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

NSF 1952247 Lead PI : Nalini Venkatasubramanian, UC Irvine Award Type IRG, Solicitation Yr 2019

IT Research

Nalini Venkatasubramanian, Sharad Mehrotra, UCI
Shangping Ren, SDSU
Andrew Chio, Roberto Yus, Praveen Venkateswaran, Roberto Yus, UCI
Zhenyu Zhang, Ryan Hildenbrant, SDSU

Social Policy Research

David Feldman, UCI (lead)
Shannon Roback, Water UCI

Water Resource Research

Ron Eguchi, Marina Mendoza (ImageCat Inc)
Paul Anderson, Illinois Inst. of Tech.
Junjie Zhu, Princeton University

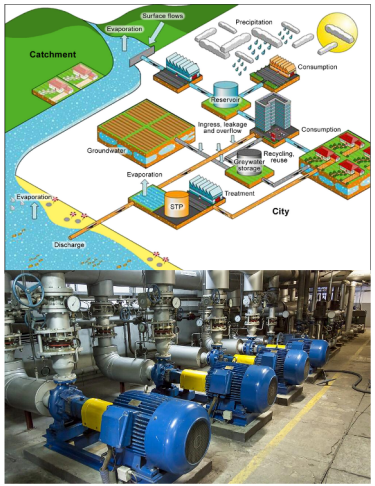
SWADE Technology Partners & Community Advisory CAB

SWADE: Project Motivation

Water infrastructure is aging and becoming increasingly complex

- Multiple Infrastructures, and dedicated regulatory and compliance agencies

Drinking water



Wastewater

Stormwater

- Data and structural information siloed within agencies and systems
- Decision-makers (agencies, policy-makers) need tools to interpret data, identify problems and take actions

SWADE: Project Overview

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

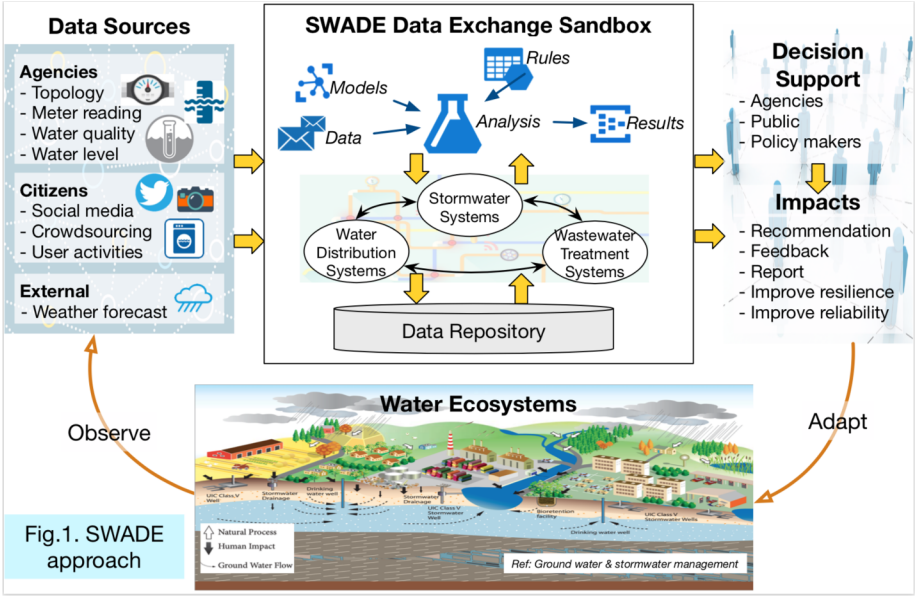
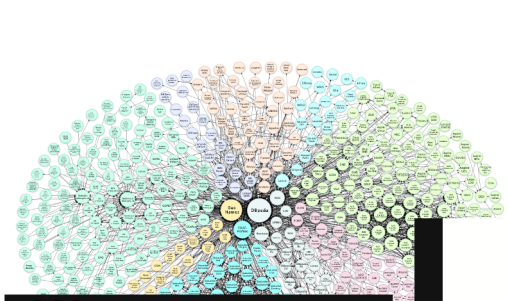
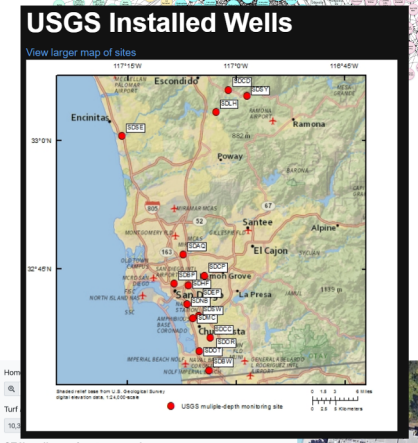


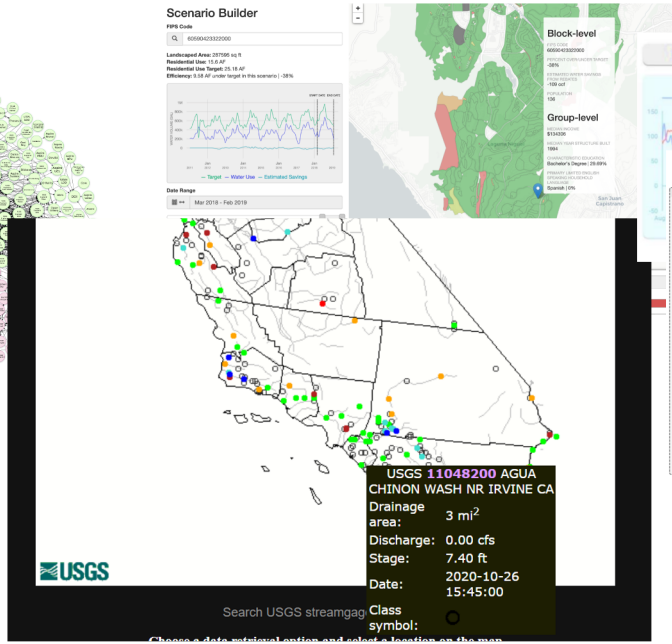
Fig.1. SWADE approach

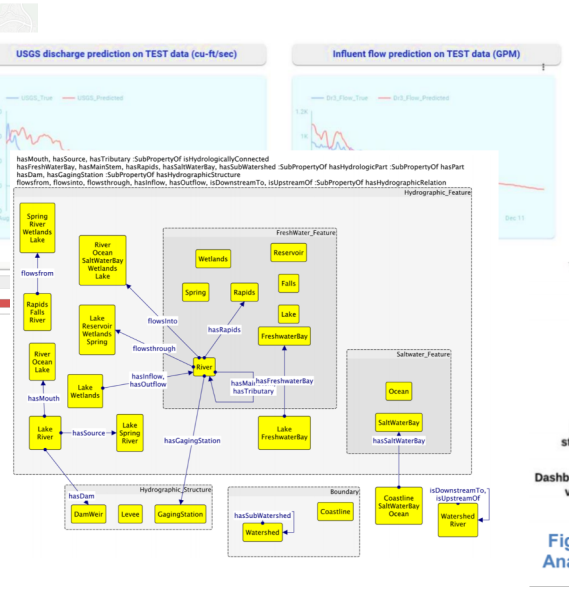
- **Model** the **different water infrastructures** and interconnections under normal operation or extreme events
- Determine barriers (societal, policy/regulatory) on gathering, sharing and usage of data.
- Identify challenges in translating (restricted) data to meaningful information for timely decision making, especially under large disruptions
- Current approaches: (1) requires significant effort to acquire and understand data; (2) information processing delays; (3) high levels of data redundancy; (4) lack of infrastructure for data exchange

SWADE: Project Evolution









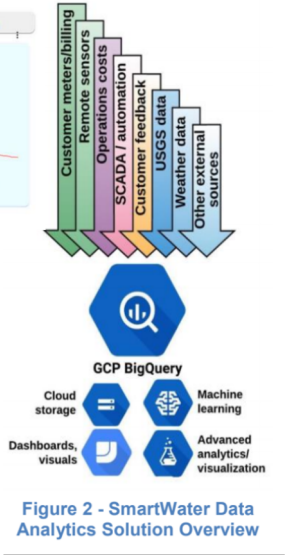


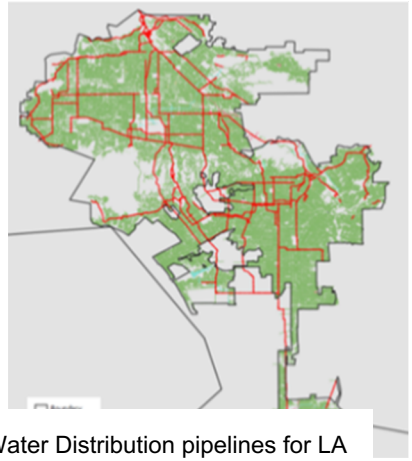
Figure 2 - SmartWater Data Analytics Solution Overview

- Many diverse, scattered sources of information
- Strategy:
  - Work with community partners to identify focus problems
  - capture and represent relevant concepts and data in a data architecture.

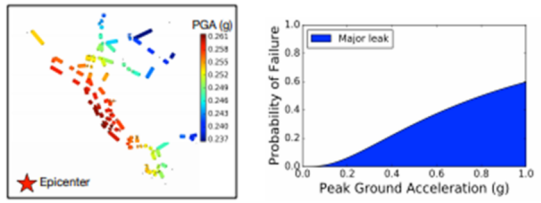
SWADE: Drinking Water Systems

Focus Problem :
Infrastructure Resilience to Extreme Events

- Exploit robust simulations to identify seismic event scenarios with significant impact on water distribution networks
- Enable additional instrumentation and retrofit to provide rapid predictions for community safety



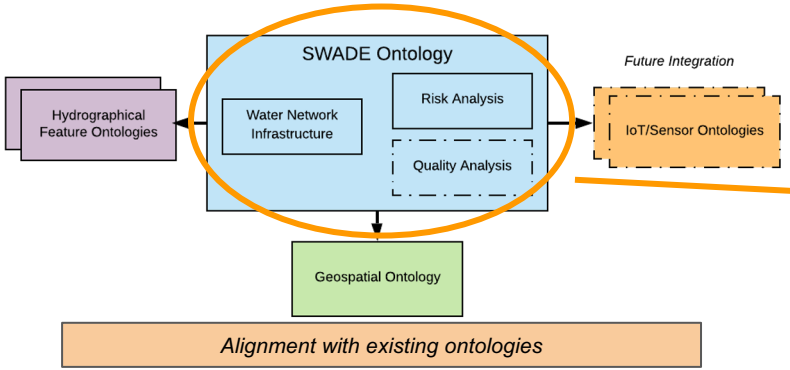
Water Distribution pipelines for LA

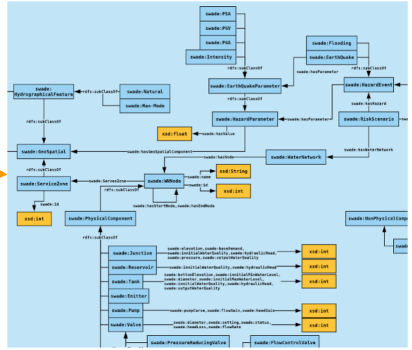


(a) Peak Ground Acceleration (b) Fragility Curve

Fig. 7: (a) PGA of pipelines for a magnitude of 5.5 earthquake. (b) Fragility curve for pipe damage.

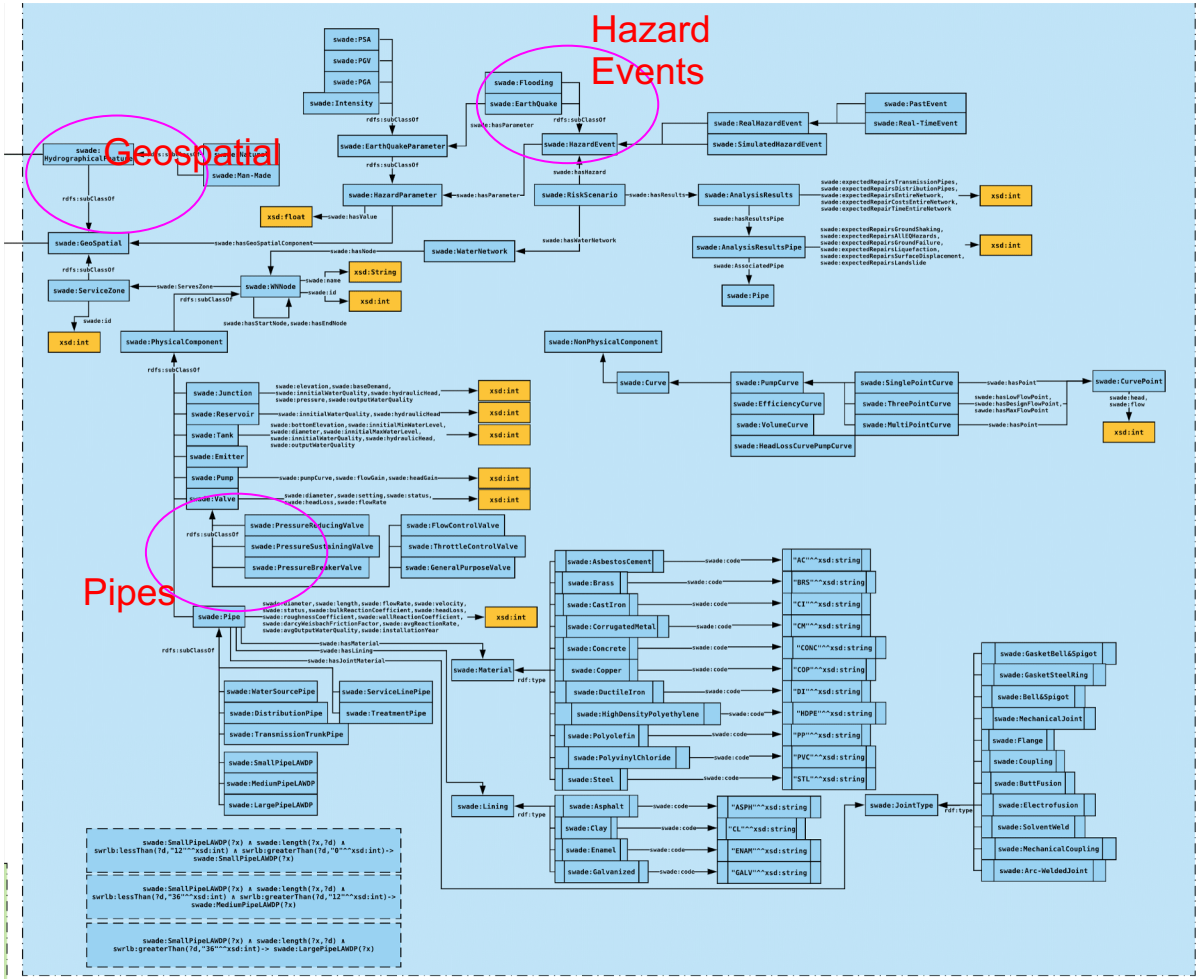
SWADE Ontology: Vocabulary to enable interoperability of water organizations





- **Current version**: Definition of water networks and components, geographical elements, hazards, and risk analysis results.

SWADE: Ontology



SWADE - Wastewater Systems

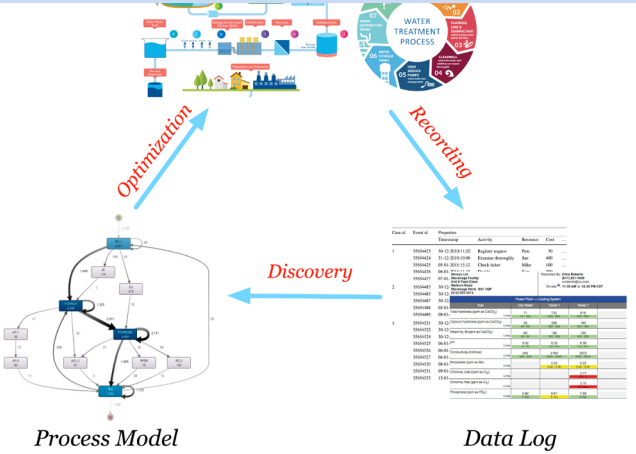
Focus Problem :
Process Mining for Wastewater Treatment Plants

Mining Implicit or Hidden Knowledge

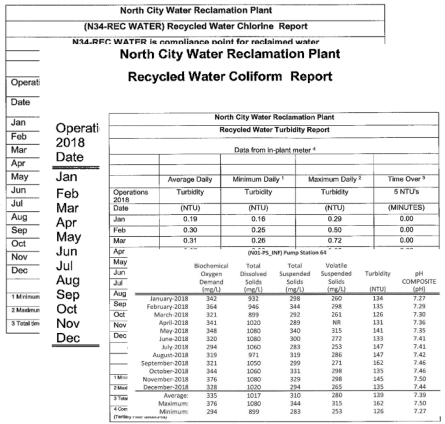
- Extract implicit knowledge and experience from historical treatment event logs
- Discover knowledge or structural properties (timing constraints, interdependent flows)
- Use to improve practical wastewater treatment workflow

Improving operational efficiency with soft-sensors

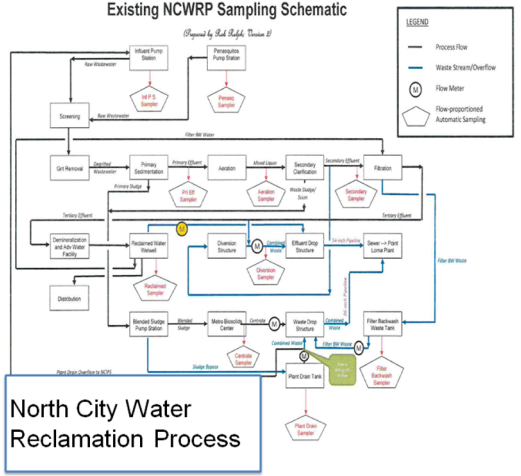
- Reduce energy demands caused by excess aeration without sacrifice in effluent quality.
- Combine data from physical sensors with regression methods.



Wastewater Treatment Event Logs



Existing NCWRP Sampling Schematic



SWADE - Stormwater Networks

Focus Problem :
Source Identification for Dry weather monitoring

Sensor Placement

- Tradeoff - Budget vs. Quality of monitoring
- Need to quickly detect high impact events

Resource-efficient monitoring

- High deployment & operational costs - underwater installation, human grab sampling.
- Coarse-grained monitoring (efficient, less accurate) vs. fine-grained monitoring (expensive, high accuracy)

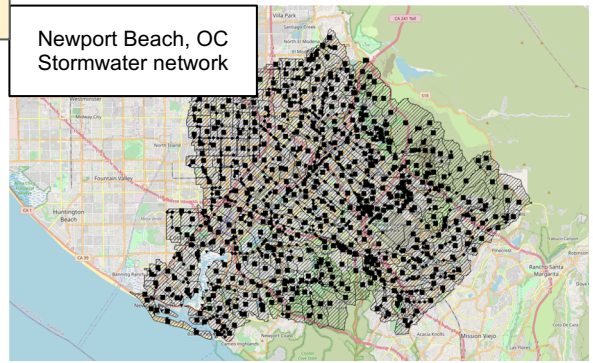
Network Structure and Time Series Analysis

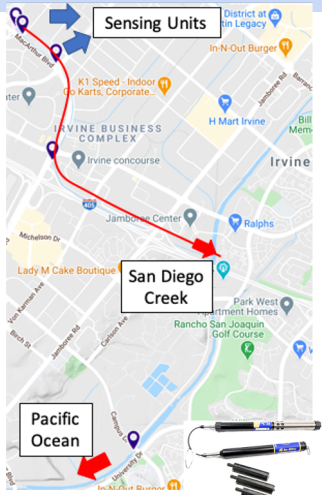
- Detect contamination events and sources rapidly
- Understand role of network structure and its relationship to data generated using time-series analytics

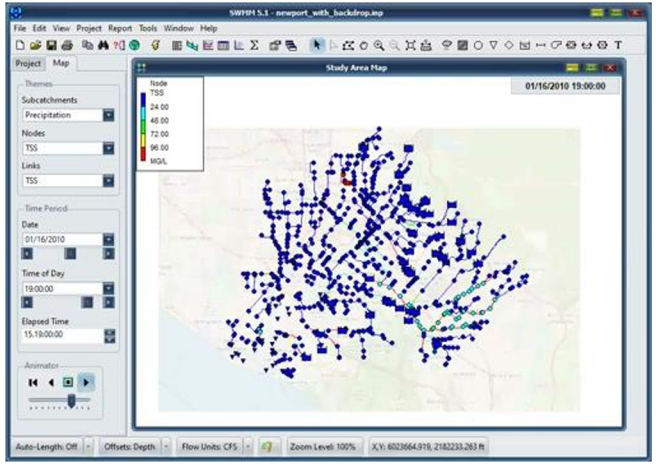
Generalizable Data Analytics

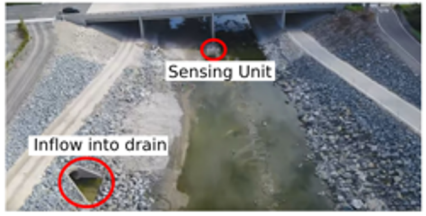
- Allow other agencies/communities to reuse successful models
- Train ML models that are robust to location-specific data biases

OC Public Works









Community Engagement

- **Stakeholder workshops and data challenge events**
  - Workshop to survey stakeholders (agencies, policy makers, academics, industry groups, etc.)
  - What tools will be valuable to them?
  - What are their concerns (privacy, data sharing, accessibility)?
  - What tools can be build to facilitate cooperation?

- **Engagement for platform creation and community data instantiation**
  - Engagement from SWADE stakeholders to help design and develop platform
  - Leverage currently used tools, datasets, and data formats.

- **Validation studies with Community Water Networks**
  - Scenario creation with partner agencies
  - Deployment of SWADE tools for analytics

- **SWADE Internships**
  - Embedding students as interns at various agencies.
  - Undergraduate and High-school research opportunities

SWADE: Ongoing work

- System-specific milestones
  - Drinking Water: Create vocabulary to represent drinking water systems and water quality analysis
  - Wastewater: Examine data representation needs for operational efficiency
  - Stormwater: Integrate time-series analytics with network structure analysis, explore generalization methods to improve robustness of ML techniques
- Identify relationships and interdependencies between different water infrastructure systems
- Expand data representation and ontology to capture multiple relationships