SWADE: Smart WAter Data Exchange

UCI: N. Venkatasubramanian(PI), S. Mehrotra(PI), A. Chio, R. Hildebrant, M. Luti, R. Rahman, P. Venkateswaran, R. Yus, G. Zhang **Water UCI**: D. Feldman(PI), C. Gim, S. Roback **SDSU**: S. Ren(PI), B. Tskhadadze, Z. Zhang **ImageCAT Inc**. R. Eguchi(PI), Z. Hu, M. Mendoza **IIT Chicago**: P. Anderson

Creating an Extensible Data Exchange and Analytics Sandbox for Smart Water Infrastructures

Motivation and Goals

- Water infrastructure (stormwater, drinking water, wastewater) is aging and becoming increasingly complex; agencies/utilities operate independently with specific regulatory compliance needs.
- Data and structural information can be used to develop tools for interpreting data, identifying problem sources, and translating it into actions, but is often siloed within agencies and systems.

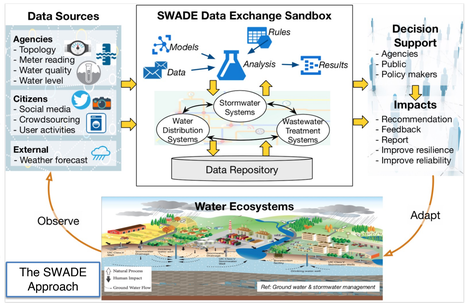
Key Premise:

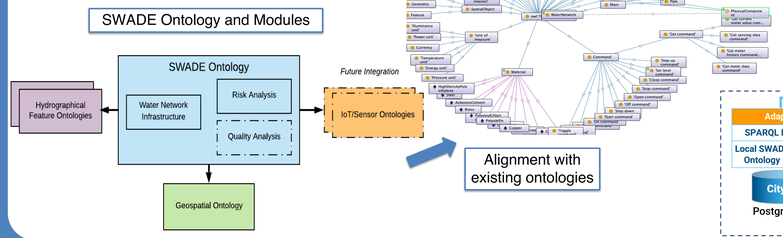
Water cycle data (historical and live) and its dependencies, a bulk of which resides within community agencies, if combined and enhanced with other geo-distributed data sources can enable new levels of efficiency and resilience.

- Goal 1: Structure interconnection of different water infrastructures and their impact on each other under normal operation and extreme events.
- Goal 2: Identify information needed to enable data-driven approaches and barriers (societal, policy) on its gathering, processing, sharing, and use
- Goal 3: Address challenges in leveraging (restricted) data to support timely decision making, possibly under large disruptions (e.g., earthquakes).

Challenges

- Modeling information from geo-distributed infrastructures whose operation is governed by physics under dynamic environments
- Heterogeneity of data sources at varying spatial and temporal granularities
- Model problems and concepts across water domains within a single platform
- Current approaches requires significant effort to acquire and understand data, handle delays in processing information, exploit innate data redundancies
- Inter-agency and community data exchange constraints and barriers (data handling, privacy, security) with multiple stakeholders
- Citizen/Community engagement for water-knowledge co-production



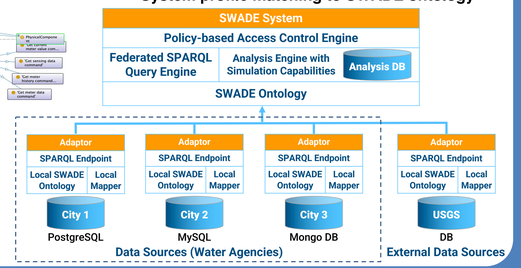


The SWADE System

- Data exchange architecture for enabling interoperability across different water organizations and external data sources
- Policy-based access control engine and federated SPARQL Query engine dictates how data is shared
- Analysis engine with simulation capabilities adapts to each organization's data format

SWADE Ontology and Information Exchange

- Vocabulary to enable interoperability of water organizations
- Definition of water networks and relevant components, geographical elements, hazards, risk analysis results, and water quality analysis
- System profile matching to SWADE ontology



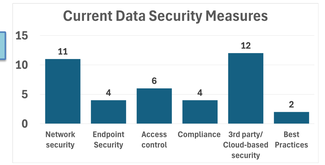
Community Engagement & Broader Impacts

Stakeholder Workshops and Data Challenge Events

- Workshop to survey stakeholders (agencies, policymakers, academics, industry partners, etc.) for their resilience preparation
- Water Data: What kinds of data are collected by water agencies? How do approaches to the collection and management of data differ among water agencies (e.g., data format, protocols for sharing, etc.)? What external factors are considered (e.g., size of customer and/or revenue base)?
- What are their concerns (privacy, data sharing, accessibility)? What tools will be valuable to them?
- Given emerging cyber-security threats, new approaches and reform to data stored and managed are necessary – water agencies, government regulators, and citizens have roles in furthering data security and management

Platform Creation and Community Data Instantiation

- Engagement from SWADE stakeholders to help design and develop platform
- Leverage current used tools, datasets, data formats
- Validation studies and scenario creation using SWADE deployed tools with partner agencies
- Survey of 15 CA water agencies on use of cyber platforms and data security concerns using ***EPA Water Cybersecurity Assessment Tool 1.0_0***



SWADE Internships

- Embedding students as interns at various agencies
- Undergraduate and high school research opportunities

Domain: Stormwater

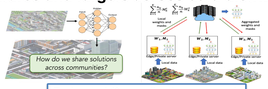
Focus Problem: Source Identification for Water Pollution Events and Anomalies

Resource-efficient Monitoring

- Detecting anomalies using existing field- and lab-based monitoring is costly (staff time, lab costs), ineffective (< 0.1%), and often delayed
- Leverage real-time water quality data from sensors for actionable decision-support, e.g., source identification
- Deployment and operational costs: underwater installation, human grab sampling, site visits, laboratory tests
- Efficiently switch between coarse-grained (efficient, less accurate) and fine-grained (expensive, more accurate) monitoring

Generalizable Data Analytics

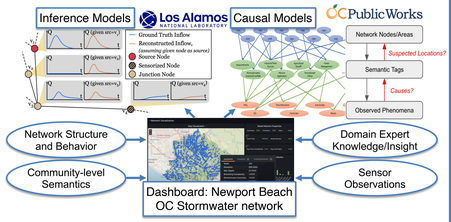
- Train ML models to discard location-specific data biases and allow other agencies and communities to reuse successful models
- Exploit stable correlations between invariant features and targets across data distributions



Federated Generalization approach

Network Structure and IoT Instrumentation

- Tradeoff between budget and quality of capturing heterogeneous anomaly events (e.g., spills)
- Leverage topological and empirical network properties to propose sample sensor deployments that balance coverage, traceability, and budget
- Integrate domain expert insights on sensor types/thresholds for anomaly ID and prioritization
- Employ physics-informed simulations and backwards inference models to identify potential sources of heterogeneous anomalies
- Ongoing: Exploit causal models, domain-expert data and semantics to reason about suspected anomaly origins



Reliable Underground IoT Communication

- Study feasibility of underground wireless data transport using analytical/empirical models and simulations for signal strength from underground nodes in channels
- Ongoing: Apply time synchronization protocols to achieve reliable communication under varying channel conditions (e.g., soil/concrete, wet, etc.)

ACM BuildSys'23, PMC'21,'23, IEEE PerCom'22, SmartComp'20, ACS'19, ICCPS'19

Domain: Drinking Water

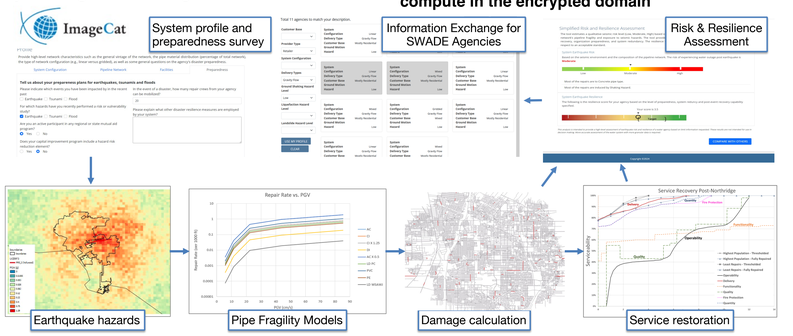
Focus Problem: Infrastructure Resilience to Extreme Events

Developing Risk and Resilience Tools

- Allow water agencies to internally assess seismic risks and resilience level, and share information / insights with other agencies
- Determine risk using only basic information from water agencies based on well-established modeling methodologies in engineering studies
- Compute resilience score for agencies as a benchmark for assessing status and refining adaptation strategies to meet or exceed industry standards
- Protect privacy by only sharing anonymized information and insights
- Ongoing: Refine tool and add practical features by working with water agencies

Improving Data Quality Across Data Silos

- Water districts vary greatly in service area (number of consumers, quality and granularity of metering), resulting in large disparities in data volume
- Ongoing: Reduce noise in metering data using imputation methods to fill in missing data readings common across networks (up to 40% of the data can be missing from network interruptions)
- Ongoing: Increase metering data accuracy in water districts and address temporal dynamics among users by leveraging deep learning; cluster similar behaviors amongst households in different districts using properties of frequency domain
- Ongoing: Develop secure, federated data-sharing strategies to encrypt data shared across silos and compute in the encrypted domain



ACM/IEEE Middleware'19, ICCPS'18, ICDCS'17

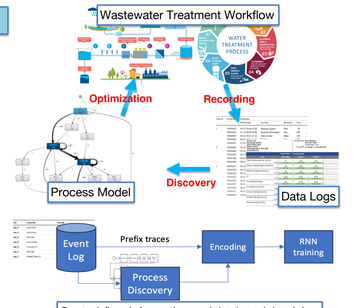
SIGSPATIAL, ARIC'20, ACM/IEEE Middleware'19, ICCPS'18, ICDCS'17

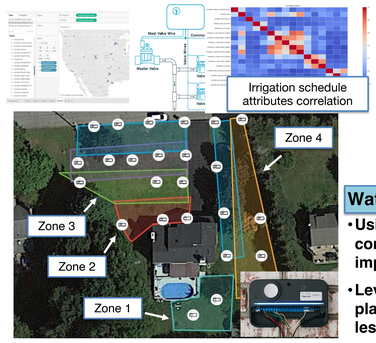
Domain: Wastewater

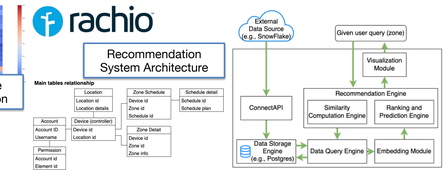
Focus Problem: Process Mining for Wastewater Treatment Plants

Mining Implicit or Hidden Knowledge

- Extract implicit knowledge and experience from historical treatment event logs
- Model discovered knowledge or structural properties (e.g., interdependent flows, timing constraints)
- Improve the practical wastewater treatment workflow based on discovered knowledge and experience
- Mining control-flow structural information from historical treatment event logs for improved predictive process monitoring
- Control-flow information encoding strategies for recurrent neural network training
- Ongoing: Generalize process mining techniques into other water domains







Water Conservation and Reuse

- Using real-life water utility data from a smart sprinkler company to understand user watering trends and identify important events that change user watering behavior
- Leverage recommendation models for irrigation schedule planning to optimize water usage, and allow users with less experience to benefit from seasoned users

IEEE COMPSAC'20, '22, JSA'22, IEEE CBI'21, IEEE AIKE'20

https://sites.uci.edu/swade