

# Chapter 2

## The Digital Earth Project: Focus and Agenda



**Roland Ruhnke, Diana Rechid, Doris Dransch, Laurens M. Bouwer, Holger Brix, Peter Dietrich, Stephan Frickenhaus, Jens Greinert, Daniela Henkel, Andreas Petzold, and Wolfgang zu Castell**

**Abstract** Digital Earth is a project funded by the German Helmholtz Association with all centers of the research field Earth and Environment involved. The main goal of the Digital Earth project is to develop and bundle data science methods in extendable and maintainable scientific workflows that enable natural scientists in collaboration with data scientists to achieve a deeper understanding of the Earth system. This has been achieved by developing solutions for data analysis and exploration with visual and computational approaches with data obtained in a SMART monitoring approach and modeling studies, accompanied by a continuous evaluation of the collaboration processes. In this chapter, the history, setup, and focus of the Digital Earth project are described.

---

R. Ruhnke (✉)

Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany  
e-mail: [roland.ruhnke@kit.edu](mailto:roland.ruhnke@kit.edu)

D. Rechid · L. M. Bouwer

Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Hamburg, Germany

D. Dransch

Helmholtz Centre Potsdam—GFZ German Research Centre for Geosciences, Potsdam, Germany

H. Brix

Helmholtz-Zentrum Hereon, Geesthacht, Germany

P. Dietrich

Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany

S. Frickenhaus

Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Bremerhaven, Germany

J. Greinert · D. Henkel

GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

A. Petzold

Forschungszentrum Jülich GmbH, Jülich, Germany

W. zu Castell

Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

© The Author(s) 2022

L. M. Bouwer et al. (eds.), *Integrating Data Science and Earth Science*,  
SpringerBriefs in Earth System Sciences,  
[https://doi.org/10.1007/978-3-030-99546-1\\_2](https://doi.org/10.1007/978-3-030-99546-1_2)

**Keywords** Data analysis · Data exploration · Monitoring · Collaboration · Interdisciplinary · Workflow · Earth science · Data science · Digitalization

## 2.1 History of the Project

The Digital Earth project was initiated by the Helmholtz Association (see Box 2.1) in advance of the joint research program 2021–2028, to take some of the ideas and challenges described in Chapter 1 regarding data science, digitalization, and Earth System Science. The vision of the project was to foster interdisciplinary collaboration and to identify and adapt in a strongly interrelated approach methods, workflows, and applications that are true “game-changers” for studying the Earth system.

To achieve this, the eight Helmholtz Centers in the research field Earth and Environment (Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research Bremerhaven (AWI), Forschungszentrum Jülich (FZJ), GEOMAR Helmholtz Centre for Ocean Research Kiel, Helmholtz Centre Potsdam German Research Centre for Geosciences (GFZ), Helmholtz-Zentrum Hereon, Karlsruhe Institute of Technology (KIT), Helmholtz-Zentrum München German Research Center for Environmental Health (HMGU), and Helmholtz Centre for Environmental Research (UFZ)) were asked in 2016 to develop a joint proposal “Digital Earth” as part of a call on future research topics within the Helmholtz Association Initiative and Networking Fund.

Already during the definition phase, it became obvious that focusing on a common direction for a joint proposal was challenging: the involved scientists of each center had different disciplinary backgrounds, expectations, and views on such a project. While some were able to contribute precise geoscientific research questions related to observation or model data from their discipline and their institutional background, others were interested in contributing methods of data science or software engineering for data exploration, and again, others were contributing a perspective on developing data infrastructures for data-intensive science within the project. Consequently, the heterogeneous interests and possible engagements were not fully harmonized toward an effective proposal development, but they rather had to be integrated into a common development and a common understanding of the project goals. To achieve this, a cross-disciplinary and cross-compartment approach was identified as useful, allowing for a holistic view on the coupled compartments of the Earth system. The strength of such an approach is the diversity of perspectives resulting in the challenge to create frameworks for fruitful long-term collaboration within the project.

### Box 2.1: The Helmholtz Association at a Glance

- Germany’s largest research organization.
- Named after Hermann von Helmholtz (1821–1894), one of the last great scientific generalists.

- Annual budget of ~ € 4.5 billion.
- ~ 39,000 employees.
- World-class science infrastructure.
- 18 independent research centers all over Germany.
- Research plays a key role in identifying reliable answers that benefit society, science, and the economy.
- Six fields of research focus on the major societal challenges of our time—such as the digital revolution, climate change, energy transition, transport in the future, and the battle against severe and widespread diseases and work on developing sustainable solutions for the future. In doing so, Helmholtz covers the entire spectrum from basic to application-oriented research while applying an interdisciplinary approach.
- The Helmholtz Association cooperates with leading research institutions at the national and international level and is committed to the highest standards of talent management at all levels and the promotion of early-career researchers.
- New knowledge can only benefit society and the economy if it is transferred and therefore made usable. For this reason, transferring knowledge and technology and promoting innovation are of extraordinary importance to us.
- 400 new patents are filed every year.
- Approximately 20 new high-tech spin-offs per year.

With the knowledge of the aspects of digitalization described in Chapter 1 and especially the challenges associated with it in mind, the discussion of redefining the aims in favor of reusable, framework-based concepts, the potential of artificial intelligence and advanced visualization methods would play an important role in the Digital Earth project. In addition, it appeared natural to manifest the goal of reusability and sustainability to support the networking process and long-term collaboration explicitly in a dedicated task, since the long-term perspective of further collaboration within the joint research program 2021–2027 was given immanently in the research field Earth and Environment (see Box 2.2).

**Box 2.2: Helmholtz Research Program “Changing Earth – Sustaining Our Future”**

Climate change, the extinction of species, environmental pollution, and the increasing vulnerability of a technological society to natural disasters are among the greatest challenges of our time. We take a systemic approach to researching our natural environment—from the land surface and the oceans to the most remote polar regions. After all, it will only be possible to plot a course

into a sustainable future with in-depth knowledge of the Earth system, innovative technologies, strategic solutions, and evidence-based recommendations for policymakers.

Seven Helmholtz Centers are collaborating to gather deep insights into the complex relationships between the processes that take place on our planet. What are the causes and effects of global environmental changes? How can natural resources be used sustainably? How can we protect ourselves more effectively from disasters and natural hazards like droughts, heavy rainfall, storms, floods, and earthquakes? We aim to develop solutions and strategies to help humankind adapt to changing environmental conditions, to minimize global threats like climate change, and to understand the potential impact of these risks—not only for the environment but also for the economy and society.

## 2.2 Focus of Digital Earth

The Digital Earth project addresses the challenge of digital transformation in Earth science. The central goal of the project is to enable Earth scientists to (a) develop methods to link data across compartmental boundaries across spatial and temporal scales; (b) establish coherent data flows and analysis workflows; and (c) develop approaches to guide data acquisition in the field by linking various field and model data. The central question of the project is: How can data science contribute to the goals and improve scientific results? This is the fundamental question asked by the natural scientists toward data science.

Therefore, the Digital Earth project is not directed to develop entirely new data science methods and technologies such as new machine learning algorithms or visualization techniques. The innovative aspect is to link natural science and data science and to develop cross-boundary approaches focusing on three main areas, as they are essential: (i) data analysis and exploration; (ii) data collection and monitoring; and (iii) collaborative interdisciplinary working, which is of special importance for the digital transformation.

Developing, advancing, and adopting means that enable this vision are the tasks of Digital Earth to transfer knowledge and close gaps between the two disciplines of Earth science and data science. Here, two scientific ambitions merge, one from the Earth science community that wants to have data science approaches available for their investigations, and the other ambition from data scientists that wants to advance data science methods in itself, but also to make them more easily adaptable to specific scientific requirements. A dialogue is required, and a long-term and sustainable cooperation and problem-solving culture need to be established, in which questions can be put forward and iteratively worked on, solutions get tested and finally adopted. This approach, which is tailored to the needs of the Earth scientists, includes faster

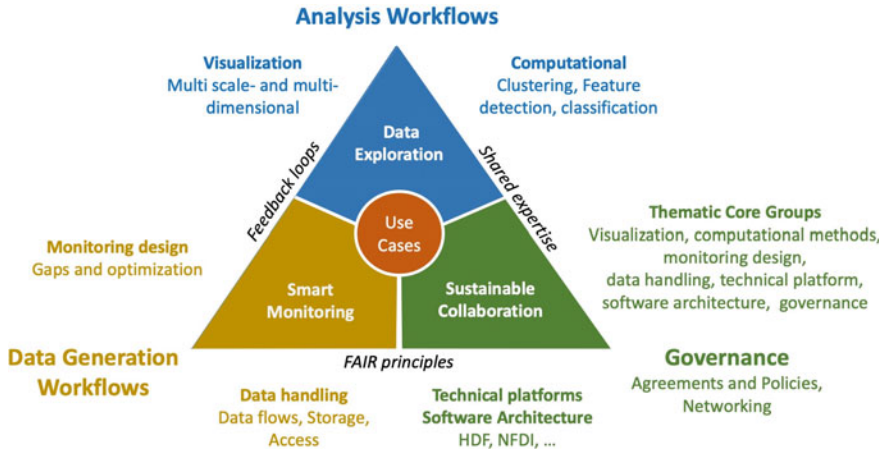


Fig. 2.1 Three cornerstones of the Digital Earth project

and easier-to-use applications, the development/promotion of best methodologies, the adjustment and extension of existing applications, and the implementation of automatization.

Within the focus triangle of Digital Earth (Fig. 2.1), we address several issues: the reuse of data and methods/tools by a broad scientific community reaching far beyond the researchers directly involved in the generation of data and methods/tools, FAIR principles (Findable, Accessible, Interoperable, and Re-usable), quality assessment, visual and computational data exploration, interpolation and integration of data from in situ measures and simulation models, and scientific workflows.

Within several showcases, we adapted and enhanced several data science methods to address challenges we have to face in the investigation of the System Earth. We address challenges related to the following three topics:

- Data analysis and exploration;
- Data collection and monitoring;
- Collaborative interdisciplinary working.

These three topics and their challenges are discussed in the next sections, and the separate chapters in this book that are dedicated to solutions developed in Digital Earth for addressing these challenges are introduced.

### 2.2.1 Data Analysis and Exploration

Challenges for data analysis and exploration addressed with **visual approaches** (see Chapter 3): The incessant processes shaping our Earth’s environment are determined by an interplay of diverse phenomena of physical, chemical, and biological nature,

with ranges of action that span from planetary scales into the microscopic realm. As such, the study of the geoscientific processes and the interplay of determining phenomena rely on the analysis of highly diverse kinds of data from many sources. The following challenges arise for analysis and exploration:

- **The need to establish connections between causes and consequences of geoscientific processes:** The connections become evident only when observations from various disciplines and sources are brought into the relationship.
- **The need to retain a sense of spatial and temporal coherence across different scales:** The sense of spatial and temporal coherence is easily lost, when simultaneously regarding information at different scales such as sediment samples that encode information at cm scale, while remote sensing data does so in km scale. In the temporal dimension, an underwater sediment plume can arise and settle in a matter of minutes, while global climatic phenomena are compared with each other across decades.
- **The need for suitable means to integrate a variety of heterogeneous spatio-temporal datasets:** Scientists have to be supported in creating a “holistic view” on processes and related phenomena.

Digital Earth addresses these challenges with visualization. A main advantage of visualization is its ability to parallel display data even if the data is heterogeneous in scale, variables or accuracy. We applied various visualization techniques and environments and adopted them to our geoscientific requirements. The outcomes are tools for interactive data exploration based on (a) multiple linked view techniques; (b) web-based technologies for real-time exploration of data across spatial and temporal scales; and (c) immersive visualization.

Challenges for data analysis and exploration addressed with computational approaches (see Chapter 4): artificial intelligence and machine learning methods are increasingly applied in Earth system research, for improving data analysis, and model performance, and eventually system understanding. Digital Earth focuses on:

- **The need to extract relevant information/features using machine learning:** For various observational features, no labeled data collections exist. Such labels are, however, important to classify specific observations using prior knowledge, for instance. Using sparse datasets and machine learning methods, alternative ways were found to broaden data availability and derive new, crucial information from existing data. We used examples to map river levees in Germany for which no consistent data was available, and for locating ammunition on the sea-bed.
- **The need to approximate complex processes with machine learning:** Some processes in the Earth system are too complex or computationally costly for large-scale or multiple simulations in models. Here, machine learning alternatives can replace some of these (partly unknown) processes. We present applications for atmospheric methane and ethane concentrations through a neural network and for combining highly heterogeneous data to simulate relations between extreme temperatures and health outcomes.

- **The need of point-to-space extrapolation:** For many applications in Earth systems research, the extrapolation of point to space and local measurements to regional or global fluxes are essential. We employ different computational approaches, for analysis and processing of point observations of methane emissions in order to be comparable with global atmospheric emissions as observed/estimated in global databases, and the functionalities of advanced approaches for point-to-space extrapolation.
- **The need of anomaly and event detection across heterogeneous datasets:** Events and anomalies are important to detect in Earth systems for scientific and practical applications. The huge amount of data and the associated heterogeneity, requires analytic approaches that automate data analysis and still provides relevant results. We present two approaches to detect and understand events in coastal and river waters that are based on this principle: one to assess the similarity of river flood events using multiple atmospheric, hydrologic, and other variables and another that combines observational and model data to detect river plumes at sea at the end of a riverine flood event chain and tracks their spatial and temporal extent.

Challenges for data analysis and exploration addressed with **scientific workflows** (Chapter 5). The challenges include:

- **The need for enhanced work environments that integrate methods and tools into seamless data analysis chains** which allow scientists to comprehensively analyze and explore heterogeneous, distributed datasets. Currently, scientific data analysis is often characterized by performing the analytical tasks in single isolated steps with several isolated tools. This isolated work environment hinders scientists to extensively exploit and analyze the available data.
- **The need for sharing and reuse of analytical methods and tools:** Scientific data analysis and exploration often require specific, highly tailored methods and tools; many of them are developed by geoscientists themselves. Often the methods and tools can hardly be shared and re-used since they lack state-of-the-art computer science methods. The analysis methods and tools are not available for others and have to be invented again and again.
- **The need to exploit data across the various scientific disciplines in Earth System Science:** To answer complex scientific questions, data from various sources has to be integrated, but also the data analysis approaches itself that extract information from the data have to be integrated across disciplines. Integration is necessary on two levels: integration on the technical executable level, but also on the conceptual scientific level.
- **The need to transform science into digital science:** The transformation of science into digital science has been an ongoing process for many years. Suitable means are required to facilitate this transformation and to support collaboration of computer- and geo-experts.

Digital Earth applied the concepts of scientific workflows and component-based software engineering to address these challenges and needs. We adapted and proved

the concepts in our geoscientific work environment and assessed how the approaches can tackle the challenges. Within the showcase “cross-disciplinary investigation of flood events,” we developed (a) several data analysis workflows on a conceptual and digital level and (b) the component-based Data Analytics Software Framework (DASF). The outcome is the Digital Earth Flood Event Explorer that allows investigating floods from several perspectives and that exemplarily shows how scientific workflows and component-based software engineering can improve scientific data analysis.

### 2.2.2 *Data Collection and Monitoring*

Challenges for data collection and monitoring addressed with **SMART monitoring approaches** (Chapter 6). The challenges include:

- **The need for SMART Sensors:** Advancing and developing sensors that have real-time data (pre)processing capacities and are linked in a self-organizing sensor network is still a challenging technological task. Automated event detection, drift correction, and failure detection are possible but still rarely done. Real-time data connections and centralized visualization and analyses are more and more established, but the real challenge is that such SMART sensors and sensor networks become easy to use and the standard way of acquiring multiparameter data in the field.
- **The need for a SMART DataFlow:** An easy to use, scalable and adaptable way of receiving data from sensors and re-distributing them through various channels and means also in real time is the challenge for an efficient SMART Monitoring DataFlow. Standardized and largely automated procedures are needed to obtain reliable data. As an essential part of the live cycle of data is the DataFlow crucial for acquiring high-quality data at the right time and location.
- **The need for SMART MetaData:** Columns of numbers of a time series alone are not useful without the context these numbers have been generated. The suitable description of data is a prerequisite for any secondary use of data. Apart from FAIR descriptions, the data trustworthiness also needs to be assessed and described to allow a correct evaluation of the data. Compiling this data in a complete manner and raising the awareness again, that MetaData are crucial for the correct use of data, is the real challenge for SMART MetaData.
- **The need for SMART Sampling:** Objectively finding the best possible sample location in space and time (most informative information for the respective research question), ideally in an automated and adapting way, is a challenging task. SMART sampling strategies are supporting this challenge. Applying state-of-the-art statistical and AI methods jointly with interactive visualization and analyses is increasing in the community. The challenge is to spread the knowledge about these methods and present easy ways of using them to lower the hurdle of their application.



Addressing these challenges was the main objective of the SMART Monitoring efforts within the Digital Earth project. The involved research centers started, iterated, and further developed the idea of an expanded SMART Monitoring Concept that finally integrates four conceptual groups of tools, each tackling one of the above-stated challenges.

### ***2.2.3 Collaborative Interdisciplinary Working***

Collaboration is essential for the success of the Digital Earth endeavor. Collaboration has to be managed on several levels: between various Earth science disciplines, between data science and Earth science, and between the involved research centers. We identified the following crucial issues in the project that we had to find solutions for:

- Establish topical working groups to shape a framework for collaboration across disciplines. For this, we defined two showcases: (a) the analysis of flood events at the Elbe River along the process cascade event generation, evolution and impact across atmosphere, and terrestrial and marine disciplines; and (b) quantification of methane emission fluxes into the atmosphere from gas exploration in the North Sea.
- Establish and implement digital collaboration platforms for information management and exchange. Mainly, we applied confluence for information sharing and GitLab for collaborative software development.
- Promote existing or upcoming infrastructures, agreements, and policies such as standards, licenses, or eScience infrastructures.

Chapter 7 presents a social science-oriented evaluation in which a World Cafe and a survey were used to evaluate the interdisciplinary collaboration and opportunities for improvement.

As Digital Earth is a pilot project, all process steps in collaboration, scientific workflow setup, method and tool development and hence scientific progress have been evaluated regularly throughout the project period using different measures (see Chapter 8). These evaluations during the project lifetime improved the process steps and produced an added value for the investigation of the Earth system and interdisciplinary collaboration.

To summarize, Digital Earth is designed as a pilot project that integrates data science methods, such as machine learning or visual data exploration into Earth and environmental science, and thus expands and enhances traditional analytical procedures. Digital Earth advances data science with the concrete application field in Earth science. Research in data science is necessary to tailor and enhance existing methods to the specific requirements resulting from Earth sciences. Furthermore, Digital Earth is a kind of a socio-cultural and organizational pilot project on collaboration between institutions and disciplines with a continuous evaluation of the progress.

In order to make the results of the described main topics of Digital Earth known to the communities of Earth sciences as well as data sciences, we have compiled them in this book. The aim of the book is to present the methods and solutions for overcoming the challenges of the three main topics in a compact way. In the following chapters, the book deals with the visual approaches (Chapter 3), the computational approaches (Chapter 4), and the developed scientific workflows (Chapter 5) of the data analysis and exploration. The collection of data using the Digital SMART monitoring approaches is described in Chapter 6. The concepts of interdisciplinary collaboration are conveyed in Chapter 7 and the evaluation of the Digital Earth approach for digitalization in Chapter 8. Finally, the lessons learned from the project are presented in Chapter 9.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

