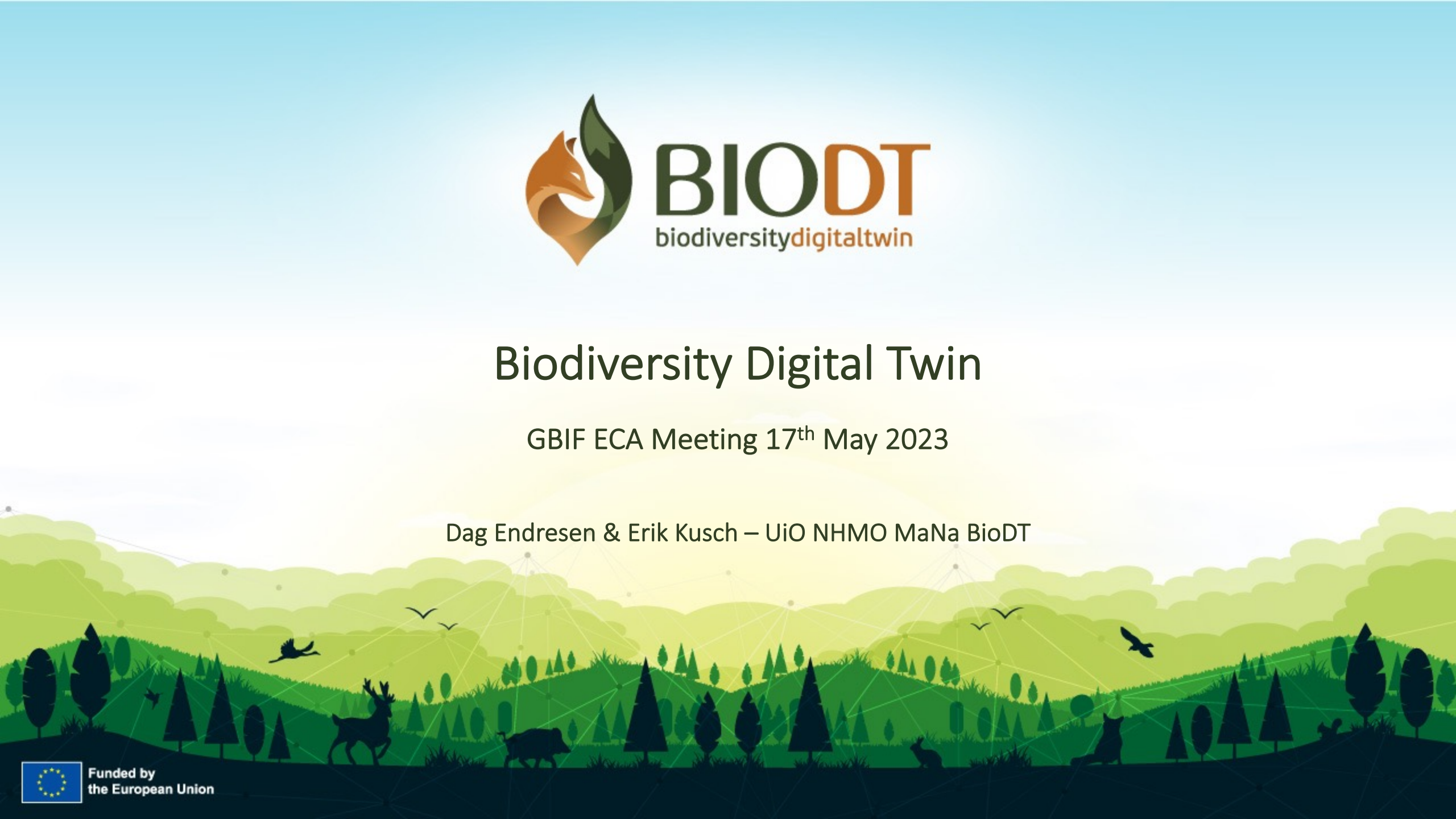


Biodiversity Digital Twin

GBIF ECA Meeting 17th May 2023

Dag Endresen & Erik Kusch – UiO NHMO MaNa BioDT

Funded by
the European Union

Outline

1. Background

- Digital twin concept
- DT concept in BioDT

2. Objectives

- Project goals and outcomes

3. BioDT Use Cases

- Practical applications
- Data streams through the bioRIs - GBIF, eLTER, DiSSCo, LifeWatch, etc.

4. LUMI

- EuroHPC supercomputer

What is a digital twin?

Virtual representation(s) of real-world entities and processes,
synchronised at a specified frequency and fidelity

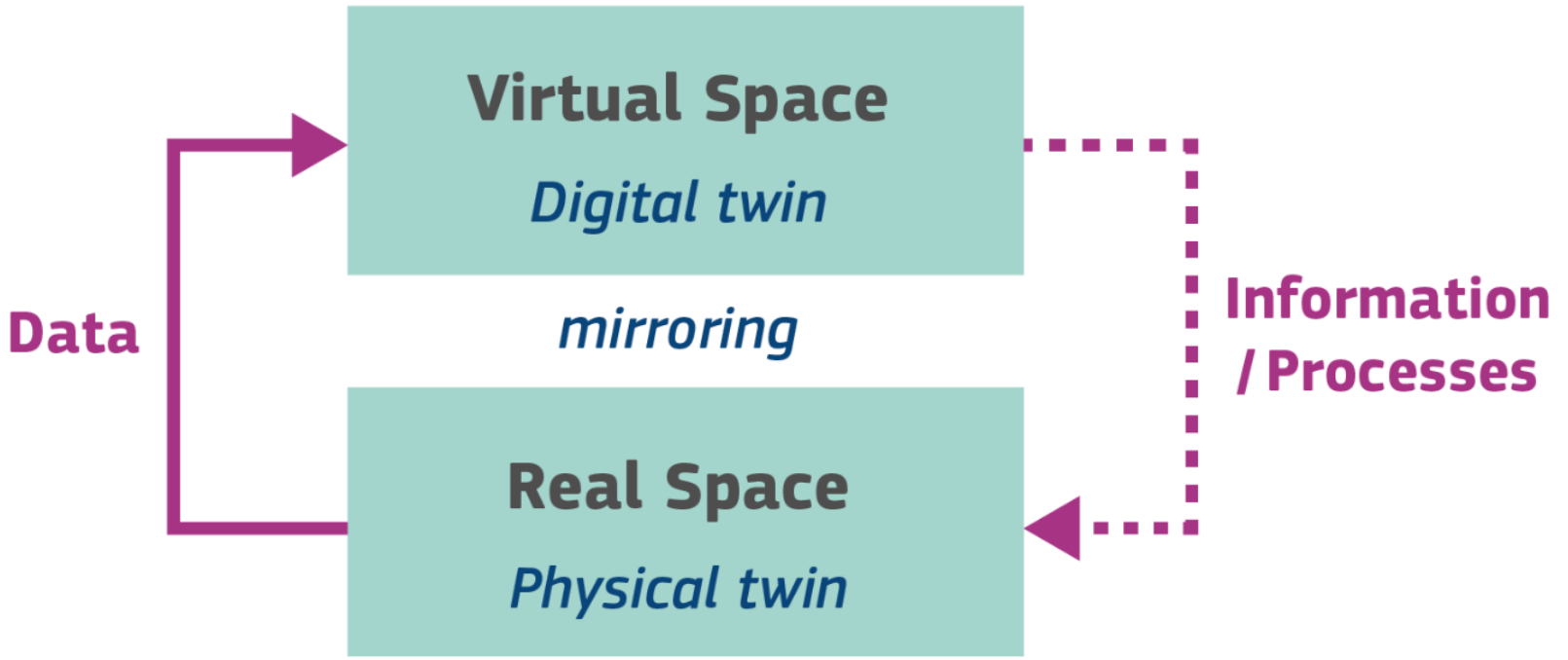
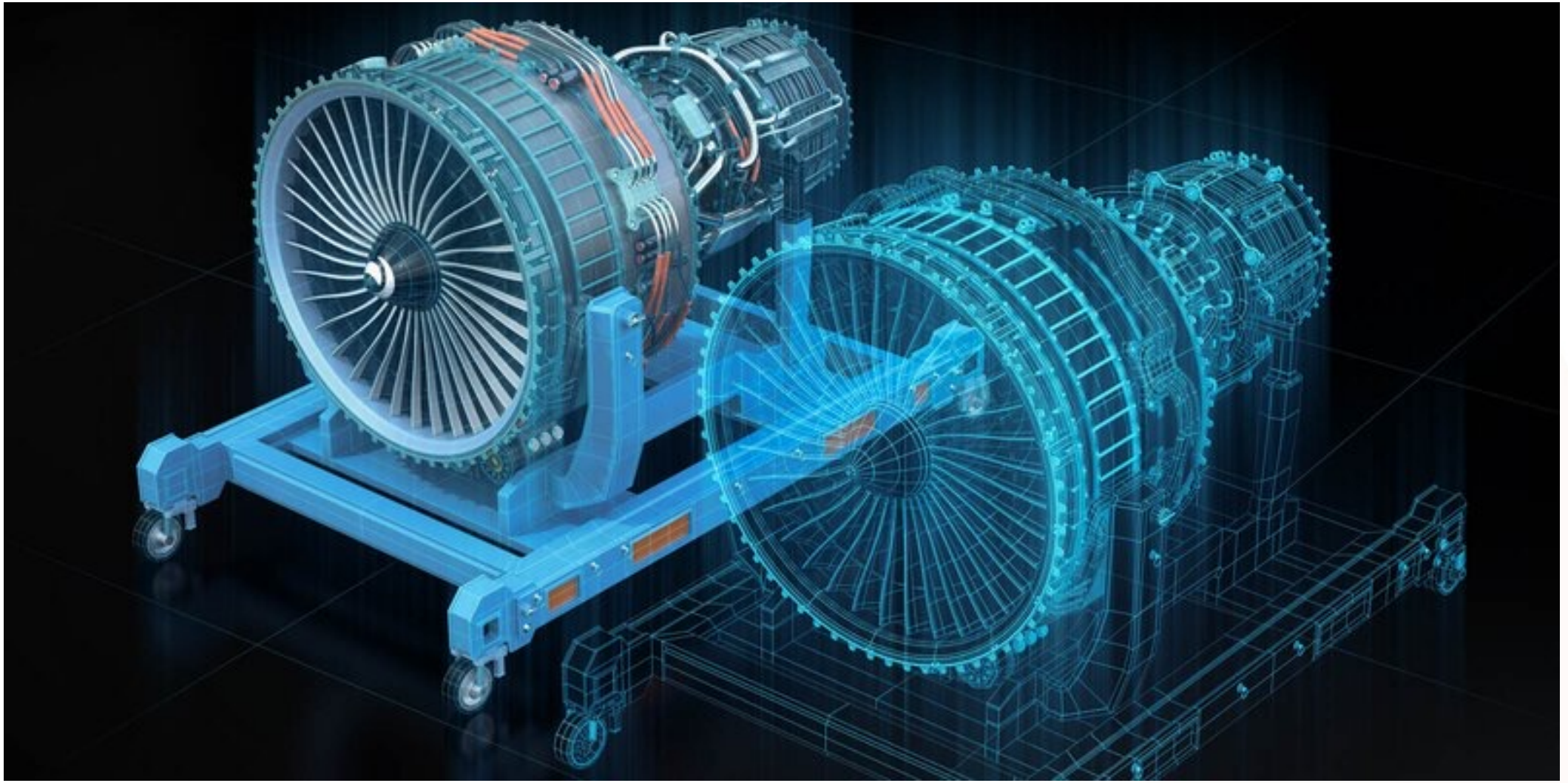


Image: digital-strategy.ec.europa.eu

Digital Twin concept is gaining widespread use in engineering



Digital Twin for crops - WUR

- Wageningen University is developing digital twins for crop plants.
- Sensors provide real-time data streams to a digital twin model.
- Simulated experiments can be made with the digital twin.

- **Industrial** DTs typically facilitate:
  - Product design
  - Operation of machinery

- In **BioDT**, DTs are used to:
  - Mimic behaviour observed in nature
  - Meet requirements of BioDT Use Cases
  - Contribute toward EC goal of devising a full DT of the Earth

Biodiversity Digital Twin (BioDT)

Horizon Europe

- **BioDT – Biodiversity Digital Twin**
  - 22 partner institutions from 13 countries
- The total project budget is 12M Euro
  - 1265 PM (105 PY FTE)
- **GBIF Secretariat** 143 PM (11.9 PY FTE) – 11% of BioDT
- **GBIF NO (UiO)** 81.5 PM (6.79 PY FTE) – 8% of BioDT

Project Information

BioDT

Grant agreement ID: 101057437

DOI

10.3030/101057437

Start date 1 June 2022

End date 31 May 2025

Funded under

Research infrastructures

Total cost

€ 11 059 061

EU contribution

€ 11 059 061

Coordinated by

CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY

Finland

Project partners

Trust-IT Services communicating to markets

COMMpla Communication Platforms and Online Solutions

SENCKENBERG world of biodiversity

GBIF Global Biodiversity Information Facility

LifeWatch ERIC

Naturalis Biodiversity Center

CSC

ECMWF

UK Centre for Ecology & Hydrology

TNO innovation for life

UFZ HELMHOLTZ Centre for Environmental Research

BAYER

umweltbundesamt

UNIVERSITY OF HELSINKI

Swansea University Prifysgol Abertawe

UNIVERSITY OF OSLO

MANCHESTER 1824 The University of Manchester

JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

VSB TECHNICAL UNIVERSITY OF OSTRAVA

KTH VETENSKAP OCH KONST

UNIVERSITY OF TARTU

MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

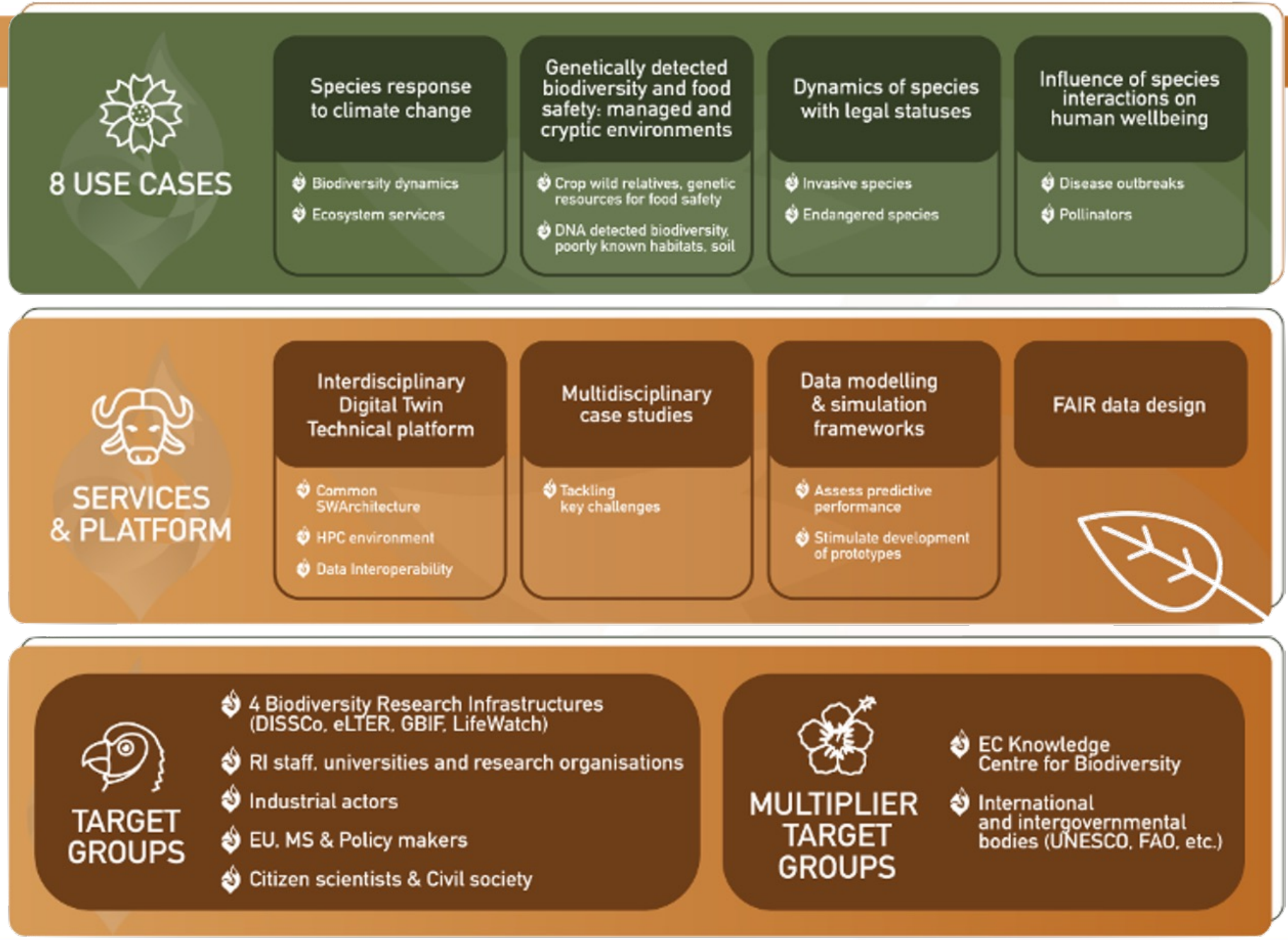
22 partners from 12 countries

1	**CSC - IT Center for Science Ltd**	Finland
2	Trust-IT Srl	Italy
3	AE - Commpla Srl	Italy
4	VSB - Technical University of Ostrava, IT4I	Czech Republic
5	TNO - Netherlands Organisation for Applied Scientific Research	Netherlands
6	JYV - University of Jyvaskyla	Finland
7	University of Tartu	Estonia
8	KTH - Kungliga Tekniska Hoegskolan	Sweden
9	Swansea University	United Kingdom
10	University of Manchester	United Kingdom
11	ECMWF - European centre for medium-range weather forecasts	United Kingdom
12	UFZ - Helmholtz-Zentrum fur umweltforschung gmbh	Germany
13	UK Centre for Ecology & Hydrology	United Kingdom
14	EAA - Umweltbundesamt gesellschaft mit beschrankter haftung	Austria
15	University of Helsinki	Finland
16	**GBIF - Global Biodiversity Information Facility, secretariat**	Denmark
17	**University of Oslo (GBIF Norway)**	Norway
18	Naturalis - Stichting Naturalis Biodiversity Center	Netherlands
19	Senckenberg gesellschaft fur naturforschung	Germany
20	E-Science European Infrastr. for Biodiv. and Ecosys. Research	Spain
21	Bayer aktiengesellschaft	Germany
22	MLU - Martin-Luther Universitat Halle-Wittenberg	Germany

- EU DestinE
- EU Digital strategy
- EU Green Deal

- 4 use case groups
- 8 use case studies

- 4 bio RIs
  - GBIF
  - eLTER
  - DiSSCo
  - LifeWatch



WP	WP title	WP Lead	PMs
1	Project management	CSC, Finland	72
2	Communication, Dissemination, Sustainability, and Impact	TRUST-IT, Italy	106
3	Digital Twin Advanced Technical Platform	CSC, Finland	153
4	**Data and Use Cases**	**GBIFS, Denmark UiO, Norway**	**301**
5	Improving Quality of Data, Workflows, and Models through FAIR Principles	Naturalis, NL	99
6	Simulation, Modelling and Data Analytics	JYU, Finland UFZ, Germany	300
7	Integration & Service Uptake with Research Infrastructure Environments	IT4I@VSB, Czech Republic	148
8	Collaboration and Integration with Strategic Initiatives and Programmes	TNO, Netherlands	86
			1265

Project objectives

- **Objective 1**
  - Build and deploy pre-operational BioDT platform
- **Objective 2**
  - Integration with RI platforms and workflows
- **Objective 3**
  - Interoperability with European DT initiatives (including DestinE) and European Data Infrastructure

Specific objectives and outcomes

Objective 1: Pre-operational BioDT platform

- Platform established on LUMI
- Prototype DTs focusing on four Use Case groups
- Model development and validation
- Including upscaling for HPC, features for interactive use
- Platform generalized to serve user needs beyond consortium

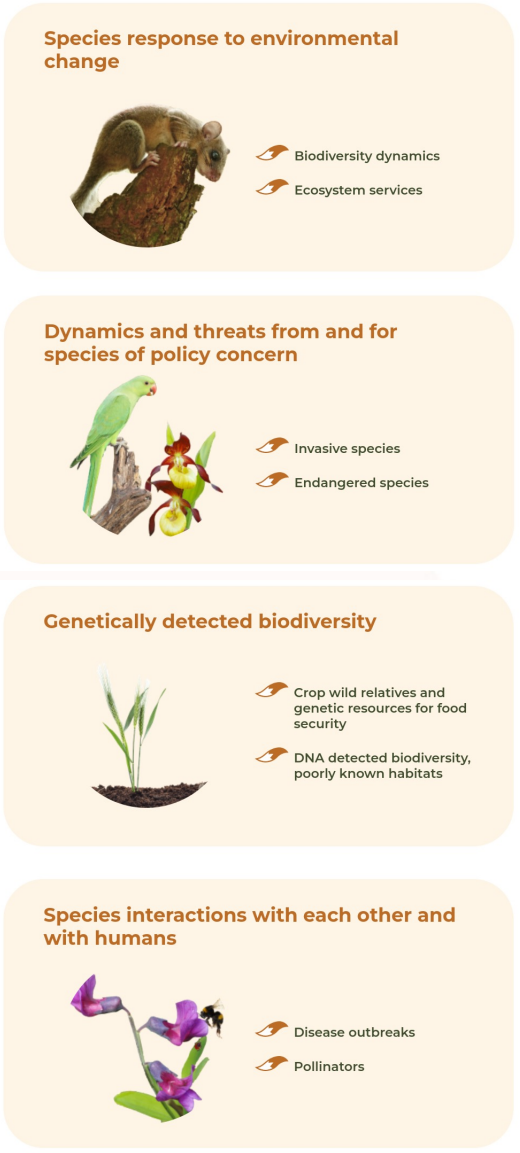
Outcomes

1 Prototype platform available *as service*

2 Four use case groups, Eight case studies

3 Improved model predictive performance

4 Increased model accuracy and precision



Specific objectives and outcomes

Objective 2: Integration with biodiversity RIs

- Interfaces, user authentication and access
- Interoperability: data, software, practices
- Uptake, new user communities, training
  - Including *Bring-Your-Own-Data* hackathons

Outcomes

1 APIs for feeding data to & from BioDT platform

2 FAIR datasets using cross-RI standards and FDOs

3 Quality indicators (e.g. FAIRness, geographic accuracy)

4 Training materials and interoperability workshops

GBIF

GBIF

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

eLTER

eLTER

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

LifeWatch ERIC

LifeWatch ERIC

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

DiSSCo

DiSSCo Distributed System of Scientific Collections

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

Objective 3: Interoperability with DestinE and European DT initiatives

- Cross-DT synchronisation and showcases
- Integration with EOSC and ELIXIR (openly available results)
- Harmonised data and data governance (EU Data Spaces)
- Contributing to EU ability to adapt actions & policies

Outcomes

1 Ingestion of DestinE by BioDT (and data outputs to DestinE)
2 Provision of agglomerated, analysis-ready data by BioDT
3 Shared service catalogue, BioDTs available to EOSC users
4 Synchronisation with other DT initiatives (e.g. Ocean DT)
5 Leveraging ELIXIR infrastructure (e.g. Tools platform)

	**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**

BIODT
biodiversitydigitaltwin

scientific use case studies




8 scientific use cases

Species response to environmental change

- Biodiversity dynamics
- Ecosystem services

Genetically detected biodiversity

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

Dynamics and threats from and for species of policy concern

- Invasive species
- Endangered species

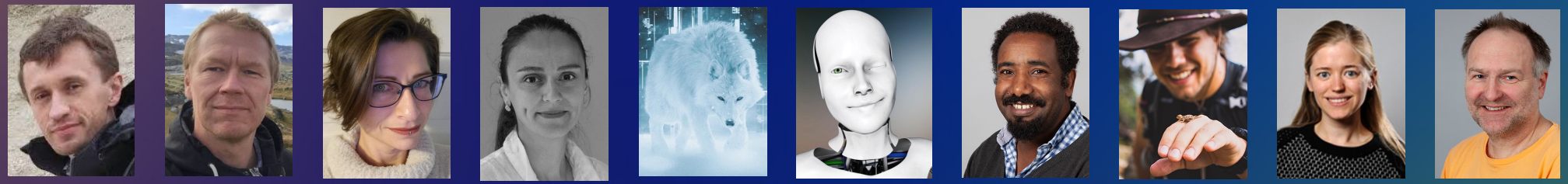
Species interactions with each other and with humans

- Disease outbreaks
- Pollinators

Scientific use case digital twin – prediction tasks

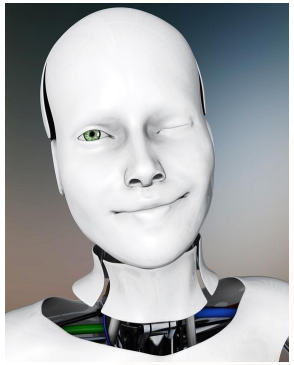
Scientific use cases	Prediction tasks
Group 1: Species response to environmental change 1. Biodiversity dynamics 2. Ecosystem services	1. Changes in species diversity, distribution, and abundance across scales. 2. Changes in species diversity, distribution, and abundance across scales due to anthropogenic effects.
Group 2: Genetically detected biodiversity 3. Crop wild relatives, genetic resources for food security 4. DNA detected biodiversity, poorly known habitats, and soil	1. DNA sequence to taxon/genotype identification. 2. DNA sequence to phenotype, trait, or property of taxon, and genotype of an organism. 3. Modelling patterns and processes in arable lands and cryptic environments. 4. Modeling changes in soil, degrading organic matter, water, and air.
Group 3: Use cases on the dynamics of species of policy concern 5. Invasive species 6. Endangered species	1. Changes in species richness and abundance of invasive species. 2. Changes in species richness and abundance of red list species.
Group 4: Species interactions with each other and with humans 7. Disease outbreaks 8. Pollinators	1. Modelling emerging diseases and their potential locations in Europe. 2. Impact of changes in pollinator populations, distribution, and/or types on the dependent plant communities, both wild and cultivated.

GBIF BioDT team – GBIFS & GBIF Norway



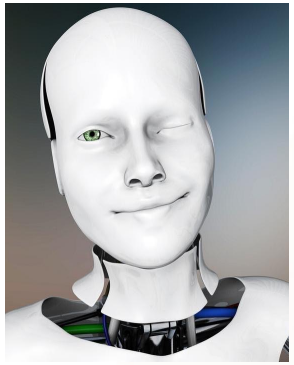
- Four full time (FTE) project positions
  - Tobias Frøslev – Task 4.1.2.2 lead – Use case: DNA detected biodiversity, soil
  - Kate Ingenloff– Task 4.1.4.1 lead – Use case: Disease outbreaks (interactons)
  - GBIF informatics – Task 4.2 data streams
  - Sanja Novakovikj – project support
  - Dmitry Schigel – project leader at GBIFS





- Three full time (FTE) project positions
  - Desalegn Chala Gelete – Task 4.1.2.1 lead – Use case: Crop wild relative digital twin
  - Erik Kusch – Task 4.3.1 lead – biodiversity research data infrastructure
  - *Open* – Task 4.3.2 – data standards and data quality indicators
  - Marcella Rydmark Orwick – liasion WP2, WP7, WP8
  - Dag Endresen – project leader at UiO Tøyen





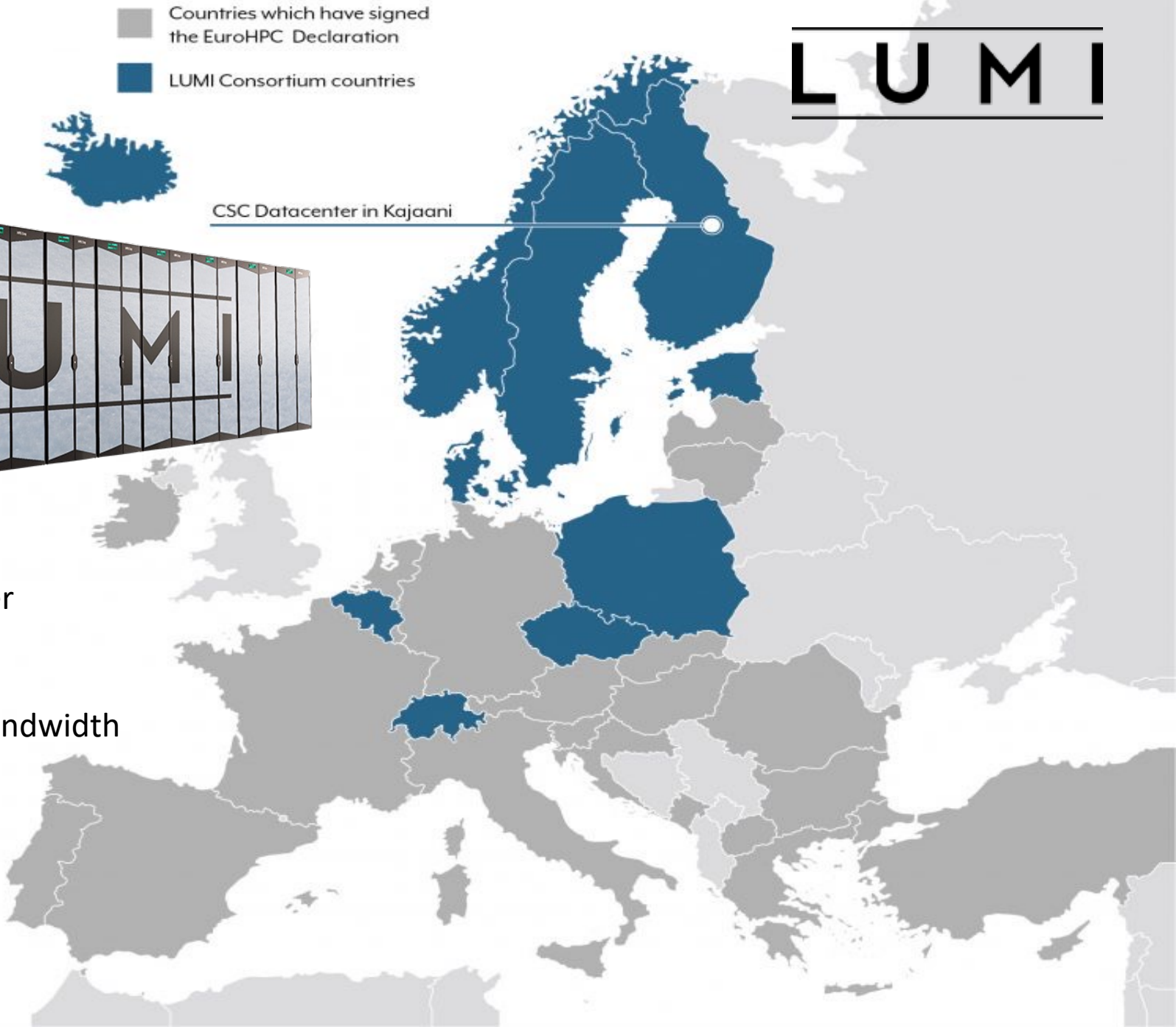

LUMI

- The **European High-Performance Computing Joint Undertaking (EuroHPC JU)** is pooling European resources to develop **top-of-the-range exascale supercomputers** for processing big data, based on competitive European technology.
- One of the pan-European pre-exascale supercomputers, LUMI, is located in CSC's data center in Kajaani, Finland.
- The supercomputer is hosted by the LUMI consortium. The **LUMI** (Large Unified Modern Infrastructure) consortium countries are **Finland**, Belgium, Czech Republic, Denmark, Estonia, Iceland, **Norway**, Poland, Sweden, and Switzerland.
- LUMI will be one of the world's best known scientific instruments for the **lifespan of 2021–2026.**

EuroHPC Joint Undertaking

LUMI

BIODT
biodiversitydigitaltwin



LUMI

LUMI is a GPU-accelerated supercomputer
GPUs with over 550 Pflop/s
32 TB memory
7 PB flash storage with extreme 2 TB/s bandwidth
30 PB object storage for staging area
80 PB parallel file system

Top500 most powerful HPC

- Top500 ranks the 500 most powerful non-distributed computer systems in the world.
- LUMI ranks as #3 (!)

TOP500 The List.

Rank	System	Cores	Rmax (Pflop/s)	Rpeak (TFlop/s)	Power (kW)
1	Frontier, Oak Ridge, USA	8 730 112	1 102	1 686	21 100
2	Fugaku, Fujitsu, Japan	7 630 848	442	537	29 899
3	LUMI, EuroHPC/CSC, Finland	2 220 288	309	429	6 016
4	Leonardo, EuroHPC/CINECA, Italy	1 463 616	175	256	5 610
68	Dardel GPU, KTH, Sweden	52 864	8	10	146
91	LUMI-C, EuroHPC/CSC, Finland	194 560	6	8	1 216
119	Betzy, UNINETT Sigma2 AS, Norway	172 032	5	6	--

https://www.top500.org/lists/top500/2022/11/

BioDT will provide infrastructure to:

- Drive long-term biodiversity research
- Maintain commitments to protect biodiversity
- Safeguard societal resilience

BioDT will be used to:

- Better observe spatiotemporal changes in biodiversity
- Develop an improved mechanistic understanding of these changes
- Push limits of predictive biodiversity modelling


BIODT
biodiversitydigitaltwin
@BiodiversityDT
BioDT
Funded by
the European Union