BioDT

Introduction and Objectives

BioDT Project Management Office
Project name: Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities (BioDT)

Call title: Next generation of scientific instrumentation, tools and methods (HORIZON-INFRA-2021-TECH-01)

Duration: 1 June 2022 – 31 May 2025 (36 months)

Consortium: 22 partners

- Experts in biodiversity, high-performance computing, artificial intelligence, digital twinning and FAIR data
- Partners from 12 countries: Finland (FI), Italy (IT), Czech Republic (CZ), the Netherlands (NL), Estonia (EE), Sweden (SE), United Kingdom (UK), Germany (DE), Austria (AT), Denmark (DK), Norway (NO), Spain (ES)
  - Incl. one Affiliated Entity and three Associated Partners

Work Package (WP) members: 140+

Coordinator: CSC – IT Center for Science

Website: www.biodt.eu
OBJECTIVE 1:

- Build and deploy a pre-operational BioDT for addressing biodiversity dynamics

OBJECTIVE 2:

- Support the interoperability of data and services through the integration of the BioDT with research infrastructure platforms and workflows

OBJECTIVE 3:

- Ensure interoperability of BioDT with Destination Earth and the European Data Infrastructure
1: Pre-operational BioDT platform

- Platform established on LUMI
- Case studies for model development
- Model development¹ and validation
- Serving user needs beyond the project consortium

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prototype available as service</td>
</tr>
<tr>
<td>2</td>
<td>Several case studies</td>
</tr>
<tr>
<td>3</td>
<td>Improved model predictive performance</td>
</tr>
<tr>
<td>4</td>
<td>Increased model accuracy and precision</td>
</tr>
</tbody>
</table>

¹ Incl. upscaling for HPC, features for interactive use
2: Integration with RIs

- Application programming interfaces (APIs), user authentication and access
- Interoperability: data, software, practices
- Uptake by research communities, new user communities, training

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>APIs for feeding data to BioDT platform</td>
</tr>
<tr>
<td>2</td>
<td>FAIR datasets using cross-RI standards and FDOs</td>
</tr>
<tr>
<td>3</td>
<td>Quality indicators, e.g. FAIRness assessment metrics, geographic accuracy</td>
</tr>
<tr>
<td>4</td>
<td>Training materials and interoperability workshops</td>
</tr>
<tr>
<td>5</td>
<td>Improved modelling, simulation and prediction capabilities for RIs</td>
</tr>
</tbody>
</table>
3: Interoperability with DT initiatives (incl. DestinE) and EDI

- Cross-DT synchronisation and showcases
- EOSC data integration, openly available results
- Showcases for interaction between the BioDT and other DTs
- Harmonised data and data governance ([EU Data Spaces](https://data.europa.eu/8json/))

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BioDT data outputs to DestinE data lakes</td>
</tr>
<tr>
<td>2</td>
<td>Interfaces and data integration for interaction with EOSC</td>
</tr>
<tr>
<td>3</td>
<td>Integration of DestinE output data for use by BioDT</td>
</tr>
<tr>
<td>4</td>
<td>Synchronisation with other DT initiatives (e.g. Ocean DT)</td>
</tr>
</tbody>
</table>
The BioDT prototype Digital Twins (pDTs) are divided into four main groups focusing on:

1. **Species response to environmental change**
   - Biodiversity dynamics
   - Ecosystem services

2. **Genetically detected biodiversity**
   - Crop wild relatives and genetic resources for food security
   - DNA detected biodiversity, poorly known habitats

3. **Dynamics and threats from and for species of policy concern**
   - Invasive species

4. **Species interactions with each other and with humans**
   - Disease outbreaks
   - Pollinators
Research Infrastructures

Data from four RIs
GBIF, eLTER, LifeWatch ERIC and DISSCo

<table>
<thead>
<tr>
<th>GBIF</th>
<th>The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure providing open access to biodiversity data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBIF</td>
<td>![GBIF Logo]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>eLTER</th>
<th>The Integrated European Long-Term Ecosystem (eLTER) focuses on critical zone and socio-ecological research.</th>
</tr>
</thead>
<tbody>
<tr>
<td>eLTER</td>
<td>![eLTER Logo]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LifeWatch ERIC</th>
<th>LifeWatch ERIC is the e-Science European infrastructure for biodiversity &amp; ecosystem research.</th>
</tr>
</thead>
<tbody>
<tr>
<td>![LifeWatch ERIC Logo]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DiSSCo</th>
<th>The Distributed System of Scientific Collections (DiSSCo) is a Research Infrastructure (RI) for Natural Science Collections.</th>
</tr>
</thead>
<tbody>
<tr>
<td>![DiSSCo Logo]</td>
<td></td>
</tr>
</tbody>
</table>

- Helmholtz Center for Environmental Research (UFZ), UK Centre of Ecology & Hydrology (UKCEH), Environment Agency Austria (EAA) and University of Helsinki (UH)
- Naturalis Biodiversity Center (Naturalis) and Senckenberg Society for Nature Research (SGN)
Leveraging high-performance computing, AI and data analytics capabilities of the LUMI supercomputer

- LUMI is the 3rd fastest supercomputer in the world
- Sustained performance: 375 petaflop/s = performs $375 \times 10^{15}$ calculations per second
- Computing power equals to the capacity of 1.5 million modern laptops
Digital twins require platforms for **computational simulation**

- Detailed and realistic simulations require large amounts of computing time

---

**Digital Twin Advanced Technical Platform**

- HPC resources from the LUMI EuroHPC computing facility
- Ensuring portability of digital twins across HPC sites and cloud environments
- Maintaining service catalogue of shared services for integration with EOSC Core services
BioDT Target Groups

**Biodiversity RIs, RI nodes, data providers and researchers**
RIs, universities, research organisations; the end-users that will contribute to developing the DT, enhancing its use cases, and testing its functionalities

**Policy makers**
EU, Member States, local governments, intergovernmental organisations (UNESCO, FAO, etc.)

**Industrial actors incl. SMEs**
Sectors related to biodiversity, such as agri-food, tourism, healthcare

**Civil society and citizen scientists**
Benefits

**Researchers** will be able to:
- Better observe changes in biodiversity in response to forces resulting from climate change or human activity
- Mechanistically understand how these changes occur
- Predict the effects of these changes

**Biodiversity RIs** will be able to:
- Improve their services

**Policy makers** will be able to:
- Make decisions based on quality data and modelling
- Better respond to societal needs and key initiatives

**Industrial actors incl. SMEs** will be able to:
- Exploit BioDT for business solutions (applications/products) in sectors related to biodiversity, such as agri-food, tourism and healthcare

**Citizen scientists and civil society:**
- Boost citizen science
- Strengthen common understanding of biodiversity dynamics and prediction models
- Foster biodiversity literacy and trust in biodiversity research