Smart Management of Urban Drainage Networks

Integrating Ontologies, Graph Theory, and Data Mining

Claude Pasquier — I3S Laboratory — Sophia Antipolis

GAMIN Workshop – June 4, 2026

Parts 1 & 2 (Ph.D. Research)
Batoul Haydar
Co-supervised by:
N. Chahinian & C. Pasquier

Part 3 (Pattern Mining)
Claude Pasquier
In collaboration with:
C. Perreira & A. Tettamanzi
B. Haydar & N. Chahinian

HSM, Univ. Montpellier — I3S, Univ. Côte d'Azur

The Sewer Network Data Challenge

- **1. Data Objectives (The Ideal)**
  - Reliable hydraulic modeling and a robust digital twin to manage urban pressure and infrastructure integrity.
- **2. The Data Reality (Current Limitations)**
  - **Heterogeneous:** Conflicting formats (SQL, Shapefiles) and inconsistent naming conventions.
  - **Incomplete:** Up to 50 % of critical attributes (diameter, material) are missing.
  - **Inconsistent:** Topological errors (reversed pipes, disconnected branches) hidden in databases.
- **3. Operational Consequences**
  - **Ineffective Simulation:** Misleading results for flood prevention.
  - **Maintenance Gaps:** Inability to identify water loss and leakage.
  - **Unresolved Risks:** Inability to detect design flaws at scale (>50 000 nodes).

Research Pipeline: Three Interconnected Sequential Steps

1. Data Integration

Establishing an OBDA system to unify semantic data views and verify initial database coherence.

2. Topology Completion

Applying graph theory to reconstruct topology and restore hydraulic flow logic.

3. Pattern Mining

Executing sequence mining on reliable data to discover operational rules and structural motifs.

Shared Motivation

Sewer network data is **incomplete, heterogeneous, and error-prone** (25–50 % of records may be unusable). Each step produces the reliable foundation required by the next.

Part 1 — OBDA Framework

Proposed solution

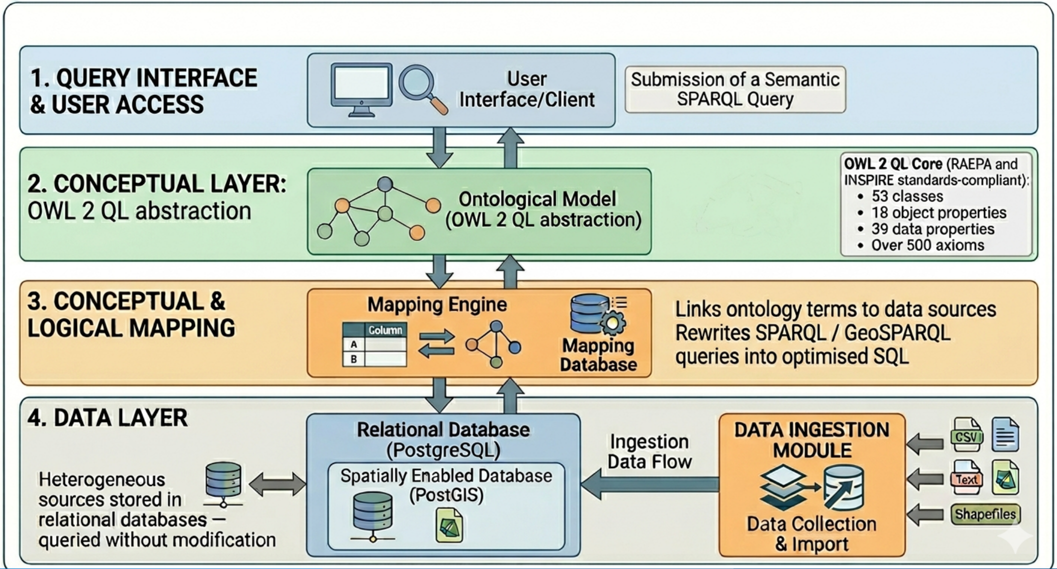
- Create an abstraction layer over existing data using OBDA.
- **Unified vocabulary:** Standardize classes (e.g., *Pipe*, *Manhole*, *Outfall*) and properties (*hasSlope*, *hasDiameter*).
- **Logic reasoning:** Define axioms for compliance (e.g., "a pipe must connect two nodes").
- **Mapping layer:** Translate data rows to meaningful entities.

Immediate benefits

a unified semantic view across heterogeneous sources.

a reasoning engine capable of detecting logical violations.

Part 1 — Application to the data sources of Montpellier metropole



Part 1 — Automated Quality Control

Topological Integrity:
- Isolated nodes/pipes
- Redundant geometries.
- Missing connections.

Hydraulic & Regulatory Logic:
- Flow mode inconsistency (e.g., Gravity vs. Forced).
- Regulatory diameter violations (< 200 mm).
- Elevation gradient vs. digitized flow direction.
- Network segregation (Wastewater → Stormwater).

Operational Consistency:
- Pumping stations connected only to gravity pipes.

Example SPARQL Query (Q5)

```
SELECT ?node
WHERE {
        ?node a :PumpingStation .
        MINUS {
                ?pipe :hasInitialNode ?
                      node ;
                :hasCirculationMode ?
                      forced .
                ?forced a :ForceMain .
        }
}
```

"Find pumps connected only to gravity pipes."

Part 1 — Case Study: Montpellier Métropole

Database Evolution & Query Results

	v1 (2023)	v2 (2024)
Pipes	51,013	51,368
Nodes	54,955	55,431
Pumping Stations	291	289
Query Result		
Q1: Isolated nodes	1,788	1,087
Q2: Inverse slope	884	896
Q3: Diam. $< 200\,\text{mm}$	12,203	12,161
Q4: WW $\rightarrow$ SW pipes	2	2
Q5: Pump-Gravity Conflict	19	19

Key Findings

Dual Correction Path: Issues resolved either through physical works (Q3) or data cleaning (Q1).

Regression Detection: New errors emerged in v2 (e.g., *inverse slopes*), likely due to manual input errors during updates.

Supporting Publication: Haydar, B., Pasquier, C., Straccia, U., & Chahinian, N. (2026). *An Ontology-based data Access Framework for Enhanced Sewer Network Management*. Proceedings of the 15th International Conference on Hydroinformatics (HIC 2026).

Part 2 — Topology Completion

Part 2 — Modeling Networks as Directed Graphs

- **Theory:** A sewer network is modeled as a Directed Acyclic Graph (DAG).
- **Main constraint:** Every node must have a downstream path to an outlet (reachability).
- **Frequent topological errors:**
  - **Inverted edges:** Pipes digitized against the flow direction.
  - **Gaps:** Missing segments creating orphan subnetworks.

Proposed correction strategy

Two successive steps: **Flow adjustment** and **Edge addition**

Part 2 — Network Correction Strategy

Pre-processing: Decomposes the graph into Weakly Connected Components.

Step 1: Flow Adjustment (edge reversal)

Objective: Ensure all nodes within a sub-network converge toward their specific *sink*.

Principle:

- Selects a target sink (real outlet or node with most predecessors)
- Reverse edges along the shortest path toward the sink

Step 2: Edge Addition (Connectivity)

Objective: Restore global connectivity to outlets of the entire system.

Principle:

- Connect the *sink* of an isolated sub-network to an "entrance" node of an already connected network.
- **Constraints:** no-loops, no gravity-to-forced transition, road geometry.
- a cost function combining **length**, **angular consistency**, and **node degree**.

Part 2 — Results

Montpellier Métropole case study

	Flow adj.	Edge add.
Unreachable nodes	19,465	13988
Initial sink nodes	152	60
Reversed or added edges	135	46
Resolved sink nodes	92	46
Unresolved sink nodes	60	14
New reachable nodes	**5,477**	**13,869**

Validation (Degradation)

Test Scenario	Jaccard/IoU
Flow adj. 1% rev.	0.98
Flow adj. 20% rev.	0.91
Edge add. 1% rem.	0.62
Edge add. 10% rem.	0.36
Combined 1%+1%	0.92
Combined 10%+10%	0.36

Interpretation

The drastic reduction of non-reachable nodes confirms substantial topology recovery. This methodology successfully restored connectivity even under high artificial noise.

Supporting Publication: Haydar, B, Chahinian, N, and Pasquier, C. (2026). *Reconstructing Sewer Network Topology Using Graph Theory*. Water, 2026, vol. 18, no 2, p. 222.

Part 3 — Pattern Mining

Part 3 — Contiguous Targeted Sequence Mining

Motivation

Sewer networks are modeled as **directed DAGs** mirroring hydraulic flow logic.

Pipes are represented as nodes, and their connections as directed edges.

Network attributes (e.g., material, diameter, slope) are discretized and attached to nodes as itemsets.

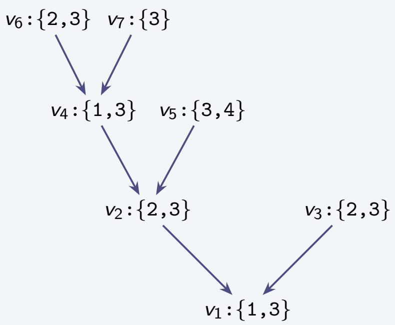
Objective: Uncover latent structural regularities within the topology to identify recurring construction patterns and detect infrastructure anomalies.

Key Challenges

Combining itemset and sequence mining within a DAG structure creates an **exponential search space**.

Standard support measures **violate monotonicity** because segments are topologically shared.

Attributed tree — example



Query $Q = \langle\{2,3\},\{1\}\rangle$ occurs at:
$v_2 \rightarrow v_1$, $v_3 \rightarrow v_1$, $v_6 \rightarrow v_4$ $\Rightarrow$ **support = 2**

Part 3 — The Pattern Explosion Problem

The Bottleneck

Traditional mining algorithms mines **all** frequent sequences.

In large datasets, this leads to a **combinatorial explosion** and the **output** of a **vast number** of irrelevant or uninteresting patterns.

The Strategic Shift

Instead of asking *"What is frequent?"*, our approach focuses the search:

"Given a critical **node** *or* **sequence of nodes** *(target Q), what are the recurring structural contexts surrounding it?"*

Part 3 — Novelty 1: Directional Contiguity Extensions

The algorithm introduces a **three-way extension mechanism** to maintain strict physical adjacency in tree structures:

i-extension

Refines the current node's attributes without moving in the tree.

s-extension

Moves **downstream** to the successor node.

p-extension (key innovation)

Moves **upstream** to the predecessor node.

Traditional sequence mining is unidirectional (left-to-right). Our algorithm treats the target Q as an anchor and grows the pattern in **both directions** along tree branches simultaneously.

Query 1 — Basic pumping sequence

Target Q: [{28, 42}] (PS + forced pipe) **Support:** 275

Key items: 28 (forced), 42 (PS), 27 (gravitational), 24 (50–400 m), 23 (<50 m)

Significant pattern	Support	% vs. Q
[{24, 28, 42}] (mid-length forced PS)	170	62 %
[{28, 42}, {27}] (PS + forced pipe → grav.)	189	68 %
[{28, 42}, {28}] (PS + forced pipe → forced)	60	22 %
[{23, 27}, {24, 28, 42}] (short grav. → PS + mid forced)	135	49 %
[{24, 28, 42}, {23, 27}] (PS + mid forced → short grav.)	108	39 %

- 62 % of those forced pipes are mid-length (50–400 m)
- 68 % of stations serve a single forced pipe followed by a gravitational one.
- 22 % of stations serve more than one forced pipe.
- Stations serving mid-length forced pipes (170) are bounded by short (<50 m) gravitational segments: 79 % upstream and 64 % downstream.

Query 2 — Station serving multiple forced pipes

Target Q: [{28, 42}, {28}] (PS + 2 forced pipes) **Support:** 60

Key items: 28 (forced), 42 (PS), 27 (gravitational), 24 (50–400 m), 23 (<50 m)

Significant pattern	Support	% vs. Q
[{23, 28, 42}, {28}] (PS + small forced pipe → forced)	29	48 %
[{28, 42}, {28}, . . . , {28}] (Pumping Station serving **9** forced pipes)	20	33 %

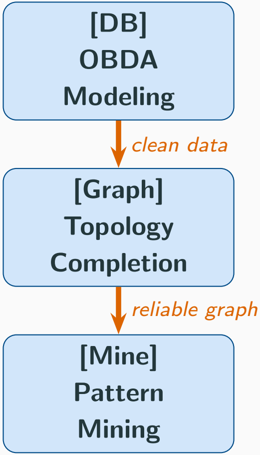
The “small pipe” shift

- **Short vs. Mid-length:** Multi-pipe stations favor short segments (<50 m) in 50 % of cases, reversing the global trend (62 % mid-length).
- **Booster Zones:** One-third of these cases serve 9 forced pipes, identifying industrial/booster clusters.

Summary & Outlook

What we have built

1. **OBDA framework** — unified semantic access, automatic inconsistency detection, and version tracking.
2. **Topology completion** — flow adjustment and road-guided edge addition; $>99\,\%$ of nodes reach outlets after correction.
3. **Pattern miner** — contiguous, targeted, closed sequence mining in attributed trees; monotonic support; no redundancy.



Thank you!

Questions?

Claude Pasquier claude.pasquier@univ-cotedazur.fr

I3S Laboratory, CNRS, Sophia Antipolis, France