D4.5 DWC Water Value Chains Ontology

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Deliverable N°4.5	DWC Water Value Chains Ontology
Related Work Package	WP4
Deliverable lead	21-SINTEF
Author(s)	Audun Vennesland, SINTEF
Contact for queries	audun.vennesland@sintef.no
Grant Agreement Number	n° 820954
Instrument	HORIZON 2020
Start date of the project	01 June 2019
Duration of the project	42 months
Website	www.digital-water.city
Abstract	D4.5 describes the development of a reference ontology for the DWC project. Based on requirements elicited from cases in DWC, the report describes the conceptualisation, implementation, evaluation, and documentation of the reference ontology. The work performed has shown that most requirements can be satisfied by existing ontologies describing water-related concepts, notably ontologies in the ETSI SAREF suite. However, a few extensions in the form of new ontology classes, object properties and data properties are needed to fully address the DWC requirements. These extensions will be forwarded to the appropriate ontology maintenance body.

Dissemination level of the document

X	PU	Public
	PP	Restricted to other programme participants
	RE	Restricted to a group specified by the consortium
	CO	Confidential, only for members of the consortium

Versioning and contribution history

Version*	Date	Modified by	Modification reasons
D1	30.06.2020	Audun Vennesland	First draft of structure and contents
D2	02.11.2021	Audun Vennesland	Draft version of full report sent to reviewers
R1	23.11.2021	Audun Vennesland	Version addressing comments from internal review
R2	24.11.2021	Nico Caradot	Review by the coordinator
S	29.11.2021	Audun Vennesland	Final version for submission

* The version convention of the deliverables is described in the Project Management Handbook (D7.1). *D* for draft, *R* for draft following internal review, *S* for submitted to the EC and *V* for approved by the EC.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Table of content

List of figures

List of tables

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Glossary

Table 1. Glossary

Term	Description
API	Application Programming Interface.
DWC	digital-water.city
ETSI	European Telecommunications Standardisation Institute
FIWARE	FIWARE is an initiative that defines reusable open-source components and standardised specifications for context data management.
NGSI	Next Generation Service Interface. A protocol to manage context information.
NGSI-LD	An NGSI information model that represents context information as entities that have properties and relationships to other entities. Every entity and relationship is given a unique IRI reference as identifier, making the corresponding data exportable as Linked data datasets.
OWL	Web Ontology Language. A Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
RDF	Describes data by defining relationships between data entities expressed using URIs (Uniform Resource Identifiers) and related via triples in the form (subject-predicate-object).
RDFS	A language for representing simple RDF vocabularies on the Web.
SAREF	The Smart Applications REFerence ontology, an ontology that specifies reccurring core concepts, relationships and constraining axioms for the smart application domain.
SAREF4AGRI	A domain ontology for the agricultural domain that extends SAREF
SAREF4SYST	An extension of SAREF that defines systems, connections between systems and connection points at which systems be be connect.
SAREF4WATR	A domain ontology for the water domain that extends SAREF.
WWTP	Wastewater Treatment Plant

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Executive summary

This report describes the development of a reference ontology in the DWC project. From explicit requirements collected in other tasks in WP4, notably task 4.1 (results reported in D4.1 [1]) and sub-task 4.3.1 (results reported in D4.4 [2]), this report describes the conceptualisation, the implementation, the evaluation, and documentation of the reference ontology. From an analysis of relevant existing ontologies related to the water domain, we have identified a set of ontologies that largely capture the requirements elicited. On this basis it is better to re-use concepts from existing ontologies than developing a new DWC ontology from scratch. Based on a detailed mapping between a conceptualisation of the elicited requirements and concepts defined in these existing ontologies, we have identified some extensions in the form of new ontology classes, object properties and data properties that have been implemented in the reference ontology. These extensions will be forwarded to the SMARTM2M technical committee in ETSI[1], which is responsible for maintaining the ontologies that we propose to further extend. The resulting reference ontology has been both automatically and manually evaluated. It is furthermore documented using illustrations of how the reference ontology addresses a list of competency questions that suggest details that the ontology must encompass. The report concludes with a section that describes how the ontology can be used in practical application.

[1] https://www.etsi.org/committee/smartm2m

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

1. Introduction

1.1. Summary of DWC and objectives of work package 4 and subtask 4.3.2

DWC aims at creating digital solutions to link water management in the physical world to the digital spheres such as sensor networks, real-time monitoring, machine learning etc.

24 partners from 10 countries work together in the case study cities Berlin, Copenhagen, Milan, Paris and Sofia and support the utilities and municipalities in improving water quality, return on investment and public information about water-related issues.

The core of Work Package (WP) 4 is to ensure that the digital solutions are designed and developed in a way in which data and information are exchanged with the utility systems in a cyber-secure and interoperable way. This includes a risk analysis and proposition of risk reduction measures to protect data and systems from unauthorized access as well as the description of semantic models and interoperability design guidelines. The tasks focus on the digital solutions and their impacts on the existing infrastructures, but not on general cyber-physical risks.

Subtask 4.3.2 has created a reference ontology for DWC. Reusing knowledge structures from existing ontologies, notably the SAREF suite[2] of ontologies, data requirements collected from a subset of the cases in DWC are mapped to these existing ontologies to consequently define the reference ontology. Where the mapping reveals that requirements from DWC cannot be fulfilled by existing ontologies, we propose extensions to existing ontologies. The reference ontology can be used to (1) provide a semantic description of water management-related entities to be stored in a FIWARE[3] context broker, (2) support semantic interoperability in the data exchange involving digital solutions developed in DWC, and (3) utilise formal semantics of water management data in downstream data analysis tasks.

1.2. Structure of this document

The structure of the remainder of this report is as follows. In section 2 we introduce ontologies and describe the set of ontologies we reuse in the DWC Reference Ontology. In section 3 we describe the approach used for developing the reference ontology. Section 4 is mainly devoted to a definition of competency questions that the ontology must be able to address given the requirements elicited in DWC. Section 5 describes a conceptualisation of relevant ontology classes and relations that should be included in the reference ontology. Section 6 describes how the reference ontology is evaluated and documented using illustrations showing how the competency questions defined in section 3 are addressed by the resulting reference ontology. Section 8 suggests how the reference ontology can be applied in practical settings. Section 9 presents conclusions from this work.

[2] https://saref.etsi.org/ : last accessed 22.11.2021

[3] https://www.fiware.org/developers/catalogue/ : last accessed 22.11.2021

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

2. Preliminaries

2.1. Introduction to ontologies

This section provides a minimal and practical description of some key aspects related to ontologies to prepare for the remainder of this report. For a more detailed explanation of ontologies and their application, the reader is referred to e.g., the book Handbook on Ontologies [3] and the W3C OWL 2 Web Ontology Language Primer[4].

An ontology is a formal definition of the concepts, properties and interrelationships of the entities that exist in some domain of discourse. It provides a shared vocabulary that can be used to describe the domain, classifying and categorising the elements contained within it. Typically, an ontology is formalised using the Web Ontology Language (OWL). OWL is a part of the W3C suite of Semantic Web standards[5], which includes among others Resource Description Format (RDF)[6], a framework for representing web data using subject-predicate-object triples, and the Resource Description Format Schema (RDFS)[7] which provides a data-modelling vocabulary for RDF data. While both OWL and RDFS offer a vocabulary for describing RDF data, OWL allows for greater expressibility than RDFS.

In an ontology, classes represent sets of individuals (also called instances or objects) with similar characteristics and are organised in a specialisation hierarchy. This hierarchy is also called a *subsumption hierarchy* in that a parent class subsumes its children classes, i.e. any individual that is a member of the subsumed (more specific) class is also a member of its subsuming (more general) class.

Figure 1 shows the extract of the SAREF4WATR ontology where its entities describe a water flow measurement performed by a water meter located in a particular spatial position. The example includes classes (ovals), individuals (orange rectangles), literal values (white rectangles), object properties (associations between individuals) and data properties (associations from individuals to literal values). The subsumption hierarchy of the ontology is made up of subClassOf relations from subclasses to more general parent classes. Individuals are instances that are members of a class, and they inherit the semantics of this class as well as those of parent classes as described above. The root of an ontology is called Thing and all individuals are members of this class. Object properties, which are named directed associations, relate one individual to another individual, where the source individual is a member of a domain class and the target is a member of the range class.

An example of this is illustrated by how the individual `WaterMeterInd_1`, which is an instance and a member of the `WaterMeter` class (domain), is related to the individual `WaterFlowPropertyInd_FlowRate_1` (range) which signifies that the property being measured is water flow. Data properties associate individuals to literal values (such as how the individual `MeasurementInd_1` which provides the actual water flow measurement is assigned both a timestamp via the `hasTimestamp` data property and a value via the `hasValue` data property).

---

4 https://www.w3.org/TR/owl2-primer/: last accessed 22.11.2021

5 https://www.w3.org/standards/semanticweb/ : last accessed 22.11.2021

6 https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ : last accessed 22.11.2021

7 https://www.w3.org/TR/rdf-schema/ : last accessed 22.11.2021

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

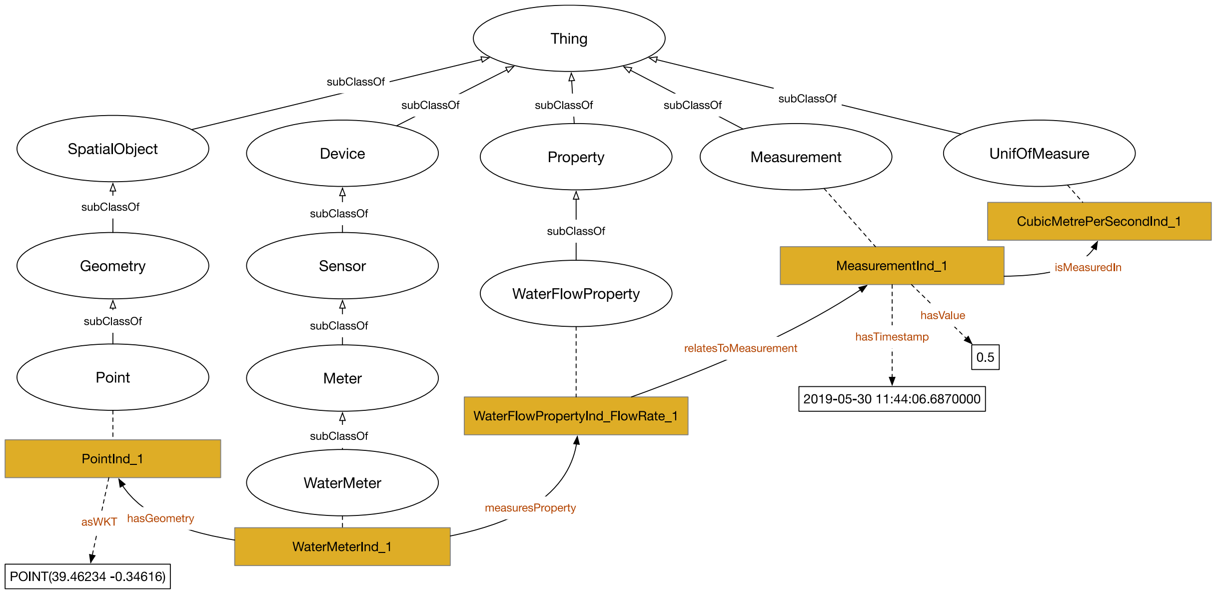

Figure 1. Example of an ontology. The example includes classes (ovals), individuals (orange rectangles), literal values (white rectangles), object properties (associations between individuals) and data properties (associations from individuals to literal values)

Ontologies come in different levels of generality. Guarino [4] suggests the following classification:

- *Top-level ontologies* describe general concepts such as time, space, events, actions, etc. These concepts are domain-independent and can be used for most application purposes. Examples of top-level ontologies are DOLCE[8] and PROTON[9].
- *Domain ontologies* and *task ontologies* describe concepts related to a specific domain or a particular task or activity respectively. These types of ontologies specialise the concepts introduced in the top-level ontologies.
- *Application ontologies* describe concepts that depend both on a particular domain and a particular task and may correspond to the roles played by domain concepts while performing a specific activity.

In an industrial setting however, ontology development and usage take a more pragmatic approach and SAREF is considered an upper ontology to which more detailed domain ontologies can further detail and be linked through [5]. SAREF itself includes relations to other more generic ontologies (e.g., the GeoSPARQL ontology[10] for describing spatial concepts, the W3C Time ontology[11] for describing temporal aspects), and Dolce Ultralite[12] via its use of the W3C Semantic Sensor Network (SSN) ontology. How ontologies include constructs from other ontologies is illustrated in Figure 2, where an extract of SAREF4WATR entities is shown along with relations to entities defined in SAREF, GeoSPARQL, SAREF4SYST [6] and W3CTime.

---

[8] http://www.loa.istc.cnr.it/dolce/overview.html
[9] http://www.ontotext.com/proton/protontop
[10] https://opengeospatial.github.io/ogc-geosparql/geosparql11/index.html
[11] https://www.w3.org/TR/owl-time/
[12] http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Domain ontologies relevant in this work are primarily SAREF4WATR [7], and SAREF4AGRI [8], but also SAREF4SYST is included. These ontologies will be described further in section 2.2.

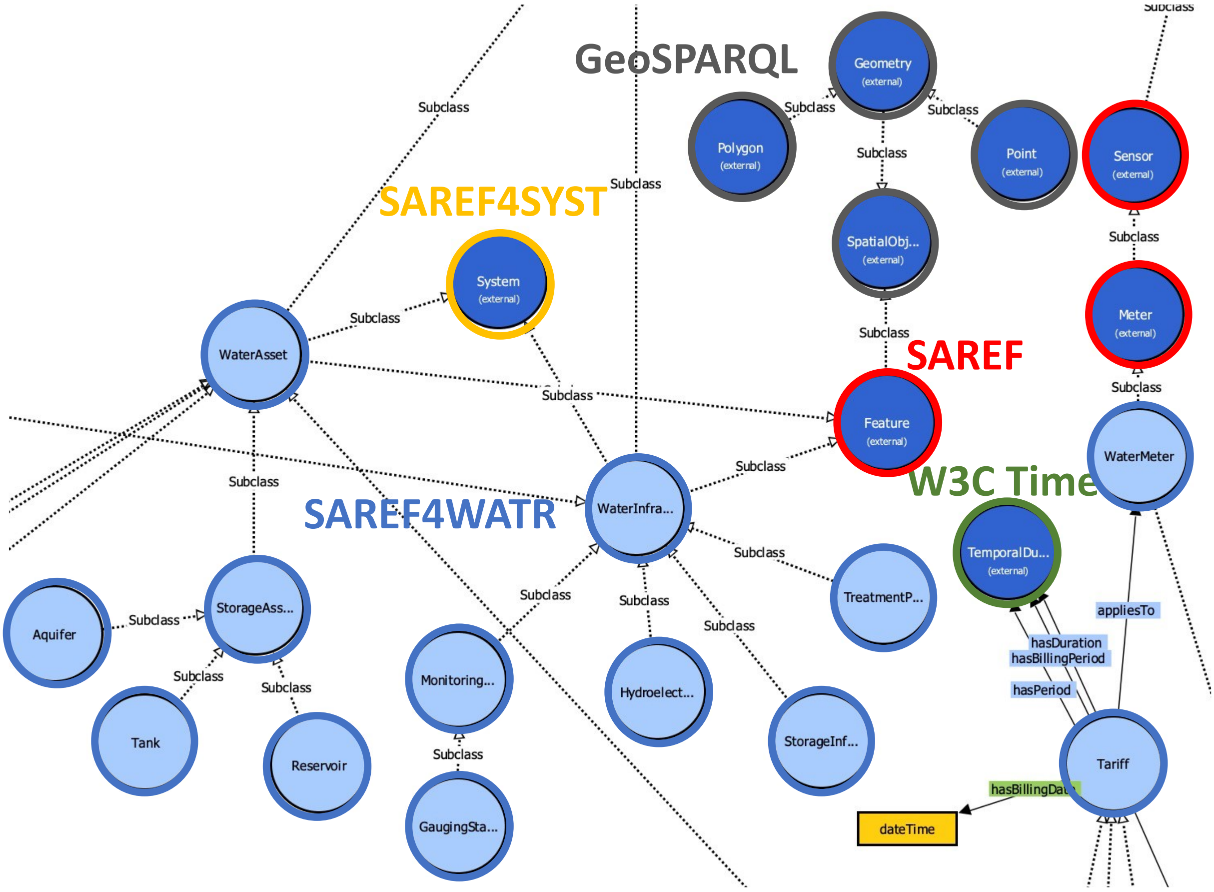

Figure 2. Illustration showing the interplay between top-level and domain ontologies. SAREF imports the GeoSPARQL and W3CTime ontologies in order to specify spatial and temporal concepts. SAREF4WATR refines water-related feature classes which extend the Feature class defined in SAREF, and re-uses design patterns and associated classes from SAREF4SYST to specify systems (in a broad sense) and relationships between systems.

2.2. ETSI SAREF Suite of ontologies

This section will include a description of the SAREF ontologies that are considered relevant in DWC. The overview provided in Table 2 describes some statistics for these ontologies.

Table 2. Ontology statistics for SAREF Core, SAREF4WATR and SAREF4AGRI

SAREF Ontology Module	Number of classes	Number of object properties	Number of data properties
SAREF Core	81	35	5
SAREF4WATR	72	40	22
SAREF4AGRI	55	49	13

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

SAREF4SYST	4	9	0

2.2.1. SAREF Core

The Smart Applications REFerence ontology (SAREF) is intended to enable interoperability between solutions from different providers and among various activity sectors in the Internet of Things (IoT). It specifies recurring core concepts in the Smart Applications domains, the main relationships between these concepts as well as axioms that constrain usage of these concepts and relationships. SAREF is built on the fundamental principles of *reuse and alignment* of concepts and relationships that are defined in existing assets; *modularity* to allow separation and recombination of different parts of the ontology depending on specific needs; *extensibility* to allow further growth of the ontology; and *maintainability* to facilitate the process of identifying and correcting defects, accommodate new requirements, and cope with changes in (parts of) SAREF [5].

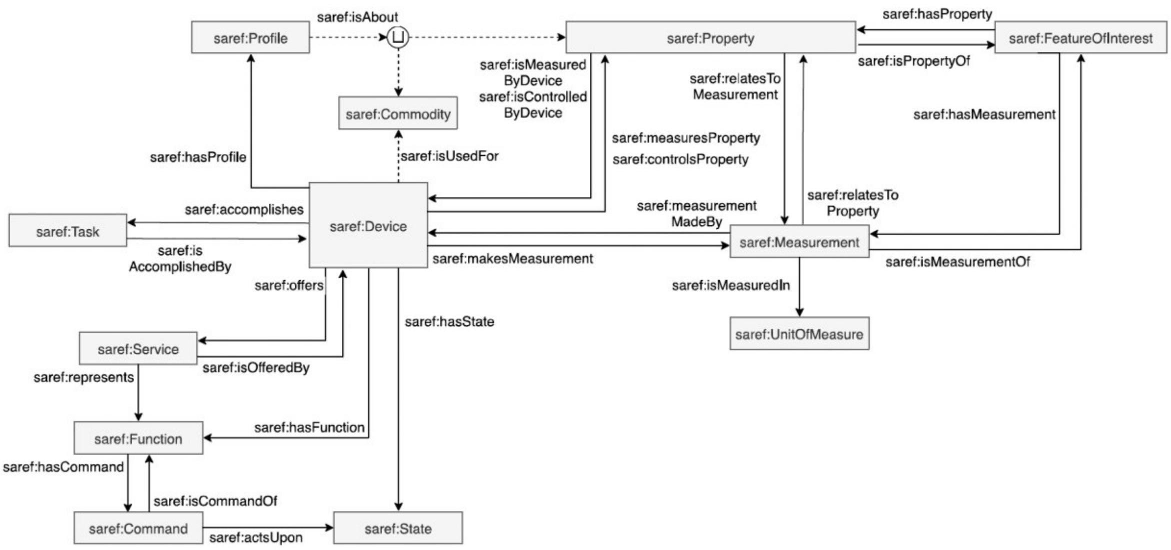

Figure 3. Overview of SAREF [5]

SAREF specifies the fundamental and generic concepts for “smart devices” such as sensors, meters, and appliances, and their reported measurements.

2.2.2. SAREF4WATR

SAREF4WATR extends the SAREF ontology for the water domain and aims to create a common core of general concepts for water data oriented to the IoT field [7]. As indicated by the blue classes in Figure 4, SAREF4WATR extends SAREF by adding water-related concepts, such as `WaterInfrastructure` being a subclass of `FeatureOfInterest` and `WaterDevice` added as a subclass of SAREF´s `Device`, while it inherits the fundamental concepts and the underlying semantics expressed in SAREF.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

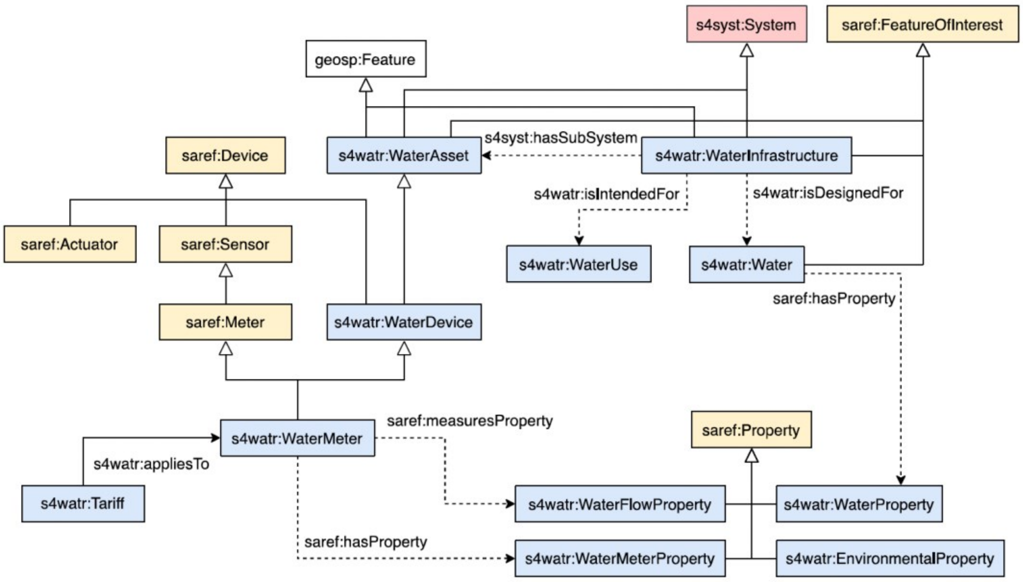

Figure 4. Overview of SAREF4WATR [7] which shows some of the core classes of the ontology, relationships between them, and relations to other ontologies (yellow, pink and white).

2.2.3. SAREF4AGRI

SAREF4AGRI is an extension of SAREF for the agriculture and food domain [8]. It is built around two examples from the agricultural domain, namely livestock farming and smart irrigation, of which the latter is relevant in the Milan case of DWC (DS5.1 and DS5.2).

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

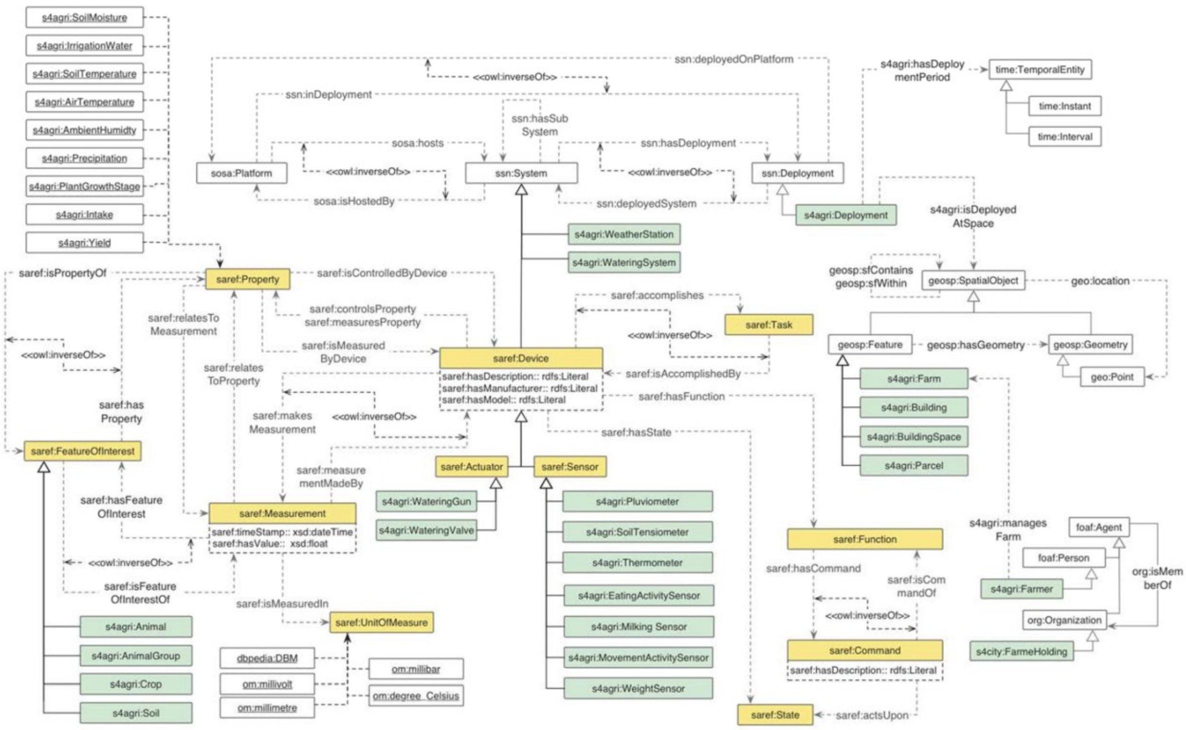

Figure 5. Overview of SAREF4AGRI [8] which shows some of the core classes of the ontology, relationships between them, and relations to other ontologies (yellow and white).

2.2.4. SAREF4SYST

SAREF4SYST [6] is a generic ontology module that defines an ontology pattern which can be instantiated for different domains. SAREF4SYST defines Systems, Connections between systems, and Connection Points at which systems may be connected. By the classes and properties of SAREF4SYST one may model how different systems relate (e.g., connecting sub-systems to main systems) and how different systems and their functions affect each other (e.g., the triggering of one function related to system A can affect a property of a system B).

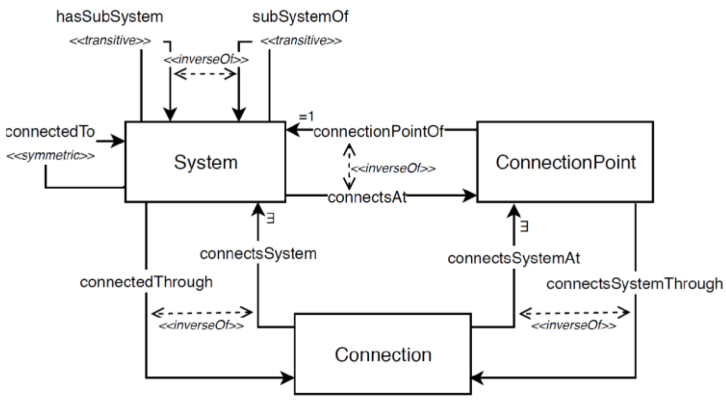

Figure 6. Overview of SAREF4SYST [6]

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

2.3. Other relevant ontologies

2.3.1. AGROVOC

AGROVOC[13] is a linked open dataset for agricultural data maintained by the Food and Agriculture Organization of the United Nations (FAO). It provides a controlled vocabulary of agricultural concepts, terms, definitions, and relationships. Currently, it consists of over 39.000 concepts and 800.000 terms in up to 40 different languages.

2.3.2. GeoSPARQL

GeoSPARQL defines a vocabulary and a query language (extending SPARQL) for representing and retrieving geospatial data. It is a standard maintained by the Open Geospatial Consortium (OGC). The GeoSPARQL ontology is available from: http://www.opengis.net/ont/geosparql

2.3.3. W3C Time

The W3C Time (OWL-Time) ontology[14] is an ontology for describing the temporal properties of resources in the world or described in Web pages. The ontology provides a vocabulary for expressing facts about topological (ordering) relations among instants and intervals, together with information about durations, and about temporal position including date-time information.

2.3.4. W3C SSN Ontology

The Semantic Sensor Network (SSN) ontology is an ontology that describes sensors and their observations, the involved procedures, the features of interest observed, observed properties and actuators[15].

[13] https://www.fao.org/agrovoc/

[14] http://www.w3.org/2006/time#

[15] https://www.w3.org/TR/vocab-ssn/

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

3. Approach

A methodology for the development of the DWC semantic model[16] is tailored from an analysis of existing ontology engineering methodologies. These existing engineering methodologies are: Methontology [9], NeOn [10], 101 [11], SAMOD [12], UPON [13], and UPON Lite [14].

Our analysis has shown that some of these methodologies are too complex and resource-demanding for our task (e.g., Methontology and NeOn), especially since the core of the task is to identify existing water-related ontologies that fulfil the demands in DWC rather than constructing a new ontology. At the same time, other methodologies do not provide enough guidance to achieve the objectives (e.g., SAMOD and UPON Lite). Therefore, we have extracted what we consider the most relevant ontology engineering activities considering the scope of the semantic model development in DWC and tailored this into the methodology illustrated in Figure 7, further described in Table 3:

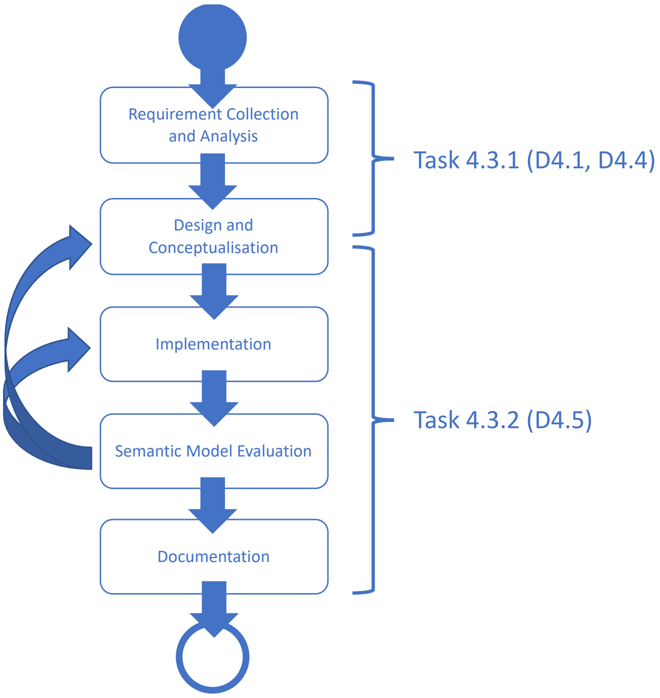

Figure 7. Methodology for developing the semantic model in DWC

The first activity of task 4.3.1 was to capture a baseline and initial requirements of the digital solutions and their interaction with water utilities involved in DWC. This is reported in deliverable D4.1 [1]. Continuing from D4.1, the goal of D4.4 [2] was to elicit and formulate more detailed design requirements by focusing on a subset of cases in DWC, namely the developments taking place in Paris and Milan. This encompasses the two first stages of the methodology: Requirements Collection and Analysis and Design and Conceptualisation. Task 4.3.2 and this deliverable (D4.5) describes how the design requirements are further elaborated as formalised concepts, implemented in a semantic model, and how this semantic model is evaluated and documented, covering partly the Design and

[16] In the document we use the terms 'semantic model' and 'ontology' interchangeably.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Conceptualisation stage as well as the three latter stages of the development approach; Implementation, Semantic Model Evaluation and Documentation.

Table 3. Description of stages in approach.

Stage	Description
Requirements Collection and Analysis	This stage begins by defining the domain of interest, the scope and purpose of the ontology to be developed. Furthermore, elicited stakeholder requirements will be described as use case scenarios that will lead to the identification of relevant concepts and properties in the ontology. Relevant non-ontological resources should be considered for re-use. Finally, a set of competency questions that the resulting ontology should address is formulated.
Design and Conceptualisation	At this stage, the "raw" set of concepts and properties resulting from the previous stage is elaborated. Classes and properties are defined with a textual description, the subsumption hierarchy of classes is established, any semantic rules (e.g., property restrictions) are defined, and properties are distinguished as object- or data properties (literals).
Implementation	In this stage the ontology is formalised in an ontological language (OWL). Relevant existing ontological resources should be re-used whenever possible and common ontology design patterns should be applied. The result from this stage is a complete ontology encompassing all concepts, properties, and semantic rules defined in the preceding stages.
Semantic Model Evaluation	In the evaluation the complete ontology is evaluated manually and automatically. The manual evaluation should involve an examination of the ontology both by the ontology engineers and domain experts to ensure sufficient coverage, relevance, and correctness. The automatic evaluation involves consistency checks using reasoning services and manually verifying the suitability of the ontology using the competency questions formulated in the Requirements Collection and Analysis stage.
Documentation	Finally, once the ontology has passed the evaluation, the ontology is made available in a persistent URL along with accompanying documentation. The documentation should in addition to appropriate definitions of the ontology constructs (classes and properties) include relevant details from each stage of the development process.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

4. Requirements Collection and Analysis

For a list of use case scenarios and detailed requirements we refer to deliverable D4.4 [2]. This deliverable continues that work by defining a set of competency questions that the DWC Reference Ontology needs to cover. These competency questions are formulated on the basis of the use case scenarios and requirements defined in D4.4 and further interaction with stakeholders from the Paris and Milan cases. These competency questions are answered in Chapter 7 as part of the evaluation and documentation of the reference ontology.

4.1. Competency Questions (CQs) from Paris and Milan

CQ1: How do you describe a river flow measurement?

River flow measurements are used as input to the statistical model developed in the Paris case.

CQ2: How do you describe a rainfall observation?

Discharge information is used as input to the statistical model developed in the Paris case.

CQ3: How do you describe an *e.coli* measurement performed by a specific sensor?

The Fluidion Alert sensor which measures *e.coli* levels in water will be implemented in both Paris and Milan cases.

CQ4: How do you describe a water quality prediction?

The output of the statistical model in the Paris case is a water quality prediction.

CQ5: How do you describe the amount of discharge water released from a wastewater treatment plant?

Discharge information is used as input to the statistical model developed in the Paris case.

CQ6: How do you describe the topological position of a particular sensor at the wastewater treatment plant?

Within a wastewater treatment plan there are two different lines into which the plant is divided. Each line has multiple stages the perform different types of treatment on the wastewater and that contain different types of sensors, meters, and samplers. Knowing the topological location of these devices is important.

CQ7: How do you describe a measurement of the amount of ammonium present in water?

There are multiple ammonium (NH4) sensors within the Peschiera Borromeo treatment plant in Milan. This CQ also serves to describe, evaluate, and document a more generic representation of sensors, their measurements, and other characteristics within a wastewater treatment plant.

CQ8: How do you describe a particular farm along with its crops and soil characteristics?

Water from the wastewater treatment plant in Milan will be used to meet the need for irrigation of crops and soil of farms. Thus, the ontology should cover concepts and properties allowing to express characteristics of farms along with their soil and crop characteristics.

CQ9: How do you describe the seeding date of a crop?

The seeding date of crops will be included into the evaluation of water stress in a given field to identify the need for irrigation water.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

CQ10: How do you describe a sampling device at a particular topological position within a wastewater treatment plant?

There are several automatic samplers at the wastewater treatment plant in Milan. Such automatic samplers may contain different types of sensors, and samples could be taken at different time intervals.

CQ11: How do you describe the soil moisture as measured by the sensors deployed on the field?

The water content of the soil as measured by sensors in the field is used to evaluate water stress.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

5. Design and Conceptualisation

This section describes the entities derived from (1) the requirements collection and analysis and (2) the design and conceptualisation stages described in deliverable D4.4 [2]. Each entity is presented in Table 4 below, along with a definition and relations that should be associated with the entity. Suggested natural language definitions are indicated by italic font. Relations include object properties and data properties. Object properties are described using the range class of the object property and an object property name underneath. Data properties are described using the data property name with a basic datatype (e.g., decimal) underneath. Along with the properties we also indicate the cardinality of them in parenthesis. There are some design choices being made, such as that the different sensors themselves are not typed as classes, the focus is on the measurements made by a sensor.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Table 4. Conceptualisation of requirements from Paris and Milan

Data element	Parent Class	Definition	Relations and cardinality when relevant	Case
Device	Thing	*An apparatus (hardware + software + firmware) intended to accomplish a particular task (sensing the environment, actuating, etc.) (source: smartdatamodels.org).* Device is the parent class of Sensor, Meter and SamplingDevice.	• Location ○ hasPoint (0..1) ○ hasPolygon (0..1) ○ hasTopologicalLocation (0..1) • Measurement ○ hasMeasurement (0..*)	Milan, Paris
Sensor	Device	*A device that detects and responds to events or changes in the physical environment such as light, motion, or temperature changes (source: smartdatamodels.org).* The following sensors are relevant: pH sensor, UV Intensity sensor, TOC sensor, Nitrate sensor, Dissolved Oxygen sensor, NH4 sensor, Cl sensor, REDOX sensor, UV sensor (Organic Load), Phosphate sensor, TSS sensor, Conductivity sensor, Water Temperature sensor (thermometer), Air Temperature sensor (thermometer), Wind Speed sensor (anemometer), Wind Direction sensor (Wind wane), Air Humidity sensor (hygrometer), Solar Radiation (pyranometer).	(Inherits relations from Device)	Milan, Paris
Meter	Device	*A device built to accurately detect and display a quantity in a form readable by a human being (source: smartdatamodels.org).* Relevant meters are flow meter and energy meter.	(Inherits relations from Device)	Milan, Paris

SamplingDevice	Device	*A device that obtains a sample of matter that will be subject to further analysis.* Relevant sampling devices are automatic samplers for wastewater treatment and porous cup loggers.	• (Inherits relations from Device) + • Sensor ○ hasIncludedSensor (0..*)	Milan
Measurement	Thing	*The measured value of some property.* The following measurements are relevant and could be proposed as subclasses to further detail the type of Measurement made: Meteorological Measurement, Water Quality Measurement, Water Observation Measurement, Agricultural Measurement	• Time ○ hasTimestamp (1..1) • Device ○ measuredByDevice (0..*) • Property ○ measuresProperty (0..*) • Unit of measurement ○ hasUOM (0..*) • value (1..1) ○ decimal	Milan, Paris
UnitOfMeasurement	Thing	*Describes different measurement units relevant for a measurement.*	• measurementType (1..1) ○ string	
Farm	Thing	*A farm is an agricultural structure which may have a farm size, farm typology, field shape, and field size (source: https://agrovoc.fao.org/browse/agrovoc/en/page/c_2804).*	• Location ○ hasPoint (0..1) ○ hasPolygon (0..1) ○ hasTopologicalLocation (0..1) • FarmTypology ○ hasFarmTopology (0..1) • farmSize (0..1) ○ decimal • fieldSize (0..1) ○ decimal	Milan

FarmTypology	Thing	*A stratification of farms that is homogenous according to specific criteria relevant to policy such as environmental performance and farm management practices (source: https://agrovoc.fao.org/browse/agrovoc/en/page/c_a4f92a63*		
Crop	Thing	*A crop is any cultivated plant, fungus, or alga that is harvested for food, clothing, livestock, fodder, biofuel, medicine, or other uses. (source:* https://agrovoc.fao.org/browse/agrovoc/en/page/c_1972*)* Should be related to seeding date and transplanting date.	• CropType ○ hasCropType (0..1) • Time ○ hasSeedingDate (0..1) ○ hasTransplantingDate (0..1)	Milan
Soil	Thing	Upper layer of the earth in which plants grow (source: https://agrovoc.fao.org/browse/agrovoc/en/page/c_7156). Should be related to different types of soil (SoilType).	• SoilType ○ hasSoilType (0..1)	Milan
Location	Thing	*Description of location in terms of different geographic data structures (e.g., point, polygon, address) and topology.*		Milan, Paris
Time	Thing	*Description of time in terms of instants or intervals.*		Milan, Paris
Irrigation system	Thing	*Describes a concrete irrigation system entity and relates to the IrrigationSystemType below for a classification of different irrigation systems.*	• IrrigationSystemType ○ hasIrrigationSystemType (0..1)	Milan

Water Quality Prediction	Thing	*WaterQualityPrediction is an independent class which describes a prediction of water quality based on a statistical model.* It should have relations to an enumerated list of prediction results (WaterQualityPredictionResult), a value related to a percentile, a timestamp, and the location for which the prediction is relevant.	• WaterQualityPredictionResult ○ hasWaterQualityPredictionResult (1..1) • Measurement ○ hasMeasurement (1..1) • Time ○ hasTimestamp (1..1) • Location ○ hasPoint (0..1) ○ hasPolygon (0..1) ○ hasTopologicalLocation (0..1) • hasPercentile ○ decimal	Paris
WaterQualityPredictionResult	Thing	*An enumerated list of prediction results that should encompass: "Excellent", "Good", "Sufficient", "Poor".*		Paris
SoilType	Thing	*An enumerated list for defining different types of soil. A classification of soil types is defined by AgroVoc: https://agrovoc.fao.org/browse/agrovoc/en/page/c_7204*		Milan
CropType	Thing	*An enumerated list for defining different types of crops.*		Milan
IrrigationSystemType	Thing	*Describes a type of an irrigation system, such as Channel, pivot, reel, or drip.*		Milan

WastewaterTreatmentPlant	Thing	*A wastewater treatment plant performs the necessary processes to remove contaminants from wastewater and convert it into an effluent that can be returned to the water cycle. Once returned to the water cycle, the effluent creates an acceptable impact on the environment or is reused for various purposes (called water reclamation) (adapted from Wikipedia)*	• Location ○ hasPoint (0..1) ○ hasPolygon (0..1) ○ hasTopologicalLocation (0..1) • Sensor ○ includesSensor (0..*) • Meter ○ includesMeter (0..*) • SamplingDevice ○ includesSamplingDevice (0..*)	Milan

6. Implementation

As suggested earlier, there exists several ontologies relevant to semantically describe the concepts of interest in DWC. This section will describe how these existing ontologies can cover and semantically annotate and structure the concepts defined in the conceptualisation described in the previous section as to represent them in an appropriate data storage and retrieve them for various types of analysis and other usage. Further, this section will describe concepts that are not covered by existing ontologies and propose how they could be added for future versions.

The reference ontology is made available online at:

https://raw.githubusercontent.com/SINTEF-SE/DWC/main/DWC_ReferenceOntology_v01.owl

The OWL file could be opened in an ontology editor such as Protégé[17].

6.1. Mapping identified concepts with existing ontologies

Relating the identified classes presented in Table 4 with the ontologies described in section 2.2, Tables 5, 6 and 7 below presents a mapping where the concepts highlighted in blue can be proposed as extensions to existing ontologies:

Table 5. Mapping of classes

Class	Relevant classes in existing ontologies
Device	SAREF Core Device: https://saref.etsi.org/core/Device
Sensor	SAREF Core Sensor: https://saref.etsi.org/core/Sensor
Meter	SAREF Core Meter: https://saref.etsi.org/core/Meter
SamplingDevice	New concept in DWC Reference Ontology which is a subclass of Device.
Measurement	SAREF Core Measurement: https://saref.etsi.org/core/Measurement
UnitOfMeasurement	SAREF Core UnitOfMeasure: https://saref.etsi.org/core/UnitOfMeasure
Farm	SAREF4AGRI Farm: https://saref.etsi.org/saref4agri/Farm
FarmTypology	New class in DWC Reference Ontology which is a subclass of Thing.
LandMeasurement	New class in DWC Reference Ontology which is a subclass of Property.
Crop	SAREF4AGRI Crop: https://saref.etsi.org/saref4agri/Crop
Soil	SAREF4AGRI Soil: https://saref.etsi.org/saref4agri/Soil
Location	To describe locations, we will reuse the SAREF4AGRI and SAREF4WATR conventions and their use of GeoSPARQL SpatialObject and its subclasses Feature and Geometry.

[17] https://protege.stanford.edu/

Time	For temporal definitions we will reuse SAREF´s hasTimestamp for defining instant time, and SAREF4WATR hasPhenomenonTime for describing time intervals.
IrrigationSystem	SAREF4AGRI: WateringSystem
Water Quality Prediction	New concept in DWC Reference Ontology. Should be a subclass of another new class Prediction that specifies more generic properties.
WaterQualityPredictionResult	New concept in DWC Reference Ontology. Should be a subclass of another new class PredictionResult that specifies more generic properties.
SoilType	New concept in DWC Reference Ontology. Different ontology design patterns can be used for such a concept, but in this version, we propose to have SoilType as a class, whose member instances represent different categories from the AgroVoc classification described at: https://agrovoc.fao.org/browse/agrovoc/en/page/c_7204
CropType	New concept in DWC Reference Ontology. Different ontology design patterns can be used for such a concept, but in this version, we propose to have CropType as a class, whose member instances represent different categories from the AgroVoc classification described at: https://agrovoc.fao.org/browse/agrovoc/en/page/c_1972
IrrigationSystemType	New concept proposed by DWC. IrrigationSystemType should be a class whose initial members (individuals) could be: "Channel", "Pivot", "Reel", "Drip".
WasteWaterTreatmentPlant	SAREF4WATR: TreatmentPlant

Table 6. Mapping of object properties

Object Properties		Relevant object properties in existing ontologies
hasPoint	Device → Location	geo:Point
hasPolygon	Device → Location	geo:Polygon
hasTopologicalLocation	Device → Location	Different structure implemented in SAREF. A Device can be associated with a Building (e.g., a wastewater treatment plant) or a BuildingSpace via the ssn:hasDeployment object property. A geographical location (e.g., polygon) and/or a name can be associated with Building/BuildingSpace.
hasMeasurement	Device → Measurement	s:makesMeasurement, s:hasMeasurement
hasIncludedSensor	SamplingDevice → Sensor	Different structure implemented in SAREF. Sensors can be

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

		associated with the new class SamplingDevice using the ssn:hasSubSystem object property.
hasTimestamp	Measurement → Time	s:hasTimestamp
measuredByDevice	Measurement → Device	The inverse object property measurementMadeBy can relate a measurement to a device.
measuresProperty	Measurement → Property	s:relatesToProperty
hasUnitOfMeasurement	Measurement → UnitOfMeasurement	s:isMeasuredIn
hasFarmTypology	Farm → FarmTypology	New object property suggested by DWC
hasCropType	Farm → CropType	New object property suggested by DWC
hasSeedingDate	Crop → Time	s4a: hasPlantDate
hasTransplantingDate	Crop → Time	New object property suggested by DWC
hasSoilType	Soil → SoilType	New object property suggested by DWC
hasIrrigationSystemType	IrrigationSystem → IrrigationSystemType	New object property hasIrrigationSystemType suggested by DWC where the domain class is WateringSystem and the range is IrrigationSystemType
hasWaterQualityPredictionResult	WaterQualityPrediction → WaterQualityPredictionResult	New object property suggested by DWC

Table 7. Mapping of data properties

Data Properties		Relevant data properties in existing ontologies
value (decimal)	Measurement	s:hasValue
measurementType (string)	UnitOfMeasurement	Different structure implemented in SAREF ontologies. An individual typed as s:UnitOfMeasure holds the unit of measurement type (e.g. cubic metre) and is associated to an individual typed as s:Measurement via the s:isMeasuredIn object property.
farmSize (decimal)	Farm	Size is a relative term, but in this case it is interpreted as the size in terms of the metric system (square

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

		meters, hectare, or similar measures). This can be accomplished reusing the measurement structure already present in the SAREF ontologies. An individual typed using s4a:Farm is related to an individual typed as s:Measurement where the data property s:hasValue is used to specify the size. The object property s:isMeasuredIn can be related to an individual such as "Hectare" typed as s:UnitOfMeasure. Proposing a new subclass of Property called LandMeasurement which is related to an individual typed as s:Measurement via s:relatesToProperty.
fieldSize (decimal)	Farm	Same as farmSize above, but instead of an individual typed as s4a:Farm an individual typed as s4a:Parcel is used instead.
hasPercentile (decimal)	WaterQualityPrediction	New data property proposed by DWC.

6.2. Suggested extensions for the DWC Reference Ontology

As described in the previous section, many of the requirements from the DWC project can be covered by the ontologies SAREF, SARE4AGRI, SAREF4WATR and their imported ontologies (e.g., GeoSPARQL, W3CTime and W3C SSN). However, the concepts summarised in Table 8 are not covered and could be proposed as extensions taken in at future versions of the SAREF suite.

Table 8. Proposed extensions of concepts

Extension	Description
New classes	
Water Quality Prediction	WaterQualityPrediction is an independent class which describes a prediction of water quality based on a statistical model. It should have relations to an enumerated list of prediction results (Water Quality Prediction Result), a value related to a percentile, a timestamp, and the location for which the prediction is relevant.
Water Quality Prediction Results	An enumerated list of prediction results that should encompass: "Excellent", "Good", "Sufficient", "Poor".

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Sampling Device	A device that obtains a sample of matter that will be subject to further analysis. Relevant sampling devices are samplers for wastewater treatment and porous cup loggers.
IrrigationSystemType	Defines different types of irrigation systems, such as channel, drip, pivot, or reel.
FarmTypology	A class that allows for a typology or classification of farms. An initial classification can be: CropFarm, LiveStockFarm, MixedFarm according to the glossary in: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Farm_typology
LandMeasurement	A type of property that can be related to measurements of measurements of areas/land.
Soil Type	An enumerated list for defining different types of soil. A classification of soil types is defined by AgroVoc: https://agrovoc.fao.org/browse/agrovoc/en/page/c_7204
Crop Type	New concept in DWC Reference Ontology. Different ontology design patterns can be used for such a concept, but in this version, we propose to have CropType as a class, whose member instances represent different categories from the AgroVoc classification described at: https://agrovoc.fao.org/browse/agrovoc/en/page/c_1972.
New object properties	
hasFarmTypology	New object property that relates a farm with its typology.
hasCropType	New object property that relates a crop with its crop type.
hasTransplantingDate	New object property that states the transplanting date of a crop.
hasSoilType	New object property that relates a soil to a particular soil type.
hasIrrigationSystemType	New object property that relates an irrigation system (watering system) to a particular type of such.
hasWaterQualityPredictionResult	New object property that relates a water quality prediction with a water quality prediction result.
New data properties	
hasPercentile (decimal)	New data property that states the percentile for which a water quality prediction value is valid.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

7. Semantic Model Evaluation and Documentation

The evaluation described in this section is performed to ensure that the requirements elicited in the project are addressed by the ontology and that the ontology is consistent. With respect to the former, a manual evaluation is performed by modelling (the response to) the competency questions described in section 4. The consistency evaluation is performed by running a reasoning service on the ontology. The reasoner applied was Hermit, which is shipped with the Protégé ontology editor, and it ran without any errors or reported inconsistencies.

As part of the Documentation stage defined in section 3, this section includes illustrative examples depicting how the reference ontology and its constructs can define the relevant concepts addressing each competency question from section 4. The convention used in the illustrations is shown in Figure 8. An entity (e.g., a particular sensor) is illustrated using rectangles where the name of the entity is presented at the bottom with white background, while the class being used to type the entity is presented at the top of the rectangle, with grey background. The source ontology is presented as a prefix (before the ":") in the concept name. Whenever we propose extensions to the existing ontologies reused, we use "dwc" as the source ontology, to indicate that this in fact is a suggested extension based on a DWC requirement. The relationship between different entities is described by means of object properties. Object properties are defined by straight and directed lines in black, while inverse object properties are in dotted, grey lines. Data properties are indicated using all white rectangles and the text within the rectangle indicates the value.

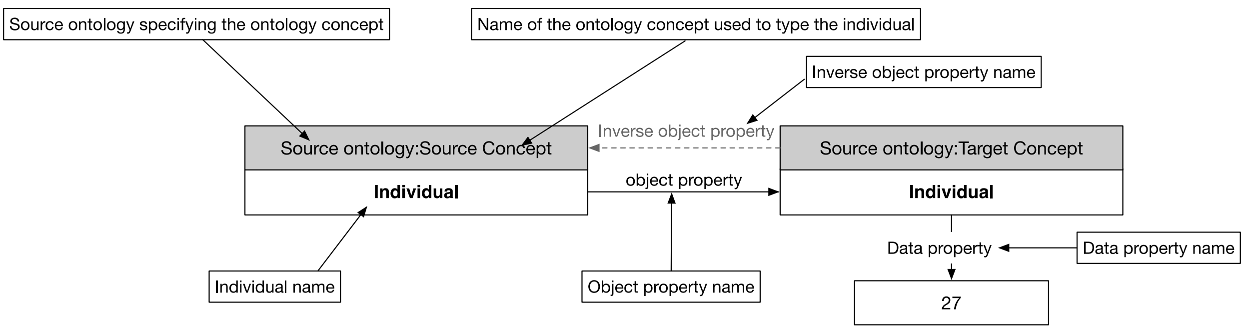

Figure 8. Explanation of notation used for illustrations

CQ1: How do you describe a river flow measurement?

Figure 9 illustrates how a river flow measurement can be described by the reference ontology. A water meter (`DWC_WaterMeter`), which has a given location expressed using the Point concept of GeoSPARQL (`geo:Point`) and latitude-longitude coordinates serialised using Well-Known Text (`geo:asWKT`), performs the measurement which is further described using the `DWC_RiverFlowMeasurement`. The water meter can (as all devices) be further described using data properties such as model, manufacturer, fabrication number, as indicated in the left part of the illustration. The measurement includes a value stating the amount of river flow (`hasValue`), a timestamp for the measurement (`hasTimestamp`), and a unit of measurement which in this case is cubic metre per second. The measurement measures the property `FlowRate`, which is of type

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

WaterFlowProperty. The makesMeasurement object property is stated to be an inverse property indicated by measurementMadeBy (in grey).

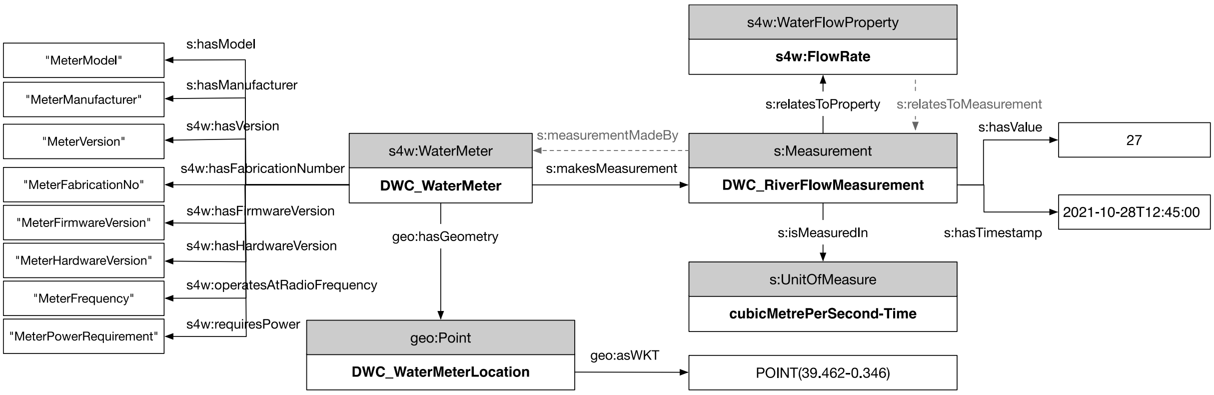

Figure 9. Riverflow measurement

CQ2: How do you describe a rainfall observation?

Figure 10 illustrates how a rainfall observation can be described by means of the reference ontology. A pluviometer (DWC_Pluviometer) is used to measure rainfall. As with the water meter in the previous section, the Pluviometer is a subclass of Device and inherits data properties allowing to express details such as model, manufacturer, and fabrication number, this will be omitted for brevity in the following sections. This pluviometer has a location which is expressed by the DWC_PluviometerLocation and coordinates described using the GeoSPARQL construct Point (geo:Point) and the serialization format asWKT described earlier. The measurement relates to the property Precipitation (typed using the Property class), and reports of 2 mm of rain measured at the indicated timestamp.

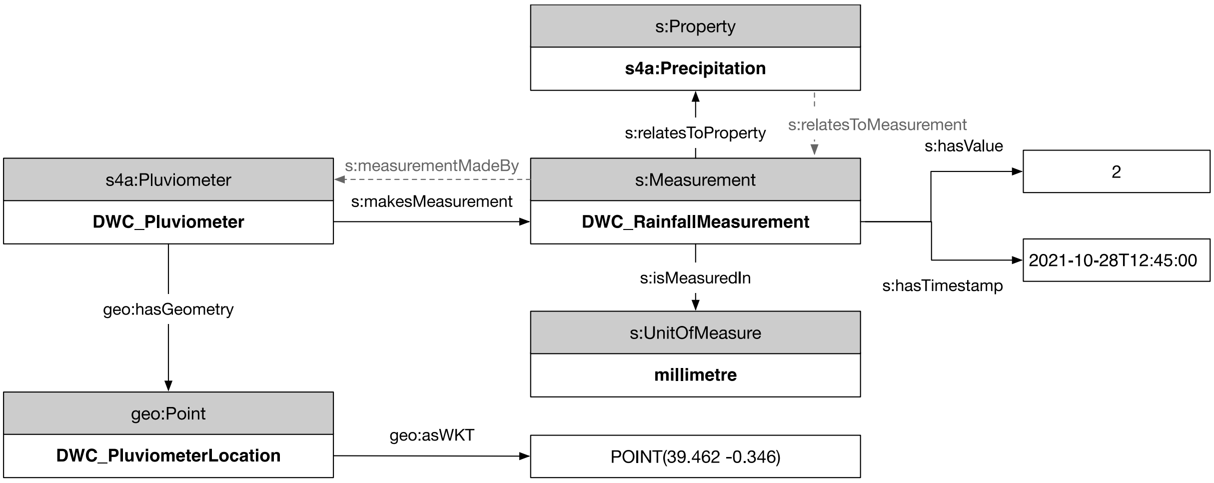

Figure 10. Rainfall observation

CQ3: How do you describe an e.coli measurement performed by a specific sensor?

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

How an e.coli measurement is described by means of the ontology is illustrated in Figure 11. The e.coli sensor (`DWC_eColiSensor`) is located according to coordinates 39.456 (latitude) and -0.353 (longitude). A measurement (`DWC_eColiMeasurement`) that reports the level of colony forming units of Escherichia Coli per millimetre is associated to the sensor. Escherichia Coli is typed as a `BacterialProperty`.

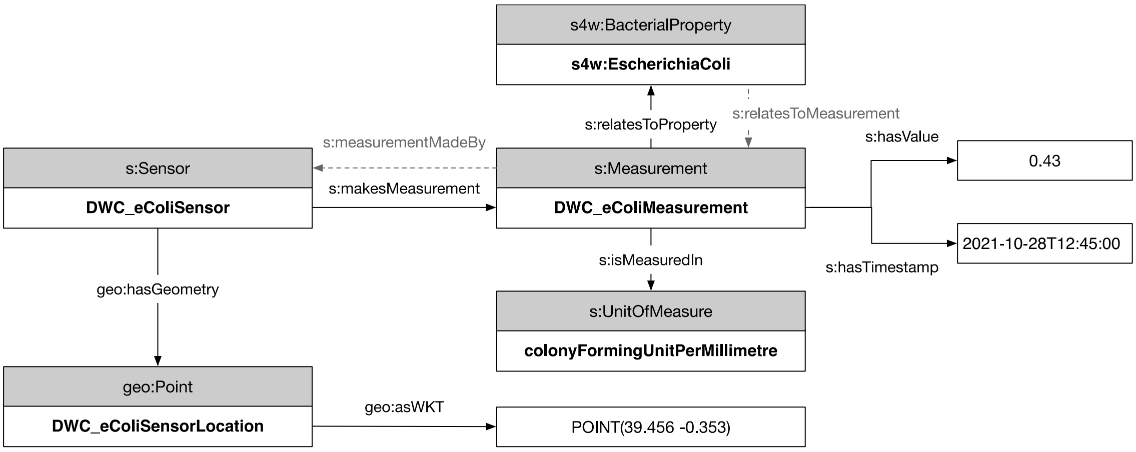

Figure 11. e.coli measurement

CQ4: How do you describe a water quality prediction?

Figure 12 shows how a water quality prediction for a part of a river (where the `River` class is a subclass of `FeatureOfInterest`) can be described by means of the reference ontology. This structure follows the convention suggested by SAREF: measurements that relate to certain properties of a feature of interest are associated with the source (e.g., sensor) performing the measurement. In this case the source is a Water Quality Prediction entity (`DWC_WaterQualityPrediction`) reporting a prediction on the basis of statistical modelling, and the related property is `BathingWaterQuality` (typed as `AcceptabilityProperty`). The literal prediction result is “Excellent” and is typed according to the suggested new class `WaterQualityPredictionResult`, which holds the following enumeration individuals: “Excellent”, “Good”, “Sufficient” and “Poor”. The prediction holds for the period 03.11.2021-05.11.2021, which is modelled using the `hasPhenomenon` time structure.

digital-water.city
digitalwater_eu
digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

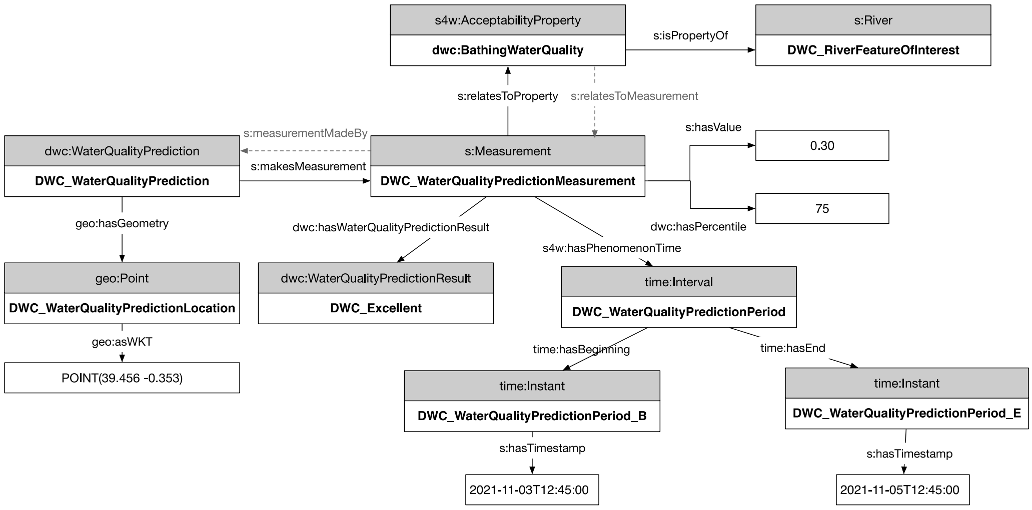

Figure 12. Water quality prediction

CQ5: How do you describe the amount of discharge water released from a wastewater treatment plant?

Figure 13 illustrates how a discharge from a wastewater treatment plant can be described by the reference ontology. This structure is very similar to how the river flow measurement described earlier, but in this case the related property is flow volume, not flow rate.

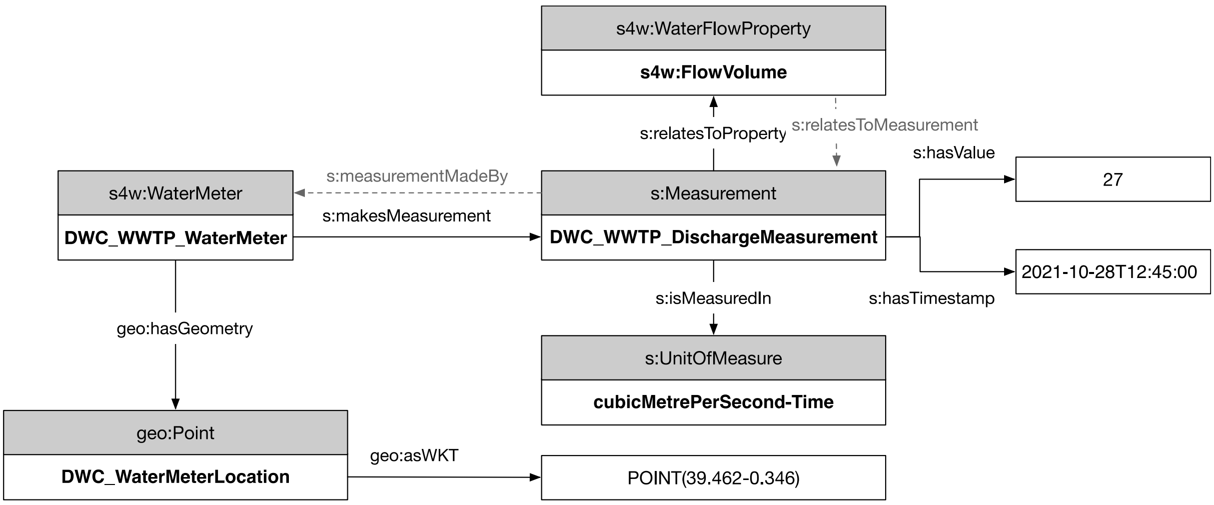

Figure 13. Discharge from wastewater treatment plant

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

CQ6: How do you describe the topological position of a particular sensor at the wastewater treatment plant?

As illustrated in Figure 14 a chloride sensor (`DWC_WWTP_ChlorideSensor`) performs a measurement of chloride content at a treatment plant. The sensor has coordinates described by means of the `Point` concept and its data property `asWKT` which serializes the sensor´s latitude and longitude coordinates. To describe the topological location of the sensor we employ the deployment constructs from the W3C SSN ontology and the building space concept and associated properties defined in the SAREF4AGRI ontology. The Deployment concept with its object properties `hasDeployment` and `isDeployedAt` links the sensor to a geographically defined building space, specified using a polygon. The building space, which in addition to the included geo structures can be defined by a name (e.g., using the data property `hasName` with value "Line1-Stage3") is then linked to the treatment plant (`DWC_WWTP_PR`) using GeoSPARQL´s `sfContains` object property. Here, the polygon of the building space is assumed to be a sub-part of the polygon defined for the treatment plant.

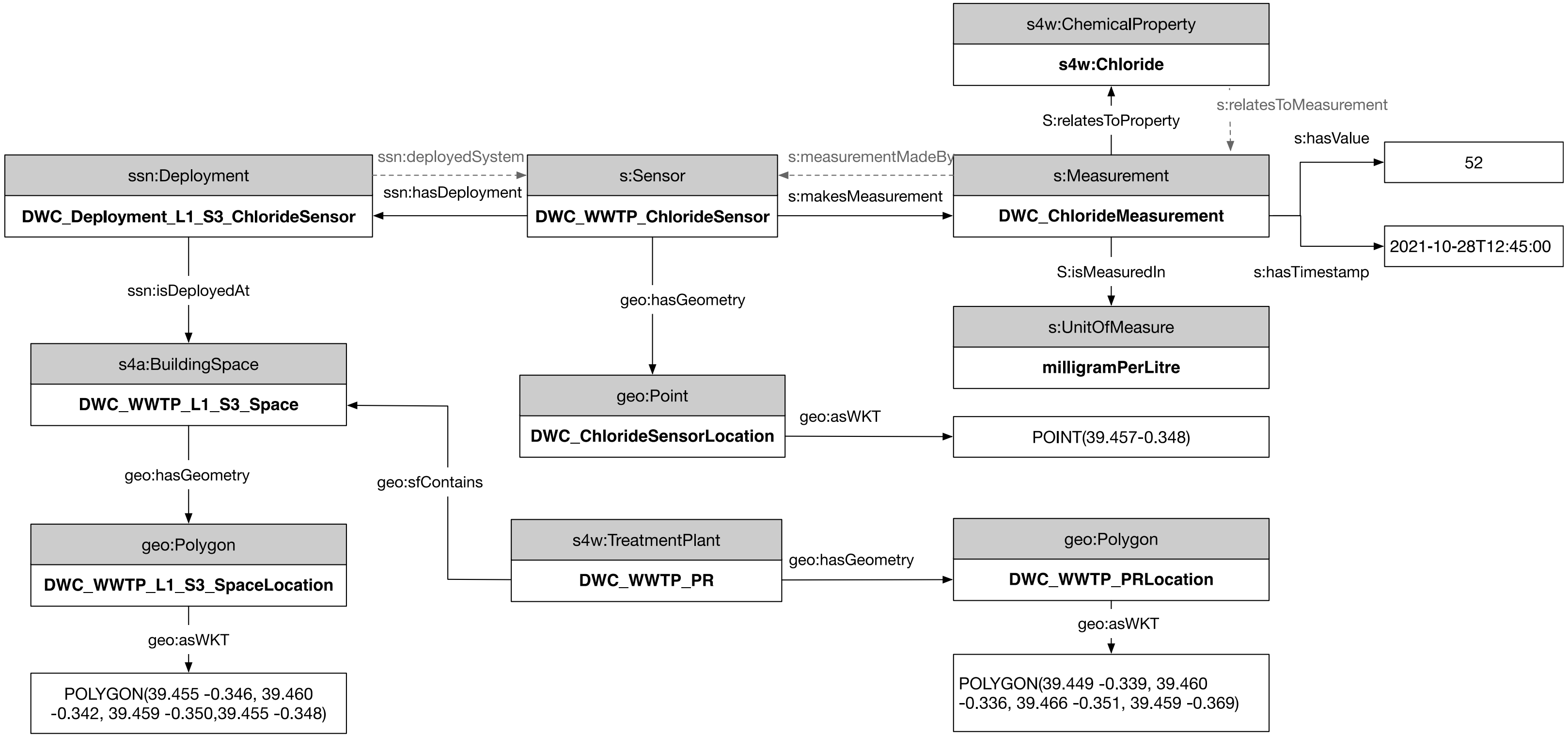

Figure 14. Topological position of a sensor at the treatment plant.

CQ7: How do you describe a measurement of the amount of ammonium present in water?

Figure 15 illustrates how an ammonium (NH4) sensor (`DWC_WWTP_NH4Sensor`) reports measurements using the reference ontology. This structure is very similar to the previous section, so we only focus on the differences. The measurement relates to the property `Ammonium` which is included in SAREF4WATR.

This property is of type `ChemicalProperty`. As in the previous example, the ammonium sensor is topologically located in a particular building space, which again is located within a wastewater treatment plant.

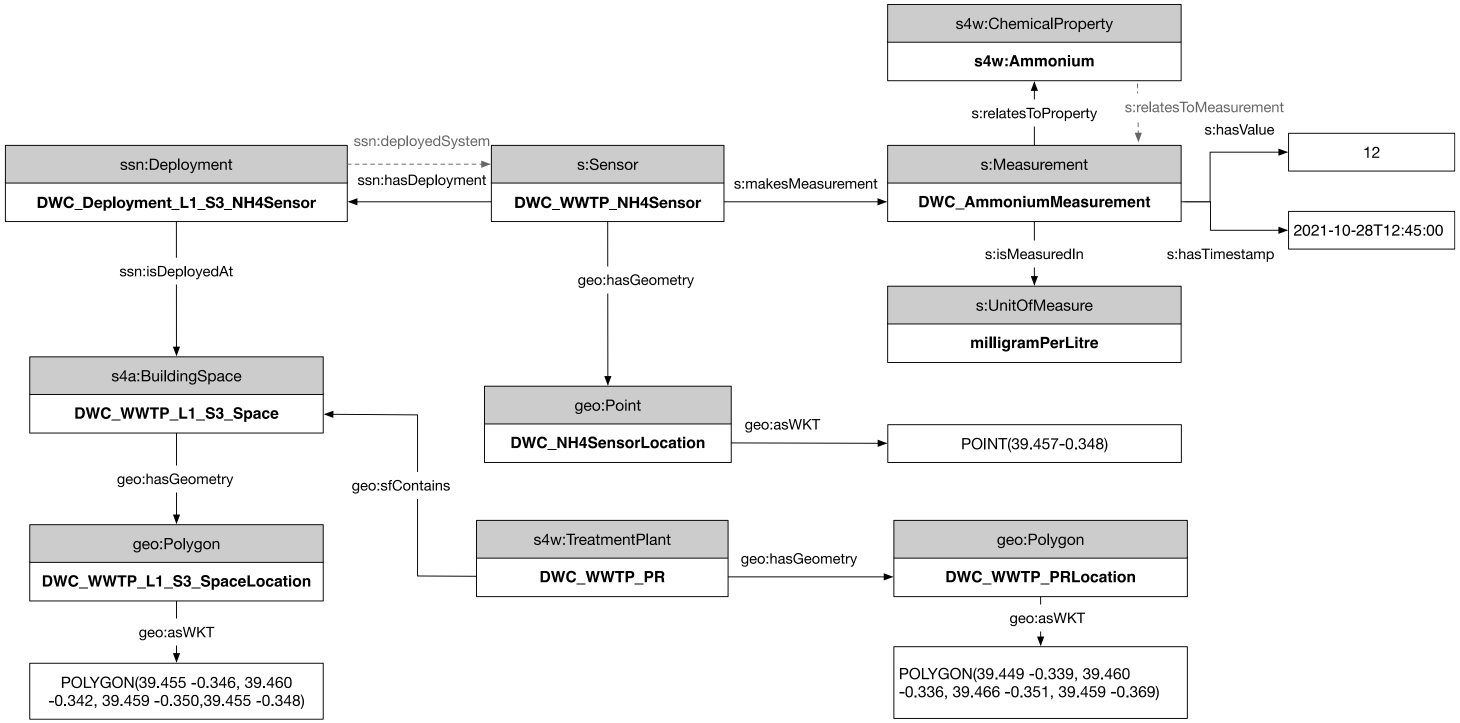

Figure 15. Measuring the amount of ammonium in water

CQ8: How do you describe a particular farm along with its crops and soil characteristics?

As Figure 16 illustrates, SAREF4AGRI includes concepts for describing farms, soils, and crops. The location of the farm `DWC_Farm` is defined by a polygon in this case, but can also be defined by the centroid of the farm using the `geo:Point` class and reference to latitude and longitude coordinates as described in earlier examples. Via the `sfContains` object property of GeoSPARQL, we can associate a parcel (typed using the `Parcel` class), which naturally is defined by a smaller polygon description than the farm. We associate a farm typology using the new object property `hasFarmTypology` that specifies this farm to be a crop farm. Further, we specify the area size of the farm via the `hasMeasurement` object property whose range is `Measurement`. The measurement relates to the property `LandMeasurement`. In this case hectare is used as the measurement unit, but other relevant units (e.g., square metres or square kilometres could also be applied. To the parcel we can associate one or more soils and one or more crops to the parcel in question. We can also specify an area size of the parcel using the same mechanisms we used to specify the area size of the farm. We can further detail the type of soil(s) using member individuals of the proposed `SoilType` class and the type of crop(s) using the proposed `CropType` class.

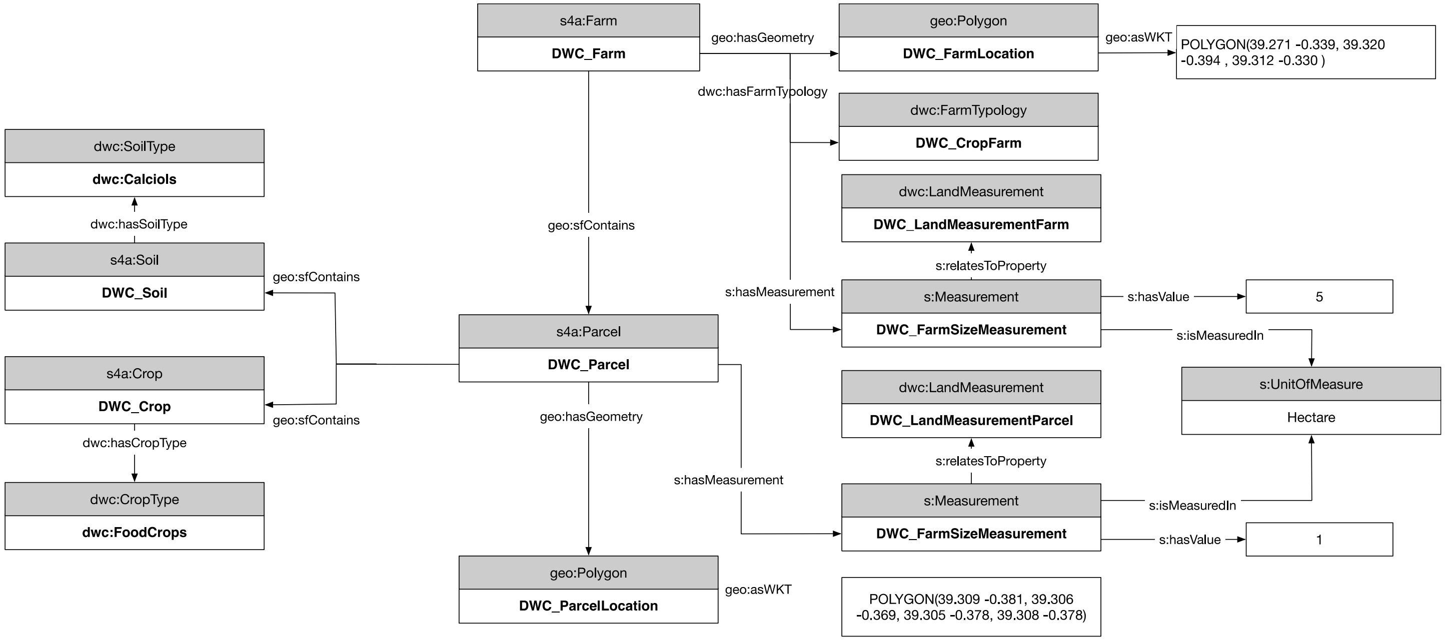

Figure 16. Describing a particular farm along with its crops and soil characteristics.

CQ9: How do you describe the seeding date of a crop?

As illustrated in Figure 17, which is very similar to the one in the previous section, we can associate a plant date to a particular crop using the `hasPlantDate` data property from SAREF4AGRI.

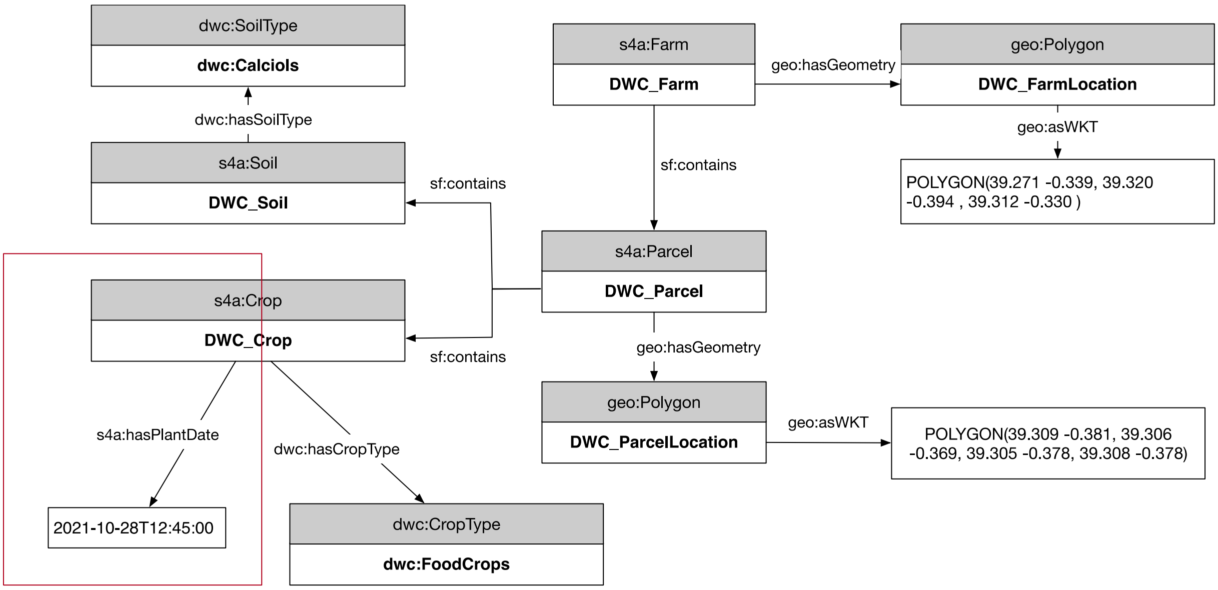

Figure 17. Describing the seeding date of a crop

CQ10: How do you describe a sampling device at a particular topological position within a wastewater treatment plant?

Figure 18 describes constructs already presented in previous sections, so we will in this section focus on the new aspects. In this case we have an automatic sampler deployed at a building space within a wastewater treatment plant. The sampler has two integrated sensors, a pH sensor (`DWC_PHSensor`) and a dissolved oxygen sensor (`DWC_DOSensor`), which measures water acidity and water oxygen level respectively. The `hasSubSystem` object property of the W3C SSN ontology connects these sensors to the sampler.

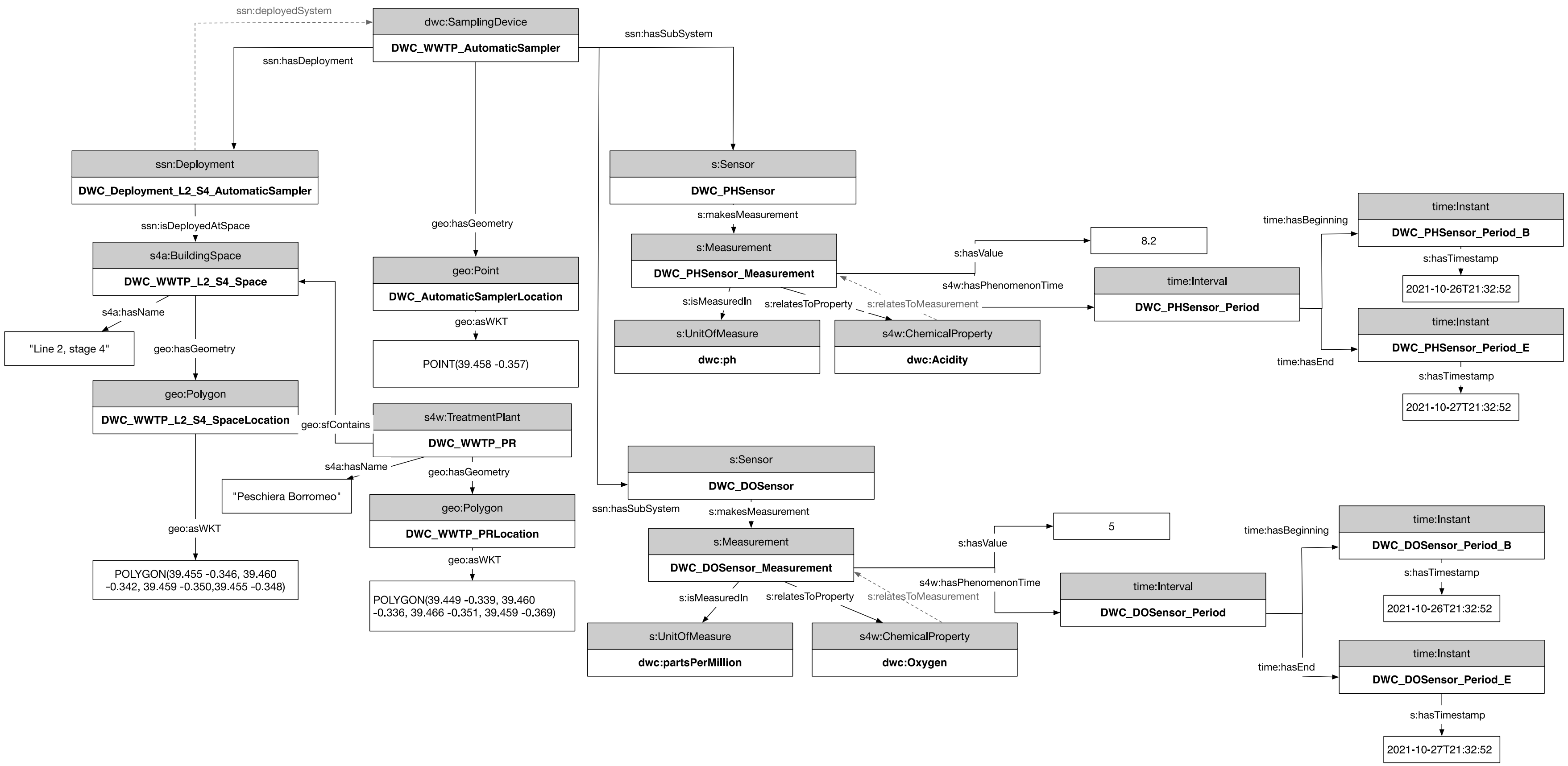

Figure 18. Describing the topology of a sampling device

CQ11: How do you describe the soil moisture as measured by the sensors deployed on the field?

Figure 19 describes the reporting of a soil moisture measurement performed by a soil tensiometer (`DWC_SoilTensioMeter`). The measurement relates to the property `SoilMoisture`, which is of type `SoilMoisture`. The measurement unit applied is kilo pascal, as expressed by kilopascal typed according to the `UnitOfMeasure` class.

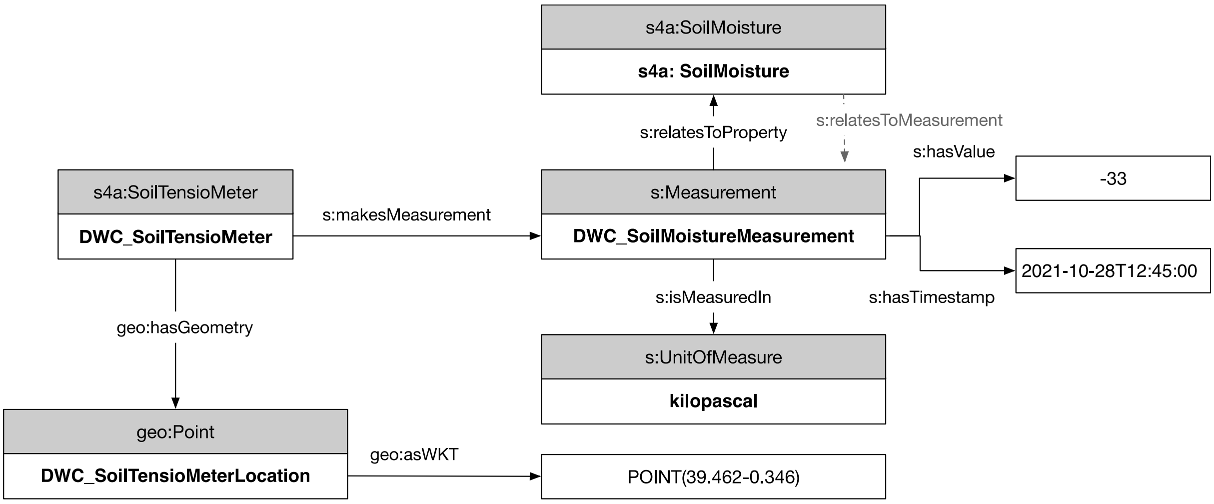

Figure 19. Describing the soil moisture

8. Possible application of the DWC Reference Ontology

This section will briefly describe some possible application areas for the reference ontology presented in the previous sections.

8.1. Provide a semantic description of water-related entities in an NGSI-LD compliant context broker

FIWARE plays an important role as an interoperability architecture in DWC as well as other research projects in the water management domain. FIWARE offers an industry-independent interoperability architecture and a set of open-source generic enablers (software components) to implement technology platforms typically related to internet of things (IoT), so-called smart domains. One of the underlying standards is NGSI, an ETSI standardised protocol for managing contextual information to describe different entities of concern. NGSI-LD is the newest version of the protocol and defines an information model (including a metamodel and a cross-domain ontology) derived from property graphs[18]. This is further described in this ETSI white paper [15] along with a set of recommendations when modelling and using NGSI-LD. When using NGSI-LD as the underlying graph model in an NGSI-LD-compliant context broker, there are certain conventions or recommendations that should be followed. Two recommendations that are especially relevant are:

- *"NGSI-LD users need to attach types to all the entities, relationships, and properties that they intend to instantiate in their implementations".*
- *"These types should be directly or indirectly derived from classes defined in shared OWL ontologies, thesauri, or taxonomies"*

The reason why these two recommendations, and the use of domain ontologies, is so important is that ontologies serve to define a standardised and formalised description of entities in a way that enables both human and machine interpretation of data associated with real-life entities. This further allows entities to be integrated with other entities in and across domains, support data retrieval tasks, and allow inference of new knowledge from the data. In other words, applying the ontologies described in this report, as well and the extensions proposed by DWC, in combination with the NGSI-LD information model will facilitate data interpretation, data integration and relevant data retrieval and knowledge creation in the water management domain.

It might be difficult to understand the relationship between ontologies and data models developed by initiatives such as smartdatamodels.org that develop data models for different domains, including water management. This relates to a challenge called semantic heterogeneity. Semantic heterogeneity occurs when e.g., a data attribute (e.g., *parcel*) can have multiple differing meanings. For example, the attribute *parcel* can be a part of a farm or something to be delivered in freight logistics. When you don´t know the context of the attribute *parcel*, it is difficult for a human to fully understand what it means, and even more so for a computer. Such context can for example be defined by vocabularies such as schema.org or more formal ontologies. Here, concepts are well-defined by a textual definition, taxonomies, and relationships to other concepts, effectively establishing a human- and machine-

---

[18] The property graph model is used by popular graph databases such as Neo4J (https://neo4j.com/) and specifies a model made up of nodes, relationships, and properties (specifying data about the nodes and relationships).

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

readable description of the context of for example the attribute *parcel*. This means that if you in a payload include a value for an attribute *parcel*, you can include a context reference that states that this value should be interpreted as "*An area of land, which might be used for grazing animals or planting crops*", according to the concept `Parcel` defined in the ontology SAREF4AGRI. A recent initiative at smartdatamodels.org is to provide links to concepts described by existing ontologies, including SAREF ontologies[19].

8.2. Enhancing semantic interoperability

Interoperability challenges occur whenever a system needs to interact with one or more other systems and these systems use different, typically non-standardised formats for describing data. In many cases the data elements in the formats intend to describe the same thing, but the syntax, the definitions, and the structure used for describing this thing is different because the data schemas prescribing the data elements are developed by different parties. This is called semantic heterogeneity.

In DWC several different data sources (sensors, weather stations, third-party legacy systems and other sources) and applications will feed data into and retrieve data from a FIWARE context broker. To do so, the data formats used by the data sources and applications must be mapped to that of the context broker. If the format used for describing the entities in the context broker is based on the NGSI information model with typing towards concepts defined in domain ontologies, each entity will be uniquely identified with a URI and its context will be described using these ontologies that provide a standardised and formal definition of the entity that can be used for human and/or machine processing. Such a scheme will reduce semantic heterogeneity and enhance interoperability.

From a more practical point of view, FIWARE offers a middleware[20] that can support interfacing at protocol level with IoT devices and third-party systems, but the mapping between data formats at semantic level (i.e., aligning the meaning and intent of the data) must still be performed manually. Identifying such mappings between the interacting systems can be very resource demanding and otherwise challenging, especially if the data formats are many and/or extensive. Task 4.3.3 in the DWC project (semantic interoperability middleware) aims to support this mapping process by using the reference ontology along with syntactic [16] and semantic [17] matching techniques. This approach aims to takes one or more source data format as input and suggest the most relevant ontology entity (class or property) for a given element in that data format, to simplify and (semi) automate the mapping task. The work performed in Task 4.3.3 will result in deliverable D4.7.

8.3. Utilising semantically enriched data for downstream data analytics tasks

Graph representation learning or graph embedding aim to learn non-explicit features of nodes and relationships in graph models[21]. These features can be used for data analysis in their own right, or they could be further processed by downstream machine learning methods. The process of graph representation learning is to represent graph structures (nodes and their context) in a numerical form, a vector representation. This is accomplished through an embedding process where the graph as input is run through a neural network (typically) where similar characteristics and contexts are of graph

[19] https://github.com/smart-data-models/data-models/tree/master/context/ontologies_files

[20] The IDAS Generic Enabler presented at: https://www.fiware.org/developers/catalogue/

[21] https://towardsdatascience.com/graph-representation-learning-dd64106c9763

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

nodes and edges are learnt. The goal of this task can be to predict (additional) properties of nodes in the graph, new relationships between nodes, or identify unseen clusters within the graph (community detection). Once nodes and relations in graphs are typed using ontologies, knowledge graphs can be derived from the graphs. Knowledge graph embedding techniques [18,19] have the same objectives as "traditional" graph representation learning techniques, but aim to preserve and use the semantics provided by ontologies during the learning phase. Such strategies have quite successfully been applied in risk assessment of biological effects of chemical release in water, as presented in Myklebust et al. [20]. The idea is that based on the known biological effects declared in the knowledge graph, the knowledge embedding model will compute/learn the probability of unknown biological effects. This is also known as link prediction. For example, the knowledge graph states that a chemical compound *X* affects (e.g., has a lethal effect) on species *Y*. How the compound *X* affects species *Z* is not known. But based on the learned vector positions of *X*, *Y* and *Z*, and the quantified effect (relationship) *X* has on *Y*, the model also predicts that *X* affects *Z*. The quantified effect (relationship) can for example be represented by some computed distance/offset between *X* and *Y*.

9. Conclusions

This report described how existing ontologies from the (1) Internet of Things domain, (2) the water management domain and the (3) agricultural domain cover the data requirements in two cases of the digital-water.city (DWC) project, namely Paris and Milan. The re-used ontologies are SAREF, SAREF4WATR and SAREF4AGRI respectively. These, and other ontologies imported by these three ontologies, have been mapped with the requirements elicited in DWC. The requirements mapping shows that these existing ontologies, along with more generic ontologies they import such as for describing spatial and temporal entities, largely address the requirements elicited in DWC. So, instead of developing the DWC Reference Ontology from scratch, we abide by the principle of re-use, which is an important principle in ontology engineering. For the requirements that are not addressed by the existing ontologies, the report describes how the needed extensions could be integrated with the existing ontologies. Most of the extensions relate to expressing details associated with agricultural entities, which is an important aspect in the Milan case specifically, but also extensions to SAREF4WATR and the more generic SAREF ontology have been proposed.

The report further describes how the proposed reference ontology has been evaluated automatically using a reasoner service and documented by illustrating how a set of competency questions derived from requirements and interaction with representatives from the Paris and Milan cases can be addressed by the ontology.

The reference ontology can be used to supplement the NGSI-LD information model with more expressive semantics in a FIWARE context broker, to support interoperability and data integration within the water management domain and with other domains, and to provide semantically enriched data for downstream data analysis tasks.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

10. References

1. Schwarzmüller H, Vennesland A, Halland Haro P, Bour G. D4.1 Interoperable and secure flow of information: Cyber-physical sphere and interoperability aspects in the utilities regarding the DWC solutions. Digital Water City; 2020 Jun.

2. Vennesland A, Haro PH, Hanssen BJ. D4.4 Semantic Interoperability Design Requirements. Digital Water City; 2020.

3. Staab S, Studer R. Handbook on Ontologies. Springer Science & Business Media; 2013.

4. Guarino N. Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. Vol. 46. IOS press; 1998.

5. ETSI TC SmartM2M. ETSI TS 103 264 V3.1.1, SmartM2M; Smart Applications; Reference Ontology and oneM2M Mapping. 2020.

6. ETSI TC SmartM2M. ETSI TS 103 548 V1.1.2, SmartM2M; SAREF consolidation with new reference ontology patterns, based on the experience from the SEAS project. 2020.

7. ETSI TC SmartM2M. SmartM2M, Extension to SAREF, Part 10: Water Domain [Internet]. 2020. Report No.: ETSI TS 103 410-10. Available from: https://www.etsi.org/deliver/etsi_ts/103400_103499/10341010/01.01.01_60/ts_10341010v010101p.pdf

8. ETSI TC SmartM2M. ETSI TS 103 410-6 V1.1.2, SmartM2M;Extension to SAREF; Part 6: Smart Agriculture and Food Chain Domain. 2020.

9. Fernández-López M, Gómez-Pérez A, Juristo N. Methontology: from ontological art towards ontological engineering. 1997;

10. Suárez-Figueroa MC, Gómez-Pérez A, Fernández-López M. The NeOn methodology for ontology engineering. In: Ontology engineering in a networked world. Springer; 2012. p. 9–34.

11. Noy NF, McGuinness DL, others. Ontology development 101: A guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and …; 2001.

12. Peroni S. A simplified agile methodology for ontology development. In: OWL: Experiences and Directions–Reasoner Evaluation. Springer; 2016. p. 55–69.

13. De Nicola A, Missikoff M, Navigli R. A proposal for a unified process for ontology building: UPON. In: International Conference on Database and Expert Systems Applications. Springer; 2005. p. 655–64.

14. De Nicola A, Missikoff M. A lightweight methodology for rapid ontology engineering. Communications of the ACM. 2016;59(3):79–86.

15. Privat G. Guidelines for Modelling with NGSI-LD [Internet]. ETSI; 2021. Report No.: 42. Available from: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp_42_NGSI_LD.pdf

16. Cheatham M, Hitzler P. String similarity metrics for ontology alignment. In: International semantic web conference. Springer; 2013. p. 294–309.

17. Vennesland A. Semantic Matching - Dynamic Composition of Matcher Ensembles for Ontology Alignment. [Trondheim, Norway]: Norwegian University of Science and Technology; 2020.

18. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

applications. IEEE Transactions on Knowledge and Data Engineering. 2017;29(12):2724–43.

19. Ristoski P, Rosati J, Di Noia T, De Leone R, Paulheim H. RDF2Vec: RDF graph embeddings and their applications. Semantic Web. 2019;10(4):721–52.

20. Myklebust EB, Jimenez-Ruiz E, Chen J, Wolf R, Tollefsen KE. Knowledge graph embedding for ecotoxicological effect prediction. In: International Semantic Web Conference. Springer; 2019. p. 490–506.

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.

Digital Water .City

Leading urban water management to its digital future

digital-water.city
digitalwater_eu

digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.