Building a world-class Research Infrastructure on RDA outputs

Dimitris Koureas
DiSSCo Coordinator
Director of International Biodiversity Infrastructures
Naturalis, Netherlands

Member, RDA TAB
@DimitrisKoureas
Natural Science Collections support **discovery** and **modelling** of all life on earth

**Europe: the global leader**

55% of the world’s assets with rich historical and global distribution
Trust lost when datasets disconnect from:

context in which they were created, or

communities who created them.

What builds TRUST in data?

Relevance

Provenance

Attribution

Completeness

Fitness-for-purpose

Agility

Branding (Datatyping)
What’s in a Museum specimen?

- Genomic data
- Biochemical data
- Morphological data
- Geographical data
- Taxonomic Information
- Species Interactions data
- Ecological data

Jointly this represents all evidence for Biodiversity
Unified evidence base, which enables semantic crosswalks between all data-types.
117 National Facilities
21 Countries

- Largest ever formal agreement between natural science collection facilities
- Centralised shared governance model already in place
- Synchronisation of facilities at access, data and policy level

a new business model: ONE EUROPEAN COLLECTION

- One European Collection of scientific assets
- Common Collections development strategy
- Economies of scope and scale
- Monitoring impact of collections (documenting ROI)
- Specialisation strategies (e.g. in alignment with national priorities, e.g. Smart Specialisation Strategies)
- Joint Research Agendas
Current model
Slow
Expensive
Inefficient
Limited

Integrated RI model
Wide access
Lower costs
Faster
New insights
Optimised
FAIR data

The first mass scale initiative to re-unite and serve genomic, chemical, geographical, morphological and taxonomic information and link it to collections objects
Institutional collections
Species distribution & genomics
Modelling / Prevention / Early detection
Facilities & information
Linked Data
Analysis / Interpretation Services

Urgent challenge

Economic costs of IAS for EU
€20 Billion / year
Kettunen et al. 2009

EXAMPLE: Alligator Weed
(Alternanthera philoxeroides)
Negative impact on native species, ecosystem services and infrastructure

Other Research Infrastructures

CC0
All data classes **unambiguously linked** to the **physical objects** they derive from.

Specimens representations become the centrepiece of the DiSSCo knowledge base – They are used as anchoring points for disperse data data classes.
Structure of a Digital Specimen Object (DSO)

PID

Kernel information
(metadata about specimen DO)

PID

PID

PID

PID

PID

metadata

metadata

metadata

metadata

metadata

metadata

metadata

content

content

content

content

content

content

Bit sequence

DO, = An envelope

Data Fabric & Terminology IG

Global Digital Object Cloud

Abstract global data architecture that is based on resolvable global persistent identifiers and globally accessible metadata
Digital Specimen: A dynamic “box” collecting links to all core information about a thing in one place.
Physical Object → Digital Surrogate

An actionable knowledge unit

- Genomic data
- Biochemical data
- Morphological data
- Geographical data
- Taxonomic Information
- Species Interactions data
- Ecological data
How do we represent a Digital Specimen and what do we include in the PID metadata?
Transforming digital entities to meaningful entities and serving them.

User

Web Services (specifications)

Data

Metadata (FAIR)

Context

Digital Object

Data Layer

Supplier

Research Data Collections WG

API Requirements Document
DiSSCo service portfolio by 2025

1. e-Science services
   - A one-stop shop for services providing unified discovery, access, interpretation and analysis of complex linked data

2. Physical and remote access services
   - A universal harmonised physical access service and digitisation on demand service

3. Support & Training services
   - Integrated user support desk and implementation of multi-modal training programmes to enhance skills & competencies
Simple DiSSCo implementation timeline

**Proposal Phase**
- 2015 - 2018
  - Governments political and financial commitments

**Preparatory Phase**
- 2018 - 2021
  - Optimum Implementation Readiness Level (IRL)

**Construction Phase**
- 2020 - 2023
  - Infrastructure in place

**Transition Phase**
- 2023-2024
  - Legal entity in place

**Operation**
- 2025 – 2035
  - Full Service portfolio
A new Data Research Infrastructure provides opportunities for a new approach to our community challenges.
Preparing through a complex programme
But the most valuable interaction comes through sharing of People
One world → One collection

Find out more at www.dissco.eu