# Show and Tell:
# **BioDT and DiSSCo Perspective**

## Claus Weiland, Alex Wolodkin & Jonas Grieb

Claus Weiland / Senckenberg, DiSSCo RI
FAIR-IMPACT Workshop: Why Mappings Matter and how to make them FAIR?
April 13. 2023

# DiSSCo - (A) Distributed System of Scientific Collections

- **Major focus: Natural Science Collections'** transformation from 16c natural history cabinets into hubs for bio- and geodiversity knowledge

- Encompasses besides preserved specimens living collections like botanical gardens, seed banks, research platforms and tissue/cell culture banks.

- **DiSSCo's objective** as RI: digitally unify today's fragmented landscape of European NSC into a single virtual data collection under common curation, access policies and practices.
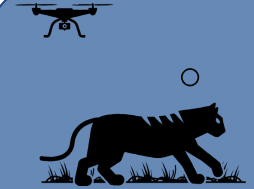
   **170+** National Facilities,
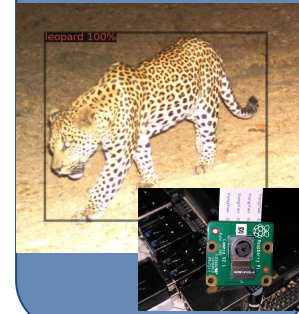   **23** Countries,
   **1** European Collection

**Physical Specimen Collections**

**Subject DBs & Aggregators**

edapho base

**Data from research platforms**

# DiSSCo's core data model: The Digital Specimen

- "A Digital Specimen [...] represents the sum of information on the Internet about a natural specimen object. The Digital Specimen acts as a processable digital twin on the Internet for the physical specimen…"
  https://know.dissco.eu/item/141

- FDO: **Unified conceptual data model**, **enables interoperability by PID/model layer** developed within the framework of FDOF (www.fairdo.org) and **community defined resource layer** (typed operations and (meta)data)

- Strong focus on self-contained exploration and processing of structured data by machines based on the interplay of descriptions/attributes and services (**machine actionability** as core objective of FAIR).
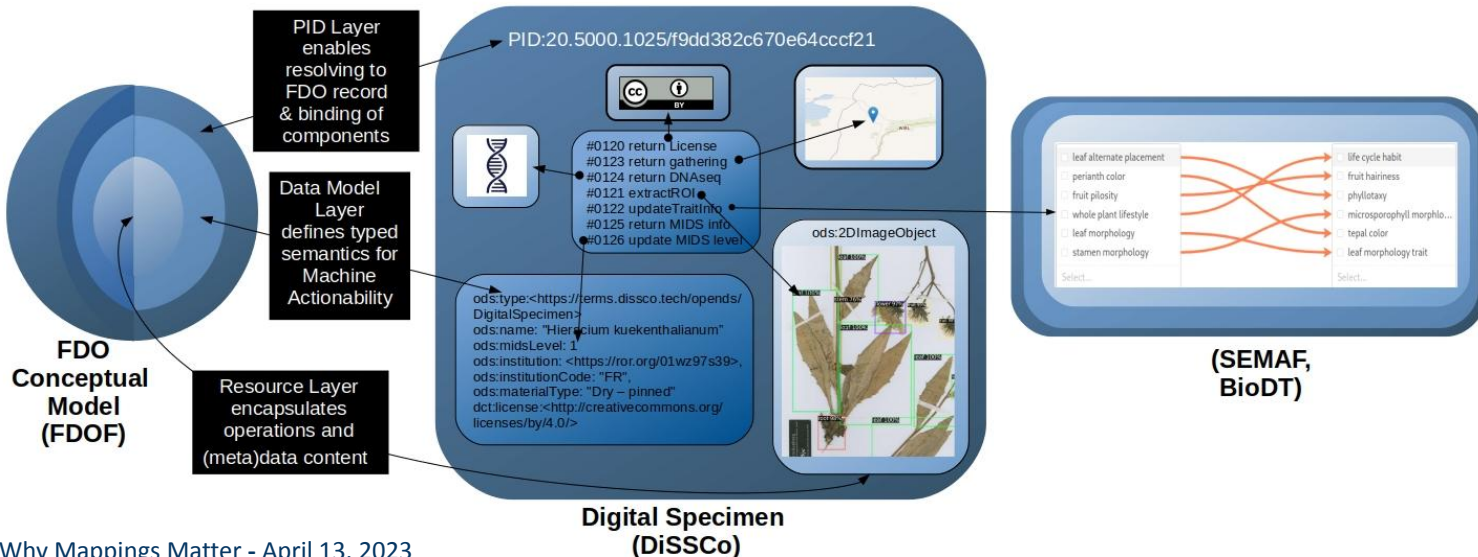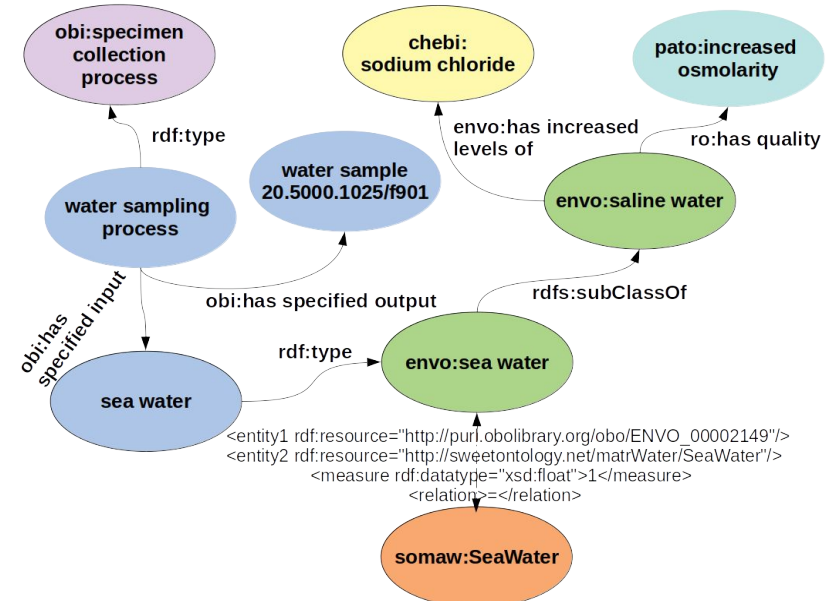
# Biodiversity Digital Twins in DestinE

- DestinE: Develop the core simulation platform for an accurate **digital model of the Earth based on thematic Digital Twins** to monitor and simulate natural and human activity and support European environmental policies (https://digital-strategy.ec.europa.eu/en/policies/destination-earth).

- BioDT: **The Game Engine for** advanced modelling, simulation and prediction capabilities with regard to **biodiversity grand challenges** (biodiversity loss due to climate/land use, alien invasive species, zoonotic/vector-based diseases).

- **Mobilize data** from environmental-related (E)RI(C)s like DiSSCo, LifeWatch, eLTER and GBIF **for state-of-the-art modeling on HPC platforms**.



Dynamic global vegetation model
(DGVM, T.Hickler/S.Scheiter, SGN)
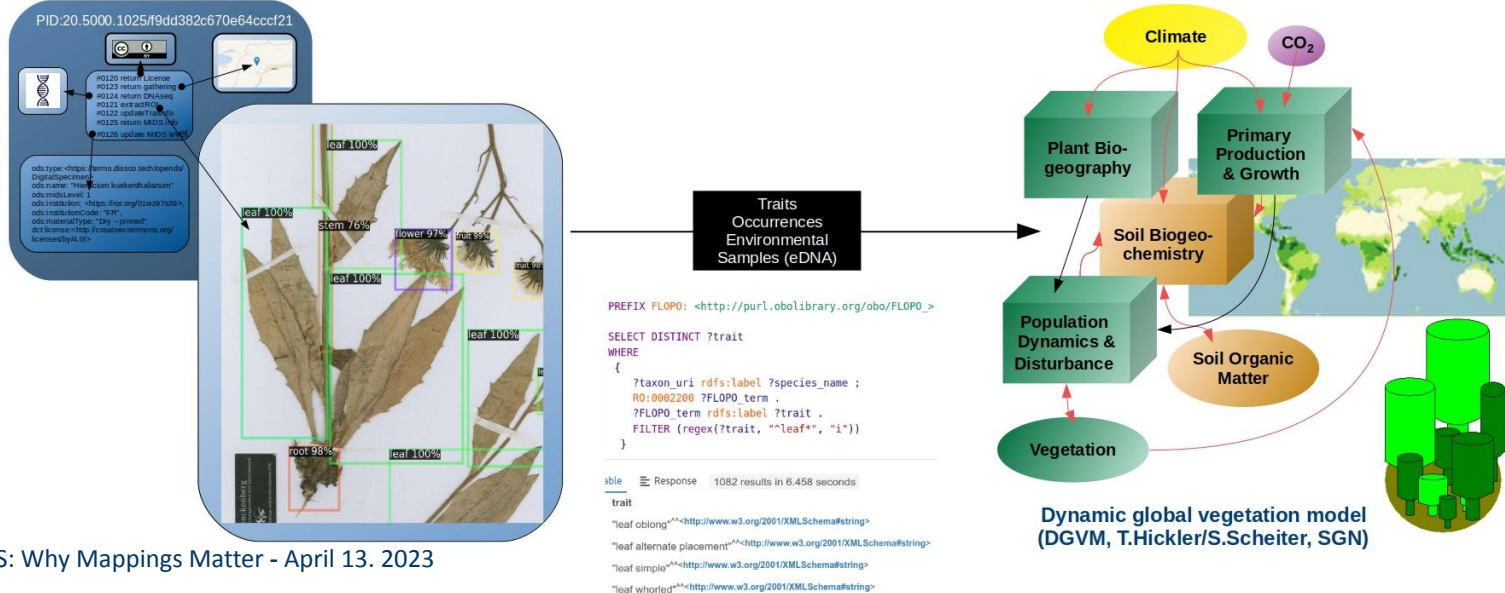
# Mappings in biodiversity semantics

- Often multiple semantic artifacts for a given domain (biodiversity: ABCD, DwC, agrosystems: AGROVOC, FoodOn) → mappings of related or similar concepts are necessary for data integration, enhance knowledge discovery.

- Alignments/crosswalks develop over time, start as baseline mappings, are partially curated ("silver standard") or completely reviewed ("gold") by domain exerts (Dahdul 2018, doi:10.1093/database/bay110).

- Consistency problem if adding any equivalence axioms to OBO ontogies Ie.g. between BCO and  DwC (e.g. dwcobo:Location owl:equivalentClass bfo:site)

- Mappings as workaround?

|   | | | | | |
|---|---|---|---|---|---|
| < | grease ice layer | ENVO_03000074 | seaice | ealmCryo/SeaIce | subclass of sea ice. (ENVO refers to freshwater ice, but sea ice is best SWEET match) |
| (=) | convergent plate boundary | ENVO_01001100 | subductionzone | realmLandTectonic/SubductionZone | equivalent (with a few exceptions) |
| (=) | transform plate boundary | /ENVO_01001101 | transform plate boundary | realmLandTectonic/TransformPlateBoundary | Found by AML equivalent |



```
<entity1 rdf:resource="http://purl.obolibrary.org/obo/ENVO_00002149"/>
<entity2 rdf:resource="http://sweetontology.net/matrWater/SeaWater"/>
       <measure rdf:datatype="xsd:float">1</measure>
           <relation></relation>
```

Mappings from Karam (2020)
doi:10.1017/S0269888920000132

Model changed from Walls (2014)
doi:10.1186/1944-3277-9-17

# Mappings as *pragmatic translations*

- Mappings are often pragmatic 'translations' between metadata descriptions and observation measurements and actually targeted at specific interoperability goals

- Require a data model to preserve the context (lab notebook, sample preparation techniques, sensor configurations, etc).

- Share and publish crosswalks using open registries and increase FAIR compliance (PIDs, specifications like SSSOM instead of xrefs)

- A data model and infrastructure draft was proposed with SEMAF (Broeder 2021 doi:10.5281/zenodo.4651421)



| fotoMerkmalWert | fotoMerkmal | fotoMerk | OBO ID | FLOPO label | remarks |
|---|---|---|---|---|---|
| glandular hairs | Important diagnostics | Habit | FLOPO:0007115 | trichome glandular | there is also "trichome glandular": FIXEd |
| red | Flower - Colour | Flower | FLOPO:0007599 | flower red | |
| spiny stems or leaves | Important diagnostics | Habit | FLOPO:0015264 | stem spiny | we may need "whole plant spiny": Added, FLOPO:0907311 |
| orbicular | Leaf - Form | Leaf | FLOPO:0017811 | leaf lamina orbicular | leaf orbicular not available |
| palmately compound | Leaf - Structure | Leaf | FLOPO:0018499 | leaf digitate | is the same as "leaf palmately compound" FLOPO:0900068: FIXED |
| pinnately compound | Leaf - Structure | Leaf | FLOPO:0907004 | leaf pinnately compound | FLOPO:0023274 might work, but the division into part and quality doesn't fit |
| capsule | Fruit - Type | Fruit | FLOPO:0900012 | fruit type capsule | here, adjectives like capsular would also be possible |
| legume | Fruit - Type | Fruit | FLOPO:0900013 | fruit type legume | |

FLOPO: (Hoehndorf 2016)
doi:10.1186/s13326-016-0107-8

# mapping.bio ↔ SEMAF

- Intended as "light-weight service" to read semantic artifacts, visualize them, add mappings as graphical connections and store the mappings as JSON-LD in a cordra (https://cordra.org) repo.

- Does not aim to provide substantial parts of the SEMAF infrastructure (registries)
  ○

# Many thanks!