
V-FOR-WaTer - a virtual research environment for environmental research

Marcus Strobl¹, Elnaz Azmi¹, Sibylle K. Hassler², Mirko Mälicke², Jörg Meyer¹, Achim Streit¹, and Erwin Zehe²

¹Steinbuch Centre for Computing, Karlsruhe Institute of Technology

²Institute of Water and River Basin Management, Chair of Hydrology, Karlsruhe Institute of Technology

Extent and diversity of environmental data are continuously increasing due to more sensor networks with higher spatial and temporal resolution. To find appropriate data for analyses and especially for large scale models and simulations in this data explosion can take up to several months. The preprocessing of these heterogeneous datasets from different research disciplines to acquire a coherent dataset, can be done with a wide range of algorithms and tools. The outcome is a base dataset that is not reproducible and in consequence, neither are the resulting analyses [3, 9]. The datasets therefore do not obey the FAIR principles [13]. The V-FOR-WaTer web portal [11] aims to improve this situation by collecting data and metadata from a wide variety of sources and by offering preprocessed data.

1 Objectives

Huge effort is made to improve the availability of data by establishing data portals and data repositories. They range from project-specific data portals, that provide access to project datasets, via data repositories for already published datasets (e.g. GFZ data services, PANGAEA), through to data portals provided from federal state offices (e.g. LUBW, USGS, NASA). However, most of these portals do not allow for proper preprocessing of heterogeneous data to prepare scientific analyses.

The V-FOR-WaTer web portal represents an enhancement for data repositories to facilitate a standardized generation of reproducible base datasets. We want to give direct access to data and metadata to some of the most important and promising data sets for environmental research, that can be scaled and preprocessed with data uploaded from the scientists. These preprocessed datasets can be downloaded to the user's local PC or processed online using tools within the portal, provided either by V-FOR-WaTer or from

Das hier beschriebene Poster ist in der Open Access-Plattform der Universität Heidelberg heiDOK unter der DOI: <https://doi.org/10.11588/heidok.00029717> veröffentlicht.

the scientific community. For efficiently working in and with the portal, we offer well documented open source tools and a simple workflow for the scientists, which can be saved and restored to perform reproducible preprocessing and analyses. All these components constitute the virtual research environment of V-FOR-WaTer.

Scientists can also upload their own data in combination with the corresponding metadata. The metadata is used for filtering, for tools to restrict which and how datasets can be combined and can be utilized to simplify data publication in established data repositories. To secure new and yet unpublished data sets, we have a fine-grained access management that enables scientists to upload data with an embargo period.

2 The V-FOR-WaTer Web Portal

The web portal is being developed with JavaScript and the Python web framework Django and is published open source [10]. These technologies are widely used and actively maintained by a large community and promise fast bugfixes and a long service life. The focus of the front end is on simple and intuitive usability. The design follows classical geographic information systems (GIS) with a map as main element on the start page (Fig. 1). Filtering of data takes place on the map and with a filter menu in the sidebar. To use data with restricted access, users can send requests to the data owner through the portal. For the identity management we use the external tool B2ACCESS [1], that is provided by EUDAT. With this identity provider researchers can reuse existing federated accounts from their home universities to log in.

The original database was designed with focus on the data sets of the Catchments as Organized Systems (CAOS) project [14]. These data sets are of special interest due to the large amount of data and their heterogeneity that represent a wide range of data used by the hydrological community. By now the metadata model is in its second version and it contains the necessary flexibility to hold all data types necessary for water and terrestrial environmental research and is continuously adapted to the needs of new data. Besides the CAOS data we already integrated data from the hydrology group of the Institute of Water and River Basin Management of KIT and from the Landesanstalt für Umwelt Baden-Württemberg (LUBW) [6]. The integration of data from more projects is work in progress. The data can be downloaded from the web portal in formats that are commonly used in environmental science, and the accompanying metadata follows the international standards of INSPIRE [4] and ISO19115 [5].

Instead of downloading, the data can also be processed within the web portal. For more flexibility to access the tools we use separate packages that are connected through a web processing server (WPS) [12], and not implemented as static part of the V-FOR-WaTer code. This way the selection of tools can easily be extended, and access is possible through the web portal or through an API from a local PC as well. The toolbox is constantly growing and already comprises among others the geostatistical toolbox SciKit GStat [7, 8]. Tools for scaling and uncertainty quantification for evapotranspiration (ET)

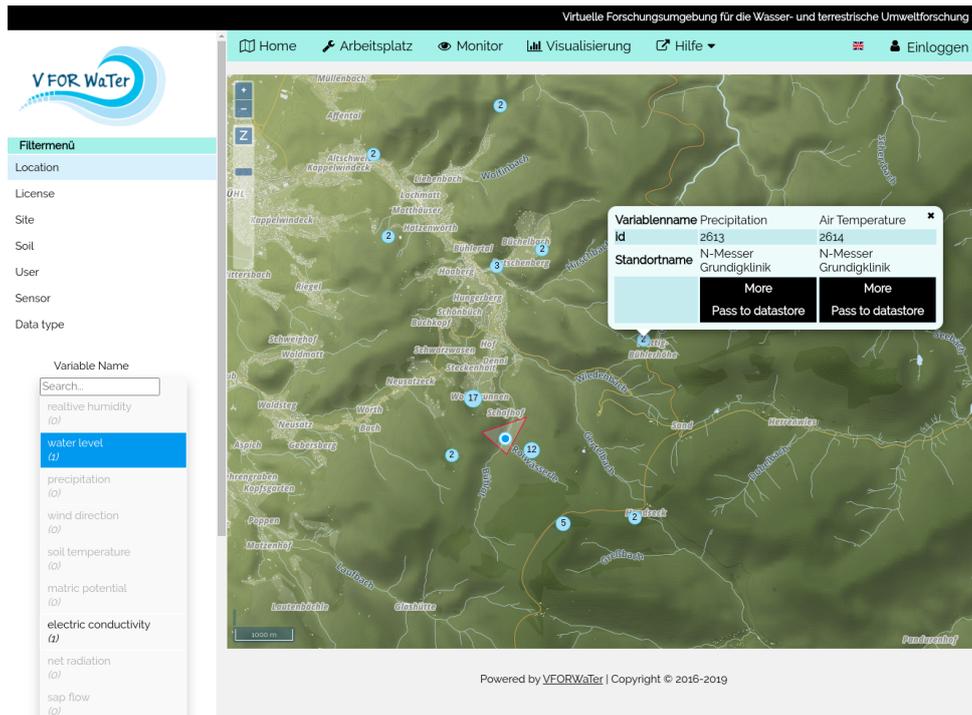


Figure 1: Screenshot of map and filter menu.

data are in development and will also be integrated in the V-FOR-Water portal. For complex workflows connecting several tools a graphical workflow editor, based on the JavaScript application framework Draw2D [2], is being integrated in the portal.

3 Conclusion

The V-FOR-WaTer web portal offers scientists centralized access to relevant data and tools, hence strongly supports them with their search, preparation, analysis and publication of data. The use of the web portal accelerates science and helps to make science reproducible.

Acknowledgments

V-FOR-WaTer has been funded by the ministry of science, research and arts of the state of Baden-Wuerttemberg, Germany.

Bibliography

- [1] B2ACCESS Service, accessed April 27, 2021, <https://eudat.eu/services/b2access>.
- [2] Draw2D, accessed May 21, 2021, <https://www.draw2d.org/draw2d>.
- [3] Hutton, Christopher, Thorsten Wagener, Jim Freer, Dawei Han, Chris Duffy, and Berit Arheimer. “Most computational hydrology is not reproducible, so is it really science?” *Water Resources Research* 52.10 (2016), 7548–7555. DOI: <https://doi.org/10.1002/2016WR019285>.
- [4] Infrastructure for spatial information in Europe (INSPIRE), accessed July 13, 2021, <https://inspire.ec.europa.eu/inspire-directive/2>
- [5] ISO 19115-1:2014 Geographic information — Metadata, accessed July 13, 2021, <https://www.iso.org/standard/53798.html>
- [6] Landesanstalt für Umwelt Baden-Württemberg (LUBW), accessed May 21, 2021, <https://www.lubw.baden-wuerttemberg.de/startseite>.
- [7] Mälicke, Mirko, Helge D. Schneider, Sebastian Müller, and Egil Möller. ”mmaelicke/scikit-gstat: A scipy flavoured geostatistical variogram analysis toolbox (Version v0.5.0)” Zenodo (April 20, 2021). DOI: <https://doi.org/10.5281/zenodo.1345584>
- [8] Müller, Sebastian and Lennart Schüler. “GeoStat-Framework/GSTools: v1.3.0 ’Pure Pink” Zenodo (April 14, 2021). DOI: <https://doi.org/10.5281/zenodo.4687075>
- [9] Stagge, James H., David E. Rosenberg, Adel M. Abdallah, Hadia Akbar, Nour A. Attallah, and Ryan James. “Assessing data availability and research reproducibility in hydrology and water resources” *scientific data* 6 (2019), 190030. DOI: <https://doi.org/10.1038/sdata.2019.30>.
- [10] V-FOR-WaTer project on github, accessed April 27, 2021, <https://github.com/VForWaTer/vforwater-portal>.
- [11] V-FOR-WaTer web portal demo instance, accessed April 27, 2021, <https://portal.vforwater.de>.
- [12] Web Processing Service, accessed April 27, 2021, <https://www.opengeospatial.org/standards/wps>.
- [13] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. ”The FAIR Guiding Principles for scientific data management and stewardship.” *scientific Data* 3 (2016), 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>

- [14] Zehe, Erwin, Uwe Ehret, L. Pfister, T. Blume, B. Schröder, M. Westhoff, C. Jackisch et al. “HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments.” *Hydrology and Earth System Sciences* 18, 11 (2014), 4635–4655. DOI: <https://doi.org/10.5194/hess-18-4635-2014>