

20th IWA Leading Edge Conference on Water and Wastewater Technologies

1-5 June 2026 | Houston, Texas – USA



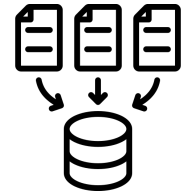
A Prototype Ontology and Data Management Platform for Water Treatment Systems

Fletcher T. Chapin, Umut Mete Saka, Lazlo Paul, Avia Anwar, Scott Struck, Meagan S. Mauter, **Gabe Fierro**

inspiring change

DATA MANAGEMENT IS A MAJOR HURDLE IN DEPLOYING DIGITAL WATER SOLUTIONS

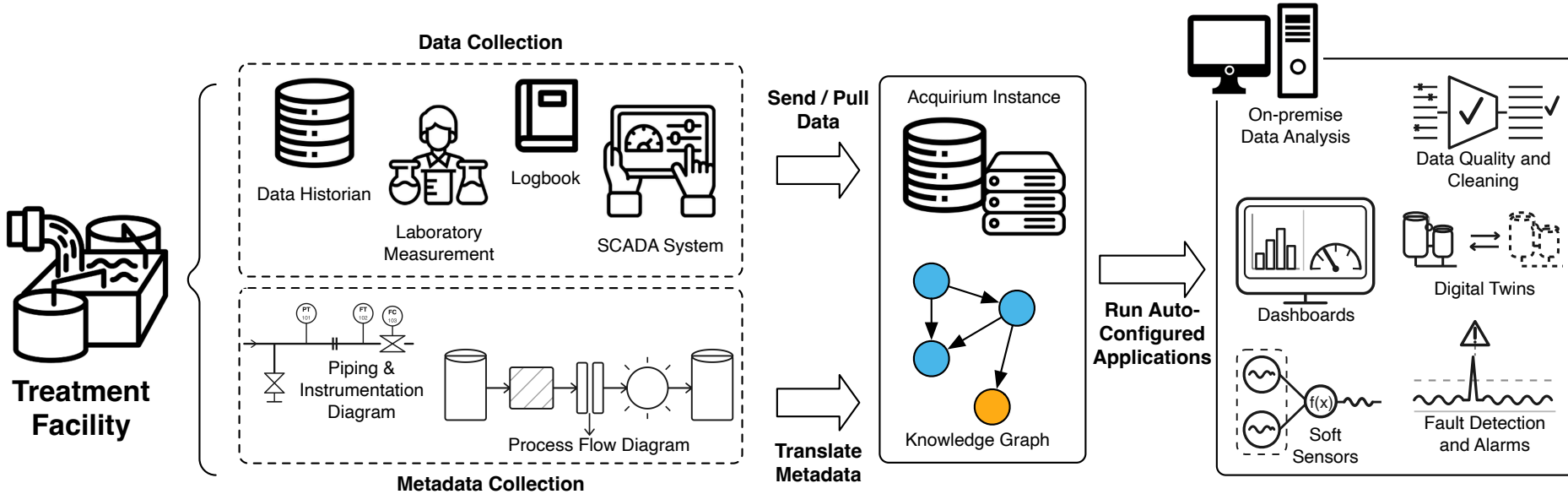
- Gathering data comprises **~40% of the effort** of research and data analyses projects in the water sector [Aguado Garcia et al., Villez et al.]
- Data management accounts for **most of the cost** of digital twin development and maintenance [Foundational Research Gaps and Future Directions for Digital Twins]
- This means experts spend valuable time, \$\$\$ on automatable tasks



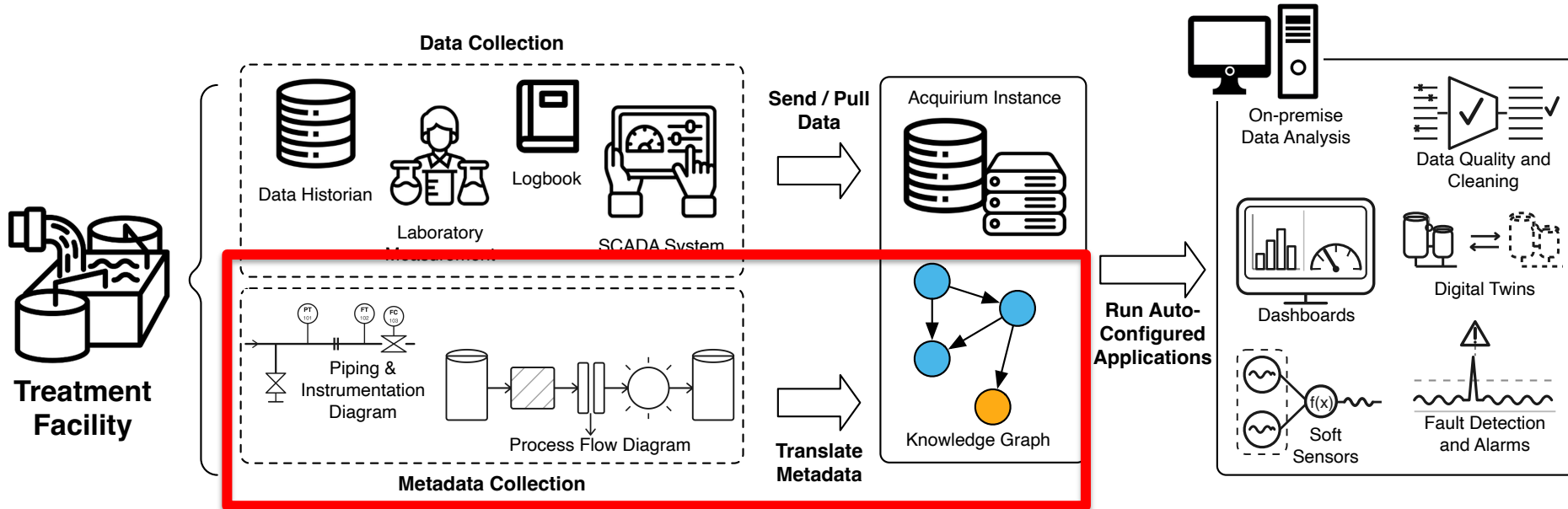
DATA MANAGEMENT CHALLENGES

- **Data (un)availability:** the ease of obtaining data in a usable, structured format
- **Data (mis)identification:** determining what data is available and which data is suitable for an application
- **Data fragmentation:** data is spread across many databases, and sometimes multiple sources measure the same quantity
- **Data quality:** broad term for frequency, precision, and correctness measurements; data sanitation, transformation, and validation

MANAGING DATA AT A REAL-WORLD WATER TREATMENT FACILITY

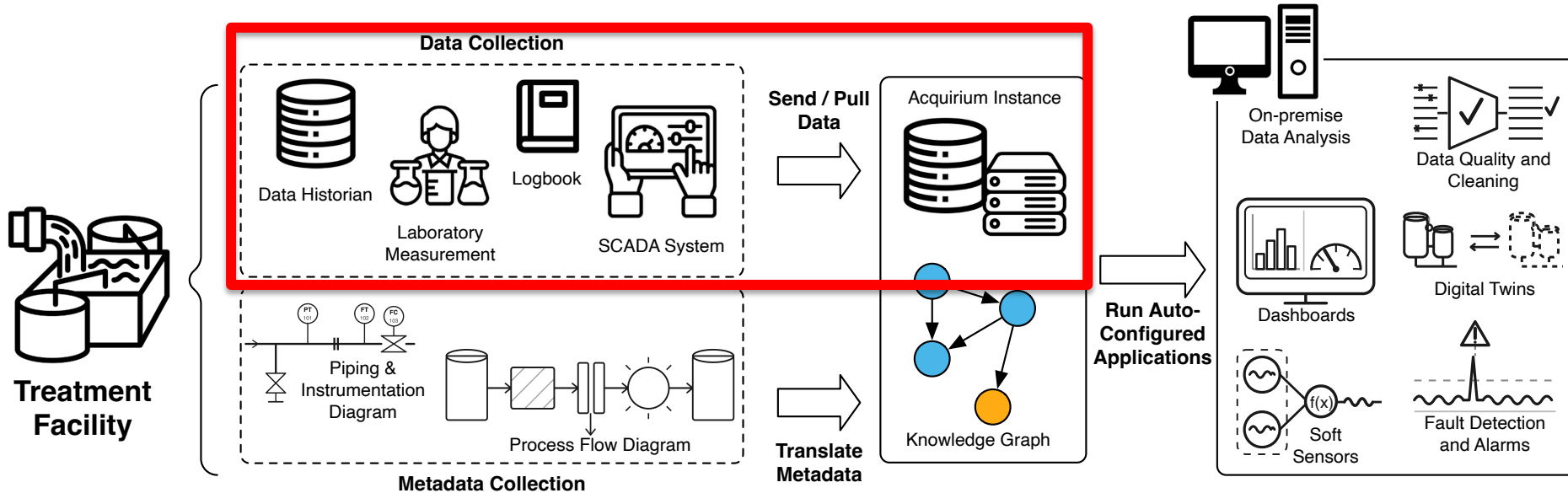


MANAGING DATA AT A REAL-WORLD WATER TREATMENT FACILITY



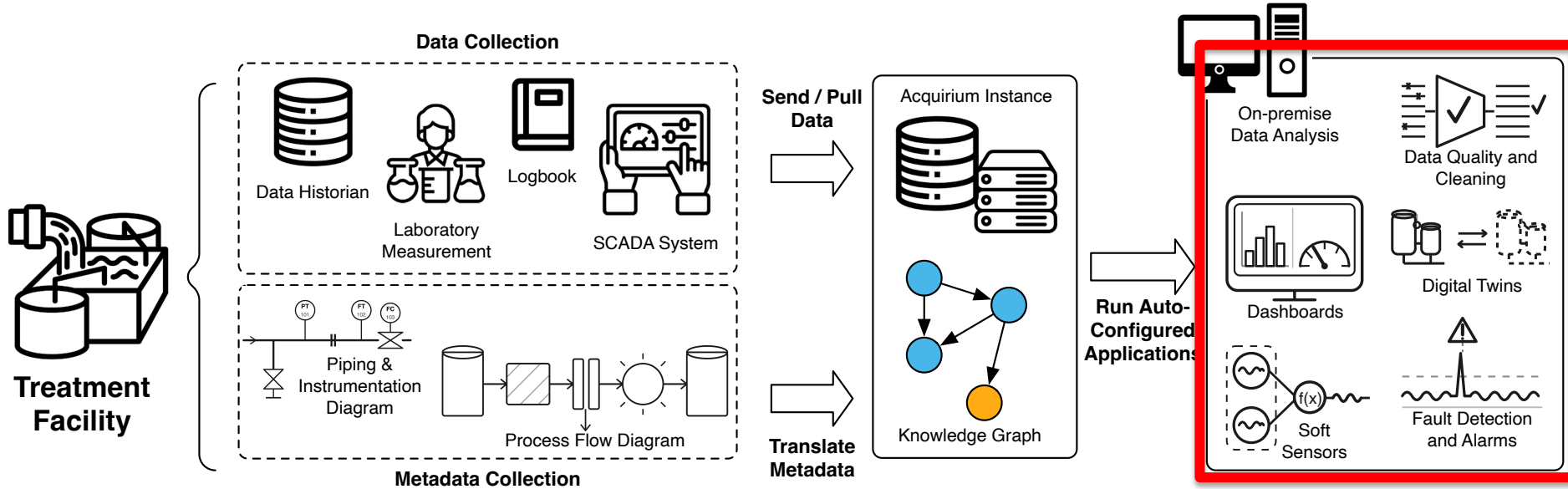
1. Standardize a single representation of site **metadata**, incorporating information from P&ID, PFD, SCADA, and other documents

MANAGING DATA AT A REAL-WORLD WATER TREATMENT FACILITY



2. Embed **data references** that connect disparate data sources (lab, SCADA, computed, etc) to their location within the site/train/process

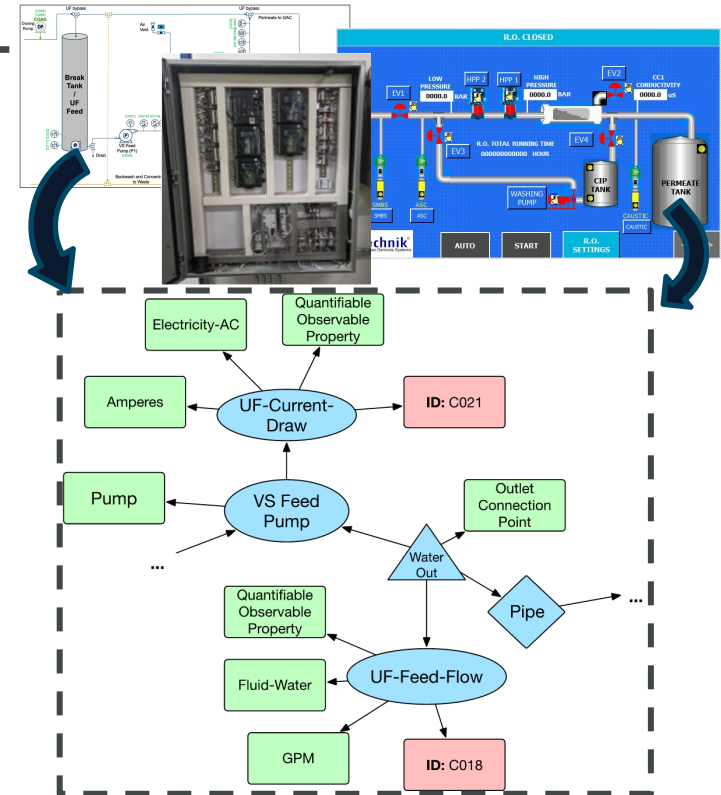
MANAGING DATA AT A REAL-WORLD WATER TREATMENT FACILITY



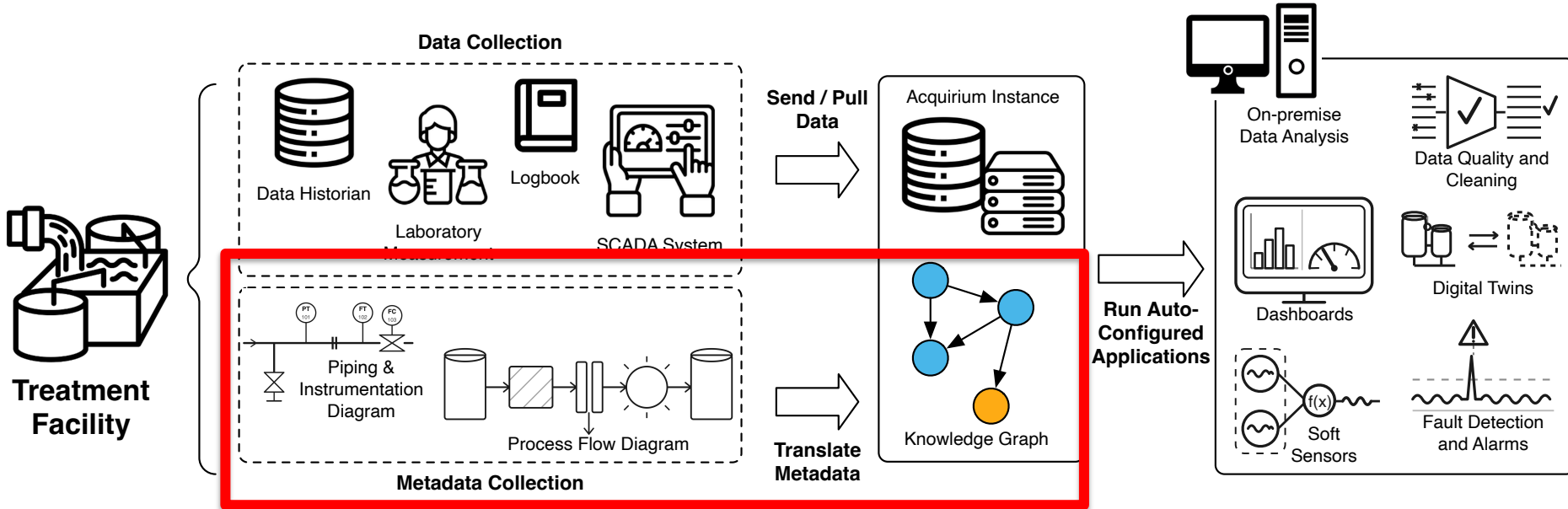
3. Develop **automated software processes** to conduct data cleaning, quality, analysis, KPIs

HOW CAN WE IMPROVE DATA MANAGEMENT PRACTICES?

- Use **knowledge graphs** as a **unified machine- and human-readable** representation of the plant and its data
 - Sourced from well-known standards, conventions
 - Machine-verifiable: enforce correct/interoperable usage
- Build on **free, open standards** for representation and exchange
- Interact through intuitive APIs and interfaces for data science, operation, etc



HOW TO DEPLOY WATR + ACQUIRIUM AT A REAL-WORLD WATER TREATMENT FACILITY

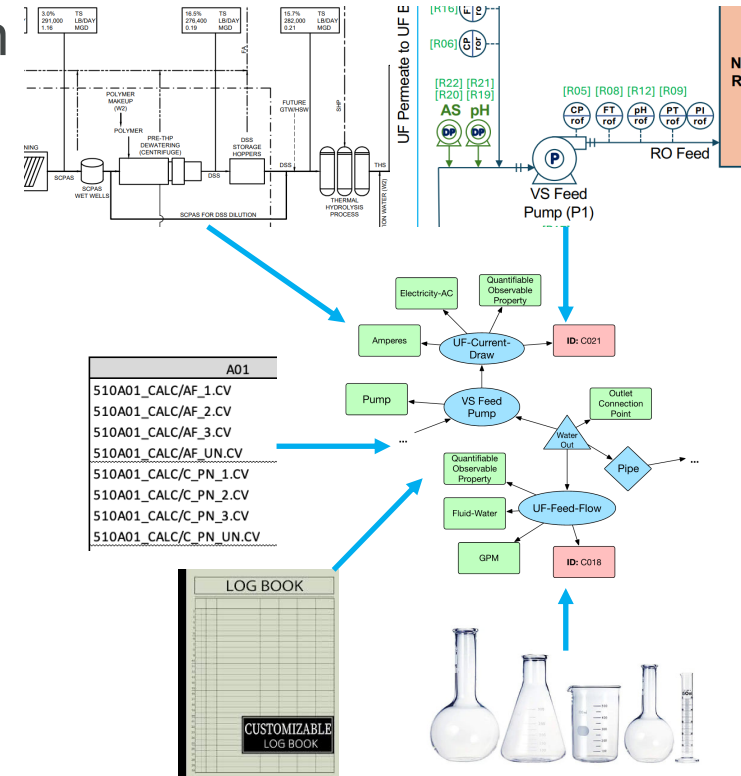


Step 1: Create
Knowledge graph

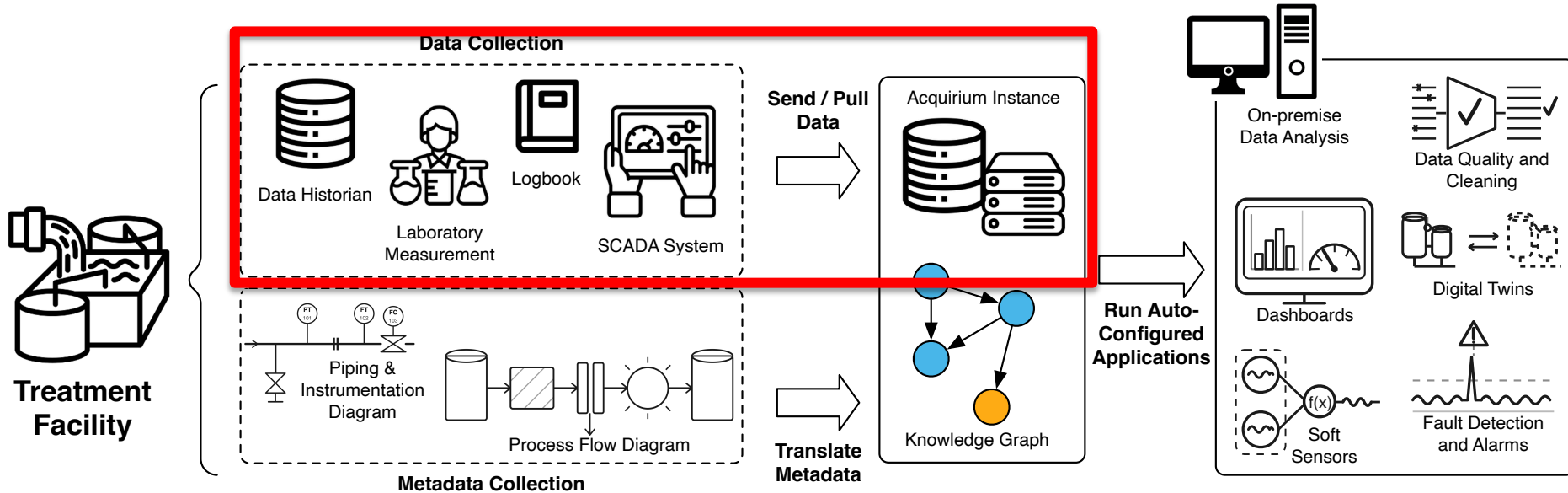
STEP 1: CREATE KNOWLEDGE GRAPH

1. Parse engineering docs to determine physical facility layout and composition
2. Cross reference telemetry and instrumentation to connect tag names to the graph
3. Incorporate logbooks, laboratory measurements, and other analog data (e.g., sampling points and frequency)

We are building tools to automate and optimize this process



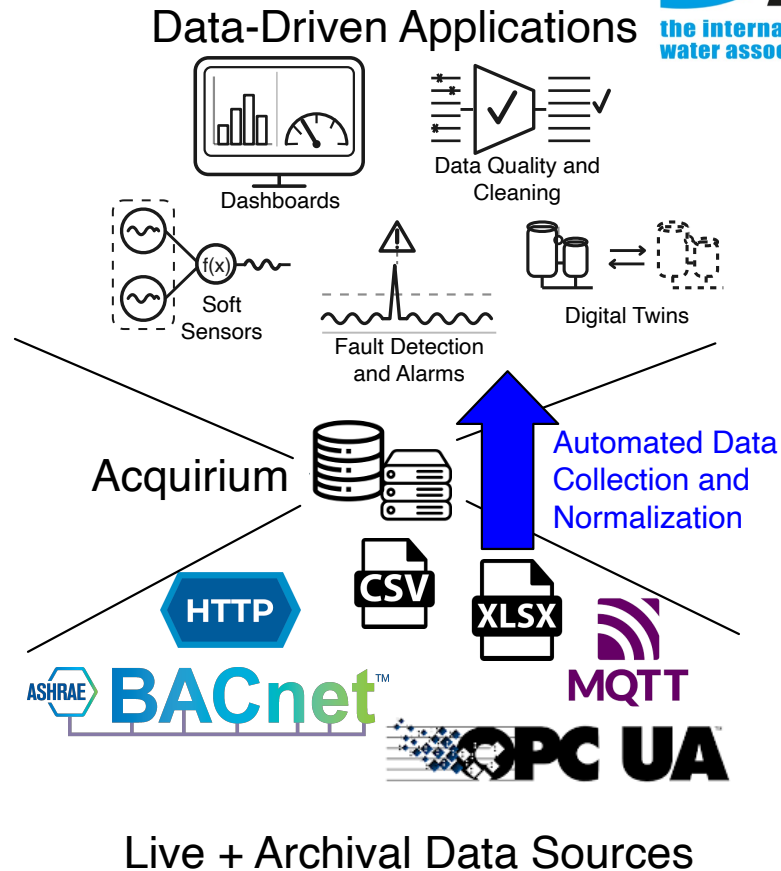
HOW TO DEPLOY WATR + ACQUIRIUM AT A REAL-WORLD WATER TREATMENT FACILITY



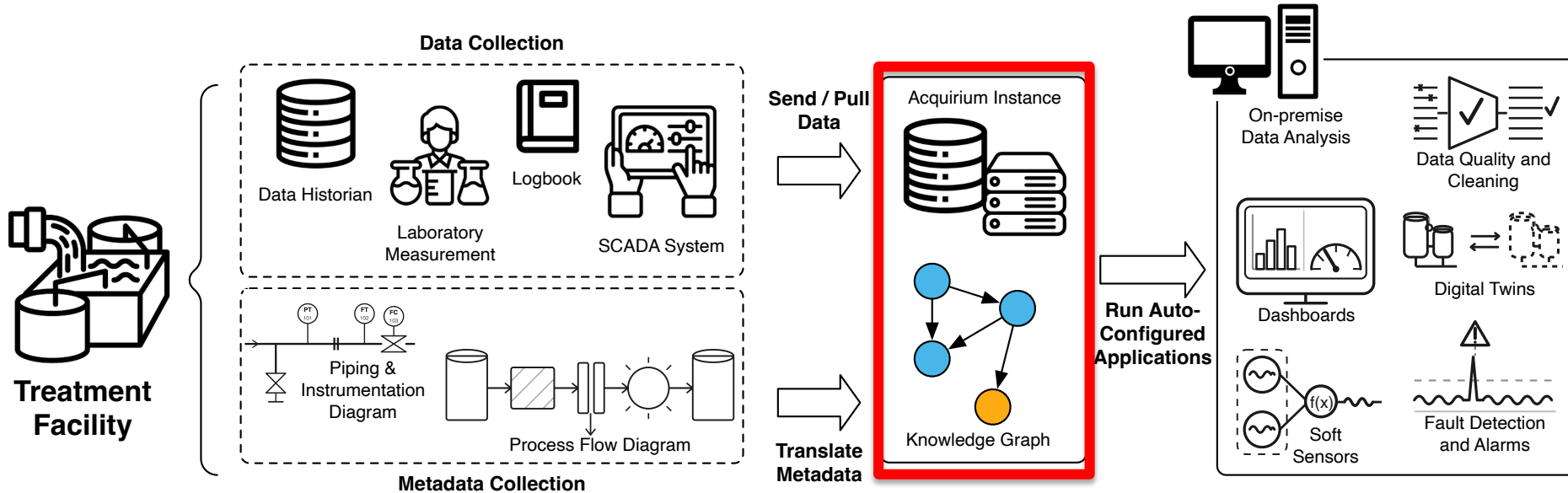
Step 2: Initialize database
with historical and live data

STEP 2: INITIALIZE DATABASE

- Automated data collection
 - Just point Acquirium at your data
 - Multiple protocol support
 - Extensible data ingestion: easily add new data sources
- Deployable on-prem, can be run "offline" → ok to ingest sensitive local data
- Tracks **data provenance**: where did data come from, how was it cleaned, transformed, etc.



HOW TO DEPLOY WATR + ACQUIRIUM AT A REAL-WORLD WATER TREATMENT FACILITY



Step 3: Query Data and Develop/Run Applications

STEP 3: QUERY DATA

User input

Query:

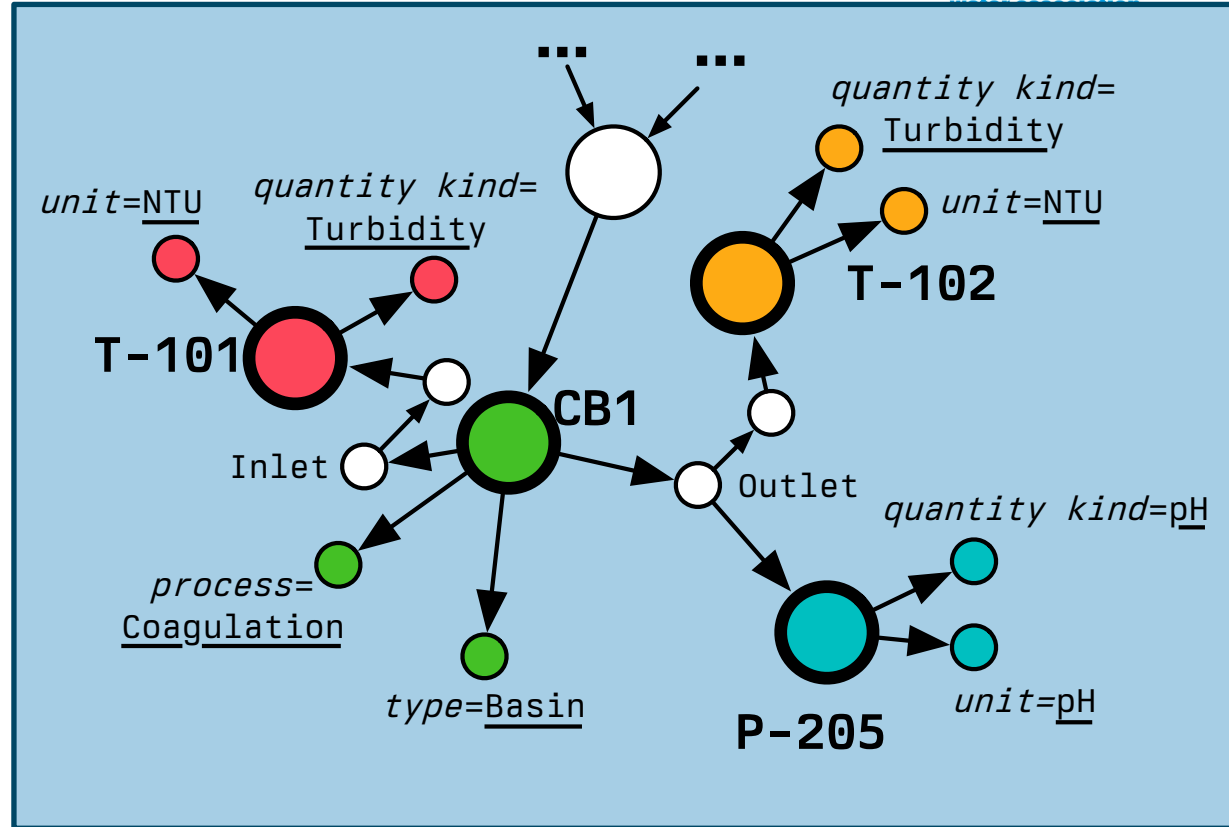
```
tanks.find_all_data()  
.downstream()  
.measuring(Turb. or pH)
```



Acquirium

Convert query to graph search

Process result of graph search



STEP 3: QUERY DATA

User input

Query:

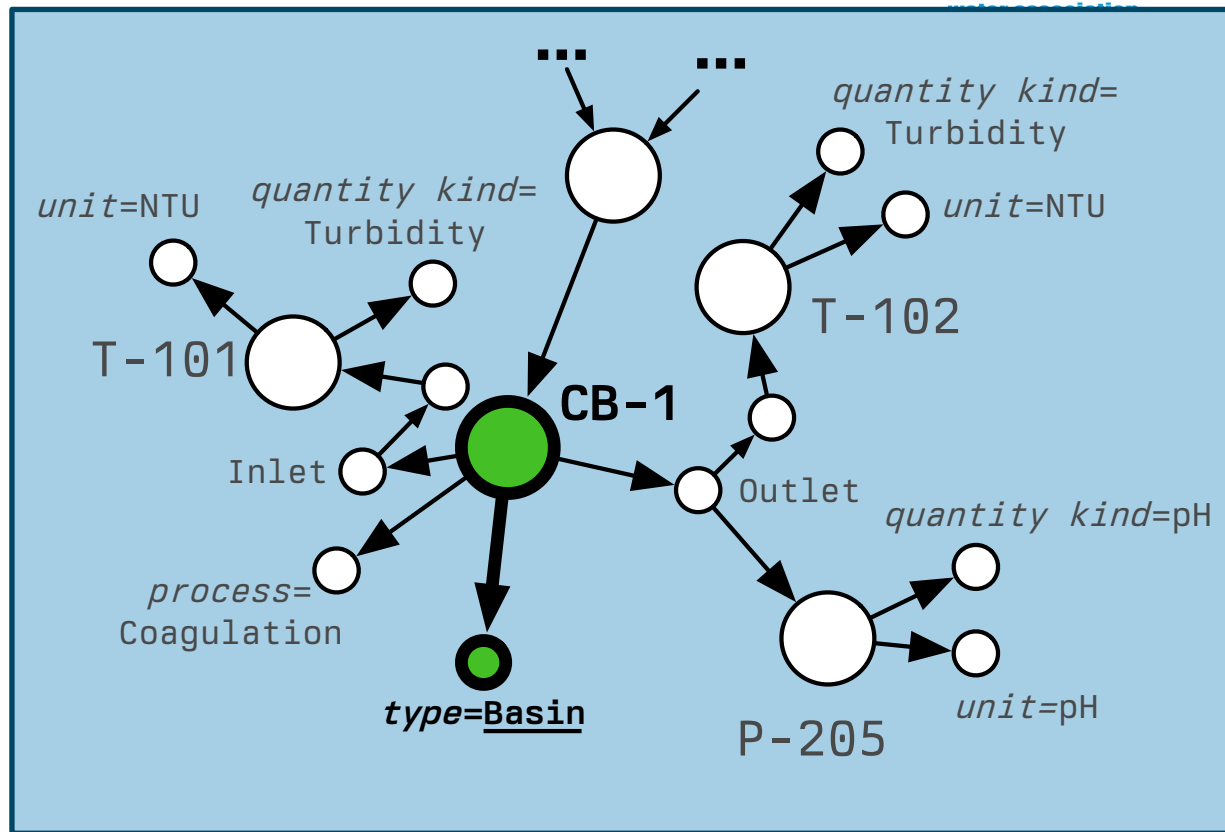
```
tanks.find_all_data()  
.downstream()  
.measuring(Turb. or pH)
```



Acquirium

Convert query to graph search

Process result of graph search



STEP 3: QUERY DATA

User input

Query:

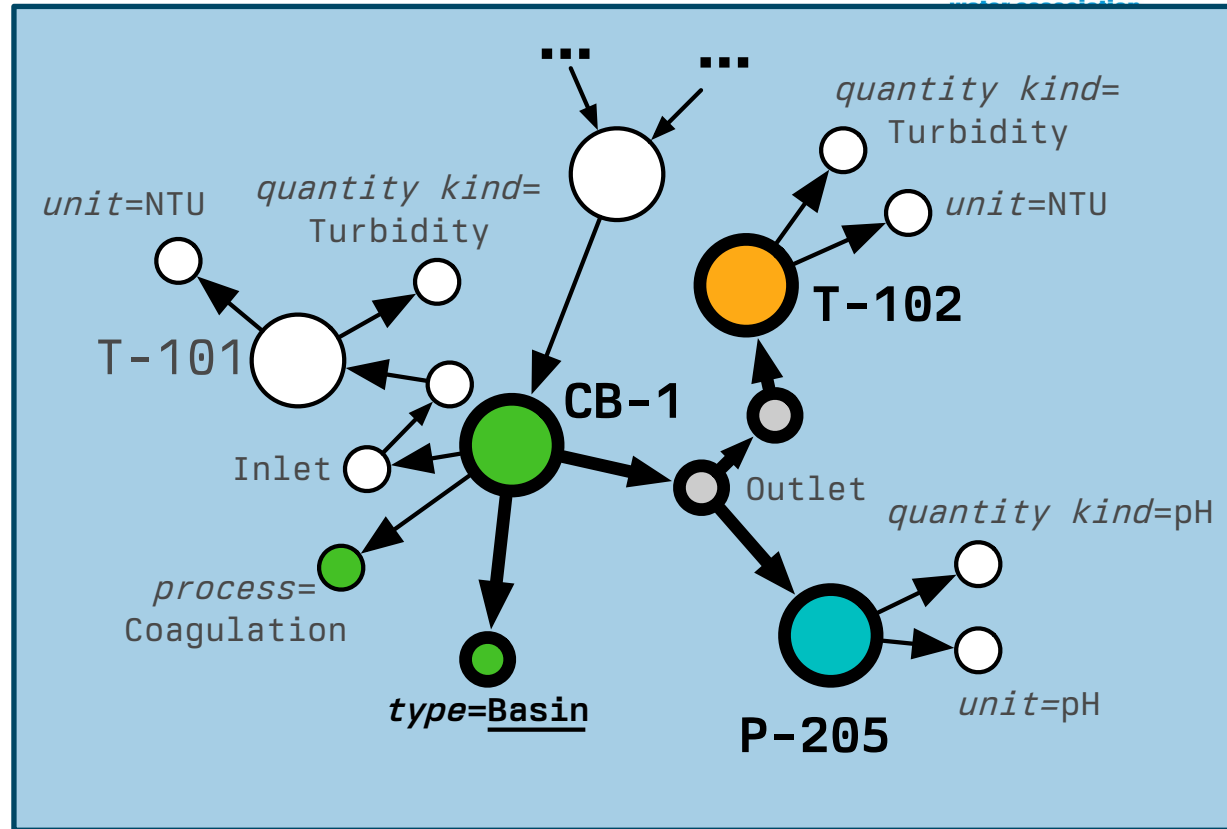
```
tanks.find_all_data()
      .downstream()
      .measuring(Turb. or pH)
```



Acquirium

Convert query to graph search

Process result of graph search



STEP 3: QUERY DATA

User input

Query:

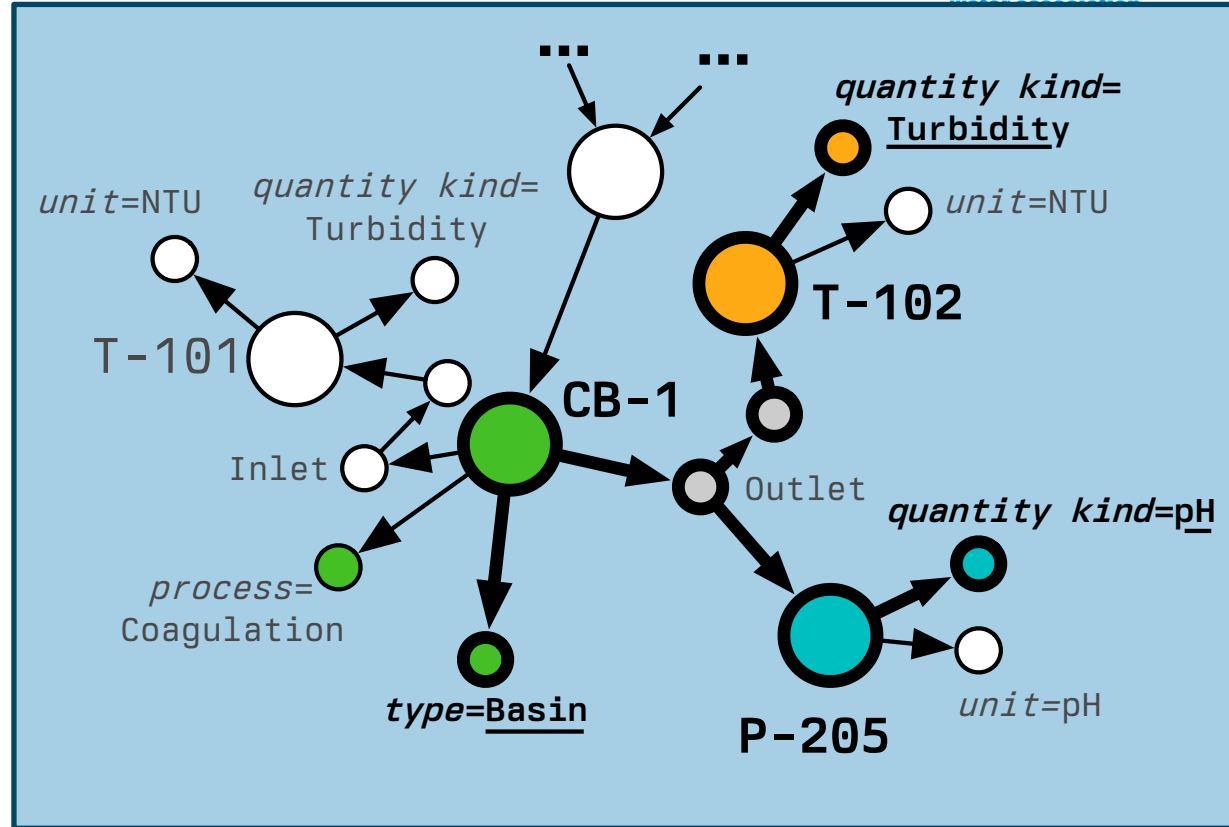
```
tanks.find_all_data()  
.downstream()  
.measuring(Turb. or pH)
```



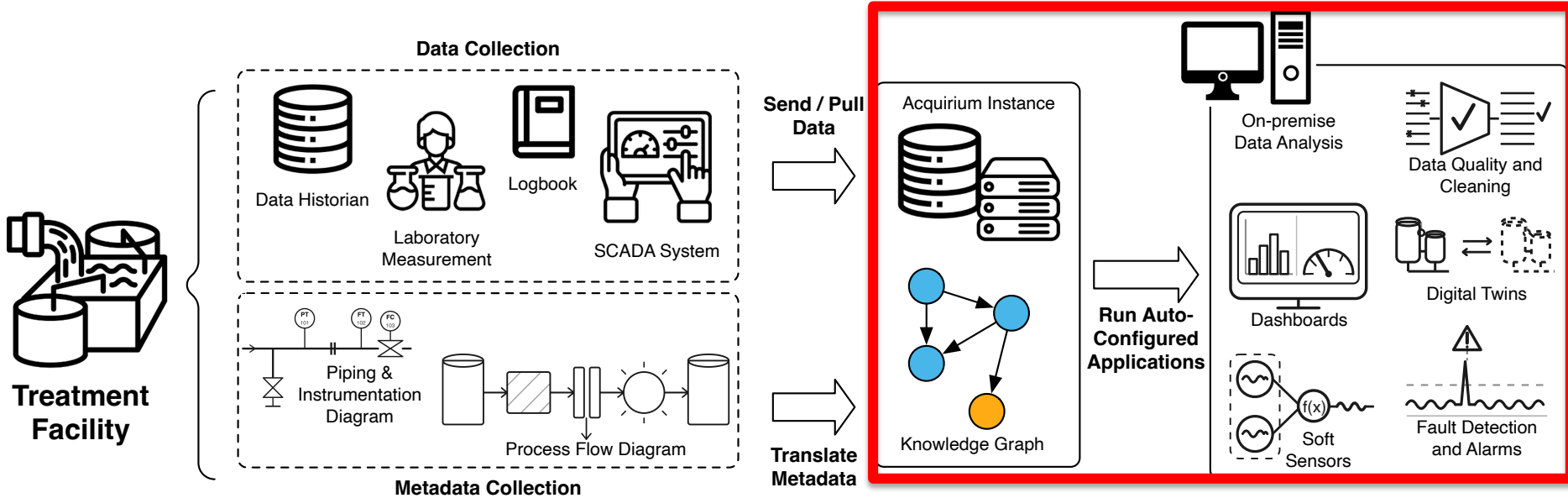
Acquirium

Convert query to graph search

Process result of graph search



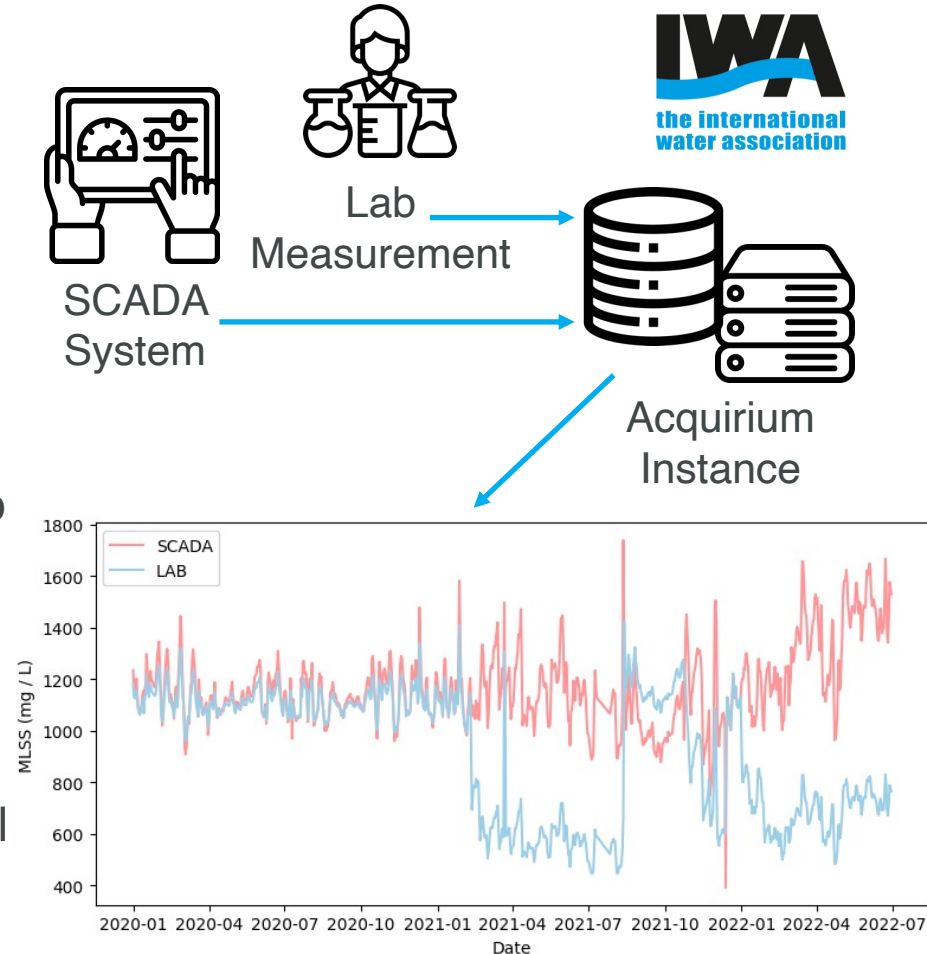
HOW TO DEPLOY WATR + ACQUIRIUM AT A REAL-WORLD WATER TREATMENT FACILITY



Step 3: Query Data and Develop/Run Applications

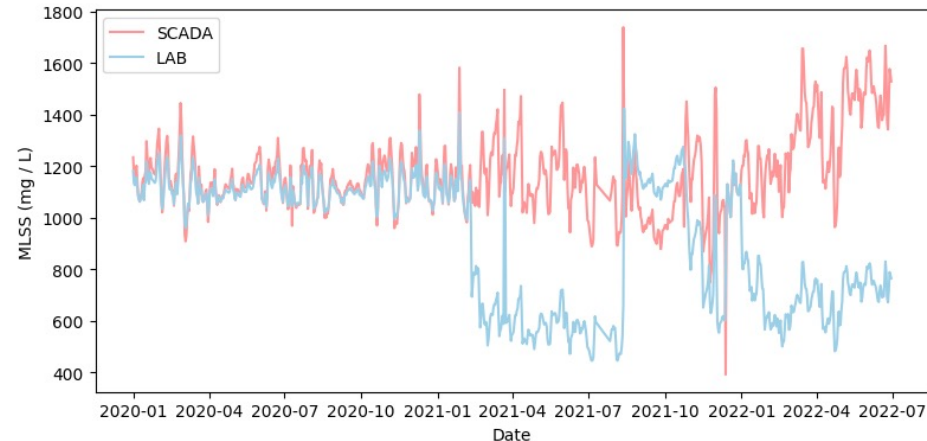
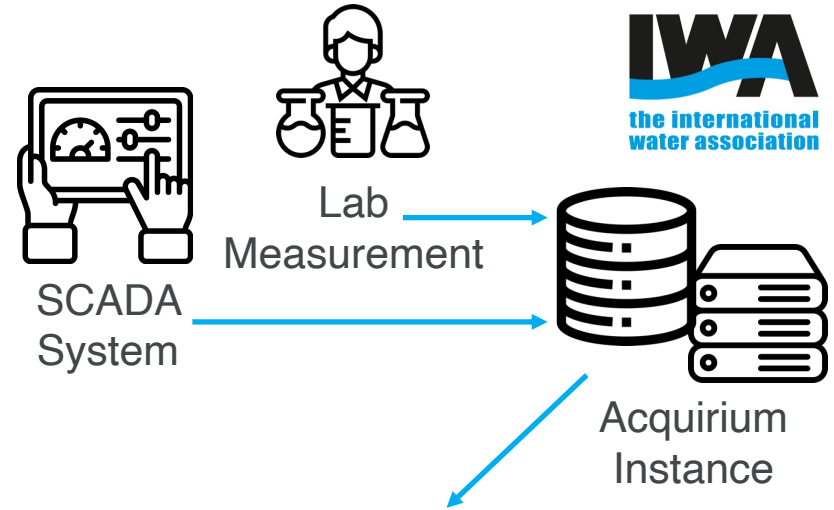
STEP 4: ANALYZE DATA

- In this example, we reconcile laboratory and sensor data for the same quantity kind and medium (i.e., mg/L of MLSS)
- Sensor data is recorded on 15-minute intervals in a SCADA system, while lab analysis is only performed every few days and stored in a separate database
- Acquirium can ingest data from both repositories and reconcile the temporal differences using built-in methods



STEP 4: ANALYZE DATA

- **Data exploration:** Locate data by substance, measured quantity, unit, upstream/downstream equipment, unit process, etc.
- **Defragmentation:** Combine logbook, laboratory samples, SCADA data, and other sources
- Apply soft sensors to clean, transform, combine, predict data
- Develop re-usable software packages
 - Mass balance, sensor recalibration, setpoint optimization, etc...



WATR + ACQUIRIUM CAN MANAGE DATA FOR A WIDE RANGE OF APPLICATIONS



- Analysis of multi-source and fragmented datasets
- Threshold based sensor monitoring
- State transition and event detection
- Automated fault detection and isolation
- Optimal sensor placement
- Data validation and quality assessment
- Automatic detection of newly installed sensors
- LLM-based question/answering over operational data
- Sensor drift detection and correction
- Building digital twins

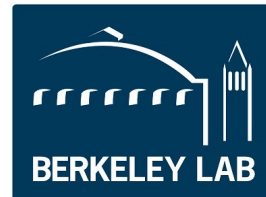
ACKNOWLEDGEMENT & CONTACT



watermetadata.org



PI: Dr. Gabe Fierro (gtfierro@mines.edu)
Assistant Professor
Department of Computer Science
Colorado School of Mines
National Lab of the Rockies



Fletcher Chapin



Mete Saka



Lazlo Paul



Avia Anwar



Scott Struck



Meagan Mauter

We thank Tzahi Cath, Lisa Tomenendal, and the rest of the NAWI 3.20 team for their help in providing data and metadata for our case studies. We also thank the team members of NAWI 3.13, 3.15, 3.17, and 3.22 for conducting investigatory interviews with us in the first stage of the project. This material is based upon work supported by the National Alliance for Water Innovation (NAWI), funded by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Industrial Efficiency and Decarbonization Office under Funding Opportunity Announcement DE-FOA-0001905.