



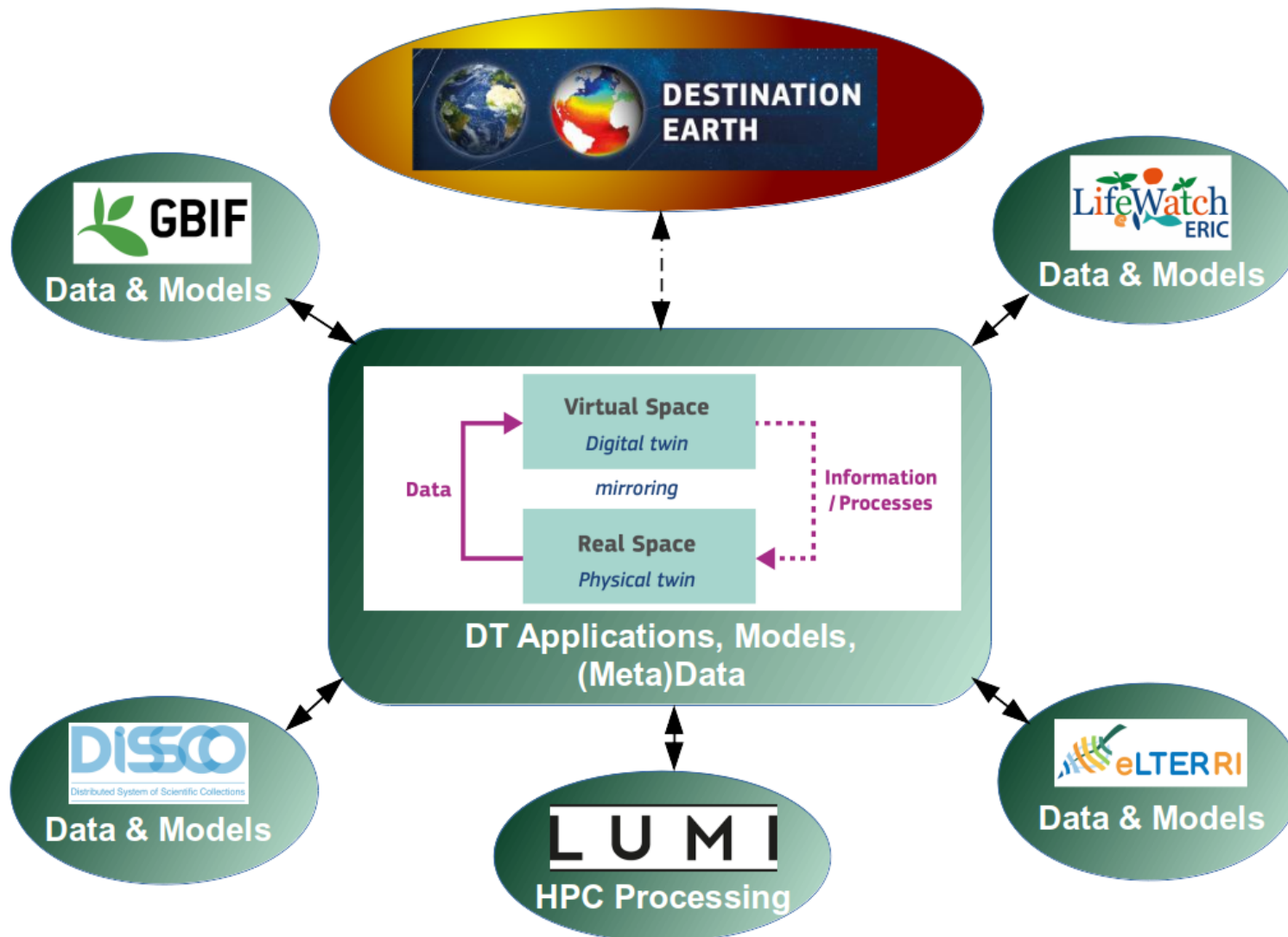
Enabling connections to EOSC, the FAIRCORE4EOSC and BioDT collaboration



**Tommi Suominen, FC4E and CSC - IT Center for Science
Contributors Hanna Koivula and Claus Weiland**



BioDT Data & Data Streams



Data Exchange from 4 Ris:

Biodiversity data is exchanged mainly with DarwinCore (DwC) standard and Ecological Metadata Language (EML) is used to describe the environment conditions at the study site. In addition many data elements can/should be *expressed with external vocabularies*

- **GBIF** - DwC and EML (+ external vocabularies are possible but not indexed in gbif.org)
- **LifeWatch** DwC, EML + vocabularies
- **DiSSCo** – DwC + EML (still evolving RI)
- **eLTER** – No chosen standards, EnvThes vocabulary (still evolving RI)

Biodiversity and Natural Environment Data

- ***Biodiversity and ecosystems*** cannot be studied without assessing the impacts of ***changing environmental conditions***.
- To bring together biodiversity data (ie data about species) BioDT needs access to data about the environment circumstances from the data collecting site(s)
- GBIF aggregates this biodiversity data with the DwC standard, which includes elements and vocabularies for expressing species observations and data gathering event (including method)
- The environmental conditions of the location are given as dataset level metadata with EML

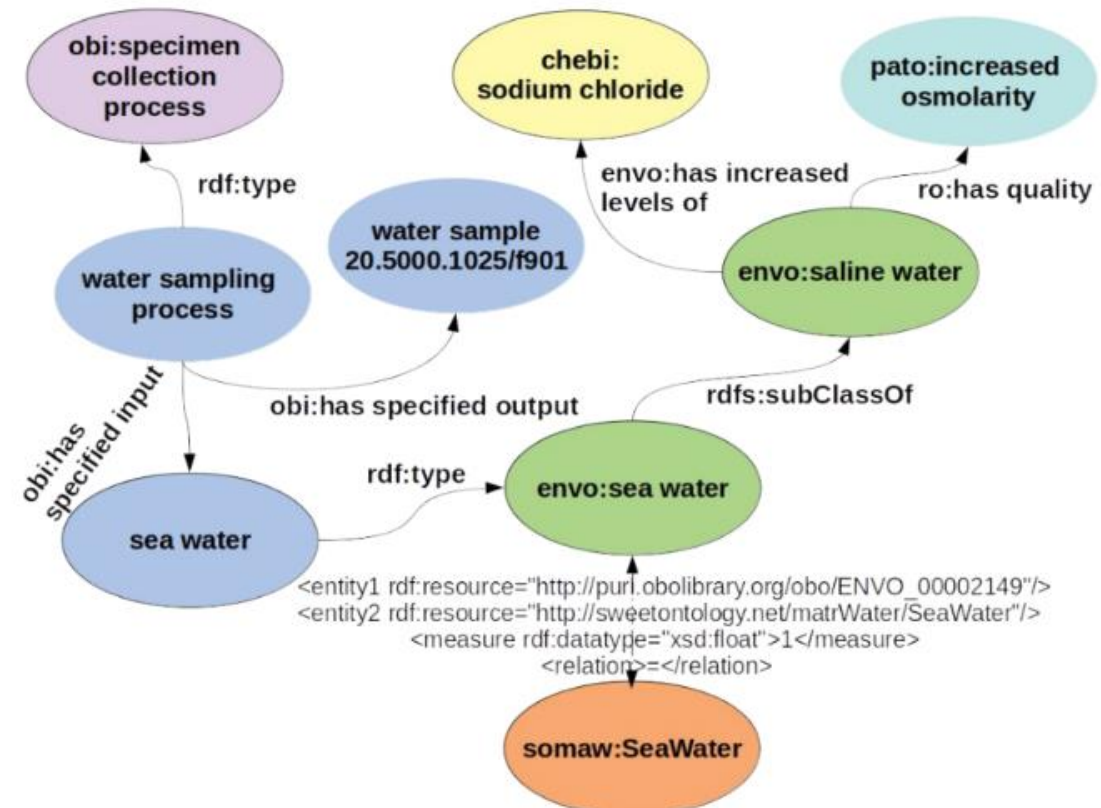
The EML-DCAT collaboration

- FAIRCORE4EOSC is developing the Metadata Schema and Crosswalk Registry (MSCR) aiming to create crosswalks between differing metadata schemas to bridge over the differences and allow conversion of metadata between differing schemas.
- FC4E aids BioDT in using this service in order to create a crosswalk between Data Catalogue Vocabulary (DCAT) 3.0 and EML
- BioDT aims to exchange metadata about sampling site natural environment in EML format, but also requires other vocabularies
- EML – DCAT conversion enables bridging data between diverse catalogues
- The BioDT is providing MSCR valuable information about needed features for the service and possibly pave the way for creating more complicated crosswalks
- Utilization of the MSCR both makes this bridging work more FAIR as well as making it more resource efficient by utilizing tools customized for the task.

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 experts (Dahdul 2018, [doi:10.1093/database/bay110](https://doi.org/10.1093/database/bay110)).
- Consistency problem if adding any equivalence axioms to OBO ontologies le.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	subduction zone	realmLandTectonic/SubductionZone	equivalent (with a few exceptions)
(=)	transform plate boundary	ENVO_01001101	transform plate boundary	realmLandTectonic/TransformPlateBoundary	Found by AML equivalent



Mappings from Karam (2020)
doi:[10.1017/S0269888920000132](https://doi.org/10.1017/S0269888920000132)

Model changed from Walls (2014)
doi:[10.1186/1944-3277-9-17](https://doi.org/10.1186/1944-3277-9-17)

The mapping.bio - MSCR collaboration

- **Integration of biodiversity data from heterogeneous sources including sensor-based monitoring tools like camera traps requires consistent mappings between different trait and environmental ontologies.**
- **SEMAF (Flexible Semantic Mapping Framework)** put forward a framework to foster sharing and publishing of pragmatic mappings and crosswalks - i.e. mappings representing translations between metadata descriptions and observation measurements
- **BioDT/DiSSCo developed [mapping.bio](#)**, a light-weight service to read semantic artifacts, visualize them, add mappings as graphical connections and store them as digital objects in a [Cordra-based](#) repo. Mapping.bio's data model is building on the concepts proposed in SEMAF leveraging on the FAIR Digital Object (FDO) approach
- **Use-case for registering mapping.bio created crosswalks to be available in the MSCR.**

Areas to further strengthen collaboration

- **Jointly explore and share expertise on the topics of common interest** such as data exchange formats, expressing data quality, workflows, interoperability
- **Develop specific activities, tools and other outputs carried out in the framework of both projects/initiatives** focusing on architecture design, customising workflow tools, PID and dynamic data
- **Collaborate on community level challenges** around use case development, user requirements collection and harmonisation
- **Stakeholder engagement, training and outreach activities:** actions for reaching various targeted audiences, to raise awareness and share training best practices and experiences, know-how and technology

