POSTER

Poster Abstract: Acquirium: A Data-Metadata Management Framework for Water Treatment Systems

UMUT METE SAKA, Colorado School of Mines, Golden, CO, United States

LAZLO PAUL, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

FLETCHER CHAPIN, Stanford University, Stanford, CA, United States

SCOTT STRUCK, National Renewable Energy Laboratory, Golden, CO, United States

AVIA ANWAR, Colorado School of Mines, Golden, CO, United States

GABRIEL TOMAS FIERRO, Colorado School of Mines, Golden, CO, United States

Open Access Support provided by:

Colorado School of Mines

Lawrence Berkeley National Laboratory

Stanford University

National Renewable Energy Laboratory

Published: 19 November 2025

Citation in BibTeX format

BUILDSYS '25: 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation
November 19 - 21, 2025
CO, Golden, USA

Conference Sponsors:
SIGENERGY

BUILDSYS '25: Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (November 2025)
https://doi.org/10.1145/3736425.3772123
ISBN: 9798400719455

Poster Abstract: Acquirium: A Data-Metadata Management Framework for Water Treatment Systems

Umut Mete Saka
saka@mines.edu
Colorado School of Mines
Golden, CO, United States

Lazlo Paul
LPaul@lbl.gov
Lawrence Berkeley National Lab.
Berkeley, CA, United States

Fletcher Chapin
fchapin@stanford.edu
Stanford University
Stanford, CA, United States

Scott Struck
scott.struck@nrel.gov
National Renewable Energy Lab.
Golden, CO, United States

Avia Anwar
avia_anwar@mines.edu
Colorado School of Mines
Golden, CO, United States

Gabe Fierro
gtfierro@mines.edu
Colorado School of Mines
Golden, CO, United States

Abstract

Increased digitization in water treatment plants leads to data management complexities. Machine-readable, semantic metadata models standardize system representations and offer easier data access. We propose a new framework, *Acquirium*, for managing data and metadata of water treatment plants.

CCS Concepts: • **Information systems → Data management systems**; **Information integration**.

Keywords: Water Treatment, Data-Metadata Management

ACM Reference Format:
Umut Mete Saka, Lazlo Paul, Fletcher Chapin, Scott Struck, Avia Anwar, and Gabe Fierro. 2025. Poster Abstract: Acquirium: A Data-Metadata Management Framework for Water Treatment Systems. In *The 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BUILDSYS '25), November 19–21, 2025, Golden, CO, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3736425.3772123

1 Introduction

Digitization of water treatment facilities has led to an explosion of IoT data from distributed sensors and devices. This data drives a wide range of applications, from predictive maintenance to energy optimization and fault detection [3], which are increasingly important given the regulatory pressure to enforce stricter water quality standards [4]. For these applications to be viable, however, they must be both technically effective and economically feasible to deploy.

BUILDSYS '25, Golden, CO, USA

ACM ISBN 979-8-4007-1945-5/25/11
https://doi.org/10.1145/3736425.3772123

A major barrier lies in categorizing, structuring, and converting raw sensor streams into usable formats which can consume over a third of the cost of data science projects in water treatment systems [5]. Mapping thousands of sensor points to their data streams while capturing relationships among components quickly becomes a significant engineering challenge. Further, these facilities often have fragmented data, i.e. the data spans over different locations, and formats.

Semantic metadata models use RDF graphs to encode metadata in a standard machine-readable form. **Ontologies** such as Brick [1] standardize the form and content of metadata models to ensure usability and interpretability. Applications use expressive graph queries (e.g. SPARQL) to extract metadata from the models. Metadata models contain links to timeseries data, not the data itself. By capturing relationships between system components, these models allow queries that combine both structure and semantics, e.g. *the flow rate of the pipe upstream of this filter*.

In practice, linking RDF metadata to underlying timeseries remains cumbersome. Bridging semantic models with database records typically requires mixing SPARQL and SQL, so metadata discovery and data retrieval become separate steps that demand expertise and manual preprocessing. Although these technologies improve interoperability, the workflow remains slow and raises barriers to adoption for engineers and site experts. To overcome these challenges we develop *Acquirium*, a data-metadata access and management framework for water treatment facilities. The framework's goal is to abstract away the details of the treatment process to allow the user to access only pertinent information.

2 Data Management Challenges

We designed *Acquirium* to address two data management challenges we sourced from interviews with engineers, researchers, and data scientists in the water treatment domain.

Data Access and Discovery: Enumerating all the data sources in a large CPS can be complex. Data sources have adhoc labels which may be hard to interpret. We use metadata graphs as indexes of all data sources in a CPS, contextualized

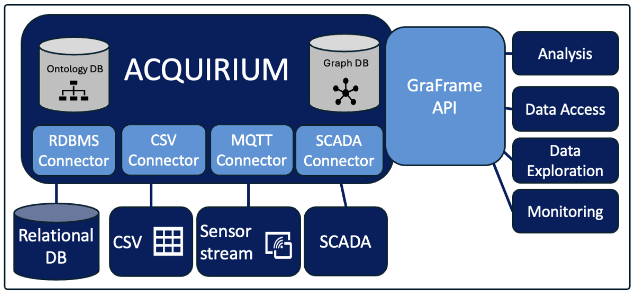

Figure 1. Acquirium system overview

by standard metadata which includes information on how to retrieve the data using connectors.

Data Fragmentation: CPS can have data spread over multiple databases and networks, each with its own protocol or format: SCADA systems, CSV files, relational databases, and others. Data also spans multimodal formats including numerical sensor streams, categorical labels, equipment metadata, and even images. We use our metadata models to integrate and contextualize each of these data sources.

We want to create a system that is usable by real practitioners, who have varying familiarity with metadata graph technologies. Acquirium's API takes inspiration from the rise of dataframes (e.g., Python's popular `pandas` library) as convenient interfaces for interactive and incremental exploration of large datasets. We hope to abstract away the essential complexity of RDF metadata models, but use that metadata to provide intuitive descriptions of data that practitioners can understand.

3 Acquirium

Acquirium (Fig. 1) is an open-source data management platform, written in Python. It wraps a graph database for metadata models and a timeseries database for CPS telemetry. It supports an array of data connectors, similar to sMAP[2], which let us integrate data from SQL databases, SCADA systems, fileshares, and other sources. Metadata models embed *external references* to these connectors to describe how to fetch data for a given data source. At the core of the system is *GraFrame*, a dataframe-inspired API which permits users to fetch metadata and data. Developers can use GraFrame to build data-driven applications without learning about RDF model. Internally, Acquirium uses the GraFrame to generate dashboards, visualizations, and fault detection pipelines.

GraFrame allows exploring a metadata graph by filtering on entity types and relationships between them. Users can discover data sources in the graph using their metadata, including the media, unit and location of the sensed physical property. Fig. 2 shows how to discover gallons-per-minute flow sensors for pumps downstream of tanks.

4 Conclusion

We introduce a data management platform, Acquirium, with an interactive dataframe-inspired API called GraFrame for exploring cyber-physical metadata and data at the same time.

```python
from acquirium import Deployment
WaTr = Namespace("urn:nawi-water-ontology#")
# Create a Deployment which represents the system
example_system = Deployment(metadata_graph="graph_model.ttl")
# Query the system for all Tanks, their Pumps and their data
(example_system.find_all(_class = WaTr.Tank)
               .find_related(_class = WaTr.Pump)
               .find_related_data_nodes()
               .filter_by_unit("GPM")) #Gallons Per Minute
pump_flow_nodes_latest_readings = example_system.latest()
for reading in pump_flow_nodes_latest_readings:
    if reading > threshold:
        print(f"Alert! High pump flow detected: {reading}")
```

Figure 2. Example Application

Acknowledgments

This material is based upon work supported by a NIST PREP fellowship and 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 DE-FOA-0001905. It was also funded by U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

[1] Bharathan Balaji, Arka Bhattacharya, Gabriel Fierro, Jingkun Gao, Joshua Gluck, Dezhi Hong, Aslak Johansen, Jason Koh, Joern Ploennigs, Yuvraj Agarwal, Mario Berges, David Culler, Rajesh Gupta, Mikkel Baun Kjærgaard, Mani Srivastava, and Kamin Whitehouse. 2016. Brick: Towards a Unified Metadata Schema For Buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, Palo Alto CA USA, 41–50. doi:10.1145/2993422.2993577

[2] Stephen Dawson-Haggerty, Xiaofan Jiang, Gilman Tolle, Jorge Ortiz, and David Culler. 2010. sMAP: a simple measurement and actuation profile for physical information. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, Zürich Switzerland, 197–210. doi:10.1145/1869983.1870003

[3] Molly C. Klanderman, Kathryn B. Newhart, Tzahi Y. Cath, and Amanda S. Hering. 2020. Fault Isolation for A Complex Decentralized Waste Water Treatment Facility. *Journal of the Royal Statistical Society Series C: Applied Statistics* 69, 4 (Aug. 2020), 931–951. doi:10.1111/rssc.12429

[4] J.B. Neethling and Holly Kennedy. 2018. *Nutrient Reduction Study: Potential Nutrient Reduction by Treatment Optimization, Sidestream Treatment, Treatment Upgrades, and Other Means.* Technical Report. Bay Area Clean Water Agencies. https://bacwa.org/wp-content/uploads/2018/06/BACWA_Final_Nutrient_Reduction_Report.pdf

[5] Leiv Rieger, Imre Takács, Kris Villez, Hansruedi Siegrist, Paul Lessard, Peter A. Vanrolleghem, and Yves Comeau. 2010. Data Reconciliation for Wastewater Treatment Plant Simulation Studies—Planning for High-Quality Data and Typical Sources of Errors. *Water Environment Research* 82, 5 (May 2010), 426–433. doi:10.2175/106143009X12529484815511