

Proposal full title	DIGITAL TWIN FOR ADVANCED SIMULATION, MODELLING AND PREDICTION CAPABILITIES
Proposal acronym	BioDT
Call identifier	HORIZON-INFRA-2021-TECH-01-01
Type of action	Research and Innovation action (RIA)
Coordinator	CSC- IT Center for Science Ltd.
Person in charge of the proposal	Dr. Damien Lecarpentier, damien.lecarpentier@csc.fi Director, CSC – IT Center for Science Ltd.

No.	Participant organisation name	Country
1	CSC-TIETEEN TIETOTEKNIKAN KESKUS OY	Finland
2	TRUST-IT SRL	Italy
3	COMMPLA SRL	Italy
4	VSB - TECHNICAL UNIVERSITY OF OSTRAVA	Czech Republic
5	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO	Netherlands
6	JYVASKYLAN YLIOPISTO	Finland
7	TARTU ULIKOOL	Estonia
8	KUNGLIGA TEKNISKA HOEGSKOLAN	Sweden
9	SWANSEA UNIVERSITY	United Kingdom
10	THE UNIVERSITY OF MANCHESTER	United Kingdom
11	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	United Kingdom
12	HELMHOLTZ-ZENTRUM FUR UMWELTFORSCHUNG GMBH - UFZ	Germany
13	UK CENTRE FOR ECOLOGY & HYDROLOGY	United Kingdom
14	UMWELTBUNDESAMT GESELLSCHAFT MIT BESCHRANKTER HAFTUNG (UBA GMBH)	Austria
15	HELSINGIN YLIOPISTO	Finland
16	GLOBAL INFORMATION FACILITY	Denmark
17	UNIVERSITETET I OSLO	Norway
18	STICHTING NATURALIS BIODIVERSITY CENTER	Netherlands
19	SENCKENBERG GESELLSCHAFT FUR NATURFORSCHUNG	Germany
20	E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND ECOSYSTEM RESEARCH	Spain
21	BAYER AKTIENGESELLSCHAFT	Germany
22	MARTIN-LUTHER-UNIVERSITAT HALLE-WITTENBERG	Germany

1. Excellence

1.1 Objectives and ambition

With our planet facing accelerating pressures on biodiversity, it is of the utmost importance to understand the influence of climate change and human activities on biodiversity and how their effects can be mitigated and even reversed. Biodiversity provides essential resources and services that we use and that directly impact human wellbeing including providing us with sources for food, nutrients, and energy resources, as well as being critical to both the production as well as decomposition of organic matter. Over 70% of essential medicines and pharmaceuticals are derived from biological sources, and there is mounting evidence that a lack of contact with natural biodiversity impacts the human microbiome with negative health consequences.

Understanding the forces shaping biodiversity is the basis for rational management of natural resources and is key to meeting the EU Biodiversity Strategy 2030¹ to put Europe's biodiversity on the path to recovery by 2030. In particular, we need to be able to ***better predict global biodiversity dynamics*** and how species interact with their environment and with each other. This is an extremely difficult task because the processes underlying biodiversity dynamics are complex. Novel ways to combine data, models and interaction processes are required to predict these dynamics and offer solutions that promote sustainable management of the earth's biodiversity and its ecosystems.

Our goal is to push the current boundaries of predictive understanding of biodiversity dynamics by developing a ***Biodiversity Digital Twin (BioDT) providing advanced modelling, simulation and prediction capabilities***. By using already existing technologies and data from relevant research infrastructures in new ways, the BioDT will be able to accurately and quantitatively model interactions between species and their environment. Scientists at Research Infrastructures (RIs) will use the BioDT to 1) better observe changes in biodiversity in response to forces that result from climate change or human activity), 2) mechanistically understand how these changes occur and 3) predict the effects of these changes. BioDT will provide a crucial infrastructure to drive long-term biodiversity research and facilitate science-driven policy and rapid-response actions to maintain our current commitments to protecting biodiversity and aid its long-term protection.

Eight use cases clustered in four groups grounded in the scientific and technical expertise of BioDT will be at the core of the BioDT prototype. These use cases focus on the species and ecosystems of highest conservation and policy concern, such as threatened species, pollinators, and grasslands, and are vital to mankind's well-being and biodiversity conservation efforts. They address global issues of critical societal interest including *climate change impacts on species and ecosystems, food security, implementation of EU and international policies and health, and contribute to the UN Sustainable Development Goals 2 - Zero hunger, 3 - Good health and well-being, 13 - Climate action, and 15 - Life on land.*

Our consortium brings together a dynamic team of experts in biodiversity, high performance computing, artificial intelligence, digital twinning and FAIR data to develop the first BioDT prototype. It combines the unique potential of the **flagship EuroHPC LUMI computing infrastructure** with the scientific expertise and existing biodiversity data from four major **research infrastructures** (GBIF, eLTER, DiSSCo, and LifeWatch ERIC). This allows us to cover several application domains such as environmental and earth science, climate science, ecology, marine biology, genomics, natural history, biodiversity informatics, computer sciences, and mathematics / statistics.

We envision that BioDT and its infrastructure will be an integral component of the Destination Earth initiative and its ambition to realise a full Digital Twin of the Earth. The long-term objectives of BioDT are also tightly interconnected with the EC vision for a robust, federated European computing and data infrastructure and initiatives such as the European Open Science Cloud (EOSC) and EuroHPC.

1.1.1. BioDT objectives and main measurable outcome

¹ https://ec.europa.eu/environment/strategy/biodiversity-strategy-2030_en

OBJECTIVE 1: Build and deploy a pre-operational BioDT for addressing biodiversity dynamics

Specific Objectives:

- Set up, develop and operate the **Digital Twin platform**, leveraging LUMI's high performance computing, AI and data analytics capabilities (WP3)
- Develop and package societally relevant **multidisciplinary case studies** so that they provide the necessary information for both developing and testing the models (WP4)
- Improve the **accuracy and predictive performance** of the **models & simulations** for biodiversity use cases in an iterative approach that will **build a methodology for twinning** and **process for quality and validity of results** (WP5, WP6)
- Generalise the BioDT platform to **serve user needs** beyond the consortium (WP3, WP7)

Pertinence to the topic of the call: advanced modelling, simulation and prediction capabilities for research data infrastructures through a convergent use of advanced digital technologies such as high performance computing, software, AI methods and big data analytics; development of more integrated systems and consistent sets of standard methods and protocols for model and data fusion and visualisation as well as AI-based knowledge generation from spatio-temporal information.

Main measurable outcome: A prototype BioDT available *as a service* for Biodiversity RIs, harnessing the power of LUMI's advanced capabilities in HPC, AI, and data analytics; eight multidisciplinary cases studies showcasing its usage and performance; improvement of models' predictive performance: increased accuracy and precision in rigorously and properly (cross)-validated models (WP6).

OBJECTIVE 2: Support the interoperability of data and services through the integration of the BioDT with research infrastructure platforms and workflows

Specific Objectives:

- Support the **interoperability of data, software and community practices** of research infrastructures involved (how to combine available data, how to integrate new data, etc.) (WP4, WP5)
- Support the **integration of the BioDT into community platforms**: Authentication and authorization interfaces (AAI), application programming interfaces (API), access policies, etc.) (WP7)
- Ensure the **uptake of the BioDT** by research communities, through the co-design approach (WP2, WP7)
- Acquire **new user communities** to increase the user base of the offered services through the project lifetime and build a foundation to scale up across audiences and realms (WP2, WP7)
- Create **training material with best practices and guidelines** for researchers to develop and interact with the BioDT, leading to more harmonisation across RI (WP7)

Pertinence to the topic of the call: co-design approach with user communities; development of quality measures and trust, development of standard quality mapping and indicators for communicating differences in qualities of inputs and outputs; delivery of breakthrough in terms of accuracy and realism, fusion of observations and models, integration of downstream sectors at the source of data production, rigorous handling of quality and confidence of information; development of capabilities enabling research communities to continuously learn and update themselves from data and information originating from different sources.

Main measurable outcome: APIs that facilitate the use of existing BioDT software infrastructure to feed data into the BioDT platform (WP3, WP7); creation of semantically annotated FAIR datasets using cross-RI standards & protocols (WP5); training material for cross-RI usage (WP2, WP7); improved machine-actionability, interoperability and reusability of data through the FAIR Digital Objects (FDO) layer (WP5); development of quality indicators, such as FAIRness assessment metrics and geographic data accuracy (WP5); interoperability

workshops with RIs and researchers (WP4); improved modelling, simulation and prediction capabilities for RIs and their research communities (WP6).

OBJECTIVE 3: Ensure interoperability of the BioDT with Destination Earth, and the European Data Infrastructure

Specific Objectives:

- Provide a **continuous interface between the DestinE developments** managed by ECMWF and the technical developments in this project such that software platforms, data handling and information extraction between **Climate DT and BioDT** can be synchronized, apparent development gaps can be identified and solutions be developed for implementation in future phases of DestinE. (WP8)
- Integrate relevant **datasets available in EOSC** and make results openly available via the EOSC Portal / Marketplace and through existing EOSC services. (WP3, WP8)
- Create showcases for interaction between the BioDT and other DTs such as **the Ocean DT** (WP8)
- Harmonize data and data governance according to the **EU Data Spaces** (WP8)

Pertinence to the topic of the call: integration into operational activities of existing research infrastructures and the EOSC platform; link to relevant actions, under the Digital Europe Programme such as Destination Earth, EuroHPC, and the data spaces.

Main measurable outcome: BioDT data outputs to the DestinE data lakes (WP8); interfaces that facilitate a fully automatic publication of installation information of BioDT data in the EOSC Portal, and automated processing of orders from the EOSC Portal (WP3); integration of DestinE output data for use by the BioDT through RIs (WP4); integration of GBIF data collection available in the EOSC marketplace with the BioDT platform (WP8).

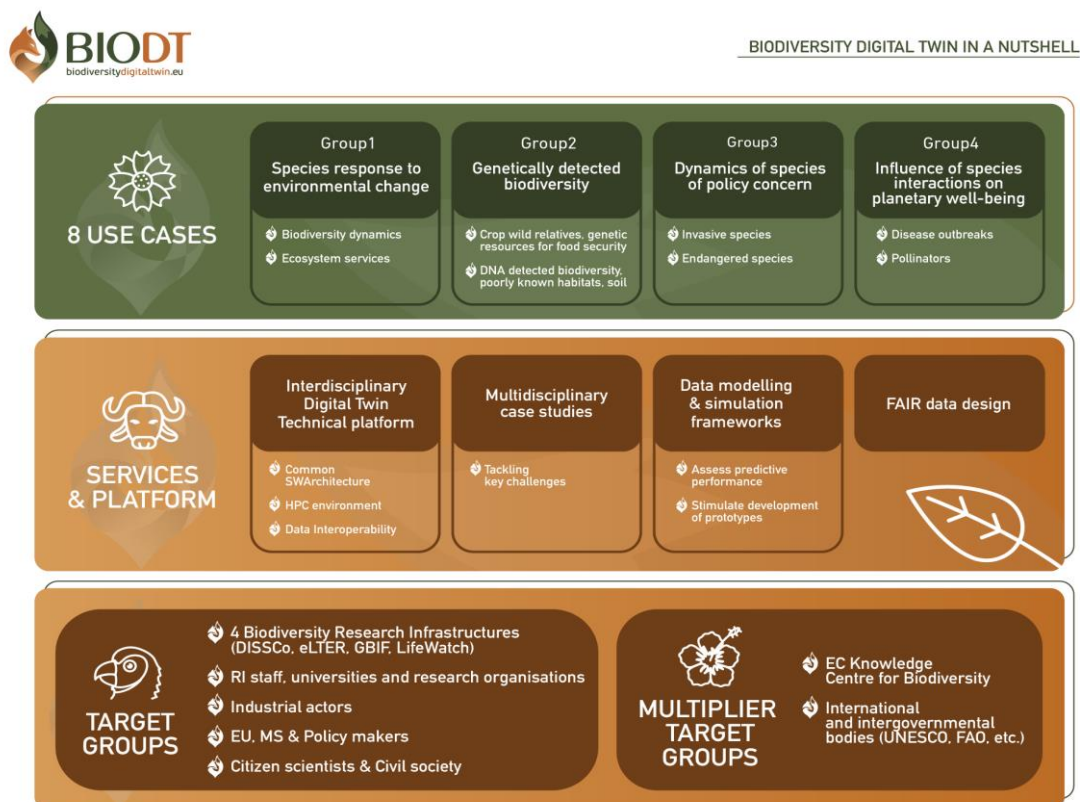


Figure 1: BioDT in a nutshell

1.1.2. Progress beyond the state of the art, innovation and ambition

The key aims of biodiversity research relate to inference of causal relations and prediction. These two are inextricably linked, as understanding the underlying mechanisms should improve our ability to make predictions. Yet, this remains largely untested in biodiversity research. At one extreme, predictions are generated by phenomenological (or statistical) models fitted to observed data, without fully understanding the causal relationship, whereas at the other extreme, predictions may be generated by mechanistic (or process-based) models fitted to experimental data.

We hypothesize that due to the complexity underlying biodiversity dynamics, state-of-the-art phenomenological models still outperform current mechanistic models in predictive tasks, except for highly simplified case studies. If true, this highlights the current gap of process-based understanding in biodiversity dynamics. We propose that hybrid models, which will combine the generality and relevance of phenomenological models with the causality of mechanistic models, can outperform both phenomenological and mechanistic models in prediction and therefore push the current boundaries of predictive understanding of biodiversity dynamics.

In BioDT, we will approach this by bringing mechanistic elements to statistical models, e.g. by informing model structures and constraining model parameters with experimental data and ecological knowledge. Conversely, we will integrate statistical elements into agent-based models (which capture the behaviour of individuals within their environment) through, for example, dependency of transition rates on covariates and spatio-temporal random fields. The major outcome of the project, from a research perspective, will be quantitative agent-based modelling of biodiversity, accompanied by a critical evaluation of the predictive performance of this new modelling framework, when compared to more traditionally used statistical approaches in biodiversity research.

Implementation of agent-based models relies on computationally intensive stochastic simulation, whereas evaluation of the predictive performance also requires data-intensive approaches including big data processing, machine learning and deep learning methods. As the amount of data to be passed between the computing modules is large, the computing environment needs to be well connected. BioDT will rely on the flagship EuroHPC LUMI supercomputer, which provides machine learning and deep learning capability, simulation capability and data processing at the highest performance, in a single environment. These capabilities are supported by storage partitions and connected with a high-speed interconnect. Using the LUMI infrastructure as the basis for building the Digital Twin will allow us to vastly speed up calculations, generate new scientific insights and improve prediction capabilities related to biodiversity dynamics.

The multidisciplinary approach adopted by the project is also unique. BioDT is to date the most comprehensive effort to bring together different data types ranging from species occurrence data, museum specimens, emerging genetically detected species data, ecological monitoring, and ecosystem status. This will be essential for understanding global biodiversity dynamics. BioDT contributes directly to advancing global interoperability in biodiversity data and research, and reduce the current fragmentation of the biodiversity research landscape. This, in turn, will pave the way for future, multidisciplinary biodiversity research.

1.2 Methodology

1.2.1 A gradient of case studies in increasing complexity

BioDT covers a set of use cases that increase in complexity, ranging from simple (e.g. lab experiments of a two-species system) to intermediate (e.g. local data on a well-studied arctic ecosystem) to complex (e.g. local data on a well-studied tropical ecosystem) to very complex (e.g. spatially extensive data covering interactions among several ecosystem types). This increase of complexity is complemented with other relevant changes, such as quality and availability of data. It is also critical that how a model is validated is fully matched to its prediction task to ensure that both the underlying mechanisms and species responses to environmental change are understood. A key element of BioDT will therefore include rigorous model validation.

Group 1: Species response to environmental change. Use cases: 1. Biodiversity dynamics and 2. Ecosystems services.

Current status and problems: Current models that predict the response of species to environmental changes rely on statistical approaches. To increase the possible range of scenario's for species response with a better accuracy, a different approach is required based on mechanistic/process-based models.

How BioDT addresses these challenges: BioDT will use mechanistic/process-based models to model species dynamics in combination with validation steps that are more rigorous than those previously used. This presents a step change in our ability to both accurately model how species and ecosystems interactions and interrelations respond to global change, as well as the underlying mechanisms that lead to these changes. These models combine the sparse data from large geographical areas (e.g. extent data) provided by GBIF with the smaller set of high-resolution time-series data provided eLTER in a single modelling framework.

Aims and beneficiaries: This work will therefore improve how accurately and efficiently BioDT users can predict temporal and spatial changes in species and ecosystems as a response to climate change, as well as quantify the uncertainty in these predictions and their sources. We hence provide a long term and strategic infrastructure for biodiversity researchers and contribute towards sound policy-based decisions for the protection of Earth's biodiversity.

Group 2: Genetically detected biodiversity. Use cases: 3. Crop wild relatives and genetic resources for food security and 4. DNA detected biodiversity and poorly known habitats, including soil systems.

Current status and problems: For over 275 years, Linnaean nomenclature has been used to inventory and monitor the Earth's biodiversity. However, in order to meet current research and policy-based needs, DNA-based methods are needed. They will allow for rapid species detection, delimitation, and quantification. Food security and management as well as the study of cryptically challenging environments in particular are two areas best monitored and studied using DNA-based methods.

How BioDT addresses these challenges: BioDT will link DNA-based methods to crop wild relatives and genetic resources (use case 3) for both farming and food security and to detect biodiversity (use case 4) in poorly known habitats, including soil. As DNA-based data comes with unique computational and data handling challenges, these use cases require further development of data stream models, including access to data derived from material samples in biobank collections. We will therefore ensure that the models and workflows for DNA-data collection in BioDT address these issues.

Aims and beneficiaries: The models and predictions based on the DNA biodiversity data will further our understanding of the biodiversity present in arable lands and soil. Use case 3 results carry particular significance in defining policies for ensuring food security. Use case 4 will open new research fronts in hard-to-study environments that will be applicable beyond the project timeline. Results from both uses cases will target SMEs involved in DNA-based biodiversity monitoring as well as potential BioDT users.

Group 3: Dynamics of species of policy concern. Use cases: 5. Invasive species and 6. Endangered species.

Current status and problems: Several key features are shared between invasive or endangered species: both are inherently difficult to model due to their scarce occurrence records, they are often not in equilibrium with their environmental conditions, and both suffer from lag effects in their responses. No modelling approach currently exists that addresses these problems with an acceptable level of accuracy and reliability.

How BioDT addresses these challenges: BioDT will use a combination of revised and improved models making use of the availability of large extent spatial data with high-resolution temporal data provided by GBIF and eLTER. This will create a new generation of model predictions that accurately and reliably provide spatial and temporal predictions on invasive and endangered species dynamics in response to a suite of interacting drivers, including climate change, land use, and socio-economic activities, such as trades and traffic flows. In addition to the data

provided by BioDT participants, data from other EU in biodiversity projects will be used, including ALARM, DAISIE, AlienScenarios, EU-MON, EU-BON, EuropaBON, and EC21C.

Aims and beneficiaries: BioDT will be a ready-to-use service that delivers real-time results and information for stakeholders in public policy and land management services. This will enable EU and member states to set up evidence-based management policies and recommendations for the early and rapid management of potential sites where alien invasive species may establish themselves, as well as protecting endangered and threatened species.

Group 4. Influence of species interactions on planetary well-being. Use Cases 7. Disease outbreaks and 8. Pollinators.

Current status and problems: Global concern about *environmental change*, not only *climate change*, is increasing, as species to species interactions, including interactions with *Homo sapiens*, have transformed. This has been painfully and abruptly realized through the current Covid-19 pandemic. Additional environmental changes are accelerating food security issues, including the decline of pollinators. Pollinators are omnipresent in ecosystems and are a critical force in our food supply, though the risks their decline pose are not fully understood.

How BioDT addresses these challenges: BioDT will focus on predicting disease outbreaks using disease vectors and pathogen distribution data (use case 7) and explore patterns and processes of insect pollinators of wild and cultivated plants (use case 8). These species interaction data and modelling will require further development of the sampling event domain model (part of the Darwin Core data standard) and data exchange models. Pipelines for extracting interactions data off collection specimens will establish temporal historic reference points for both cases. Particularly important are pipelines that deliver maps of forage (nectar, pollen) availability in agricultural landscapes, based on land use, weather and phenology data. These maps provide input for pollinator models, such as BEEHAVE and Bumble-BEEHAVE for honey bees and bumblebees, respectively.

Aims and beneficiaries: BioDT will contribute to quantitatively modelling and predicting emerging diseases and their potential locations in Europe and can serve as an important source of information for public health officials, as well as fill the current knowledge gap on the impact of changes in pollinators. Digital Twins of pollinators in agricultural landscapes will provide testbeds for greening measures and can be linked to other twins, such as crop and grassland models.

The table below summarises the type of data which will be used in each study as well as the prediction tasks.

Groups and use cases	Prediction tasks	Data types to be used
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.	- Species occurrence and abundance spatiotemporal data - Abiotic data - Land use data - Carbon stocks - Water quality
Group 2: Genetically detected biodiversity 3. Crop wild relatives, genetic resources for food security 4. DNA detected biodiversity, poorly known habitats, soil	1. DNA sequence to taxon/genotype identification. 2. DNA sequence to phenotype, trait, or property of taxon, and genotype of organism. 3. Modelling patterns and processes in arable lands and cryptic environments 4. Modelling changes in soil, degrading organic matter, water, air.	- Species occurrence data - DNA sequence monitoring data including barcodes, metabarcoding data, and eDNA - Habitat data - Agrobiodiversity specific data
Group 3: Dynamics of species of special / policy	1. Changes in species richness and abundance of invasive species	EU and national lists of alien invasive species ("Union list",

concern such as threatened and invasive alien species 5. Invasive species 6. Endangered species	2. Changes in species richness and abundance of red list species.	national “black lists”, if applicable) • Data on climate • Land use • Trade exchange and traffic
Group 4: Influence of species interactions on planetary well-being 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.	• Diseases vector and reservoir distribution data • Pathogen distribution data • Plant and pollinator quantitative and interaction web data

Each case defines the working hypotheses, sets of data required, models needed, validation method and scenarios that can be explored. (WPs 3, 4, 5, and 6). These use cases will mobilize and involve the relevant scientific communities and major target groups. Dissemination, exploitation and communication strategies will actively engage SMEs and the general public to take part in BioDT as citizen scientists (WPs 2 and 7). To ensure that the approaches most commonly used by different user disciplines and domains provide comparable results, FAIR data and workflows will be used (WPs 5). This will provide common grounds for learning and cross fertilization of ideas between the scientific and technical communities to innovate and develop the tools that will solve our current limitations in modelling life on earth.

1.2.3 From use cases to modelling

The use cases will be used to develop the Digital Twin, test its predictive performance, and apply the BioDT to address biodiversity challenges through scenario simulations, predictions and biomonitoring methods. We envisage the simultaneous development of several modelling approaches (jointly or independently by several research groups) so that their predictive performances can be systematically compared, stimulating an iterative approach to developing next-generation prototypes. Initial research into this topic will be conducted using small-scale pilot studies, followed by model upscaling, implementation and comparisons using the EuroHPC LUMI supercomputer.

To connect the modelling approaches and associated methods with different use cases, standardized pipelines will be constructed for data processing, model execution and data visualization. To ensure maximal compatibility of the pipelines between use cases, individual pipeline components will be packaged as reusable building blocks. The pipelines will implement state-of-the art approaches including joint species distribution models, agent-based models (e.g. GRASSMIND) and theoretical models²³ to explore causality, and to develop theory allowing for predictions across use cases and systems. Crucially, the proposed project will seek to improve upon the performance of existing modelling methods through the development of hybrid modelling approaches that combine mechanistic and phenomenological approaches (Section 1.1.2).

A key property of a trustworthy model as well as of a credible theory is the ability to produce predictions that match observations outside of the training data. We will test for the limits of current predictive biodiversity science by developing a systematic toolbox of predictive model comparison methods for evaluating the transferability of ecological models. Beyond classical cross-validation approaches where data are split into training vs. test sets, we will develop a novel perspective on how cross-validation optimally mimics the actual predictive task, such as the change in biodiversity or ecosystem processes induced by climate change. Identifying where and why knowledge on underlying mechanisms improves prediction is central for understanding the current limits of predictive biodiversity dynamics, and this will be heavily investigated during the BioDT development.

² Cornell, S. J., Suprunenko, Y. F., Finkelshtein, D., Somervuo, P. and Ovaskainen, O. 2019. A unified framework for analysis of individual-based models in ecology and beyond. *Nature Communications* 10, 4716.

³ Thompson et al. 2020. A process-based metacommunity framework linking local and regional scale community ecology. *Ecology letters*, vol 23, issue 9. <https://doi.org/10.1111/ele.13568>

A critical challenge in the transition from use cases to implementing the models will involve identifying ways in which the proposed approaches can be upscaled to handle massive data sets and exploit the resources available on LUMI (Section 1.2.4). To address this challenge, we will pursue two independent lines of investigation including theoretical model exploration and incorporating support for parallel computing and/or GPU acceleration. To provide direct, easy access to the upscaled models, end-user features for interactive use of the BioDT (e.g. for biomonitoring) will be developed. The proposed sequence of steps to connect the use cases with actions related to model design and implementation is summarized in Figure 2.

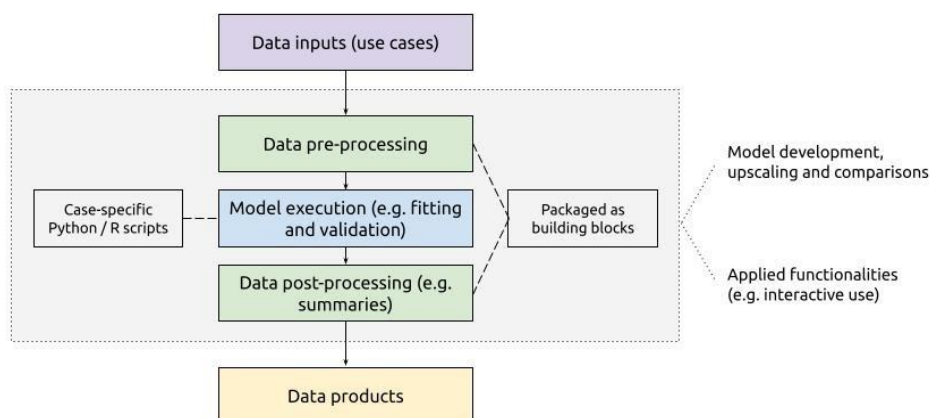


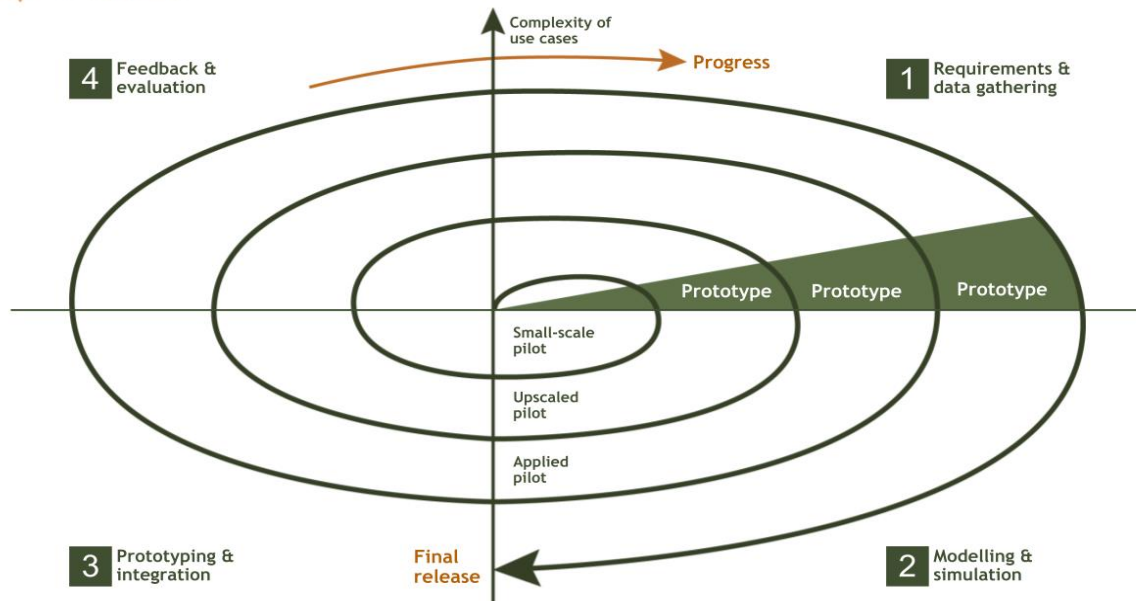
Figure 2. Flow chart illustrating the sequence of steps to connect use cases with actions related to model and data analysis pipeline development, model upscaling and implementation.

1.2.4 From modelling to prototyping

The proposed BioDT platform will be developed so that it can benefit from the advanced computational techniques provided by LUMI. The development of the BioDT technical platform will be done in an incremental and iterative process that is synchronized with the gradient of complexity in use cases. We recognise the inherent difficulty of building fully fledged Digital Twins for the complex and loosely bounded phenomenon of biodiversity. Hence, the process for developing the BioDT will be designed to be self-correcting: the platform will be built in increments so that experiences and feedback from previous increments can be used to guide the development of the forthcoming increments.

We will adopt a flexible architecture that allows rapid implementation of the first use cases, while remaining open for adaptation and customisation as the understanding of the technical priorities increases. The availability of the LUMI EuroHPC strongly supports the rapid development in the beginning of the project. The consortium brings in a deep understanding of the LUMI platform, together with the associated Puhuri project for resource allocation, tracking and federated group management. Building on the high-level services available the first increment of the technical platform can be deployed quickly.

The incremental development of the platform is driven by the gradient of use cases. By initially focusing on a less complex use case while building on an existing technological stack, we can cut down the lead time of the first prototype and thus maximise the time available for incremental further iterations. As a result, the later prototypes can benefit from the build-up of knowhow from the earlier ones. For example, it is hard to foresee the performance bottlenecks of HPC workloads in advance. As first models are built early and quickly integrated into a technical platform for prototyping it is possible to also start the performance profiling work early and have more time for the demanding task of HPC optimisation (Section 1.2.3).



The BioDT platform is intended to be a clean and lightweight system connecting existing and newly developed shared services for Digital Twins with the planned users of the platform through well-defined Application Programming Interfaces (APIs). We will ensure that users as well as the Biodiversity platform and shared services are supporting a common AAI solution and the system as a whole is compliant with GDPR to ensure that the DTs are easily accessible in a fully digital manner to researchers, BioDT will leverage LUMI's API-based access layer provided by NeIC's Puhuri project allowing easy and GDPR-compliant access to supercomputing facilities for pan-European collaborations, national research infrastructures as well as commercial entities.

1.2.5 Involving the community of users through a co-design approach

Keeping a user-centric approach to the BioDT design is critical to its success and is a key principle for our consortium. We will use an active and dynamic engagement strategy with research communities following the true principles of co-design. To ensure the creation of tools desired by the users, we will engage user communities and research infrastructures to provide their requirements and feedback during the integration process. These will be provided to all the technical partners as part of the co-design. In all the work packages, we have achieved a balance of IT experts and domain experts to ensure that the solutions designed meet the requirements of the future users.

We will perform Bring Your Own Data (BYOD) and Bring Your Own Model (BYOM) hackathons to reach out to the communities beyond the consortium and quickly familiarize them on the use of the Digital Twin services, data, and model. We will target life science (e.g. ELIXIR), marine, and climate science communities, and make use of our partner networks and ongoing working relationships to engage with the relevant organisations and data holders.

1.2.6 Involving other (Digital Twin) communities

Beyond the scope of biodiversity, other Digital Twins are being created in other projects and initiatives, such as Destination Earth (DestinE) creating more accurate Twins for weather, climate and hazards & risks, covering the atmosphere (ECMWF). Another initiative is the Iliad project where a Digital Twin of the Ocean will be created covering the marine part of our planet. Several nations have initiated Digital Twins of the topography and sub-surface areas. All these different Digital Twins will be part of the future wider picture to create a Digital Twin for the Earth

where Twins can be linked together. BioDT will cover the species and nature aspects for the earth and will need to fit into this family of Twins and be accessible for any other future initiative.

This requires aligning with the other Twin initiatives in terms of infrastructure, data, interoperability, metadata, etc to make sure the collaborations are possible. During the iterations of developing the BioDT, proof of concept demonstrators will be created that feature interactions between Twins from different initiatives, such as the climate model. The BioDT will also be made available in the European Open Science cloud using APIs that will allow other researchers to use the BioDT. Another link will also be made to the EU Data Spaces initiative to ensure data and the BioDT models are available for anyone in the EU with FAIR rules and the EU principles and values for use of data.

1.3 Open Science & FAIR and interoperable data

Building Digital Twins to study global biodiversity dynamics requires data from a cross-domain perspective and data derived from a wide spectrum of sources. Over the years, the FAIR principles⁴ have emerged as a guideline for addressing such heterogeneous and complex data sources. We use FAIR principles as a reference point to ensure that data within the BioDT platform can be organised and managed to be easily *Findable*, *Accessible* as widely and easily as possible, in a manner that is user-friendly and machine-readable, machine-actionable, *Interoperable* so it can be linked with other data and Digital Twins; and *Reusable* and *Reproducible* so it is easy to re-purpose and re-create. We look at data integration issues by addressing the variety in community data standards, schemes, protocols and quality of the data sources by consolidating the RI FAIR data designs through organising the data with the Fair Digital Objects (FDO⁵) layer. FDO offers an object-oriented framework that is modular, actionable, meaningful, technology independent and FAIR by design.

The FDO layer and the FAIR approach will underpin the architecture blueprint for the BioDT technical platform and will have practical and long-term implications for data, workflows, and models. For example, usage of persistent identifiers, typed objects and machine actionable metadata as components of FDO will improve Findability and Accessibility in alignment with the EOSC Persistent Identifiers (PID) Policy⁶ and the EOSC Interoperability Framework⁷.

Along with using FAIR principles as guidelines, a cross-domain semantic mapping will be piloted between digital objects to further improve Interoperability. As the discoverability and accessibility are essential aspects of the FAIR principles, semantic mapping ensures data and other entities can be integrated and used despite heterogeneity of the baseline datasets. Various data staging, processing, and sharing activities in the LUMI and other HPC environments can make use of semantic mapping for integrating data using different ontologies and vocabularies. The semantic mapping approach can be combined with existing tangible results produced by the ESFRI cluster projects (e.g. ENVRI FAIR, EOSC-Life).

The level of FAIRness of research data output will be measured through appropriate FAIR assessment tools⁸. These tools along with maturity indicators define core criteria to assess the implementation level of the FAIR data principles and provide a structured yet flexible approach in assessing FAIRness. These tools and methods will be used in the development of semi-automated indicators (e.g. usage of persistent identifiers, ensuring that metadata is represented using a formal knowledge representation language) of quality for the datasets. Quality indicators, such as FAIRness

⁴ M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* [Internet] 3 (2016), 160018. Available at <http://www.nature.com/articles/sdata201618>.

⁵ <https://fairdo.org/>

⁶ <https://doi.org/10.2777/926037>

⁷ <https://doi.org/10.2777/620649>

⁸ Examples of such tools and methods: <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>, FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. DOI: 10.15497/RDA00050

assessment metrics and geographic data accuracy can provide processes for consistency that are needed in different stages of workflows using a Digital Twin. These indicators are also crucial for compliance with recognized community standards and benchmarking for performance tests. The level of FAIRness and other quality indicators based on the quality enhancements coming from RI data streams (e.g. geospatial and temporal data) will be used to build the quality mapping framework. The FDO layer and data provisioning through the RIs will provide the researcher with reliability and verifiability of the data, while the quality mapping framework will also guide the user about the quality of the data and expected outputs of the BioDT. The use cases from WP4 will be aligned with the FAIR framework and will provide input to develop quality indicators in WP5 for the accuracy of the developed models.

The FAIR principles are not only applicable to data but also to other infrastructure components such as services, software, workflow and compute engines as discrete, yet interconnected digital objects. Applying FDOs and FAIR semantics and mapping services to all digital objects provides a harmonised and interlinked interoperability fabric for the BioDT infrastructure. FAIR principles for workflows⁹ and models will be applied, and findability and accessibility supported by adopting community-developed schemas¹⁰ and EOSC-Life deployed registries such as ELIXIR's WorkflowHub¹¹. Work will exploit the latest developments using RO-Crate¹² to deploy FDOs for workflows and models in VRE platforms such as Galaxy¹³, LifeWatch ERIC's Tesseract, and Jupyter notebooks. The FDOs facilitate integration with the Common European Data Spaces and the European Open Science Cloud, and will lower the barrier for developers to compare and exchange models of Digital Twins and for researchers to use them.

Data Management Plan & Open Access Strategy

The BioDT project will operate in line with the Commission's Open Access to research data policy for facilitating access, re-use and preservation of its research data. A Data Management Plan (DMP) will be developed by month 6 and documented in a deliverable according to the EC guidelines on FAIR data management. This will outline how research data will be handled during and after the project, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. The DMP will also provide the dataset metadata specification that will be used in the data registry, following an appropriate relevant standard. It will specify the recommended licensing schemes, preferably using the Creative Commons Public Domain (CC0) and Attribution (CC BY) licenses as suggested by Horizon Europe. In the cases where the datasets cannot be publicly shared, the reasons will be mentioned in its metadata description (e.g. ethics, rules on personal data use, intellectual property, commercial, privacy-related, security-related).

Knowledge management & IPR strategy

In achieving this aim, the BioDT project recognizes the importance of intellectual property and its proper treatment. A project Intellectual Property Policy will be drafted by WP1 at the beginning of the project. The purpose of this policy is to clarify the rules, principles and recommendations in regards to the intellectual property aspects of the Collaboration as well as to set recommendations regarding such aspects. The following principles will be put forward: a) a party of the BioDT project owns the IP created by it and, b) the BioDT consortium will (non- exclusively) manage and exploit the IP for the benefit of the project. Regarding the licensing, the policy will recommend that, if

⁹ Carole Goble, Sarah Cohen-Boulakia, Stian Soiland-Reyes, Daniel Garijo, Yolanda Gil, Michael R. Crusoe, Kristian Peters, Daniel Schober; FAIR Computational Workflows. Data Intelligence 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

¹⁰ <https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE/>

¹¹ <https://workflowhub.eu>

¹² <https://www.researchobject.org/ro-crate/>

¹³ <https://galaxyproject.org/learn/advanced-workflow/>

permitted by the license terms and conditions applied, the software IP is licensed under an open source license. Particular license recommendations are given: a) The MIT License for software, and b) Creative Commons BY 4.0 license, Open Data Commons Attribution License v. 1.0 or Open Data Commons Open Database License (ODbL) v. 1.0 for non-software.

2. Impact

The Biodiversity Digital Twin will provide both **advanced models for simulation and prediction capabilities, via the practical use cases** to address critical issues related to global biodiversity dynamics. BioDT will enhance the protection and restoration of Europe's ecosystems and biodiversity by managing natural resources sustainably. This will ensure clean and healthy environments, as well as contribute to increasing food security and sustainability of the food system.

Efficient facilitation of evidence-based solutions for biodiversity loss and restoration is at the heart of BioDT by employing accessible and interoperable data combined with digital infrastructure, predictive modelling and AI solutions. BioDT responds to key EU and international policy initiatives, including the EU Biodiversity Strategy 2030, EU Green Deal, UN Sustainable Development Goals (in particular *SDG 2* Zero hunger, *SDG 3* Good health and well-being, *SDG13* Climate action, and *SDG15* Life on land), Destination Earth, and capitalises scientific and technological excellence the frontline of European innovation (e.g. LUMI Supercomputer).

2.1 Project's pathways towards impact

Involving leading biodiversity RIs in the co-creation of the BioDT will expand BioDT result impacts. BioDT will strengthen the operational services already provided by RIs as well as create new opportunities for novel uses of the RIs. Novel BioDT capabilities will be mediated by the national RI/GBIF node teams established in most European countries. Integration of BioDT outputs with the RIs and interoperability with and influence on the other digital twins under development in Europe will be the primary pathways towards impact created by the BioDT.

2.1.1 Unique contributions in relation to the specific topic of the call, and wider impacts

The table below summarises the main impacts of BioDT and positions our initiative in relation to the specific call topic scope of developing "Next generation of scientific instrumentation, tools and methods".

<i>Contributions to the outcomes of the topic and wider impacts HORIZON-INFRA-2021-TECH-01-01 Call</i>
<i>[...] a pre-operational prototype of an interdisciplinary Digital Twin [...]</i>
<u>Impact #1 - Technological</u> : 4 biodiversity RIs will contribute to developing, testing and using the BioDT prototype built upon various advanced technologies such as HPC and AI, agent-based modelling, statistical modelling and hybrid models combining mechanistic and phenomenological approaches.
<u>Impact #2 - Scientific</u> : European high-resolution maps of land use as basis for any research on biodiversity in agricultural landscapes.
<u>Wider impacts</u> : Become a model for other thematic digital twins
<i>[...] latest modelling and prediction technologies [...] and supporting interoperability of data and software, integration and collaboration across different scientific domains, disciplines and across the different research infrastructures involved;</i>
<u>Impact #3 - Scientific</u> : 8 BioDT use cases aim to involve end-users and create an iterative approach to developing a set of modelling approaches and to test, compare and evaluate their predictive performance through cross-cutting scenario simulations and biomonitoring methods.
<u>Impact #4 - Scientific</u> : Target groups will have access to enhanced and integrated data as well as modelling tools to provide new innovative services.

Impact #5 - Scientific: BioDT will engage research communities through a series of web tools such as fora, wikis, polls, and brainstorming boards, establishing a bottom-up channel where members can communicate and express their needs, as well as ensuring a multi-directional space for discussion and exchange where different stakeholders can meet virtually. This will be hosted on the LifeWatch ERIC Community Platform, which is projected to become a flourishing online community focused on biodiversity and ecosystem research, monitoring and management.

Impact #6 - Scientific & Societal: National RI/GBIF nodes will mediate and expand the reach of BioDT outputs to data publishers and scientific end-users in their respective national RI node networks. RI nodes provide essential hands-on support services to catalyse the uptake of investments in research infrastructure such as BioDT by respective national researchers.

Wider impacts: Cross domain interoperability opens the possibility for hybrid innovation based upon diverse data sets from multiple domains.

[...] ensure the quality, reliability, verifiability of the data, information and outputs of such Digital Twins [...] exploit [...] existing and new data made available through the Common European Data Spaces and the European Open Science Cloud.

Impact #7 - Scientific: BioDT will ensure a FAIR approach is applied to the information that the project will produce by combining FAIR principles and FDO layer, in alignment with the EOSC interoperability framework.

Impact #8 - Technological: Guiding users to understand the quality of data with BioDT that will provide reliability and verifiability of the data.

Impact #9 - Scientific: By bringing together different initiatives from EOSC, EuroHPC and Destination Earth, BioDT will ensure that the prototype is aligned with biodiversity policies and become available via key platforms, increasing Europe's multi-disciplinary capacity.

Impact #10 - Societal: Enabling the advancement of effective conservation for plant genetic diversity through use cases, BioDT will contribute to the conservation of crop wild relative gene pools that will be required to secure future food production under the challenges created by climate change. Addressing SDG 2 Zero Hunger, in the medium to longer time perspective.

Wider impacts: Improved biodiversity literacy and knowledge on biodiversity status, pressures, impacts and responses, in support of the lifelong learning envisioned in SDG 4 Quality Education.

[...] provide advanced modelling, simulation and prediction capabilities [...] through [...] use of advanced digital technologies such as high-performance computing, software, AI methods and big data analytics.

Impact #11 - Technological & Societal: Our development of ground-breaking RI technologies will enable new discoveries and keep Europe's RIs at the forefront of scientific excellence. This will pave the way to the development of innovative solutions to societal challenges and new industrial applications, products and services.

Impact #12 – Economical & Societal: Technology transfer will open new areas of R&D for industrial applications/products across biodiversity-related sectors such as agri-food, tourism and healthcare.

Wider impacts: Results from the advanced data modelling will help assess the effects of global climate change and implement appropriate policy decisions and management options.

[...] standard methods and protocols in the areas of (a) model and data fusion for optimal synergy between observations and models, including provisions to include information from the entire digital continuum[...] (b) visualisation and artificial intelligence-based knowledge generation from spatio-temporal information.

Impact #13 – Scientific: Spatiotemporal responses of biodiversity patterns and processes by combining high-resolution dynamic biodiversity data with large extent geographic biodiversity data available for modelling.

Impact #14 – Economic: Reduced cost and increased time from observation to data to model to scientific analysis to recommendation to decision to action.

Wider impacts: By improving tools to analyse climate change impact on ecosystems, BioDT will contribute to address the challenges of SDG 13 Climate action.

[...] development of quality measures and trust, development of standard quality mapping and indicators for appropriately communicating differences in qualities of inputs and outputs from digital twins [...]

Impact #15 – Scientific: By combining FAIR principles and semi-automated indicators (e.g. usage of persistent identifiers, metadata that is represented using a formal knowledge representation language), BioDT will ensure progressive improvement in the quality and usefulness of the datasets across the different projects with standards, interoperability and/or programmatic access.

Wider impacts: BioDT will ensure a closer collaboration between public authorities, researchers and citizens.

Table 1 – Specific contributions of BioDT to the expected outcomes and scope of the call

2.1.2 Overcoming barriers for impact

The main barriers for achieving the impacts identified within the BioDT project are detailed below.

Barriers, strategies and mitigation measures
Barrier: Difficulties to determine what data is most important to collect and analyse in the long run. Strategy for barrier overcoming: BioDT will allow predictions on spatio-temporal responses by combining high-resolution dynamic (i.e. changes in abundance over time) biodiversity data with large extent geographic biodiversity data available. BioDT will showcase how to combine massive data without unnecessary simplification, and disseminate this information through papers and conferences. This will lay groundwork for future EU policy that will revolutionise how modelling is used in global level decisions and actions, including IPCC and IPBES taking into account urgency, speed and scale needs. Mitigation measures: By using novel methods, HPC, AI and big data, BioDT will demonstrate the effectiveness and reliability in its approach to using data-intensive modelling for conducting research, and how it can be upscaled for analysing massive amounts of data.
Barrier: Ensuring technology, policy and markets are working at the same time. Strategy for barrier overcoming: Markets, Technology and Policy are three levers that must be synchronised and work in complementary ways. Policies that will be drawn from the BioDT policy papers will spark new technologies that shape the market systems. Simultaneously however, policies should be shaped by the technologies we develop. The consortium will tap into both the policies and the markets, through the identified case studies, and engaging with communities with real, demonstrable needs. Mitigation measures: Eight real world use cases will be used to test BioDT in terms of accuracy, precision, reliability and repeatability, playing a key role towards the mobilisation of relevant scientific communities and the involvement of the main stakeholders.
Barrier: Limitation of current simulation models. Strategy for barrier overcoming: To date global simulation models are limited in spatial resolution, representation of biodiversity and interaction dynamics as well as their explicit drivers. Agent-based models are often too simplistic to model the complexity in real ecosystems, while statistical models lack explicit causal descriptions. BioDT will develop a hybrid model combining theoretical and statistical approaches with a critical evaluation of the predictive performance of this new modelling framework, especially in comparison to the more traditionally used statistical approaches in biodiversity research. Mitigation measures: A task will be dedicated to model validation (T6.3). The model outputs will be scanned using approaches including AI methods and Approximate Bayesian Computation to identify regularities (patterns) which were not known or used for model development or parameterization, and to enable systematic cross-comparisons of the different modelling approaches across the 8 use cases.

Table 2 – Barriers and BioDT approach to them

2.1.3 Scale and significance of the project's contributions to the expected outcomes and impact

Impact: # of countries to which BioDT will be disseminated # (Scale): BioDT will ensure geographical coverage with at least 15 RI member countries with the idea of supporting an additional 5 countries on <i>widening the user base</i>. Reaching approximately 700 publishers Significance: This will cover all European countries connected to the biodiversity infrastructures Measuring Method: Number of events & webinars (with BioDT results presentation) organized by national RI nodes, WP7 & WP4 (and estimated number of people & data publishers attending).
Impact: # citations of BioDT use-cases # (Scale): 10 citations per use case Significance: Demonstrating increasing trend of research uptake of BioDT use case outputs Measuring Method: Assigning one DOI to each use case (e.g. through a data/software paper) and ask researchers to cite the BioDT datasets/resources.
Impact: Policy recommendations generated by use-case # (Scale): 1 Published policy paper per use-case group to be presented at either National or EU event Significance: Creating awareness of BioDT possibilities among policy agencies Measuring Method: DOI-citation in published papers. Number of events where the paper is presented.
Impact: # of new BioDT trainers inside the RI nodes # (Scale): 20 Significance: Top-Tier biodiversity researchers participate Measuring Method: Number of participants in workshops that will train further students.
Impact: # of students trained # (Scale): Approximately 500 Significance: Moving a new generation of biodiversity researchers into advanced informatics data analysis skills Measuring Method: Keeping track of authenticated /validated users registered through the BioDT website, LifeWatch ERIC and GBIF platforms.
Impact: BioDT outputs presented at conferences # (Scale): 40 conferences visited, 50 publications Significance: Creating awareness of BioDT among established researchers. Bringing BioDT technical development to the conference where the experts we need to involve are present. Measuring Method: Conferences attended by BioDT partners. DOI-citation in published papers
Impact: Engaged SMEs # (Scale): 10 companies to use the prototype with DNA Use Case Significance: Environmental survey SMEs can run impact studies on the twin Measuring Method: Number of companies using the twin. Reaching out to EITDigital¹⁴ & European Innovation Council¹⁵ to invite SMEs and startups to contribute to the use cases.
Impact: Engaged General Public

¹⁴ <https://www.eitdigital.eu/our-community/purpose/>

¹⁵ <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/european-innovation-council-eic-pilot>

(Scale): 1500 monthly visitors

Significance: Generating awareness among the general public

Measuring Method: Number of Internet users visiting BioDT website.

Table 3 – Scale and significance of BioDT contribution

2.2 Measures to maximise impact - Dissemination, exploitation, and communication

Our approach to developing **digital marketing and content strategies** is highly audience-centric, addressing the “What’s in it for me?” question in every message. A multichannel communication and dissemination strategy will be adopted to ensure maximum uptake and exploitation of the project’s results (see section 2.2.1, 2.2.2. and 2.2.4). Coupled with our stakeholder-tailored activities across strategic channels, the project aims to combine this with outreach to highly-relevant **multipliers** (see section 2.2.5).

Target groups addressed

	Biodiversity RIs, RI nodes, data providers and researchers (T1) Role: RIs, universities, research organisations; the end-users that will contribute to developing the DT, enhancing its use cases, and testing its functionalities. Challenges: Data collection and accessibility, lack of well-functioning digital infrastructures and technologies, lack of model sharing and common data model standards as well as sub-optimal data processing and analysis. RI nodes have very different capacities and resources and require coordination and training before new services can be made operative across the community. Benefits: RI nodes and researchers from different disciplines can use the results of BioDT to improve services. RI nodes will provide training for their networks, including BioDT results. RI nodes are an important link between BioDT outcomes and data providers holding the data for BioDT. Onboarding strategy: Researchers and RIs will be involved directly through training events as well as Bring Your Own Data (BYOD) and Bring Your Own Modules (BYOM) workshops.
	Policy makers (T2) Role: EU, Member States, Local governments, intergovernmental organisations (UNESCO, FAO, etc.) Challenges: Policy makers, governments and the public sector need appropriate tools to perform reliable, comprehensive and timely analyses and implement data-driven decisions to protect and restore biodiversity. Benefits: Make decisions based on quality data and modelling, responding better to societal needs and to key initiatives including the EU Biodiversity Strategy 2030 and the UN SDGs. Improved interconnection between core EU digital initiatives such as Destination Earth, EOSC, EuroHPC, and the EU Data Spaces. Onboarding strategy: Policy makers are the main target stakeholders for the BioDT Policy Papers, which will include medium and long-term recommendations for the employment of the DT.
	Industrial actors and SMEs (T3) Role: Develop and/or exploit advanced technologies for business solutions in sectors related to biodiversity, such as agri-food, tourism, healthcare.

Challenges: The lack of powerful and innovative simulation models is a main obstacle to the implementation of optimised strategies that bring greater business results with less negative impacts on biodiversity. The Agri-Food industry faces new challenges from climate change to maintain or increase food production. Solving food security challenges will require introducing novel genetic diversity (including from crop wild relatives) into crop breeding programs.
Benefits: Enhanced sustainability of European industry through co-development with industrial actors of advanced RI technologies and technology transfer; Opening up of new areas of research and development of new industrial applications/products.
Onboard strategy: Engage industrial actors through technology transfer to create new applications and the data spaces to explore new Research & Development solutions.

	Civil society and citizen scientists (T4) Role: Civil society will contribute to drive future data collections that will improve the predictive ability of the digital twin.
Challenges: Citizens and citizen scientists are willing to contribute to biodiversity-related research but are lacking the tools to understand the value of the data they collect and its impact on the actual research.	
Benefits: Boost citizen science, strengthen common understanding of biodiversity dynamics and prediction models, have an impact on policy making. Foster biodiversity literacy and trust in biodiversity research among the general public.	
Onboarding strategy: Citizens will be engaged through dedicated sessions at citizen science-related events, as well as through the EOSC network. Interactive tools will be shared with citizen science projects across Europe to highlight the need for new data, such as via Zooniverse.org¹⁶ and Weobserve¹⁷.	

2.2.1 Draft plan for dissemination, exploitation, and communication

BioDT will implement a multi-faceted and wide-ranging dissemination, exploitation and communication plan combining:

- **Dissemination of results** tailored to diverse stakeholder groups, spanning open-access publications and scientific conferences targeted at the biodiversity research community; policy papers targeting policy makers and local governments on key findings and best practices. Key results from use cases will also be disseminated thanks to the organisation of dedicated workshops and webinars.
- **Exploitation of results** will be carried out through: (a) the scientific community which needs to see the work made inside BioDT and adopt this way of work and thinking; (b) students in university curricula, summer schools and webinars in order to adopt the use of the BioDT prototype; (c) relevant stakeholders and policy makers who need reliable evidence for the management of the natural capital and its services.
- **Stakeholder engagement:** Extending partner networks across all primary groups through events and synergies established with related projects funded by the INFRA EOSC calls, DestinE, and the EU Data Infrastructure.

¹⁶ <https://www.zooniverse.org/> “People-powered research” zooniverse is the world’s largest and most popular platform for people-powered research. This research is made possible by volunteers — more than a million people around the world who come together to assist professional researchers

¹⁷ <https://www.weobserve.eu/about/citizen-observatories/> Citizen Observatories (COs) are community-based environmental monitoring and information systems, that invite individuals to share observations, typically via mobile phone or the web.

- **Communication strategy:** Ensuring strong visual identity, brand reputation. Tailored and content-rich communications and marketing campaigns, website design and evolution, social media daily activities and pay-per-click campaigns.

A sample of the channels that will be used in the dissemination and communication plan is reported below.

Type & KPI Per Year	Target group	KPI & Month
Tailored promotional materials for different target groups (Flyers, posters, infographics, etc.)	T1, T2, T3, T4	6 (M1, M2, M4, M8, M21, M24)
Podcasts (<i>dedicated interviews with end-users</i>)	T1, T4	4 (M7, M27, M29; M35)
Scientific publications	T1	50 (M18, M24, M32, M34, M36)
Website releases	T1, T2, T3, T4	3 (M2, M16, M26)
Number of authenticated /validated users achieved through the outreach events, (webinars, workshops, training, summer school, final event)	T1, T2, T3, T4	Over 500 by M36
Newsletters (<i>newsletter subscriptions +350</i>)	T1, T4	12 (M4, M7, M10, M13, M16, M19, M22, M25, M28, M31, M34, M36)
Press releases (<i>impressions on traditional media 3</i>)	T2, T4	4 (M1, M11, M20, M36)
Webinars	T1, T3	10 with average attendance of 40 participants (M2, M4, M11, M13, M18, M20, M23, M25, M28, M33)
Videos (<i>visualisations 2000 sponsored video, 500/organic video</i>)	T2, T4	4 Video documentaries (M3, M5, M14, M33) - 10 Video interviews (M4, M5, M6, M7, M8, M9, M13, M21, M24, M32)
Social media channels & campaigns	T1, T2, T3, T4	1.500 combined social media followers (M36).

Table 4 – Communication channels and KPIs

2.2.2 Dissemination of results

BioDT dissemination strategy aims at three main objectives:

- Transferring BioDT knowledge and results to stakeholders that can best make use of them.
- Raising awareness and openly demonstrating clear economic, social, and technological benefits of employing BioDT solutions in the biodiversity community.
- Maximising the impact of research and enabling the value of results to be wider than BioDT creators and end users.

BioDT will disseminate its results through different channels listed below:

Type	Target group	KPI & Month
Use case workshops + training sessions	T1	8 with average attendance of 60 participants (M6, M12, M12, M14, M16, M17, M22, M24)
R&I nodes coordination workshops	T1	2 events (M12, M24)
Summer school	T1	1 event (M34)
Third-party events	T1, T3	> 40 Presence events (M36)
Bring Your Own Data event	T1, T3	2 events (M18, M28)

Policy Papers	T1, T2, T3	4 policy papers (M24)
Final event with Destination Earth	T1, T2, T3, T4	1 joint Destination Earth event (M35)

Table 5 – Dissemination channels and KPIs

2.2.3 Project exploitation

The BioDT exploitation plan is based on three complementary and strategic paths to exploitation:

1. A joint exploitation path

Strategic goal: Sustain core, interdependent assets beyond the lifecycle of the project to enable widespread adoption by the biodiversity community covering also additional other domains such as Life Science, Marine ecology and Earth Sciences.

Measures to be taken: Define and implement a joint exploitation plan, dealing with all relevant IPR aspects, operational and maintenance costs.

2. Detailed exploitation plans defined by each individual partner, in line with each partner's business and research strategy.

Strategic goal: Ensuring BioDT becomes an integral part of the business portfolios/institutional services for guaranteed availability for uptake beyond the project lifecycle.

Measures to be taken: Innovation and IPR management; implementation of a go-to-market strategy. A commercialisation strategy will be defined and implemented in both cases.

3. Exploitation of the pilots

Strategic goal: Using the Use Cases as a lever to bring onboard other RIs and increase the use of the BioDT prototype.

Measures to be taken: Large-scale marketing campaigns, video pills, workshops and training sessions.

2.2.3.2 Joint exploitation plan

Joint exploitation plans mostly revolve around sustaining the integrated BioDT platform, its components and major outputs as central to the future sustainability strategy.

BioDT KER #1	Asset: Digital Twin Platform - TRL Initial 4, Final 7
Joint Exploitation Measures - Adoption guides, training sessions, presentations targeting scientific community to demonstrate the work produced in BioDT and share this way of work and thinking. Summer schools and webinars in order to increase the usage of BioDT prototype. Joint Dissemination Measures - Scientific publications, web-based media (e.g. podcasts, newsletters), media coverage (newspapers, radio, TV, and video documentaries).	
BioDT KER #2	Asset: Data modelling & simulation frameworks - TRL Initial 4, Final 8
Joint Exploitation Measures - Workshops to involve RI nodes in the development of new data types and domain models required for advancing the modelling solutions for the use cases. Exchange of capacity between nodes to enable the RI nodes across the network to work efficiently together with their data publishers to establish new and enhanced data streams. Joint Dissemination Measures - Dedicated marketing packages on concrete benefits. Workshops for RI nodes will mostly be organized back-to-back with the annual RI nodes meetings and RI nodes seminar series and webinars.	
BioDT KER #3	Asset: Use Cases - TRL Initial 3, Final 7
Joint Exploitation Measures - Dedicated workshops to develop, present, and pilot the Use Case outputs. Training and education sessions will be organised based on the Use Cases results in order to pave the way for BioDT service uptake	

Joint Dissemination Measures - Scientific publications, tutorials, animated demonstrations, training software and courses to inform and involve the scientific community, industry and policy makers.	
BioDT KER #4	Asset: Integration APIs - TRL Initial 4, Final 7
Joint Exploitation Measures - Promotion and distribution of APIs to pave the way for future sustainability of BioDT through its integration with EU Data Spaces, EOSC and the other DTs under Destination Earth.	
Joint Dissemination Measures - Dedicated training events will be organised in order to increase users' competencies required to effectively run their data projects on the DT. Training will involve researchers, coders, data providers and RI nodes.	
BioDT KER #5	Asset: FAIR data design - TRL Initial 6, Final 8
Joint Exploitation Measures - All data will be available via a trusted, FAIR data repository, accessible via open protocols. The methods, design and process of the use cases will be published as open access publications. Training materials (available in open access format) will be produced for RIs, and the wider research community detailing how to use the BioDT prototype.	
Joint Dissemination Measures - Existing and established networks of natural science collections and biodiversity communities (such as CETAF, TDWG, SPNHC) will be used to exploit the value of BioDT use cases. Networks include mailing lists, periodic conferences, social media and other interactive channels (such as Slack).	
BioDT KER #6	Asset: Policy Papers
Joint Exploitation Measures - Research and development activities based on the outcomes of the use case will result in key recommendations and best practices that will be communicated to policy makers during specific meetings and events.	
Joint Dissemination Measures - Dedicated marketing package and distribution both in printed and digital copy. Dedicated webinars and press releases.	
BioDT KER #7	Asset: Training - TRL Initial 6, Final 9
Joint Exploitation Measures - A Summer School mainly addressing RIs and their research and ICT staff is planned by M34 with the aim of empowering practitioners in the use of BioDT as well as work on its scalability.	
Joint Dissemination Measures - Training will be delivered online through interactive and recorded lessons, available on the LifeWatch ERIC. RI nodes will receive capacity to multiply the dissemination of training materials (train-the-trainers model) and provide BioDT courses and direct end-user support to researchers in the respective country of the RI node.	

Table 6 - Joint Exploitation Plan

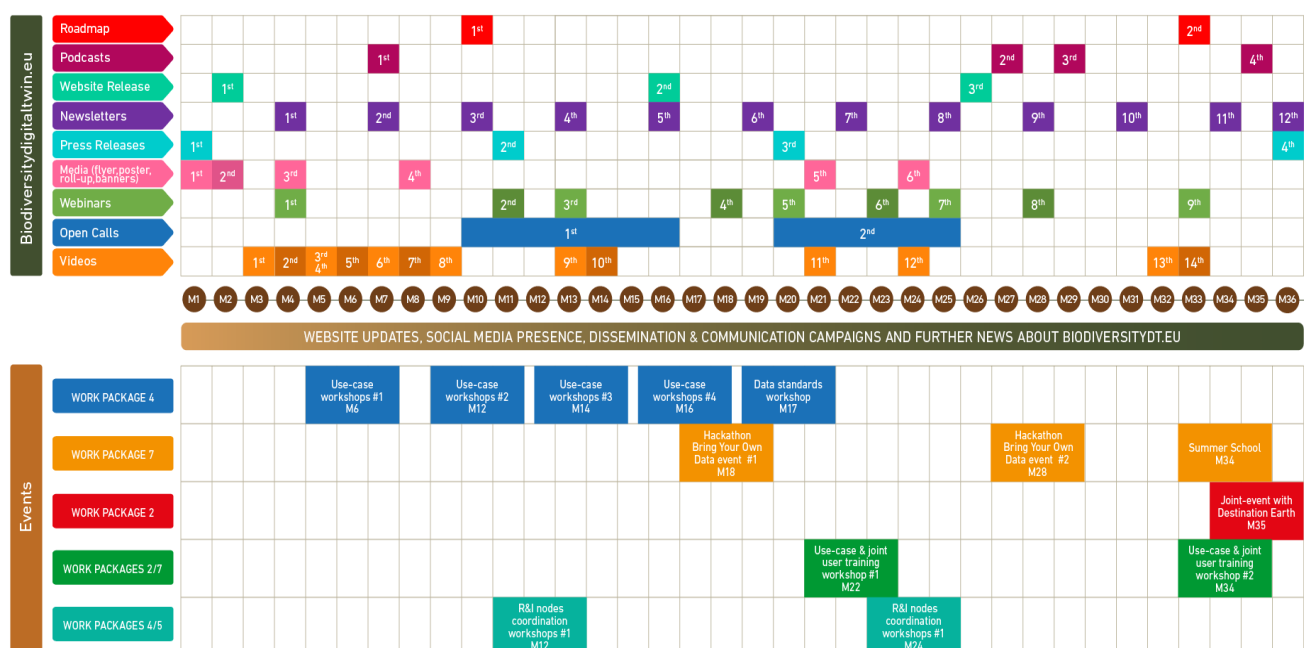
2.2.4 Communication activities, channels and means

The 36-month communication strategy will be pragmatically organised in a regular timeline with a breakdown of specific activities, channels and monitoring elements to allow the project to gauge the effectiveness of specific campaigns and make necessary adjustments where necessary.

The communication strategy and its channels described above will also avail of two novel features **Involvement of natural history museums, botanical gardens** for events, meetings & exhibitions.

Documentaries - If budget permits we would draw upon the experience of LifeWatch ERIC documentary series to support the key impact pathways and showcase results by producing a documentary series telling the success story of the project through the voices of the different stakeholders and beneficiaries involved. The LifeWatch ERIC Multimedia Production Centre will create specific videos, with a storytelling approach to communicate the project journey. This will be showcased at public events, sent to appropriate festivals and distribution platforms, and disseminated through the LifeWatch ERIC Web TV, which will be launched in spring 2022.

A preliminary timeline with the main dissemination and communication activities is reported below:



2.2.5 Multiplier networks

BioDT will spread its results across partner networks as well as stakeholders and multipliers:

Associations linked to Biodiversity RIs, data providers and researchers: @NFDI4BioDiv, @BExplo_research, @CopernicusLand, @eurotaxonomy, @evoltreeNW, @porbiota, @CIBIO_InBIO, @CECT_UV, @RJBOTANICO, @BccmCollections, @INBOVlaanderen, @RBINSmuseum, @AgroBRC_RARe, @INRAE, @latjitieto, @EoscLife, @ENVRIcomm, @ELIXIREurope

Associations relevant for Policy makers: @UNBiodiversity, @unepwcmc, @Biodiversity_be, @OFBiodiversite, @OECD_ENV, @IPBES, @GEOSEC2025, @EUClimateAction, @EUgreenresearch, @LIFEprogramme, @BiodivERsA, OISPG - Open Innovation Strategy and Policy Group.

Industrial actors and SMEs: @EUROPARC, @OrganicsEurope, European Green Digital Coalition, @WeValueNature, @EEN_EU, @CapsCoalition, @epe_asso, Biodiversity in Good Company, @valuebalancing, @ProBiodiversity, @biconsortium

Associations relevant for Civil society and citizen scientists: @Green_Europe, @SDG2030, @WWF, @FAO, @FSC_IC, @ReinforceEU, @WeObserveEU, @LandSense, @growobservatory, @CSEOLab, @SCENT_EU

Multipliers linked to relevant initiatives such as EOSC, digital twins, EuroHPC, etc.: @EoscPortal, @EuroHPC_JU, @eoscassociation, @EOSCFuture, @BioExcelCOE, @esiwace, @EUScienceInnov, @DataEcoEU, @HorizonEU, @ScienceEurope, @PRACE_RI, @EOSC_Nordic, @EoscPillar, GAIA-X, UK Digital Twin Programme

News and media portals related to research and biodiversity: @ResearchEurope, @scibus, @CORDIS_EU, @agrochemie, @biobasedpress, @guardianeco, @BBCEarth, @EUEnvironment, @euronewsgreen

Table 7 – Multiplier networks and press & media channels utilised

2.3 Summary

The key elements of the impact section are summarised in the table below, which also includes some practical examples.

<i>SPECIFIC NEEDS</i>	<i>EXPECTED RESULTS</i>	<i>D + E + C MEASURES</i>
<i>(1) Need for technological solutions to ease data collection and accessibility</i> <i>(2) Increase of model sharing and common data model standards</i> <i>(3) Improve cross-domain interoperability between RIs</i> <i>(4) Improve interconnection between EU digital policy initiatives</i> <i>(5) Lack of powerful and innovative simulation models in biodiversity</i> <i>(6) Ensure quality measures, indicators and standards to evaluate datasets</i> <i>(7) Promote Open Science & FAIR and interoperable data</i>	<i>BioDT aims to push the current boundaries of predictive understanding of biodiversity dynamics by providing:</i> <i>(1) A flexible software platform and computing services that can be easily integrated into diverse user communities.</i> <i>(2) Data modelling, prediction technologies and simulation frameworks to support data and software interoperability, integration and collaboration across different scientific domains</i> <i>(3) 8 use cases of global importance, to showcase BioDT usage and performance as well as test its predictive performance through scenario simulations, predictions and biomonitoring methods.</i> <i>(4) A set of APIs to facilitate data sharing and the use of existing BioDT software infrastructure to a wide variety of research communities.</i> <i>(5) Consolidate the RIs FAIR data designs through a layer of FAIR Digital Objects to improve findability and accessibility with the use of machine actionable metadata.</i> <i>(6) A lesson learned and a series of RI skills mapping & good practices policy papers will be developed to guide efforts aimed at addressing the EU Biodiversity Strategy priorities.</i> <i>(7) Training programme aiming to increase users' competencies required to effectively run their data projects on the Digital Twin during the project's lifetime and after the project ends.</i>	<i>Dissemination Measures:</i> <i>Scientific conferences, open-access publications.</i> <i>Policy papers and tailored content for policy makers and citizen scientists.</i> <i>Workshops, webinars, podcasts and documentaries.</i> <i>Exploitation Measures:</i> <i>Training events and exchange of capacity sessions with RIs, data providers, coders and researchers, Summer schools and Bring Your Own Data events.</i> <i>Integration with EU Data Spaces, EOSC and the other DTs under Destination Earth.</i> <i>Mechanisms for innovation management and IPR.</i> <i>Communication and Stakeholder Engagement</i> <i>Website set-up (M3) with SEO-based and UX design, BioDT branding and visual identity. Social media set-up (M1), e.g. YouTube, Twitter and LinkedIn.</i> <i>Communication campaigns using diverse formats, e.g. press releases, podcasts, video pills.</i> <i>Targeting and profiling key target groups recruited at BioDT and 3rd-party events, as well as social and professional networks.</i> <i>SMART-based campaigns with clear calls to action and/or with best practice highlights.</i>

TARGET GROUPS	OUTCOMES	IMPACTS
Expected to Uptake (1) Biodiversity RIs, RI nodes, data providers and researchers Expected to Use (2) Industrial actors and SMEs in biodiversity-related sectors such as agrifood, healthcare, tourism Expected to benefit (3) Policy makers (EU, Member States, local governments) (4) Civil society and citizen scientists	(1) RI nodes and researchers from different disciplines can use the results of BioDT to improve their services. RI nodes will provide training for their networks, nd communicate BioDT results. This is a key link between BioDT outcomes and data providers holding the data for BioDT. (2) Enhanced sustainability of EU <i>industry</i> through co-development with industrial actors of advanced RI technologies and technology transfer. Opening up of new areas of research and development of new industrial applications/products. (3) Policy makers will be able to make decisions based on quality data and modelling, responding better to societal and key initiatives including the EU Biodiversity Strategy 2030 and the UN SDGs. Improved interconnection between fundamental EU digital policy initiatives such as Destination Earth, EOSC, EuroHPC, and the EU Data Spaces. (4) Boost <i>citizen science</i>, strengthen common understanding of biodiversity dynamics and prediction models, have an impact on policy making. Foster biodiversity literacy and trust in biodiversity research results among the general public.	Scientific impacts: High-resolution maps of land use for research on biodiversity in agricultural landscapes. Test and evaluate the predictive performance of the developed set of modelling approaches through scenario simulations and biomonitoring methods. Access to enhanced and integrated data as well as modelling tools to provide new innovative services. An online biodiversity community where users can interact and express their needs. Hands-on support services for the uptake of BioDT by national researchers. FAIR principles applied in alignment with EOSC. Coordination and interoperability with key initiatives such as EuroHPC and DestinE. Spatiotemporal responses of biodiversity patterns and processes by combining dynamic biodiversity data with geographic biodiversity data. Improve the quality of datasets with standards, interoperability and/or programmatic access. Technological impacts: Prototype to become a model for other DTs. Enhance reliability and verifiability of the data for users. Societal impacts: Enable new discoveries paving the way to the development of innovative solutions to societal challenges. Efficient crops, increasing richness and productivity of arable lands, contributing to a sustainable agriculture. Economical impacts: Open new areas for R&D applications in biodiversity-related sectors such as agrifood and healthcare. Reduced cost and increased time from observation to data to model to scientific analysis to recommendation to decision to action.

3. Quality and efficiency of the implementation

3.1 Work plan and resources

In order to achieve the continual interactions between different parts of the project, we have divided the main project tasks into eight work packages. **Project Management (WP1)** will cover the administrative and financial management of the project. WP1 will be also responsible for the management of quality and risk aspects, as well as for liaising with the EC. **Communication, Dissemination, Sustainability and Impact (WP2)** will promote the project's activities and publish regular information about the resources made available by the project and how to access them. **Digital Twin Advanced Technical Platform (WP3)** will set up, develop and operate the digital twin platform and ensure the availability of the required IT resources for the users. **Data and use cases (WP4)** will develop and package the use cases and content streams so that they provide the necessary information for both developing, testing, and implementing the models. **Improving quality of data, workflows and models through FAIR principle (WP5)** will consolidate the FAIR data designs of the involved RIs through a layer of FAIR Digital Objects (FDO) and ensure interoperability of data. **Simulation, modelling and data analytics (WP6)** will use the case studies to develop the Digital Twin, test their predictive performance, and apply the BioDT to address biodiversity challenges through scenario simulations, predictions and biomonitoring methods. **Integration with research infrastructure environments (WP7)** will provide the technical expertise and enabling support required by the research communities and users to run their data projects on the Digital Twin. Finally, **Collaboration with other initiatives and programmes (WP8)** will ensure the effective connection between the Biodiversity Digital Twin and other related initiatives and programmes, in particular Destination Earth (DestinE) and the European Data Infrastructure.

Table 3.1a: List of work packages

WP	Work Package Title	Lead	Short Name	PMs	Start	End
1	Project Management	1	CSC	72	1	36
2	Communication, Dissemination, Sustainability and Impact	2	TRUST-IT	106	1	36
3	Digital Twin Advanced Technical Platform	1	CSC	153	1	36
4	Data and Use Cases	16	GBIFS	301	1	36
5	Improving Quality of Data, Workflows and Models through FAIR Principles	18	Naturalis	99	1	36
6	Simulation, Modelling and Data Analytics	6	JYU	300	1	36
7	Integration & Service Uptake with Research Infrastructure Environments	4	IT4I@VSB	148	1	36
8	Collaboration and Integration with Strategic Initiatives / Programmes	5	TNO	86	1	36
			Total PMs	1265		

M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15 M16 M17 M18 M19 M20 M21 M22 M23 M24 M25 M26 M27 M28 M29 M30 M31 M32 M33 M34 M35 M36



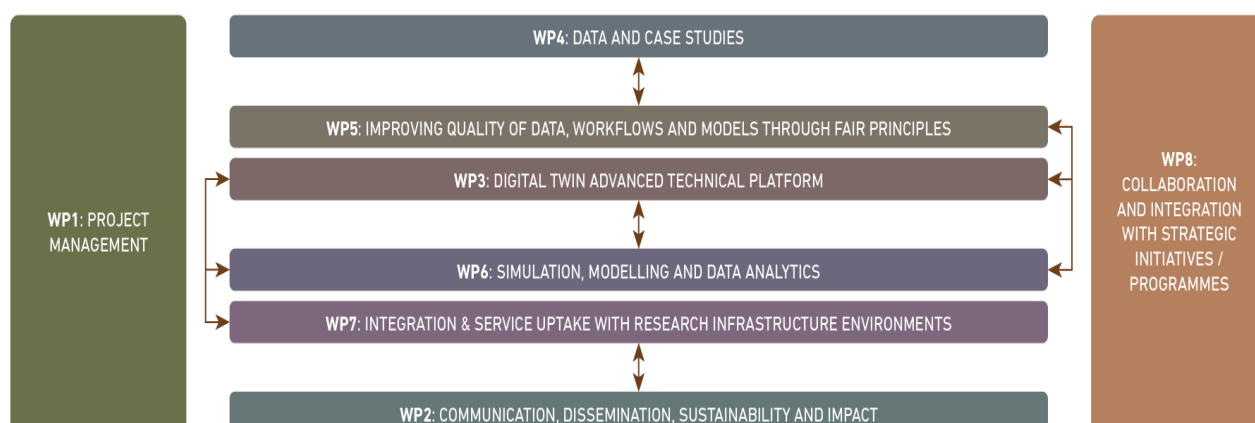


Table 3.1b: Work package description

WP1	Lead Beneficiary: CSC						
Title	Project Management						
Participant number	1	2	4	5	6	16	18
Short name of participant	CSC	TRUST-IT	IT4I@ VSB	TNO	JYU	GBIFS	Naturalis
Person months per participant	36	6	6	6	6	6	6
Start month	1		End Month		36		
Objectives: Organise the overall project administration, finance and project management including the coordination of the quality and risk management; establish effective framework and procedures for efficient communication, collaboration and fast and accurate decision-making and project orchestration; set up, coordinate and ensure efficient interaction with the governing bodies (council, project management board, and advisory board).							
Task 1.1 - Project management and financial operations (M1-M36). <u>CSC</u> , TRUST-IT, IT4I@VSB, TNO, GBIFS, Naturalis, JYU. The day-to-day operations will be performed by the Project Management Office (PMO) led by a designated experienced Project Manager (PM) and the Project Management Board (PMB) including all WP leaders. The PD will ensure that work assigned to the different work package teams progresses according to schedule and within the allocated budget and resource plan. Project internal meetings will be organised within this task.							
Task 1.2 - Quality control, ethics, risk management and DMP (M1-M36). <u>CSC</u> . The PMO will ensure that project results, milestones, and formal deliverables undergo a quality control process prior to submission to the EC. Potential risks will be analysed and mitigation will be formulated. A Data Management Plan will be developed for making the project’s data/research outputs findable, accessible, interoperable and reusable (FAIR).							
Task 1.3 - Governance support (M1-M36). <u>CSC</u> . The PMO will provide any relevant information required by the governance bodies, as well as physical and/or virtual facilities and for documenting their discussions.							
D1.1 Quality Plan, Risk Register, Ethics & Compliance (M3). This will document (1) the quality control processes; (2) risks and their corresponding mitigation plans; and (3) ethics & compliance issues.							

D1.2 Data Management Plan (M6). The DMP will evolve during the lifetime of the project in order to present the status of the project's reflections on data management.

D1.3 First, second and third Periodic Report. (M12, 24, 36)

WP2		Lead Beneficiary: TRUST-IT									
Title	Communication, Dissemination, Sustainability and Impact										
Participant number	1	2	3	4	5	6	7	8	9	10	11
Short name of participant	CS C	TRUST-IT	Com mpla	IT4I @VS B	TN O	JYU	UT	KTH	UniSw ansea	UNI MAN	ECM WF
Person months per participant	12	23	11	4.5	2.5	0.5	0.5	1.5	0.5	0.5	6
Participant number	12	13	14	15	16	17	18	19	20	21	22
Short name of participant	UF Z	UKCEH	EAA	UH	GBI FS	UiO	Natu ralis	Sencken berg	LW ERIC	Bayer	MLU
Person months per participant	8	3.5	3.5	5.5	6	3.5	6	0.5	6	0.5	0.5
Start month	1					End Month	36				

Objectives: This WP develops, updates and implements the communication, dissemination & exploitation strategy for the Project, carrying out the related activities, including all digital, web and social media presence. It will organise and roll-out workshops and events with all key stakeholders of the project for widening the user base, and create impact and project results' sustainability beyond the consortium boundaries and promote innovation.

Task 2.1 - Outreach, Communication, dissemination & Digital Presence (M1-36): TRUST-IT, all partners.

This task comprises the development of the Plan for dissemination & exploitation including communication activities and execution of the resulting activities, targeting all the relevant target groups. Branding and positioning of BioDT will be covered. Moreover, the task will cover the effective design, implementation and continuous update of the project website as well as the production of communication materials, such as brochures, press releases, videos, podcasts etc. The results from the identified use cases (such as prototypes, software, publications, data) will be disseminated under this task. Regular social media activities will be carried out. The task will report on the BioDT impact as per the set of KPIs indicated in Section 2 in close collaboration with the targeted efforts of user uptake carried out in WP7 and with use cases in WP4.

Task 2.2 - Events (Months: M1-36) Lead: TRUST-IT, all partners; The task will be responsible for the organisation of the BioDT events both in physical and virtual format. In collaboration with other WPs, this task will ensure a proper engagement with all relevant target audiences thanks to dedicated event formats spanning, webinars, use cases workshops, training sessions and hackathons, mixing informative and hands-on, interactive sessions. The task will help BioDT reach a wider user base and will contribute to paving the way for future exploitation and sustainability.

Task 2.3 - Biodiversitydigitaltwin.eu for widening the user uptake (M1-36). GBIFS, Trust-IT, CSC, IT4I@VSB, ECMWF, LW ERIC, Naturalis, UiO, UH, EAA, UKCEH, UFZ. This task is carried out in coordination with WP7 with the involvement of new user communities during the second half of the project. The activities will encompass: a) synthesis of the value proposition(s) brought forward by the Project; b) scouting of potential early adopters and nurturing engagement; c) onboarding of users; d) development of synergies, linked

with WP8. The BioDT value-proposition will be leveraged to the various target users and communicated during the scouting & nurturing user communities. To aid in the organisation of communications activities, a dedicated online engagement space will be made available to foster and engage a community of scientists and users.

Task 2.4 - Sustainability (M12-36): CSC, TRUST-IT, IT4I@VSB, KTH, ECMWF, UKCEH, UH, GBIFS, Naturalis, LW ERIC. A sustainability model for the BioDT will be developed in the project's second year, drawing from requirements from the research infrastructures and lessons learned from the introduction of the BioDT services into their environments and as part of the LUMI service portfolio. The model will also draw on the existing arrangements at European level, in particular related to the EuroHPC and Destination Earth, and from relevant activities across the project (notably tasks WP7 and WP8). It will assemble the business case for the future BioDT covering the following elements: costs, funding sources, and business models.

D2.1 Plan for dissemination & exploitation including communication activities (M6): This report presents the plan for targeted engagement and all outreach, exploitation & communication strategies.

D2.2 Final plan dissemination & exploitation including communication activities (M35): Outlines stakeholder groups and targeted engagement activities & achievements.

D2.3 Services Uptake report (M36): The deliverable reports experiences, results and lessons learnt in the process of involving and engaging new users of BioDT prototype.

D2.4 Sustainability report (M36): The report will include the sustainability model identified for the BioDT project

WP3			Lead Beneficiary: CSC			
Title	Digital Twin Advanced Technical Platform					
Participant number	1	4	5	7	8	
Short name of participant	CSC	IT4I@VSB	TNO	UT	KTH	
Person months per participant	48	24	12	48	21	
Start month	1		End Month		36	
Objectives: Set up, develop and operate the advanced technical platform required by the digital twins; provide world-class HPC resources from the LUMI EuroHPC computing facility; ensure portability of digital twins across HPC sites and cloud environments; maintain service catalogue of shared services for integration with EOSC Core services.						
Task 3.1: Platform architecture and implementation (M1-36): Partners: <u>CSC</u>, IT4I@VSB, TNO, UT, KTH. This task will create an extendable and secure architecture for the Digital Twin platform, integrate selected core components and shared services for digital twins based on requirements collected in T7.1 and the FDO layer provided through T5.1. Defined architecture will be used for selecting services and components for implementing the platform core. The architecture will be aligned with EOSC and other relevant European infrastructures. The task will also be responsible for maintaining a service catalogue of shared services for the Biodiversity community which will be integrated with EOSC Core services. Security topics will be considered throughout the project and internal audits will be used to verify the critical components. As a result, the platform will provide API-first services to the integrators and developers of the Biodiversity RIs. These services will be then integrated by WP7 into applications of Biodiversity RIs.						
Task 3.2: Integration with HPC sites (M1-36): <u>CSC</u>, IT4I@VSB, UT: This task will ensure the provision of the resources from the LUMI EuroHPC system, as required by the Digital Twins developed by WP6, including large scale GPU computing and data processing resources. The resources will be allocated from the Finnish allocation on the LUMI system. Storage resources will be provisioned and optimised to match the computing						

needs. Services developed in the Puhuri project (<https://neic.no/puhuri/>) will be used to make the environment available as early as possible and platform agnostic.

Task 3.3: Operations of the platform (M1-36): UT, CSC, IT4I@VSB, KTH: This task will take care of operating and maintaining the Digital Twin technical platform. It will follow the information security principles set in Task 3.1 and ensure the availability of the platform. Communication channels and technical issue tracking will be established for the application providers in WP7. This task will liaison with LUMI operations teams to maintain situational awareness with regards to the computing and storage resources.

Task 3.4: Integration of platform services (M1-36): UT, CSC, IT4I@VSB, TNO, KTH: This task will integrate the shared services selected by Task 3.1 into the Digital Twin platform. All integrated services will become available through a common API for integration with applications by WP7. Service integration will follow specifications from T5.1 and feedback from WP7.

D3.1 Architecture and implementation plan (M9). Describes the architecture for the Digital Twin platform, core components and shared services. Documents information security practices and more detailed implementation plan.

D3.2 Software for Digital Twin Advanced Technical Platform (M24). Components and integrations that form the technical platform are released with an open source license to make it widely available, support sustainability of the platform and adhere with the principles of open science.

D3.3 Productisation roadmap (M34). Roadmap describes the steps needed for productisation and large-scale adaptation of the Digital Twin Advanced Technical Platform. The roadmap will be aligned with the developments of the European Open Science Cloud, as well as the activities under the EuroHPC joint initiative.

WP4		Lead Beneficiary: GBIFS					
Title	Data and Use Cases						
Participant number	1	4	5	6	12	13	14
Short name of participant	CSC	IT4I@VSB	TNO	JYU	UFZ	UKCEH	EAA
Person months per participant	6	6	6	6	42	26	12
Participant number	15	16	17	18	20	21	22
Short name of participant	UH	GBIFS	UiO	Naturalis	LW ERIC	Bayer	MLU
Person months per participant	6	96	58	18	15	2	2
Start month	1	End Month				36	
Objectives: This work package will: 1) ensure the execution of the biodiversity use-cases by providing the necessary information to develop, test, and implement the prototype digital twins in research and at the science/policy interface, 2) make available and ready for use the dataset(s) required for the implementation of the use cases, 3) provide best practices, expertise, and lessons learnt, and transfer them to the researchers and research infrastructures.							
Task 4.1: Use cases in research at the science-policy interface (M1-36): <u>GBIFS</u>, UiO, UFZ/ UKCEH, JYU, MLU, Naturalis, LW ERIC, Bayer.							
<i>Task 4.1.1: Use cases on species response to environmental change: UFZ, GBIFS, UKCEH, JYU, MLU.</i> These use cases focus on the interactions and interrelations between species and ecosystems. A leap will be made by incorporating temporal dynamics rather than pure space-for-time substitutions. This will improve temporal							

predictions and accuracy. Additionally, different sources of uncertainty will be quantified. The work uses the large geographic extent (GBIF) combined with a smaller set of high-resolution time-series data (eLTER) in a single modelling framework.

Task 4.1.2: Use cases on genetically detected biodiversity: GBIFS, UKCEH, UiO. Food security, managed and cryptic challenging environments, and many other research frontiers will be approached by adding genomic methods, based on DNA data to those of phenomics, which are based on traditional biodiversity data (diversity and composition). Use case 3 addresses crop wild relatives and other genetic resources for farming and food security and use case 4 explores DNA detected biodiversity and poorly known habitats including soil.

Task 4.1.3: Use cases on the dynamics of species of policy concern. UFZ, GBIFS, Naturalis, UKCEH, LW ERIC. Modelling and HPC will be applied to use case 5 on invasive and alien species recognized at the EU and national levels, and to use case 6 on endangered species (IUCN data and national Red Lists). Data models and primary data streams based on the current species occurrence data, as well as metabarcoding data, are structurally sufficient, while further work will be required on the complementary data models for taxon checklists with conservation or invasiveness statuses. In addition, there are crucial problems related to the detection of the responsible environmental conditions as well as their synergistic (melt-down) effects on taxa and ecosystems, which will be tackled with the novel modelling techniques (WP6).

Task 4.1.4: Use cases on the influence of species interactions on planetary well-being. GBIFS, UFZ, LW ERIC, UiO. These use cases will focus on predicting disease outbreaks using disease vectors (use case 7) and explore patterns and processes of insect pollinators of plants, both wild and cultivated (use case 8). Further development of both the sampling event domain model (part of Darwin Core data standard) and data exchange models are critical for these use cases. Pipelines for extracting interactions data (digitisation on demand) off collection specimens will establish temporal historic reference points.

Task 4.2: Data streams (M1-36). GBIFS, UKCEH, CSC. This task will gather, translate and combine the existing European and global biodiversity data necessary for carrying out the modelling applications across the above use cases. It will re-use and combine data products and services from various sources including GBIF.org (EOSC service). Datasets needed for the models and applications will require a new combination of existing biodiversity and abiotic data, increased data quality and information confidence and processes (in collaboration with WP5) to ensure the right fit for purpose in terms of data formats or data mobilisation. To ensure transparency, credit and attribution, existing data citation mechanisms will migrate from system-specific solutions to the community norm in ecology (in collaboration with Task 4.3). For composite/aggregated datasets (input data streams to DT models) including source data without dataset DOIs, data papers will provide additional data citation solutions. Opportunities for linking the ecology community further closer to the RI nodes community (and developing shared norms) will emerge as a result of joint work on data streams (with Task 4.3). Collaboration between WP4, WP5, and WP6 will ensure the integration of data-producing sectors with data use, at the source of data production.

Tasks 4.3 Biodiversity RI nodes network (M1-36). UiO, GBIFS, TNO, EAA, Naturalis, LW ERIC, UH. The following sub-tasks are targeted to the interaction of this WP with the RI nodes to create awareness and adoption of the prototype digital twin developed through the use cases, the datasets used, and workflows for the further development of the digital twins. This will capacitate RI nodes to work together with their data publishers to transform data and create data streams.

Task 4.3.1: Capacity and awareness building among RI nodes and European data publishers to deliver the new data streams. UiO, GBIFS, UH, Naturalis, LW ERIC. Progressing from early prototype use cases towards the implementation of digital twins will require capacity building, outreach, and training for the RI nodes. Training events, workshops will enable and inform the change of data management culture to deliver biodiversity data streams while simultaneously progressing towards more advanced data formats. This task is essential to develop

capabilities and to enable research communities supported by RI nodes to deliver dynamic data and information originating from different sources as well as stay up to date with evolving analytical and modelling methods.

Task 4.3.2: Biodiversity informatics expertise. UiO, GBIFS, UH, Naturalis, LW ERIC, TNO, CSC. This task will create a high-profile expert network to ensure the interoperability of data and software across the research infrastructures involved. This will lead to a robust framework of data flow, e-services, and modelling solutions. This interactions task will ensure the harmonization of know-how across WP4 with WP5, WP6, and WP7. The respective international expert communities will be approached through their society conferences and respective task groups to harmonize and extend data standards to BioDT requirements. In particular, further work with Biodiversity Information Standards (TDWG) on the data standards for publishing sampling event data is needed.

D4.1 Report on the availability of consolidated, improved, fit for use data streams, data model availability, improvement, and application (M24). A variety of heterogeneous streams of biodiversity data from a range of sources and scales, as well as accompanying abiotic data is secured and developed, through interactions with RIs and expert communities, to ensure the fuelling the BioDT functions with FAIR attribution and data citations.

D4.2 Report of results on applying models, data, and data streams covering the use cases (M32). BioDT modelling outcomes are developed, tested, prototyped and applied, with workshops / training events, across the eight use cases covering species response to environmental change, genetically detected biodiversity, threatened and invasive species, as well as species interactions and planetary well-being.

WP5				Lead Beneficiary: Naturalis					
Title	Improving Quality of Data, Workflows and Models through FAIR Principles								
Participant number	5	10	13	14	16	17	18	19	20
Short name of participant	TNO	UNI MAN	UKC EH	EAA	GBIFS	UiO	Naturalis	Senckenber g	LW ERIC
Person months per participant	4	12	9	8	8	12	16	15	15
Start month	1			End Month				36	
Objectives: The aim of WP5 is to facilitate the interoperability of the Biodiversity Digital Twin, in particular the modelling processes by adding quality indicators, improving the FAIRness and the reusability of the datasets, services and models. In this work package, we will: consolidate the FAIR data designs of the involved Research Infrastructures (RIs) through a layer of FAIR Digital Objects (FDO); pilot semantic mapping to improve interoperability and reusability of semantic artifacts; apply FAIR Digital Objects to the services; and research outputs for improved reusability and develop a framework to indicate data quality as part of the data product.									
Task 5.1: Consolidate RI FAIR Data designs into a FAIR Digital Object (FDO) layer (M1-M36). <u>Naturalis, GBIFS, LW ERIC, UKCEH</u> : This task will organise data from the content streams in WP4 through FDO to provide improved data findability along with typed digital objects and machine-actionable metadata for improved accessibility to underpin the architecture blueprint for the DT in WP3. The consolidation of data design through FAIR principles and FDOs provide a foundation to organise multidisciplinary heterogeneous data by improving machine-actionability, interoperability and reusability. This consolidation ensures that with enough metadata we can provide information for agents (both human and machine) and support operational requests on the digital objects. With this foundation, the RIs will provide a unified data layer to accommodate computational workflows, interlinking and reusability for the DT technical platform created in WP3.									
Task 5.2: Pilot FAIR semantic mappings and provide them as a service (M1-M24). <u>Senckenberg, EAA, LW ERIC</u> : This task aims to foster joint usage and FAIR sharing of data from different domains. In contrast to pragmatic approaches focussing on the comparison of two datasets based on a tacit matching of the underlying									

ontologies, Task 5.2 aims to process and open up the matching results for subsequent quality evaluation, FAIR compliant reuse as well as estimation and quantification of semantic similarity. The mappings will be provided through various services and tools such as OxO (EBI) and EcoPortal (LW ERIC). The approaches applied in the EOSC Future project will be adopted here to achieve maximum interoperability between the data services from the Biodiversity Digital Twin technical platform and different EOSC services. The task will also experiment with LifeBlock, a cutting-edge technology based on BlockChain provided by LW ERIC, to investigate if this can accelerate the results of the semantic mapping service and enhance the integration of data from cross-domain disciplines.

Task 5.3: Provide data quality indicators and map these to the datasets through FDO (M1-M34). UiO, GBIFS, LW ERIC, EAA, Naturalis: This task provides quality indicators and maps these semi-automated to the FAIR Digital Objects as the data product. Standardization of both data and services is essential for ensuring seamless, programmatic access to data for the technical platform. The task will also identify and address gaps or issues with current community standards that need to be solved to enable the implementation of the quality indicators. The goal of the task is to ensure progressive improvement in the quality and usefulness of the datasets across the different projects with standards, interoperability and/or programmatic access. The quality mapping will serve not only the DT technical platform but will also serve the modellers directly to make decisions which datasets to include or exclude for a specific use case.

Task 5.4: Supporting FAIR workflows and models (M1-M36). UNIMAN, Senckenberg, LW ERIC, UKCEH, TNO. This task will create FAIR metadata for workflows, models, their packaging, interlinking and lifecycle so researchers can understand what is there, what they do and how they operate. All DT objects (data, models, etc) will be linked, creating support for model ensembles and the modelling lifecycle, and enable comparison, exchange, archiving and publishing with persistent identifiers. The integration will address the specific nature of different types of Digital Twins, different levels of the computation tasks, various VRE platforms, heterogeneous data sources, multiple software stacks and data provenance to ensure seamless integration and usage. The work will leverage infrastructure developed in the ELIXIR tools platform and the EOSC-Life Tool and Workflow Collaboratory. This means community-developed schemas, registration in resources (WorkflowHub) and standards (RO-Create / Common Workflow Language) will also be integrated.

D5.1: Pilot Service to open up FAIR semantic mappings for FDO (M24)

D5.2: Pilot integration of workflows and data products organised through FDO with common community workflows and data products used within the RIs (M34)

WP6				Lead Beneficiary: JYU							
Title	Simulation, Modelling and Data Analytics										
Participant number	1	5	6	9	10	12	13	16	20	21	22
Short name of participant	CSC	TNO	JYU	UniSwansea	UNIMAN	UFZ	UKCEH	GBIFS	LWERIC	Bayer	MLU
Person months per participant	30	36	60	36	2	60	12	9	18	3.5	33.5
Start month	1			End Month					36		
Objectives: In this Work Package, use cases defined in WP4 will be used to develop the Digital Twin, to test its predictive performance and to apply the DT to address biodiversity challenges through scenario simulations,											

predictions and biomonitoring methods. This work package aims to integrate and compare the predictive performance of several modelling approaches, stimulating an iterative approach to developing next-generation prototypes.

Task 6.1: Developing pipelines: small-scale pilot studies (M1-18): JYU, UFZ, UniSwansea, MLU, TNO. This task will establish standardized pipelines (workflows from input data to data products) for each use case developed in WP4. The inputs consist of species composition data merged with data on environmental covariates, species traits, phylogenetic relationships, the spatio-temporal context of the study, and the array of management and policy scenarios. The pipelines comprise scripts for model fitting and validation, prediction tasks, scenario simulations, and visualization. Scripts for data pre- and post-processing will be packaged as building blocks that can be applied across use cases. Further to implementing joint species distribution models, agent-based models and theoretical models, this Task will involve the development of hybrid modelling methods combining mechanistic and phenomenological approaches.

Task 6.2: Upscaling and implementing the DT (M6-30). UFZ, CSC, JYU, UniSwansea, MLU, GBIFS, TNO. This task will upscale the modelling approaches developed and tested in Task 6.1. Two lines of investigation will be pursued: 1) developing "transfer functions" that identify subsets of parameters driving local dynamics, and 2) identifying performance bottlenecks using benchmark tests and implementing support for GPU acceleration and/or parallel computing where required by the use cases being addressed. Transfer functions will be identified using scripts developed in T6.1 and high-resolution soil, climate and vegetation data provided, for example, by eLTER. To meet the computational demands of upscaling the DT, models will be implemented on LUMI.

Task 6.3: Testing the predictive capacity of the DT (M12-30). CSC, JYU, UniSwansea, GBIFS. This task will apply the DT to model validation tasks (predictions testable against independent data) for each use case. For this, model outputs will be scanned using AI methods and Approximate Bayesian Computation to identify regularities which were not known or used for model development or parameterization, and to enable systematic comparisons of the modelling approaches across use cases (4.1). Existing data (4.1) and expert knowledge (4.3) will be used to check if the corresponding patterns exist in the real counterpart of the DT. Such independent predictions are critical to acceptance of the DT as a tool for developing decision support and policy development.

Task 6.4: Applying the DT to scenario simulations, predictions and biomonitoring. (M24-36). UFZ, UKCEH, JYU, CSC, UniSwansea, Bayer, MLU, GBIFS, TNO. This task will develop functionality that enables users to apply the DT to scenario simulations and predictions that address the applied aims of the DT, as described under the WP4 use cases. Further, this task will develop functionality enabling users to apply the DT to biomonitoring. By continuously refitting itself to the updating input data (with 4.1 and 4.2), the DT will update the predictions of the global present state of biodiversity in the forms of interactive maps and summaries.

D6.1 Pipeline release and validation (M18). Description and release of pipelines and building blocks for model fitting and validation, prediction and simulation tasks, and data visualization. These will cover each of the use cases and methods being addressed, with validation of the pipelines based on small-scale pilot analyses.

D6.2 Report on model upscaling approaches (M24). Report on approaches implemented for model upscaling. Description of relative performance improvements gained through executing the upscaled models on LUMI. Updated model source code will be released through a public repository.

D6.3 Report on DT predictive capacity (M30). Report on the predictive capacity of the DT in relation to its real-life counterpart, based on validation test results. The report will focus on the performance of the DT as a next-generation tool for aiding decision support and policy development within the field of biodiversity research.

D6.4 Release of end-user features for simulations, predictions and biomonitoring (M36). Release of end-user features for interactive use of the DT, focusing on simulation and prediction tasks related to the use cases. Description of features for producing data summaries and online visualisations (e.g. maps) for now-casting optimal data collection locations for citizen scientists.

WP7				Lead Beneficiary: IT4I@VSB						
Title	Integration & Service Uptake with Research Infrastructure Environments									
Participant number	1	4	5	7	13	15	16	17	18	20
Short name of participant	CSC	IT4I@VSB	TNO	UT	UKCEH	UH	GBIFS	UiO	Naturalis	LW ERIC
Person months per participant	16	43	13	22	4	16	6	8	4	16
Start month	1			End Month					36	
Objectives: This WP provides the technical expertise and enabling support required by the research communities and users to run their data projects on the Digital Twin. The task will initially focus on the uptake plans defined by the community partners and extend later to additional data projects coming from outside the consortium. The main objectives are a) to ensure the uptake of the Digital Twin by the selected research communities, and b) to acquire new user communities to increase the user base of the offered services through the project lifetime. This activity is complemented by a dedicated training programme to increase users’ competencies required to effectively use the Digital Twin during and after the project’s lifetime.										
Task 7.1 Identification of Research Infrastructures’ requirements (M1-M24). <u>UH</u>, <u>IT4I@VSB</u>, <u>TNO</u>, <u>UT</u>, <u>UiO</u>, <u>Naturalis</u>, <u>LW ERIC</u>. This task will execute a continuous survey of scientific communities’ environments and service requirements, with an initial focus on the RIs involved in the consortium. For each RI, we will: (a) characterize the research community environment and extract technical requirements to be fed to the possible extent into the projects’ service developments in T3.1, (b) document usage scenarios, and (c) identify capacity building and training needs and opportunities in dialogue with the use cases through a fitness for use approach (WP4) and co-designing with external user groups. The requirements will be documented and communicated with technical WPs (3-6) and Task 7.3 to make sure they are factored into the service development agendas.										
Task 7.2 - Service uptake & integration with community platforms. (M6-M36). <u>IT4I@VSB</u>, <u>CSC</u>, <u>TNO</u>, <u>UT</u>, <u>UKCEH</u>, <u>GBIFS</u>, <u>Naturalis</u>, <u>LW ERIC</u>. This task is responsible for the uptake of the Digital Twin by the research infrastructures and its integration into communities’ workflows and platforms. Resources in this task are allocated to infrastructure and community technical experts involved in case studies. These experts are responsible for integrating and supporting the necessary Biodiversity Digital Twin platform services within their own community environments. The task will initially focus on the uptake plans defined by the consortium partners and their support for FAIR workflows as created in WP5.4, and extend later to additional user groups coming from outside the consortium. Depending on the requirements collected in Task 7.1, the most relevant tools to support the ease of access and to improve the adoption rate will be considered. These may be for example Puhuri for the AAI, HEAppE for the HPC-as-a-Service, Galaxy for the workflows’ deployment and/or existing RI PaaS such as virtual labs from eLTER and LifeWatch ERIC containing Jupyter Notebook and Tesseract.										
Task 7.3 - Tailored training for identified audiences (M12-M36). <u>LW ERIC</u>, <u>CSC</u>, <u>IT4I@VSB</u>, <u>TNO</u>, <u>UT</u>, <u>UH</u>, <u>UiO</u>, <u>Naturalis</u>. This task will perform a concerted effort during the second phase of the project to engage with research and engineering communities providing them with the necessary knowledge and building on their skills to make them true users of the project’s outcomes. A training plan will be outlined to provide the strategic framework for this task, identifying the main goals and different formats to deliver training sessions, associating and adapting them to the different categories of users, and finally detailing a work plan, media mix and measures to assess and evaluate its performances. The plan will build on the work done in task 7.1 and move from the needs										

expressed by users and RIs to design an adapted training offer based on the identified requirements. Training will focus on Data FAIRness and FAIR principles, and on the use and application of the Digital Twin within use cases T4.1.1-T4.1.4, and will mainly be delivered online via guidelines provided through interactive and recorded lessons, available on a dedicated and project-branded training platform provided by LifeWatch ERIC, which will remain accessible after project end. Moreover, a Summer School addressing RIs and their research and ICT staff is planned by month 34 with the aim of empowering practitioners in the use of DT as well as work on its scalability.

D7.1: First stage community platform integration report (M18). The initial integration process will be defined and related tools will be analysed based on the requirements, architecture development, and data and models development.

D7.2: Final community platform integration report (M36). Demonstration of the Digital Twin integration with external community environments.

D7.3: RI skills mapping & Good Practices and training report (M36). Map the skills & Good Practices identified and developed in other work packages throughout the project and report of the training activities, including a dedicated report on the Summer School.

WP8			Lead Beneficiary: TNO					
Title	Collaboration and Integration with Strategic Initiatives / Programmes							
Participant number	1	5	7	8	11	16	20	
Short name of participant	CSC	TNO	UT	KTH	ECMWF	GBIFS	LW ERIC	
Person months per participant	6	18	6	12	12	12	20	
Start month	1		End Month					36

Objectives: This WP aims to ensure the effective connection between the Biodiversity Digital Twin and other related initiatives and programmes, in particular Destination Earth (DestinE) and the European Data Infrastructure (EOSC, GaiaX, EuroHPC, etc.). In the first year we will focus on integration scenarios, which will then be implemented in the second and third year. The main objective of the WP is to make the DT available in the other European platforms and infrastructures thereby increasing its uptake and use in research workflows.

T8.1 Alignment with Destination Earth (M1-M36). ECMWF, CSC, TNO. This task will ensure full alignment of the BioDT with the DestinE initiative and all its procured actions, in particular those related to the Extreme Earth and Climate Adaptation digital twins. We will make sure system design can be quickly adjusted on either side, and according to the respective developments. This also allows early testing of new inputs generated by ECMWF and procured DT components and testing their impact on biodiversity dynamics for example.

T8.1.1 Interface with DestinE DTE configuration (M1-M24). This subtask will prepare the BioDT data streaming requirements for integration into the DestinE DTE specification. This includes biodiversity DT data processing for exchanging simulation data (input and output) with the Climate DT and providing suitable APIs that facilitate the use of existing DestinE software infrastructures for feeding data to the DestinE data lakes so that the BioDT data becomes accessible to the entire DestinE user base. This includes facilitating the harmonized data access, data fusion and analyses of a wider range of DTs with DestinE external data (e.g. from Copernicus).

T8.1.2 Demonstration of BioDT data stream through DestinE (M18-M36). This subtask will provide a TRL 6-7 type proof of concept demonstration: (i) ingesting DestinE climate DT data for constraining the BioDT, (ii) ingesting the BioDT output in the DestinE data stream feeding into the DestinE data lake, (iii) pushing the DestinE Climate DT contribution into BioDT quality indicators. The demonstration involves early access to provisional

DestinE DT datasets as they are being generated in phase 1 of DestinE for testing data interoperability and workflow set-ups before DestinE becomes operational.

T8.1.3 Interface with the Ocean Digital Twin (M6-M24). This subtask will connect to the Ocean Digital Twin (Iliad) project to allow ocean data and models relevant to biodiversity to be re-used by the BioDT. This will provide a concept demonstration of ingesting ocean data and model output that can be used through the existing ocean research infrastructures and researchers.

T8.2 Alignment with the European Data Infrastructure (M1-M36). TNO, CSC, UT, KTH, GBIFS, LW ERIC.

T8.2.1 Interface with European Open Science Cloud (M1-M36). This subtask will ensure that the architecture of the BioDT is aligned with the EOSC developments and that the proposed services can be fully integrated in the EOSC platform. The goal is to make the Biodiversity Digital Twin available *as a service* for EOSC users, and this task will analyse EOSC technical and policy requirements that need to be considered by the project to realise this. We will also seek to exploit data made available in EOSC by relevant communities as input to the Digital Twin. This will be achieved via a close interaction with the EOSC Future project. A roadmap will be agreed upon by engaging with the project technical team. The ESFRI science clusters and the training team will be engaged to maximise outreach and impact.

T8.2.2 Interface with European Data Spaces Science (M1-M36). This subtask will translate the current state-of-the-art from the European Data Spaces, such as the standards from the International Data Spaces Alliance, Gaia-X and OpenDEI, to the work done in this proposal in terms of 1) Governance on data, 2) Services created with data based on existing core functional blocks and 3) standards used for interoperability. This task will follow further developments and link to other initiatives based on the current Data Spaces Radar (Data Space Radar | International Data Spaces) to ensure that the BioDT and the proposed service can fully ingest and exploit data from relevant domains and the experience from this BioDT flows back to the European Data Space knowledge base.

T8.2.3 Interface with European HPC infrastructure (M1-M36). This project will be a pilot for digital twins specifically in the area of earth system modelling as well as a more general pioneering effort for realising digital twins on top of HPC-based e-infrastructures. Making this effort sustainable requires on one hand the alignment of the design efforts (in particular within WP3) with the evolution of the European HPC-based e-infrastructures ecosystem as well as feedback of lessons learned into this ecosystem. This ecosystem is most prominently represented by the Joint Undertaking EuroHPC (including its advisory groups), the different sites hosting HPC systems owned or co-funded by EuroHPC, and various research and innovations as well as coordination actions funded by EuroHPC. The objective of this subtask is to establish a dialogue with the relevant EuroHPC stakeholders to ensure alignment of this project with the EuroHPC roadmap as well as to feed lessons-learned on the realisation of digital twins back to the EuroHPC ecosystem

D8.1: Report on how the twins of different initiatives can be connected (DestinE, Ocean, EOSC, EUDS, etc) (M24)

D8.2: Overview of lessons learned by connecting multiple twins with (DestinE, Ocean, EOSC, EUDS, etc) that can be adopted by RIs and other work packages (M28)

Table 3.1c: List of Deliverables

Number	Name	WP	Lead participant	Type	Dissemination level	Delivery date
D1.1.	Quality Plan, Risk Register, Ethics & Compliance	WP1	CSC	R	SEN	M3
D1.2	Data Management Plan	WP1	CSC	DMP	PU	M6

D1.3	First, second and third Periodic Report	WP1	CSC	R	SEN	M12/24 /36
D2.1	Plan for dissemination & exploitation including communication activities	WP2	TRUST-IT	R	PU	M6
D2.2	Final plan dissemination & exploitation including communication activities	WP2	TRUST-IT	R	PU	M35
D2.3	Services Uptake report	WP2	TRUST-IT	R	SEN	M36
D2.4	Sustainability report	WP2	TRUST-IT	R	SEN	M36
D3.1	Architecture and implementation plan	WP3	CSC	R	PU	M9
D3.2	Software for digital twin advanced technical Platform	WP3	CSC	DEM	PU	M24
D3.3	Productisation roadmap	WP3	CSC	R	PU	M34
D4.1	Report on the availability of consolidated, improved, fit for use data streams, data model availability, improvement, and application	WP4	GBIFS	R	PU	M24
D4.2	Report of results on applying models, data, and data streams covering the use cases	WP4	GBIFS	R	PU	M32
D5.1	Pilot Service to open up FAIR semantic mappings for FDO	WP5	Naturalis	DEM	PU	M24
D5.2	Pilot integration of workflows and data products organised through FDO with common community workflows and data products used within the RIs	WP5	Naturalis	DEM	PU	M34
D6.1	Pipeline release and validation	WP6	JYU	DEM	PU	M18
D6.2	Report on model upscaling approaches	WP6	JYU	R	PU	M24
D6.3	Report on DT predictive capacity	WP6	JYU	R	PU	M30
D6.4	Release of end-user features for simulations, predictions and biomonitoring	WP6	JYU	DEM	PU	M36
D7.1	First stage community platform integration report	WP7	IT4I@VSB	R	PU	M18
D7.2	Final community platform integration report	WP7	UT	DEM	PU	M36
D7.3	RI skills mapping & Good Practices and training report	WP7	IT4I@VSB	R	PU	M36
D8.1	Report on how the twins of different initiatives can be connected	WP8	ECMWF	R	PU	M24
D8.2	Overview of lessons learned by connecting multiple twins with (DestinE, Ocean, EOSC, EUDS, etc) that can be adopted by RIs and other work packages	WP8	KTH	R	PU	M24

Table 3.1d: List of milestones

Number	Milestone name	WP	Date	Means of verification
M1.1	Project kick-off meeting	WP1	M1	Minutes approved by the partners.
M1.2	Advisory board is established	WP1	M3	All board members appointed and details available to the consortium
M2.1	BioDT Branding	WP2	M1	Branding material available
M2.2	BioDT Web platform	WP2	M3/12/24	Website up and running
M2.3	BioDT Use-case workshops #1, 2, 3 & 4	WP2 WP4	M6/12/14/ 16	Workshop agenda & attendance list
M2.4	BioDT Use-case workshops + joint user	WP2	M22/34	Workshop agenda & attendance

	training workshop #1 & 2	WP7		list
M2.5	BioDT R&I nodes coordination workshops #1 & 2	WP2 WP4	M12/24	Workshop agenda & attendance list
M2.6	BioDT Data standards workshop	WP2 WP4	M17	Workshop agenda & attendance list
M2.7	BioDT Summer school	WP2 WP7	M34	Summer school agenda & attendance list
M2.8	BioDT Webinars x10	WP2	M2/4/11/1 3/18/20/2 3/25/28/3 3	Webinar agenda & attendance list
M2.9	BioDT Bring Your Own Data event #1 & 2	WP2 WP4 WP7	M18/28	Event agenda & attendance list
M2.10	BioDT Joint-event with Destination Earth	WP2 WP8	M35	Event agenda & attendance list
M3.1	Computing and storage environment is available on the LUMI HPC site	WP3	M6	Resources have been earmarked
M3.2	Platform operations are launched and technical communication channels are set up	WP3	M10	Operations and technical communications up and running
M3.3	First Digital Twin as a service is available for Biodiversity RIs.	WP3	M16	Digital twin up and running
M3.4	Second HPC site is connected to the platform	WP3	M24	Site is connected to the platform
M4.1.1	Four workshops / training events and a hackathon on data management and publication for the data holder / use case communities	WP4	M16	Workshop/event agendas & attendance lists
M4.1.2	Report on specifications and research data use by thematic communities	WP4	M18	Specifications described and available
M4.1.3	Submission of high impact methodological open access research publications based on the application of Digital Twin across use cases	WP4 WP6	M32	Publications submitted
M4.2.1	Digital availability of data streams as standardized open data and citable objects	WP4	M6	Data streams are available
M4.2.2	Data papers and policy white papers published on the improvement of European and global data related to agrobiodiversity and on cryptic environments	WP4	M24	Data and policy papers are published
M4.3.1	RI nodes coordination workshops carried out	WP4	M12/24	Workshop agenda & attendance list
M4.3.2	Guidelines on collaboration principles for RI nodes (M16) and a strategy on reduced fragmentation of the data actions nationally (RI nodes) and in Europe published online	WP4 WP2 WP7	M16/28	Guidelines & strategy are published
M4.3.3	Publication on advancement and application of the event core, material samples, and the species interaction/resource relationship domain data model	WP4	M24	Publication submitted
M4.3.4	Two webinars on data citation	WP4	M18	Webinar agendas & attendance lists

M5.1	Implementation plan for the FDO layer by the RIs	WP5	M12	Plan is ready
M5.2	Semantic standards used for data, quality and workflows	WP5	M30	Standards are available
M5.3	Prototype of the FAIR semantic mappings for use cases reusing available mappings (ENVO) and the Semantic Web for Earth and Environment Ontology (SWEET)	WP5	M12	Prototype is ready
M5.4	Prototype FDO service online for the FAIR semantic mapping	WP5	M18	Prototype is ready
M5.5	Fully functional FDO implementation	WP5	M30	FDO is functional
M6.1	Interim report on pipeline development	WP6	M9	Report is published
M6.2	Pilot study results and model bottlenecks	WP6	M12	Pilot conducted
M6.3	Mid-term code release	WP6	M18	Released code is available
M6.4	Toolbox of model comparison methods	WP6	M18	Toolbox is available
M6.5	Interim report on applied functionalities	WP6	M30	Report is published
M7.1	First stage research infrastructures' requirements collected	WP7	M6	Requirements collected
M7.2	Training plan defined	WP7	M18	Training plan is ready
M7.3	Consolidated RIs, external user groups requirements	WP7	M24	Requirements are consolidated
M7.4	User training workshops #1 and #2	WP7	M22/34	Trainings conducted
M7.5	Release of the training package #1 and #2	WP7	M24/36	Training package released
M8.1	Identification of all relevant stakeholders for interaction between Biodiversity DT and other initiatives	WP8	M6	Relevant stakeholders are identified
M8.2	First collaboration agreement between BioDT and other initiatives	WP8	M12	Agreement is done
M8.3	First link between BioDT and other DTs	WP8	M18	Link is established

Table 3.1e: Critical risks for implementation

Description of risk (indicate level of (i) likelihood, and (ii) severity: Low/Medium/High)	WP	Proposed risk-mitigation measures
Lack of response and cooperation from one individual partner preventing timely delivery of project results, including deliverables. Likelihood: Low / Severity: Medium	WP1	Early intervention of Project Manager, leading to escalation to Project Management Board and decision over corrective actions, such as reassessing responsibilities and tasks among the consortium.
Communication and engagement failures. Likelihood: Low / Severity: Medium	WP2	The communications partner has many years of experience in the field & focusing on the impact of the communication and a long, positive track-record in previous EC-funded RIA, IA & CSA initiatives. As a contingency measure, the continuous communication and engagement monitoring activities will ensure timely corrections.
Lack of external user uptake Likelihood: Medium / Severity: High	WP2	BioDT will ensure a proper engagement with all relevant target audiences thanks to dedicated and diverse event formats and engagement channels. Dedicated tasks for scouting and nurturing user uptake, coupled with support for innovation management will take place in WP2 to reduce both likelihood and impact.
Difficulty on attracting attendees to events/workshops	WP2	Advanced planning of the events and daily monitoring of the registered attendees, plus specific social media activities, and

Likelihood: Medium / Severity: High		one-on-one recruiting will be performed. The Partners Network will be tapped into. Organisation of (on site, in person) events together with other community events will reduce travel load (and expenses) for event participants.
Activities delay (due to Covid-19 pandemic) especially in terms of physical meetings. Likelihood: Medium / Severity: Medium	WP2	Events can be run in blended 'live/online' formats. As a contingency measure, alternative ways to reach the stakeholders and achieve the foreseen impact (i.e.: webinars, podcasts) will be proposed and implemented.
Lack of clarity of copyright issues, open standards on interoperability guidelines Likelihood: Medium / Severity: High	ALL	Foster the adoption of relevant standards and good practices, underlying open-source practices already supported by large communities. Partners have experience on standards efforts and lead the CSA StandCIT.eu on this for Europe. Adoption of open source & non-restrictive SW licenses will reduce the likelihood and impact of this risk.
Limitations in data availability Likelihood: Low to moderate Severity: Medium	W P 4 / 5	Securing the flows of biodiversity and abiotic data is interwoven with a number of aspects of evolving academic culture to data openness across the use case / data holder communities. The situation is expected to range from very low risk (species occurrence data, climate data) to moderate risk of data inaccessibility (ecology, agrobiodiversity) to patchy and disjunct data species interaction data with emerging data models. RI partners have 40+ years of experience in promoting data openness and driving the events to overcome such obstacles. Tasks 4.3 are set to preventively address the community and data flow risks. The range of 8 use cases for BioDT across the data complexity and availability gradient ensures the distribution of risks and immunizes the project from data failures. Mitigation actions include refocusing activities to the data-secure use cases.
Difficulties in gathering requirements for semantic mapping Likelihood: Low / Severity: Medium	WP5 WP6 WP7	Requirement gathering and analysis will be addressed before the service implementation. The partners in task 5.2 have expertise in understanding the complexity of semantic mapping and building services. Also, close collaboration with WP4 and WP7 will help with gathering requirements before implementation.
Risk of unwanted noise and bias in the data (low quality data) Likelihood: High / Severity: Low	WP5	Data quality indicators from 5.3 will ensure FAIRness and a feedback loop for reporting issues back to the data publisher. Coordination across the work packages is essential for mitigating this risk. Sensitivity of data modelling approaches to bias across taxon groups, spatial and temporal coverage will be managed through the choice of algorithms and validation of modelling results.
Gaps in FAIR data elements (such as lack of machine-actionable metadata, Persistent Identifiers, community standards) from the data providers Likelihood: High / Severity: Medium	WP5	Regular communication with the data providers, adherence to the design specification and requirements will be used for mitigation. Also, the FDO layer and quality requirements in place can mitigate some of the gaps in the data.
Lack of engagement and commitment on FAIR data from the RIs Likelihood: Low / Severity: High	WP5	FAIR data design requires regular engagement and long-term commitment from the RIs. The existing resources and expertise around FAIR within the RIs will be utilised to ensure partners understand the importance of FAIR data. (add metadata, community standards)
Lack of FAIR data from other digital twin projects Likelihood: Low / Severity: Low	WP5 WP8	The advantage of multidisciplinary digital twins is the ability to combine data and models. Outreach and collaboration will be in place to ensure cross project / domain data FAIRness.

Failure to collect appropriate requirements in time for Digital Twin implementation Likelihood: low / Severity: Medium	WP7	Increase of the number of the meetings related to the requirements collection from RIs and users
Delays in API or software module implementations. Likelihood: medium / Severity: medium	WP3 WP4 WP5 WP6 WP7	Project implementation will start with existing high TRL components. Continuous feedback from the RIs and users will help to early identification of the potential problems and their mitigations.

Table 3.1f: Summary of staff effort

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	Total PMs per Participant
CSC	36	12	48	6		30	16	6	154
TRUST-IT	6	23							29
COMMPLA		11							11
IT4I@VSB	6	4.5	24	6			43		83.5
TNO	6	2.5	12	6	4	36	13	18	97.5
JYU	6	0.5		6		60			72.5
UT		0.5	48				22	6	76.5
KTH		1.5	21					12	34.5
UniSwansea		0.5				36			36.5
UNIMAN		0.5			12	2			14.5
ECMWF		6						12	18
UFZ		8		42		60			110
UKCEH		3.5		26	9	12	4		54.5
EAA		3.5		12	8				23.5
UH		5.5		6			16		27.5
GBIFS	6	6		96	8	9	6	12	143
UiO		3.5		58	12		8		81.5
Naturalis	6	6		18	16		4		50
Senckenberg		0.5			15				15.5
LW ERIC		6		15	15	18	16	20	90
Bayer		0.5		2		3.5			6
MLU		0.5		2		33.5			36
Total PMs	72	106	153	301	99	300	148	86	1 265

Table 3.1g: ‘Subcontracting costs’ items

Not applicable

Table 3.1h: ‘Purchase costs’ items

Participant Number/Short Name: 2 / Trust-IT		
	Cost (€)	Justification
Other goods, works and services	60 350 €	WP4&WP5&WP7: Use case workshops (including training) for 70 participants @70€ per participants, logistics costs 1.5k€ x 4 events (€25.600); 2 Workshops for BYOD @ 70 € per participant, 2k€ for logistics for 70 participants (€11.800); 2 RI nodes coordination workshop for 70 participants @70€ per package @1.5€ logistics costs (€11.300); 1 Data Standards workshop for 70 participants @70€ package, 1.5€ logistics costs (€6.400); Dissemination Pay-per click campaigns /PPC 105 x €50 (€5.250) = Total €60.350

Remaining purchase costs (<15 % of pers. costs)	21 900 €
Total	82 250 €

Table 3.1i: ‘Other costs categories’ items (e.g. internally invoiced goods and services)

Not applicable

3.2 Capacity of participants and consortium as a whole

The BioDT project brings together a unique consortium of e-Infrastructure providers, research infrastructure operators and researchers from a wide range of scientific disciplines and backgrounds.

The consortium is coordinated by **CSC-IT Center for Science** (Finland) which has been implementing the Finnish Open Science agenda since 2011, and provides nationally the services funded under the Finnish national open science initiative which is directly steered by the Finnish Ministry of Research and Education. CSC operates ‘end-to-end’ across the networking, grid, HPC and data-management layers of the e-infrastructure and hosts and operates the EuroHPC pre-exascale machine LUMI, on which the BioDT will be built. CSC is one of the drivers of the development of the European data infrastructure and is particularly active in the EuroHPC, the European Open Science Cloud, and the Gaia-X initiatives, which makes it an ideal partner for bridging the DT with the wider infrastructure ecosystem. In this project, CSC’s main role, in addition to coordinating the project (WP1), will be to provide and operate the technical platform that will support the digital twin (WP3). CSC will also contribute with its data analytics and data management teams to the FAIR data (WP5), modelling (WP6), sustainability and exploitation (WP2), integration with research infrastructure environments (WP7) and collaboration with DestinE and the European data infrastructure (WP8) activities.

Several members of the **LUMI** consortium have come together in this proposal with a common desire to exploit the LUMI machine to support the Digital Twin. All these partners are leading computing resource providers in their own countries and will provide expertise in infrastructure design, data analytics, and user community support. **IT4I@VSB** (Czech Republic) operates the national Czech supercomputing infrastructure for science and industry and has a long experience of supporting a wide range of scientific communities. In this project, IT4I@VSB will lead the activities related to integrating the digital twin platform with research infrastructure environments (WP7), in addition to participating in the development of the platform (WP3), and community outreach activities (WP2). The Estonian Scientific Computing Infrastructure (ETAIS), represented in this project by the **University of Tartu** (Estonia) provides computing infrastructure and application support services for researchers and companies in Estonia and will contribute to WP3 and WP7. The **Royal Institute of Technology in Stockholm – KTH** (Sweden) is one of Europe’s leading technical and engineering universities and will also contribute to the development of the technical platform (WP3 through the involvement of its Center for High Performance Computing – PDC. In addition, it will lead the task on collaboration with the EuroHPC initiative (WP8)

In terms of the research infrastructures, which are contributing about 50% of the total project effort, four large ESFRIs research infrastructures are directly involved in the consortium. This strong RI involvement is essential for realising the digital twin following a co-design approach. Indeed, the RIs involved will not only provide data but also actively contribute to the technical development of the platform and the models. The **Global Biodiversity Information Facility (GBIF)**—, represented in this project by **GBIF Secretariat** (hosted by Denmark), and the **University of Oslo** (Norway), hosting GBIF Norway — is an international network and data infrastructure funded by the world’s governments and aims at providing anyone, anywhere, open access to data about all types of life on Earth. Coordinated through its Secretariat in Copenhagen, the GBIF network of participating countries and organizations, working through participant nodes, provides data-holding institutions around the world with common standards and open-source tools that enable them to share information about where and when species have been recorded. In BioDT, GBIF will lead the Data and Use cases work package (WP4), contribute to the work on FAIR data (WP5), model

development (WP6), integration activities with RIs (WP7), and linkage with the European Open Science Cloud (WP8). The **Integrated European Long-Term Ecosystem, critical zone and socio- ecological system Research Infrastructure (eLTER)** is a new distributed Research Infrastructure which joined the ESFRI Roadmap in 2018. It aims at integrating traditional natural sciences and holistic ecosystem research approaches, including studies of the human component, to better understand ecosystems. eLTER will provide indispensable integrated data sets – abiotic, biotic, societal, covering all system structures and functions – for system model development and validation, hence supporting predictions and decision-making in an interdisciplinary framework and in collaboration with other *in situ*, domain specific RIs. In the BioDT project eLTER is represented by the **Helmholtz Center for Environmental Research – UFZ** (Germany), the **UK Centre of Ecology & Hydrology - UKCEH** (United Kingdom), the **European Environmental Agency - EEA** (Austria), and the **University of Helsinki - UH** (Finland). e-LTER will contribute to use cases (WP4). UKCEH and EEA will work on FAIR data (WP5), and UFZ and UKCEH will contribute directly to the development of models (WP6). The University of Helsinki will lead the work on research infrastructure environments and requirements with contribution from UKCEH (WP7). The **Distributed System of Scientific Collections – DiSSCo**, represented by **Naturalis Biodiversity Center** (Netherlands), and **Naturmuseum Senckenberg** (Germany) is a new world-class Research Infrastructure (RI) for Natural Science Collections which joined the ESFRI Roadmap in 2018. DiSSCo mobilises, links and delivers currently fragmented biodiversity and geodiversity information at a scale, form and precision required, tooling scientists to address major challenges of the Anthropocene. DiSSCo represents the largest ever formal agreement between natural history museums, botanic gardens and collection-holding universities in the world. The DiSSCo RI also aims to create a new business model for one European collection that digitally unifies all European natural science assets under common access, curation, policies and practices that ensure that all the data is easily Findable, Accessible, Interoperable and Reusable (FAIR principles). In the BioDT project DiSSCo will contribute to use cases (WP4) and lead the FAIR data work package (WP5). The **e-Infrastructure for Biodiversity and Ecosystem Research (LifeWatch ERIC)** is the fourth RI in the project. Also on the ESFRI roadmap, LifeWatch ERIC is a landmark distributed Research Infrastructure to facilitate advancement of biodiversity research and addressing the relevant big environmental challenges. It also supports knowledge-based strategic solutions to environmental preservation. This mission is achieved by providing access to a multitude of datasets, services and tools, provided through dedicated Virtual Research Environments (VREs). It will contribute to our communication and dissemination activities (WP2), the use cases (WP4) and the models (WP6), as well as FAIR data activities (WP5), integrating activities with RI environments (WP7) and collaboration with strategic initiatives, in particular EOSC (WP8).

The consortium also includes a number of scientists and scientific teams which have specialised in biodiversity dynamics and modelling. These teams have been working together on biodiversity models for several years and will allow the project to remain scientifically sound and nurture its research agenda. The **University of Jyväskylä** (Finland) will be leading the work on simulation and data modelling (WP6) and will be supported by scientists from the **University of Swansea** (UK) and the **Martin Luther University of Halle-Wittenberg** (Germany). Aspects of the proposed work build on ideas and data generated from an ERC funded project called LIFEPLAN, which aims to establish the current state of biodiversity across the globe, and generate accurate predictions of its future state under future scenarios.

Another set of partners ideally complements the consortium by bringing specific expertise, discipline and geographical outreach. The **Netherlands Organisation for Applied Scientific Research - TNO** (Netherlands) connects people and knowledge to create innovations that boost the competitive strength of industry and the well-being of society in a sustainable way. In this project, TNO will bring expertise in the digital twinning business as part of WP3 and WP6, collaborate on data interoperability (WP5), and lead our strategic collaboration work package (WP8) developing the links between the project and the EU Data Spaces. The **University of Manchester** (UK) will contribute to the development of FAIR activities and semantic data standards in WP5, bringing also expertise from and linking with the bioinformatics community. **Bayer** (Germany) will contribute to scenario developments for case studies, leveraging past experience and expertise in grassland modelling (WP4 and WP6). The **European Centre for Medium-Range Weather Forecasts - ECMWF** will facilitate the link with DestinE. ECMWF is foreseen to be

in charge of the design and deployment of the Digital Twin Engine (DTE) of DestinE that will deliver the Earth-system simulation and Earth-observation data assimilation workflows feeding the DestinE high-priority digital twins (DT) for weather induced and geophysical hazards (Extremes DT) and climate change adaptation (Climate DT). ECMWF's role in this project aims at providing a continuous interface between the DTE developments managed by ECMWF and the technical developments in this project such that software platforms, data handling and information extraction between Climate DT and biodiversity DT can be synchronized, apparent development gaps can be identified and solutions be developed for implementation in future phases of DestinE (WP8). Finally, we have chosen **TRUST-IT services (Italy)** as our communication partner (WP2) to deliver our message effectively to the widest audience and make sure that awareness of the project is optimal. **TRUST-IT** will avail **Commpla srl** as an associated entity.

Industrial and commercial involvement in the project

Digital twinning is a concept that first emerged in the industry business and which is slowly entering the scientific sector. Industry is progressing ahead with its own offerings and capabilities in this space, and might have an influential role in the definition of data standards and new technologies which BioDT must be cognisant with. Spreading data management capabilities among commercial providers could also be a cost-efficient approach for BioDT and will be addressed as part of our exploitation plan in WP2. It is also possible that BioDT becomes attractive to a wide range of users, drawn not only from academia and the public sector, but also from industry, and we will organise specific actions to target this segment. The project includes one commercial company (Bayer) which will contribute with its expertise to the development of the digital twin (see above).

External advisory board

The purpose of the EAB is to provide advisory strategic support to the project and feedback on the Project, helping to identify European priorities and existing challenges, and help connect the BioDT project with other strategic initiatives. It is composed of the following experts: Jerry Lanfear - Elixir Chief Technical Officer (UK), Erja Turunen - Executive Vice President, Digital Technologies VTT, member for GAIA-X Board of Directors (Finland), and Sarah Jones - EOSC engagement manager at GEANT, member of the EOSC Board of Directors.