FEATURE

Moving from Information to Insight by Linking Urban and Hydrologic Systems through the Urban Flooding Open Knowledge Network

J. Michael Johnson, Siddharth Saksena, Lilit Yeghiazarian, Venkatesh Merwade, Sankar Arumugam, Scott Back, Jerad Bales, Ximing Cai, Doug Fils, Torsten Hahmann, Jeffery S. Horsburgh, Zhenyu Huang, Renke Huang, Amirhossein Mazrooei, Kyle Onda, Ranji Ranjithan, M. Sadegh Riasi, Shawn Rice, Majid Shafiee-Jood, Adam Shepherd, Peter Singhofen, Shirly Stephan, David Tarboton, and Alexandre Tartakovsky

Floods are the costliest type of disaster in the United States. While evaluating flood risk remains a critical issue for the scientific community, assessing the total impact across complex urban systems is an even greater challenge. During flood events, interconnected urban infrastructure like the power grid, transportation network, drinking water and sewerage systems are all impacted, and experience cascading failures. We call this interconnected system of systems the Urban Multiplex. Even though we have data about individual components of the Urban Multiplex and an abundance of hydrologic forecasts, this information is not yet connected in a way that promotes clear, immediate insight about existing and potential problems.

OVER THE LAST THREE DECADES, three prominent ideas have fundamentally changed the way we store, manage and interact with data and documents. We believe they can help elevate our representation of flood impacts. These include:

WWW and Semantic Web 3: The World Wide Web was designed by Tim Berners-Lee in 1989 to connect documents on the internet through navigable hyperlinks. The vision for the Web 3.0, or the semantic web, is that the network of hyperlinked human-readable web pages will include machine-readable metadata that describes the fundamental content and relationships of the data. Through these links, connections to other related resources can be derived providing a "linked" or collective view of the world's knowledge (*e.g.* knowledge graph). This transition from linear connections (user clicks through links) to semantic relationships (metadata is used to connect diverse documents) marks a large step toward the Internet's ability to provide not just information, but also insight.

Digital Earth: In 1998, US Vice President Al Gore proposed the idea of a Digital Earth in which information about a collective cultural and scientific understanding of the world (imagery, text, geospatial objects) could be stored in an interactive 3D globe. At the same time, significant investments were made in creating geospatial datasets that represent our spatial understanding of where things are in the world and what attributes are associated with them. In 2008, a working group published a paper on the *Next Generation of Digital Earth* emphasizing that a truly Digital Earth must not be a single system, but a collection of integrated infrastructures built on open data.

National Water Model: In 2016, the National Ocenic and Atmosperice Administration (NOAA) implemented version 1.0 of the National Water Model (NWM; https://water.noaa.gov/about/nwm) which provides current and forecasted streamflow simulations for the entire Continental United States (CONUS). Such a system is built on a rich collection of national scale geospatial products, which fundamentally changes the landscape of federated hydrologic forecasting.

Combined, these developments offer a new and exciting opportunity to better connect our understanding of the built and natural environments through the way we document, structure, store and share knowledge about the human-environment interface.

No Clear Semantic Understanding of Flooding

The Google search bar provides a unique lens into the way human knowledge about any topic is modeled, and how well relevant data can be connected in the semantic web. In a familiar example, a search for 'coffee shop near

me' is able to link our understanding of 'coffee shop', and 'near me', to provide not only locations but also relevant information such as hours, reviews, and menus. Today, a similar search for "flooding near me" (Figure 1 left panel; made in Santa Barbara on February 21st, 2020) produces less informative results. Instead of a rich set of information, the first entry is a link to a general question of *"where can I get flood maps"* while the second and third entries advertise an Australian service called "*Floods Near Me"*. Moreover, the videos are from 2017 and 2018 events in the Midwest and Texas. The limitations in connecting the idea of flooding to the time and place of the query, and to other compatible information (like weather) highlights an opportunity to transform the way flood information is communicated.

To illustrate, if a clear semantic understanding of flooding and its relationship to space and time were contained in a knowledge graph, the quality of these results could be improved to resemble something like Figure 1 (right panel) where the connections between geospatial products are defined, the connections to hydrologic models and observation systems are explicit, and the contextualization of flooding as a weather-driven phenomenon is concrete. In this way, flooding would not give us reviews, hours and menus like a 'coffee shop', but maps (Digital Earth), warnings (NWM), and weather reports.

When these links and relationships are defined, knowledge can be traversed in such a way that we immediately connect streamflow values to the structures at risk, and to the entities those structures serve. That is, we can identify if a streamflow is large enough to impact a power station; the address points and infrastructure served by that power station; and the role those entities play in the Urban Multiplex. Such connections not only help homeowners prepare for imminent flooding risks but also anticipate other problems such as possible power outages, sewage backflows, and service closures. More importantly, such connections enable the analysis and optimization of the interactions among various urban infrastructures such that damages can be minimized, and recovery expedited. In this way a forecast becomes more than just information, but insight that can inform context-based decisions at all scales.

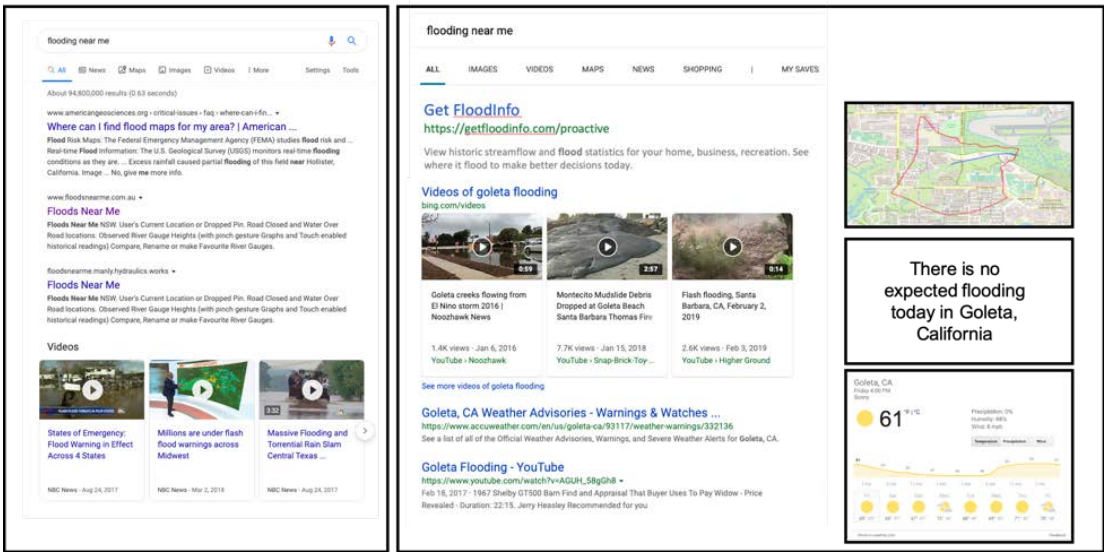

Figure 1. Searching for "flooding near me": current results (left panel); and expected results enabled by the Urban Flooding Open Knowledge Network (right panel).

The Urban Flooding Open Knowledge Network (UFOKN)

In 2019, as part of the National Science Foundation's (NSF) push to harness the Big Data revolution through the Convergence Accelerator pilot, a team of researchers set out to model our understanding of flooding by developing the semantic connections between the Urban Multiplex and hydrologic predictions.

In the first phase of this project, two stakeholder workshops were organized to bring together industry, academic, and government participants to better understand their day-to-day challenges. From these, a robust list of user personas was identified to guide the initial content and relationships needed in a urban flooding-focused knowledge graph. To date, we have successfully linked OpenStreetMap infrastructure and OpenAddress points with the NWM to provide real-time assessments and historic synopses of impacted infrastructure allowing users to search for flooded roads or at-risk homes. As this graph grows, additional infrastructure types like power grids, bridges, and Superfund sites are being added along with their high-level relationships to one another (e.g. power grids serve homes).

Like Digital Earth, a truly successful open knowledge network will be a constantly evolving entity, able to integrate a wide array of resources from other architectures and users. To achieve this, we are developing an extendable conceptual model that allows others to introduce their own model output, observation networks, and urban information to what we hope will be a community resource. In this vein our team involves a host of modelers, engineers, geographers, social scientists, data and computer scientists, seeking to generalize the type of information that can be ingested and extracted from the graph. More broadly, we hope that through this work our semantic understanding of flooding will become rich enough so that finding

relevant flood information will be as simple as finding a good coffee shop. ■

Acknowledgment: We gratefully acknowledge the National Science Foundation for supporting this project (OIA grant # 1937099).

Michael Johnson is a PhD student in Geography at the University of California, Santa Barbara (UCSB). He focuses on hydrologic and geographic data science and is the current Dangermond Fellow at UCSB. He is a former student and course coordinator for the NOAA National Water Center Summer Institute and a 2020 recipient of the CUAHSI HydroInformatics fellowship.
Contact: jmj00@ucsb.edu

Co-Authors

Siddharth Saksena - Department of Civil and Environmental Engineering, Virginia, Polytechnic Institute and State University, Blacksburg, VA

Lilit Yeghiazarian and M. Sadegh Riasi - Department of Chemical and Environmental Engineering, University of Cincinnati, Cincinnati, OH

Venkatesh Merwade and Shawn Rice - Lyles School of Civil Engineering, Purdue University, West Lafayette, IN

Sankar Arumugam and Amirhossein Mazrooei Ranji Ranjithan - Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC

Scott Back, Jeffery S. Horsburgh, and David Tarboton - Utah Water Research Laboratory, Utah State University, Logan, UT

Jerad Bales - Consortium of Universities for the Advancement of Hydrologic Science, Inc., Cambridge, Massachusetts

Ximing Cai and KMajid Shafiee-Jood - Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

Doug Fils - Consortium for Ocean Leadership, Washington, DC, USA

Torsten Hahmann and Shirly Stephan - School of Computing and Information Sciences, University of Maine, Orono, ME

Zhenyu Huang, Renke Huang, and Alexandre Tartakovsky - Pacific Northwest National Laboratory, Richland, WA

Kyle Onda - Nicholas Institute for Environmental Policy Solutions, Duke University, Durham, NC

Adam Shepherd - Woods Hole Oceanographic Institution, Woods Hole, MA

Peter Singhofen - Streamline Technologies, Inc., Winter Springs, FL