Ref. Ares(2024)6949276 - 01/10/2024

Project Title	Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities
Project Acronym	BioDT
Project Number	101057437
Type of Project	RIA - Research and Innovation Action
Topics	HORIZON-INFRA-2021-TECH-01-01
Starting Date of Project	1 June 2022
Ending Date of Project	31 May 2025
Duration of the Project	36 months
Website	www.biodt.eu

D8.1 - Digital Twins Ecosystem Interest Group Manifesto

Work Package	WP8 | WP Collaboration and Integration with Strategic Initiatives / Programmes
Task	T8.1 | Alignment with Destination Earth
Lead Authors	Thomas Geenen (ECMWF), Nick Oostervink (TNO), Timea Biro (CSC)
Contributors	Christos Arvanitidis (LifeWatch ERIC), Cristina Huertas (LifeWatch ERIC), Tero Aalto (CSC), Adam Warde (ECMWF), Dirk Pleiter (KTH), Ilja Livenson (UTARTU)
Peer Reviewers	Dirk Pleiter (KTH), Ilja Livenson (UTARTU), Dag Endresen (UiO)
Version	V1.0
Due Date	31/05/2024
Submission Date	31/05/2024
Resubmission Date	01/10/2024

Dissemination Level

X	PU: Public
	SEN: Sensitive – limited under the conditions of the Grant Agreement
	EU-RES. Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON. Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC. Classified Information: SECRET UE (Commission Decision 2005/444/EC)

Funded by the European Union

Version History

Revision	Date	Editors	Comments
0.1	16/04/2024	Thomas Geenen (ECMWF)	Initial document, ToC, paper outline
0.2	18/04/2024	WP8 members	Work Package (WP) discussion and adjustments on the outline of the deliverable
0.3	19/04/2024	WP8 members	Collaborative edits of the document
0.4	25/04/2024	Nick Oostervink (TNO), Timea Biro (CSC)	Content in collaboration sections
0.5	26/04/2024	Thomas Geenen (ECMWF), Timea Biro (CSC), Nick Oostervink (TNO), Dirk Pleiter (KTH)	Content in all sections
0.6	29/04/2024	Thomas Geenen (ECMWF)	Document for peer review
0.7	15/05/2024	Thomas Geenen (ECMWF), Timea Biro (CSC), Nick Oostervink (TNO)	Reviewer comments implemented
0.8	17/05/2024	Thomas Geenen (ECMWF), Timea Biro (CSC), Nick Oostervink (TNO)	Internal WP8 comments considered and document sent to BioDT Project Management Board (PMB) review
0.9	25/05/2024	Thomas Geenen (ECMWF), Timea Biro (CSC), Nick Oostervink (TNO)	Processed final comments and details
1.0	31/05/2024	Gabriela de Paula Souza Zuquim (CSC), Anna-Liisa Allas (CSC)	Submission-ready version incorporating BioDT Project Management Office (PMO) editorial changes
1.0	01/10/2024	Anna-Liisa Allas (CSC)	Deliverable name corrected from “Report on Connecting Multiple Digital Twins” to “Digital Twins Ecosystem Interest Group Manifesto” as per the BioDT Description of the Action (DoA)

Glossary of Terms

Item	Description
DEDL	Destination Earth Data Lake
DestinE	Destination Earth
DT	Digital Twin
GBIF	Global Biodiversity Information Facility
HPC	High Performance Computing
pDT	Prototype Digital Twin
WP	Work Package

Keywords

Biodiversity, Digital Twin, Destination Earth, Collaboration, Modelling, Simulation, Prediction, High-Performance Computing, Artificial Intelligence

Disclaimer

The BioDT project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057437 (https://doi.org/10.3030/101057437).

Executive Summary

This deliverable contains information about the process and work done in Task 8.1, specifically on the information collected on how Earth System digital twins and digital twins, which are part of the wider Green Deal Initiative, could operate together in a federation of digital twin systems. It is composed of six sections: The first section provides an introduction to the deliverable. The second section sets the scene for the rest of this document and describes the digital twin initiatives that are in the scope of this deliverable, and outlines the major commonalities and differences relevant to the work being done in Work Package 8 Task 8.1. The third section addresses technical alignment. The fourth section closes the report with a high-level roadmap towards interoperability.

Table of Contents

1. Introduction

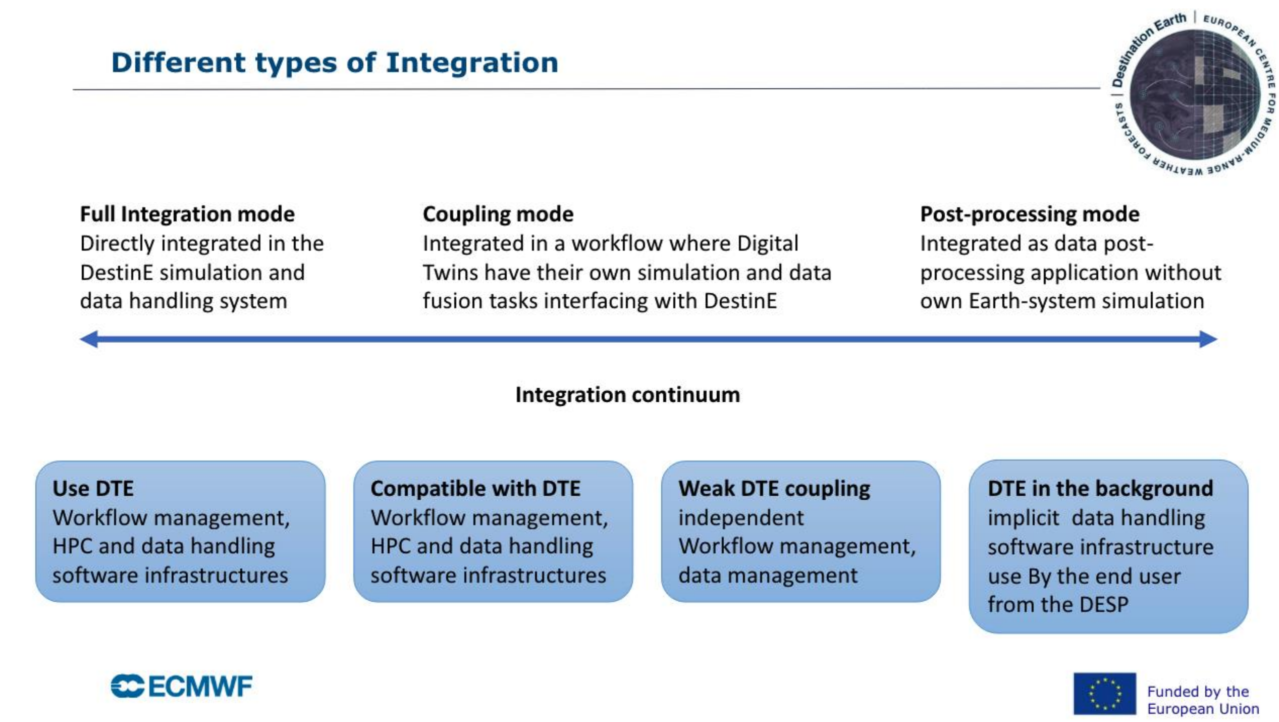
The Biodiversity Digital Twin (BioDT) project aims to push the current boundaries of our understanding of biodiversity dynamics by developing Biodiversity Digital Twin (BioDT) prototypes that provide advanced modelling, simulation and prediction capabilities. The project exploits existing technologies and data available across relevant biodiversity and broader environmental data research infrastructures in new ways, making it possible to accurately and quantitatively model interactions between species and their environment. The BioDT project does not operate in a silo but aims to leverage the broader context and forge valuable synergies with similar efforts, and especially other Earth System digital twins and initiatives. The aim of this document is to describe how Earth System digital twins and digital twins that are part of the wider Green Deal Initiative could cooperate in a federation of digital twin systems. For this purpose, we explain in detail the processes, technical implementation and ontology alignment that need to be established in order to allow for the interoperability of digital twin systems stemming from different communities and initiatives that are consolidated under the wider Green Deal initiative.

This document is not intended to provide a generic interoperability framework but works from the assumption that the most value can be derived from providing specific solutions, driven by use cases, that are generic by design but not designed to be generic. We, therefore, define specific use cases in which the connection of data and information, directly feeding from digital twin data, creates value. Clearly, not only technical aspects are in the scope of this effort but also policy, governance, ontologies alignment and stakeholder engagement.

2. Setting the Scene & Scoping the DTs

2.1. From Connection to Integration

We are aiming for integration through federation, in which each system can be integrated through integration functions or services that allow interoperability between digital twin system components when required. We are not aiming for integration through aggregation. The level of integration requirement between digital twins can span a wide range of functions and services, from full integration on the physics or algorithmic level (tightly coupled digital twins, for instance coupling ocean processes to atmospheric processes) to integration through DT outputs (loosely coupled). The diagram below provides further details on how this could be conceptualised for digital twins in the Destination Earth context.



Source: Digital Twins of the Earth system - Thomas Geenen et al. at 20th ECMWF Workshop on high performance computing in meteorology 2023; https://ecmwfevents.com/assets/presentations/hpcws2023-geenen1697008723.pdf

To capture these requirements, we leveraged an integration continuum, in which we can map integration requirements between digital twins, digital twin platforms, and digital twin systems. From these pilots, part of broader community efforts, we aim to evolve a shared high-level architectural view and, also a shared glossary that can describe the implementation of each participating project. This document highlights the BioDT contributions and interactions with other DTs in support of this ambition.

2.2. Scoping the Digital Twins

The digital twins that are in the scope of this deliverable have all been funded under the same Horizon Europe call HORIZON-INFRA-2021-TECH-01-01 "Interdisciplinary digital twins for modelling and simulating complex phenomena at the service of research infrastructure communities"[1] or are closely associated with the Green Deal initiative. Specifically, this document is now oriented around the initial collaboration with *DT-GEO: A Digital Twin for GEOphysical extremes*[2] and *interTwin: An interdisciplinary Digital Twin Engine for science projects*[3], in addition to the joint efforts to establish a link with the Destination Earth Data Lake[4] and data handling systems through use cases from EUMETSAT and efforts around the Digital Twin Ocean[5] via the DTO-BioFlow[6] sister project.

BioDT, DT-GEO and interTwin projects have a special mention of or a task on integration with the Destination Earth initiative[7], interest reciprocated as in the DestinE broad objectives there also is the notion of connecting

[1] https://cordis.europa.eu/programme/id/HORIZON_HORIZON-INFRA-2021-TECH-01-01

[2] https://dtgeo.eu/

[3] https://www.intertwin.eu/

[4] https://destination-earth.eu/destination-earth/destines-components/destination-earth-data-lake/

[5] https://digitaltwinocean.mercator-ocean.eu/

[6] https://dto-bioflow.eu/

[7] https://destination-earth.eu/

to other digital twins in the context mentioned above but also for future Horizon Europe calls and even potentially for initiatives from outside the European Union. This document describes the important steps taken towards connecting digital twins, digital twin platforms, and digital twin systems. The approach we follow is to define an area of shared interest and connection that clearly illustrates the added value of the connections. Sometimes there is already a strong business case for bringing digital twin information together in a single platform, through federation, without the explicit coupling of the twins themselves.

2.3. Scoping the Alignment

Digital twins from different communities tend to be very different in nature. The digital twin concept implementation for the early pilots in different fields is typically an evolution of existing simulation and data handling implementations which is domain-specific. Here, we have an interesting population of digital twin systems under consideration and it is clear that they could benefit from being able to interoperate, or at least exchange data and information between them. Through the work in BioDT Work Package 8, a wide range of digital twin initiatives have been identified and collaborations continue to be sought and activated.

A first collaboration that has been established and is part of the focus of this deliverable, is the work of BioDT with DestinE, DT-GEO, interTwin, among others, to develop a Glossary for DTs and align it with the Digital Twin Consortium Glossary[8], together with a set of cross-DTs use cases (see further detailed in section 2.4.1).

Although not a digital twin or a project *per se*, for the purpose of this deliverable we also consider a set of use cases funded[9] by EUMETSAT[10] in the context of the Destination Earth programme. The use cases developed in collaboration with BioDT pDTs are meant to demonstrate the benefits of the Destination Earth Data Lake and potentially pilot a mechanism for the sharing of biodiversity data within the DestinE context and other federated data holdings. Destination Earth is viewed as the most relevant and strategic digital twins initiative within the EU, as it strives to establish an infrastructure that will facilitate the integration of multiple data sources from across the EU. Hence, for BioDT, the interoperability with the DestinE and other digital twin projects will prove to be crucial for future successful integrations.

It is important to note that the choices on the DTs to consider in the scope of this deliverable are the result of the discussions and efforts of the partners in the WP, not meant to exclude potentially relevant DT projects as we foresee future collaborations with other DT initiatives too. One example is the DTO-BioFlow[11] in which many of the partners from BioDT also participate, and we see the potential for alignment later on as developments within the DTO-BioFlow advance. BioDT and DTO-BioFlow are both EU projects on digital twinning related to biodiversity supporting the objectives of the Mission and the EU Biodiversity Strategy. The BioDT project focuses on terrestrial ecosystems while DTO-BioFlow relates to marine ecosystems and is viewed as an important link for the future alignment with the Digital Twin Ocean, as detailed in section 3.3 below.

2.4. Digital twin Commonalities and Complementarities

This section will highlight some of the key commonalities and differences across the four DT projects to indicate where synergies were found and where these projects differ but complement one another.

[8] https://github.com/digitaltwinconsortium/dtc-glossary/blob/main/glossary.md

[9] At the time of writing, the two contracts are being negotiated.

[10] EUMETSAT | Monitoring the weather and climate from space | EUMETSAT

[11] Biodiversity Data for Digital Twins of the Ocean | DTO-BIOFLOW

2.4.1. BioDT, DT-GEO, DTO-BioFlow, and interTwin

For each of the projects, there are clear commonalities and differences that have consequences for the extent to which the different projects can interconnect, exchange knowledge, and orient towards Destination Earth.

Commonalities

- **Strategic Goals:** All four projects aim to support decision-making processes, though each project has a different end-focus. BioDT, DT-GEO, and DT-BioFlow aim to (among others) support policy-makers and implementers, while interTwin aims to create interoperability among different DTs in different sectors. Put differently, all four projects aim to learn how to build digital twins that can become able to predict future scenarios and outcomes through the use of DTs. Whether this is about the impact of climate change on biodiversity or the occurrence of natural disasters.
- **Contributing to Destination Earth:** Each project contributes differently to the overall DestinE project: BioDT by creating detailed digital twins for specific aspects of the Earth's ecosystems, DT-GEO by creating a dynamic and detailed model of the Earth's geophysical processes, and interTwin by focusing on the capability of different digital twins to communicate and operate together seamlessly, which is crucial for the integration of diverse digital twins under the Destination Earth system.

The four projects, though each with their unique focus, partners and intricacies, have some major commonalities in trying to improve decision-making for policy-makers who want to tackle climate change, and all four are focused on contributing significantly to the DestinE project.

Complementarities

- **Project focus:** Whereas BioDT focuses exclusively on biodiversity, DTO-BioFlow on marine biodiversity, DT-GEO concentrates on the earth’s geophysical system, and interTwin focuses on the interoperability of various DTs themselves.
- **Complexity & Specificity of Data:** All projects require high-quality data but the type and specificity of the data vary greatly. BioDT focuses on biological and ecological data, DT-BioFlow on marine biodiversity, DT-GEO on geophysical data, and interTwin on data interoperability across various DTs.

2.4.2. DestinE focused EUMETSAT Use Cases

Under the umbrella of the EUMETSAT use cases the BioDT Crop Wild relatives pDT and the Forestry biodiversity pDT pioneer and demonstrate the benefits of the usage of the DEDL services and federated datasets.

BioDT and the EUMETSAT use cases share a common interest in the development of a data adaptor that would enable a federation between one of the key BioDT research infrastructures, GBIF, and the DEDL. This would help establish a mechanism for the sharing of biodiversity data with DestinE.

2.5. Benefits for Alignment of DTs and DT Users

Aligning multiple DT projects in Europe provides a range of benefits for all parties involved and assures, in a broader sense, an improved service and user experience for potential end users. During the project, a list of beneficial collaborative activities has been outlined. The list below has also been part of the different letters of intent that BioDT has established (and will continue to establish) with different DT projects. This list highlights where the DT project sees possible benefits for alignment.

- Explore, discuss and share expertise on topics of common interest (e.g. biodiversity, digital twin technology, workflows, interoperability etc.).

- Present specific activities, tools, results and other outputs carried out in the framework of both projects/initiatives (e.g. architecture design, customising workflow tools, PID and dynamic data).
- Share relevant good practices in the topics of common interest (i.e. through the distribution of key project deliverables).
- Collaborate on community-level challenges (e.g. use case development, user requirements collection and harmonisation).
- Participate in and/or develop joint training activities, supporting collaboration on the topics of training materials and sharing of training best practices and experiences, know-how and technology.
- Contribute to disseminating the results of the projects in the framework of both communities of interest (e.g. in conferences, meetings, news and logos on the website, communication tools, etc.).
- Communication and outreach activities: actions for reaching various targeted audiences as well as wider ones and the general public, that will raise awareness around shared areas of interest.
- Stakeholder engagement: joint effort for expanding the stakeholders' community, spanning from professionals, and policy-makers, to researchers and citizens as well as other strategic initiatives.
- Joint workshops or conference sessions to present developed outcomes.
- Preparation of joint publications.

To illustrate how these collaborative activities have been pursued, we note that several trilateral meetings between BioDT - DT-GEO - interTwin were held starting from September 2023. The effort was kicked off following a joint session BioDT organised at the European Open Science Cloud Symposium[12] which highlighted the needs and opportunities such a collaboration can bring forward.

The session was followed through joint applications submitted for the EOSC Symposium 2024 Unconference and the EGI Conference 2024. The collaboration sees further highlights during several events such as the EGU2024[13] or the RDA 22nd Plenary.[14] A joint position paper is also planned to reflect on common challenges, shared approaches to user engagement and impact of digital twins from a social and economic perspective.

3. Technical Alignment

The effort to connect different digital twin initiatives has been greatly supported by the technology transfer effort initiated by DG-CNECT. Especially in the context of how these digital twin initiatives relate to DestinE, guidance and templating are required to make this effort manageable with a steadily expanding digital twin landscape. Although this effort has a significant technical component, explored through running interoperability pilots with the different DT projects, and alignment on glossary and high-level architecture gives guidance for interoperability implementation, we recognise that the engagement model and the establishment of an overarching governance model is essential to make interoperability a sustainable effort. For instance, we identified that there is funding required to execute interoperability efforts and, also that alignment on goals and mutual benefits requires a governance model where the business case for interoperability can be established. We recognize that this still has to be established and we are not yet in a position to define a governance model to support this.

---

[12] https://symposium23.eoscfuture.eu/symposium/unconference-session-digital-twins-and-eosc-advancing-data-driven-science-through-optimal-data-use-data-management-computing-and-research-infrastructures-insights-from-biodt/

[13] https://meetingorganizer.copernicus.org/EGU24/session/48007

[14] https://www.rd-alliance.org/digital-twins-supporting-data-driven-science-through-optimal-data-management-and-sharing/

3.1. Cross-DT Alignment - Technical Glossary and Selected Pilots

A first collaboration that has been established and is part of the focus of this deliverable is the work of DestinE with BioDT, DT-GEO[15] and interTwin[16], among others to establish a Glossary for DTs and align it with the Digital Twin Consortium Glossary[17]. The benefits of exchanging data and information and running digital twins from different communities side by side.

More specifically, there are several activities that have proven highly beneficial for alignment among DT projects.

- **DT Glossary**

  One important activity has been the establishment of a glossary of DT concepts and definitions among DestinE, BioDT, DT-GEO, and interTwin. Working on this glossary has proven to strengthen the understanding of the different terms, concepts, and definitions in the projects. This helps to understand the differences and similarities, but more importantly, also provides fertile ground to solidify a common understanding in the DT space of certain main concepts and their definitions. This will help future projects, preventing the need to reinvent the wheel and assuring there is less misunderstanding across projects and consortium partners.

- **Co-development of pilots**

  As mentioned above, the level of integration requirement between digital twins can span a wide range of functions and services, from full integration on the physics level (tightly coupled digital twins) to integration through DT outputs (loosely coupled). A benefit of a more loosely coupled approach is that pilot projects (e.g., data exchange between similar but different use cases) can be co-developed. Through regular interaction, the different projects can exchange ideas and approaches on how to, for example, exchange or reuse data from each other's use case(s). This assures synergy where each project can retain its authentic identity and at the same time synchronise with other projects to create more value. Loosely coupling different DT projects can also create value itself. As the different DT projects are aligned, the different parties involved can find and utilise the right information and data more efficiently.

For these studies, we treat the digital twin of one domain as an application in the other domain. What does this mean? Typically, we consider a digital twin as a full implementation of a digital twin system with all features present to use the system as a digital twin. Applications on the other hand are typically implementations that leverage several digital twin services in the context of a use case. In the context of interoperability and the selected use cases, this can exemplified as follows. A meteotsunami digital twin from the DT-Geo project e.g., interoperates with DestinE by consuming its stored representation, in this case, extreme low -pressure fields, and consumes that through the data handling service and service interfaces of DestinE. In the future, we can foresee that some digital twin systems will be integrated as digital system- of-systems where multiple digital twin systems form a larger digital twin system- of- systems. In the context of this study, we recognize that we are still some distance away from such a setup and focus on integration via integration interfaces that operate on specific digital twin services for data handling and information retrieval.

[15] DT-GEO (dtgeo.eu)

[16] interTwin project - Home

[17] dtc-glossary/glossary.md at main · interTwin-eu/dtc-glossary · GitHub

3.2. Alignment with Destination Earth via EUMETSAT Use Cases

For the technical alignment, two BioDT prototype DTs were considered for more encompassing use cases in response to EUMETSAT DestinE Data lake pilot projects: (1) the Crop Wild Relatives pDT[18] and (2) the Forest/Bird Biodiversity pDT[19].

3.2.1. *Crop Wild Relatives*

Develop mitigation options based on using genetic resources from crop wild relatives (CWR) for mitigating the consequences of climate change on agricultural crops over Europe.

Output:

a. (Prototype) Web Interface visualising improved capabilities for crop simulation and decision management (i.e. habitat suitability maps on a crop-by-crop basis for their CWRs).
b. Data transfer between DEDL and data consumer/pDT CWR based on DEDL and GBIFs federation and BioDT's model infrastructure (RO-Crate-based).

3.2.2. *Forest/Bird Biodiversity*

Forest Biodiversity modelling under different climate conditions. The Forest Biodiversity use case will demonstrate the potential of detailed predictions of forest state variables for estimating the temporal trajectories of boreal forest biodiversity and carbon stock.

The key ideas demonstrated in the use case are:

a. EO data available in the DEDL can be used to produce high-resolution maps of forest structural parameters.
b. The DT-predicted future weather (climate) data available on DEDL can be used to evaluate the development course of forest structure and productivity.
c. The predicted forest attributes can predict its biodiversity. We will use Finnish birds as the case study.
d. The processing relevant to this can be implemented in a cloud platform similar to that available in DEDL.

3.3. Pathway for the Alignment with the Digital Twin Ocean via DTO-BioFlow

BioDT and DTO-BioFlow are both EU projects on digital twinning related to Biodiversity supporting the objectives of the Mission and the EU Biodiversity Strategy. The BioDT project focuses on terrestrial ecosystems while the second relates to marine ecosystems and is viewed as an important link for the future alignment with the Digital Twin Ocean.

Both projects aim to understand the interaction of species with each other, with their environment and how these change in response to environmental and anthropogenic pressures through time. Both projects will provide virtual representations of the ecosystems by integrating Earth observations, modelling, and digital infrastructures. Both projects will allow us to simulate and study "what if" scenarios enabling effective conservation, management and policy development. Both projects will provide workflows that simulate end-to-end processes in marine biological systems, their interactions and the consequences of environmental and

[18] Crop wild relatives and genetic resources for food security | BioDT

[19] Biodiversity Dynamics | BioDT

anthropogenic pressures on them making use of powerful computational capacity and artificial intelligence systems. Both will develop different use cases (in BioDT, 10 Use Cases; In DTO-BioFlow 7 DUCS - Demonstrator Use Cases). Both projects have selected one use case on Invasive Alien Species.

4. High Level Roadmap towards Interoperability

This section highlights some of the points of attention on the road towards future interoperability. First the challenge of alignment scalability, second some challenges regarding balancing generic versus specificity.

4.1. The Challenge of Alignment Scalability

For this initial interoperability study, we have focused on a limited number of projects and pilots. In the future, we foresee that more interoperability pilots could be run and also we observe that the digital twin landscape in the context of the Green Deal is steadily expanding. This requires a scalable approach to develop these interfaces. This approach has both a technical component, making sure that digital twin services are small and well-defined (with a specific purpose), as well as an organisational element. Here, the focus should be to guarantee that, with the addition of new development teams working on specific pilots, the amount of time that teams spend on alignment between them does not grow. This requires establishing a product management board that sets priorities and orchestrates product development. In addition, clear guidance to implementation teams should be provided to make sure they deliver a shared architectural view and product vision. Periodically digital twin development teams should be brought together to discuss mutual dependencies and product evolution vision. This allows teams to work independently for periods of time without resulting in significant alignment efforts when changes are merged back into the main interoperability product.

4.2. Balancing Generic versus Specificity

Supporting the interoperation of digital twins as well as digital twin systems, should happen at multiple levels. These range from the level of model design, data and metadata standards, to the definition of data management service interfaces and service quality. Accordingly, different levels of technical specificity are required and useful, i.e. trade-offs need to be considered, for which co-design approaches are known to be helpful. Two generic co-design use cases can be considered in this context, where different types of interoperation are required. The first concerns the coupling of two digital twins, which can be one-directional, i.e. one digital twin produces data that is used as input by another digital twin, or even bi-directional. To ensure the sustainable operation of those digital twins, which depend on data from another digital twin, a precise definition of (meta-)data formats, data quality, service quality, and data exchange mechanisms are required. A different approach is needed considering the generic use case of a user, who is interested in combining and analysing data produced by different digital twins. Given the diversity of users, and the expected and targeted growth and expansion of the user communities, a more agile approach is required based on generic design principles.