Chapter 3
Data Analysis and Exploration with Visual Approaches

Everardo González Ávalos, Doris Dransch, Nicola Abraham, Valentin Buck, Daniel Eggert, Tom Kwasnitschka, Daniela Rabe, Flemming Stäbler, and Viktoria Wichert

Abstract A comprehensive study of the Earth system and its related processes requires a holistic examination and understanding of multidimensional data acquired with a large number of different sensors or produced by various models. To this end, the Digital Earth project developed a set of software solutions to study environmental data sets using visual approaches. In the following chapter, we present three data visualization products developed to deal with the challenges of the analysis and exploration of environmental data.

Keywords Data visualization · Data exploration · Spatiotemporal exploration · Linked views · Immersive visualization

3.1 Challenges

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. Several challenges arise from this situation:

1. The need to establish connections between causes and consequences of geoscientific processes. The mechanisms that link these processes together might be

E. González Ávalos (✉) · V. Buck · T. Kwasnitschka · F. Stäbler
GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
e-mail: egonzalez@geomar.de

D. Dransch · D. Eggert · D. Rabe
Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany

N. Abraham · V. Wichert
Helmholtz-Zentrum Hereon, Geesthacht, Germany

few and scattered; thus, the connections become evident only when observations from various disciplines and sources are brought into relationship with each other.

2. 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 regarding information across different scales simultaneously. This can be the case with sediment samples that encode information in cm scale, while remote sensing data does so in km^2 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.

3. The need for suitable means to integrate a variety of heterogeneous spatiotemporal data sets. 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 simultaneously display data even when it differs in scale, variables, or accuracy. We use interactive visualization to display heterogeneous data sets in a unified 4-dimensional environment and to interactively explore the data with respect to context and connections.

We applied different visualization techniques and environments and adapted them to our geoscientific requirements: We developed and respectively utilized following visualization tools:

– The Data Analytics Software Framework (DASF) which provides linkable visualization components (multiple linked views),
– The Digital Earth Viewer, an engine for 4D data contextualization and visualization, and
– The ARENA2, an immersive visualization infrastructure.

3.2 The Data Analytics Software Framework (DASF) Providing Linkable Visualization Components

3.2.1 Introduction

The Data Analytics Software Framework (DASF) (Chapter 5.2.3) which we have developed in Digital Earth aims at implementing scientific data analysis workflows. Besides the module to integrate components into a scientific data analysis workflow, it also provides a visualization module to present the data and results that are used and created in the workflow.

3.2.2 Visualization Concept

In order to enable scientists to simultaneously show and explore the data in its multiple dimensions (space, time, different variables, accuracy), the DASF visualization module consists of a variety of visualization components that are linked according to the multiple linked view visualization approach (Roberts 2005; Spence 2007). The visualization components can be all types of views on data: maps, diagrams, tables, animations, or calendar maps to name a few. The visualization components are presented in several windows simultaneously. The windows are linked; this means that operations in one window affect all other related windows. For example, if a user selects a subset of data in one window (e.g., in a map), all other windows visualize information for the same selected data subset. The operations are executed interactively with mouse-based interaction techniques, such as brushing, highlighting, or filtering. Multiple linked views are a widely used technique in data and information visualization to present multivariate and multidimensional data sets.

3.2.3 Technical Implementation

Our architecture to technically implement the DASF visualization module utilizes well-established techniques. To create the single visualization components, we used for instance *Openlayers, (*https://openlayers.org) or *leaflet* (https://leafletjs.com/) for maps, *D3* (https://d3js.org/) or *chartjs* (https://www.chartjs.org/) for charts (e.g., histogram, time diagram, radarplot), and *Vueftify* (https://vuetifyjs.com/en/components/data-tables/) for tables. To implement the multiple linked view approach, we applied the "reactive properties" concept which is provided by the *Vuejs and Vuetifyjs* software package (https://vuejs.org/v2/guide/reactivity.html). For more information on the complete DASF implementation see, Chapter 5.2.3.

3.2.4 Application

The DASF visualization module has been applied in all workflows contributing to the Digital Earth Flood Event Explorer (Chapter 5.3). For each workflow, an aligned visual interface was assembled from the various visualization components and the multiple linked view approach. Exemplarily, we show the visual interface of the "River Plume Workflow" which supports geoscientists to investigate the impact of river floods on the marine environment (Chapter 5.3.3). It should enable scientists to detect the spatiotemporal influence of the river plume on the sea due to chemical anomalies and to answer the question: Where and when can we detect the flood river plume in the sea? Several data sets have to be used and combined to answer this question. These include observations of chemical characteristics of the waterbody,

such as salinity, which are collected by a sensor at regular intervals along a ferry route, and a data set from a physical model that calculates model trajectories of the waterbodies observed on the ferry route up to 10 days into the past and 10 days into future from a reference day. On the basis of these data sets, anomalies of salinity, chlorophyll or surface temperature, and thus the spatiotemporal behavior of the river plume can be detected. Chlorophyll anomalies related to the river plume occur when the deviation is above the expected range, while salinity and surface temperature anomalies occur when the deviation is below the expected range.

The multiple-view visualization consists of following components (Fig. 3.1): The map view (V1) shows the spatial distribution of the observed ferry box data and calculated model data; the color encodes the concentration of a chemical or physical parameter such as salinity. Another view presents a comparison of the quantities for each chemical or physical parameter measured inside and outside of a user-defined region of interest. The bar chart (V3) gives an overview of all chemical/physical parameters; the table below (V4) shows detailed information for one selected parameter. A further view (V5) presents temporal information and visualizes the occurrence of anomalies of the parameters in time in a calendar-heatmap. These anomalies are
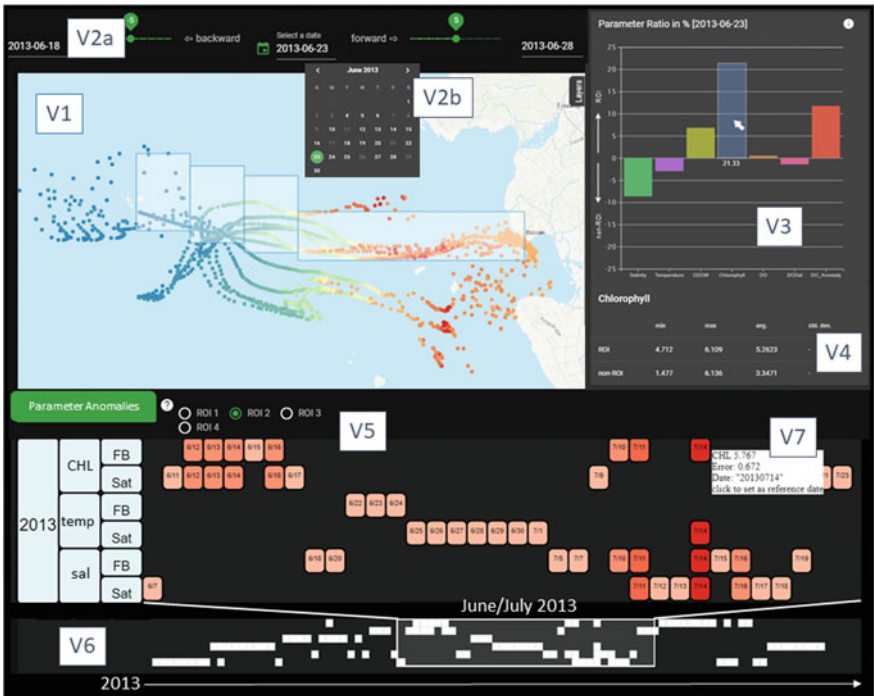

Fig. 3.1 Multiple linked views to determine the spatiotemporal behavior of the flood's river plume in the waterbody, such as the North Sea (Interface-in-Action Video: https://youtu.be/yl8ngubBxYY)

candidates for detections of the river plume in the observational data and are determined automatically through a Gaussian Regression algorithm (Chapter 4.5.1). In the calendar-heatmap, days with deviations from the expected range of each variable are shown with different color intensity. More details about the anomaly can be added for each entity in the heatmap through a mouse-over action to load the relevant observational and model data into the interactive map (V7). An additional overview presentation (V6) shows the whole data set and puts the data subset presented in V5 into context. The links between the single views are realized with interactive operations like the filtering of a region of interest (V1), the selection of a continuous time interval (V2a), or discrete time step (V2b), or by mouse-over actions for presenting additional information (V7).

Added Value: The visual interface of the "River Plume Workflow" supports scientists to visually put the various data sets into context: anomalies of chemical/physical parameters and their spatial and temporal distribution. The overall view on the one hand and the capability of interactive data exploration on the other hand assist scientists to finally detect the river plume and its behavior in space and time. The combination of the visual interface and the Gaussian Regression algorithm to detect the river plume that was also developed in Digital Earth (Chapter 4.5.1) provides a novel approach for scientists to comprehensively analyze the various data sets and to detect the river plume.

3.3   The Digital Earth Viewer

3.3.1   Introduction

The Digital Earth Viewer is a web application for spatiotemporal contextualization and visualization of heterogeneous data sources. It was developed with the goal of enabling real-time exploration of geoscientific data sets across spatial and temporal scales. To this end, it is capable of ingesting data from a large variety of types that are usually found in the geosciences, and it deploys a user interface, which allows for interactive visual analysis. At the same time, online and offline deployment, cross-platform implementation, and a comprehensive graphical user interface are all capabilities that make the Digital Earth Viewer particularly accessible to scientific users.

3.3.2   Visualization Concept

This infrastructure provides a framework in which new visualizations for heterogeneous data set can be created with relative ease. Since no reduction in dimensionality

is undertaken and the four-dimensional data (space and time) is displayed in a four-dimensional context, temporal or spatial distortions are mostly avoided; this leads to an improved interpretation capability and supports the understanding and contextualizing of information in an intuitive process. The Digital Earth Viewer enables the user to visualize assorted geoscientific data in three spatial dimensions and across time. It projects the spatial coordinates latitude, longitude, and altitude into a virtual globe and builds an ordered registry of temporal events. Both of these features can be accessed through user interface elements in real time.

Different data sources can be added simultaneously to form individual layers, each with its own data basis and transformation pipeline. Global parameters, such as the position in time and space or the intrinsic scale of the visualization, are implicitly and explicitly communicated to the user in the graphical interface while specialized parameters can be set through a menu. The transformations for each data source happen independent from one another and are composed together into one final result, allowing the blending of multiple data sources.

Data is grouped into several different categories for display. Traditional 2D maps can be projected onto a spherical surface and displacement along the sphere's normal vector can be applied. Scalar values are mapped onto one of a set of color maps, while precolored maps are passed through. Sparse data can be displayed as a point cloud which is projected, colored, and culled according to the global projection parameters. For the intuitive representation of vector fields, an animated particle system is created in which the particles follow the vector field which is projected onto the virtual globe.

3.3.2.1 Technical Implementation

The tool is a hybrid application, which is split into a server back-end and a client front-end. The rust (https://www.rust-lang.org/) back-end handles data extraction from different file formats as well as re-gridding into viewable areas and caching. It can be hosted remotely on a server or locally for offline access. The front-end consists of an HTML interface component and is responsible for the 3D data rendering using the WebGL API (https://www.khronos.org/webgl/) and the implementation of the graphic user interface controls which uses Vue.js (https://vuejs.org/).

For each data type that the server needs to ingest, a specific data type is built that transforms the incoming data format into an internal representation optimized for computation operations. This representation is then passed over to the client, which applies graphical transformations to compute a visual representation of it.

A running instance of the application can be accessed under the following web address: https://digitalearthviewer.geomar.de.

The Digital Earth Viewer is an open-source software licensed under the EUPL (https://joinup.ec.europa.eu/collection/eupl/eupl-text-eupl-12).

3.3.3   Applications

3.3.3.1   Methane Budget in the North Sea

The Digital Earth showcase "Methane Budget in the North Sea" set up to build a methane gas budgeting for the North Sea region. The showcase makes use of the Digital Earth Viewer to unify a large number of data sets under a single visualization interface. Boreholes from fossil fuel production are known to be important sources of methane that is released into the atmosphere. The GEBCO (https://www.gebco.net) bathymetry from the North Sea region is displayed in a 3D elevation model. The aerosol and trace gases dispersion model ICON-ART (Rieger et al. 2015) is used to calculate the contribution of these boreholes to the atmospheric methane concentration. The viewer's interface allows to quickly compare the resulting data product with existing methane estimates from the EDGAR (https://data.jrc.ec.europa.eu/collection/edgar) emissions database and provides a visual assessment of their accuracy. In a similar way, measurements of geochemical water properties from the expedition POS 526 (https://oceanrep.geomar.de/47036/) are displayed in spatial context of other measurement compilations from the Pangea (https://www.pangaea.de) and MEMENTO (Bange and Bell 2009) databases. The observation of their development over time is further supported by the visualization of three-dimensional physical water properties like current velocities and pycnocline depth obtained from the NEMO Model (Madec 2016). An instance of the Digital Earth Viewer displaying the Methane showcase can be found under following web address: https://digitalearthviewer-methane.geomar.de.

Added value: Using the Digital Earth Viewer, scientists can simultaneously access and visualize data from all the sources mentioned above. Seamless spatial navigation allows them to directly compare the global impact that regional methane sources have in the atmosphere, while temporal components enable them to do so across the different seasons of an entire year (Fig. 3.2).

3.3.3.2   Explorable 4D Visualization of Marine Data

The expedition Mining Impact-II started a series of experiments to answer some of the most important questions regarding profitability and sustainable exploitation of resources in deep sea mining, such as mining for manganese nodules. Deep sea exploration is a challenging endeavor that can be greatly aided by the use of modern visualization techniques. In this work, we aim to recreate a sediment plume which resulted from an underwater dredging experiment. This will help to quantify similar sediment depositions from mining that could impact deep sea ecosystems at a depth of 4,000 m. Sensor data fusion allows for the virtual exploration of experiments on the seafloor; the following is an overview of the different data sources acquired during the expedition that come together within one visualization:
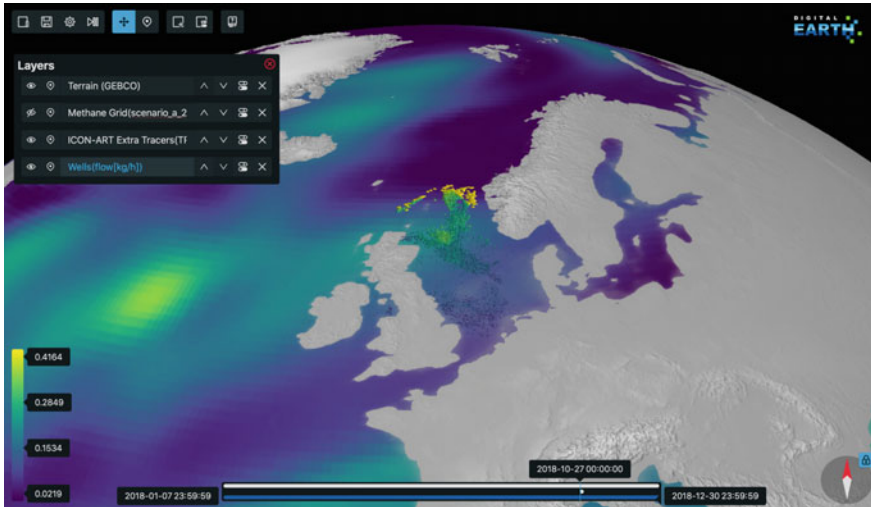
Fig. 3.2 Digital Earth Viewer used to display the North Sea area, atmospheric methane calculations from the ICON-ART model, and methane flows from oil and gas wells

- Turbidity sensors calculate this optical property of water by measuring the scattered light that results from illuminating the water.
- Current sensors use the Doppler effect to measure the velocity of particles suspended in the water column and thus calculate the speed and direction of water currents.
- Multi-beam echosounders emit fan-shaped sound waves and use time of flight to reconstruct the seafloor bathymetry.

For the dredging experiment, an array of 8 turbidity sensors and 8 current sensors was placed on the sea floor in an area previously scanned with a high-resolution multi-beam echosounder mounted beneath an autonomous underwater vehicle. Moreover, a virtual sediment plume was modeled and integrated into the experiment. The spatiotemporal contextualization of all data sources took place allowing for a real-time simultaneous analysis of heterogeneous data sources in 3D and across time. An instance of the Digital Earth Viewer displaying the Sediment Plume showcase can be found under following web address: https://digitalearthviewer-plume.geomar.de.

Added value: Using the Digital Earth Viewer, the experiment grounds were recreated. This virtual environment allowed scientists to peer beyond the darkness of the deep sea and explore the impact of a simulated mining endeavor. The numerical model of the sediment plume was superimposed and compared to the in situ data obtained by the turbidity and current sensors resulting in a visual confirmation of the sensor placement and model correctness. Deployment as a web-based application accounted for cross-platform portability across devices and operating systems, allowing scientists to visually explore the phenomena that take place in this virtual abyssal plain, and to share their discoveries with a wider audience (Fig. 3.3).
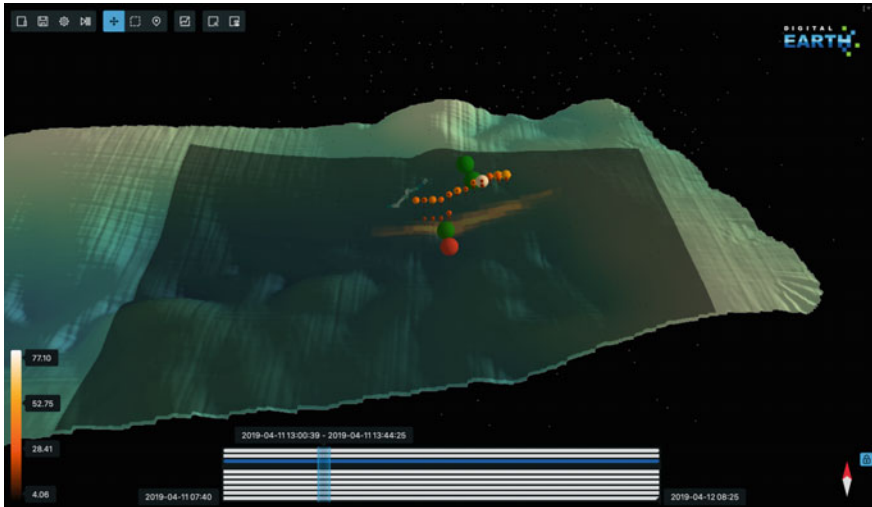
Fig. 3.3 Parallel exploration of multiple marine data types: the Digital Earth Viewer is used for the visual corroboration of a 3D plume model. This is done comparing the plume dispersion values (small orange dots) with the readings of the turbidity sensors (larger green and red dots) and the water currents (green lines)

3.4 Spatially Immersive Visualization of Complex Seafloor Terrain

3.4.1 Introduction

To a large extent, field geologists derive their mental models of complex outcrops and depositional features (e.g., volcanoes) from situational awareness and the first-person perception of an environment via the bodily senses. To date, this is still reflected in the way geologists are trained, even in light of emerging digital technologies. Moreover, economies of scale (e.g., geospatial correlation of small outcrops) unfold only across large "industrial scale" survey areas, as opposed to isolated, confined demonstrator missions. In order to bridge the scales of these two different planes of operation and to retain a true sense of dimensionality while doing so, the ARENA2 spatially immersive visualization laboratory was developed at GEOMAR.

3.4.1.1 Visualization Concept

The ARENA2 projects geospatial data onto an elevated dome and allows one or multiple users to navigate freely across all three dimensions of a virtual environment. A faithful recreation of live exploration of the geological features is enabled by this.

Notably, most of the ARENA2 applications are also available in desktop environments, or even through web browsers. This creates a continuum of visualization infrastructure scaling from opportunistic, personal access all the way to collaborative, structured visualization campaigns serving the entire spectrum of academic use cases: data sets that were previously cured and preprocessed using commonly available software applications on a desktop PC can be displayed and analyzed by multiple users with a new sense of immersion.

3.4.2 Technical Implementation

The ARENA2 features a tilted, stereoscopic projection dome covered by a five-channel projection system, which is in turn fed by a node-based visualization cluster of five computers. We follow a three-tiered approach on visualization software: First, the graphical output of the photogrammetric post processing software itself is ported to the dome environment by means of OpenGL buffer distribution and re-interpretation across the cluster, known as openGL-hooking. Second, a dedicated, distributed visualization software loads a statically exported point cloud in a georeferenced virtual globe context. Third, parallel WebGL-based visualization tools are synchronized across the cluster, with dome-specific warping and blending applied on the operating system level. Here, a preprocessed level-of-detail pyramid of the point cloud is dynamically streamed to all web clients.

3.4.3 Application

To exemplify the procedure, an inaugural data set demonstrator was created: We surveyed a 500×500×80m volcanic crater in 2016, hosting an extensive, active hydrothermal field at the Niua South Volcano, Tonga, by means of the remotely operated vehicle (ROV) ROPOS (https://www.ropos.com/index.php/ropos-rov). During 100 hours of a methodical survey, some 220.000 photographs of the seafloor were collected using a single-lens reflex camera and transformed into a 3D, color textured terrain model. Its scale and detail form a considerable challenge for agile, interactive visualization procedures.

Added value: Scientists studying these hydrothermal vents will be able to do so in life-like detail. The virtual environment can be freely explored without the physical constraints of the ROV used for the original acquisition. For the wide majority of the population that will never be able to take part in an expedition such as the one where these geological features were captured, the ARENA2 can help bring these people closer to the richness of the underwater landscapes with an immersive experience (Fig. 3.4).
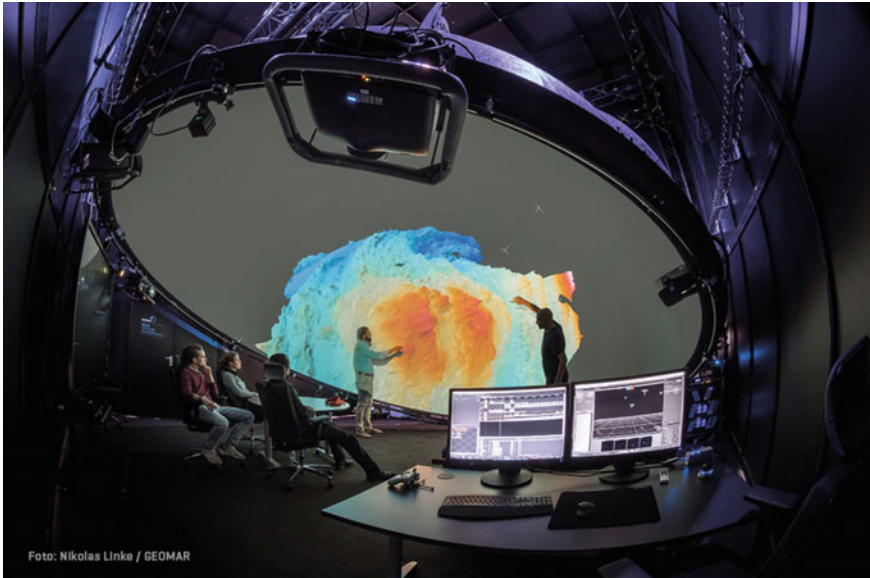
Foto: Nikolas Linke / GEOMAR

Fig. 3.4 ARENA2 projection dome displaying a bathymetric relief of the deep sea. The infrastructure provides an immersive experience for the presenter and the audience alike, opening new narrative possibilities and discussion planes

3.5 Assessment of the Three Visualization Approaches and Techniques

The three data visualization techniques developed in the Digital Earth Project respond to the challenges formulated in 3.1. All integrate a variety of heterogeneous spatiotemporal data sets, but each does so in a different way. The linked views of the DASF relate directly to the first challenge and propose an efficient answer with the use of graphs and charts for data conceptualization. These provide rapid access to insights of interlaced behaviors which a user can then recognize as causal links. The 3D data representation capabilities are the core strength of the Digital Earth Viewer which excels at exploring all dimensions of spatiotemporal data from multiple sources. It addresses specially the second challenge, which is to retain a sense of spatial and temporal coherence across different scales. Both the applications based on DASF and the Digital Earth Viewer are deployed as web applications and thus are accessible across platforms. Considering the ARENA2 consists of a building-sized infrastructure, accessibility and deployment are by far its largest limitations. To make up for this, it provides unparalleled data immersion capabilities and a revolutionary way of experiencing data, during both the exploration and presentation of geoscientific data sets. With this, it also addresses challenges 1 and 2.

References

Bange HW, Bell TG (2009) MEMENTO: A marine methane and nitrous oxide database, Solas News, Issue 9, Spring 2009, p 42

Madec G, NEMO System Team, 2016: NEMO ocean engine. Scientific Notes of Climate Modelling Center (27) - ISSN 1288–1619

Rieger D et al (2015) ICON–ART 1.0—A new online-coupled model system from the global to regional scale. Geosci Model Dev 8:1659–1676

Roberts JC (2005) Exploratory visualization with multiple linked views. In: Dykes J, MacEachren AM, Kraak M-J (eds) Exploring geovisualization. Elsevier, pp 159–180

Spence R (2007) Information visualization. Design for interaction. Pearson