EuroVis Workshop on Visual Analytics (2023)
M. Angelini and M. El-Assady (Editors)

A Practical Approach to Provenance Capturing for Reproducible Visual Analytics at an Ocean Research Institute

A. Bernstetter[1,2], T. Kwasnitschka[1], and I. Peters[2,3]

[1]GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany
[2]Kiel University, Germany
[3]ZBW Leibniz-Information Centre for Economics, Kiel, Germany

Abstract

Reproducibility - and the lack thereof - has been an important topic for some time in the field of Human-Computer Interaction. Visual analytics workflows and in extension immersive analytics workflows are no exception there and benefit from being more transparent and reproducible. At our research institute, domain scientists in ocean research are using interactive visualization workflows for sensemaking processes. We are building a framework that supports these workflows by shifting the focus from solely lying on the end-product (i.e. published insights and visualizations) towards the generation process. We do this by capturing, organizing, and visualizing provenance artifacts using a modular and extensible web-based application. We not only apply this framework to conventional 2D display-based work but also workflows inside a unique and spatially immersive projection dome.

CCS Concepts

• ***Human-centered computing*** → ***Scientific visualization; Visual analytics;*** *Collaborative interaction;*

1. Introduction

While by no means a new trend, the topic of reproducibility [FW21] is continuously getting attention in many scientific fields, and computational science is no exception here [Pen11]. Data processing and workflows in computational science need to be transparent and reproducible to foster open and accessible research and the field of Human-Computer Interaction (HCI) has a role to play in this area [FDTWS19]. Visual analytics as an HCI approach that uses interactive visualization for sensemaking [PC05] is certainly included in this call for improving reproducibility. Already in 2006 for example Kindlmann argued that "visualization must be reproducible if it is to be scientific" [Kin06]. As Dou et al. [DRC10] state, typically the focus is on the product, i.e. the produced visualization, but the process itself also contains important information about how the product was created and the analysts' workflows.

For the purpose of transparency and reproducibility of these processes, information on the provenance of workflows and visual analytics products is needed. Provenance is broadly defined by the World Wide Web Consortium (W3C) as "[...] the sources of information, such as entities and processes, involved in producing or delivering an artifact " [MBC13]. Making such information available may help turn a "reproducibility crisis" into a "credibility revolution", as Vazire puts it [Vaz18].

In this work we introduce the *Digital Lab Book* (DLB), a tool designed to support visual analytics workflows by connecting to visualization applications, and capturing and managing the provenance information that is produced during interaction processes. The first iteration of this tool is a prototype built with use cases of ocean researchers in mind. At the GEOMAR Helmholtz Centre for Ocean Research in Kiel, Germany, scientists from the broad area of ocean sciences are using interactive visualizations for visual data exploration. Beyond traditional 2D workstations, GEOMAR has a spatially immersive visualization laboratory available that offers an environment for immersive analytics. Such an environment yields a multitude of possibilities for interaction, immersion, and collaboration but it also comes with special preconditions regarding the reproducibility of visual analytics workflows.

2. Motivating Example: Visual Analytics at an Ocean Research Institute

Visualization has a long history in geosciences and ocean research. For large portions of the public, the oceans are far away, literally and figuratively, and visualizing information about them helps to make these topics graspable. Ocean models, bathymetric maps, 3D seafloor model reconstructions, ocean currents visualizations are all examples of the usage of visualization in ocean research. In their workflows, ocean researchers are using a multitude of different tools and programs, ranging from programming languages and related technologies (R, Python, Jupyter Notebooks) over traditional general-purpose tools (e.g. ParaView), to applications built specifically for this research area (e.g. OceanDataView [Sch15]).

DOI: 10.2312/eurova.20231095

The topic of immersion and immersive analytics comes into play where it is necessary or at least beneficial to simulate an environment that cannot easily be visited and examined in the real world. One of these use cases is that of marine geology and volcanology. Traditionally, geologists would go on field work excursions to quarries and other sites, and directly interact with their studied objects. They would visit geological outcrops, visually inspect, touch, and measure, directly form hypotheses and insights on-site, immediately note down their thoughts and contextualize the gathered information (e.g. put it in relation to weather effects).
Marine geologists often do not have these opportunities. When the only way to come close to the seafloor is either via remotely operated or autonomous underwater vehicles, or expensive submersibles, there is a need for an alternative to re-instate in-depth visual analysis.
The ARENA2 [KSK*23] is a digital projection dome, simulator, and visualization laboratory located at GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany.

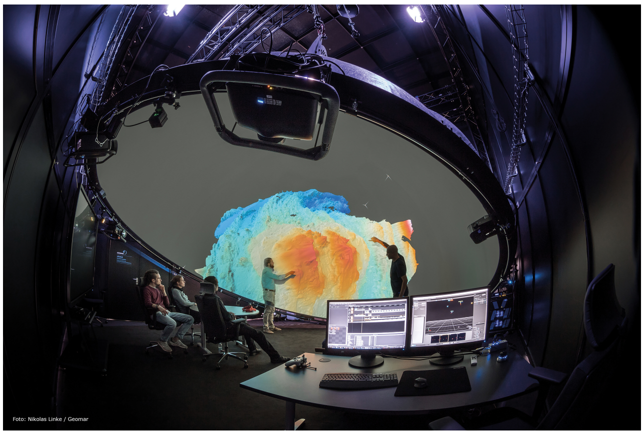


Figure 1: *Demonstrating the ARENA2's capabilities for immersive analytics by interactively exploring a 3D seafloor model.*

Applying the concept of a hemispheric projection screen found primarily in digital planetariums, this dome is used for the visualization of geospatial data, enabling collaborative work in small groups inside an immersive environment. It facilitates visual analytics by allowing domain scientists working in marine or geoscientific research to interact with visualized data. The ARENA2 is able to simulate field work for cases as described above where scientists cannot get to physically remote or inaccessible places.
This environment aims to solve this problem by allowing domain scientists to not only view pre-recorded underwater video footage in real-life scale but also to visualize data in various ways such as 3D models or point clouds, and display them in an immersive environment enabling interactive visual data exploration. For this purpose, a multitude of different visualization applications and geospatial analysis tools have been integrated including planetarium software, Fledermaus (`https://qps.nl/fledermaus/#`), the Unreal Game Engine, and the Digital Earth Viewer [BSM*22], an application developed in-house at GEOMAR. This unique context presents the fundamental challenges for this work. How can provenance be used in such a spatially immersive environment to bring collaborative immersive analytics towards a quantitative and reproducible utilization?
The need for tracking provenance of geospatial analyses is certainly present, as Ziegler and Chasins discover in a need-finding study with users of GIS systems where participants struggled to create reproducible and shareable workflows [ZC23].

3. Related Work

According to a definition by the W3C similar to the one introduced before, "provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" [20113]. This definition mentions the issue of quality assessment which is an important topic in the field of visualization. Van Wijk for example raises this exact question of how to assess the value of a visualization [vW05]. What are the requirements for a visualization to be beneficial to a user, the public, or the respective research field itself?
A good overview of provenance research is given by Ragan et al. [RESC16] who characterize the field of provenance in visualization and data analysis, as well as Xu et al. [XOW*20] who published an extensive survey on the topic in 2020.
The concept of provenance in visual analytics has been given thought before, Dou et al. [DRC10] for example build on van Wijk's model and address the topic of how to capture the reasoning processes of users during visual analytics workflows. They differentiate internal and external capturing. The former includes artifacts coming directly from the visualization application such as the visualization state and interaction logging. The latter includes either results externalized by users themselves (e.g. notes, dictations) or information captured using additional methods and hardware (e.g. observers, eye and motion tracking).
A hierarchy of how much information is contained in different levels of captured artifacts is described by Xu et al. [XAJK*15]. They distinguish between *events*, *actions*, *subtasks*, and *tasks* in a hierarchical analytic provenance model. *Events* refer to interactions such as key presses or mouse clicks with low semantic richness, *actions* are the next level including actions such as zooming, rotating, or loading a dataset. *Subtasks* then are analyses performed during a workflow, and on the highest level *tasks* describe the overall endeavor why the analysis is being done. Our application currently mainly operates on the level of actions in a transferred sense by determining the difference between two sequential visualization states. By offering users options to externalize information via note taking and mind maps, however, it reaches into the subtasks category. If a visualization application offers direct logging of low-level events itself, the DLB is also able to adapt to that captured information as well.
The aim of our application is not to implement an event-driven history management system and simply offer undo-redo functionality. Many visualization applications already implement this themselves. The Digital Lab Book is meant to offer a way to record and externalize thought processes and insights emerging during visual analytics workflows and to be able to export these. This supports publication as supplemental material and reproducibility of such workflows.
There are several earlier works which approach the issue of

("third-party") provenance capturing for visual analytics. VisTrails [CFS*06] describes early work in this area and is a provenance management system that can be combined with multiple existing systems. It has also already been used in ocean research at an ocean observatory, the Center for Coastal Margin Observation and Prediction in Portland [HLB*08]. Since 2018, however, VisTrails is no longer maintained.

Kreuseler et al. [KNS04] describe a history management system for visual data mining and propose a branching history of states which the DLB also implements due to its usage of Git for version management. Nancel and Cockburn [NC14] dive into more theoretical detail on conceptual considerations about interaction history. They draft a conceptual model of interaction history which allows users improved flexibility regarding temporal interactions with the history of past commands and states. Both Nancel and Cockburn, and Kreuseler et al., however, mostly target improving the interaction with the visual analysis application itself by offering improved ways to undo and redo actions. The DLB's goal is rather to support gathering and externalizing thought processes and insights together with the visualization states, as mentioned above.

Stitz et al. nicely describe the general concept of a tool that combines visualization, provenance tracking, and retrieval of information and visualization states from the provenance graph with the example of their - also web-based - application *KnowledgePearls* [SGP*18]. They also offer the functionality to export and import the provenance data in the JSON format. In contrast to the DLB, however, the provenance capturing part of the application is built into the visualization application. This approach is less flexible and generally applicable as the visualization application itself has to be adapted which is often not possible with software that is not open source. Cutler et al. [CGL20] implemented a web-based provenance tracking library. Their approach - being a software library - can be included in a web-based visualization application itself when programming it, and captures its provenance whereas the DLB acts as third-party and is able to support stand-alone applications.

A study in which provenance information of (visual) data analysis workflows is used for collaborative hand-off scenarios between researchers has been conducted by Block et al. [BER*22]. They suggest that representations that are able to summarize and reduce the complexity of the provenance information will be beneficial in hand-off analysis scenarios. This implies that it is not sufficient for a provenance capturing tool to simply gather an interaction log. It has to manage the information and make it available to the users in a way that maximizes the user experience. Their visual analysis task and tool, however, consists of analyzing text-based documents in contrast to the interactive (immersive) 3D visualizations which the DLB targets.

The Digital Lab Book faces challenges which - to our knowledge - have not been addressed in the same way before by having the goal of supporting multi-user collaborative (immersive) visual analytics workflows with an additional focus on ocean research.

4. The Digital Lab Book

The Digital Lab Book is a web-based application which is currently a work in progress. The concept behind the DLB is that of a tool which captures the provenance of visual analytics workflows, and organizes and visualizes this provenance.. Such a tool is able to connect to one or more visualization applications, either by being directly "baked into" the application itself or acting as a third party. In the case of the DLB, this communication is done via web sockets and APIs, depending on which software it is connecting to. Its source code is available at `https://git.geomar.de/digital-lab-book/digital-lab-book`.

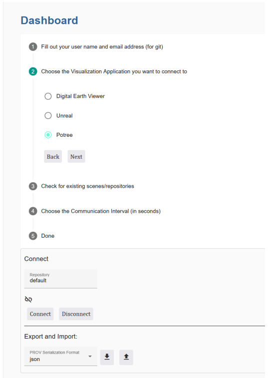
4.1. Implementation

The design choice of a standalone application was made to be as independent from visualization applications as possible with portability and universality being important design criteria. Researchers in geoscience are using many different applications depending on which specific field they are working in. Some of these applications are closed-source and proprietary, which means directly integrating a provenance tool into them is not possible. Some of those applications, however, offer interfaces for third-party access. The choice of building a web application was made because it allows to potentially run the DLB as "Software as a Service" on a hosted server in the future so that users do not have to install it themselves. Apart from that, a web app increases the accessibility because many modern devices include a web browser. The interaction requirements for the DLB were originally set by a domain expert (geologist) for the above described use case of simulating geological field work.

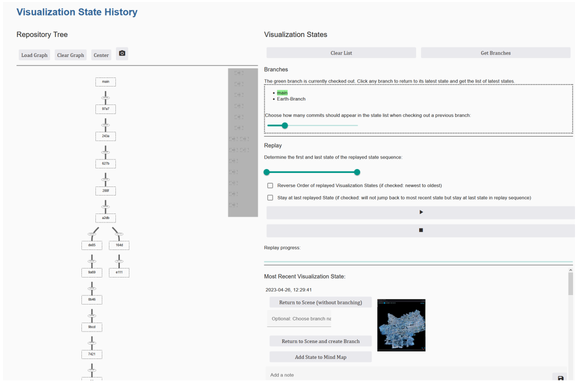
The main visualization software that a user is working with communicates information about its visualization state and possibly other provenance artifacts to the DLB if it is requested. The DLB in turn stores and visualizes the state history as a provenance graph. A user is then able to retrieve states from this provenance history and cause the visualization to return to the retrieved state from inside the DLB.

Generally speaking, the minimum requirements for a visualization application to be supported are an interface that allows the DLB to a) request the current visualization state e.g. by sending an HTTP GET request for the data describing the state (e.g. JSON data) and b) trigger the reloading of a previous visualization state e.g. by sending an HTTP POST or PUT request including the state's save data.

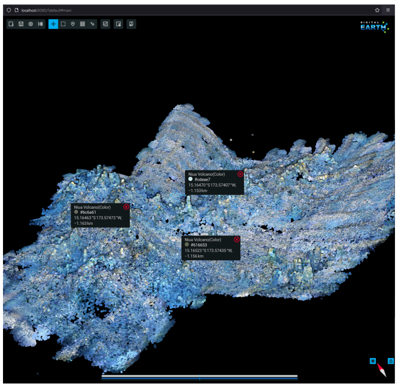
The software structure of the DLB is that of a full-stack web application. A frontend written in *TypeScript* using the *Angular* framework connects to a backend written in *Python* using the *Flask* framework via a REST API. The provenance persistence is handled by the Python library *pygit2* which offers bindings to *libgit2*. This means, projects in the DLB persist as repositories of the version control system Git instead of as a custom data format or e.g. a relational database. This decision was made as Git already offers much of the functionality that is desired from a provenance management system. It allows tracking of entities (e.g. files), actors (author, committer), and activities (adding, deleting, and modifying entities, committing, branching, merging etc.) [DNMV*13]. Furthermore, Git repositories can not only be represented as a directory on a local file system and thus also packaged and distributed, but also uploaded to a remote Git platform such as GitHub or GitLab and accessed there in similar fashion as a source code repository. The DLB allows to export the generated Git repository as a compressed zip file to a user-defined location, which in turn allows



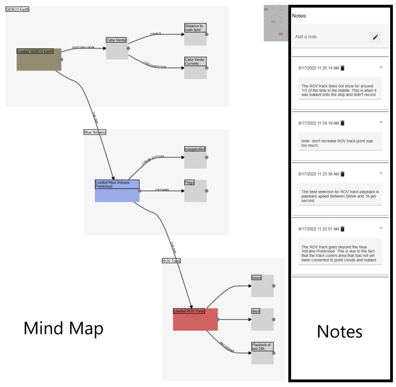
(a) *The dashboard view of the DLB allowing to adapt several settings before connecting to a visualization application including: i) credentials with which the git commits in the backend will be authored, b) the visualization applications to connect to, c) whether to create a new repository or connect to an existing one, d) the communication time interval.*



(b) *The visualization state history view of the DLB. It displays the provenance tree on the left and a linear view of the current branch's states on the bottom right side. Above this are displayed the list of branches in the backend repository and the component allowing to "replay" a selection of states.*



(c) *A screenshot of the Digital Earth Viewer displaying point cloud photogrammetry data of a deep sea black smoker vent field*



(d) *A section of the Mind Map component of the DLB showing the mind map and notes created by the user to externalize insights.*

Figure 2: *Components of the Digital Lab Book during an example session with the Digital Earth Viewer used as visualization application.*

the provenance package to be uploaded as supplementary material. Such a packaged project can also be imported to the DLB again.

4.2. Supported Applications

The DLB currently supports three different visualization applications or families of applications which a user can select in the interface of the DLB. Each of the supported applications is independent from the DLB in its usage which means for the actual interaction with the application a user would not have to change their workflow.

Digital Earth Viewer: The Digital Earth Viewer (DEV) is a web-based tool for visualizing and exploring geospatial data in real time, which is being developed at GEOMAR [BSM*22]. The DEV is usable inside the dome of the ARENA2 which allows its use in immersive analytics workflows. Communication between the DLB and the DEV happens via a websocket API which also handles the synchronization between different instances of the DEV.

Potree: Potree is a WebGL based application developed for rendering large point clouds [Sch16, SOW20]. It offers many different settings, and interaction- and measurement options for LAS/LAZ point clouds. Similar to the method used with the DEV, the DLB connects to a simple websocket API which we added to Potree's electron-based desktop application, and exchanges visualization state information. Potree is not yet integrated into the ARENA2.

Unreal Engine: A family of applications which are usable in the ARENA2 and can be supported by the DLB are immersive, scientific visualizations built in the Unreal game engine. Such game engines are increasingly being used for the purpose of scientific visualization [RCM*20] and allow a multitude of possibilities for immersive analytics. The communication with an Unreal Engine application happens via a UE Blueprint actor which has to be included in the game and which exposes functions that can be accessed using the Unreal Engine Remote Control API. This Blueprint actor can potentially be included in any visual analytics application built in Unreal.

4.3. Workflow Example

Figure 2 shows screenshots of an example session in which the user is working with the Digital Earth Viewer (DEV) as main visualization application and the Digital Lab Book as supporting provenance management tool.
The detailed workflow below mostly describes the user's interaction with the DLB and not specific interactions with the DEV. The DLB is implemented in a way that it is generally agnostic to which application it is used with.

- The user starts the Digital Earth Viewer (see 2c) and the Digital Lab Book executable files, opens two browser windows and navigates to the respective local web server addresses
- In the DLB UI, the user chooses which visualization application to connect to (currently DEV, Potree, or Unreal Engine), and clicks a button to connect to the chosen option and start the communication between the applications (see 2a)
- Additional settings are available to define a repository name and a user name to keep projects of different users separated, as well as the frequency in which the DLB sends requests (2a)
- The DLB now records the visualization state in user-defined intervals and takes screenshots of the DEV. Since the DLB is working with Git in the backend, these states are represented as Git commits with the diffs between the state data persisting as commit messages (see 2b)
- In the DLB UI, the user is presented with a list of the states including thumbnails of the scene, as well as the provenance history, which is a graph of the Git commits and branches (see 2b)
- From there on, the user can look at details of each state, add notes, trigger a simple return to a previous state, or alternatively trigger a return to a previous state that includes a branching action (see 2b)
- Additionally, the DLB offers a mind map functionality for external provenance capturing by explicit externalization of results by the user (i.e. associating and labelling different states, taking notes etc) [DRC10], which can be seen in Fig. 2d)

Included in the supplemental material is a short screencast video demonstrating the interaction with the DLB interface and the DEV, as well as the exported provenance package of this interaction. This provenance repository can also be viewed here, uploaded to our institute's GitLab: https://git.geomar.de/digital-lab-book/Prov-Demo-Repository.

5. Future Work

Additional to the implemented git2prov approach, we would like to support the W3C PROV model directly. This could be done by directly defining entities, actors, and activities during DLB runtime according to the PROV specifications.
The DLB is currently mainly recording provenance artifacts which do not contain information about the environment (e.g. 2D monitor, virtual reality head-mounted display, ARENA2) the visualization workflow is happening in. The spatially immersive environment of the ARENA2 was, however, essential to the original specification of the project. Therefore it is of importance to work on capturing the unique provenance artifacts that arise in such an immersive environment, including motion capturing and potentially eye tracking. Important as well is the collaborative nature of work in the ARENA2 which means the provenance tool also needs to support the provenance of collaborative work and not just single-person workflows. To evaluate the DLB, it will be examined in qualitative case studies with use-cases of ocean researchers at our institute, as well as during quantitative user studies.

6. Conclusion

This work introduces the Digital Lab Book, an application for the capturing, management, and visualization of provenance from visual analytics workflows. It has a special focus on supporting workflows of domain scientists in ocean research, and aims at supporting collaborative immersive analytics inside a unique spatially immersive environment, though not exclusively. We argue that making visual analytics and visualization creation processes transparent and traceable is good scientific practice, and the DLB is built to help with this endeavor. The DLB enables making recorded provenance accessible e.g. as supplementary material to publications. If the analyzed data, a static image of the visualization, and insights are being published, so should be the process that happened in between.
Having such a provenance capturing tool supports the visual analytics process itself by offering an interactive history of the visualization which can be used to compare and return to previous states and manage "What-If?"-scenarios by branching. It also supports especially asynchronous collaboration between analysts and it supports extended analysis processes by allowing to more easily pick up the work at a later point in time. The Digital Lab Book meets these requirements by being a light-weight and extensible web-based application which acts as a third-party to visualization applications, records provenance information of visual analytics workflows, allows interaction with the provenance history, and is able to save and export this information in an open non-proprietary format.
While the proposed use-cases at our institute aim at making use of the spatial immersion offered by the ARENA2, the DLB is designed to apply to non-immersive analytics workflows at first, with immersive analytics being a special case. Whether or not the DLB supports immersive analytics depends on whether the visualization application supports immersive analytics. Concrete recording of artifacts of immersive work is only experimentally implemented but will be an inclusion that can be enabled or disabled. The DLB will therefore not only be usable in an immersive scenario.
Similarly, the scientific domains in which the DLB is used and can be used are not predetermined. The currently supported applications generally focus on geospatial data, which not only includes ocean research but multiple other geosciences. Nevertheless, even this is not a hard requirement because it depends on the third-party visualization application. If the visualization application supports a different domain and supports the requirements of interacting with the DLB, the DLB is agnostic to the domain or data format.

Acknowledgements

Armin Bernstetter is a doctoral researcher funded through the Helmholtz School for Marine Data Science (MarDATA), Grant No. HIDSS-0005.
The initial program code with which the Digital Lab Book sends requests to the Digital Earth Viewer API is based on a code snippet by Valentin Buck, one of the Viewer's developers.

References

[20113] *PROV-Overview. An Overview of the PROV Family of Documents*. Project report, April 2013. URL: https://eprints.soton.ac.uk/356854/. 2

[BER*22] BLOCK J. E., ESMAEILI S., RAGAN E. D., GOODALL J. R., RICHARDSON G. D.: The influence of visual provenance representations on strategies in a collaborative hand-off data analysis scenario. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–11. doi:10.1109/TVCG.2022.3209495. 3

[BSM*22] BUCK V., STÄBLER F., MOHRMANN J., GONZÁLEZ E., GREINERT J.: Visualising geospatial time series datasets in realtime with the digital earth viewer. *Computers & Graphics 103* (2022), 121–128. doi:10.1016/j.cag.2022.01.010. 2, 4

[CFS*06] CALLAHAN S. P., FREIRE J., SANTOS E., SCHEIDEGGER C. E., SILVA C. T., VO H. T.: Vistrails: Visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 2006), SIGMOD '06, Association for Computing Machinery, p. 745–747. doi:10.1145/1142473.1142574. 2

[CGL20] CUTLER Z., GADHAVE K., LEX A.: Trrack: A library for provenance-tracking in web-based visualizations. In *2020 IEEE Visualization Conference (VIS)* (Oct. 2020), pp. 116–120. doi:10.1109/VIS47514.2020.00030. 3

[DNMV*13] DE NIES T., MAGLIACANE S., VERBORGH R., COPPENS S., GROTH P., MANNENS E., VAN DE WALLE R.: Git2prov: Exposing version control system content as w3c prov. In *Proceedings of the 12th International Semantic Web Conference (Posters & Demonstrations Track) - Volume 1035* (Aachen, DEU, 2013), ISWC-PD '13, CEUR-WS.org, p. 125–128. 3

[DRC10] DOU W., RIBARSKY W., CHANG R.: Capturing reasoning process through user interaction. In *EuroVAST 2010: International Symposium on Visual Analytics Science and Technology* (2010), Kohlhammer J., Keim D., (Eds.), The Eurographics Association. doi:10.2312/PE/EuroVAST/EuroVAST10/033-038. 1, 2, 5

[FDTWS19] FEGER S. S., DALLMEIER-TIESSEN S., WOŹNIAK P. W., SCHMIDT A.: The role of hci in reproducible science: Understanding, supporting and motivating core practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI EA '19, Association for Computing Machinery, p. 1–6. doi:10.1145/3290607.3312905. 1

[FW21] FIDLER F., WILCOX J.: Reproducibility of Scientific Results. In *The Stanford Encyclopedia of Philosophy*, Zalta E. N., (Ed.), Summer 2021 ed. Metaphysics Research Lab, Stanford University, 2021. URL: https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/. 1

[HLB*08] HOWE B., LAWSON P., BELLINGER R., ANDERSON E., SANTOS E., FREIRE J., SCHEIDEGGER C., BAPTISTA A., SILVA C.: End-to-end escience: Integrating workflow, query, visualization, and provenance at an ocean observatory. In *2008 IEEE Fourth International Conference on eScience* (Dec. 2008), pp. 127–134. doi:10.1109/eScience.2008.67. 3

[Kin06] KINDLMANN G.: Lack of reproducibility hinders visualization science. In *Is there science in visualization (Panel)* (2006). Jankun-Kelly, TJ and Kosara, Robert and Kindlmann, Gordon and North, Chris and Ware, Colin and Bethel, E Wes. IEEE Visualization Conference Compendium. 1

[KNS04] KREUSELER M., NOCKE T., SCHUMANN H.: A history mechanism for visual data mining. In *IEEE Symposium on Information Visualization* (Oct. 2004), pp. 49–56. doi:10.1109/INFVIS.2004.2. 3

[KSK*23] KWASNITSCHKA T., SCHLÜTER M., KLIMMECK J., BERNSTETTER A., GROSS F., PETERS I.: Spatially Immersive Visualization Domes as a Marine Geoscientific Research Tool. In *Workshop on Visualisation in Environmental Sciences (EnvirVis)* (2023), Dutta S., Feige K., Rink K., Zeckzer D., (Eds.), The Eurographics Association. 2

[MBC13] MISSIER P., BELHAJJAME K., CHENEY J.: The w3c prov family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology* (New York, NY, USA, 2013), EDBT '13, Association for Computing Machinery, p. 773–776. doi:10.1145/2452376.2452478. 1

[NC14] NANCEL M., COCKBURN A.: Causality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Apr. 2014), ACM. doi:10.1145/2556288.2556990. 3

[PC05] PIROLLI P., CARD S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (2005), vol. 5, McLean, VA, USA, pp. 2–4. 1

[Pen11] PENG R. D.: Reproducible research in computational science. *Science 334*, 6060 (2011), 1226–1227. doi:10.1126/science.1213847. 1

[RCM*20] REINA G., CHILDS H., MATKOVIĆ K., BÜHLER K., WALDNER M., PUGMIRE D., KOZLÍKOVÁ B., ROPINSKI T., LJUNG P., ITOH T., GRÖLLER E., KRONE M.: The moving target of visualization software for an increasingly complex world. *Computers & Graphics 87* (2020), 12–29. doi:10.1016/j.cag.2020.01.005. 4

[RESC16] RAGAN E. D., ENDERT A., SANYAL J., CHEN J.: Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 31–40. doi:10.1109/TVCG.2015.2467551. 2

[Sch15] SCHLITZER R.: Data analysis and visualization with ocean data view, 2015. URL: https://epic.awi.de/id/eprint/37570/. 1

[Sch16] SCHÜTZ M.: *Potree: Rendering Large Point Clouds in Web Browsers*. Master's thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/E193-02, A-1040 Vienna, Austria, Sept. 2016. URL: https://www.cg.tuwien.ac.at/research/publications/2016/SCHUETZ-2016-POT/. 4

[SGP*18] STITZ H., GRATZL S., PIRINGER H., ZICHNER T., STREIT M.: Knowledgepearls: Provenance-based visualization retrieval. *IEEE Transactions on Visualization and Computer Graphics (VAST '18) 25*, 1 (2018), 120–130. doi:10.1109/TVCG.2018.2865024. 3

[SOW20] SCHÜTZ M., OHRHALLINGER S., WIMMER M.: Fast out-of-core octree generation for massive point clouds. *Computer Graphics Forum 39*, 7 (Nov. 2020), 1–13. doi:10.1111/cgf.14134. 4

[Vaz18] VAZIRE S.: Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science 13*, 4 (2018), 411–417. PMID: 29961410. doi:10.1177/1745691617751884. 1

[vW05] VAN WIJK J. J.: The value of visualization. In *IEEE Visualization* (2005), IEEE Computer Society, pp. 79–86. doi:10.1109/VISUAL.2005.1532781. 2

[XAJK*15] XU K., ATTFIELD S., JANKUN-KELLY T., WHEAT A., NGUYEN P. H., SELVARAJ N.: Analytic provenance for sensemaking: A research agenda. *IEEE Computer Graphics and Applications 35*, 3 (May 2015), 56–64. doi:10.1109/MCG.2015.50. 2

[XOW*20] XU K., OTTLEY A., WALCHSHOFER C., STREIT M., CHANG R., WENSKOVITCH J.: Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum 39*, 3 (2020), 757–783. doi:10.1111/cgf.14035. 2

[ZC23] ZIEGLER P., CHASINS S. E.: A need-finding study with users of geospatial data. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Apr. 2023), ACM. doi:10.1145/3544548.3581370. 2