©EUMESAT

2023

Synthesis paper

Digital Twins of Planet Earth

Second draft

HELMHOLTZ Earth and Environment

SynCom HELMHOLTZ

SynCom
HELMHOLTZ

Digital Twins of Planet Earth

Sabine Attinger, Thomas Jung, Martin Visbeck, Franz Ossing

Peter Braesicke, Jean Braun, Sascha Brune, Mauro Cacace, Holger Gohlke, David Greenberg, Judith Hauck, Harrie-Jan Hendricks-Franssen, Corinna Hoose, Marion Jegen, Patrick Jöckel, Joana Kollert, Stefan Kollet, Eric von Lieres, Klas Ove Möller, Annemarie Müller, Magdalena Scheck-Wenderoth, Martin Schultz, Beate Slaby, Joanna Staneva, Claas Teichmann, Stephan Thober, Jan Vanderborght, Harry Vereecken, Sergiy Vorogushyn, Wolfgang Wiechert

Second Draft

Draft for the General Assembly

The term "Anthropocene" refers to a geological epoch shaped by human activities, resulting in an all-encompassing sustainability challenge that has impacted all parts of the Earth system. At the same time, there has been a significant increase in our knowledge and understanding of the planet in recent decades. Moreover, through advances in digital information technology we can begin to combine Earth system data and models to generate highly accurate digital representations of the real planet Earth, a digital twin. Digital twins (DTs) can be used to explore various human development scenarios by varying different factors, foreshadowing anticipated changes, and evaluating their consequences. This can provide essential information as a decision-support system to inform future developments, ensuring human prosperity while minimizing the negative impact of human intervention in the natural environment. The objective is to create a family of applied digital twins that utilize diverse models, Earth system datasets, prediction systems, and forecasts, which can serve as a tool for science, decision-makers, and public engagement.

What are Digital Twins?

A digital twin (DT) is a virtual representation of a real process or object with a two way connection between the two. This means that changes in the real-world object can be incorporated into a DT in near real-time, and varying boundary/initial conditions or parameters of the DT can be used to simulate potential 'what-if' scenarios of the real object.

The use of a digital twin concept has a long tradition in industrial engineering and design to optimize desirable outcomes. The rapid development of digital technology, with the exponential increase in the performance of processors and electronic memories, is now enabling us to also build digital twins of the highly complex Earth system full of interconnected processes arose, providing an effective decision-making tool to optimize human interventions.

Digital Twins in Earth System Research

Planet Earth can be divided into five natural components: geosphere, atmosphere, hydrosphere, cryosphere, and biosphere. In the course of an extremely short time in Earth history of just a few hundred years, collective human activity has become such a significant force with noticeable consequences such as climate change, loss of biodiversity, pollution, and habitat destruction that we speak of a new subsystem – the anthroposphere (Fig. 1).

Not only climate: Planet Earth in the Anthropocene

For the first time in human history, human interventions in the natural world have led to globally noticeable impacts. Humans themselves have become a geofactor and have noticeably changed all of the natural subsystems of our planet. The United Nations have formulated 17 "Sustainable Development Goals" to address this globally threatening development. The goal is to recognise the human-induced changes, to understand their effects, to grasp the resulting threats, and to take countermeasures. The Earth system sciences will provide critical scientific insights to form sustainable options for the coming decades.

All of these subsystems are closely interrelated, and each of them is nonlinear, dynamic, interactive, and full of feedbacks, both internal and external. As a result, our Earth system as we know it now is a highly complex structure.

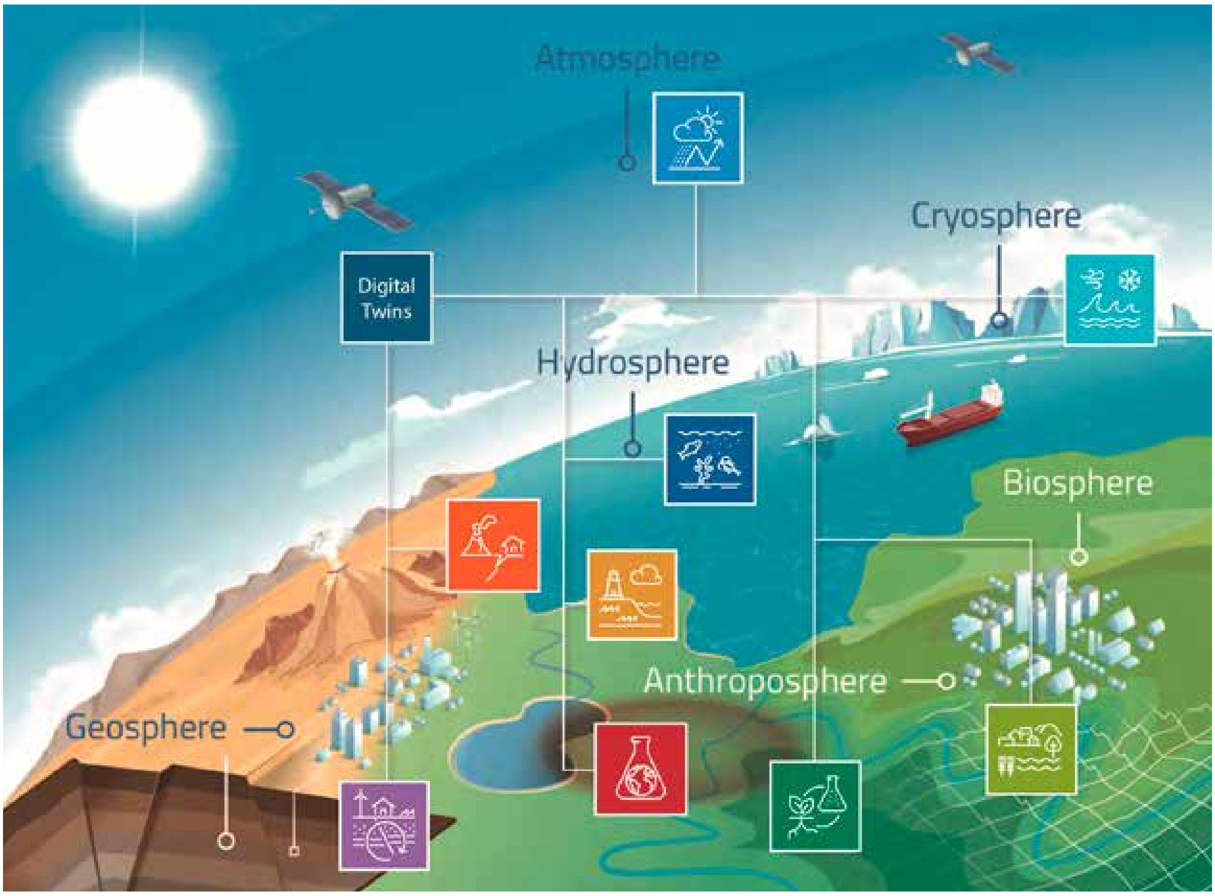


Fig.1: The research program of the Helmholtz Research Field Earth & Environment with its nine Topics (icons) reflects the complexity of the Earth system with its six subsystems and their interconnectedness. (Graphics: Helmholtz Research Field Earth & Environment) ©UFZ / dieaktivisten

Draft for the General Assembly

"What if...?" A tool for knowledge-based decisions

DTs are an ideal tool to explore the consequences of interventions by providing interfaces to data (from observations and models, including on demand simulations and diagnostics) to guide real world decisions and to explore their potential impact immediately. They enable the answering of "what if..." questions. In an iterative process users can explore options and optimise them ahead of implementing an intervention by exploring potential impacts in near real time. In a continuous loop, the DT is updated, has a potential to grow and thus helps decision-makers to monitor the interaction between natural phenomena and human activities.

For certain DT applications as a planning and decision-making tool, it might be crucial to be able to make real-time or near real-time predictions about the future state of the system with high reliability. Thus, to be able to simulate the future under different "what-if" scenarios, a DT must be well tested and validated against observations and show (in this process) strong predictive capabilities.

Draft for the General Assembly

Crucial: the architecture

Operationalised DTs are first and foremost a tool for decision-makers and, in addition, provide opportunities for communication and engagement with many audiences. The spectrum of potential end-users ranges from law makers and public authorities to regional administrations and national governments in the public sector. In the private sector it might range from those who use natural resources to those who offer knowledge services to enhance Earth system literacy and public awareness. Each group expects a user-friendly interface tailored to their needs, or even a supportive environment to build their own twin. Thus, a flexible and interoperable twin architecture is essential. Data systems need to be accessible, prediction systems interoperable, and user interfaces standardized. The center piece of DTs is the so called "data lake": a place where all the data and information become accessible. These data must obey the FAIR and CARE principles and should be as openly available as possible.

The user interface is an important factor of DTs (regardless of the modules that form the DT) and therefore needs to be user-friendly. Here, interventions can be controlled, and the consequences of different choices can be explored. However, limitations of DTs are set by the laws of physics, process understanding, current parameterisations, and the available data and computational resources.

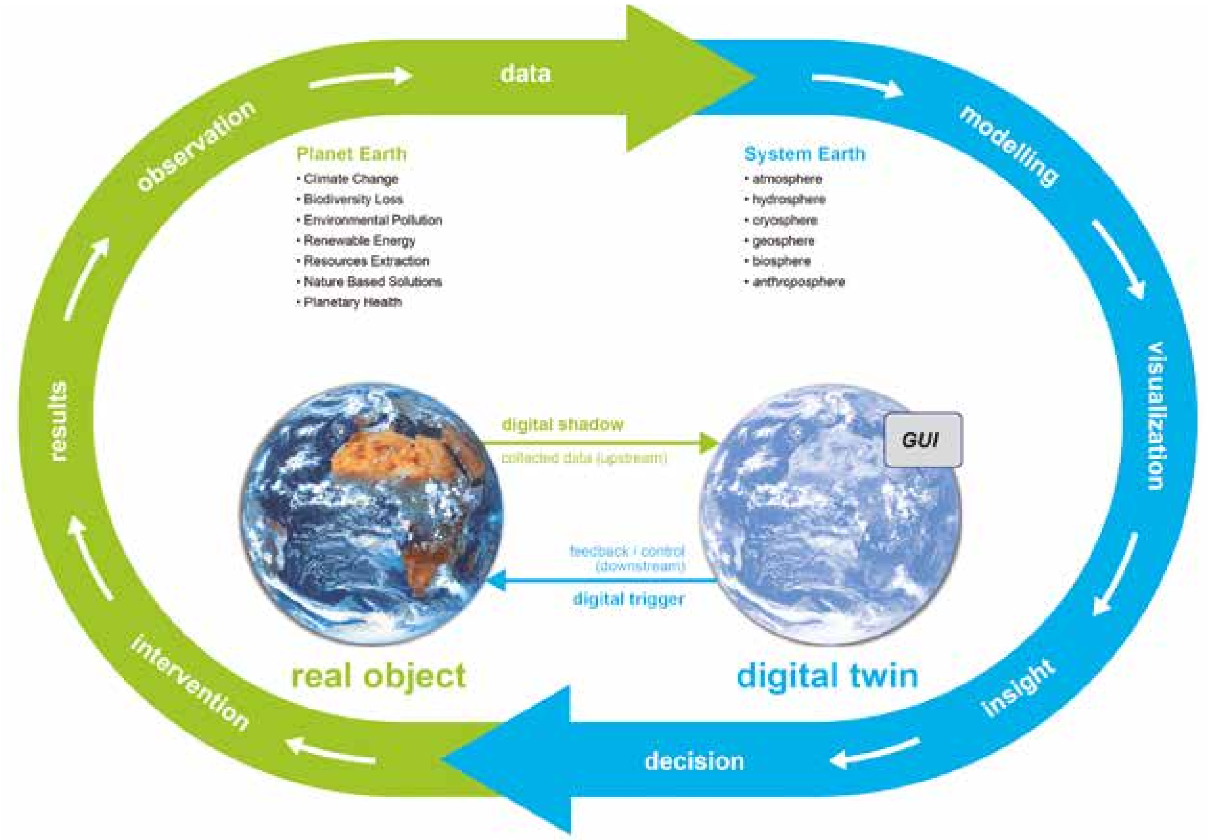
GUI: the Digital Twin User Interface

The knowledge that a DT generates must be made available to users on an easy-to-use interface. Today's Graphical User Interfaces (GUI) provide such a tool. Decision-makers who use a DT assume that the knowledge contained within the DT is scientifically validated. At the user interface, they can explore their decision parameters and derive decisions from them. Therefore, the biggest challenge in interacting and visualizing a DT is to provide the necessary data through appropriate visualizations.

How does a digital twin work?

The concept of the DT thus has four basic components: The Earth system observing system, an Earth system prediction capability, a central data hub, and the user interface GUI as an intervention instrument (Fig. 2).A real object, e.g. planet Earth, or a real process (e.g. sea ice change) is described by observation, measured data, first principles, theories and hypotheses. The digital representation of this fundamentals is called a "digital shadow". These "ingredients" are shaped to a virtual object, the DT. DT thus reproduces the real object. For very complex systems, the real object can also be represented in connected twin components (DTC).

Fig.2: Concept of intervention driven relation real object and DT The (data) describing the state of the real Planet Earth create a digital shadow forming the virtual object System Earth (DT). The natural processes as well as the impact of humans on the planet are analysed, evaluated, and converted into instructions for "intervention testing" within the DT at the GUI. The most suitable interventions in a decision process (involving stakeholders) as feedback/control (digital trigger) from the virtual System Earth to the real object, almost always in an iterative process. (Graphics: G.Schwalbe/F.Ossing, GFZ, after Leopoldina Nationale Akademie der Wissenschaften (2022): „Zukunftsreport Wissenschaft. Erdsystemwissenschaft - Forschung für eine Erde im Wandel", Halle (Saale), 101 S., https://doi.org/10.26164/leopoldina_03_00590 ; ISBN: 978-3-8047-4255-0)



In a DT, changes can be made with the aim of testing hypotheses by changing control parameters or boundary conditions and other core variables of the digital representation, exploring changes or optimizing planned measures in terms of effectiveness or unintended consequences. A user-friendly GUI supports effective workflows. For some "smaller" systems/objects (e.g., lab testing) a fully connected DT where human interventions are executed as a "digital trigger" downstream from the virtual to the real object/process is possible. From the comparison and analysis of the deviations between the real world and its virtual, DT, the changed processes in the real objects can be checked for their effectiveness and adapted; these changes in turn flow back into the DT to regulate the processes there accordingly.

The difference between a DT and an Earth System forecast system lies in its purpose. Traditional forecast systems try to capture a state of the environment and then use the underlying Earth System dynamics to generate deterministic or probabilistic forecasts. A DT uses a similar engine but explores how a forecast is changing when certain inputs are altered, e.g., representing human interventions. Given the immense amounts of data and information and the interconnectedness of the Earth system's components, automated data processing, preparation and delivery procedures remain significant challenges for both more "traditional" forecast systems and DTs.

Certain classes of DTs, however, will be retionally or temporally restricted, focusing on the highest possible resolution of a particular process or the impacts of a single variable. This reduces certain resource requirements,

Draft for the General Assembly

e.g., amounts of data, and computational resources. Additionally, the complexity of different DTs can vary greatly, from small, compact systems to some of the most complex systems, such as Earth system models at kilometer scales.

Challenges in Big Data

Already large geo-data sets will exponentially grow fed by new types of remote sensing data and in-situ observations. Wireless sensors, fibre optic cables, environmental DNA, and other measurement and observation technologies provide relevant information. Even for exascale-class computing systems like JUPITER, it is an enormous challenge to simulate the entire Earth system. The data volumes are too extensive. With the National Research Data Infrastructure and in particular NFDI4Earth, the geosciences have a good starting point, but for Big Data to be processed, (synthesized) digital infrastructures beyond NFDI4Earth will need to be developed.

Platforms and users

All relevant data need to be made available in the DT world on an easily accessible, user-friendly platform that collects and harmonises it. The greatest challenge is the spatial, temporal, and content-related diversity of the data, spanning from observations, modelled data and other materials such as videos, images to even citizen science data. This diversity initially sets limits to the interoperability of the different data and therefore requires specific preparation of the data before it can be used in a DT. Because the amount of available data is growing much faster than the capacity for fast data transmission, data should be prepared, retrieved, and processed close to the source. Therefore, working with distributed data is becoming a common challenge for many applications. Consequently, international digital standards must be defined and applied to harmonise the data sets and make them available for global DT applications.

Finally, the wealth of information produced by a DT must be made available to the respective stakeholders in a comprehensive and engaging way. Thus, a DT must visualise data and provide access to analytical tools that meet the needs of a diverse user audiences with different goals and interests. The main challenge for interaction and visualisation in a DT is to provide the necessary data from the respective models and/or fsensor networks using appropriate visualisations (and interfaces) to allow users (and other automated systems) to connect to the DT in a meaningful way.

Data standards

One objective of the SynCom Digital Twin Project within the Helmholtz Research Field E&E is the promotion of interconnectivity concepts relating to DTs of the Earth system. It is obvious that appropriate data structures for models and observations are required for this. However, the current state of data sets in Earth system research is characterized by a vast variety of formats, time and space scales, infrastructures, and isolated solutions. The interoperable use of this valuable data requires traceable metadata information. Thus, the Helmholtz Open Science Office promotes open science and FAIR Data concept. Using FAIR principles, characteristics are defined that should include data resources, tools, terminology, and infrastructure to facilitate discovery and reuse by third parties. FAIR stands for Findable, Accessible, Interoperable, Reusable. The FAIR principles apply not only to the data itself, but also to data storage, material and digital infrastructures and services. Data also need to be used ethically and responsibly, and to achieve this, the CARE principles (Collective benefits, Authority to control, Responsibility and Ethics) have been articulated.

These standards are not a scientific end in themselves. Above all, they serve to discover sources of uncertainties or even errors in order to rule out misinterpretations and incorrect instructions in the workflow of the DT.

Draft for the General Assembly

Automated Analysis Strategies and Innovations in AI

Even with high-performance computing infrastructures (HPC /DIC), the huge Earth system data amounts can no longer be managed using traditional data processing methods. Thus, automated analysis strategies are required to handle the data efficiently. On top, novel dynamical analysis methods and machine learning algorithms are a suitable means of finding hidden structures in this data that are not easily recognizable. The extension of Machine Learning (ML) to Deep Learning uses artificial neural networks with which the ML algorithms can be trained. This can help timely responses when complex/surprising patterns are discovered and can be acted upon.

Artificial Intelligence (AI) tools and processes offer further opportunities to uncover hidden structures, connections, and correlations in big data.

The HAICU Helmholtz Artificial Intelligence Cooperation Unit (Helmholtz AI) platform promotes the application of AI in all Helmholtz Research Fields through the development and dissemination of AI methods. Already AI projects here cover a wide geoscientific range, from geothermal energy to ocean research and atmospheric science.

Earth and Environment in the DT

In view of the complexity of our Earthsystem, it makes sense to develop DTs for subsystems. The different subsystem DTs need to be able to integrate "neighboring" data streams - e.g., as boundary conditions or fully interactively. With this in mind, the nine Topics of the Helmholtz Research Field Earth & Environment can provide an integrated overall strategy for ES DTs. The goal is to create a family of applied DT engines including a variety of data from models and observations. But that is only the first step: the long-term vision is to digitally represent the entire Earth system with high temporal and spatial resolution.

The nine Topics of the Helmholtz Research Field Earth & Environment

- The atmosphere in global change
- Oceans and cryosphere in climate change
- Restless Earth
- Coastal transition zones under natural and human pressure
- Landscapes of the future
- Marine and polar life
- Sustainable bioeconomy
- Georesources
- One healthy planet

Draft for the General Assembly

The following examples do not claim to be complete, but show the range of possible applications of DTs.

EU and UN promote digital twins

The most important frame of reference in this context is currently the EU. In its flagship projects Destination Earth (DestinE) and Digital Strategy, the European Commission is funding the development of Digital Twins of the Earth (DTE) in order to better understand the effects of climate change and extreme events, to run through future scenarios and to be able to take measures quickly.

Due to the ocean's essential role in the Earth system, it is clear that there is also a need for digital ocean twins (DTO). The United Nations has put the issue on its agenda with the Decade of Ocean Research for Sustainable Development and the international Digital Twin of the Ocean Program DITTO.

Draft for the General Assembly

a) Digital Twin of the Climate System

A digital climate twin is intended to help decision-makers to explore, analyze and gauge possible development paths, and plan their response to climate change-related impacts and events in their decision-making area. Anthropogenically induced climate change ultimately has impacts down to local scales, so climate adaptation measures start here. However, current climate models do not have the temporal and spatial resolution needed to assess local/regional impacts of the climate crisis and downscaling had to be employed in the past. It is therefore difficult for regional planners and decision-makers to obtain climate information at the fine resoluti-on they need. DT Climate Adaptation will create a climate information system based on Earth system models. These will have spatial resolutions on the kilometer scale and provide the necessary information for adaptation scenarios and measures.

Advanced treatment of physics in models

Because of the immensely increased resolution, these models incorporate less assumptions for treating the physics within them. In previous models, important subscale quantities, like eddies in the ocean, were not explicitly described in physical terms, but were parameterised, i.e. introduced into the model calculation with physically motivated approximations. However, this leads to physical and temporal/spatial structural uncertainties. The now possible modeling approaches, on the other hand, envisages treating some important small-scale processes (e.g. ocean eddies and atmospheric convection) as described by the fundamental equations.

More frequent updates

Climate DT organises the workflow for users in such a way that their planning scenarios can be adjusted much more frequently (increased interactivity) than was the case previously (e.g. as part of the Coupled Model Intercomparison Project, CMIP). This interactivity, in turn, can also be used to tailor additional simulations not foreseen in the CMIP context. In this way, different adaptation scenarios can be run through, always interactively and on the basis of updated data. Ultimately, impact chains can be created and checked down to the regional scale: hydrological events such as floods can be directly linked to precipitation and the hydrological cycle, their emergence can be reconstructed and their impact can be estimated.

DTs thus output a state of the climate system that a user can easily compare with measured observations. The comparison, in turn, results in possibilities for readjustment. The method of data assimilation, which has been used in numerical weather forecasting for decades, is extended decisively here. The user receives a data system in real-time that corresponds to his or her experience and action environment. This in turn enables interactive action and user involvement with the aim of optimising the system.

Data streams and data sets

Due to the sheer volume, some high-resolution data can only be streamed to provide the basis for applications (e.g. hydrology and wind energy). Therefore, the Climate DT of the Destination Earth initiative incorporates the use of streaming. Streaming refers to how nature presents its information as a state that changes over time. Applications must capture, record, and process this streamed state to make sense of the world. By streaming data instead of saving it, the output models are forced to present the simulated world to applications in physical terms, which the applications can then capture, record, and process. This approach greatly expands the amount of data that can be shared, enables demand-driven use of climate information, and allows for interactivity.

Streaming thus means, that data will be available at high resolution, with full information content, for a few days only (buffering). The data will come from different Earth system models (Integrated Forecast System IFS coupled to FESOM and NEMO, as well as Icosahedral Non-hydrostatic Model ICON).

The large amounts of data can no longer be handled with the traditional form of data processing. They are constantly updated and are only available in this high resolution for a certain time in the working memory. At the same time, the large data volume is reduced in size, e.g. by coarsening the space/time grid or by using efficient lossy or lossless compression algorithms (depending on application), and can then be transferred to a long-term archive.

DT Climate is thus being developed into a constantly updated system for decision-makers. Initially, use cases are planned in five climate-relevant areas: energy, hydrology, hydrometeorology, forestry, and urban environment. There is an interface with the UFZ's DT Water.

Production runs will run on the EuroHPC supercomputer LUMI of the Finnish CSC - IT Center for Science, a pre-exascale class computer. The AWI Alfred-Wegener-Institute contributes its expertise in the areas of model development, in particular the next generation sea ice ocean model (FESOM) based on an irregular grid, as well as in the execution and analysis of the simulations and the development of storyline scenarios.

b) Digital Twin of the Ocean

The oceans play a dominant role in the Earth system. The world's oceans cover 70% of the Earth's surface, play a crucial role in the climate system and form the basis of life for billions of people. In view of the threatening man-made changes in the oceans, the United Nations have declared the years from 2021–2030 as the Decade of Ocean Research for Sustainable Development. They have formulated ten concrete challenges to reduce the pollution of the world's oceans; challenge 8 calls for a DT of the ocean to support marine research, marine protection, marine governance and sustainable marine management.

DTOs in the heterogeneous data world

Sustainable use of the oceans requires reliable knowledge. A digital twin of the oceans (DTO) requires data that describes the current state of the ocean with high precision. Although the network of marine observing systems has made great strides in recent years (globally GOOS, IODE and EMODnet for Europe), most of the ocean regions remain under-monitored. In addition, there are large observation gaps for many important ocean variables. Added to this is the great heterogeneity of the data: The marine data are based on an international patchwork of observation and remote-sensing platforms. Successful DTOs will require more rapid progress on making this data landscape clearer and setting up an interoperable system with replicable data structures (metadata), formats and standards as advanced by ODIS under IODE. Building on this, DTO frameworks provide the ability to identify where important data and information is missing; they can provide concrete plans for closing these gaps or information on where better observation methods are needed. There are several well-established data curation activities around the world that address these challenges; furthermore, there are international efforts to share information, such as Blue-Cloud.

Draft for the General Assembly

Draft for the General Assembly

From model to twin

Ocean models have evolved from a largely hydrology/physics-based description of ocean dynamics to models which now include a variety of biochemical and ecosystem-related processes. In this way, they record the connection between oceans as nature, living space, and economic space in an increasingly detailed manner, thus enabling retrospective analyses and forward-looking planning. However, the available data and information is also growing immensely. These circumstances suggest that the complex system ocean cannot be mapped in a single DT, but will have regional and thematic sub-component DTOs.

DTOs not only offer the possibility of processing the enormous amounts of data using HPC /DIC technologies; they can also help track down information and structures that are often not openly recognizable in the data and make them usable. Hybrid approaches of measurement data and AI-supported data processing (machine learning and deep learning) can uncover these relationships and will thus lead to increasingly realistic model results and more reliable scenarios. In this way, administrative and political decision-makers are supported by DTOs with dynamically updated data. Ecosystem-based management, marine spatial planning or the implementation of nature-based solutions and marine protection measures can be continuously checked and optimized.

Ocean models also need to be regionally embedded, so DTOs need to include the knowledge of local actors and the involvement of local communities. DTOs thus also form the interface for co-designing and improving marine observation systems, in particular through the exchange between the developers of the observation networks and the end users. To do this, the DTOs must offer an appropriately designed, practical, user-friendly interface, as currently GUIs offer.

DITTO: A international program to advance digital twins of the ocean

Oceans as natural systems do not stop at national borders. Building a DTO requires coordinating international activities, promoting best practice examples, raising awareness of the use of DTs and demonstrating their potential for decision-making.

Coordinated at GEOMAR, the international UN Decade Project DITTO is dedicated to these tasks. DITTO is developing a common understanding of DTOs and spreading awareness of their benefits and potential applications. This is done through the collaboration of six working groups that address different building blocks of DTs. These include: Ocean observations, modelling, interoperability, visualisations, a digital framework, and education and capacity development. By involving international partners from various sectors (research institutions, private sector, local governments, etc.), the global DTO network is strengthened. The aim of DITTO is to create a digital framework for DTOs that will enable ocean professionals from all sectors worldwide to create their own DTs.

c) Digital Geo-Twin

The EU project DT-GEO (A Digital Twin for GEOphysical extremes) is an integral part of the Destination Earth initiative. Here, the prototype for a DT for geophysical extremes is being developed through collaboration among 18 European institutions. The aim is to analyze and predict the effects of such extreme events. The motivation is that large parts of Europe and the world are threatened by earthquakes, volcanic eruptions, and tsunamis. Extreme geophysical events are the cause of more than 10 % of the damage caused by natural disasters of all kinds worldwide.

DT-GEO's prototype consists of interlinked Digital Twin Components (DTC) that address different types of geohazards: natural and anthropogenically caused earthquakes, volcanoes and tsunamis triggered by earthquakes and landslides.

The German Research Centre for Geosciences (GFZ) is involved in the DT-GEO sub-components Tsunami, Earthquakes and Anthropogenic Earthquakes.

Tsunami

Tsunamis are caused by earthquakes, volcanic eruptions, and landslides. Especially when tsunamis occur near the coast, a fast and effective response is required. In already existing tsunami early warning systems like GI-TEWS, the evaluation process is run partly automatically. The integration of real-time or near-real-time data in a DT, their assimilation, including the dynamic development of earthquake parameters, sea level and GNSS data streams into the running model allows successively more precise mapping of the situation, reduces uncertainties and has an effect on the decision module. In this twin component, the GFZ is responsible for collecting and harmonising the requirements and specifications developed for the tsunami twin. There is an interface to the DT component earthquake, from which data and information on earthquake rupture features and information on ground shaking are provided as input to the tsunami twin.

Earthquake

Earthquakes are very difficult to predict. To mitigate their catastrophic consequences, it is necessary to know where they occur and what magnitude they can develop. The European Seismic Hazard Map displays the hazard in Europe for earthquake-induced ground motion. From this, the earthquake risk is derived, which describes what possible effects earthquakes can have on people, on facilities and utilities, and on nature. Precautionary measures can be taken based on these quantities.

However, DT-GEO-Earthquake is intended to go beyond precautionary measures. Its goal is to better recognize what is happening in the event of an earthquake from the information that is constantly being received and to draw conclusions about the maximum size of aftershocks, for example. In addition to the traditionally recorded physical measurements, large amounts of new data are being added: for example, cellphone apps are already used to record the intensities of an earthquake in the wider area and send them to the authorities. The shake maps and the seismic hazard maps based on them can thus be updated and the deployment of rescue forces better planned.

The GFZ is developing Europe-wide ground motion models (GMMs) for this twin component. These site-specific GMMs are further developed into hybrid GMMs by combining data-adapted GMMs and existing shake map libraries. Empirically collected ground motion data will be combined with physics-based model calculations at a previously unattainable resolution. Hence, now the entire chain of effects – from tectonics and hazard to the damage effect in the event of an earthquake – can be calculated for all kinds of conceivable earthquakes today or in the future.

Anthropogenically induced earthquakes

Some geothermal projects have caused noticeable earthquakes. Therefore, the prediction of the effects of anthropogenic geophysical extremes (Anthropogenic Geophysical Extremes Forecasting AGEF) is of particular interest for geothermal plants. In this DT Geo twin component, the possibility of predicting the remote seismic impact of georeservoirs and the prediction of the delayed geological-physical responses of georeservoirs is investigated. The GFZ models the maximum possible magnitudes of such induced earthquake events. This DT component is currently under research at the Strasbourg geothermal site in France, an important European location for earthquake research.

d) Digital Twin of Terrestrial Systems

Terrestrial systems are the immediate habitat of humans. They are shaped by a multitude of abiotic and biotic processes and interactions. Human interventions in natural systems have caused impairments that are now globally noticeable and have altered material flows, posing a threat to vital ecosystem services. Biodiversity, landscape degradation, water quality – these are critical variables that are exacerbated by and feed back into climate change.

Draft for the General Assembly

Draft for the General Assembly

The management of terrestrial areas, agricultural and forestry production, and sustainable soil health require an appropriate database. The provision and evaluation of such data primarily encounters the problem of an almost unmanageable heterogeneity, which is compounded by a wide range of temporal and spatial scales. Just to record the abiotic material flows of carbon, phosphorus and nitrogen, a comprehensive and, above all, spatially far better resolution is required than is possible with the existing observatories. In addition, there is the diversity of life that makes our planet unique: biotic systems can vary greatly even at the meter scale, plant and animal species are directly dependent on each other, and biotic material cycles are closely interwoven with abiotic material flows.

Terrestrial parametrization: the clever linking of the known and the unknown

Europe, with a land area of 10.5 million km², cannot be covered extensively with terrestrial observatories. Terrestrial parameters such as soil porosity are not known over the entire area and must therefore be skilfully linked with existing area-wide information (from observatories, satellites, field measurements) and then brought to the required model scale. This link can be established by means of suitable statistical methods (e.g. non-linear regression), but in the future also increasingly with the help of AI. Parametrization in terrestrial models is therefore not a description of sub-scale processes, but rather the linking of known with unknown model parameters and the scaling of these parameters to the model scale. The goal here is to arrive at a uniform but seamless parameterisation of the models at kilometer scales throughout Europe - an interface to the resolution of the climate models envisaged in EU-DestinE. At this interface, the terrestrial DT exchanges information with DT Climate.

A twin for water

The link between abiotic and biotic processes is the element of water. Water supply and distribution, precipitation up to extreme events, and changes in the terrestrial water cycle due to climate change are essential decision variables for terrestrial future planning. The National Water Strategy for Germany is a reaction to the change in water supply. Hydrological modelling provides the necessary data, analyzes the current state and helps with planning. In addition to the spatial grid size, such modelling also contains a great deal of temporal variability: extreme events in particular demonstrate the need to provide updated hydrology information in the input and output every hour. Forestry, on the other hand, needs reliable data over annual periods for its planning. And this is exactly the goal in Destination Earth: a water twin will communicate continuously with the weather/climate twin and can thus be used as an analysis and decision-making tool that adequately reflects the different landscape types. Here, there is close cooperation with AWI via DT Climate.

A broad field: biotics

Terrestrial systems are the home of life, biogeochemical processes the interface between biotic and abiotic processes, and thus also between their modelling. Forest, grassland, and agricultural landscapes are inhabited by rather different life forms. Human intervention and climate change have decisively changed this system and led to the current biodiversity crisis.

The twin BioDT contributes to combating the biodiversity crisis by providing the necessary data for specific use cases in a highly precise and interactive manner; the UFZ is involved in this project.

Forests worldwide are under stress. In Germany, too, the drought of recent years has not only caused damage to forestry, but has also threatened the ecosystem. SpaceTwin is a DT that uses high-resolution forest models to investigate how droughts, fires and deforestation affect tropical and boreal forest ecosystems and how they interact with global climate change.

e) Digital Twin for Urban Development

Cities and municipalities are focal points of the sustainability crisis. The necessary transformation of energy supply, mobility, and housing confronts urban planners with complex questions that cannot be solved independently. At the urban planning level, flexible and user-friendly tools are needed that dynamically link climate and environmental sciences with urban requirements. For decision-makers, this also means considering what is required in terms of (natural) science in the regulatory cooperation between municipalities, the state, and the federal government when developing the necessary guidelines.

In the city as a living space, environmental parameters such as thermal comfort, wind comfort, and air quality are important parameters for well-being. The functionality of the city in terms of quality of life is determined by urban infrastructure parameters such as transport, energy and water supply. This complex structure is under increasing pressure from the climate crisis. Urban systems react sensitively to climate change, but urban planning thinks in terms of years, if not decades, especially in the case of major structural reconstruction measures.

Regional climate models on ever higher time/space resolution, such as those used in the international EURO-CORDEX initiative, can provide the necessary data to move from global to regional to urban space. While a climate model grid width of 12 kilometers still applies throughout Europe today, which can be refined regionally to 13 km and scaled to 15 m for individual representative days within cities, the resolution is to be further increased in future at the European level within EURO-CORDEX and globally with DT Climate. In the future, urban planners will have methods and data at their disposal in a resolution that will allow them to run through scenarios of future climate on spatial and temporal scales of urban reality, to test suitable measures virtually, and to analyze their feedback on the overall city system.

Urban DTs reflect the reality of a city

The EU project DestinE explicitly provides for the development of DTs for urban areas. Within the EURO-CORDEX initiative, there will also be an ensemble of regional climate simulations taking into account urban structures on which Urban DTs can build. Ideally, twin components for the urban subsystems of housing, economy, and mobility can be developed here, which as a coupled system interweaves the respective planning variables of development, traffic, roads, planting, and water areas. The FONA project ProPolis is a step in this direction.

The Climate Service Center Germany (GERICS), an institute of the Helmholtz Centre Hereon, is a ProPolis project partner. The further development of the PALM model, a meteorology model for simulating the near-ground atmosphere, into a high-resolution urban climate model (PALM4U) is the goal of ProPolis. The aim is to create a practical, user-friendly urban climate model that is tailored to the needs of municipalities, urban planning, and practical users. Since it is also used in research, the practical application can be fed back into science.

Draft for the General Assembly

Draft for the General Assembly

Data for real-time

The integration of such city models into climate models that resolve parts of the city down to a few meters requires the mutual scale adjustment of space and time (upscaling, downscaling) and a precise definition of the interfaces. Completely new information flows in from urban planning - if possible in real-time - such as data from traffic control systems, energy and water supply. The resulting planning and control possibilities are almost unlimited. For example, they allow a rapid response to extreme weather events. But they also make it possible to extract certain events, such as heat waves, from the larger-scale models and to recalculate them for the urban area for certain critical days in order to determine the effect down to the street scale. Playing out future climate scenarios allows the development and testing of adaptation measures that can, for example, prevent urban heat islands from heating up to health-threatening levels.

Future Urban DTs thus reflect the reality of a city. They can be used by urban authorities and political decision-makers with the help of real-time visualisations. Their range of applications goes far beyond adaptation to climate change. In fact, they encompass all Helmholtz Research Fields on the urban scale.

f) Digital Twins for Sustainable Bioeconomy

Bioeconomy is concerned with the establishment of sustainable material cycles based on renewable raw materials. It makes a significant contribution to combating the causes of climate change and littering of our planet. One focus within Helmholtz is on plant-based food production for a growing world population, another focus is the sustainable replacement of fossil carbon sources. Both areas are closely linked and face specific challenges in the context of climate change and sustainability.

The bioeconomists of the Helmholtz Association are concerned with two-way feedbacks between natural element cycles in the earth system and anthropogenic cycles generated in a bio-based economy. These cycles interact in the soil, which is the basis for plant growth. Understanding the interactions is important to develop new plant production systems under changing climate conditions that will sustain food and materials production. Based on such renewable resources new biotechnological production methods are developed to replace fossil fuels in chemical industry by sustainable plant-based feedstock and, thus, develop circular material cycles.

The acquisition of quantitative data for the biological, technical, economic and societal evaluation of material cycles forms the backbone of bioeconomy research. This requires complex experimental facilities that allow the study of plants and microorganisms under highly controlled conditions. Three experimental facilities at Forschungszentrum Jülich are presented here as examples.


Jülich Biofoundry: Industrial biotechnology is about replacing fossil raw materials for the production of chemicals. A biofoundry is a fully automated facility for the generation of genetically modified microorganisms for chemical substance production and their subsequent quantitative characterization. Here, thousands of mutants (so-called strains) are produced and tested in an assembly line analogous to an industrial production plant. Thus, generic engineering is driven in a fully digitalized Design-Built-Test-Learn (DBTL) cycle. Biofoundries can significantly accelerate the development process for new biotechnological production processes, which normally takes about 10 years.

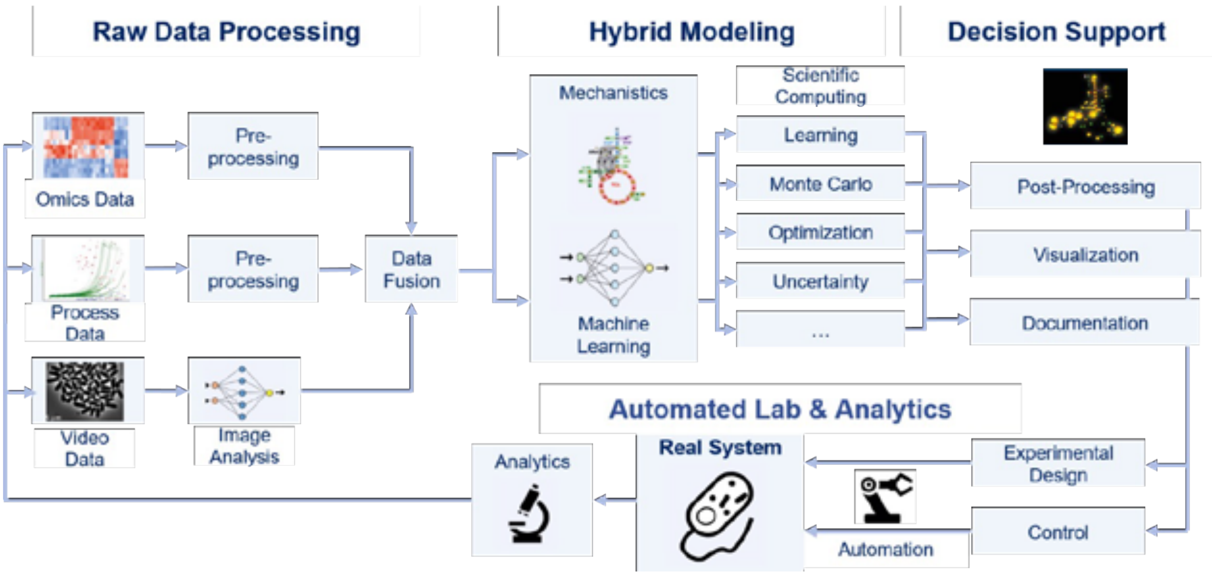
Plant phenotyping: Driven by the increasing interest in quantitative data for breeding research and physiology, and the need for production of plants in engineered systems, extensive automated green house facilities for the automated analysis of structural and functional plant properties with high spatial and temporal resolution have been built up at Jülich. They allow environmental conditions to be gradually or rapidly varied by moving the plant to a different growth zone and recording plant parameters. This includes sensor networks and control systems for determining acute environmental conditions and actuators for spatio-temporal control of input parameters. Digital devices quantify dynamic plant properties in space and time, perform on-line analysis of valuable substances as well as key performance parameters. These technologies and expertise are used to develop and establish DTs of crops that allow to dynamically analyze the spatial and temporal interaction with the biotic and abiotic environment of plants above- and belowground.

AGRASIM, the soil-plant time machine: The changing climate and changing atmospheric CO2 concentrations will influence plant growth, plant water use, and element cycles. In the AGRASIM ecotrons, future scenarios, which are based on model simulations (Climate DT), will be simulated experimentally and the response of crops and fluxes (H2O, C, N) between the soil, plant, and atmosphere, will be monitored. The experimental facility will be mirrored by a numerical simulator, the AGRASIM digital twin. Sensitivity analyses using the DT will be used for experimental design and data obtained from the ecotrons will provide information to improve the parameterization and model conceptualization of the DT. The insights obtained from this twinned experiment will subsequently be implemented in soil-plant-atmosphere modules of Earth System models.

All these largely automated facilities have in common that complex heterogeneous data sets are generated in periods ranging from a of a few days to several years. Experiments must be planned carefully and data must be evaluated promptly in order to be able to plan the next experiments with respect to a maximum information yield. In many areas of the bioeconomy, data evaluation has become the real bottleneck for information acquisition and development processes.

Draft for the General Assembly

DTs are the key to the efficient use of our experimental facilities. The biofoundry, greenhouses and agricultural simulator were developed from the outset together with a corresponding digital infrastructure that enables central data management, brings together all measurement data, orchestrates evaluation algorithms in complex workflows and finally analyzes all data in the context of overarching system models. In this way, the wealth of information is condensed to the essentials with the help of suitable data visualization tools, so that timely human decisions are possible. Ideally, the system will make decisions on its own with the help of AI methods and modern experimental design algorithms. This vision of fully automated experimentation, in which humans only have a monitoring function, is being consistently pursued, especially in the Jülich Biofoundry (see figure). Here, the closed loop between system and digital image, i.e. the bidirectional coupling, is already realized.



At Forschungszentrum Jülich, the institutes IBG-1 (Biotechnology), IBG-2 (Plant Research), IBG-3 (Agricultural Sciences), IBG-4 (Bioinformatics) and IAS-8 (Data Analytics and Machine Learning) have joined forces in the Simulation and Data Lab "Digital Bioeconomy" to work systematically on the development of multi-scale models for the bioeconomy and specifically on DTs. From a technological point of view, DTs represent a synthesis of data-based (especially machine learning) and model-based (relying on known physics) description methods. The development of hybrid data and physics driven approaches is a main research area of modern data science. The Jülich IBG institutes are also involved in the Helmholtz Graduate School HDS-LEE (Helmholtz School for Data Science in Life, Earth and Energy), where currently more than 60 PhD students are developing new concepts, methods and applications. DTs will be a focus topic of HDS-LEE in the coming years.

Draft for the General Assembly

Outlook: Future tasks for Digital Twins

The sustainability crisis, which is becoming more and more evident, touches all compartments of the entire planet Earth in measurable size. DTs have an enormous potential to support data-informed decision-making to address our environmental problems and explore optimized pathways towards a safe and sustainable way of living and interactions with our planet. International environmental goals such as the UN SDGs, the European Green Deal and the findings of UN panels (IPCC, WOA, IPBES, Ozone Assessment) require major efforts in global data collection, analysis, and sharing. This requires the development of a decentralised, international knowledge platform that enables multi-stakeholder collaboration and partnerships, best practice sharing and policy advice.

DTs and their associated Data Lakes are particularly suited for this purpose. They enable decisions based on simulations of future "what-if" scenarios. They can learn from and respond to system developments and reactions in the real world. They will be able to connect the global patchwork of heterogeneous and autonomous data infrastructures and can capture the complex interactions of both natural and social phenomena. However, much remains to be done before DTs become a wide-spread reality.

Academia and research, in collaboration with governments, the private sector and civil society should improve modelling techniques and increase data optimisation and interoperability. Policy actors are called upon to recognize the enormous potential of DTs at an early stage and to promote their use for appropriate policies. The DT approach creates an opportunity for us to reflect on our role in the natural system and to develop action plans and counterstrategies.

Draft for the General Assembly

Abbreviations and acronyms

AI	Artificial Intelligence
AGEF	Anthropogenic Geophysical Extremes Forecasting
AWI	Alfred Wegener Institute for Polar and Marine Research
CARE	Data standard: Collective benefit, Authoritytocontrol, Responsability, Ethics
CMIP	Coupled Model Intercomparison Project
BioDT	Bioiodiversity Digital Twin prototype
DestinE	EU-Initiative Destination Earth
DITTO	Digital Twin of the Ocean
DT	Digital Twin
DTC	Digital Twin Component
DTE	DigitalTwin Earth
DT-GEO	Digital Twin for GEOphysical extremes
DTO	Digital Twin Ocean
E&E	Helmholtz Research Field Earth & Environment
EMODnet	European Marine Observation and Data Network
EuroHPC	European High Performance Computing Joint Undertaking
FAIR	Data standard: Findable, Accessable, Interoperable, Reusable
FESOM	Finite-Element/volumE Sea ice-Ocean Model
GEOMAR	GEOMAR Helmholtz Centre for Ocean Research Kiel
GERICS	Climate Service Center Germany
GFZ	Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences
GITEWS	German-Indonesian Tsunami Early Warning System
GMM	Ground Motion Model
GOOS	Global Ocean Observing System
GUI	Graphical User Interface
HAICU	Helmholtz Artificial Intelligence Cooperation Unit
Hereon	Helmholtz-Zentrum Hereon
HPC	High Performance Computing
ICON	Icosahedral Non-hydrostatic Model
IFS	Integrated Forecast System
IODE	International Oceanographic Data and Information Exchange
JUPITER	Joint Undertaking Pioneer for Innovative and Transformative Exascale Research
LUMI	Large Unified Modern Infrastructure
NFDI4Earth	Nationale Forschungsdaten-Infrastruktur für Erdsystemwissenschaften
ODIS	Ocean Data and Information System
PALM	PALM model system for atmospheric and oceanic boundary-layer flows
RF	Helmholtz Research Fields
UFZ	Helmholtz-Centre for Environmental Research

Draft for the General Assembly

Suggested citation of the publication:

Franz Ossing (2023): Digital Twins of Planet Earth, with contributions by Attinger, S., Jung, T., Visbeck, M., Brune, S., Cotton, F., Teichmann, C., Potsdam: GFZ German Research Centre for Geosciences.

https://doi.org/...

This work is licensed under a Creative Commons Attribution 4.0 International License.
(CC BY-NC 4.0) https://creativecommons.org/licenses/by-nc/4.0/

Imprint

SynCom, Helmholtz Research Field Earth and Environment

Helmholtz Centre Potsdam

GFZ German Research Centre
for Geosciences

Telegrafenberg
D-14473 Potsdam

Published in Potsdam, Germany
2023
DOI: https://doi.org/...

Draft for the General Assembly