DIGITAL EARTH

March 2020

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Newsletter of the *Digital Earth* Project

Contributions of all Helmholtz Centres to Digital Earth

This newsletter of Digital Earth presents recent results of joint activities of all participating Helmholtz Centres within Digital Earth.

2nd Digital Earth Interim Meeting

Daniela Henkel[1], Jens Greinert[1], Holger Brix[2], Diana Rechid[2], Stephan Frickenhaus[3], Angela Schäfer[3], Doris Dransch[4], Bruno Merz[4], Wolfgang zu Castell[5], Ralf Kunkel[6], Andreas Petzold[6], Peter Dietrich[7], Thomas Kalbacher[7], Roland Ruhnke[8], Ottmar Moehler[8]
1 Helmholtz Centre for Ocean Research Kiel (GEOMAR)
2 Helmholtz-Zentrum Geesthacht Centre for Materials and Coastal Research (HZG)
3 Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)
4 German Research Centre for Geosciences Helmholtz Centre Potsdam (GFZ)
5 German Research Center for Environmental Health Helmholtz Zentrum München (HMGU)
6 Forschungszentrum Jülich (FZJ)
7 Helmholtz Centre for Environmental Research (UFZ)
8 Karlsruhe Institute of Technology (KIT)

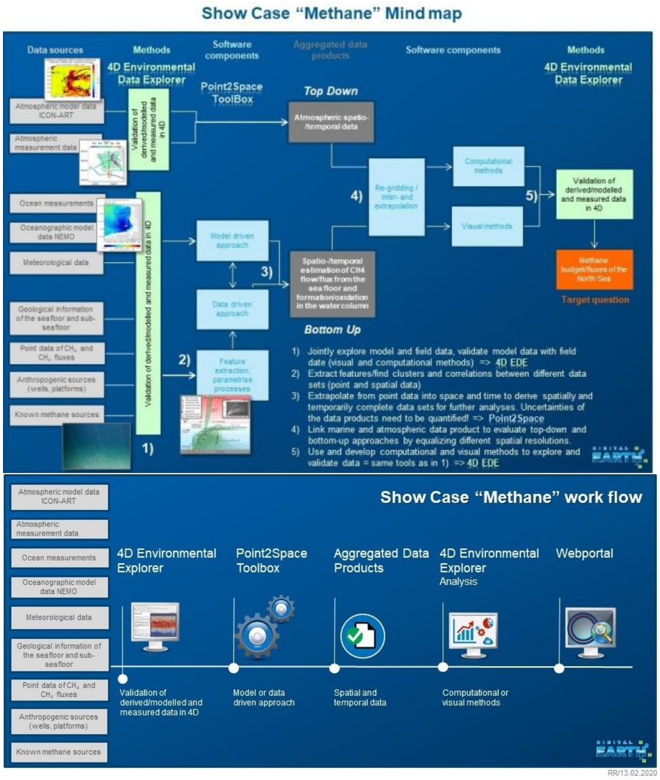
From 28 -30 January 2020 more than fifty natural scientists and data scientists from eight Helmholtz Centres met at the Helmholtz Centre Potsdam German Research Centre for Geosciences (GFZ) to discuss the progress of the project at the 2nd Digital Earth Interim Meeting. "The unique feature of the concept of Digital Earth is that we really live multi and interdisciplinarity; through these meetings we bring together colleagues with extensive expertise from all areas of earth and environmental sciences with data sciences in order to address a common scientific issue", explains Jens Greinert from GEOMAR and scientific coordinator of the project.

A further milestone for the Digital Earth project is the organisation of a summer school to make young scientists aware of the necessity of this exchange and will take place in Gummersbach at the end of September. "The concept of this summer school is an important step towards implementing the awareness of the 'next generation' for the necessity of integrating data science in natural and environmental sciences", emphasizes Doris Dransch, GFZ. Further information on the summer school can be found at (https://www.digitalearth-hgf.de/de/summer-school).

Show Case Methane - Current Activities

Roland Ruhnke[1], Jens Greinert[2], Valentin Buck[2], Flemming Stäbler[2], Everardo Gonzáles[2], Patrick Michaelis[2], Christian Scharun[1], Andrey Vlasenko[3], Wolfgang zu Castell[4], Mahyar Valizadeh[4], Carlos Garcia Perez[4]
1 Karlsruhe Institute of Technology (KIT)
2 GEOMAR Helmholtz Centre for Ocean Research Kiel
3 Helmholtz Zentrum Geesthacht Centre for Materials and Coastal Research (HZG)
4 Helmholtz Zentrum München German Research Center for Environmental Health (HMGU)

Although Methane (CH_4) is the second most important anthropogenic greenhouse gas its total budget as well as the absolute contributions of the different natural and anthropogenic emission sources to the total atmospheric budget are still under debate. Within Digital Earth natural and data scientists aim to quantify the total emission from gas and oil exploration in the North Sea which is not fully represented in current emission data bases like EDGAR (Emissions Database for Global Atmospheric Research) and thus reduce the uncertainty on the overall methane budget and source specific contributions. To achieve this goal a holistic view of the methane exchange and transport from the North Sea seafloor into the atmosphere has to be developed.



1) Jointly explore model and field data, validate model data with field date (visual and computational methods) => 4D EDE
2) Extract features/find clusters and correlations between different data sets (point and spatial data)
3) Extrapolate from point data into space and time to derive spatially and temporarily complete data sets for further analyses. Uncertainties of the data products need to be quantified! => Point2Space
4) Link marine and atmospheric data product to evaluate top-down and bottom-up approaches by equalizing different spatial resolutions.
5) Use and develop computational and visual methods to explore and validate data = same tools as in 1) => 4D EDE

RR/13.02.2020

Figure 1: Mind Map and workflow of the "Methane" Show Case.

This includes data science questions as: how to extrapolate emissions from distinct local source into the grid-space of the atmospheric model; how to parameterize the free gas transport through the water column depending on environmental conditions as tides, seasonal variations in temperature; how to link the impact of wind and storm events at the sea surface to the gas exchange between sea and atmosphere. Comparisons between top-down and bottom-up budget estimates will be used to explore which parameters have the greatest impact for adjusting bottom-up estimates to the top-down budget. The work flow and the goal of the show case are shown in Figure 1.

In the first year of Digital Earth the natural scientists have compiled their data for the reference year 2018 and have made them available to the Digital Earth public. Due to the size of the data the data have not been transferred to one center; marine cruise data (ocean and atmospheric) can be retrieved from a cloud server at GEOMAR, GCOAST35-NEMO model data of the North Sea are stored at DKRZ and atmospheric simulation results of the ICON-ART model are stored at KIT with giving access to members from Helmholtz centers by using the Helmholtz Data Federation Authentication and Authorization Infrastructure (HDF-AAI). In addition, data visualization tools like the 4D Data Viewer of GEOMAR and scripts for ParaView have been developed to display point measurement and gridded model data within the ocean and atmosphere at the same time and their variability over time.

A show case methane workshop at KIT in April 2020 will be used to connect the knowledge of methane emissions and transport pathways in different compartments with temporal and spatial changes to achieve the overall goal of this show case to link bottom-up and top-down fluxes.

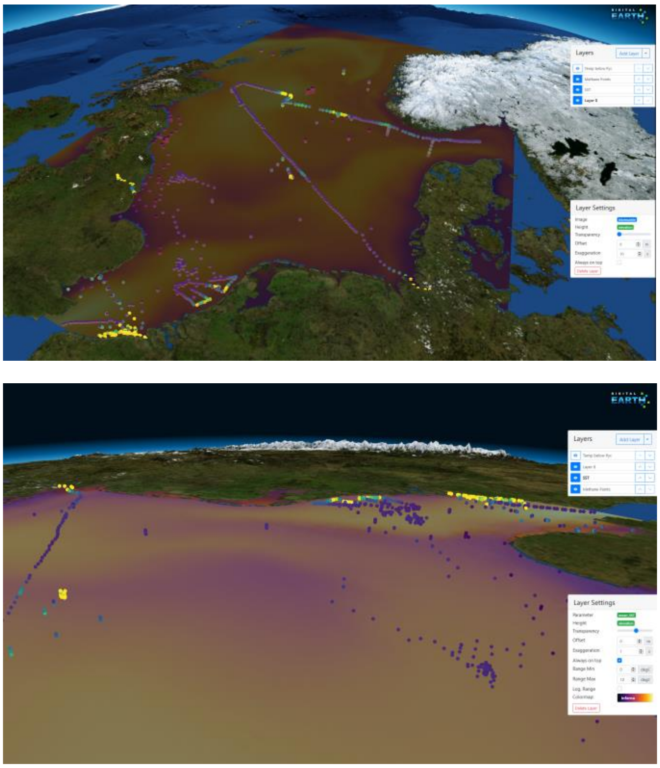


Figure 2: A view on the North Sea with sea surface temperature semi-transparent on top of the ETOPO2 elevation data set and colour coded methane concentrations extracted from the PANGAEA data base.

Show Case FLOOD - Current Activities

Stefan Lüdtke[1], Daniel Eggert[1], Daniela Raabe[1], Bruno Merz[1], Doris Dransch[1], Viktoria Wichert[2], Philipp Fischer[2], Peter Dietrich[3], Erik Nixdord[3], Christin Nam[4], Laurens Bouwer[4]

1 German Research Centre for Geosciences Helmholtz Centre Potsdam (GFZ)
2 Helmholtz Zentrum Geesthacht Centre for Materials and Coastal Research (HZG)
3 Helmholtz Centre for Environmental Research (UFZ)
4 Climate Service Center Germany (GERICS)

River floods and associated adverse consequences are caused by complex interactions of hydro-meteorological and socio-economic preconditions and event characteristics. Floods, as exceptional events at different spatio-temporal scales, often reveal an unexpected system behaviour and offer unique opportunities to gain system insights across multiple compartments and to improve existing methods and models. On the other hand, increasing flood losses indicate the immediate need of science-based facts to support flood risk management as well as disaster relief and recovery.

Within the Show Case Flood, we address these research questions with different workflows that approach challenges like the determination of a suitable location and time of sensors for even driven monitoring, the quantification of nutrients and pollutants contributing to the coastal system by a flood event or the comparison and analysis of large model ensembles.

As an example, Figure 3 shows a WEB-interface to analyse the change of precipitation days per year with respect to a baseline and a set of climate model predictions based on three different RCPs for a user defined region of interest. Similar WEB-interfaces are developed for all workflows in the Show Case Flood. All follow the same guidelines in terms of data provision and access and the included software libraries. This ensures that the different components are compatible with other workflows and thus reusable- a key feature to enable the linkage between the single workflows and the creation of a cross compartment system understanding within the Show Case Flood.

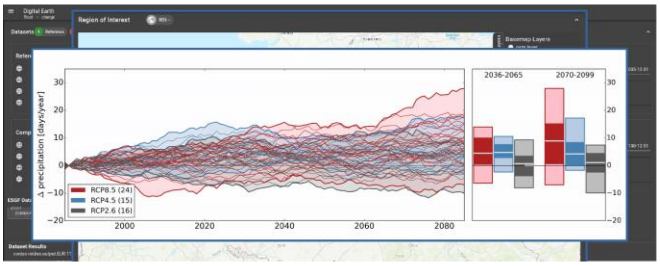


Figure 3: Web application to explore the results of climate model projections in comparison to a baseline for a user defined region of interest.

During the 2nd Digital Earth Interim Meeting last January, the participants of the contributing centres started and made the first steps to join the single workflows in order enable a holistic view and a cross compartment system understanding. There are multiple possible intersections between the different workflows. The methods, tools and infrastructure developed might be directly transferred to other workflows and research questions. For example: An optimal sensor deployment for hydrological monitoring, currently developed for the Weißeritz catchment in the Ore mountains in Germany, is likewise as useful to plan measurement campaigns in the North Sea to detect nutrients and pollutants. These connections between the current workflows and their associated research questions will be further investigated and established during the next months. Especially the contributions of the Bridging PostDocs, that provide a link between two different Helmholtz-Centers, will unlock huge potential and provide the basis for fruitful discussions.

Discussing a data-driven Monitoring for Digital Earth

Robert Wagner[1] & Tobias Goblirsch[2]

1 Helmholtz Zentrum Geesthacht Centre for Materials and Coastal Research (HZG)
2 Helmholtz Centre for Environmental Research (UFZ)

Due to technological limitations, current methods have so far failed to allow a holistic assessment of varying, large-scale environmental phenomena in an appropriate manner. One reason is the lack of suitable interfaces between data collection, analysis and modelling of complex system interrelationships, while the corresponding scientific disciplines have developed further apart. Another reason can be found in the fact that modelling in environmental sciences is often based on assumptions, as the data density and interlinkages between different datasets are insufficient in many cases.

To overcome this lack, an in-depth sampling paradigm was developed and initially tested to ensure that a single measured value can be unambiguously assigned and described. The **P**redictive **O**bject **S**pecific **E**xposure (POSE) serves as a basic measurement paradigm for data collection, ensuring that each measured variable, here the content of dissolved organic matter (DOM), has a clear spatial and temporal reference but can also be assigned to a specific context or prior information (Figure 4).

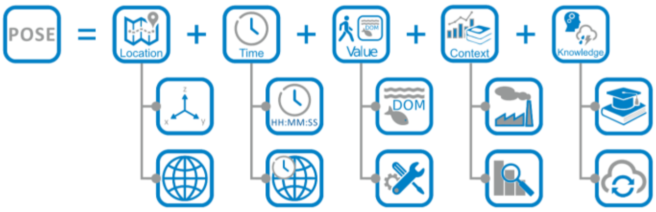


Figure 4: In the future, sensors will be based on a sophisticated data model in order to create a cutting-edge data basis for high performance environmental modeling.

The connection to *a priori* information is of particular importance in order to acquire data according to existing knowledge. The POSE approach also serves to derive a corresponding data model. Starting with the sensor system design and data collection, this data model also serves to efficiently design the database as well as the data provision and finally the data processing as well as the visualization.

This procedure serves to build up an entire architecture, which in this context will be called data stream architecture. Figure 5 shows an example of this architecture.

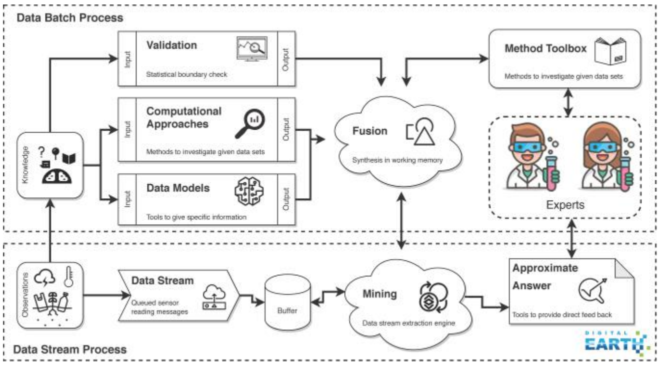


Figure 5: Exemplary presentation of a data stream architecture for the provision of real-time data based on IoT-capable sensors and sensor systems and consideration of server-side services for the valorisation and contextualization of environmental data.

The need for the development of a data stream architecture for the provision of real-time data is made clear in particular by the fact that with current monitoring approaches there is too much time lag between data collection and provision and finally also between the analysis and valorisation of the data. As a consequence, the approach to be developed in Digital Earth follows a data-driven monitoring strategy. This is based on the already mentioned POSE paradigm and on coordinated data processing methods. A fundamental distinction can be made between the pure data stream process and the data batch process. This differentiation brings decisive advantages in the provision of, e.g., near real-time data. In practice, sensor readings can thus be recorded, processed and finally provided in an environmental information system. It is irrelevant whether the data are time series or spatially distributed data formats. The most important criterion is the rapid provision of information to the user in the field. To achieve this, data is only recorded in conjunction with a unique time stamp and georeferencing. If this paradigm is not fulfilled, the measurement point is not considered valid and will not be considered for further processing. However, if the quality criterion is met, each sensor reading enters the so-called data stream.

Technically, this can happen via an IoT-capable sensor — i.e. a sensor sends a message to an end point via Wi-Fi. The message consists of all attributes necessary to describe the measured value (e.g. time stamp, measured value, measurement method, sensor, sensor ID, ...) and can be transmitted as text (JSON string). The endpoint is a data sink on a local server or a web server and can be implemented e.g. via an MQTT broker or a REST interface. However, as soon as the measured value message is available in the data stream and thus on the server, the real advantage of this architecture comes into play. So-called micro-services can now be triggered on the basis of the measured value. These are specialized processes that, for example, send a request to a stored criteria investigation on the basis of the geographic coordinates of the measurement. In the same way, a model could be triggered in order to optimize the validity of the modeling results in the study area on the fly, based on new measured values.

Another advantage is that the incoming data of the data stream can be analysed for distribution, quality or consistency using computational approaches. This would provide information about anomalies in the data but also help to characterize and identify the dynamics and phenomenological nature of a study area much faster. In the same way, validation of the incoming data sets would be possible by using stochastic methods or by comparison with historical data sets or a priori information.

The scientific and technological advantage of this approach lies in the fact that the fusion of these micro-services creates new methods that can also be used for other issues. This is possible because the micro-services communicate with each other exclusively via defined interfaces and represent independent and autonomous processors. The scientific users as experts finally provide the test and decision criteria according to which process steps are passed through. This means that initial analyses can already be carried out in the field with few technological resources. As a result, knowledge of the ecosystem can be gained much more quickly.

Within Digital Earth, this approach is implemented by setting up a measurement infrastructure using the example of a water quality probe.

Besides the integration of the water quality probe into an environmental information system for processing, storage and presentation of the data in the field, e.g. on a ship, it is particularly important to be able to valorise the aggregated sensor data with other, parallel collected data or historical data sets. This also includes the coupling of real-time data and modeling approaches in order to trigger models with the incoming data from the data stream.

In order to deal specifically with the show-case flood, a water quality probe upgraded according to the IT architecture or IT infrastructure described above could be used as a scenario covering the entire course of the Elbe up to the Elbe estuary in the North Sea. The data thus obtained could be merged with other data in real time. For this purpose, real-time access to all water levels of all gauging stations along the Elbe already exists. In addition to the water level, parameters such as conductivity, temperature and turbidity are also available in isolated cases. In addition, remote sensing data is also available via micro services and can be integrated within the data stream architecture as mentioned above.

A prototypical setup of this measuring configuration in connection with a field test to the extent described above could thus provide essential information on how data-driven monitoring up to the coupling to models could be carried out in the future.