

# Bridging Research Infrastructures: Digital Specimens as a Catalyst for Cross-RI Synergies

**Workshop on Synergies and Overlaps between  
ESFRI ERICs, RIs and perspective RIs  
March 11, 2025  
Online**

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

Current global challenges demand that we *scale up and speed up* biodiversity research.

Natural Science Collections provide a crucial base of primary reference data but some changes are needed:

- Embrace new developments (e.g. Artificial intelligence) and new forms of data (e.g. genomics).
- Increase integration and interoperability to “break” the boundaries of individual institutions and work together.





Distributed System of Scientific Collections

**DiSSCo brings together European Natural Science Collections (bio and geological) into a digital infrastructure that will function as a unified collection for *all* of Europe.**

**200+ Institutions**

**23 Countries**

**1 business model:** shared curation and access policies across countries.



# DiSSCo's value proposition

One-stop digital infrastructure for:

- Physical and virtual access to the Collections (FAIR) data, facilities and expertise.
- State-of-the-art user services: digitisation on demand, machine annotation, research support...  
[diissco.eu/services](https://diissco.eu/services)

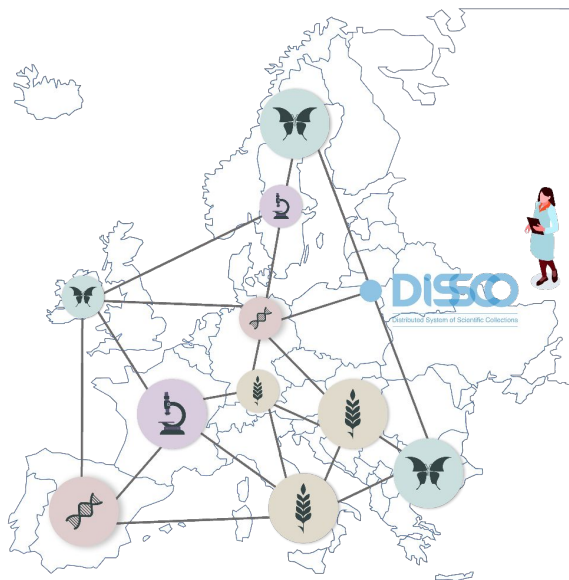
↓  
*A digital transformation of today's  
fragmented landscape of NHCs.*





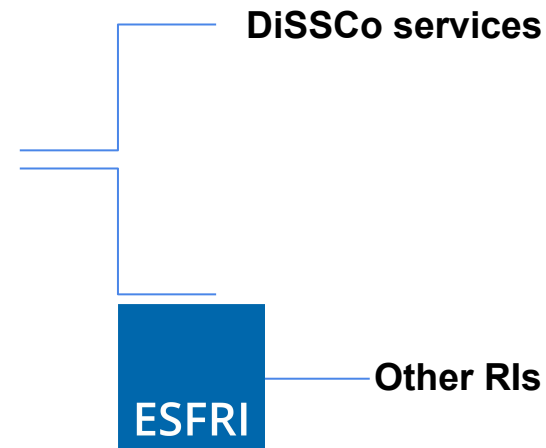
## Current model

Fragmented  
Expensive (CO2 footprint)  
Slow  
Limited to physical access



## Integrated RI model


Global access  
Lower costs  
Faster  
FAIR data and AI ready



# Building the bridge to the ERIC

## DiSSCo Timeline



A woman with blonde hair in a braid, wearing a white lab coat and orange gloves, holds a glass jar containing a preserved specimen. In the background, a man in a lab coat and safety glasses is working at a lab bench. The scene is a laboratory with large windows.

Enabling a new model

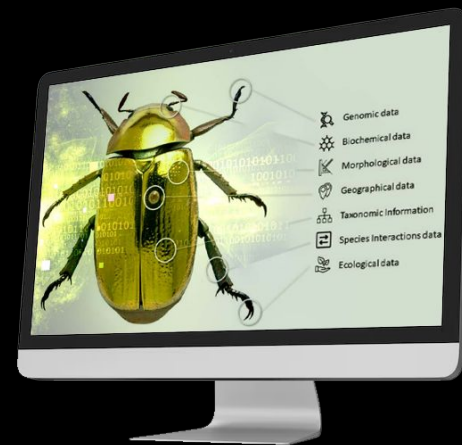
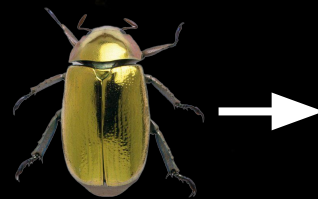
## DiSSCo's technical architecture

- It seeks to transform static records about specimens into dynamic, actionable objects in an integrated landscape.
- DiSSCo's bet: Digital Extended Specimen (based on FAIR principles, current standards, and global consultations)



# Digital Extended Specimen

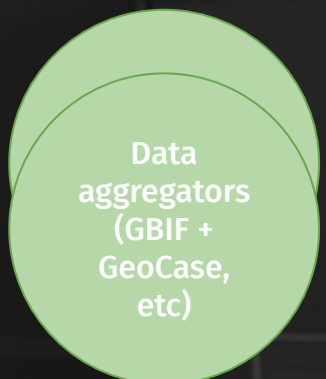
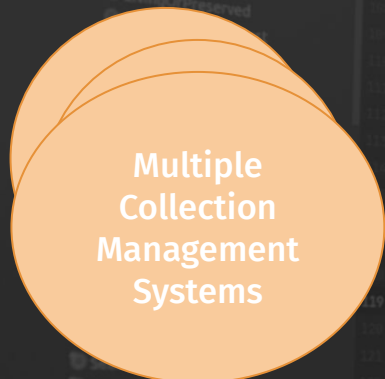
A FAIR digital object that acts as a “Digital Twin” of a physical specimen, but one that includes links to all known information derived from the specimen and is actionable, i.e. can be curated and annotated by humans but also by machines (AI)



**FAIR** Findable, Accessible,  
Interoperable, Reusable

Hardisty, A.R., Ellwood, E.R., Nelson, G., Zimkus, B., Buschbom, J., Addink, W., Rabeler, R.K., Bates, J., Bentley, A., Fortes, J.A. and Hansen, S., 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *BioScience*, 72(10), pp.978-987.





Automated Annotations

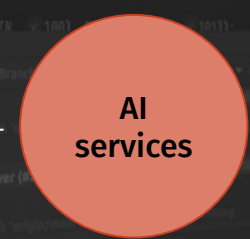
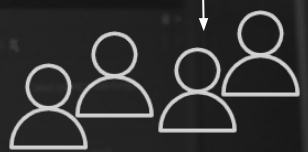


image recognition, data quality (e.g. fix dates, taxonomy), processing, storage.



# The building blocks

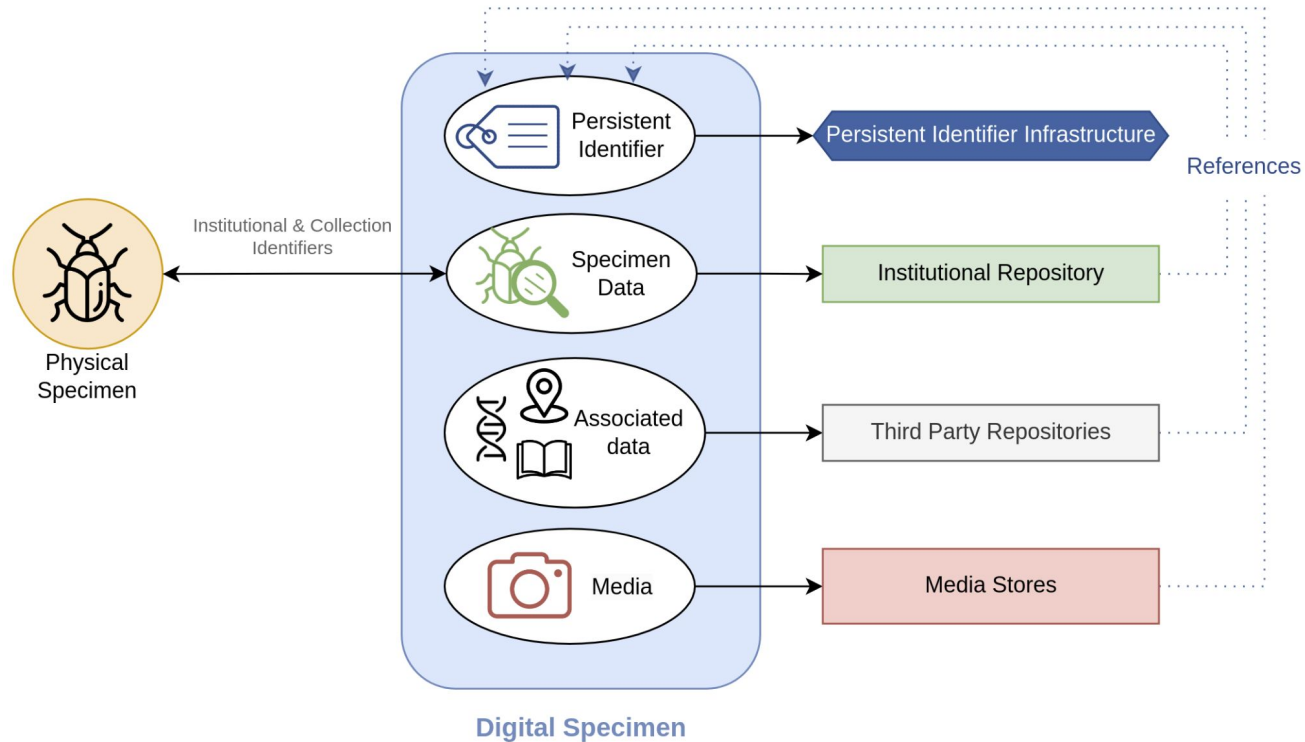


- A way to describe the Digital Specimen:
  - Data schema for open Digital Specimen (openDS: <https://github.com/DiSSCo/openDS> )
- Ingestion, data harmonisation, and annotation platform: <https://discover.dissco.eu/>
- Persistent Identifier service for Digital Specimen
- Authentication and Authorisation (ORCID integration), Provenance
- Machine Annotation Services

Image credit: [Konkapp - Flaticon](#)

<https://dissco.github.io/mas-developers-documentation/> "A Machine Annotation Service (MAS) is an automated service that annotates a target in DiSSCo. These services are scheduled by users on individual specimen or media in DiSSCo. What a MAS offers is broad. From sophisticated AI services to taxonomic services to linking to other infrastructures, MASs add value to natural science collections data in all sorts of ways."

# Digital Specimen and Persistent Identifiers



# Demonstrating scientific value

- Deeleman-Reinhold CL, Addink W, Miller JA (2024)  
The genera *Chrysilla* and *Phintelloides* revisited with  
the description of a new species (Araneae, Salticidae)  
using digital specimen DOIs and nanopublications.  
Biodiversity Data Journal 12: e129438.

<https://doi.org/10.3897/BDJ.12.e129438>

- Persistent reference of Digital Specimens using the DOI
- PID record metadata usage in ARPHA to provide context about the DOI
- Use of Nanopubs (RDF triples) to create a FAIR assertion. Example “**G0G-G7D-N5J is identified as *Chrysilla lauta* Thorell, 1887**”

[https://w3id.org/kpxl/pensoft/bdj/np/RADCQw2GTQ6neJ7ahujuTTrUDfkFpQQIAt\\_tmVewkBdwk](https://w3id.org/kpxl/pensoft/bdj/np/RADCQw2GTQ6neJ7ahujuTTrUDfkFpQQIAt_tmVewkBdwk)





# Minimal Viable Product

<https://discover.dissco.eu/>

Citation: Natural History Museum of Denmark. (2025). *Ploceus burnieri* N. E. Baker & E. M. Baker, 1990. Distributed System of Scientific Collections. [Dataset]. <https://doi.org/10.3535/KTW-4LF-X0S>

(alignment with DataCite metadata schema. resourceType as “dataset”)

The screenshot shows the dataset page for *Ploceus burnieri* N. E. Baker & E. M. Baker, 1990. The page is titled "Specimens" and "Ploceus burnieri N. E. Baker & E. M. Baker, 1990". It features a "Digital Specimen" tab, a "Geographical Map" showing the location in Tanzania, and a "Specimen Host" section. The "Specimen Host" section includes the name "Natural History Museum of Denmark" and the collection "Ploceus burnieri N. E. Baker & E. M. Baker, 1990". The "How to cite" section provides the citation: "Natural History Museum of Denmark (2025). Ploceus burnieri, N. E. Baker & E. M. Baker, 1990. Distributed System of Scientific Collections. [Dataset]. https://doi.org/10.3535/KTW-4LF-X0S".

DataCite  
Commons

10.3535/ktw-4lf-x0s

Works People Organizations Repositories

1 Work

**Ploceus burnieri N. E. Baker & E. M. Baker, 1990**

Natural History Museum of Denmark  
Digital Specimen published 2024 in DISCO  
Digital Specimen for the physical specimen hosted at Natural History Museum of Denmark.

Other Identifiers  
primarySpecimenObjectid: 582bffa9-820f-4488-af91-17a7e700ba90

DOI registered December 18, 2024 via DataCite.

<https://doi.org/10.3535/ktw-4lf-x0s>

Publication Year

2024

Work Type

Dataset

License

CC0-1.0

Field of Science

Biological sciences

Registration Agency

DataCite

Repository Type

Repository

# Cross RI synergies

- Use of persistent identifiers, standards mapping tools and templates are useful for linking specimens to sequences and annotations.
- See Biodiversity Genomics Europe output (<https://github.com/bge-barcoding/StayingMapped>). For example relationship between [https://dwc.tdwg.org/list/#dwc\\_materialSampleID](https://dwc.tdwg.org/list/#dwc_materialSampleID) and `mixs:source_mat_id` <https://genomicsstandardsconsortium.github.io/mixs/0000026/>

# Cross RI synergies

- DiSSCo provides **links** to digitised herbarium data and annotations. This includes data associated with the accession but also the herbarium voucher enables **identification** of the material to be verified, and taxonomic changes to be **tracked**, while also providing a valuable ecological and historical record.
- Annotations can come from various sources. **Phenotypic data** (e.g., morphological traits, stress responses, growth patterns) can be linked to a digital specimen as a structured annotation.

# Cross RI questions

- How can we reach agreement on common metadata profiles? For instance, taking advantage of schema.org, DCAT vocabularies, bioschema profiles.
- Check Green Deal Data Space discussions  
<https://www.greatproject.eu/great-project-resources/public-deliverables/>
- How can we create a shared, cross-infrastructure framework for tracking, negotiating, and reporting **Material Transfers** that works across seed banks, biobanks, and natural history collections — while also complying with the diverse legal landscapes like Nagoya Protocol, ITPGRFA, and national ABS laws? (*this is also a matter of common metadata profiles*)



# Thank you

**Acknowledgement:  
DiSSCo Development team, Naturalis  
DiSSCo Coordination and Support Office**



Distributed System of Scientific Collections



**Funded by  
the European Union**

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.