Module 4:
Sharing: Making Library, Archive & Museum Collections FAIR

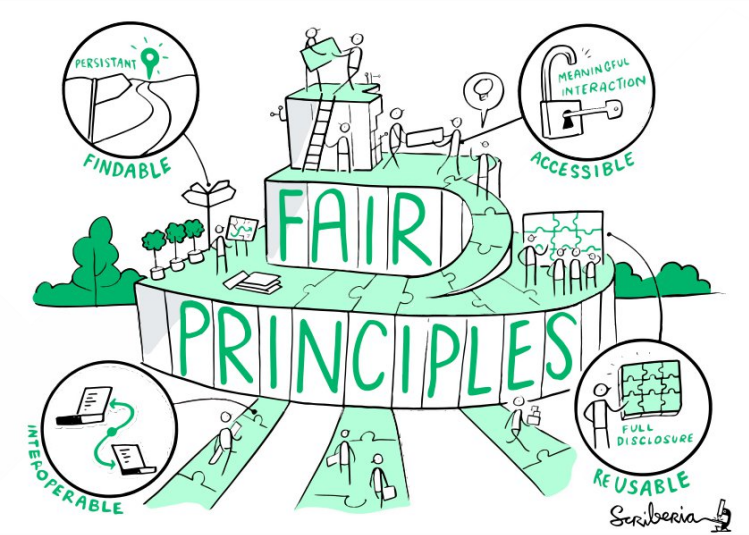
FAIR Implementation Starts at Home

Collection Data Curation and FAIR Playbook

Sharif Islam, Data Architect, Naturalis Biodiversity Center, The Netherlands
https://orcid.org/0000-0001-8050-0299

Digital Collection Curator Training Pilot
24th September 2024

Supporting eosc
Co-funded by the European Union
UK Research and Innovation



1. For both Humans and Machines (machine-actionability)
2. FAIR Principles apply to both data and metadata. FAIR can be applied to software, algorithms and workflows
3. FAIR does not necessarily mean open
4. These are principles not standards. Adoption and implementation can be gradual

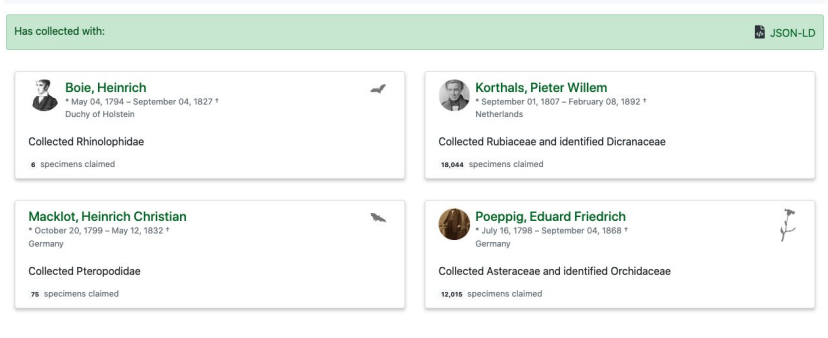
The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: https://doi.org/10.5281/zenodo.3332807.

Collection of links: facts, objects, people, locations, narratives ..

Tomistoma schlegelii (Müller, 1838);
collected in Indonesia
Naturalis Biodiversity Center (Creative Commons Public Domain Dedication, CC0)

RMNH.RENA.7935

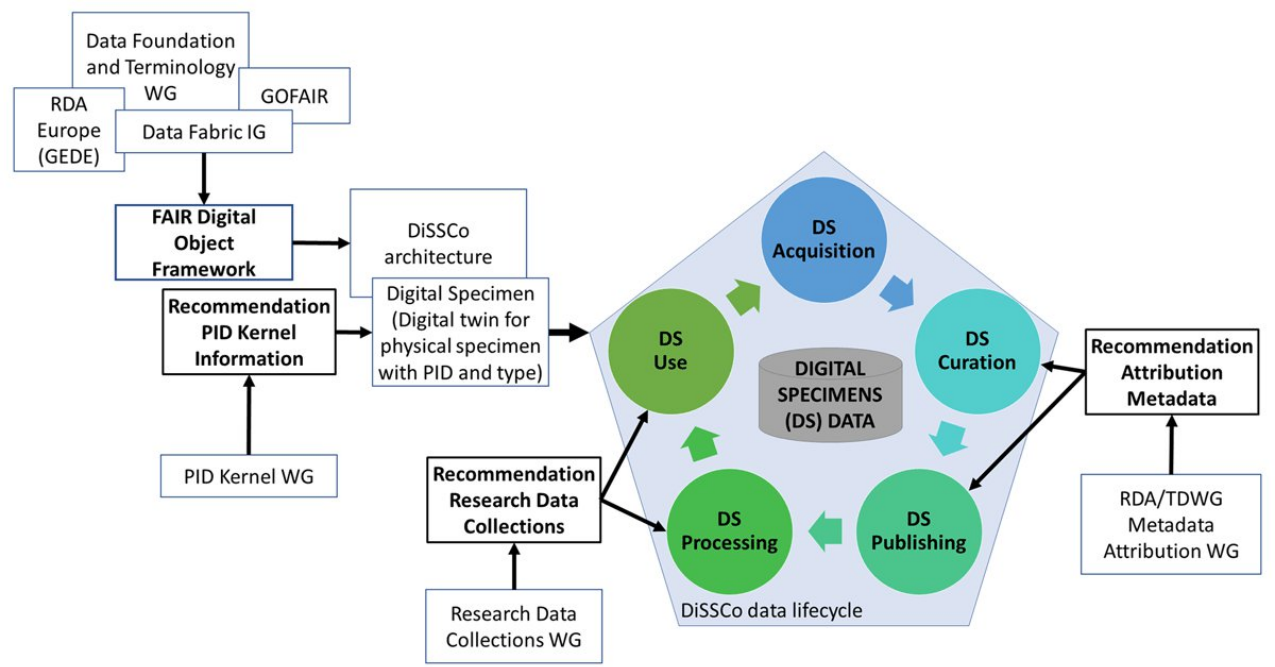
Salomon Müller Q61390 (wikidata)



Other actors:
Illustrators, local guides, non-European historical actors

Wikidata graph: https://w.wiki/6hBD

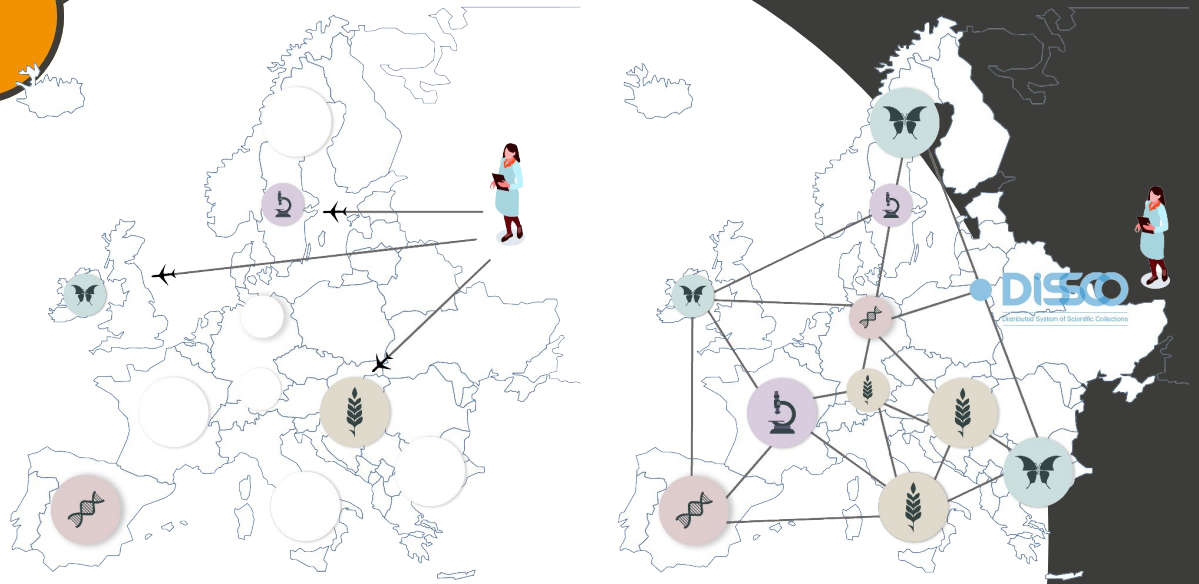
Digitisation and Digital Specimen Data Lifecycle



Source: Islam, S., Hardisty, A., Addink, W., Weiland, C. and Glöckler, F., 2020. Incorporating RDA outputs in the design of a European research infrastructure for natural science collections. *Data Science Journal*, *19*(50), pp.1-14. http://dx.doi.org/10.5334/dsj-2020-050

Supporting eosc
Co-funded by the European Union
UK Research and Innovation
Skills 4 eosc

Skills 4 eosc



DiSSCo services

ESFRI

Other RIs

Current model

Fragmented
Expensive (CO2 footprint)
Slow
Limited to physical access

Integrated Research Infrastructure (RI) model

Global access
Lower costs
Faster
FAIR data and AI ready

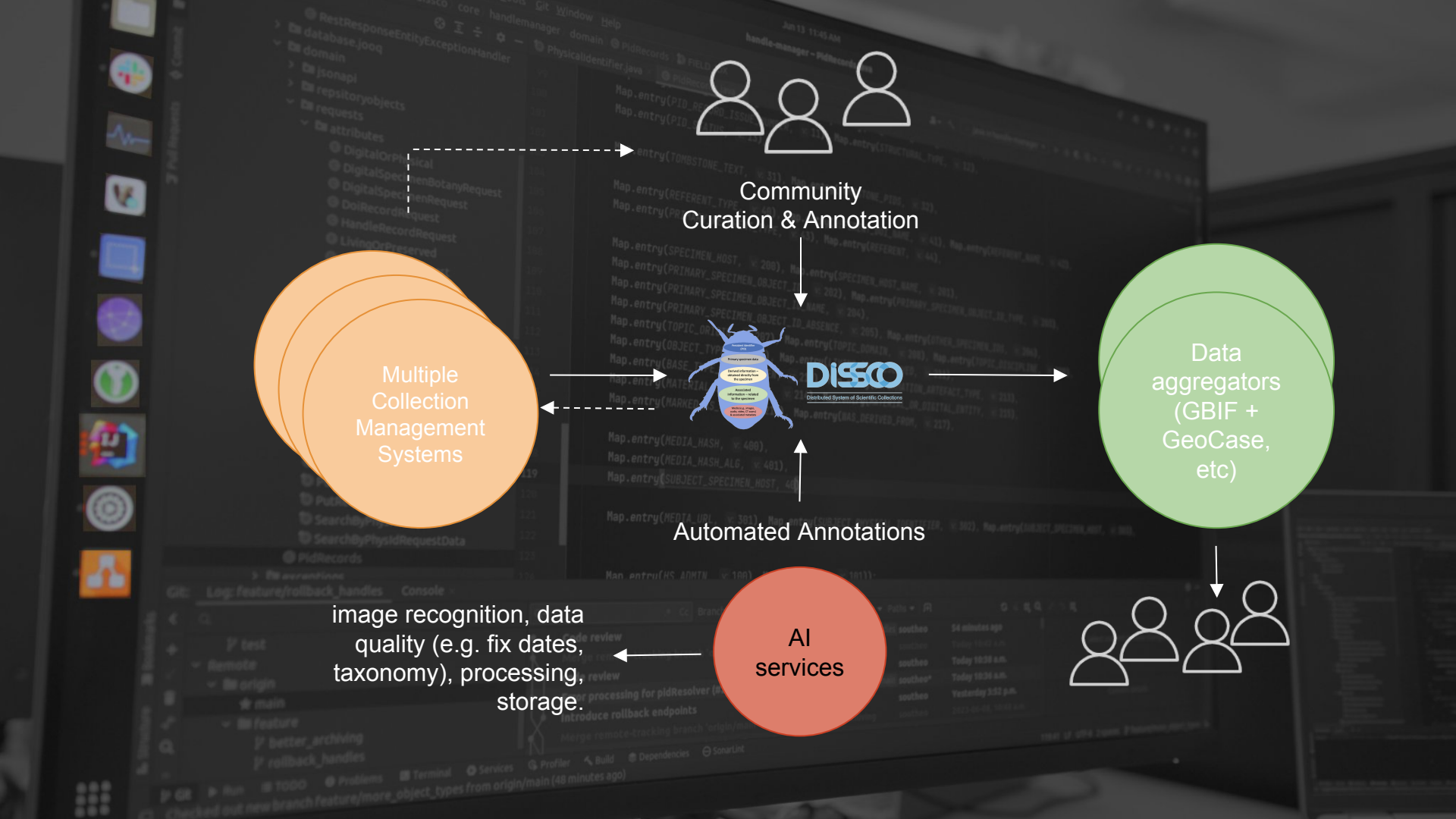
https://dissco.eu/

Funded by the European Union

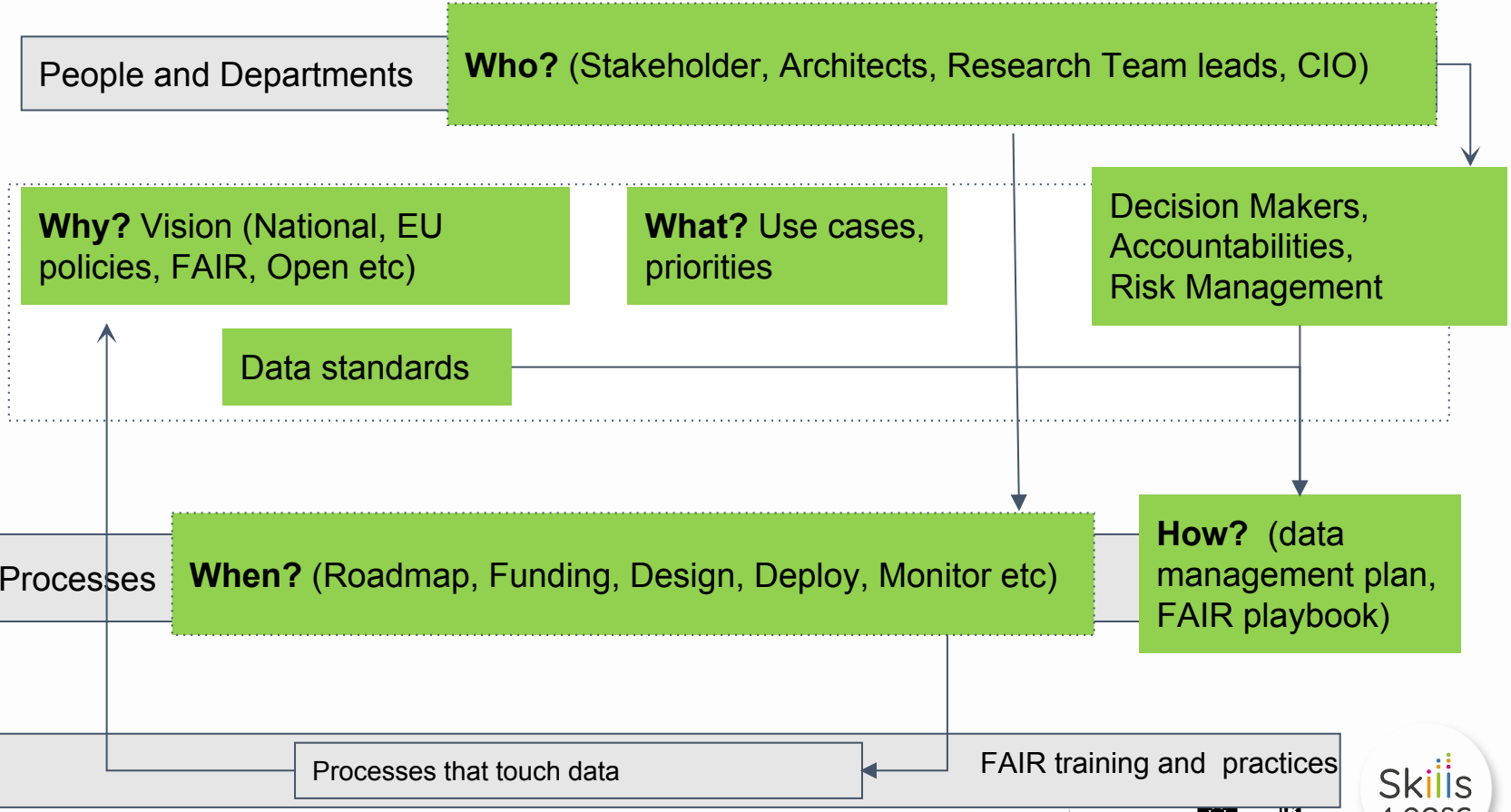
Supporting eosc

Co-funded by the European Union

UK Research and Innovation


Community
Curation & Annotation
Multiple
Collection
Management
Systems
DiSSCo
Distributed System of Scientific Collections
Data
aggregators
(GBIF +
GeoCase,
etc)
Automated Annotations
AI
services
image recognition, data
quality (e.g. fix dates,
taxonomy), processing,
storage.

Naturalis Research Data Management Tasks



Naturalis Biodiversity Center

Supporting eosc

Skills 4 eosc

What is a FAIR playbook?

A guideline for bringing data management into a actionable and manageable practice

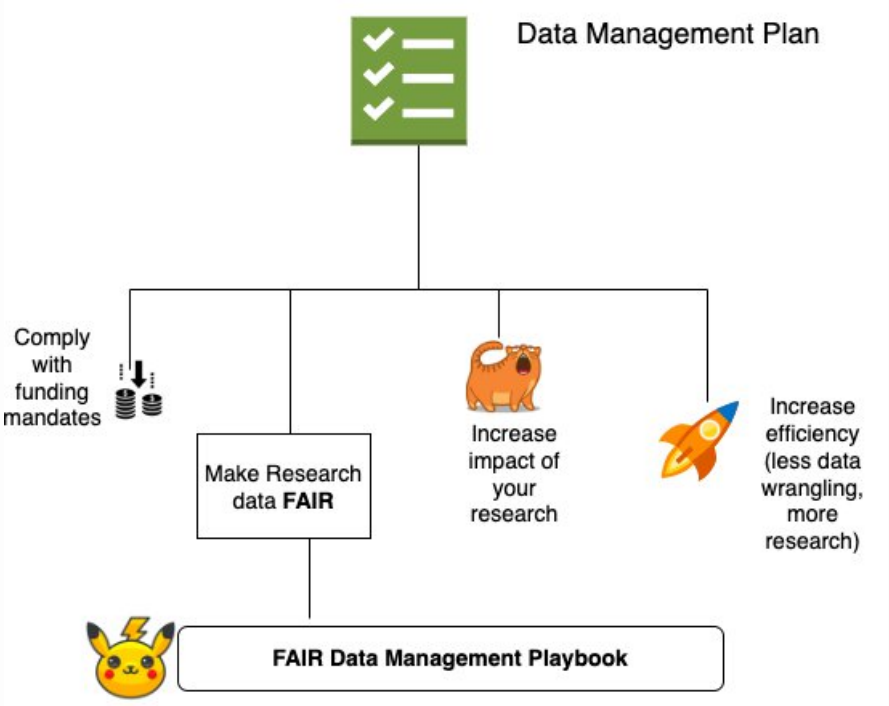


Image Credit: Naturalis FAIR data playbook

FAIR playbook in action

FAIR Research Data Management Playbook

Background 2

- Who is this playbook for? 3
- How to use this playbook? 3
- A few external resources: 3

Play One: Creating and Collecting data 4

Play Two: Processing and interpreting data 5

Play Three: Preserving data 7

Play Four: Giving access to data 7

Play Five: Re-using data 8

FAIR playbook in action

Play Two: Processing and interpreting data

When to use this play?

- After data has been collected and you are ready for the next step
- When manual or automated data curation and fixing is needed
- You want to understand and the interpret data

Key activities for Play # 2

1. Think about temporary storage needs during your data processing phase.
2. Think about the external software, tools needed for processing and data validation. Are these open source, freely available? Do you need to purchase these tools?
3. Think about data backup needs (during and beyond the processing time frame).
4. Think about data versioning procedures.
5. Do you need statistical and visualisation tools? If yes, do you know where to go for help?
6. Are there external data links and repositories that you need to rely on for this part?
7. Do you need HPC or Cloud computing access? Do you need ICT support for this?
8. Do you need specific infrastructure deployment help such as creating virtual machines or docker images?

Resources related to this play

1. A Guide to Environmental DNA (eDNA) by Biomeme
2. Biodiversity Data Use
3. Versioning Data Is About More than Revisions: A Conceptual Framework and Proposed Principles
4. Version control – FAIR in (biological) practice

Common pitfalls to avoid

) ahead for the resources you need to work with your data.
u need support from Naturalis ICT, contact them sooner than later.

Challenges

- Lack of training and expertise in data curation, stewardship and FAIR implementation
- Limited budget allocated to digital skills and data curation activities (DiSSCo partner capability survey, 2017)

Working towards solution

- Easy to use, modular FAIR playbook implementation (needs buy-in from management and department heads, having a data steward helps)
- Coordinated training and collaboration on FAIR data curation and stewardship (targeted for collection management)

Working towards solution

- Example initiative: *Community Action: Planning for Specimen Management in Funding Proposals* https://doi.org/10.1093/biosci/biae032 *from The Biodiversity Collections Network (BCoN, U.S.A)*

  - *we recommend that funding agencies require SMPs for biological specimens collected or generated during government-funded research and that SMPs provide detailed information regarding the requirements for collection, digitization, curation, long-term maintenance, and funding of accessioned specimens as a part of all research projects in which specimens of any type are generated or collected.*

Conclusion

- Starts with you and your organizations
- Requires coordinated effort, commitment, and dedicated budget
- User friendly, sustainable data management processes and tools

Acknowledgement

Naturalis Research Data Management working group
DiSSCo Development team, Naturalis
DiSSCo Coordination and Support Office

Skills4EOSC has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101058527 and from UK Research and Innovation (UKRI) under the UK Government's Horizon Europe funding guarantee, Grant No. 10040140

This presentation is released under a CC-BY 4.0 license

Funded by the European Union

Islam, Sharif. *FAIR Implementation Starts at Home: Collection Data Curation and FAIR Playbook*. Digital Collection Curator Training Pilot, 24th September 2024

Supporting eosc

Co-funded by the European Union

UK Research and Innovation