



**BIODT**  
biodiversitydigitaltwin

# **The BioDT use cases:** unique demonstrators to test the biodiversity Digital Twin prototype

Report of webinar 27 September 2022










# **The BioDT use cases:**

## **unique demonstrators to test the biodiversity Digital Twin prototype**

The BioDT project aims to develop the biodiversity digital twin prototype to help better predict global biodiversity dynamics and how species interact with their environment and with each other.

In order to test the work developed by the BioDT team in the three years of the project, groups of use cases will showcase how BioDT can examine its predictive performance and address biodiversity challenges through scenario simulations, predictions and biomonitoring methods.

In order to speed up with the development of the BioDT use cases, the project team has met from 5 to 7 September 2022 in Potsdam, Germany, for an internal technical workshop. The aim of the event was to:

-  Consolidate the use cases leads and project teams;
-  Develop the scientific questions for each use case;
-  Start developing requirements for each use case;
-  Identify data types and sources to be used;
-  Prioritise UCs and exact scenarios to be addressed by initial development work;
-  Define the process of data flow from Research Infrastructures (RIs) to modellers;
-  Define data format and quality requirements between 1) RIs and 2) RIs and modellers.

The workshop was organised by UFZ and welcomed 17 participants from BioDT partners.

Following this workshop, and in order to present the use cases developments also to external partners, a free BioDT webinar was organised on 27 September 2022. The event showcased through presentations, interactive sessions with the audience and a Questions & Answers session the main achievements of the use cases and the next steps. During the 1-hour webinar, it was possible for attendees to learn more about the scope, challenges, solutions and benefits brought by the BioDT use cases and how their development can help scientists in predicting future trends in biodiversity. The webinar was organised by Trust-IT in cooperation with the use cases leaders.

The following sections present the main objectives and characteristics of the use cases and the main outcomes of the BioDT webinar.



# The BioDT four use cases groups

The BioDT use cases are eight in total and are divided into four main groups, which are focused on:



## Group 1

Species response to environmental change:

This group puts attention on biodiversity dynamics and ecosystems services.

## Group 2

Genetically detected biodiversity:

This group aims to bring research forward for crop wild relatives and genetic resources for food security and DNA-detected biodiversity, poorly known habitats.



## Group 3

Dynamics and threats from and for species of policy concern:

This group explores invasive species and endangered species.

## Group 4

Species interactions with each other and with humans:

studies disease outbreaks and pollinators.





The topics investigated by each group are then tackled in more specific areas by the two use cases in each group. These use cases are designed to build on the biodiversity Research Infrastructures involved in BioDT (DiSSCo, eLTER, GBIF, LifeWatch) to develop the first prototype biodiversity digital twins. The objective is to demonstrate that these digital twins are fit for scientific research.

An overview of the scientific use cases for BioDT was presented by Dag Endresen, GBIF Node Manager for Norway and BioDT Use Cases Co-Leader, at the use cases webinar, where three of the eight use cases, covering three of the four case groups, were presented. In particular, the webinar presented:

- 🔥 Genetically detected biodiversity, presented by Desalegn Chala Gelete, Post-doc researcher at Universitet i Oslo;
- 🔥 Dynamics and threats from and for species of policy concern, explained by Ingolf Kühn, Professor at Helmholtz Centre for Environmental Research - UFZ;
- 🔥 Species interactions with each other and with humans, introduced by Marcella Orwick Rydmark, Research Coordinator at Universitet i Oslo.

The following paragraphs present the characteristics of each use case group.



# Group 1

## Species response to environmental change



### Short Introduction

Global change is increasingly threatening nature's biodiversity as well as services for human well-being. Species across various taxa and trophic levels respond differently to environmental changes (e.g. climate extremes and change, management and land use change) which can lead to undesired changes of biodiversity and related ecosystems' functioning and services. To mitigate negative effects, interactions of climate, management and ecosystem dynamics have to be understood and future ecosystem states have to be projected reliably, in order to timely intervene by adequate management actions. These topics are covered in the Group 1 of the BioDT Use Cases, which cover "biodiversity dynamics" and "ecosystem services", that were presented only during the workshop.

### Scope

In the use case on "biodiversity dynamics", the BioDT team focus the analysis on the spatial(-temporal) dynamics of various terrestrial ecosystems: (A) temporal grassland observations (eLTER) and mechanistic modelling will be combined in real-time exchange at several local sites across Europe focusing on plant species diversity under management and climate change, (B) biodiversity models (jSDM), forest simulators and real-time observations (forest inventory, bird data) will be combined with interactive multi-objective optimisation methods on management options with the focus on tree and bird species in continuous forest areas of Finland, and (C) citizen science records on bird sound will be combined in real-time exchange with machine learning and jSDM approaches to assess bird migration (e.g. shifts in spring arrival to Europe due to climate change).

In the use case on "ecosystem services", the focus is on cultural ecosystem services related to recreation and biodiversity. In particular the landscape model ESTMAP-Recreation is combined with species occurrence models

to predict where and when a species may be sighted. In addition, adaptive models on species occurrence will incentivise new learning opportunities by encouraging citizen scientists to search for species in poorly monitored areas. The resulting digital twin will be aimed at landscape managers and policy makers as well as to the general public.

## Challenges

Some use case examples are first proof-of-concept studies in Finland and could be, if successful, extended to other European countries or other species (e.g. UC examples on forest bird biodiversity and citizen science bird monitoring), with the latter example especially requiring high demands for data flows between mobile phones & HPC. For the UC example on grassland biodiversity dynamics, the availability, accessibility and completeness (e.g. management, soil and weather data) of eLTER observational sites will be decisive for its accuracy and coverage across Europe. While the ambitious UC on ecosystem services will test combining citizen science, camera traps and acoustic biodiversity monitoring with human preference data.

## Benefits

Different management and climate scenarios will be covered in the use cases on “biodiversity dynamics” to explore their interactive influence on the spatio-temporal dynamics and future states of biodiversity in terrestrial ecosystems, which in turn can support stakeholders’ decisions on management options for an intervention of undesired effects (also in an ecological, social, and economic manner). The combination of large-scale and long-term observations as well as citizen science data (mobile app, JYU Mobile) in real-time with modelling approaches and methods of different complexity (mechanistic modelling, jSDM, machine learning, interactive multi-objective optimization) provide a range of demonstration examples for real-time biodiversity digital twins.

## Partners Involved



# Group 2

## Genetically detected biodiversity

### Short Introduction

Food security, managed and cryptic challenging environments, and many other research frontiers will be approached by adding genomic methods. Use case group 2 includes one use case for crop wild relatives and genetic resources for food security and another on DNA detected biodiversity in poorly known habitats. Grass pea is a legume crop which is very resilient to drought and thus one of the last remaining food resources when all other crops fail. Unfortunately grass pea also produces a neurotoxin which causes a gradual and irreversible paralysis of the knees (known as Lathyrism). The other use case was not presented at the meeting, but includes examples based on DNA detected biodiversity,, one focussing on multitaxa soil microbiota in one country (Denmark), the other on fungi at a global scale. A third case is also in planning, on soil bacterial phylogenetic diversity.



### Scope

The grass pea (*Lathyrus sativum*) use case is designed for the digital twin to support the hunt for alleles in crop wild relatives of *Lathyrus* in Ethiopia linked to a lower toxicity to be used for crop breeding of the grass pea crop. The use cases on cryptic habitats focus on species groups (fungi, bacteria, protists) that are not readily detectable by other than genetic methods. The geographic scope is regional (one case) to global (two cases), and the taxonomic scope is very broad, covering microbial biodiversity in the broad sense (bacteria, fungi, protists and other microbial eukaryotes). The modelling approach will be community composition based.



## Challenges

The main challenges of the “cryptic habitats” use case(s) are likely that datasets are large, and that a very large proportion of the diversity does not carry unambiguous taxon labels/names.



## Benefits

The main benefits of the “cryptic habitats” modelling will be to showcase how large datasets of cryptic biota can be modelled and visualised at large geographic scales, with the possibility of adding new data as it is produced.



## Partners Involved



UiO : University of Oslo





# Group 3

## Dynamics and threats from and for species of policy concern

### Short Introduction

We focussed on the presentation of alien invasive species, which are covered in the EU [Regulation 1143/2014](#) on Invasive Alien Species and the [List of Alien Invasive Species of Union concern](#). Invasive alien species (IAS) negatively affect biodiversity, ecosystem functioning, and human well-being, bearing a substantial economic cost to European and global society (Pyšek et al. 2020; Diagne et al. 2020). The impacts of IAS will likely exacerbate in the future due to new species introductions (Seebens et al. 2017) and the synergy with other drivers of global change (Pyšek et al. 2020). To mitigate the impacts of IAS, it is essential to accurately predict where and under which conditions new invasions might occur and what might be the overall extent of those invasions.

### Scope

We will restrict our analysis to IAS in Europe, but try to cover as much temporal information as possible, to allow for modelling spatio-temporal dynamics. With respect to taxa, we will start with vascular plants and add additional groups depending on data availability and ecological expertise. We will model the probability of occurrence, taking climatic suitability and habitat affiliation of the species into account.



## Challenges

Data sources are very heterogeneous in several aspects: spatial and temporal extent as well as resolution, taxonomic coverages, biases in the distribution of occurrence data. Different countries will have different lists of IAS, but some scientific consensus is available through [GRIIS](#). Early 2023, a new checklist of alien invasive plant species will be published.

Despite having some good knowledge on spatial distribution and some long-term data sets being available (e.g., via eLTER), we still lack good knowledge for most taxa in most locations on their temporal dynamics.

The input data will result in large rectangular and triangular (distance) matrices.

Land Management (intervention) scenarios are difficult, because any given policy change (e.g., EU Common Agricultural Policy) can result in unforeseen changes of land use.

Spread scenarios will be difficult, because different species will have different spread rates (dispersal kernels) in different habitats (land cover types).

## Benefits

We will deliver work flows that unify so far decentralised and heterogeneous data sources, to be directly fed into the models. This will allow for the presentation of “real time” dynamics in the levels of biological invasion and provides sound foundations to quantify changes. It may also serve as an “Early Warning” tool.

Intervention scenarios (based on assumed changes in land cover and land management) will help managers and policy makers in the decision and deliberation processes.

## Partners Involved



et al.



# Group 4

## Species interaction with each other and with humans



### Short Introduction

Group 4 includes two use cases, on specific species and dynamics, i.e. “pollinators” under multiple stress and disease dynamics mediated by “invasive species”. The very species to be tackled as “invasive species” is still under discussion, but a potential candidate is African swine fever transferred by wild boars and possibly also domestic pigs. For “pollinators”, the honey bee model BEEHAVE (Becher et al. 2014) will be used as it is well-tested and has been applied in more than 20 publications. While honey bees cannot fully represent all issues concerning pollinators in general, it is a species of high public concern and a key species for pesticide risk assessment of bees. BEEHAVE was designed to represent multiple stressors and their interaction: mites, diseases, forage gaps, pesticides, and beekeeping practices.

### Scope

BEEHAVE simulates the dynamics of a single honey bee colony in an agricultural landscape of 5x5 km<sup>2</sup>. As input, it requires weather data that determine whether the bees will go out for foraging, and the area, position, and crop type of all agricultural fields in the modelled landscape. For each crop, data on begin and end of the flowering period are needed as well as the amount of pollen and nectar provided each day by each field. The scope of BEEHAVE within BioDT is to be run, pixel by pixel, for the whole of Germany, where each pixel has the size of the original BEEHAVE landscape, 5x5 km<sup>2</sup>.

## Challenges

Weather data are easily available, but for specifying the land use, maps produced from satellite data (Sentinel 2) via machine learning will be used. Such maps are produced, for example, at UFZ with a resolution of 20 m, but the most recent maps are from 2019. Furthermore, BEEHAVE needs to be tested and recalibrated for different regions. For this, hive weight data of more than 500 hives in Germany, which are registered every 5 minutes, will be used. Both running BEEHAVE for the whole of Germany, and running 1000s of simulations of calibration and uncertainty analysis, will require a super-computer, such as LUMI. The data base for the amounts of nectar and pollen provided by different crops, and plant species in general, is scattered, incomplete, and shows huge variation.

## Benefits

The ultimate goal is to develop a map visualising the risk that in a certain landscape forage gaps will weaken the colony. Such a map can then inform greening scenarios and the risk management of pesticide applications. Moreover, treatment of the varroa mite can be regionalised.

## Partners Involved



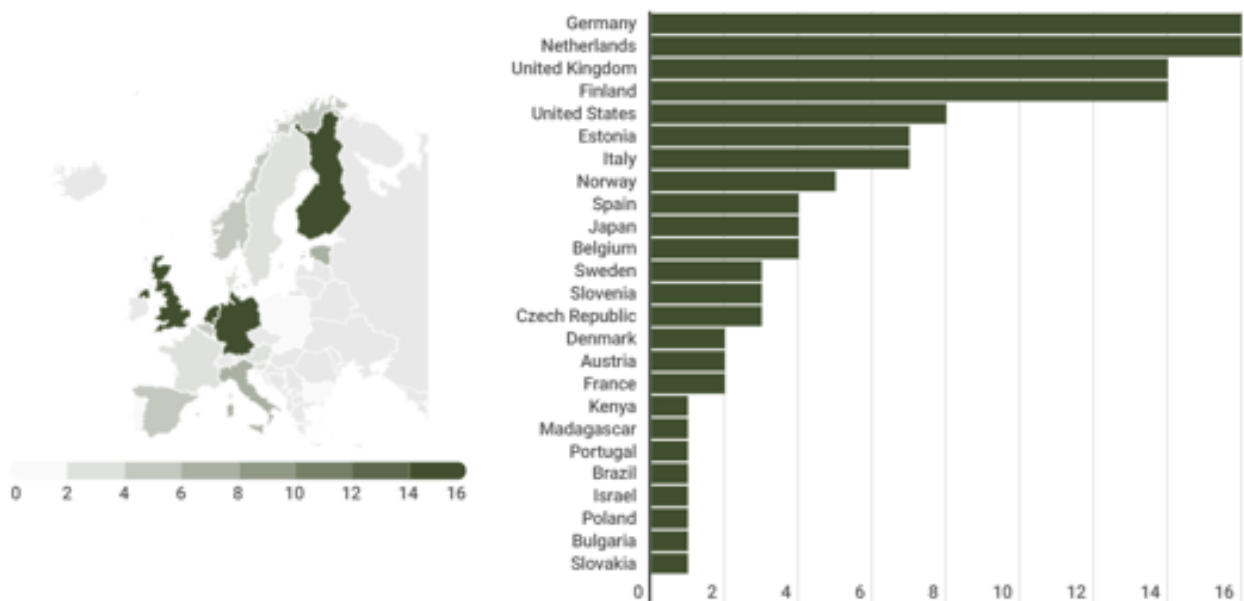

---

Reference: Becher, M. A., Grimm, V., Thorbek, P., Horn, J., Kennedy, P. J., & Osborne, J. L. (2014).

BEEHAVE: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure. *Journal of Applied Ecology*, 51(2), 470-482.

# Interactive session with the audience

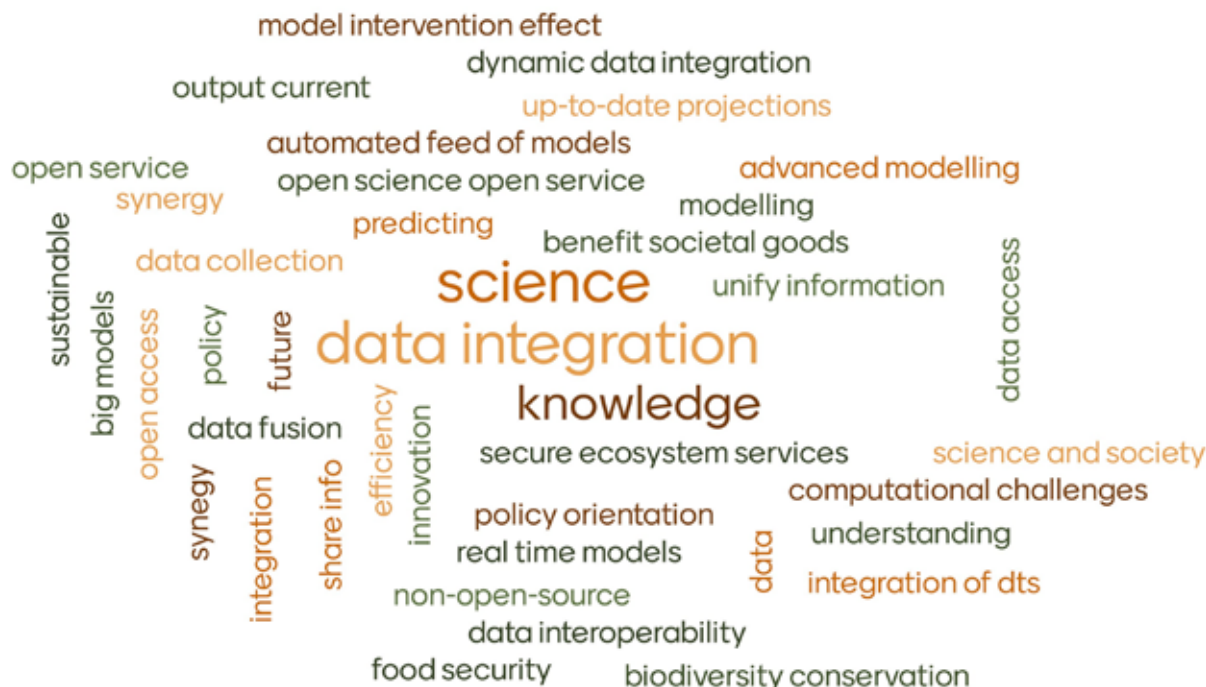
The BioDT use cases webinar attracted a total of 93 participants from 21 countries.







## Describe with max 3 words the main benefits brought by the BioDT use cases



## For which purpose have you ever dealt with the data connected to the topics of the BioDT use cases?



To monitor terrestrial fauna and adaptation to its habitats for conservation and restoration initiatives

To map EO remote sensing for habitat

To model crop traits in landraces

To analyse macroecological resilience

To manage molecular biodiversity data

## Which challenges have you experienced in developing use cases built on digital twin models?



Understanding what the digital twin is

Scalability of data

Difficulties in implementation

Biodiversity is too complex to be covered with use cases

Data acquisition

Obtaining comparable data from different datasets

Data standardisation and interoperability

Lack of common language between scientists and computer engineers

Need of high-quality FAIR reference datasets

Lack of analysis and decision making tools

Non open-source data

## What are you expecting from the Biodiversity Digital Twin?



Data accessibility and sharing for predictive analysis

Improvements in data integration

Extension of use cases to freshwater ecosystems

Platform to run massive virtual experiments

Better understanding of data integration

Interactive modelling across parallel aspects of biodiversity dynamics

Be potentially integrated into the framework of genomics observation

Learning workflows to transfer best practices to other areas

Basis to make progress in the biodiversity research field

Possibility to offer a platform to move modelling from methods to services

Following the interactive session, Dmitry Schigel, Scientific Officer at GBIF Secretariat and BioDT Use Cases Co-leader, moderated the Q&A one. The topics addressed during this session are summarised below:



Following the interactive session, Dmitry Schigel, Scientific Officer at GBIF Secretariat and BioDT Use Cases Co-leader, moderated the Q&A one. The topics addressed during this session are summarised below:

- 🔥 Future scenarios and topics observed by the Biodiversity Digital Twin;
- 🔥 Interoperability rules to integrate the Digital Twin into the LUMI Supercomputer;
- 🔥 Digital Twin and species distribution modelling;
- 🔥 Data and semantic rules adopted to model the use cases;
- 🔥 Benefits and future scenarios of the use cases;
- 🔥 Possible collaborations with external institutions for the development of the use cases.





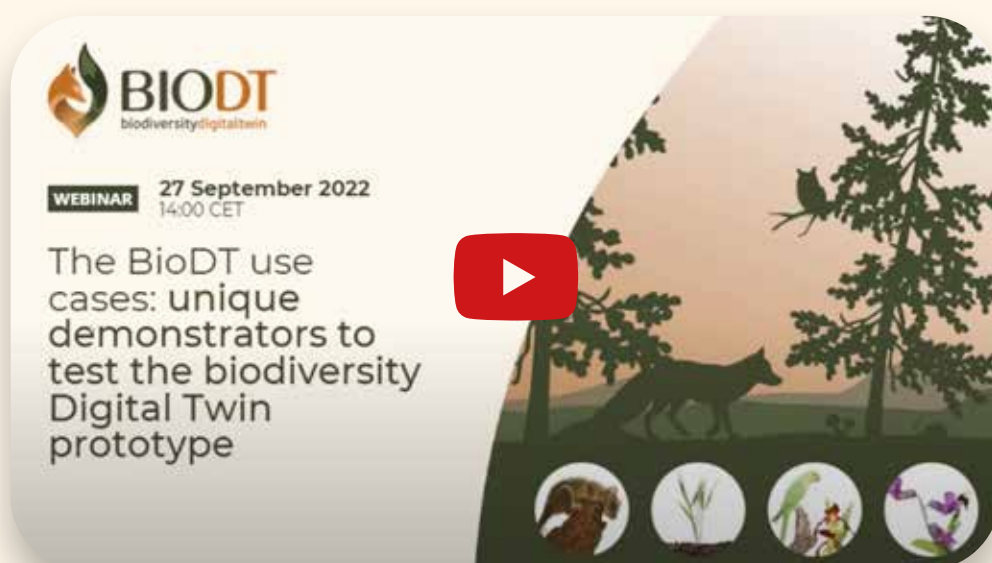
## Conclusions and next steps

The main technical advancements of the BioDT use cases are going to be presented in a series of dedicated workshops and webinars in 2023 - 2024.

Follow BioDT on the main project's social media channels and visit the website to follow the development steps of the use cases.

 [biodt.eu](https://biodt.eu)    [@BiodiversityDT](https://twitter.com/BiodiversityDT)    [/company/biodt](https://www.linkedin.com/company/biodt)    [BioDT](https://www.youtube.com/BioDT)

In the meantime, we invite you to watch the webinar recording on the [BioDT YouTube channel](https://www.youtube.com/BioDT).





# BIODT

biodiversitydigitaltwin



[biodt.eu](https://biodt.eu)



[@BiodiversityDT](https://twitter.com/BiodiversityDT)



[/company/biodt](https://www.linkedin.com/company/biodt)



[BioDT](https://www.youtube.com/BioDT)



Funded by  
the European Union

