Deployment of Real-Time Analytics and Modeling at the City of Flint, Michigan Water System

INTRODUCTION

The City of Flint, Michigan owns and operates the public water system that provides drinking water to its nearly 100,000 residents. In 2014 the City of Flint, Michigan ("City") ceased drawing treated water under its agreement with the Detroit Water and Sewage Department and began treating water from its Flint River source. The change in source water and the lack of corrosion control treatment following that change allowed lead to leach into the drinking water. The U.S. Environmental Protection Agency (EPA) found that the drinking water provided by the City of Flint to its residents posed an imminent and substantial endangerment to residents' health.

EPA, the City, and the utility recognized that a detailed analysis of the drinking water distribution system was needed. Drinking water distribution systems are complex systems, typically encompassing hundreds or thousands of miles of pipes linked together through a complex array of valves, pumps, and storage facilities, to efficiently treat and convey water from a raw water source to the customer. The ability of a water utility to successfully manage such a feat under the best of circumstances is difficult but when an emergency occurs, that difficulty is multiplied. Thus, when an emergency occurs, proper understanding of the system and how it works are key to limiting damage and furthering recovery. An accurate water distribution system network model can provide this understanding.

EPA formed a Safe Drinking Water Task Force to support the City on October 16, 2015. On January 21, 2016 the EPA issued a Safe Drinking Water Act Emergency Administrative Order to the City and the State of Michigan to put in place necessary actions to protect public health. The Safe Drinking Water Task Force included providing support from EPA scientists and technical experts. The Task Force identified the need for an updated and accurate water distribution system network hydraulic model for the City. EPA began working with the City to update and calibrate their network model in March 2016. To help improve and optimize operations in the system, EPA worked with representatives from the utility, the state of Michigan, the City, CitiLogics, and ARCADIS.

Maintaining an updated and accurate water distribution system network model can be difficult for any water system. It was recognized to be a particularly difficult problem, given future uncertainties regarding their long-term water source and a myriad of operational issues and needs that were being addressed by the City and the utility. As a result, EPA suggested the need for a data integration

EPA/600/S-18/353
October 2018

framework that could allow the utility to easily assess model accuracy and more efficiently update their network model, e.g., by incorporating new or modified infrastructure elements (e.g., pumps and reservoirs) or revising operational rules. To meet this need, EPA deployed the EPANET-RTX (RTX) real-time data integration framework to update, calibrate and confirm the accuracy of the City's network hydraulic model and to provide a lasting data integration framework [1, 2, 3, 4].

This technical brief summarizes the deployment of the prototype RTX real-time analytics (RTX:LINK) and data integration framework technologies to the City. It presents a high-level summary of the network hydraulic model calibration and accuracy assessment that were conducted and describes the benefits to the City once the technologies are fully deployed and adopted by the City. Finally, this technical brief demonstrates that RTX and RTX:LINK technologies can enable more efficient model calibration and accuracy assessment and make it easier for water utilities to update their network model as their operations change or are updated.

DEPLOYMENT OF REAL-TIME ANALYTICS - RTX:LINK

The first step for EPA in assisting the City with updating, calibrating and assessing the accuracy of their network model to reliably predict water system behavior was to access to the city's SCADA database. The City initially provided their SCADA data for network model calibration as comma-separated values (CSV) files, collected from the data stored in the SCADA historian database, and provided to EPA on compact disks or email attachments. After some discussion and interactions with City staff, the project team worked with utility staff to perform quality assurance on the data and implement some configuration changes to the way SCADA was being collected, resulting in shorter and more consistent data retrieval (e.g., one-minute data intervals for tank and reservoir levels and pump statuses, and 15-minute intervals for control valve positions, pressure measures, and flow measures). These changes were important for constructing diurnal demand curves and properly regulating the water supplies into the network model. As these CSV files were obtained, they were uploaded into a cloud database equipped with dashboard views for easy viewing, analyzing and management.

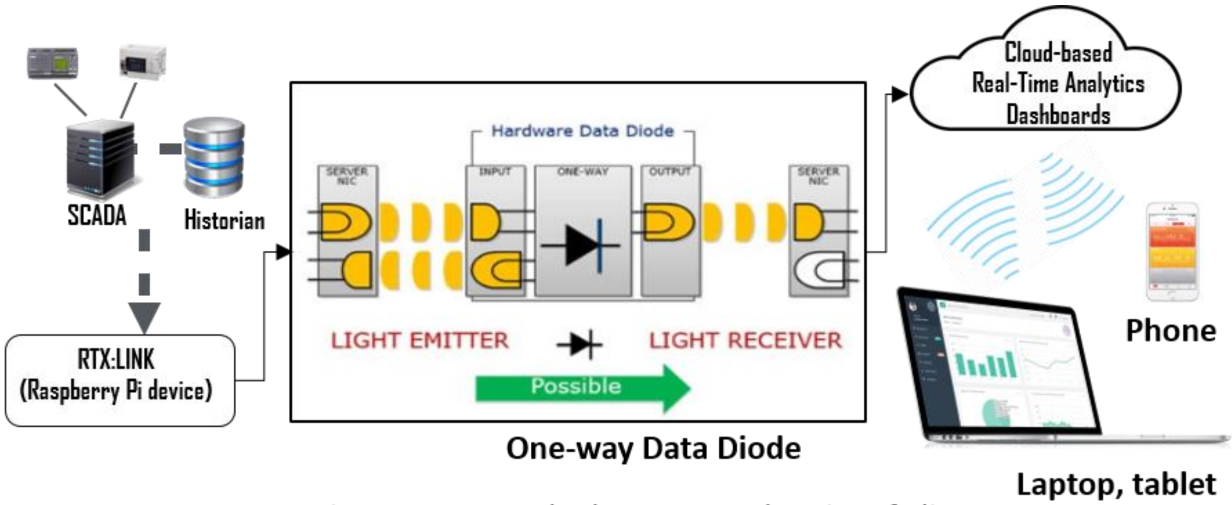

Fig. 1. RTX:LINK deployment to the City of Flint

However, it was quickly recognized that the City needed a data integration framework that could allow the utility to easily assess model accuracy and update their network model more efficiently, e.g., by incorporating new or modified infrastructure elements (e.g., pumps and reservoirs) or revising operational rules. To meet this need, the manual CSV file transfer process was replaced by the installation of RTX:LINK. RTX:LINK is an RTX application that provides a simple, secure, read-only access to key operational data streams (e.g., those stored within a SCADA historian), through web-based dashboards for trending, analysis, and alerting [5]. RTX:LINK was deployed to the City along with a one-way data diode between the RTX:LINK SCADA historian replication device (raspberry pi) and the city's public internet connection, to prevent malicious attacks on the SCADA network from the public internet (Fig. 1). The deployment of RTX:LINK to the City provides managers and operators:

- Anywhere and anytime access to their SCADA data analytics via a mobile dashboard,
- Ability for easy increased attention to potential operational and emergency problems related to hydraulic data, equipment, and facility operations. This allows the utility to quickly catch potential problems observable through SCADA data (e.g., potential problems related to pump operation, flows, pressures, tank operation and disinfectant management), and
- Improved water quality management including the ability to monitor disinfectant loss and disinfection byproduct formation in the distribution system. Such data are a function of residence time, which is affected by day-to-day system operational policies and decisions, particularly those associated with water stored in tanks. By providing real-time, accurate information about operational effects on tank residence time and mixing, real-time analytics can improve awareness of tank water quality, which can lead to operational changes that improve disinfectant management.

RTX:LINK also enabled the City to more easily deploy real-time modeling, e.g., real-time water age mapping. RTX:LINK also provided the City with the capability to more easily adopt model-based predictive, real-time analytics (using real-time SCADA data and an updated network model) which is critical in allowing the early detection and awareness of operational problems in the distribution system.

EPANET-RTX BASED NETWORK MODEL

The next step was to update the City's network model with EPANET-RTX to provide the functionalities for data access, data transformation and data synthesis (including real-time analytics). The RTX data integration framework that was deployed to the City uses CitiLogics' Polaris™ Work Bench data integration environment (Fig. 2)[1]. Fig. 2 is a conceptual illustration of how the RTX library can be harnessed to build a data integrated model. The RTX-based data integrated, real-time model (real-time model) deployed to the City is depicted as the light blue shaded box within Fig. 2. In Fig. 2, the inputs to this real-time model are shown in the boxes to the left of the blue shaded box while the outputs, specifically "Prediction" and "Accuracy Assessment" and "Network Diagnostics and Calibration Needs," are shown below. The setup of the real-time model began by importing the City's existing network model, developed in 2009. The original model was obtained as an EPANET file that was exported from the City's commercial software application[2] [6].

The setup of the real-time model consists of three principle tasks as shown in Fig. 2:

- "Data Access and Integration" is the process for integrating a network model with real data. The initial, manual SCADA data integration process was later replaced by a live SCADA data feed with the deployment of RTX:LINK.
- "Asset Mapping" is the process in which all the key infrastructure assets (e.g., pumps, valves, tank, and reservoirs) for the City were associated with their appropriate SCADA data streams.
- The "Data Pipeline Configuration" process is the final step. Raw SCADA data streams are filtered and processed by removing errors, bad data, and dropped signals. Data pipeline configuration refers to the process of converting raw SCADA data streams into more accurate and useful network model measures and boundary conditions. The design of RTX incorporates a data time-series processing library for efficiently removing data outliers, filtering noisy data, and bridging data gaps, as well as transforming and combining data by basic mathematical operators.

[1] Polaris Work Bench is an RTX-based commercial software program.

[2] An EPANET file is a computer text file with a ".inp" extension in its file name which describes a water distribution system. An .inp file specifies all the infrastructure of the water system, including their characteristics, operating instructions, as well as other hydraulic and water quality configuration parameters.

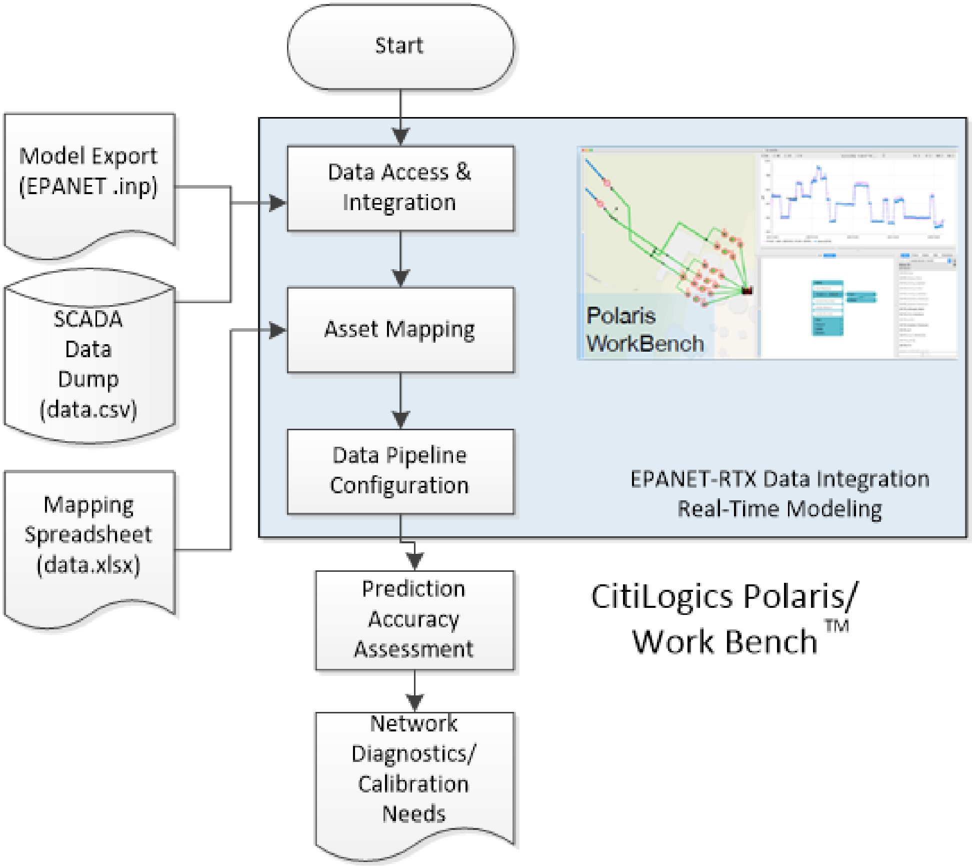

Fig 2. Commercial application using EPANET-RTX

EPANET-RTX: EASIER CALIBRATION AND ACCURACY ASSESSMENT

Once the configuration of the City's real-time model was completed, the resulting network model was calibrated and then tested for its ability to predict operational behavior. The real-time model made testing, calibration, and accuracy assessment much easier. For instance, an RTX-based real-time model allows the user to simulate any historical time period supported by the underlying data, while producing accuracy metrics for every SCADA measure and comparing model predictions with observations. The resulting error statistics are more meaningful because they reflect the true operational record from the SCADA historian database. Through this systematic process of replacing operations' assumptions with data[3], any significant output errors can be interpreted as misrepresentations of the infrastructure or its condition, and not of the operational decisions. The capability to select and evaluate a particular simulation period, e.g., a week in August 2017, with a detailed understanding of the associated operations taking place (an understanding supported by

[3] Typical network models (e.g., EPANET) assume a pump turns on, for example, at 6 AM every day and off at 3 PM; an RTX-based real-time model can automatically incorporate the actual pump statuses that occurred during the time period of interest.

data), provides the basis for effective and efficient model calibration and a continuous assessment of model accuracy.

CITY OF FLINT – NETWORK MODEL CALIBRATION

The calibration of the City's real-time model was a collaborative effort between EPA, the City, CitiLogics, and Arcadis. The calibration consisted of a 16-week period from July 29, 2016 until November 18, 2016. For the calibration period, field pressure and flow data and the City's SCADA historical record for the water system were accessed and used to calibrate the network model using CitiLogics' Polaris Work Bench.

Multiple site visits were made by project team representatives to ensure an accurate representation of the pumping and control facilities on the treatment plant grounds, at the primary booster pump station, and the two ground level reservoirs and associated pumping facilities. These data were then used to support the modifications made to key facility infrastructure representations and operating characteristics (e.g., pump head-discharge and valve loss curves) in the network model.

Multiple field studies were also implemented to collect additional hydraulic data. Pressure and flow data were collected by the Michigan Department of Environmental Quality. EPA and CitiLogics personnel deployed pressure monitors and collected pressure data. Network model node elevation data were reviewed and updated. Fire hydrant flow tests were conducted by ARCADIS to evaluate the significant differences that were observed between actual and expected head losses[4] (ranging from 3 to 8 ft./mile) in transmission mains feeding the city's two ground level reservoirs. Due to these field studies, the network model's C factors[5] were reduced significantly for all pipe size ranges, including several transmission mains.

Network model base demands were updated to reflect billing records as late as 2015. Water usage billing data from approximately 50,000 individual customer accounts (representing approximately 900,000 customer water meter readings) from February 2013 to September 2016 were put into a billing database. The data was then processed to estimate accumulated average water usage rates that were used to update the model base demands in the EPANET model file. Along with other changes, the calibration process resulted in a significantly improved network model.

[4] Head loss is the resistance to water flow due to friction.

[5] A value used in a hydraulic network model to describe the smoothness of the interior of a pipe. The higher the C factor, the smoother the pipe, the greater the flow capacity, and the smaller the friction or energy losses from water flowing in the pipe.

Comparing the SCADA historical record with the calibrated model's output over the 16-week calibration period displayed a high level of model accuracy. Table 1 below shows the accuracy results for the City's network model over the 16-week calibration period in 2016.

Table 1. Calibrated Model Accuracy Results Over 16-Week Calibration Period in 2016*

	Hydraulic Head (FT)		Tank Level (FT)		Pipe Flow (GPM)		Tank Flow (GPM)	
Period	Avg MAE	Avg R	Avg MAE	Avg R	Avg MAE	Avg R	Avg MAE	Avg R
2016	3.20	0.87	0.76	0.99	273	0.94	217	0.97

*Results expressed as the average (Avg) mean absolute errors (MAE) over all measurements within a category. R corresponds to Pearson correlation coefficients. FT, feet; GPM, gallons per minute.

CITY OF FLINT – NETWORK MODEL ACCURACY ASSESSMENT

The accuracy assessment for the updated and calibrated City 's network model was performed for a similar but slightly shorter period in 2017 (July 29th to Oct. 27th)[6]. The selection of this time period was to test the network model's ability to simulate hydraulic events for a different operational period. The period selected in 2017 included a significant operational difference from 2016. Specifically, one of the City's two major ground reservoirs was out of service for the entire time. Thus, the 2017 accuracy assessment period tested the ability of the network model to recreate the behavior observed in 2017 without the hydraulic interaction that existed between the two ground level storage facilities in 2016, during the time period for the network model calibration. Table 2 below shows the accuracy results for the City's network model over the 13-week accuracy assessment period in 2017.

[6] During week 14 of the accuracy assessment period, a SCADA reservoir level signal at the operating ground level reservoir experienced a string of invalid values (sudden and sustained decreases and increases of 5-15 feet without reason) that continued for more than 24 hours. Due to the nature of these errors it was impossible to recover the true signal using available filtering operations, and because of the large capacity of the operating reservoir, the errors propagated to the calculated diurnal demand which, in turn, had a significant impact on model simulation accuracy that then continued for the duration of the accuracy assessment period. Rather than have such errors potentially bias the interpretation of the accuracy assessment results, it was decided to shorten the accuracy assessment period from 16 to 13 weeks.

Table 2. Accuracy Assessment Results Over 13-Week Accuracy Assessment Period in 2017

	Hydraulic Head (FT)		Tank Level (FT)		Pipe Flow (GPM)		Tank Flow (GPM)	
Period	Avg MAE	Avg R	Avg MAE	Avg R	Avg MAE	Avg R	Avg MAE	Avg R
2017	3.08	0.79	0.57	0.97	304	0.93	348	0.92

Avg, average; FT, feet; GPM, gallons per minute; MAE, mean absolute errors over all measurements within a category; R, Pearson correlation coefficients.

BENEFITS TO THE CITY OF FLINT

Real-Time Data Analytics - The deployment of RTX:LINK hardware and software to the City provides utility managers and operators with an easy, web-based access to all their water distribution system operational data through a highly secure connection. Fig. 3 shows an example view of the RTX:LINK dashboards illustrating system pressures and water quality. These Grafana-based dashboards are easily user- configured to provide the analytics of most interest to the City or water utility.[7]

The real-time analytics dashboards provide utility managers and operators their key utility operational data streams via a smart phone APP – fostering improved communication about operations between operators, engineers, and managers.

Real-Time Model - The City now has an updated, calibrated, and validated (through the accuracy assessment) network model. More importantly, the City now has an updated, modern SCADA data-based real-time model of their water system operations that can be more easily maintained and used in day-to-day management of their water system operations. The RTX-based hardware/software system integrates the City's SCADA data with their hydraulic and water quality network model for real-time prediction of system wide flows, pressures, and water quality. The real-time model includes a continuous accuracy evaluation of the real-time predictions, using an unprecedented volume of streaming, real-time (SCADA) operational data. These accuracy metrics can now be tracked automatically, greatly decreasing the effort and costs required to maintain the City's network model while, at the same time, enhancing confidence in its use.

[7] Grafana - https://grafana.com/ is an open source platform for analytics and monitoring that is well-suited for time-series data.

Fig. 3. City of Flint real-time analytics dashboards

Operations and Management - During the model calibration process, the real-time model was instrumental in identifying excessive head losses in city's transmission and distribution mains and ineffective reservoir pumping operations that were resulting in excessive water age in storage.

The real-time model can be used to assist with asset management and sensor placement (e.g., to support improved operations or compliance monitoring). Finally, the real-time model can now be used by the City to support future model upgrades and engineering studies.

The real-time model can also be used to monitor and help manage water quality. Water age is a parameter that can be monitored for an indication of water quality. Water age is a major factor in in the loss of water quality within drinking water distribution systems. Water quality deteriorates from the time treated water leaves the treatment plant until it is used by the customer. As the water travels through the distribution system, it undergoes various chemical, physical and aesthetic transformations due to its interactions with pipe walls (including biofilms, pipe scale, and contaminants) thus impacting water quality. A longer water age can allow disinfectants more time to react with naturally occurring materials in the distribution system to form disinfection by-products (DBPs). Water age is generally driven by customer demand and water system design and operation.

Fig. 4 shows a snap-shot of a prototype, real-time water age map for the City. This snap-shot map of real-time water age provides an illustration of a continually updated view of the last 24 hours of water age for the City, as a function of actual operations. A continuously updated, real-time model can provide utility managers with the information and the tool needed to set specific water quality goals (e.g., lowering water age) and help them achieve them.

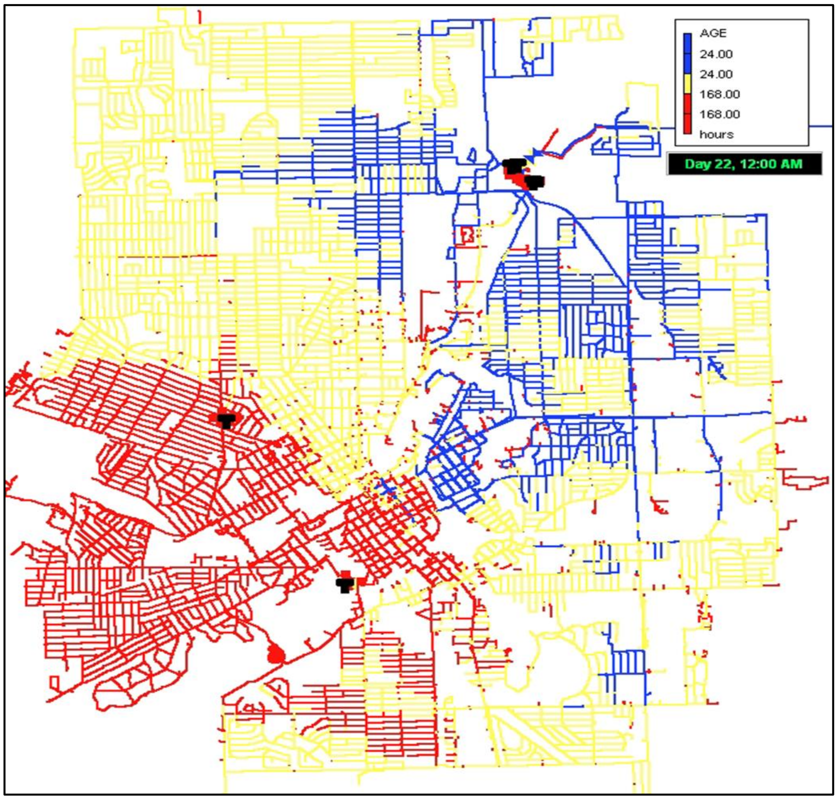

Fig. 4. Real-time water age map for City of Flint.

Note: Water age is not directly correlated with disinfectant chlorine residuals or other water quality parameters and does not reflect the City's efforts to perform booster chlorination at its storage tanks.

FOR MORE INFORMATION

The intended audience for RTX:LINK is water utilities and their consultants and software developers. The intended audience for EPANET-RTX is software developers. EPANET-RTX and RTX:LINK are open-source software projects. If collaborators are interested, there are various ways to get involved (e.g., connecting to the code repository, looking over coding conventions and using the issues tracker to make a feature request and communicate with the developers). To learn more about RTX:LINK and real-time modeling using EPANET-RTX or ongoing enhancements, visit the repositories of EPANET-RTX and RTX:LINK at https://github.com/OpenWaterAnalytics.

REFERENCES

[1] Uber, J., Hatchett, S., Hooper, S., Boccelli, D., Woo, H., and Janke, R. (2014). Water Utility Case Study of Real-Time Network Hydraulic and Water Quality Modeling Using EPANET-RTX Libraries. Cincinnati, Ohio: U.S. Environmental Protection Agency. EPA 600-R-14-350.

[2] U.S. Environmental Protection Agency. (2014). "Enhancements to the EPANET-RTX Open Source Libraries 2014," (Technical Brief). EPA/600/S-14/310.

[3] U.S. Environmental Protection Agency. (2015). "Enhancements to the EPANET-RTX (Real-Time Analytics) Software Libraries (Fiscal Year 2015)," (Technical Brief). EPA/600/S-14/441 https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=309673 accessed on July 27, 2018.

[4] U.S. Environmental Protection Agency. (2012). http://wateranalytics.org/epanet-rtx/index.html and https://github.com/OpenWaterAnalytics/epanet-rtx, accessed on July 26, 2018.

[5] U.S. Environment Protection Agency. (2016). "RTX:LINK – An EPANET-RTX Software Tool For Water Utilities To Implement Real-Time Analytics," (Technical Brief). EPA/600/S-16/287 https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=328356, accessed on July 26, 2018.

[6] Rossman, L.A. (2000). EPANET 2 Users Manual. Cincinnati, OH: U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory Report, EPA/600/R-00/057.

CONTACT INFORMATION

For more information, visit the EPA Web site at https://www.epa.gov/homeland-security-research

Technical Contacts: Robert Janke (janke.robert@epa.gov)
Regan Murray (murray.regan@epa.gov)

General Feedback/Questions: Amelia McCall (mccall.amelia@epa.gov)

DISCLAIMER

The U.S. Environmental Protection Agency funded and collaborated in the work described here. It has been subjected to the Agency's review and has been approved for publication. Note that approval does not signify that the contents necessarily reflect the views of the Agency. Mention of trade names, products, or services does not convey EPA approval, endorsement, or recommendation.

U.S. EPA's Homeland Security Research Program (HSRP) develops products based on scientific research and technology evaluations. Our products and expertise are widely used in preventing, preparing for, and recovering from public health and environmental emergencies that arise from terrorist attacks or natural disasters. Our research and products address biological, radiological, or chemical contaminants that could affect indoor areas, outdoor areas, or water infrastructure. HSRP provides these products, technical assistance.