Biodiversity Digital Twin:

A novel and transformative approach to biodiversity research and application

Sharif Islam
Hanna Koivula
Dag Endresen
Erik Kusch
Dmitry Schigel
Wouter Addink

Funded by the European Union

BIODT biodiversitydigitaltwin

The Biodiversity Digital Twin prototype will provide advanced models for simulation and prediction capabilities, through practical use cases addressing critical issues related to global biodiversity dynamics

- 1 June 2022 – 31 May 2025 (36 months)
- 22 partners
- Experts in biodiversity research, high-performance computing, artificial intelligence, digital twinning and FAIR principles coming together

More information: www.biodt.eu

Funded by the European Union

The Digital Twin (DT) concept in BioDT

A digital twin is a virtual representation of real-world entities and processes, synchronised at a specified **frequency** and **fidelity***

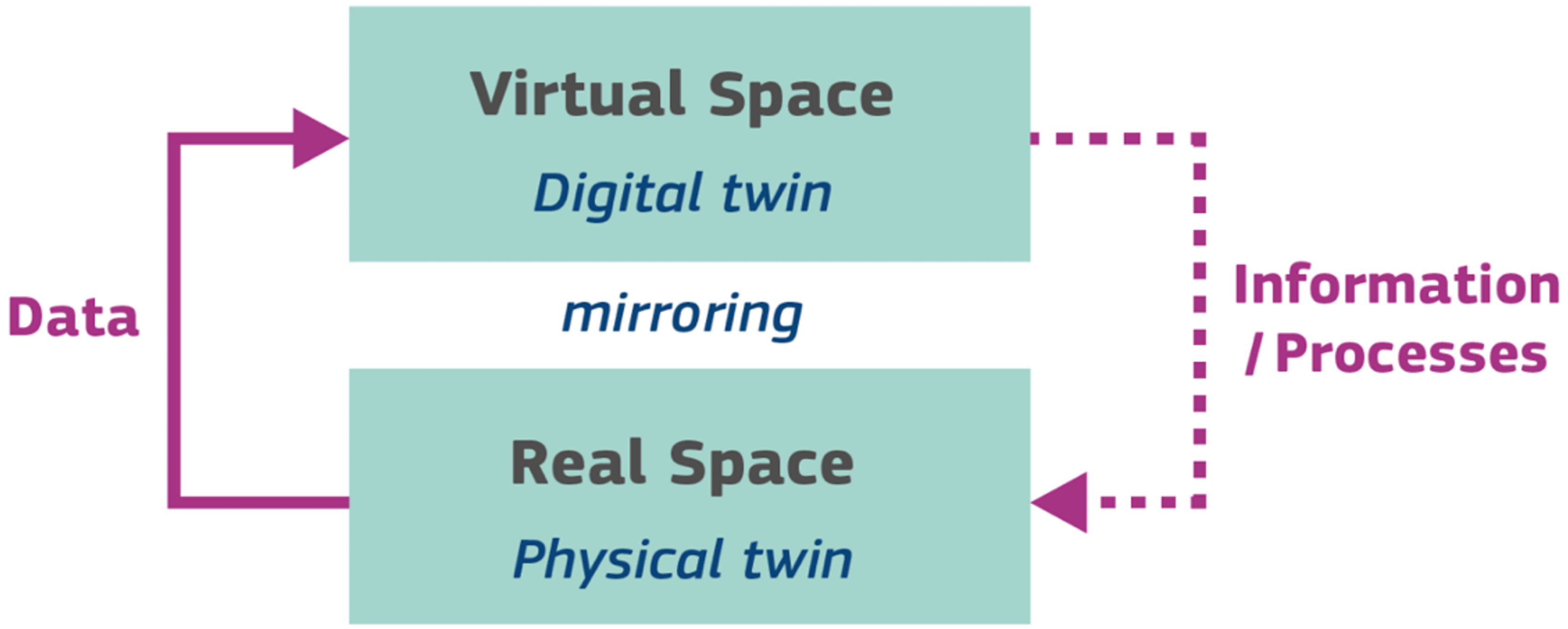


Image: digital-strategy.ec.europa.eu

*Here, fidelity refers to the level of precision captured by the DT in comparison with its physical counterpart.

The Digital Twin (DT) concept in BioDT

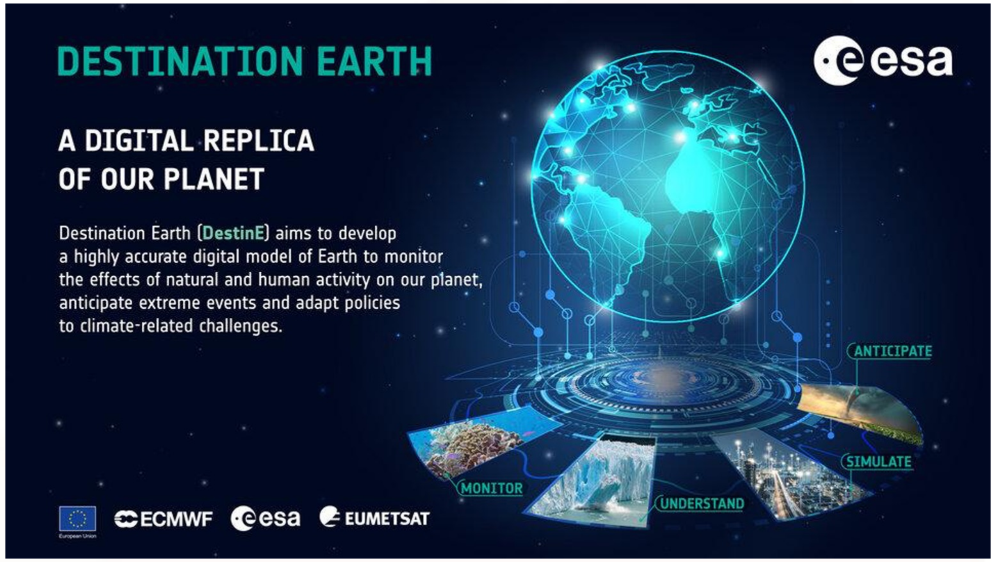
A DT is composed of:

- Data
- A model that is the representation in terms of behaviour and
- An application that connects the data and model in a way that makes the outputs of the model relevant, given the specific purpose of the DT

Since different scopes require different behaviour and fidelity, there cannot be a single twin answering all possible questions

Destination Earth

The BioDT project responds to key EU and international policy initiatives, including the EU Biodiversity Strategy 2030, EU Green Deal, UN Sustainable Development Goals, Destination Earth



Data and services from four Research Infrastructures

GBIF

The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure providing open access to biodiversity data.

LifeWatch ERIC

LifeWatch ERIC is the e-Science European infrastructure for biodiversity & ecosystem research.

eLTER

The Integrated European Long-Term Ecosystem (eLTER) focuses on critical zone and socio-ecological research.

Helmholtz Center for Environmental Research (UFZ), UK Centre of Ecology & Hydrology (UKCEH), Environment Agency Austria (EEA) and University of Helsinki (UH)

DiSSCo

The Distributed System of Scientific Collections (DiSSCo) is a Research Infrastructure (RI) for Natural Science Collections.

Naturalis Biodiversity Center (Naturalis) and Senckenberg Society for Nature Research (SGN)

Challenges and paths towards solution

- **Metadata**
  - Metadata enhances data findability and context
  - Important for the FAIR principles
- **Controlled Vocabularies**
  - Consistency and clarity in data interpretation
- **Semantic Mapping**
  - Enables harmonisation and meaningful integration of diverse data
- **Utilisation of Existing Standards**
  - Standards like Darwin Core and EML serve as foundational frameworks

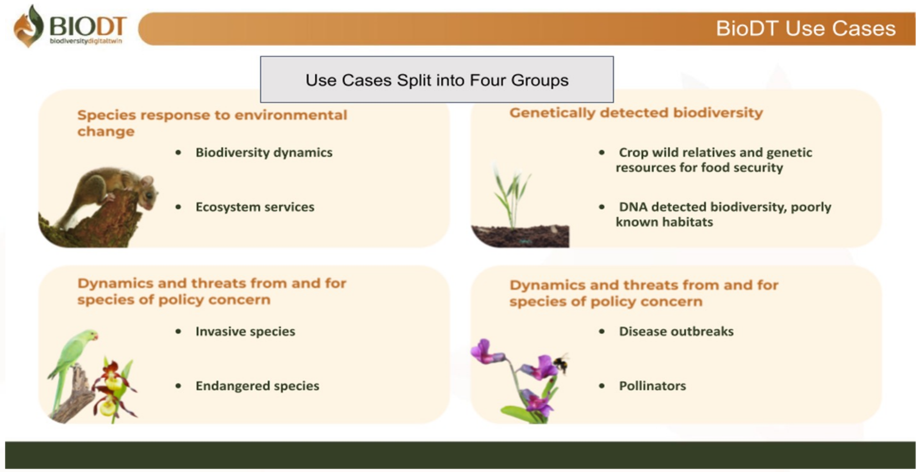


Image credit: BioDT

Challenges and paths towards solution

- Combining biodiversity data, which includes event and evidence information, with direct environmental (such as temperature, precipitation, soil composition) measurements presents a multifaceted challenge
- This integration is essential to understand how environmental factors influence biodiversity and the behavior of models

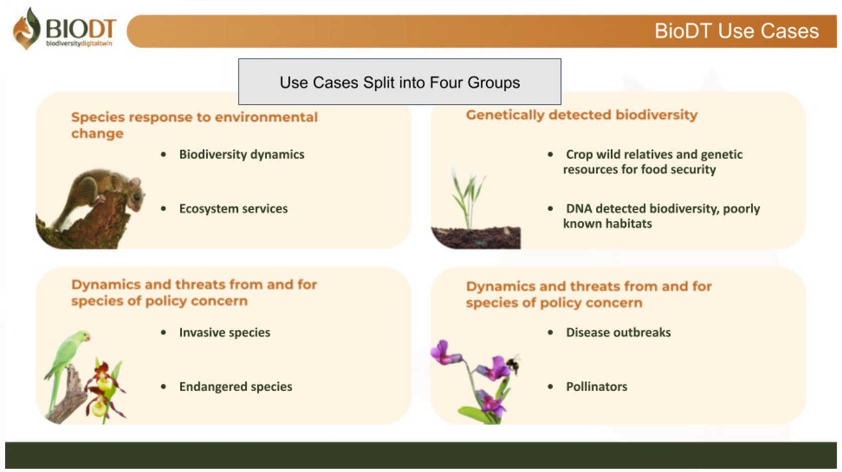
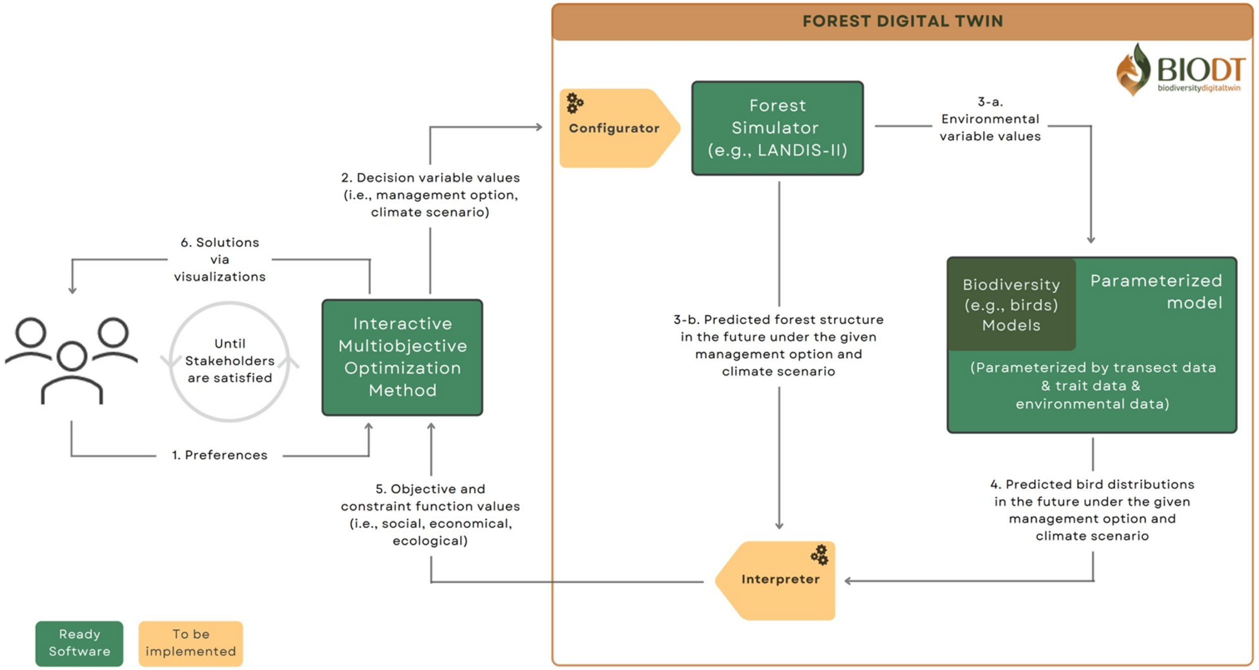


Image credit: BioDT

Example

How will forest biodiversity change under different forestry and climate change scenarios, and how can these predictions be utilised in conservation and adaptive forest management?



Digital Twin Application: Finding the most appropriate forest management strategy

Afsar, Bekir, & Ovaskainen, Otso. (2023). Forest biodiversity under different management and climate change scenarios. Zenodo. https://doi.org/10.5281/zenodo.8100210

Towards a solution: FAIR Digital Objects

- Foundation of DT applications: Data + Models
- Harmonised abstraction layer using Persistent Identifiers (PID) and FAIR Digital Object (FDO) records
- Semantic mapping and crosswalk techniques to provide machine-actionable metadata

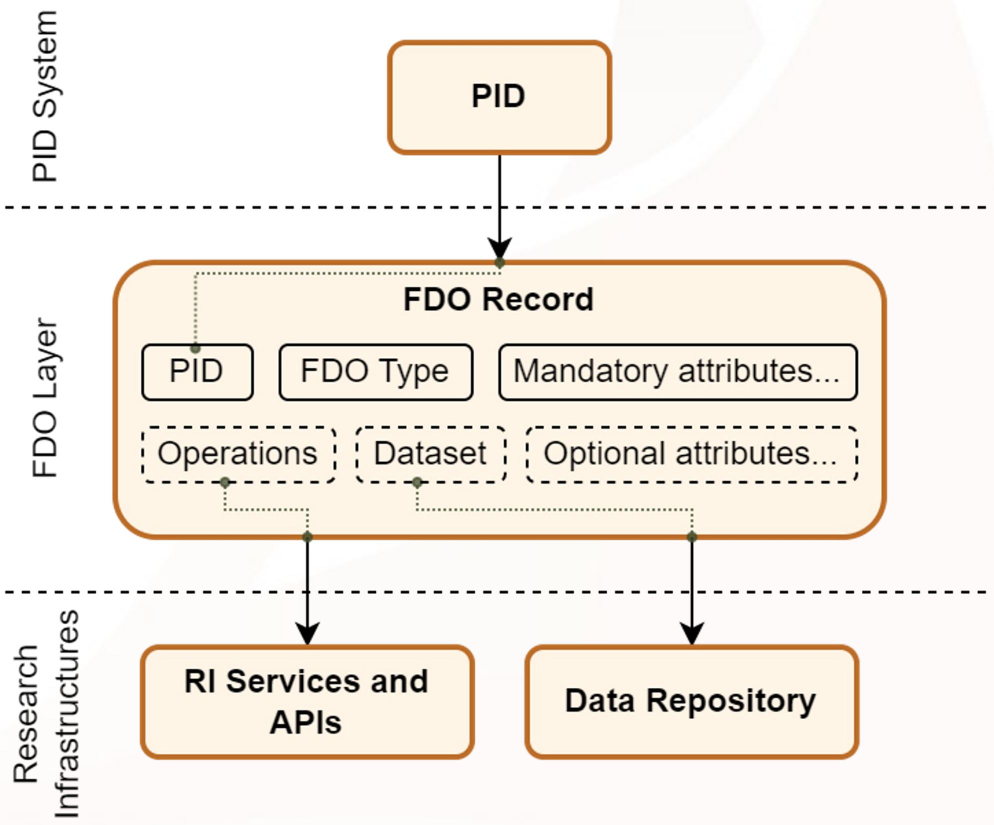


Image credit: BioDT Work Package 5

Models in BioDT

- Formal representations of problems or processes, implemented through equations, algorithms, or a combination of both

- Metadata example:
  - What is the **format** of the **input** data?
  - What **types** of data does the model accept?
  - What is the format of the **output** data?
  - **Where and how** can the model be accessed?
  - What **steps** are needed to run and test the model?
  - What were the **parameters** used to generate a particular version of the model?

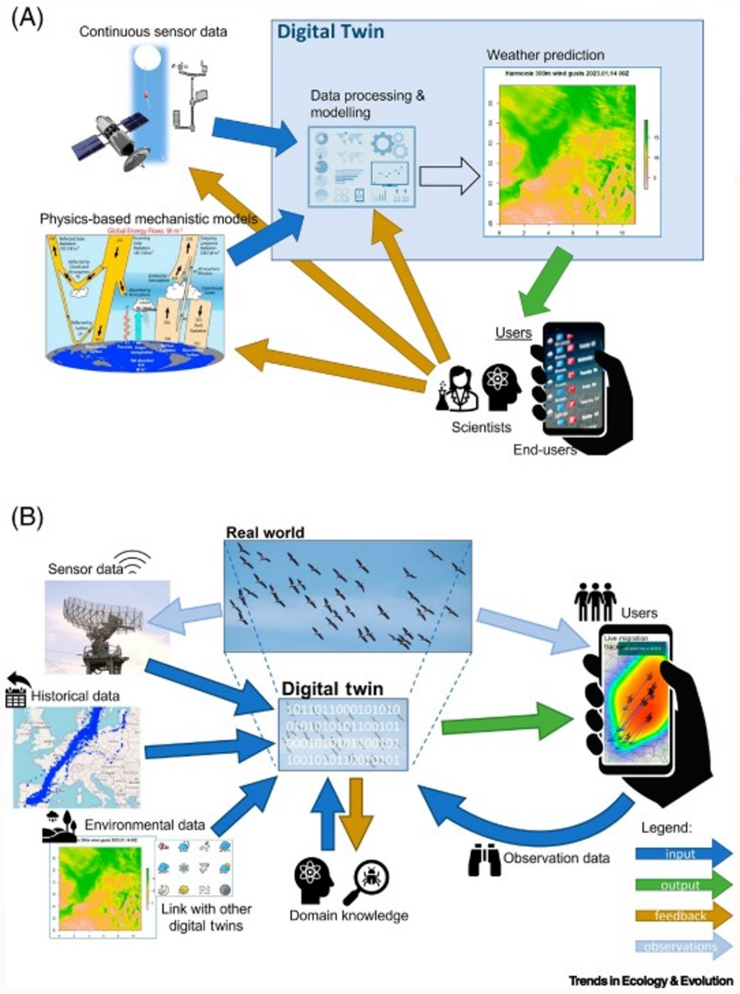


Image Credit: de Koning, K., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D. and Grimm, V., 2023. Digital twins: dynamic model-data fusion for ecology. Trends in Ecology & Evolution. https://doi.org/10.1016/j.tree.2023.04.010

More information: https://github.com/BioDT/biodt-fair

Infrastructure and software

- **Infrastructure**
  - Data and models require a robust computing infrastructure
  - BioDT is utilising LUMI for the prototype digital twins

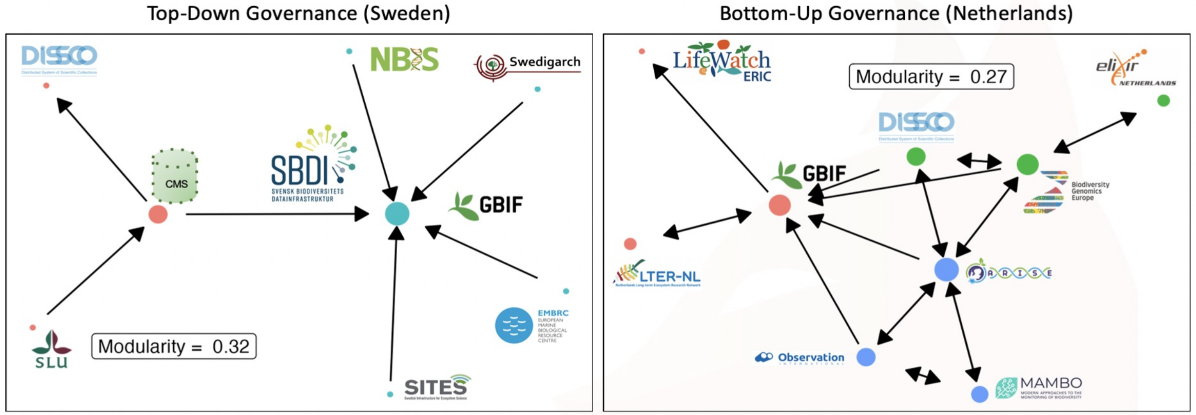
- **Software**
  - The use cases require a comprehensive suite of software tools
  - Also need software for data acquisition, storage, processing, and analysis
  - Coordinating diverse software components is a crucial aspect of BioDT

- **Workflow Representation**
  - Standardised procedures and protocols
  - Implementation of Research Object Crate (RO-Crate) to organise and manage digital objects within different workflows

https://www.csc.fi/en/lumi

Governance and stakeholder management

- Research infrastructures vary in governance approaches, from top-down to bottom-up models
- BioDT places a strong emphasis on aligning technical and data standards discussions among research infrastructures and stakeholders
- This alignment fosters synergy, enabling the project to leverage diverse expertise and resources effectively



Erik Kusch. (2023, May 18). The RI Landscape of BioDT – Plans for Assessing Fragmentation and Improving Communication. Zenodo. https://doi.org/10.5281/zenodo.8070318

Conclusion

- Groundbreaking and transformative approach to biodiversity research

- BioDT leverages collaborative research infrastructures and partnerships (GBIF, DiSSco, eLTER, and LifeWatch)

- The foundation of BioDT rests on robust data standards, and the ability to refer to multiple data originators by using PIDs for data citation

BIODT
biodiversitydigitaltwin

@BiodiversityDT

BioDT

Funded by
the European Union