

EGU2020-9251

<https://doi.org/10.5194/egusphere-egu2020-9251>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



## An integrative framework for data-driven investigation of environmental systems

**Daniel Eggert** and Doris Dransch

German Research Center for GeoSciences GFZ, Potsdam, Germany

Environmental scientists aim at understanding not only single components but systems, one example is the flood system; scientists investigate the conditions, drivers and effects of flood events and the relations between them. Investigating environmental systems with a data-driven research approach requires linking a variety of data, analytical methods, and derived results.

Several obstacles exist in the recent scientific work environment that hinder scientists to easily create these links. They are distributed and heterogeneous data sets, separated analytical tools, discontinuous analytical workflows, as well as isolated views to data and data products. We address these obstacles with the exception of distributed and heterogeneous data since this is part of other ongoing initiatives.

Our goal is to develop a framework supporting the data-driven investigation of environmental systems. First we integrate separated analytical tools and methods by the means of a component-based software framework. Furthermore we allow for seamless and continuous analytical workflows by applying the concept of digital workflows, which also demands the aforementioned integration of separated tools and methods. Finally we provide integrated views of data and data products by interactive visual interfaces with multiple linked views. The combination of these three concepts from computer science allows us to create a digital research environment that enable scientists to create the initially mentioned links in a flexible way. We developed a generic concept for our approach, implemented a corresponding framework and finally applied both to realize a "Flood Event Explorer" prototype supporting the comprehensive investigation of a flood system.

In order to implement a digital workflow our approach intends to precisely define the workflow's requirements. We mostly do this by conducting informal interviews with the domain scientists. The defined requirements also include the needed analytical tools and methods, as well as the utilized data and data products. For technically integrating the needed tools and methods our created software framework provides a modularization approach based on a messaging system. This allows us to create custom modules or wrap existing implementations and tools. The messaging system (e.g. pulsar) then connects these individual modules. This enables us to combine multiple methods and tools into a seamless digital workflow. The described approach of course demands the proper definition of interfaces to modules and data sources. Finally our software framework provides multiple generic visual front-end components (e.g. tables, maps and charts) to create interactive linked views supporting the visual analysis of the workflow's data.

