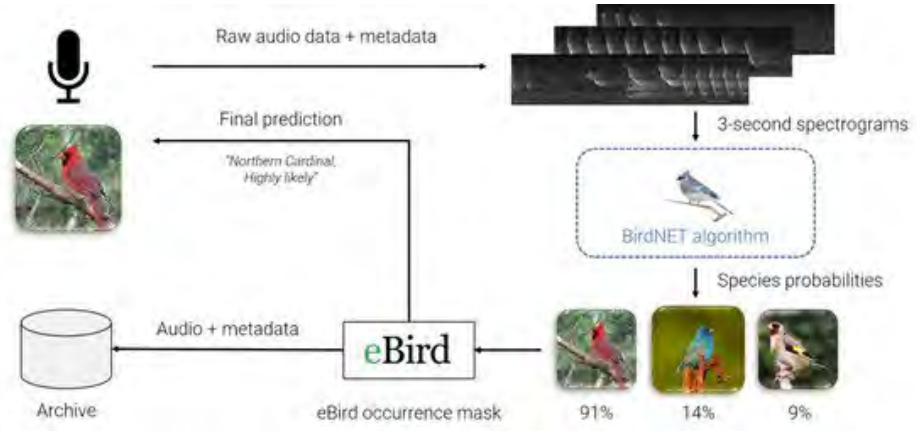
AI for data collection and data-analysis

- Tools like camera traps, satellites and acoustic sensors generate large volumes of data
- AI can help in pre-sort, extract or classify species in photos and sounds - e.g., iNaturalist, BirdNET
- AI might be useful if:
  - There exists lots of data (e.g. >10k data points)
  - Some level of mistakes are OK
  - There are resources for tailor-made solutions

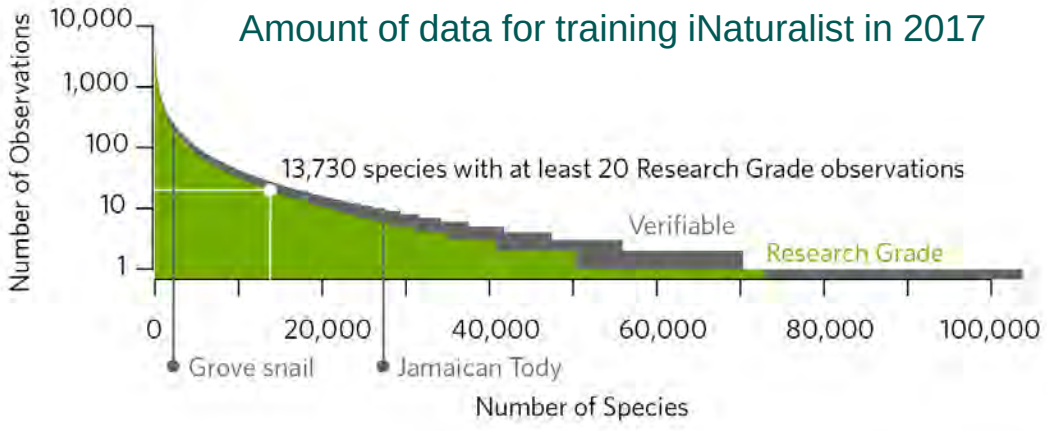
EUNIS classes of Austria from Sentinel 2 satellite data:



Work scheme of BirdNET:



Amount of data for training iNaturalist in 2017



AI in database management

- Anomaly detection: ML can flag unusual entries (e.g. coordinate far outside known range, new species)
- AI-enhanced databases learn from usage: e.g. predictive indexing or query optimization to speed up common database queries.
- Self-managing systems: cloud databases use AI to auto-scale, balance loads, and optimize storage for biodiversity repositories.

- Only few real life examples: Mostly basic research and custom solutions until now

Anomaly detection from camera data

Finlands miljocentral
Finnish Environment Institute

AI in master data management (MDM)

- Master data (like species lists) must be consistent

- Unifying reference data (especially taxonomy): AI can reconcile synonyms, spelling variants, and common names across datasets.
    - Example: GBIF's Backbone Taxonomy (plus AI fuzzy matching) ensures consistent species names.

- Recent advances in Natural Language Processing (NLP), like Large Language Models (LLMs) will potentially help
    - "LLM for Data Management": https://doi.org/10.14778/3685800.3685838
    - "How Large Language Models Will Disrupt Data Management": https://doi.org/10.14778/3611479.3611527

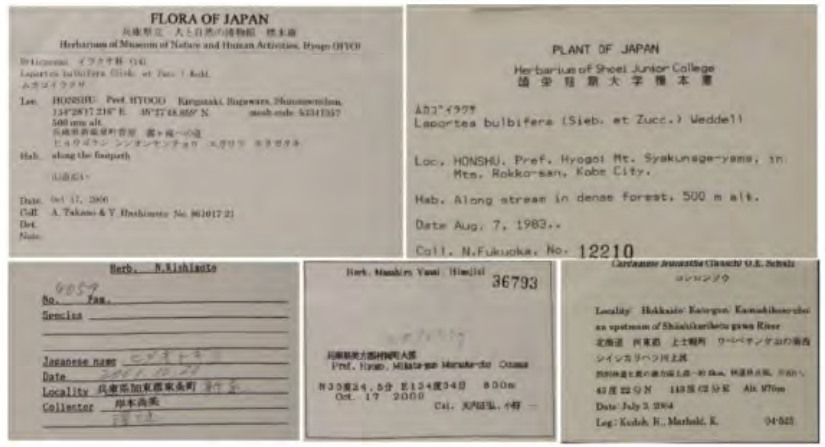
AI for reference data & metadata management

- AI can automate metadata creation: e.g. extract scientific names, locations, and keywords from text or images

- Intelligent tagging: NLP models classify datasets by theme (ecosystem, methodology), improving search/discovery

- Smart catalogs: AI-powered data catalogs (e.g. Alation) infer metadata and suggest relationships between datasets

Automatic metadata creation from text labels

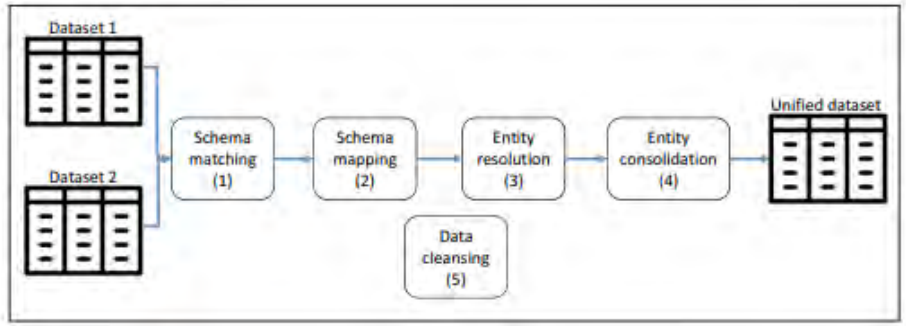


Suomen ympäristökeskus
Finlands miljöcentral
Finnish Environment Institute

Ai for data transformations and ETL pipelines

- ETL = extract, transform, load
- Smart ETL pipelines: AI/ML suggests mappings between formats and generate code for transformations
- Continuous integration: With ever-changing schemas, ML can update workflows automatically.
- While concrete examples are emerging, the trend is clear: ML will make data wrangling less manual and more intelligent.

Dataset integration with AI:



Suomen ympäristökeskus
Finlands miljöcentral
Finnish Environment Institute

AI in Biodiversity Data Management

Storage & lifecycle management:

- Data organization: AI can index and tag datasets, suggest metadata to improve Interoperability, and compress or summarize data for efficient accessibility (= easier and better FAIR)
    - This makes it easier to re-use existing data

- Preservation and arciving: AI can help to assess which data merit long-term storage (based on usage, quality, scientific value)
    - Example: EOSC EDEN project (Horizon Europe project, 2025->) is developing frameworks for this

- Cloud optimization: AI is increasingly being used by the cloud environment providers to optimize all kinds of processes

Storage and lifecycle management currently remains mostly manual for researchers

Ethical and legal concerns

- Transparency & explainability: Complex AI models are “black boxes.”
  - It’s essential to document, share code/models, and use explainable AI techniques
- Data bias and fairness: Biased training data (e.g. over-sampling common species or regions) can skew AI models.
  - Bias can lead to under-detection of rare or understudied species.
- Privacy and sensitivity: Some biodiversity data (e.g. the location of an endangered species nest) are sensitive.
  - AI tools must respect data privacy laws (like GDPR) and ethical guidelines.
- Environmental impact: Training large AI models consumes energy
- Regulation (EU AI Act): New laws will classify AI tools by risk level.
  - “low-risk” tools (like AI-assisted analytics) might have fewer restrictions, while “high-risk” cases (e.g. wildlife monitoring systems affecting policy) may need strict governance.

Future directions in AI research

- Foundation models: Models that are trained on vast datasets so they can be applied across a wide range of use cases
- Agentic AI: AI systems that can perform tasks independently and collaborate with other AI agents. These systems are expected to handle more complex tasks with minimal human intervention
- Resource efficiency: Large AI models will become more resource efficient
- Explainable AI (XAI): Research of interpretable models is and will stay popular due to increasing regulation
- Collaboration and community: Shared AI resources (open models, large labeled datasets like iNaturalist) will accelerate progress and revolutionalise some fields (E.g. recent development in meterology)
- EU AI and data legislation impact: The upcoming EU AI Act and related laws (GDPR, data protection) will shape AI use

Conclusion

- Key messages: AI is a powerful set of tools that can greatly enhance biodiversity data workflows.
    - Its success depends on quality data management
    - There are not many practical tools available

- Action points to be able to utilize AI for your data management:
    - Make sure your data is accessible and of good quality
    - Expect things to take more time and resources than the hype leads you to believe
    - For best results, collaborate with ML/AI experts
    - Keep your eyes open
        - Lots of attention on the field: Safe to expect fast progress

Questions/discussion?

eero.siivola@syke.fi

Suomen ympäristökeskus
Finlands miljöcentral
Finnish Environment Institute

Research/sources

https://arxiv.org/html/2504.09288v1

https://peerj.com/articles/18853/

https://www.inaturalist.org/pages/computer_vision_demo

https://bou.org.uk/blog-granados-birdnet/

https://blog.vito.be/remotesensing/habitat-mapping-ai

https://doi.org/10.3389/fmars.2023.1283265

https://www.nature.com/articles/s41598-023-50179-0

https://doi.org/10.1525/elementa.418

https://doi.org/10.3389/focsu.2024.1528837

https://www.frontiersin.org/news/2025/03/03/reimagining-fair-for-an-ai-world-frontiers-introduces-fair-data-management

https://link.springer.com/article/10.1007/s42514-022-00101-3

https://eosc.eu/eu-project/eosc-eden/