Circular water specification

Version 1,0

# Introduction

The circular water specification is designed to outline the creation process of a dataspace that allows for the sharing and consumption of standardized data produced from different sources. The data should follow a set of semantic and technical specifications. The semantic specifications outline a set of agreed semantics, terms, shapes, metadata, and attributes. While the technical specifications include, but are not limited to hosting, mapping, and linking.

# Versioning system

* Uses a two-number format (x,y):
  + Left number (x): Semantic version (major updates, structural changes).
  + Right number (y): Technical version (minor updates, implementation details).
* Current Version: 1.0:
  + Establishes ontology and semantic standards.
  + Defines technical requirements like hosting and interoperability.

# Version

The current version is 1,0. This version lays the foundation for the ontology and semantic standards. It also lays the foundations for the technical standards,

# The difference between a specification and a standard

A specification is simply a technical document that describes how a thing that has been built works. This includes documenting interfaces, ontologies, etc., to support others in reusing them, and getting valuable early feedback.

At a later stage, a small subset of those specifications could be put on a standards track. The difference is that a specification represents a technical description by one party by itself, whereas a standard is a specification that reflects the consensus of multiple parties.

# Key elements in standard

* Metadata structure: define how datasets should be described and cataloged. This includes standardized metadata models (e.g., DCAT, RDF Connect), and dataset descriptions (e.g., sources, update frequency, spatial-temporal coverage). (e.g., OGC, RDF Connect). Metadata will be divided into a tiered system according to importance.
* Data discoverability: Ensures data can be found, accessed, and utilized efficiently. Includes cataloging mechanisms (e.g., DCAT-based metadata registries) and indexing and searchability across datasets.
* Data relationships: specify the linking mechanism between data resources, especially with real-time sensor data (e.g., Tree).
* Data Cataloging: Organizes, indexes, and manages metadata about datasets to improve data discoverability, accessibility, and usability (e.g., DCAT).
* Mapping rules: specify mapping rules for harmonizing heterogeneous datasets to standardized ontologies. (e.g., RML).
* Hosting rules: determine how the data will be distributed, hosted and how it will be accessible through the host (e.g., Solid).
* Compliance requirements: define mandatory and optional elements for interoperability and consistency.
* Data linking: ensuring data is web accessible and allowing data exchange between data silos sustainably and cost-effectively (e.g., LDES, YARRRML).
* Access control policy: Specify access and usage control policy for each data source and for each user
* Data formats: ensuring data conforms to a single linked data compliant format (RDF).

# Use Cases

Define the practical scenarios where the specification will be applied. This is determined by the actual use cases obtained from stakeholders and data producers throughout the project’s life cycle.

Currently, use case scenarios are not concretely defined, they will be defined later.

# Ontology design

* Semantic interoperability: ensure different systems interpret and use data consistently.
* Data integration: integrate heterogeneous data sources by defining shared vocabularies and relationships.
* Querying and Reasoning: Enable advanced queries and inferencing to derive new insights.
* Alignment with existing standards: align with existing standards (e.g., RDF, DCAT, TREE)
* Alignment with existing ontologies: align with existing ontologies (e.g., SSN SOSA, OGC WaterML2, OWL-Time)
* Concepts and Entities:
  + Core resources (e.g., rainwater, wastewater).
  + Core processes (e.g., treatment, energy generation).
  + Core actors (e.g., households, industries).
  + Core quality and compliance standards (e.g., parameters, values).
* Relationships: how entities are related and hierarchical structures where applicable.
* Metadata integration: integrate metadata to describe data (3-tier level)
* Data properties: specify data attributes (e.g., quality metrics, locations)

# validation and compliance

shapes ensure that the data used to create the knowledge graph conforms to rules and constraints to avoid having outlier data (e.g., SHACL).

# Security and Privacy

Usage access control policy ensures secure access to data by tracking and controlling access privileges of users querying data.

# Query and Reasoning

Querying would allow for a direct interface with data to derive new insights.

# Real-time data

Support integration with sensors and real-time monitoring systems by defining APIs or protocols for ingesting real-time data.

# Role of stakeholders

Stakeholders have the role of producing data that conforms to the agreed semantic standards of the specification. They also have the role of providing direct access to the data utilizing technical-compliant specifications.

# Governance framework

The governance framework defines how the specification will be maintained and updated over time. The governance framework is as follows:

* Advisory board: the waterframe advisory board and UGent IDlab are experts in water management and semantic web technologies.
* Review and approval: the steering committee proposes future updates and approve changes into new versions.
* Long-term viability: the specification remains adaptable to new technologies

# Performance metrics

Performance will be evaluated using the following metrics:

* Data Quality: Percentage of datasets meeting semantic and technical standards.
* Interoperability: Number of successful data exchanges across systems.
* System Efficiency: Query response times and data integration speeds.

# Testing and Validation Framework

To ensure data is compliant with pre-defined semantic standards, the following framework will be used:

Data validation: utilize automated tools (e.g., SHACL) to check for data compliance.

Ontology validation: check for correct implementation using RDF validation tools.

Interoperability: simulate exchanges between different systems to verify compatibility.

# Global Alignment

The specification aims to align with international standards and adopt existing principles. For the semantic specification, the goal is to align with ISO and OGC. The technical specification aims to align with linked data event streams (LDES), Tree, Dcat, and others.

# User Documentation and Support

To facilitate the adoption, this document serves as a thorough explanation of the circular water specification along with a GitHub repository that contains the following:

* Circular water ontology
* SHACL shapes
* RML mapping rules
* LDES data fetching pipeline

# Sustainability Aspects

The circular water specification promotes sustainability in data management and water resource utilization.