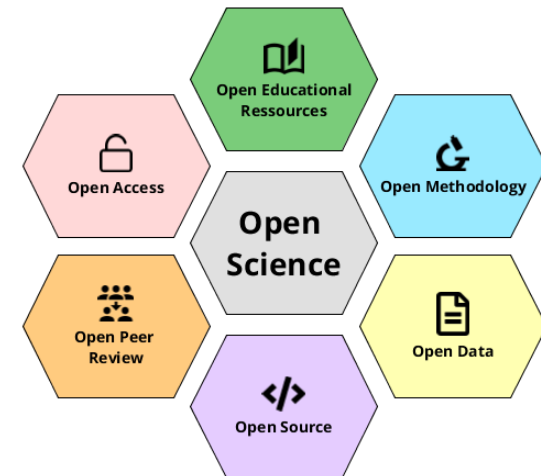


Data and software publications for open science @ UFZ

Katharina Klemm, Paul Meißner, Julia Sperfeldt (Library)
Stephanie Jurburg (UMB)
Ronny Gey (Research Data Management)



29 Nov 2023

Agenda

- **Open Science (RDM)**
- **Helmholtz Open Science Policy (RDM)**
- **Helmholtz Quality Indicators for Data and Software Products (RDM)**
- **Data- and Softwarepublications (RDM)**
- **Workflow for submitting your publications to UFZ (Library)**
- **PIDs (ORCID, ROR) (Library)**
- **Open Science – a practitioners perspective (Stephanie Jurburg)**
- **Q&A**



The world needs open science now. Why?

Open science can accelerate our ability to help solve the complex challenges of our interconnected world. We need it because:



Global challenges such as poverty, armed conflict, the climate crisis, environmental degradation and health and humanitarian crises are **urgent**—and science, technology and innovation can better respond to them if they collect and apply ideas from diverse contributors and knowledge systems.



Open science has the potential to accelerate the achievement of the **Sustainable Development Goals** by reducing or ending inequalities in access to science, technology, innovation and their applications.



Research practices that are more transparent, collaborative and **inclusive** are subject to more effective peer review, increased scrutiny and critique which in turn increases the verifiability and reproducibility of the science produced. This ultimately leads to better science, more trust in science and more relevant and positive impacts of science on society.

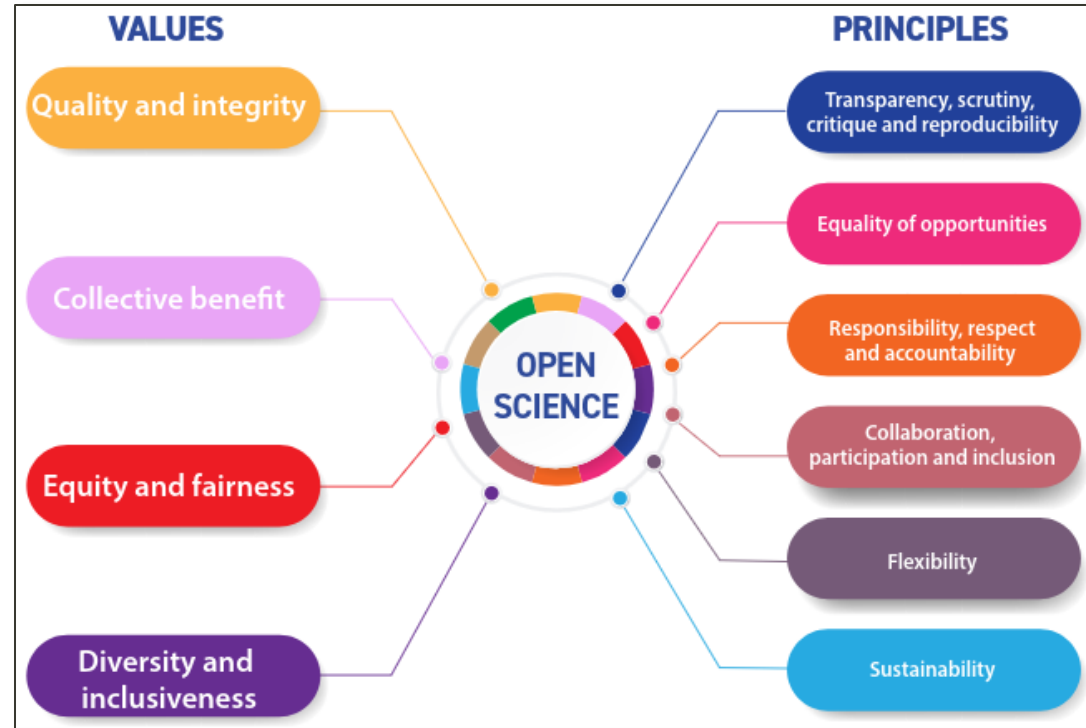
- Movement to make **scientific research** and its dissemination **accessible to all levels of society**.
- Global challenges, inequality
- Verifiability, reproducibility, efficiency => better, trustful science
- **UNESCO Recommendation on Open Science**
<https://doi.org/10.54677/XOIR1696/>
<https://doi.org/10.54677/MNMH8546>

Open Science

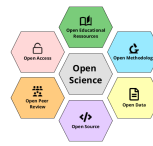
Motivation




UNESCO Recommendation on Open Science – Values and Principles



CC BY-SA 3.0 IGO: <https://doi.org/10.54677/MNMH8546>



- **Access:** Research outputs are distributed online, free of access charges or other barriers. 
- **Data:** Openly accessible, exploitable, editable data, shared by anyone for any purpose.
- **Software:** *Source code* freely available for possible modification and redistribution.
- **Educational resources:** teaching, learning, and research materials free to own, share, and modify.
- **Methodology:** Open application of methods and the entire process of a research project.
- **Peer review:** *Principles:* Open identities, Open reports, Open participation.
- **(Hardware):** open design specifications of a phys. object which can be studied, modified, and distributed.
- Converging overlap with Good Scientific Practices!

Open Science is Science just done right.

Open Science

Motivation – Why would *you*?



- *Open* data and software publications count! (in HGF since reporting yr 2022)
- Visible, transparent, reproducible, reusable, sustainable research
- Inspire others
- Open science requirements increasingly important for career in science (job promotion, journal requirements)



Open Science

What about HGF?

- **Helmholtz Open Science Office:** <https://os.helmholtz.de>

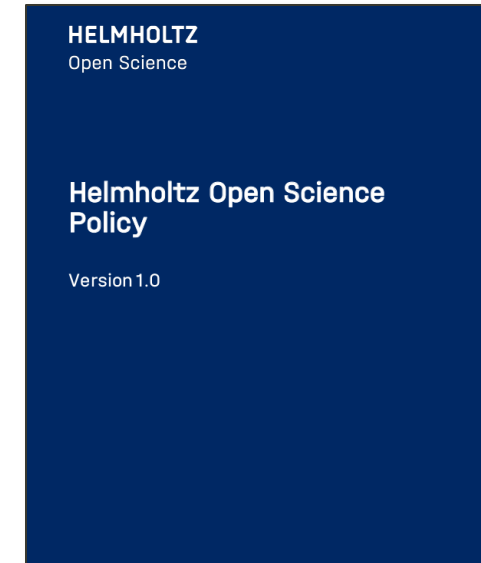
- Workshops, presentations, conferences, work & task groups
- Recommendations, **guidelines**, white papers
- (Inter)National Initiatives/Consortia (EOSC, RDA, RSA, deRSE, ...)

- **Helmholtz WG Open Science:**

- Task Group Open Access Transformation
- Task Group for the Implementation of Research Data Policies
- Task Group Research Software
- **Task Group Helmholtz Quality Indicators for Data and Software Products**
- Task Group Guidelines for Research Software
- Task Group Recommendations for the Implementation of the Helmholtz Open Access Guideline
- Task Group Addition to the Criteria for Operating OA Publication Funds
- Task Group Open Access Guideline
- Task Group Rules of Procedure

Enabling Open Science
practices at Helmholtz

- **2022 Sep 20** adopted by the General Assembly of the Helmholtz Association
- <https://doi.org/10.48440/os.helmholtz.056>
- **Monitoring**
 - Specific and verifiable *objectives*
- **Implementation & common requirements**
 - Open science *practices*
- **3 Pillars:**
 - Open Access
 - Open Research Data
 - Open Research Software



- **WG Open Science & WG Library and Information Management**
- **Monitoring**
 - 2020: 60% OA
 - 2025: 100% OA
- **Implementation & common requirements**
 - Peer reviewed OA publications are funded by the Helmholtz center
 - Electronic copy after 12 months in institutional repository, freely available (CC-BY)
 - Metadata: CC0, FAIR, PIDs
 - publication (author(s), title, date of publication, publication organ); name, acronym, and number of the funding project; licensing conditions; persistent identifiers for the publication (DOI) and, if possible, their organizations (ROR) and the funding ().

Helmholtz Open Science Policy

Open Access - UFZ

- [Publication Guideline](#) (2022)
- [Publication fund of the UFZ](#)
- OA Ambassadors
 - Dr. Johannes Förster (UPOL)
 - Dr. Anja Miltner (UBT)
- [Open Access](#) (BIBL)



UFZ-Regulation | IR-47/12 | 10.01.2022

Publication Guidelines of the Helmholtz Centre for Environmental Research – UFZ

Please note: This is a translation of the German version of the above-mentioned regulation and is intended solely as a convenience to the non-German-reading employees. Any deviations from the original German regulation are not binding and have no legal effect for compliance or enforcement purposes.

Enforced by		
Executive Management		
	Prof. Dr. Georg Teutsch	Dr. Sabine König
Executive Management meeting	Executive Management meeting no. 363, resolution no. 3.2	
Date	24.03.2022	

Document history

Date	Version	
01.01.2007	1	Entry into force of the version of 18.12.2006
24.03.2022	2	Resolution on the entry into force of the version of 10.01.2022

- **Monitoring**

- Development of research data policy
- 2023: basic indicator
- 2024: quality indicator

- **Implementation & common requirements**

- Orientation: DFG - Guidelines for Safeguarding Good Scientific Practice, [Recommendations for Policies of the Helmholtz Centers on Research Data Management](#)
- DMP, Repository ([re3data](#) indexed)
- CC-BY or CCO; as open as possible, as closed as necessary
- Metadata: CC0, FAIR, PIDs
 - data set (description, date of deposit, authors, publication organ, and embargo); name, acronym, and number of the funding project; licensing conditions; persistent identifiers for the data set and, if possible, organizations and funding

- [Principles for the Responsible Handling of Research Data at the Helmholtz Centre for Environmental Research – UFZ](#) (2018 Sep 13)
- Policy Update in progress ([hopefully] coming into force: beginning 2024)
- Practice? [RDM Guidelines](#)
- [RDM Team](#)
- [RDM „Community“](#)

PANGAEA Community Workshop
"FAIR Data Publishing with PANGAEA"
Recordings

UFZ-Regulation | IR-5/18 | 13.09.2018

Principles for the Responsible
Handling of Research Data at the
Helmholtz Centre for Environmental
Research - UFZ

	Released		
Department	Executive Management		Meeting of the Executive Management 06/2018 (consecutive number 333)
Name	Prof. Dr. Georg Teutsch	Prof. Dr. Heike Graßmann	resolution Nr. 3.2
Date			13.09.2018

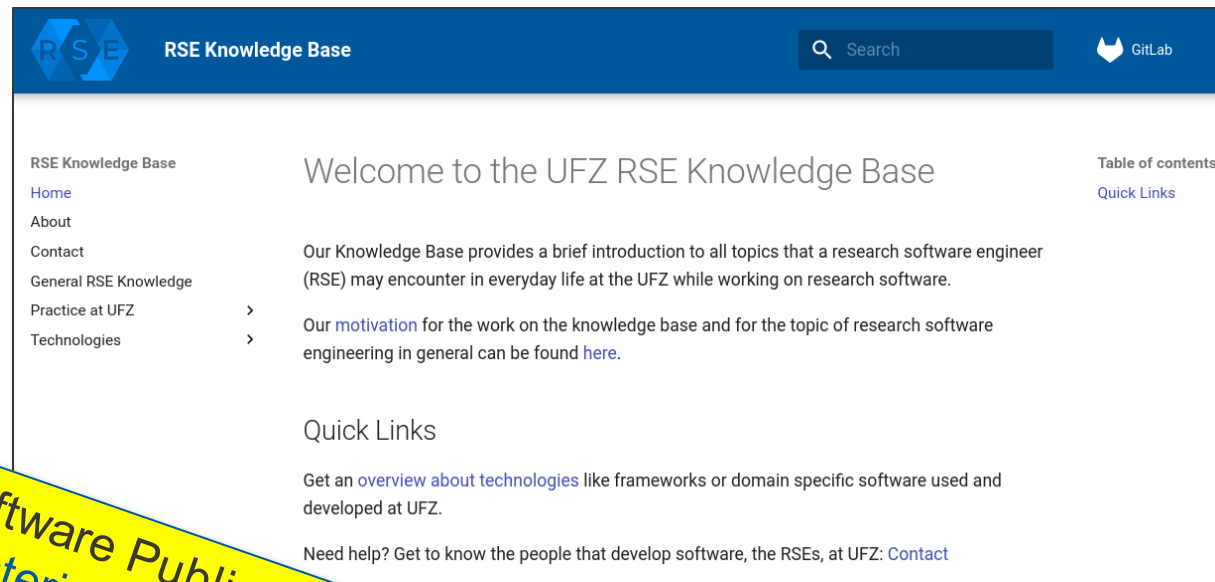
- **Monitoring**

- Development of research software policy
- 2023: basic indicator
- 2024: quality indicator

- **Implementation & common requirements**

- Orientation: [Model Policy on Sustainable Software at the Helmholtz Centers](#)
- Openly accessible repository
- *“researchers are supported in publishing research software”*
- Metadata: CC0, FAIR, PIDs
 - program code (description, date of deposit, version, authors, repository); name, acronym, and number of the funding project; licensing conditions; persistent identifiers, and, if possible, organizations and funding

- [RSE Policy](#) (coming into force in 2024)
- [RSE Knowledgebase](#) (WIP!)
- [RSE Community](#)



Foundations of Research Software Publication
[HIFIS Training Material](#)

Data and Software Publications

Make them count...

- **Why do we need indicators?**
 - Consistent reporting over the 18 HGF centers (criterias, processes)
 - Making data and software count!
- Task Group Helmholtz Quality Indicators for Data and Software Products
- HGF Open Science Policy => **Monitoring**
- **2022/23: *basic* indicator**
 - 1 quantitative variable: **DOI registered** data/software publication
- **2024/ff: *quality* indicator**
 - Quantitative **and qualitative** variables
 - *In progress ...*

■ Data

- Publishing, Metadata, Curation, Openness, External view

Published with Information on Access to the Data (Openness)

Example

- 0 - No Metadata available
- 1 - Metadata available, but no Data Access-Information available in the Metadata
- 2 - Metadata available, Data Access information available only in human-readable form
- 3 - Metadata available, Data Access information available only in human readable form, including general license information
- 4 - Metadata available, Data Access information available in human-readable and machine-readable form*, including license information

■ Software

- Findable, Accessible, Interoperable, Reusable, Scientific basis, Technical basis

Code Structure (Technical Basis)

Example

- 0 - No information given.
- 1 - Is every developer free to use his/her own style of coding?
- 2 - Are there general recommendations for coding, albeit every developer being able to follow his/her own style?
- 3 - Is there some harmonization of code style being enforced following common standards including meaningful naming of functions/variables etc.?
- 4 - Is the code style being checked when accepting changes into the repository?
- 5 - Is the code style being enforced via a review process (e.g. failed pipelines or auto-formatting)?


- **Procedure**

- Manually
- Lists of UFZ publications per repository (e.g. UFZ publication database, Zenodo, Pangaea)
- DOI check
- Disambiguation

- **Data Publications:** 127

- **Software Publications:** 44

- **Feedback**

- High number of undiscovered cases expected
- ~a dozen different spellings for 'UFZ' ( to the rescue...)
- Unconsistencies with authorship
- How do we deal with double affiliations?
- Authors join/leave UFZ within reporting year?

- No manual procedure
- Indicators based on the publications **registered with the library/publication database** only
- Probably again changes in 2024 (quality indicators)

Data and Software Publications

Takeaway message...

- Register a DOI for your D&S publications
- Register your D&S publication in the UFZ reporting process



An overview of the workflow for submitting data and software publications

Katharina Klemm - Library

Julia Sperfeldt - Library

November 2023

Submission process

- The collection of all UFZ publications is carried out by the library for all departments in the Institutional Repository of the UFZ
- In addition to text publications, **software and data publications** are now included in our database **from 2022** (year of publication) onwards
- Please notify us of your software and data publications with an **email** to **publikationen@ufz.de**
- After submission, all important data will be entered into the database promptly
 - your bibliographic data are subjected to extensive quality checks, standardization, validation and is enriched with structural data from our center
- Based on the publication directory, the **publication lists** of the POF pages, department homepage and your own homepage are created
- Furthermore, it forms the **basis for various evaluations**
 - it is in your own best interest to submit your publications as soon as possible!

Submission process

Mandatory information for registration in the database

Software

- valid POF number
- DOI
- Version


Data



- valid POF number
- DOI

Please include a **link** to the repository / publisher

Submission process

View in publication database

Category	Data Publication
DOI	10.1594/PANGAEA.895383
Licence	
Title (Primary)	Results from lake water analysis from Rappbode Reservoir in Saxony-Anhalt during the Inland Water Remote Sensing Validation Campaign 2017
Author	Hieronymi, M.; Bumberger, J. ; Friese, K. ; Herzog, M. ; Kormann, B.; Kuehn, B. ; Werther, M.; Wieprecht, M. ; Rinke, K.
published in	PANGAEA
Year	2023
Department	SEEFO; MET
Language	englisch
Topic	T5 Future Landscapes
ID	27253
UFZ intranet only	
Persistent UFZ Identifier	https://www.ufz.de/index.php?en=20939&ufzPublicationIdentifier=27253
<p>Hieronymi, M., Bumberger, J., Friese, K., Herzog, M., Kormann, B., Kuehn, B., Werther, M., Wieprecht, M., Rinke, K. (2023): Results from lake water analysis from Rappbode Reservoir in Saxony-Anhalt during the Inland Water Remote Sensing Validation Campaign 2017 PANGAEA 10.1594/PANGAEA.895383</p>	

Category	Software Publication
DOI	10.5281/zenodo.7664005
Licence	
Title (Primary)	EE-Monitor - Monitoring for a nature-friendly energy transition in Germany
Version	1
Author	Thrän, D. ; Manske, D.  Schinkel, B. ; Schmiedt, J. ; Mittelstädt, N.
published in	ZENODO
Year	2023
Department	BIOENERGIE
Topic	T5 Future Landscapes
ID	27402
UFZ intranet only	
Persistent UFZ Identifier	https://www.ufz.de/index.php?en=20939&ufzPublicationIdentifier=27402
<p>Thrän, D., Manske, D., Schinkel, B., Schmiedt, J., Mittelstädt, N. (2023): EE-Monitor - Monitoring for a nature-friendly energy transition in Germany Version: 1 ZENODO 10.5281/zenodo.7664005</p>	

PIDs at UFZ

PIDs (ORCID, ROR)



- **Open Researcher and Contributor ID**
- Persistent identifier for persons who contribute to research output
- ORCID record provides information about education, workplaces and a list of research output
- UFZ part of ORCID DE consortium
 - More information on our [UFZ ORCID website](#)
- [Register now!](#)

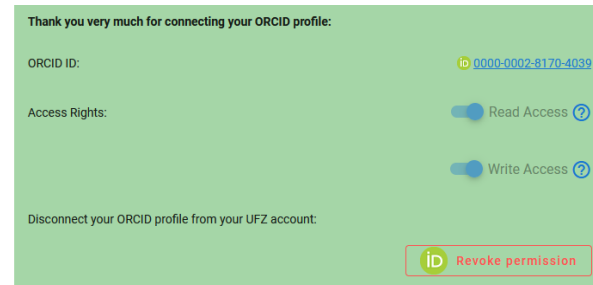
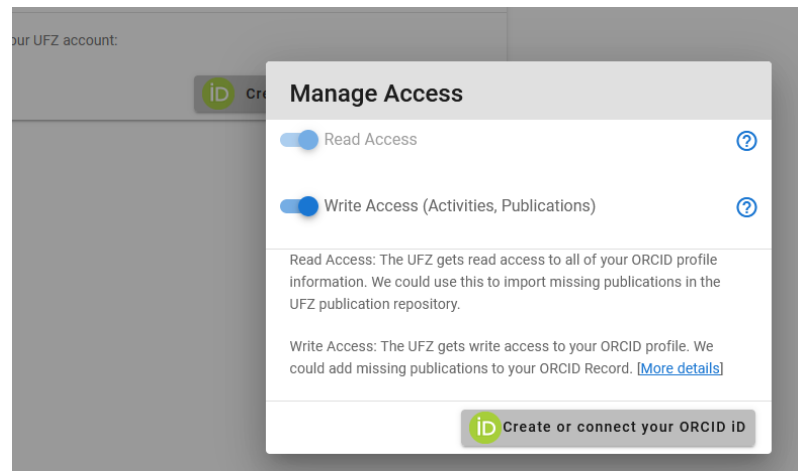


- **Research Organization Registry**
- Disambiguates institution to prevent mistaking for another Helmholtz Centre
- UFZ: <https://ror.org/000h6jb29>

Image source: <https://doi.org/10.5281/zenodo.4702037>

PIDs (ORCID, ROR)

Connecting UFZ account to ORCID



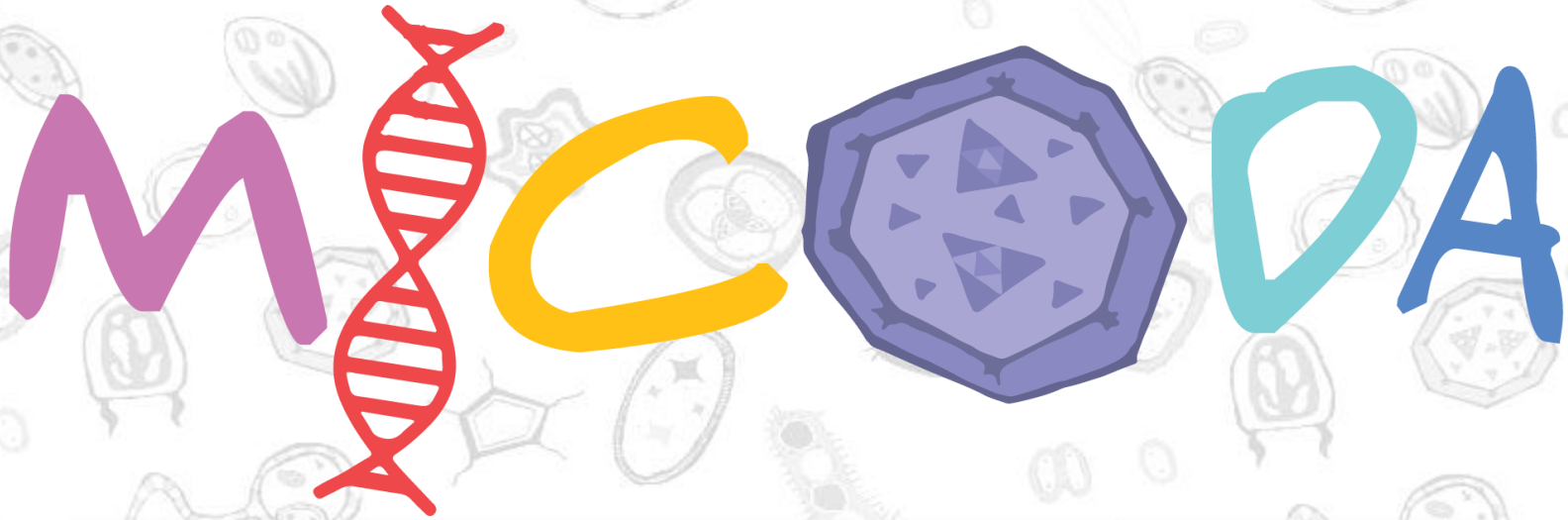
Use the following button to log out:

ORCID Profile vs UFZ Publication Repository

Found in UFZ Repository: 0
Found in your ORCID Profile: 0

DOI	10.1016/j.watres.2021.117776
Document	publication document UFZ intranet only
Title (Primary)	Superabsorbent polymer as a supplement substrate of constructed wetland to retain pesticides from agricultural runoff
Author	Jing, Y.; Krauss, M. ; Zschieschang, S.; Miltner, A. ; Butkovskyi, A.; Eggen, T.; Kästner, M.; Nowak, K.M.
Journal	Water Research

A practitioners perspective

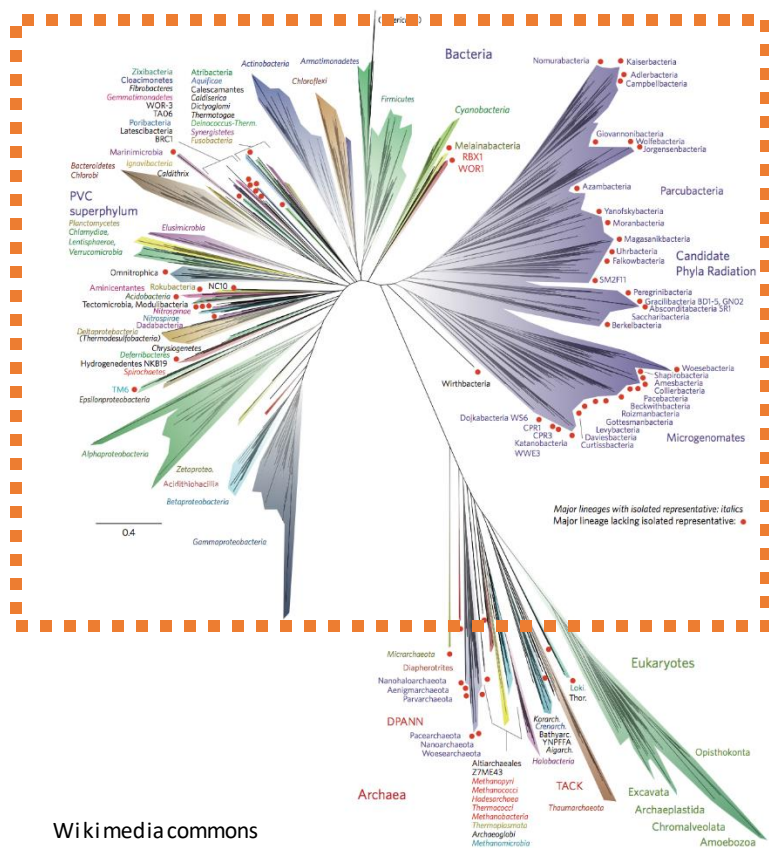


The microbial community database

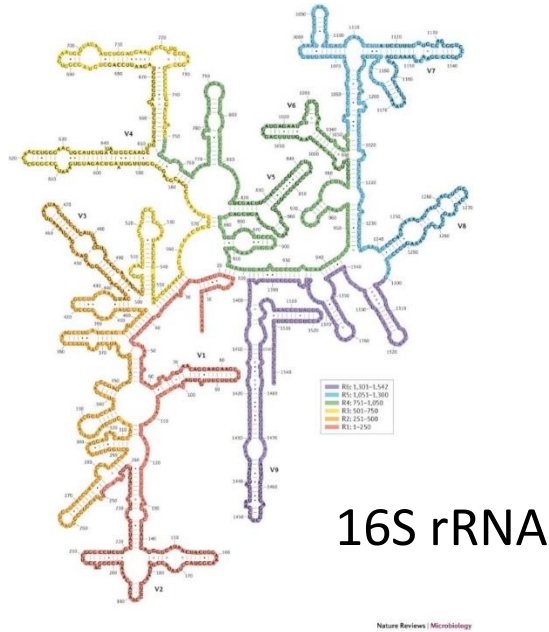
Steph Jurgurg

UMB, Helmholtz Centre for Environmental Research

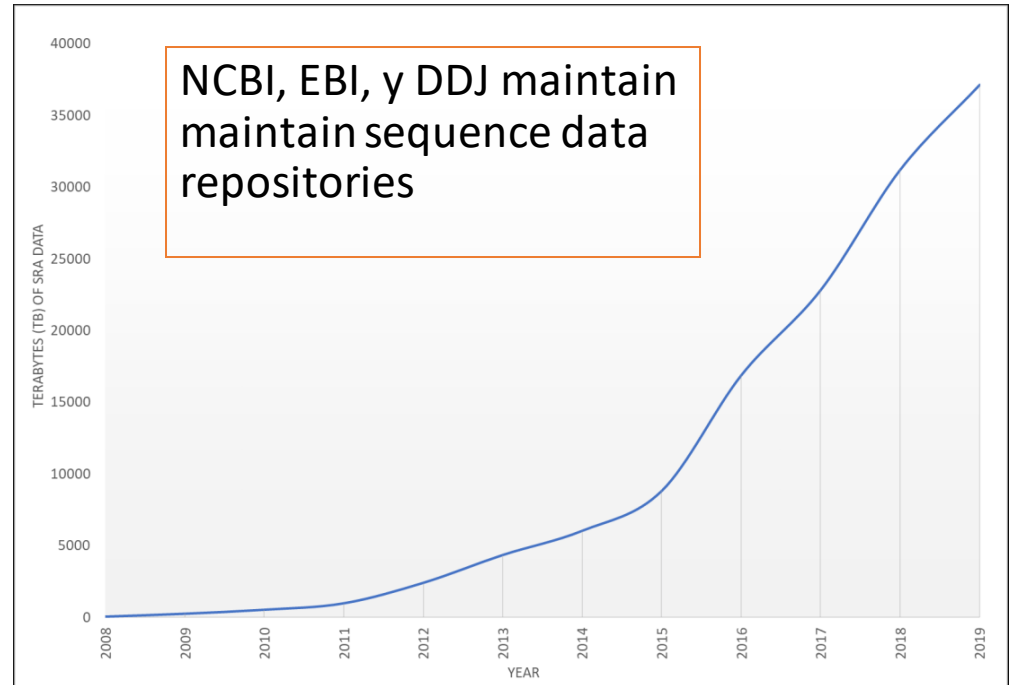
Microbial communities are the oldest



... and to us, they are the newest



Yarza *et. al*, 2014



NCBI, 2019

Data reuse should be easy in microbial ecology

*D54–D56 Nucleic Acids Research, 2012, Vol. 40, Database issue
doi:10.1093/nar/gkr854*

Published online 18 October 2011

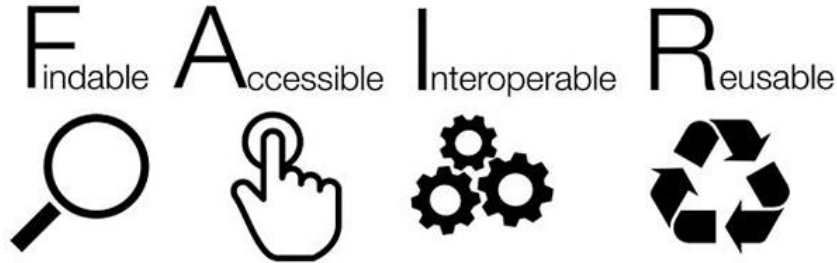
The sequence read archive: explosive growth of sequencing data

Yuichi Kodama^{1,*}, Martin Shumway² and Rasko Leinonen³ on behalf of the International Nucleotide Sequence Database Collaboration

¹Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Yata, Mishima 411-8540, Japan, ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA and

³European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Data archiving in our field is mandatory



Where are the data?



Maximilian Konzack



Anna Heintz-Buschart

Where are the data?

- Census of the availability of 16S rRNA gene data
- Reusability after each step of archiving

17 journals

29 235 articles

2660 16S rRNA
gene datasets



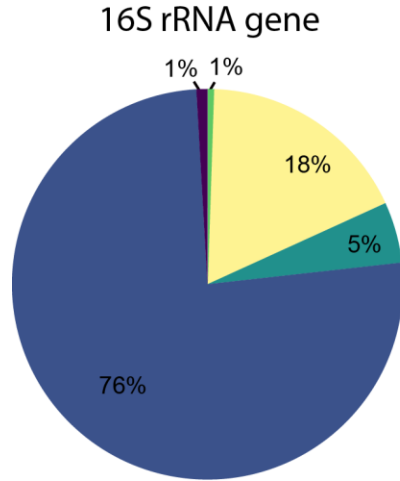
Maximilian Konzack



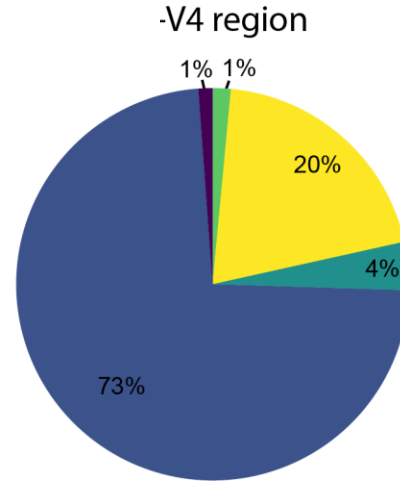
Anna Heintz-Buschart

Where are the data?

- Census of the availability of 16S rRNA gene data
- Reusability after each step of archiving



a



b

Storage location

- figshare
- INSDC databases
- MG RAST
- qiita
- no data

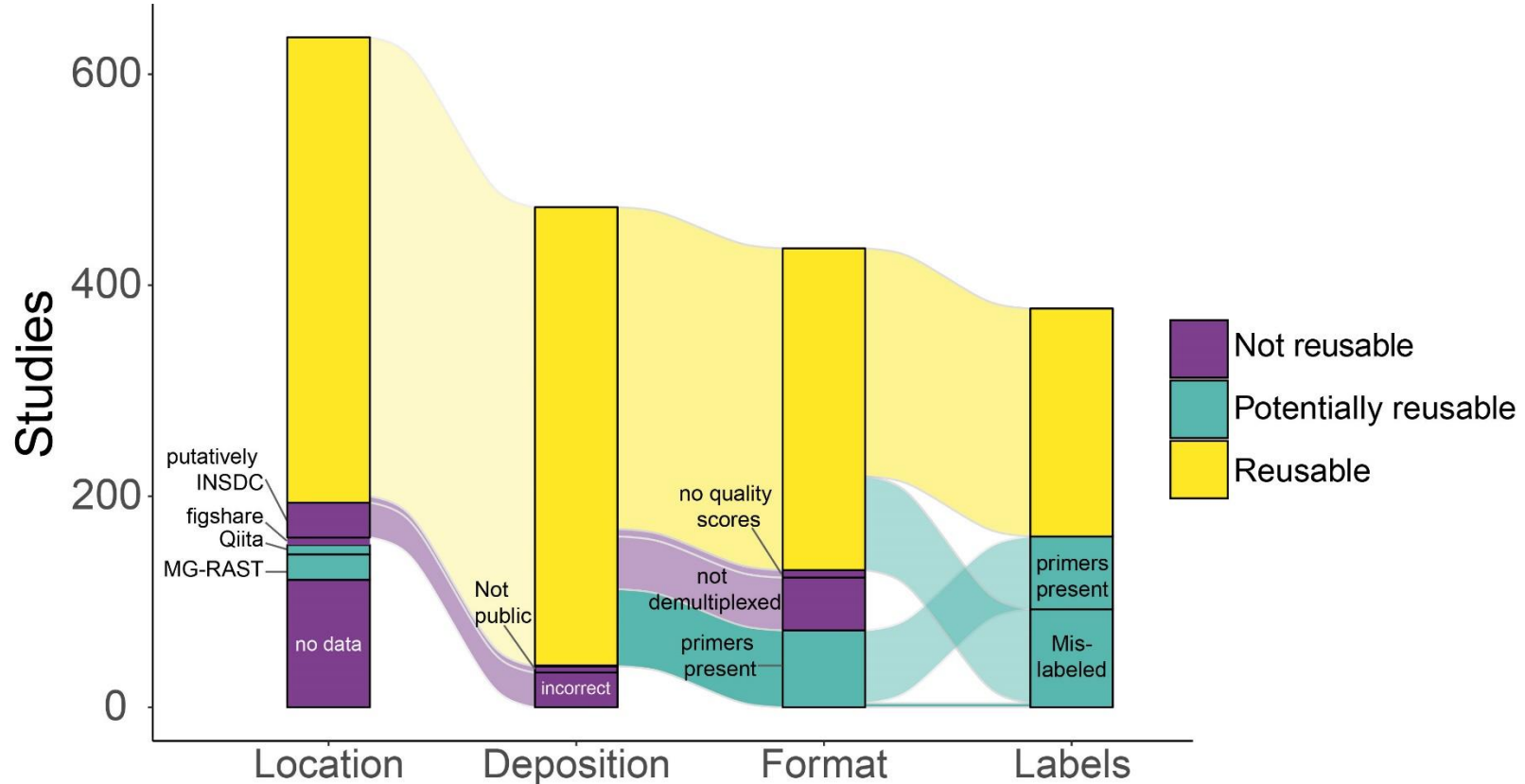


Maximilian Konzack



Anna Heintz-Buschart

How are the data?



MiCoDa: the Microbial Community Database



iBiD

Biodiversity data
support unit

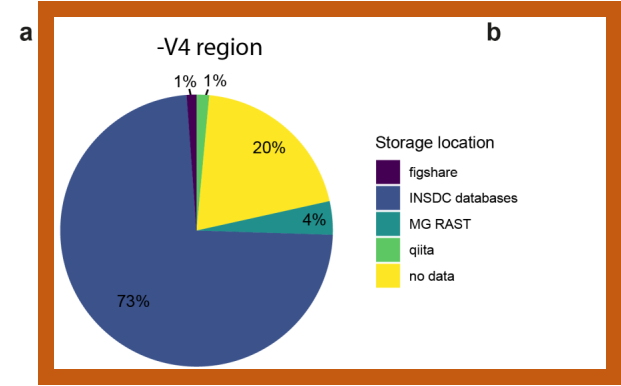
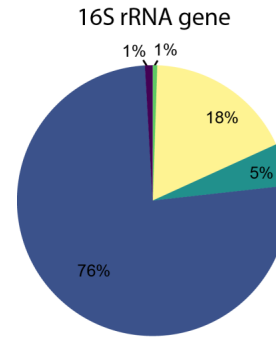
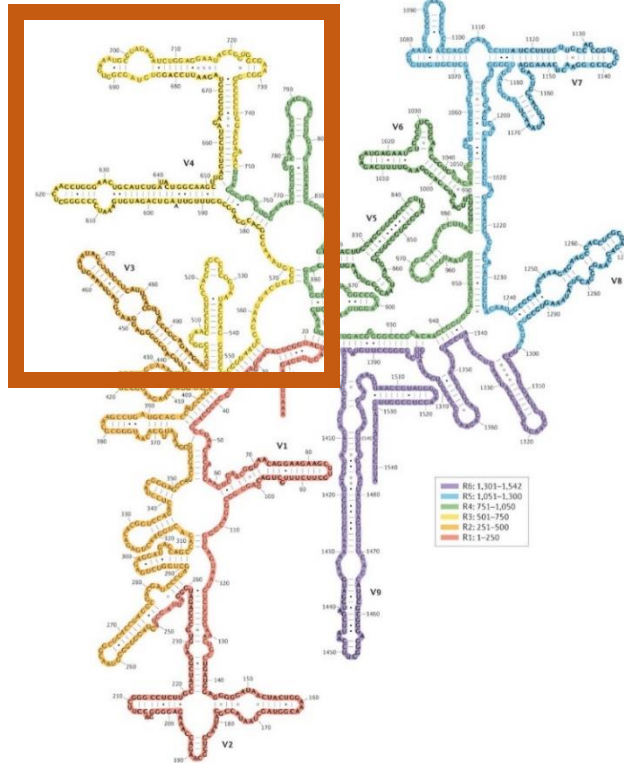


Desi Langer

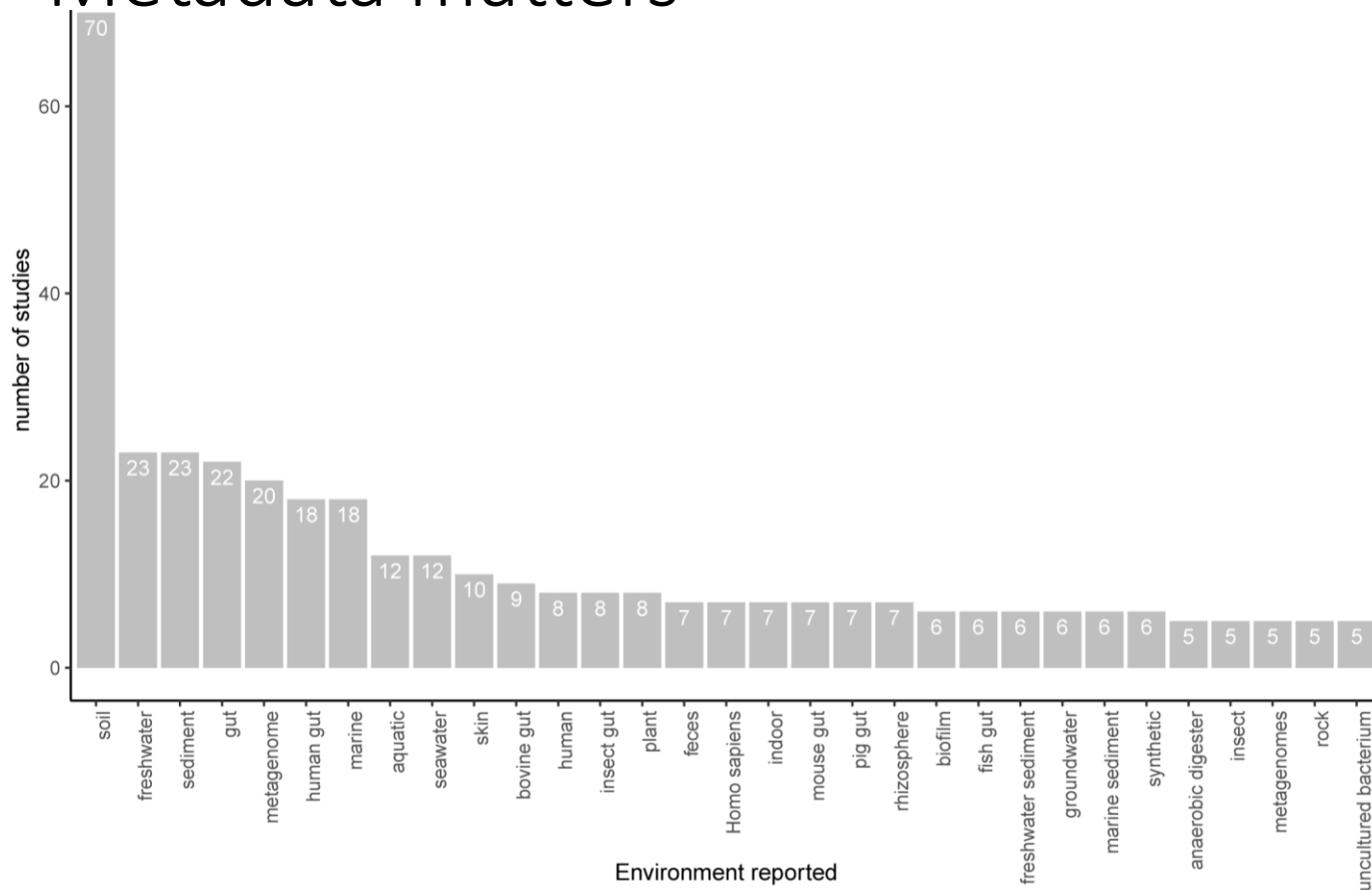


Anna Heintz-Buschart

Improving reusability of collected data

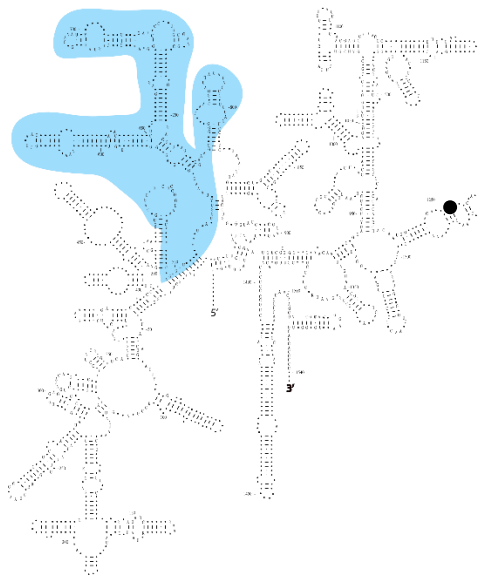


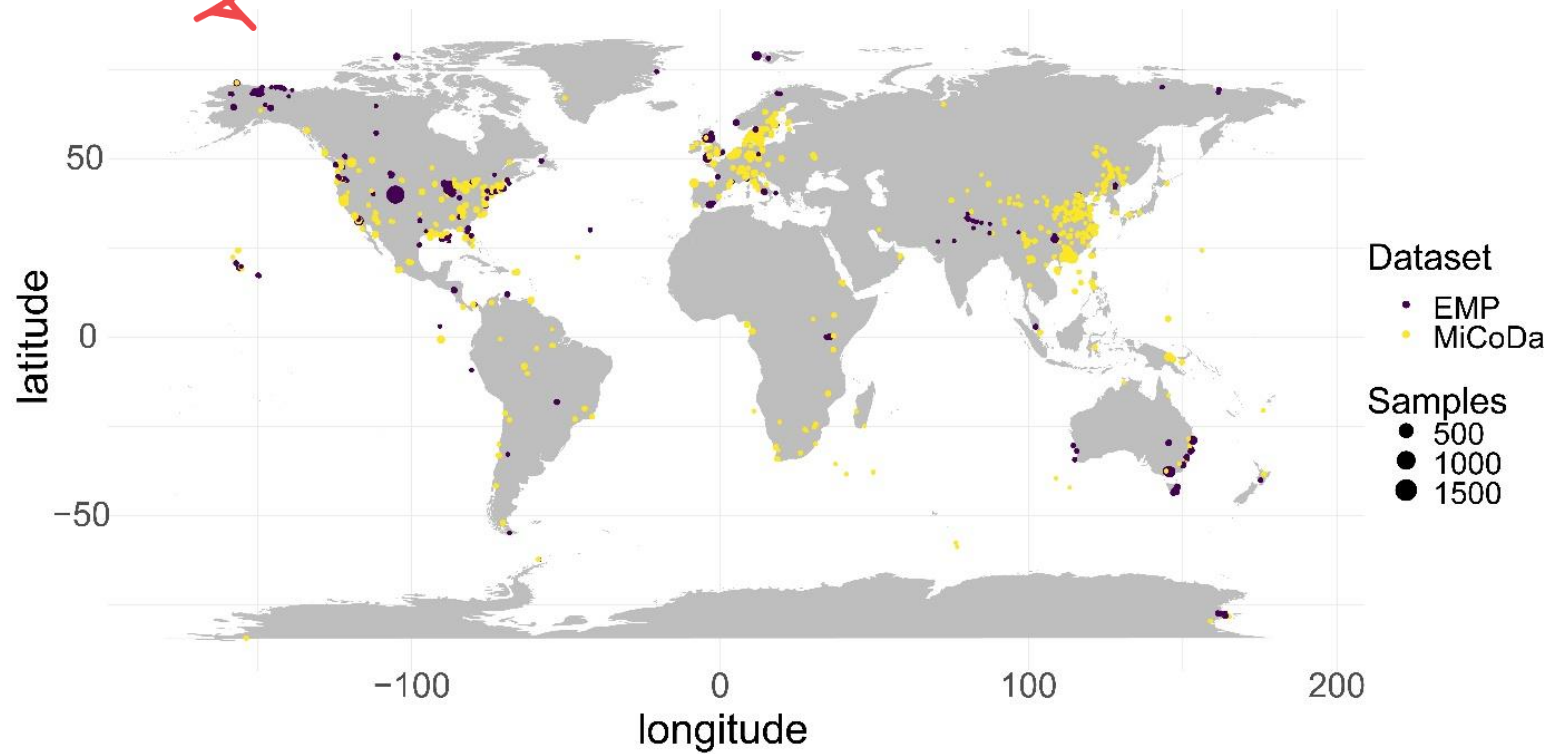
Metadata matters



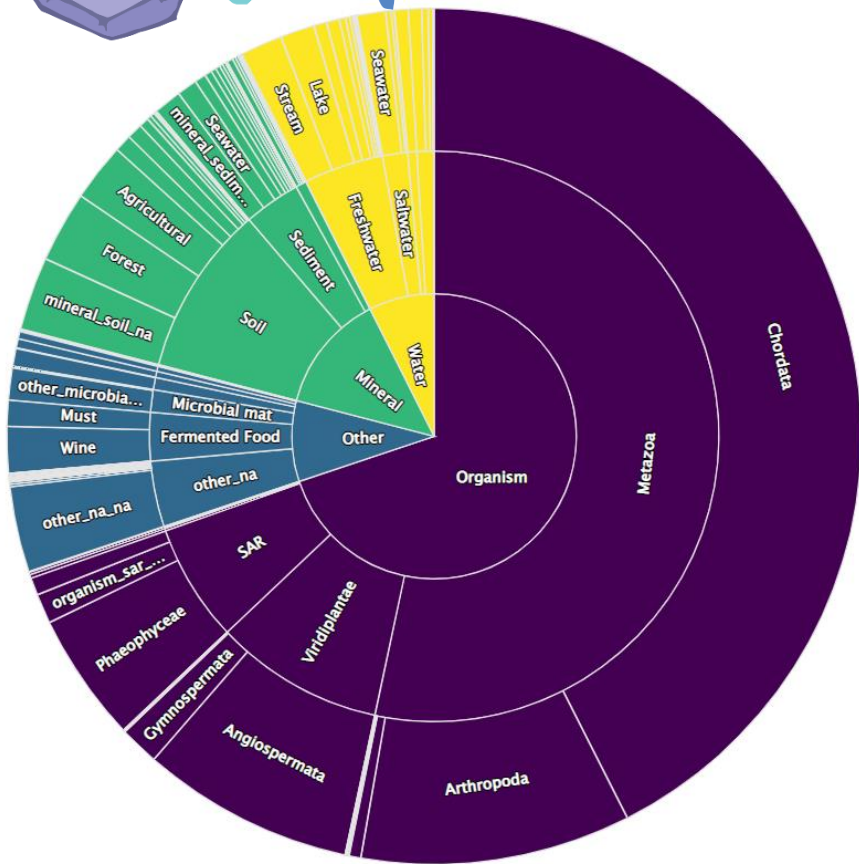


- We collected:
 - 252 articles
 - ~16,000 samples from control or observational studies
 - Same gene region (comparable for microbiologists)
 - Collected technical metadata (comparable for ecologists)
- Integrated with the Earth Microbiome Project (~15,000 samples)

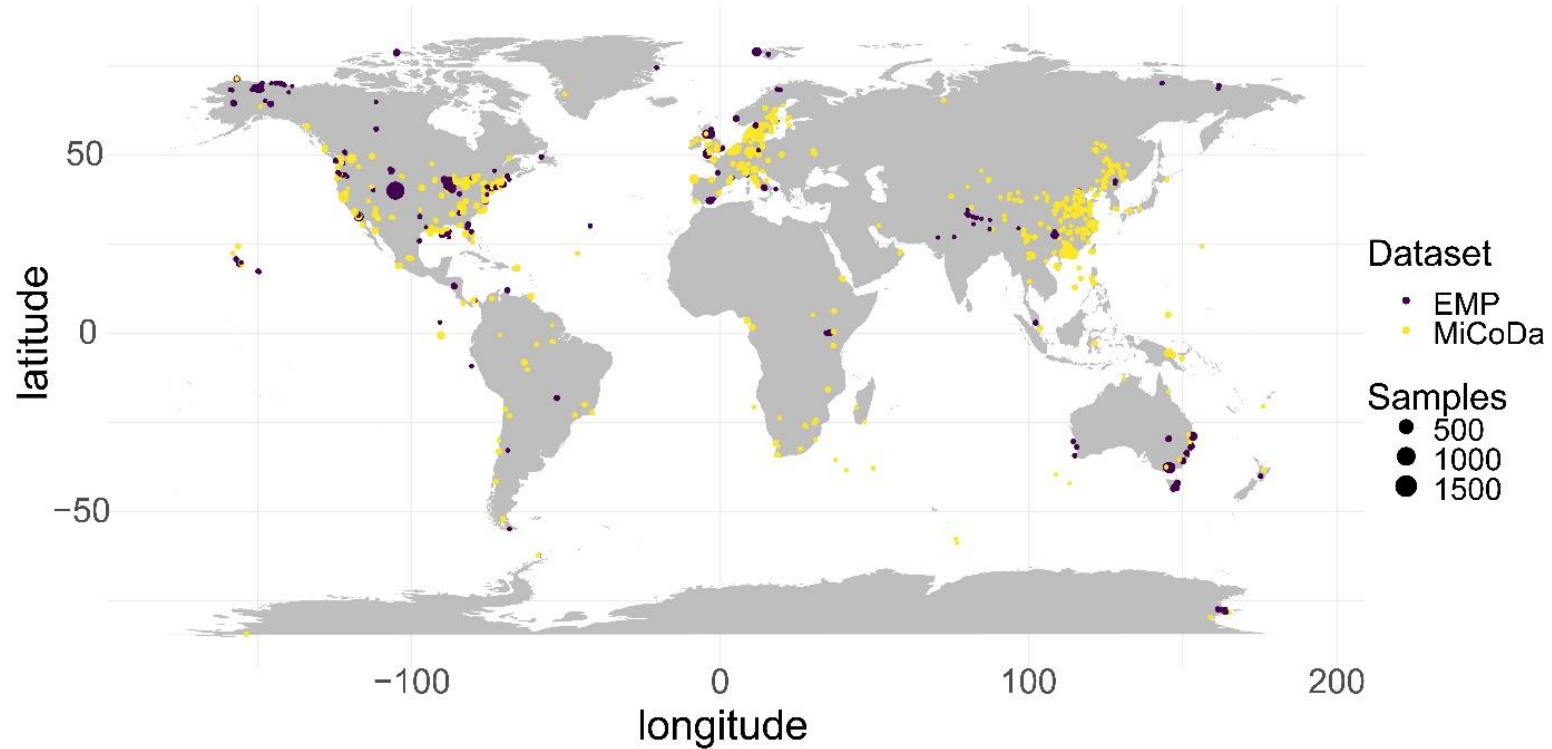




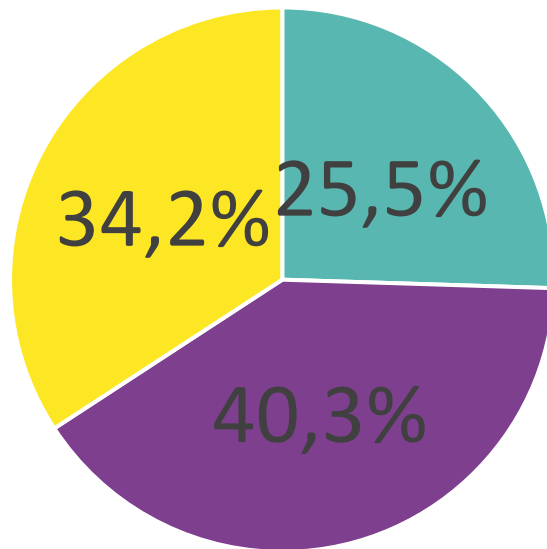
The logo for MCMCDA (Multi-Criteria Monte Carlo Data Analysis) features the letters 'M', 'C', 'D', and 'A' in a stylized, colorful font. The 'M' is purple, the 'C' is yellow, the 'D' is light blue, and the 'A' is dark blue. A red double helix DNA structure is positioned between the 'M' and 'C'. A purple, multi-faceted geometric shape is placed between the 'C' and 'D'. The entire logo is set against a background of a colorful, abstract pattern resembling a globe or a stylized map.



Data vacuum in the Global South



How can we improve data archives?



Could be improved
through training of
scientists

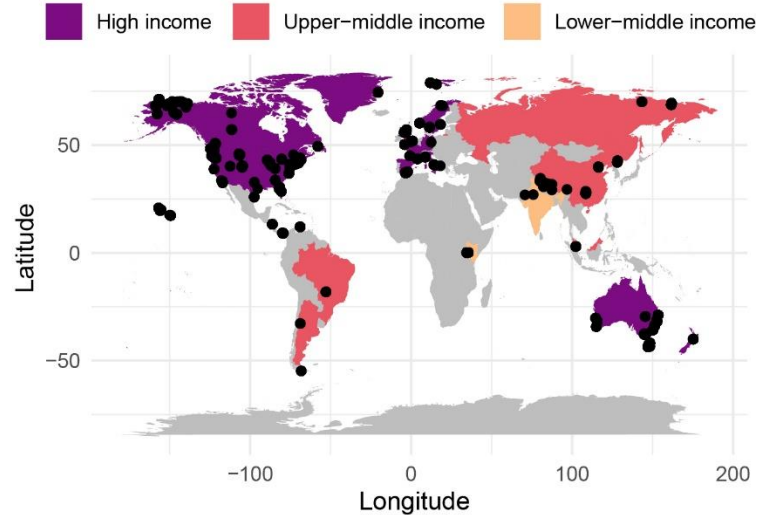
■ Partially reusable ■ Not reusable ■ Available

1. Data is expensive
2. Data archiving requires practice and guidance

a

The plot shows the percentage of ecological meta-analysis (p/author, p/country) for different income levels, categorized by % GERD of GDP. The x-axis ranges from 0.00 to 0.20. The y-axis lists income levels: NA, High income, Upper-middle income, Lower-middle income, and Low income. A color scale for % GERD of GDP is shown at the top, ranging from 1 (dark purple) to 4 (light green).

Income Level	% GERD of GDP	% of ecological meta-analysis (p/author, p/country)
NA	1	0.00
High income	1	0.00
High income	2	0.01
High income	3	0.02
High income	4	0.03
High income	1	0.04
High income	2	0.06
High income	3	0.07
High income	4	0.10
High income	1	0.21
Upper-middle income	1	0.00
Upper-middle income	2	0.01
Upper-middle income	3	0.02
Upper-middle income	4	0.06
Lower-middle income	1	0.00
Lower-middle income	2	0.01
Low income	1	0.00



Datathon 2022: Uruguay+Argentina

a



Symposium with six talks focused on: data reuse in other fields of ecology, the history of synthetic research in ecology, and the concept of 'Open Data' in biodiversity research.

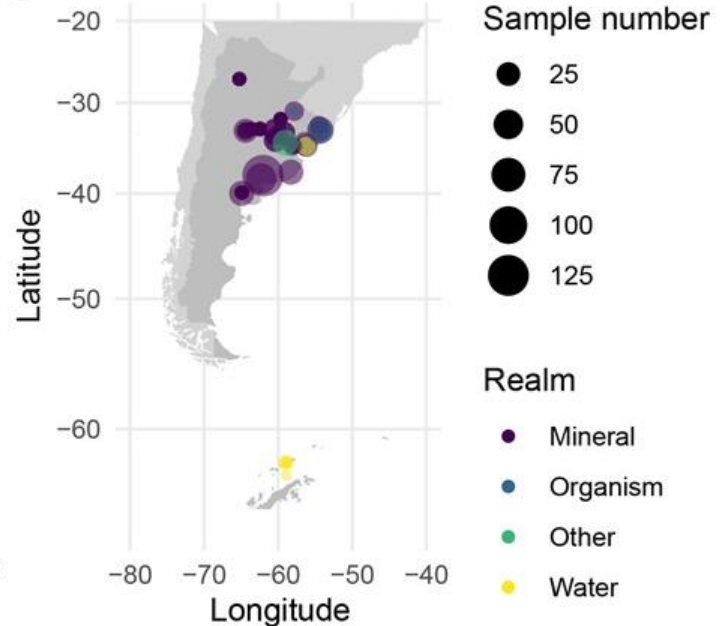


Collective data deposition to The National Center for Biotechnology Information (NCBI) with a unified format using step-by-step guides adapted to first-time users and custom metadata tables.



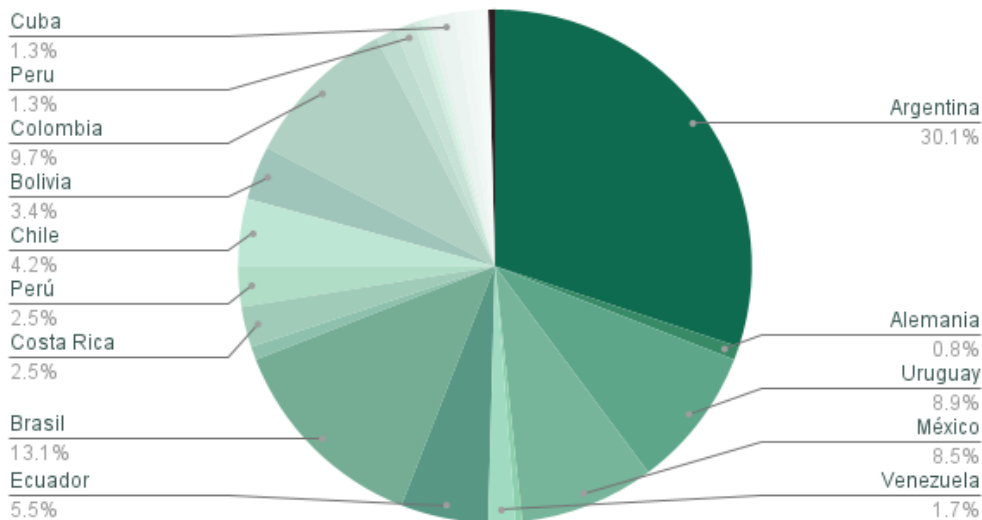
Virtual forum for proposing and organizing new projects using the data collectively deposited, and forming novel collaborations

b



Datathon 2023: Latin America

- ~250 participants
- >2000 samples archived
- Broad interest in data reuse



Final points

- Data archiving is a burden on researchers.
 - How can we make it easier?
 - Are data producers benefitting from archiving their data?
- There is a great need to close the gap between data producers and re-users.
 - Especially when these divisions align with access to funding.
- Solutions likely need to take into account local barriers.

Thank you!



iDiv



Thank you!

Slides: provided afterwards through HIFIS Events platform

Katharina Klemm
Julia Sperfeldt
bibinfo@ufz.de

Paul Meißner
paul.meissner@ufz.de

Stephanie Jurburg
stephanie.jurburg@ufz.de
 0000-0002-7701-6030

Ronny Gey
ronny.hey@ufz.de
 0000-0003-1028-1670