FEATURED COLLECTION INTRODUCTION: OPEN WATER DATA INITIATIVE

Jerad Bales


INTRODUCTION

Large-scale data sharing has been an accepted practice in the scientific community since at least 1873, when an international standard for weather observation was adopted by the Vienna Congress (Sieber, 2015). Nevertheless, open data sharing is not universally practiced. For example, in one survey more than 20% of doctoral students in life sciences were denied access to information, data, materials, or code associated with published research (Vogeli et al., 2006).

At the dawning of the Internet age in 1997, the National Research Council (NRC) addressed data sharing in the following way:

The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research. The public-good interests in the full and open access to and use of scientific data need to be balanced against legitimate concerns for the protection of national security, individual privacy, and intellectual property (National Research Council, 1997).

The NRC acknowledged numerous challenges associated with data sharing, including complications arising from sharing data with different levels of quality and quality assurance, institutional control, documentation, sharing data across scientific disciplines, intellectual property rights, privacy concerns, and the role of government versus the private sector, including commercialization of scientific data. Most of these challenges remain today.

The America COMPETES Act of 2007 (U.S. Congress, 2007) codified the open sharing of United States (U.S.) federal civilian scientific data, stating that agencies shall develop and implement an “overarching set of principles to ensure the communication and open exchange of data and results to other agencies, policy-makers, and the public of research conducted by a scientist employed by a federal civilian agency and to prevent the intentional or unintentional suppression or distortion of such research findings. The principles shall encourage the open exchange of data and results of research undertaken by a scientist employed by such an agency and shall be consistent with existing federal law.” Subsequently, President Obama in 2013 issued an executive order making open, interoperable, machine-readable data the “new default for government information” (The White House, 2013).

According to Crosas (2012), there are two imperatives around scientific data sharing: (1) replication and (2) data citation. Citing King (1995), Crosas notes replication requires “sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.” This standard can be difficult to meet but should be the goal of all scientific data-sharing activities. A persistent reference to the data also is needed to ensure proper citation, with the most common reference being the digital object identifier.
Responding to a variety of requests and opportunities, the U.S. Office of Science and Technology Policy, through the Subcommittee on Water Availability and Quality, began in 2012 to identify key needs in the water resources community that could be addressed through concerted federal activities. Out of these discussions and others, the Open Water Data Initiative (OWDI) emerged in 2014, and was chartered under the Department of the Interior’s Advisory Committee on Water Information (http://acwi.gov/spatial/index.html). Since that time, exemplifying the momentum of the initiative, the OWDI has been identified as a key activity in proposed Congressional legislation (e.g., H.R. 291, https://www.congress.gov/bill/114th-congress/house-bill/291) and in the 2016 Presidential Memorandum on Drought Resilience and associated federal commitments (The White House, 2016).

This JAWRA featured collection presents a broad range of perspectives on the OWDI. A number of articles discuss activities that are mature, having been ongoing since well before the OWDI was formalized. Other articles document the momentum in data sharing and associated applications that have been generated by the OWDI. The water challenges of today and of the future demand all relevant water data and ancillary information be shared quickly, seamlessly, and with applications that add value to the data. The OWDI is helping to meet this need.

**OVERVIEWS**

The featured collection opens with three papers which explore some broader applications of an open water data infrastructure and community. The National Spatial Data Infrastructure (NSDI) is defined as “technologies, policies, criteria, standards, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and nonprofit sectors, and the academic community” (Federal Geographic Data Committee, 2016). Since 1994, the NSDI has improved access to and common standards for geographic data in the U.S. Maidment (2016) explores some of the elements of an OWI, and examines whether a National Water Data Infrastructure (NWDI) might be developed with similar goals for water to those that the NSDI embodies for geospatial information. Such an infrastructure could, for example, underlie the operation of a National Water Model, in which land-atmosphere hydrology and streamflow discharge are computed and forecast continually in a near real-time, high spatial resolution manner across the continental U.S.

A “Digital Divide” in data representation (Maidment et al., 2010), exists between the common way of data archival by earth science data centers and the preferred way of data access by communities that mainly deal with discrete spatial objects (e.g., watersheds) through time. Teng et al. (2016) describe an effort to bridge the Divide, by developing “data rods,” which enable operational access to long time series (e.g., 36 years of hourly data) of selected National Aeronautics and Space Administration (NASA) datasets. The “data rods” project leverages existing NASA capabilities, along with parallel processing, to efficiently generate these time series files. As a result, access to NASA data has been significantly facilitated for the hydrology user community. This activity demonstrates the importance of strong and transparent linkages between the data collection community and the user community. Michelsen et al. (2016) describe the U.S. Geological Survey’s (USGS) Water Availability and Use Science Program (WAUSP). The WAUSP is striving to provide a wide array of data, including detailed water budgets, water availability and use studies, groundwater assessments, improved evapotranspiration and water use information, and geographic focus area studies information. Coordination of data activities with partners, data providers, and federal advisory organizations such as the Advisory Committee on Water Information as part of the OWDI will help ensure the data are useful to a wide range of users and purposes.

**DATA-SHARING SYSTEMS**

Data-sharing systems allow individuals or groups to collaboratively share, edit, and archive locally curated data. A conceptual foundation for water data sharing through the Open Water Web (OWW; similar to the concept of the NWDI), which is to be an outcome of the OWDI, is provided by Blodgett et al. (2016). The OWW is described in terms of four conceptual functions: water data cataloging; water data as a service; enriching water data; and enabling a community for water data. Three OWI-focused use cases (flooding, drought, and contaminant transport) are examined to identify successful practices and needed enhancements. A challenge of any data-sharing system is to facilitate open and agile data exchange while maintaining high levels of data quality, which is explored by Larsen et al. (2016). The challenges associated with ensuring data quality that are specific to the OWI are addressed, along with the current state of the research on this topic.
HydroShare is an example of a data-sharing system which is intended to democratize and expand initiatives such as OWDI (Horsburgh et al., 2016). The portal can be thought of as a “YouTube” for water data that includes data sharing, metadata, and social networking elements. HydroShare’s core concept is the data “resource” which can be, for example, a time series, spatial data, a technical report, or even a video. These resources provide a framework around which tools or web apps can be built to enable cloud-based data storage and analysis. Systems such as HydroShare seem likely to become more widespread in the future as data and modeling moves to the cloud.

SPECIFIC DATA SETS

Interoperable hydrologic datasets, which either enable the OWDI or are examples of open water data, are discussed in four articles. Digital stream networks first began to emerge in the 1970s, with the modern National Hydrography Dataset Plus (NHDPlus) for the U.S. being the latest incarnation (Moore and Dewald, 2016). Digital stream networks with associated catchments provide a geospatial framework for linking and integrating water-related data. An enormous amount of technical GIS development, as well as a great deal of interagency coordination to ensure consistency across organizations, was associated with the data compilation required for the development of NHDPlus. Advancements in the development of NHDPlus are expected to continue to improve the capabilities of this national geospatial hydrologic framework. One such advancement, currently under development, is NHDPlus High Resolution (NHDPlusHR), which is a new generation hydrographic framework for the U.S. (Viger et al., 2016). NHDPlusHR will support a number of new applications, most notably methods for robustly scaling between local to national representations of the network. Currently NHDPlusHR is a bit of a patchwork underpinned by consistent 1:24,000-scale geospatial data, but updated with a growing number of patches of much higher-resolution data being added as the data become available. Ongoing efforts to add functionality to NHDPlusHR include delineation of catchments, addition of flow direction and flow accumulation grids, and other attributes.

A number of applications for serving downscaled global climate model simulations have been developed (e.g., Alder and Hostetler, 2015). Woodbury et al. (2016) describe a new Coupled Model Intercomparison Project phase 5 (CMIP5) based database which has been prepared through scaling and weighted averaging for use at the level of USGS HUC-8 watersheds (approximately 1,800 square kilometers). The new dataset is deployed through HydroShare (Horsburgh et al., 2016), using WaterOneFlow web services in the WaterML format. Two use case scenarios, applications with the Climate Analysis Toolkit (an extension to HydroDesktop) and rapid comparison of model forecasts across watersheds, are provided.

Although a number of open access web services exist for viewing snow cover, Kadlec et al. (2016) present a new, open-source application based on existing imagery for accessing time series data of snow cover and probability of snow cover at particular point locations. The application is made available, using the Tethys platform web user interface, and a WaterML programming interface provides third party applications direct access to data.

DATA TOOLS AND MODELS

The final articles present examples of data tools and models which demonstrate value from application of the OWDI concepts, with four of the five papers addressing some aspect of flood modeling or forecasting. The first example (Harpham et al., 2016) demonstrates how a flexible modelling architecture that integrates models with observational data has been constructed, using a set of standards and a Model MAP (Metadata, Adaptors, Portability) gateway concept to prepare numerical models for use in flood forecasting. Hydraulic results, including impact to buildings and hazards to people, are given for the use cases of severe and fatal flash floods which occurred in Genoa, Italy in 2011 and 2014.

A second example (Snow et al., 2016) presents a method for routing global runoff ensemble forecasts and global historical runoff generated by the European Centre for Medium-Range Weather Forecasts (ECMWF) model, using the Routing Application for Parallel computation of Discharge (RAPID) to produce high-spatial resolution 15-day stream forecasts, approximate flood recurrence intervals, and warnings at locations where streamflow is predicted to exceed the recurrence interval thresholds. The ECMWF model is unique in that it provides a longer-range ensemble forecast to be routed and evaluated through the rest of the flood modeling system. In addition, the Streamflow Prediction Tool web application was developed for visualizing results at both the regional level.

JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

813

JAWRA
and at the reach level of high-density stream networks. The application formed part of the base hydrologic forecasting service available to the National Flood Interoperability Experiment (Blodgett et al., 2016) and can potentially transform the Nation’s forecast ability by incorporating ensemble predictions at the nearly 2.7 million reaches of the NHDPlus into the national forecasting system.

The third flood-related example (Perez et al., 2016) examines issues of downscaling (global and regional forecasts) relative to the needs of emergency managers for flood warnings that are detailed and specific to local conditions. An approach for the combination of distributed hydrologic models with downscaled forecasts to produce high-resolution flood forecasts is demonstrated. Three modeling strategies are compared for addressing downscaling issues: downscaling of coarse resolution global runoff models to high-resolution stream networks and routing with RAPID; the use of hierarchical distributed models; and precomputed distributed models. By demonstrating the analyses in the context of an open-source, cloud-based computing environment, the important practical challenges of providing tools and computing power for local emergency managers at the local level are also addressed.

The fourth example is an informative example on flow estimation (Selvanathan et al., 2016), using the concepts of the OWDI. Specifically, the study explores the development of 1 and 10% exceedance probability peak discharge flows throughout the U.S. based on seven selected climate models representing the upper and lower bounds for possible future discharge conditions. A weighted regression approach applies climate model results to 7,302 stream gauge locations, using regression equations developed uniquely for each Level 2 USGS Hydrologic Unit Code (HUC-2) region. The work serves as an example of using nationally available OWDI-related data to generate a new and potentially significant data product.

Sensors and enabling technologies are becoming increasingly important tools for water quality monitoring and associated water resource management decisions. In particular, nutrient sensors are of interest because of the need for accurate and timely information about drivers of adverse effects of water quality degradation such as nutrient enrichment on coastal hypoxia, harmful algal blooms, and impacts to human health. Using nitrate sensors as the primary example, Pellerin et al. (2016) highlight applications in freshwater and coastal environments that are likely to benefit from continuous, real-time nutrient data. The concurrent emergence of new tools to integrate, manage, and share large datasets is critical to the successful use of these types of real-time sensors. Near-term opportunities that will help accelerate sensor development, build a national network, and develop open data standards are highlighted.

**SUMMARY**

The OWDI, in many respects, is a rebranding of what most in the water-resources science community have been advocating and moving toward for a number of years. For example, USGS water web services have been in place for more than a decade, with new capabilities being regularly added, and HydroShare was operational well before OWDI was instituted. Nevertheless, as demonstrated by the articles in this collection, the OWDI has created a more concerted and organized effort to provide water data and community applications built on those data, especially in the federal community. As noted by Larsen, “In the future, your performance metric will not be how many people visit your website, but how many applications your data support” (ClimateWire, 2015).

**LITERATURE CITED**


