Application of open hydro-meteorological data for web-based analysis and processing

Stefan Wiemann
Technische Universität Dresden
Chair of Geoinformatics
Dresden, Germany
stefan.wiemann@tu-dresden.de

Abstract

The work described in this paper aims to facilitate the application and processing of open spatiotemporal data on the web, with particular focus on hydro-meteorological data. The trend towards open government data on the web leads to a rapid increase of available data sources with great potential for environmental analysis and decision-making. First, the paper addresses the data preparation phase required to provide flexible means to access and analyse data in both spatial and temporal dimension. Second, the paper describes an information system that exploits the enhanced access means to provide customizable interfaces for ad hoc data visualization, retrieval and processing. It combines selected hydro-meteorological open data sources, more specifically ground radar data from Germany's National Meteorological Service (DWD) and stationary river discharge data provided by local authorities, for the purpose of flash flood analysis, monitoring and prediction.
Keywords: information system, web processing, hydro-meteorological data, radar precipitation.

1 Introduction and Motivation

Since hydro-meteorological processes have a huge impact on human life, it does not come as a surprise that corresponding measurements have already been documented for over a century. Access to historic and current hydro-meteorological measurements is a prerequisite for the analysis, assessment and prediction of various environmental phenomena, such as precipitation (e.g. Kirstetter *et al.*, 2015), river floods (e.g. Krajewski *et al.*, 2017) or air pollution (e.g. Rolph, Stein & Stunder, 2017). In recent years, more and more of this data becomes available as open data to the public, due to the growing number of open government initiatives. Even though there are still various issues connected to this process (see Attard *et al.*, 2015), it already provides a valuable source for spatiotemporal information retrieval, analysis and decision making.

In Europe, one of the major drivers for open environmental data is the INSPIRE directive, which requires the provision of essential hydro-meteorological measurements by the responsible authorities in compliance with the Observation and Measurements standard (INSPIRE, 2013). Another prominent example in the context of this study is the amendment to the *Deutscher Wetterdienst Act* in force since 25 July 2017, which induced the German National Meteorological Service to offer a huge amount of their data, products and services on their open data platform[1].

Besides the provision of open data, open capabilities for ad hoc data analysis and processing are required to support effective information retrieval and decision-making. Here, the trend is towards distributed web-based environments, which allow for thin clients and ubiquitous access to data and information via well-defined interfaces, ideally based on existing software solutions (Swain *et al.*, 2015). Especially in the field of weather forecasting, there are plenty of applications available for the general public. However, professionals often demand for access to underlying raw data, metadata as well as customizable data analysis and processing capabilities, all of which are rarely offered.

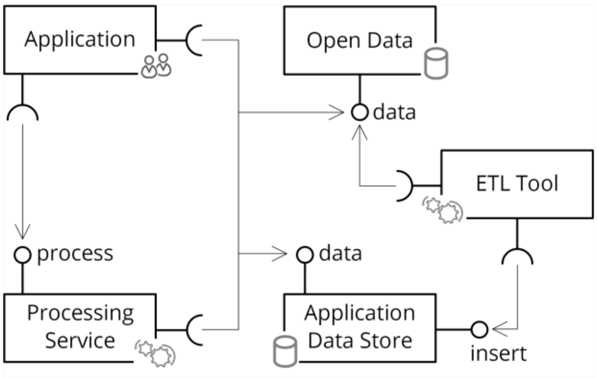
This article presents an approach to combine open hydro-meteorological data sources for information retrieval on the web. Section 2 describes the creation of an intermediate storage for faster and more flexible access to the data. Subsequently, section 3 introduces the components of the envisaged information system. Based on this, section 4 describes an application use cases for the assessment of flash flood events. The conclusion in section 5 summarizes the findings of the presented study and provides a brief perspective on further research issues.

2 Data preparation

Leaving aside the non-technical aspects, the consumption of distributed open data is often hampered by the lack of common data formats, data structures and provision means (Attard *et al.*, 2015). The data is, usually on purpose, not specially processed to support a wide range of potential applications and to reduce the costs for data provision. Moreover, simple data downloads from websites is still the predominant means for the provision of open data on the web.

[1] https://opendata.dwd.de

Figure 1: Components for application-oriented data preparation and provision.



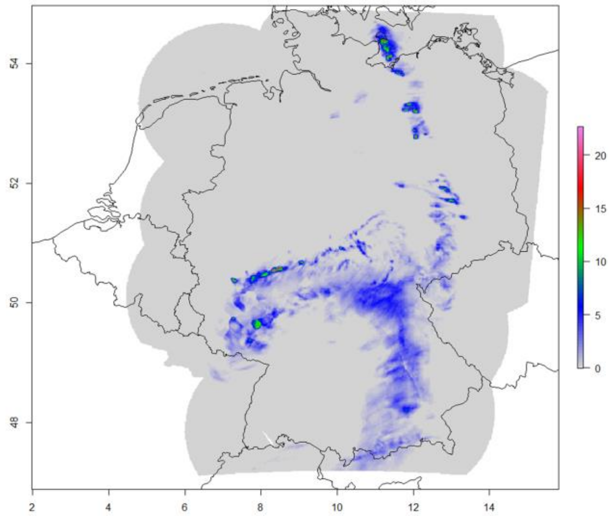
Therefore, an intermediate data storage is required to provide means for fast data access and processing. Even though this leads to redundant data storage and necessary synchronization efforts, it significantly improves the performance of data-intensive applications.

The typical components for the creation and use of an intermediate storage for open data are shown in Figure 1. An ETL (Extract Transform Load) tool downloads relevant data from the original open data platform once or in a regular interval, tailors the data to an application and finally uploads the data to a target data store. While the data content largely remains unchanged, its provision is optimized for certain applications, e.g. in terms of fast data access, enhanced query options or extended metadata. Different applications and services can now access the data for visualization, analysis and processing.

The focus for data preparation in this study is on ground radar data for precipitation measurement, more specifically on a number of RADOLAN (Radar Online Adjustment) products, provided by the DWD through its open data platform. While current data is provided using a custom binary format for each time step, historic data is stored using the same format, but in zip archives. Both are available as file downloads, which makes it hard to request large amounts of data on demand within a reasonable amount of time. To enhance data access and facilitate direct filtering in both spatial and temporal domain, an array-based data structure serves as an intermediate application data store.

Two approaches are implemented and evaluated: the NetCDF format and a single instance of the array-database SciDB. As data source, the RADOLAN RW product is used. It provides quality-checked precipitation data based on radar reflectivity adjusted with in situ precipitation measurements, with a spatial resolution of one square kilometre (total 900x900 km) and a temporal resolution of one hour. Whereas historic data is downloaded in archives and transformed once, real-time data is directly synchronized upon provision by the DWD open data platform. A sample RADOLAN image is shown in Figure 2. In total, those are ~8760 raw binary files (~13GB) per year. The setup of both NetCDF and SciDB, primarily adressing chunk size and compression, is optimized to achieve reasonable response times for both time slice and time series requests. However, a higher priority is given to time series, as they are requested more frequently for the estimation of hydro-meteorological conditions in a particular

Figure 2: RADOLAN image (RW Product), 2016-07-21 07:50 UTC, units=mm/h.



area. The results of this study show that a single SciDB instance could not perform close to NetCDF, which is also in line with previous findings by (Liu *et al.*, 2016). While the performance for the retrieval of whole image time slices is comparable, NetCDF is found to be considerably faster when querying long time series for a particular location. To facilitate further testing and evaluation, both approaches are maintained. However, the NetCDF approach is currently the preferred solution for data querying, analysis and processing in the information system.

A similar data preparation procedure is applied to the second data source used in this study: stationary hydro-meteorological measurements. In this case, the data is transformed and provided using the OpenSensorWeb platform[2], which acts as a proxy to a number of open sensor networks on the web using the Elasticsearch API. The platform inter alia includes a large number of official meteorological and river monitoring measurements, which are provided by the responsible governmental authorities.

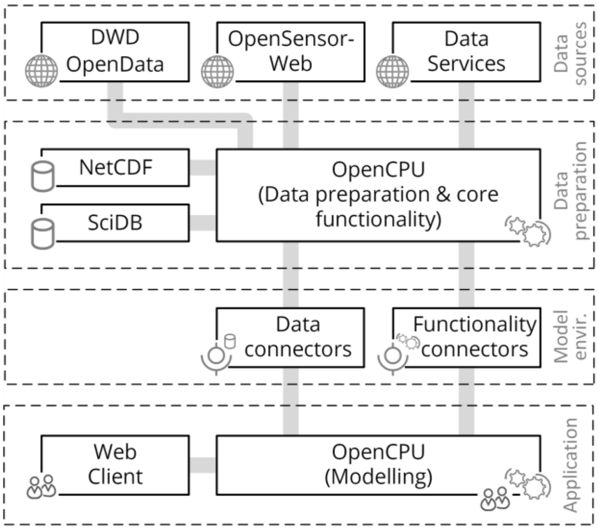
3 Information System

The main idea of the information system is to provide means to access and utilize open hydro-meteorological data on the web.

As shown in Figure 3, the core of the information system is the OpenCPU platform[3] providing a HTTP API to R, which is quite popular within both the meteorological and the hydrological domain. In this case, it is used to provide an environment to connect and expose hydro-meteorological models online. Therefore, it provides accessible connectors for both data and functionality. The corresponding R package implemented for this study builds on top of existing packages for spatiotemporal data handling and is available online as open source[4].

[2] https://opensensorweb.de
[3] https://www.opencpu.org
[4] https://github.com/GeoinformationSystems/xtruso_R

Figure 3: main components of the described information system.



Data connectors provide access to underlying data sources in its original or preprocessed form. In the current implementation, stationary measurements from the OpenSensorWeb platform and contextual data from Web Feature Services (WFS) are directly accessed and filtered via their respective interfaces. Access to ground radar data is supported by three different means: 1) direct download from the DWD open data platform, 2) querying a NetCDF file on the server and 3) querying a local or remote SciDB instance.

Functionality connectors on the other hand provide means to derive data and information from the underlying data sources in a way that they become more useful to downstream applications and services. The current implementation covers inter alia:

- Time series extraction from radar data based on either single coordinates or catchment areas. While the first results in precipitation time series, the latter additionally produces zonal statistics from the corresponding radar measurements.
- Disaggregation of hourly radar precipitation data (RADOLAN RW product) using 5-minute radar reflectivity data (RADOLAN RX product).
- Computation of zonal statistics from continuous raster inputs, such as radar precipitation data or elevation models, within an area of interest.
- Computation of percentage distribution for classified raster inputs, such as land cover, within an area of interest.
- Descriptive statistical analysis of measurement time series for the derivation of hydro-meteorological indices.

Both data and functionality connectors are used to dynamically interlink and execute hydro-meteorological models within the information system. As shown in Figure 4, the corresponding model integration is composed of two steps. First, an expert interactively registers, calibrates and parameterizes a model based on historic measurement time series and available contextual data. Subsequently, this parameterized model can be executed by a user on demand, using either historic or real-time measurement data. Both steps, model registration and model application, are supported through the provision of corresponding R functions, which can be accessed either directly via the OpenCPU platform or indirectly using the built-in JavaScript API.

For the invocation of a registered model by a user or application, a reasonable degree of process granularity and interface complexity is required to facilitate its application and transferability. The provided data and functionality connectors support this requirement by means to pre-parameterize registered models, e.g. with original or preprocessed hydro-meteorological data sources and contextual data, and thus limit the required inputs to the most essential.

4 Application

To demonstrate the feasibility of the above-described approach, the information system is prototypically implemented to provide historic and current hydro-meteorological information for the assessment of flash flood risks. The current application focuses on small and medium sized catchments in the study area of Saxony, Germany. As depicted in Figure 5, it is based on a three-step workflow. First, the general hydro-meteorological status of a catchment area is determined in a regular interval. This includes essential pre-conditions, such as soil moisture and predicted rainfall. If a critical state is reached, the update interval is increased and

Figure 4: Registration and application of model implementations within the information system.

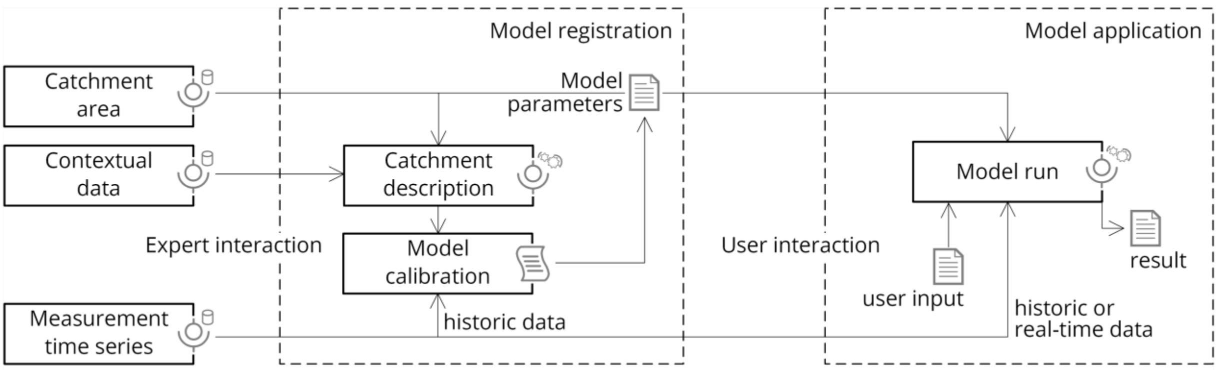
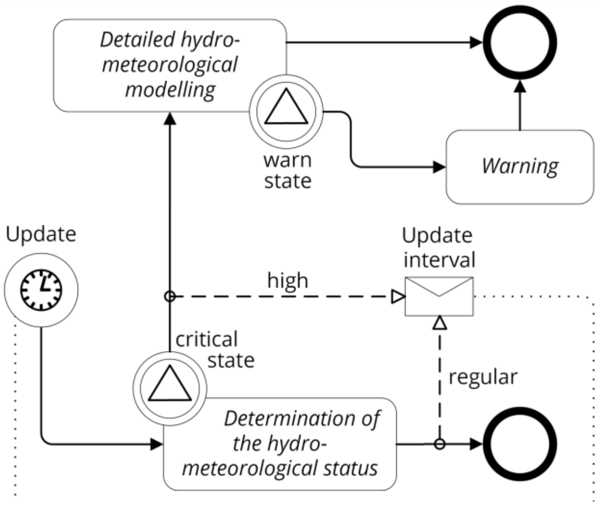


Figure 5: workflow for the assessment of flash flood risks (Business Process Modeling Language Notation).



a detailed hydro-meteorological modelling is conducted to assess the actual risk of a flash flood in the corresponding catchments. This is done by either conceptual or data-driven modelling, both able to estimate potential discharge and flood prone areas. If it exceeds a certain warning threshold, a corresponding warning is issued to responsible decision makers or peoples at risk.

For data visualization, the implemented web client uses OpenLayers[5] to visualize maps served by time-enabled Web Map Services (WMS) and to render vector features from Web Feature Services (WFS). Time series graphs are generated using the D3 JavaScript library[6]. The client also uses the provided Javascript API provided by the underlying OpenCPU platform to access data and functionality. Figure 6 shows a sample screenshot of the current application.

With respect to the main application workflow shown in Figure 5, the following models are currently registered and provided by the information system:

- Historic and current soil moisture estimation (high, medium, low) based on thresholds for the accumulated 24h and 72h rainfall data for both ground radar and stationary measurements. As input, the user must select a catchment and a timestamp, and can optionally define a custom threshold.
- Estimation of river discharge (high, medium, low) using the current discharge, 2h precipitation forecast data from ground radar reflectivity (RADOLAN FX product) and a runoff coefficient. The latter is predefined based on the historic rainfall-discharge ratio. As input, the user must only select a catchment.
- Determination of historic and current soil moisture condition based on historic precipitation, temperature discharge, global radiation, temperature and air pressure time series using an R implementation of the BROOK90 hydrologic model (Federer, Vörösmarty & Fekete, 2003). The current parameterization is based on the work conducted by (Luong *et al.*, 2017).

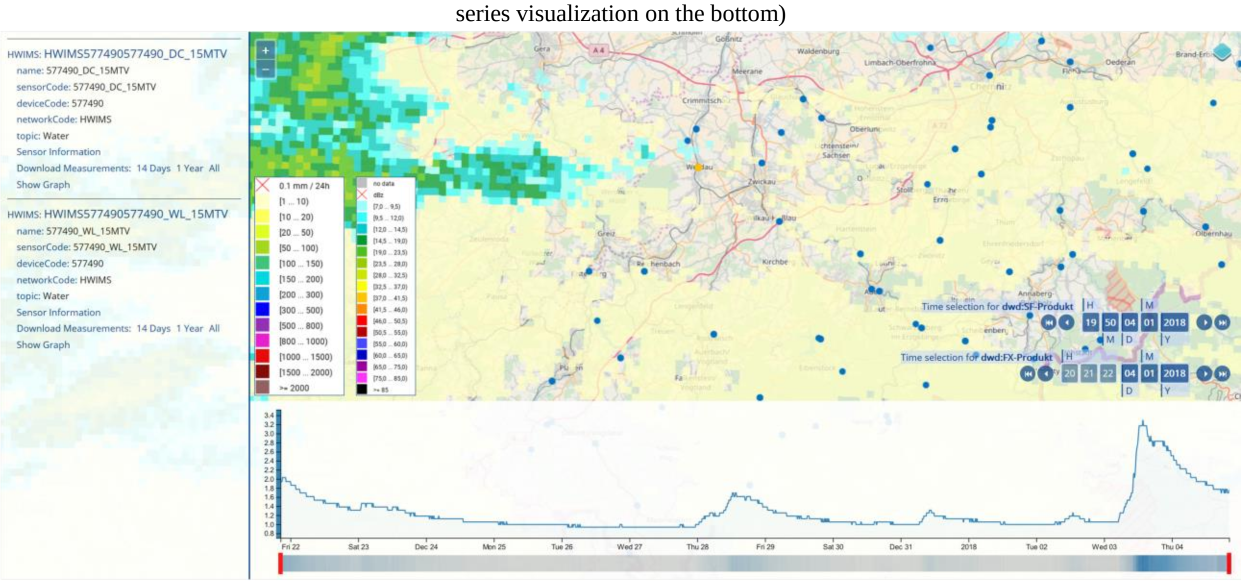
As described in the previous section, all of the registered models are using selected data and functionality connectors to pre-parameterize individual model runs. Since the areas of interest are primarily the catchments, the zonal statistics functionality is frequently used. The same is true for the provided hydro-meteorological data sources: precipitation from ground radar and stationary measurements.

Currently, the information system is mainly used to support the ex post identification and analysis of historical flash flood events. However, it can also provide first estimates on the hydro-meteorological condition of catchments in the same time interval as data is provided through the connected open data portals.

5 Conclusion

The presented information system not only allows for accessing and filtering open hydro-meteorological data

Figure 6: Screenshot of the client application (map with legend and WMS time selector; feature information on the left; time series visualization on the bottom)



sources on the web, but also enables the ad hoc analysis and processing of this data for information retrieval and decision making using built-in or custom R functionality. To extent the scope of the application, further data sources will be integrated in the near future, including the DWD COSMO-DE product for short-term meteorological forecasting (up to 72h) and precipitation measurements collected by low cost sensors. Moreover, additional models will be added to eventually enable the generation of ensemble outputs for both the determination of the hydro-meteorological status and the discharge modelling. Since single measurements and modelling results need to be considered inherently uncertain (cf. Abbaspour *et al.*, 2015; Cecinati *et al.*, 2017), a combination of multiple input sources and models will also facilitate the assessment and feedback of uncertainty information.

The conceptualized information system addresses a number of different user groups. While decision makers and interested citizens can use the provided, pre-configured data and functionality for information retrieval, scientists and modellers can leverage the power of R within the provided environment for advanced statistical analysis and adaptive modelling.

Even though R represents a very popular means for hydro-meteorological modelling, the current restriction to R functions has its limits in terms of interoperability. Further abstraction is accordingly required to allow for model integration using standardized interfaced for web-based data processing, such as the Web Processing Service (Castronova, Goodall & Elag, 2013). In addition, while many authoritative open data sources are readily available on the web, the spatial density of traditional sensor networks is rather low. Thus, crowdsourcing becomes an increasingly important source for hydro-meteorological measurements (Muller *et al.*, 2015). Further research will thus focus on the collection, quality assessment and assimilation of crowdsourced observations within the modelling workflows conducted through the information system.

References

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., et al. (2015) A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology*. [Online] 524, 733–752. Available from: doi:10.1016/J.JHYDROL.2015.03.027.

Attard, J., Orlandi, F., Scerri, S. & Auer, S. (2015) A systematic review of open government data initiatives. *Government Information Quarterly*. [Online] 32 (4), 399–418. Available from: doi:10.1016/J.GIQ.2015.07.006.

Castronova, A.M., Goodall, J.L. & Elag, M.M. (2013) Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. *Environmental Modelling & Software*. [Online] 41, 72–83. Available from: doi:10.1016/j.envsoft.2012.11.010.

Cecinati, F., Rico-Ramirez, M.A., Heuvelink, G.B.M. & Han, D. (2017) Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach. *Journal of Hydrology*. [Online] 548, 391–405. Available from: doi:10.1016/j.jhydrol. 2017.02.053.

Federer, C.A., Vörösmarty, C. & Fekete, B. (2003) Sensitivity of Annual Evaporation to Soil and Root Properties in Two Models of Contrasting Complexity. *Journal of Hydrometeorology*. [Online] 4 (6), 1276–1290. Available from: doi:10.1175/1525-7541(2003)004 <1276:SOAETS>2.0.CO;2.

INSPIRE (2013) *D2.8.III.13-14 Data Specification on Atmospheric Conditions and Meteorological Geographical Features – Technical Guidelines*.

Kirstetter, P.-E., Gourley, J.J., Hong, Y., Zhang, J., et al. (2015) Probabilistic precipitation rate estimates with ground-based radar networks. *Water Resources Research*. [Online] 51 (3), 1422–1442. Available from: doi:10.1002/2014WR015672.

Krajewski, W.F., Ceynar, D., Demir, I., Goska, R., et al. (2017) Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the American Meteorological Society*. [Online] 98 (3), 539–554. Available from: doi:10.1175/BAMS-D-15-00243.1.

Liu, H., Van Oosterom, P., Hu, C. & Wang, W. (2016) Managing Large Multidimensional Array Hydrologic Datasets: A Case Study Comparing NetCDF and SciDB. In: *Procedia Engineering*. [Online]. 1 January 2016 Elsevier. pp. 207–214. Available from: doi:10.1016/ j.proeng.2016.07.449.

Luong, T.T., Kronenberg, R., Bernhofer, C., Al Janabi, F., et al. (2017) Comparative estimation and assessment of initial soil moisture conditions for Flash Flood warning in Saxony. In: *Geophysical Research Abstracts*. 2017 p. 3728.

Muller, C.L., Chapman, L., Johnston, S., Kidd, C., et al. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*. [Online] 35 (11), 3185–3203. Available from: doi:10.1002/joc.4210.

Rolph, G., Stein, A. & Stunder, B. (2017) Real-time Environmental Applications and Display sYstem: READY. *Environmental Modelling and Software*. [Online] 95, 210–228. Available from: doi:10.1016/ j.envsoft.2017.06.025.

Swain, N.R., Latu, K., Christensen, S.D., Jones, N.L., et al. (2015) A review of open source software solutions for developing water resources web applications. *Environmental Modelling & Software*. [Online] 67, 108–117. Available from: doi:10.1016/j.envsoft.2015.
01.014.