Graph-based river network analysis for rapid discovery and analysis of linked hydrological data

Matt Fry¹ and Jan Rosecký²

¹UK Centre for Ecology and Hydrology, Water Resources, United Kingdom of Great Britain and Northern Ireland (mfry@ceh.ac.uk)
²Epimorphics Ltd

Hydrological analyses generally require information from locations across a river system, and knowledge on how these locations are linked within that system. Hydrological monitoring data e.g. from sensors or samples of the status of river flow and water quality, and datasets on factors influencing this status e.g. sewage treatment input, riparian land use, lakes, abstractions, etc., are increasingly available as open datasets, sometimes via web-based APIs. However, retrieving information, for data discovery or for direct analysis, based on location within the river system is complex, and is therefore not a common feature of APIs for hydrological data.

We demonstrate an approach to extracting datasets based on river connectivity using a digital river network for the UK, converted to a directed graph, and the python networkX package. This approach enables very rapid identification of upstream and downstream reaches and features for sites of interest, with speeds suitable for on-the-fly analysis. We describe how such an approach could be deployed within an API for data discovery and data retrieval, and demonstrate linking data availability information, capturing observed properties and time series metadata, from large sensor networks, in a JSON-LD format based on concepts drawn from SSN/SOSA and INSPIRE EMF. This approach has been applied to identify up- and downstream water quality monitoring sites for lakes within the UK Lakes Database for nutrient retention analysis, and production of hierarchical datasets of river flow gauging stations to aide network understanding.