Pattern-based design for modelling an ontology network in the water and health domains

Anna Sofia Lippolis[1,2,†], Giorgia Lodi[1,†] and Andrea Giovanni Nuzzolese[1,*,†]

[1]Semantic Technology Laboratory, Istitute of Cognitive Sciences and Technologies, National Research Council, Rome and Bologna, Italy
[2]Department of Phylosophy and Communication Studies, University of Bologna, Bologna, Italy

Abstract

Recently, an increasing interest in the management of water and health resources has been recorded. This interest is fed by the global sustainability challenges posed to the humanity that have water scarcity and quality at their core. Thus, the availability of effective, meaningful and reusable ontologies and data model is crucial to address those issues in the broader context of the Sustainable Development Goals of clean water and sanitation as targeted by the United Nations. In this paper, we present the ontology network developed in the context of the Water Health Open knoWledge project (WHOW) along with its design methodology. The ontology network consists of five pattern-based modules that we extensively describe in this paper by also including a review of the state of the art in terms similar works in both domains water and health.

Keywords

Ontology Design Patterns, Pattern-based Design, Ontology Network, Modular Ontology Design, Water Quality, Health, Environmental Data, Clean Water and Sanitation

1. Introduction

Interest in water and sanitation management has grown in recent years driven by global sustainability challenges that prioritise, among the others, clean water and sanitation, as outlined in the UN Sustainable Development Goals[1].

To provide effective responses to these global issues, the availability of high quality and open data models becomes an essential requirement. However, the heterogeneity and complexity of water and health data, when available, can pose significant challenges. This paper introduces the ontology network of the Water Health Open knoWledge project[2] (WHOW), which aims at building the first European open distributed knowledge graph for linking, using a common semantics, data on water consumption and quality with health parameters (e.g., infectious diseases rates, general health conditions of the population). Designed to understand the impact

14th Workshop on Ontology Design and Patterns (WOP 2023)

*Corresponding author.
†The authors are sorted alphabetically as they equally contributed to this paper.
annasofia.lippolis@istc.cnr.it (A. S. Lippolis); giorgia.lodi@cnr.it (G. Lodi); andreagiovanni.nuzzolese@cnr.it (A. G. Nuzzolese)
0000-0002-0266-3452 (A. S. Lippolis); 0009-0000-9032-4831 (G. Lodi); 0000-0003-2928-9496 (A. G. Nuzzolese)

CEUR Workshop Proceedings (CEUR-WS.org)

[1]https://sdgs.un.org/goals.
[2]https://whowproject.eu/.

of water-related climate events, water quality, and water consumption on health, the ontology network provides a harmonized knowledge layer that can be re-used for analysis, research, and development of innovative services and applications in the water and health domains.

The ontology network consists of five pattern-based modules that we extensively describe in this paper by also including a review of the state of the art in terms similar works in both domains water and health.

The rest of the paper is organized as follows: (i) Section 2 presents the related work; (ii) Section 3 discusses the design methodology; (iii) Section 4 addresses the results achieved in terms of the ontology network; (iv) ; finally, (v) Section 5 concludes the paper, discusses the limitations, and defines future directions of research.

2. Related Work

In the context of the monitoring a pillar is the Semantic Sensor Network Ontology (SSN Ontology)[6]. It allows one to represent sensors and observational processes and implements, for the majority of its semantic elements, the ISO 19156 Observations and Measurements (O&M) standard, used also as reference model in the INSPIRE context.

Other European projects target water monitoring data models. This is the case of the ODALA[3] project that created the ODALA Air & Water application profile[4]. The profile builds on a core module derived from both O&M and the SSN Ontology. ODALA presents concepts similar to those defined in the WHOW water monitoring ontology; this creates the prerequisites for a semantic alignment between these knowledge graphs. In the same direction, [15] describes a knowledge-based approach aiming at water quality monitoring and pollution alerting through the proposed Observational Process Ontology (OPO). Similarly, [8] presents a three-module water quality ontology that combines numerous standards from different domains to obtain a comprehensive approach to the issue. These standards are, among others, GeoSPARQL[5], the O&M and SSN cited above, the RDF Data Cube[6] as well as non-ontological resources associated with standards (WaterML[7]). At the European level, the European Environmental Agency publishes a Linked Open Data section[8] that comprises data on water quality monitoring. This data is currently under investigation in order to enable possible links with the proposed WHOW knowledge graph.

As far as the health domain is concerned, although it is difficult to find (linked) open data available for the re-use, interesting resources were taken into account when creating the WHOW-KG. In particular, we mention here the Snomed standard[9] for health terms, that has been re-used in order to create proper links with our produced controlled vocabulary on infectious diseases.

In essence, although a variety of works in both domains can be identified, it is still difficult, to the best of our knowledge, to get access to a resource capable of linking the two domains

[3] https://odalaproject.eu/.
[4] https://purl.eu/doc/applicationprofile/AirAndWater/Water.
[5] https://www.ogc.org/standard/geosparql/
[6] https://www.w3.org/TR/vocab-data-cube/
[7] https://www.ogc.org/standard/waterml/
[8] https://www.eea.europa.eu/data-and-maps/daviz/eionet/data/eea-data.
[9] https://www.snomed.org/.

together as we propose with our ontology network.

3. Method

The methodology we used for modelling the ontology is inspired by the one defined in [5] and relies on eXtreme Design [2] (XD) for ontology modelling. XD emphasises the reuse of ontology design patterns [10] (ODPs) into an iterative and incremental process. More interestingly, XD is a collaborative methodology that fosters the cooperation among multiple actors with different roles (e.g. knowledge engineers, domain experts, etc.) to make sure all the modelling requirements are first captured and then effectively covered. Hence, we opted for XD since it fits our collaborative setting based on the co-creation programme. Furthermore, there is evidence il literature [1] that the reuse of ODPs (i) speeds up the ontology design process, (ii) eases design choices, (iii) produces more effective results in terms of ontology quality, and (iv) boosts interoperability.

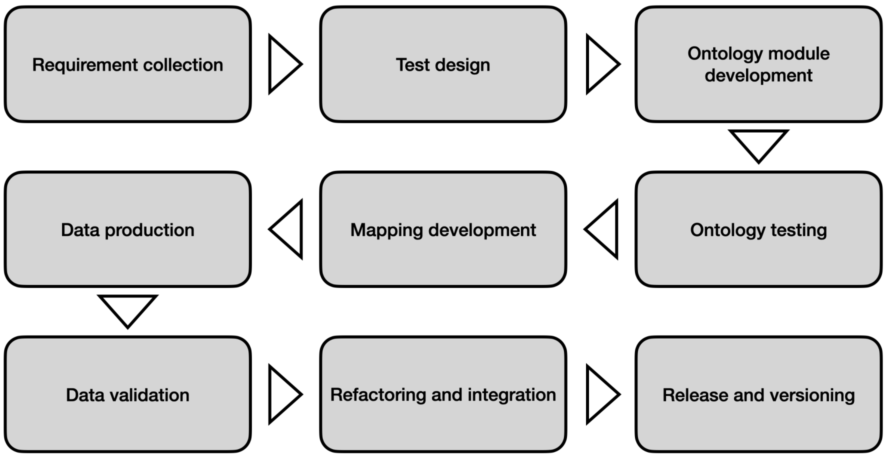

Figure 1: Methodology implemented for constructing the knowledge graph of the WHOW project.

Figure 1 shows the methodology implemented for constructing the knowledge graph of the WHOW project. The methodology encompasses data production steps that are part of our future work. In this paper we focus on the steps related to ontology design.

Ontology design. In such a figure, the activities named *requirement collection*, *test design*, *ontology module development*, and *ontology testing* come from XD and focus on ontology design. The requirement collection activity aims at eliciting the requirements as *competency questions* [11] (CQs). CQs are natural language questions conveying the ontological commitment expected from a knowledge graph (KG) and drive both ontology modelling and validation. In fact, on the one hand CQs are a means for ontology development. On the other, they can be converted to formal queries in order to assess the effectiveness of the resulting KG to cope with the requirements. We implemented the validation into the *ontology testing* activity. This was done by converting defined CQs into SPARQL and executing the latter as unit tests with toy data following the solution defined in [3]. The ontology development we applied is modular (cf. activity named ontology module development) allowing us to generate a set of networked ontologies. Each ontology of the network is a separate module designed with the purpose of minimising coupling with other ontology modules and maximising the internal cohesion of its conceptualisation. The re-use of external ontologies and ODPs was done by applying both the direct and indirect approach [14, 4]. Direct re-use is about embedding individual entities or importing implementations of ODPs or other ontologies in the network, thus making it highly dependent on them. Instead, indirect re-use is about applying relevant entities and patterns from external ontologies as templates, by reproducing them in the ontologies of the network and providing possible extensions. We opted for direct re-use in case of widely adopted vocabularies, such as SKOS, the Time ontology available in the Italian national catalog of semantic assets for public administrations[10], aligned with the W3C time ontology, and the top-level[11] (TOP) and environmental monitoring facilities[12] (EMF) ontologies of the Linked ISPRA project[13]. TOP is used as a top-level ontology that provides general concepts and relations, whilst EMF provides core domain concepts and relations for modelling environmental monitoring data. On the contrary, we opted for the indirect approach for re-using patterns and to support interoperability with other pertinent ontologies, e.g. SSN/SOSA[14] [12]. The latter case was realised by means of alignments axioms, such as rdfs:subClassOf and owl:equivalentClass in dedicated alignment ontologies.

4. Ontology Network

The WHOW ontology network consists of 5 ontology modules. In Figure 2 each ontology is represented as a circle, whilst the arrows represent `owl:imports` axioms among the ontologies. The ontologies represented as white circles are external ontologies we re-used with the direct approach. The ontologies represented as gray circles are the novel contributions. The base namespace defined novel ontologies is `https://w3id.org/italia/whow/onto/`. From this base namespace each module defines its local namespace following the table of prefixes reported in Figure 2. Table 1 reports core metrics about the ontology network, which is: (i) under version control on GitHub[15]; (ii) shared on Zenodo[16] with a CC-BY 4.0 International licence; and (iii) findable on Linked Open Vocabularies[17].

Hydrography module. The *Hydrography* ontology (prefix `hydro:`[18]) represents a general-purpose hydrological taxonomy following the definitions given in the European Directive

---

[10] https://schema.gov.it

[11] https://github.com/whow-project/semantic-assets/blob/main/ispra-ontology-network/top/latest/top.rdf.

[12] https://github.com/whow-project/semantic-assets/blob/main/ispra-ontology-network/inspire-mf/latest/inspire-mf.rdf.

[13] https://dati.isprambiente.it/

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

[15] https://github.com/whow-project/semantic-assets/tree/main/ontologies.

[16] https://doi.org/10.5281/zenodo.7916179.

[17] https://lov.linkeddata.es/dataset/lov/.

[18] The prefix `hydro:` stands for the namespace https://w3id.org/whow/onto/hydrography.

Table 1
Statics of the ontology network.

Metric	Value	Metric	Value
Axioms	2,672	SubObjectPropertyOf axioms	137
Classes	120	Inverse object properties	61
Object properties	161	Transitive object properties	10
Datatype properties	21	Declared property domains	155
DL expressivity	SRIQ(D)	Declared property ranges	153
SubClassOf axioms	255	Property chains	6
Disjoint classes	22	Annotation assertions	1,412

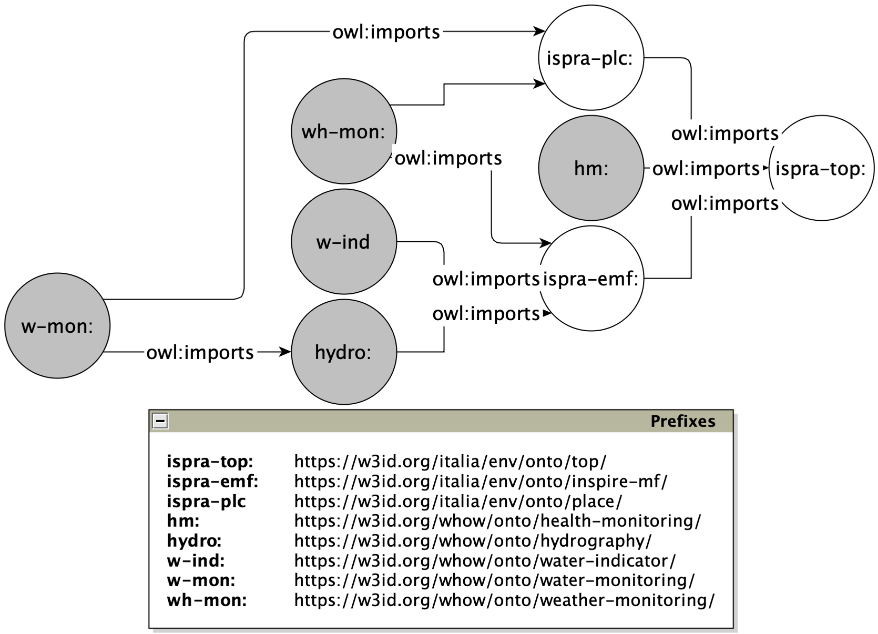

Figure 2: The WHOW ontology network.

2000/60/EC[19]. The `hydro:` ontology is depicted in Figure 3 using Graffoo as reference notation [9]. With white rectangles we indicate classes directly re-used from external ontologies and with grey rectangles new defined classes. The top-level class is `hydro:WaterFeature`, a subclass of the ISPRA ontology `ispra-emf:FeatureOfInterest` with `hydro:WaterBasin` and `hydro:WaterBody` as subclasses. A `hydro:WaterBody` further specialises into a number of subclasses defining a clear classification among the different types of water bodies. Those subclasses are `hydro:TransitionalWaterBody`, `hydro:MarineWaterBody`, `hydro:RiverWaterBody`, `hydro:LakeWaterBody`, `hydro:GroundWaterBody`, and `hydro:CoastalWaterBody`. In this on-

[19]https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0060&rid=2.

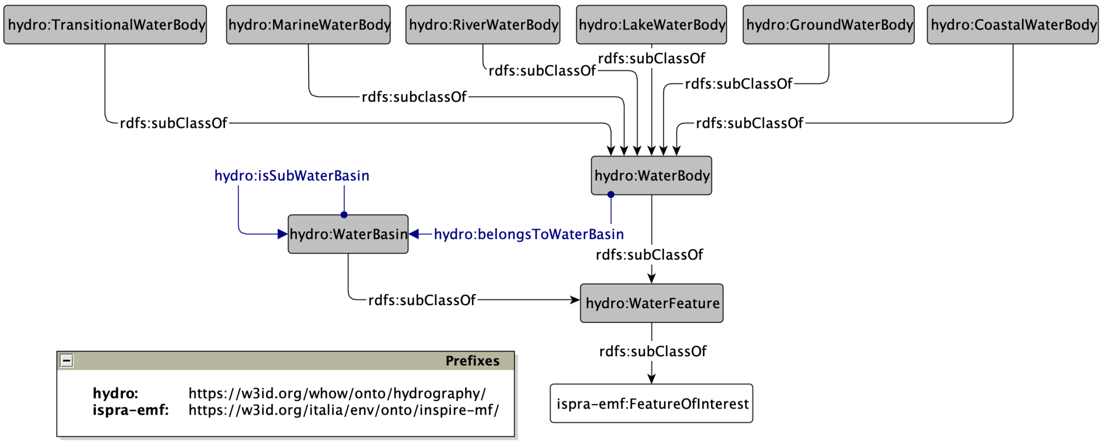

Figure 3: The Hydrography ontology.

tology we reused the PartOf ODP[20] for expressing parthood between water basins (cf. the object property `hydro:isSubWaterBasin`).

Water Monitoring module. The *Water Monitoring* ontology is identified by the prefix `w-mon:`[21]. It provides means to represent observations related to the quality of water courses, such as chemical and biological substances found in water bodies. The requirements for the representation of water observations are defined according to the data provided by the data providers involved in the project and the standards and directives in terms of observations and water-related assessments. For what concerns the representation of water observations, it is possible to refer to European directives: (i) those deriving from taxonomies from European Directive 98/83/CE (and subsequent ones)[22], confirmed by the Italian Ministry of Health[23], concerning parameters of the waters for human consumption, and (ii) those deriving from the European Directive 2009/90/EC[24], concerning parameters of surface waters. Thus, water quality monitoring requires the integration of heterogeneous types of both observations and observation objects derived from samplers. As a result, in the ontology (cf. Figure 4), a `w-mon:WaterObservation` is divided into `w-mon:DrinkingWaterObservation`, `w-mon:SurfaceOrGroundWaterObservation`, and `w-mon:RadioActivityObservation`, which are, in turn, further divided into subclasses based on the specific parameter being observed. In fact, the observations that have as an object a microbiological agent or a chemical substance, monitor it through its concentration in the water. On the contrary, observations on properties of water, such as hardness, density or pH, do not imply the presence of an object being observed sinse no chemical substance or microbiological agent is implied there. The ontology follows the

[20] http://ontologydesignpatterns.org/wiki/Submissions:PartOf.
[21] The prefix `w-mon:` stands for the namespace https://w3id.org/whow/onto/water-monitoring.
[22] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31998L0083.
[23] Water quality parameters published by Italian Ministry of Health: https://www.salute.gov.it/portale/temi/p2_6.jsp?lingua=italiano&id=4464&area=acque_potabili&menu=co.
[24] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0060&rid=2.

Stimulus-Sensor-Observation Ontology Design Pattern (SSO ODP) [13], which is a standard for the Infrastructure for Spatial Information in Europe [7], and the Specimen model of ISO 19156:2011[25], which outlines the properties of sampling process features.

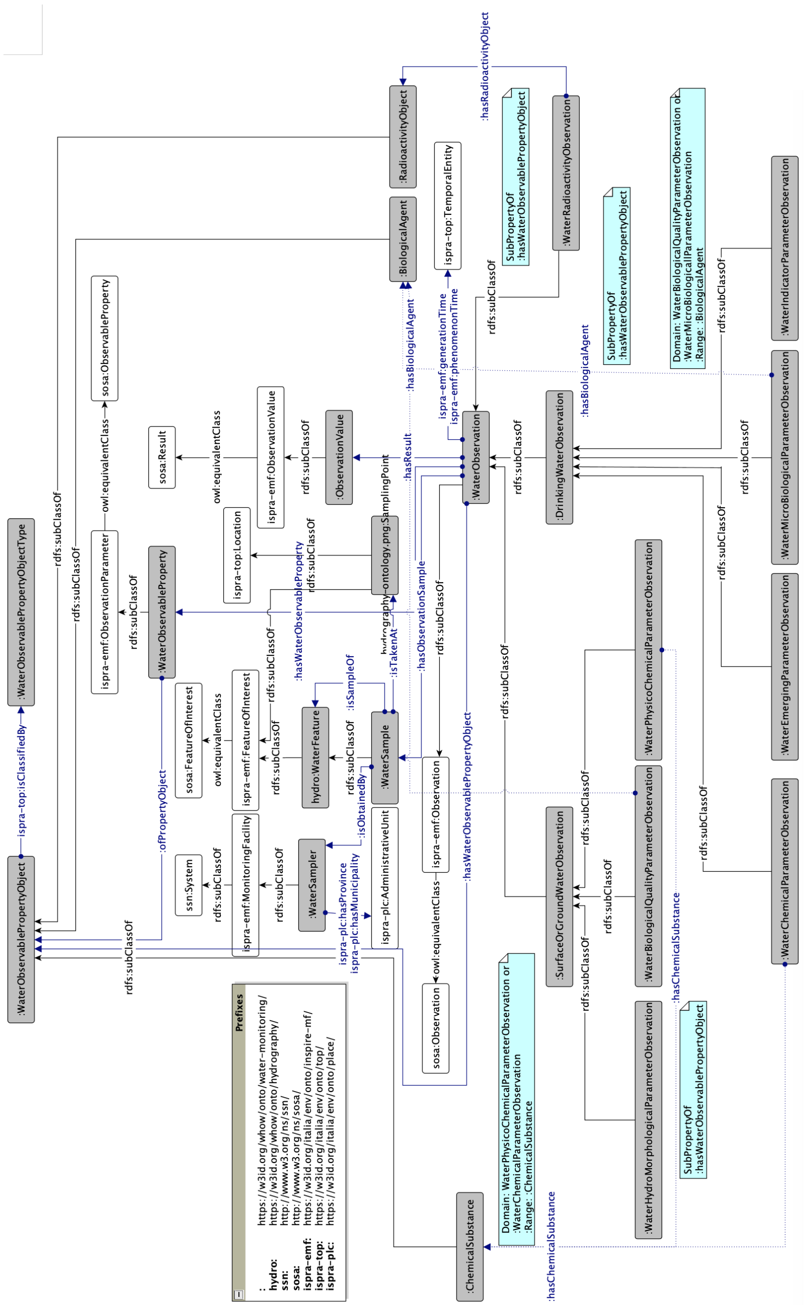

Figure 4: The Water Monitoring ontology.

[25]http://www.iso.org/iso/catalogue_detail.htm?csnumber=32574.

Water Indicator module. The **Water Indicator** ontology, with prefix `w-ind:`[26], re-uses the Indicator ontology design pattern[27] defined in OntoPiA[28], which is the Italian national network of ontologies and controlled vocabularies. This pattern is re-used to address indicators and metrics for the indicator calculation of water quality. As shown in Figure 5, the indicators can be bathing water quality classes or indicators of lakes' chemical status.

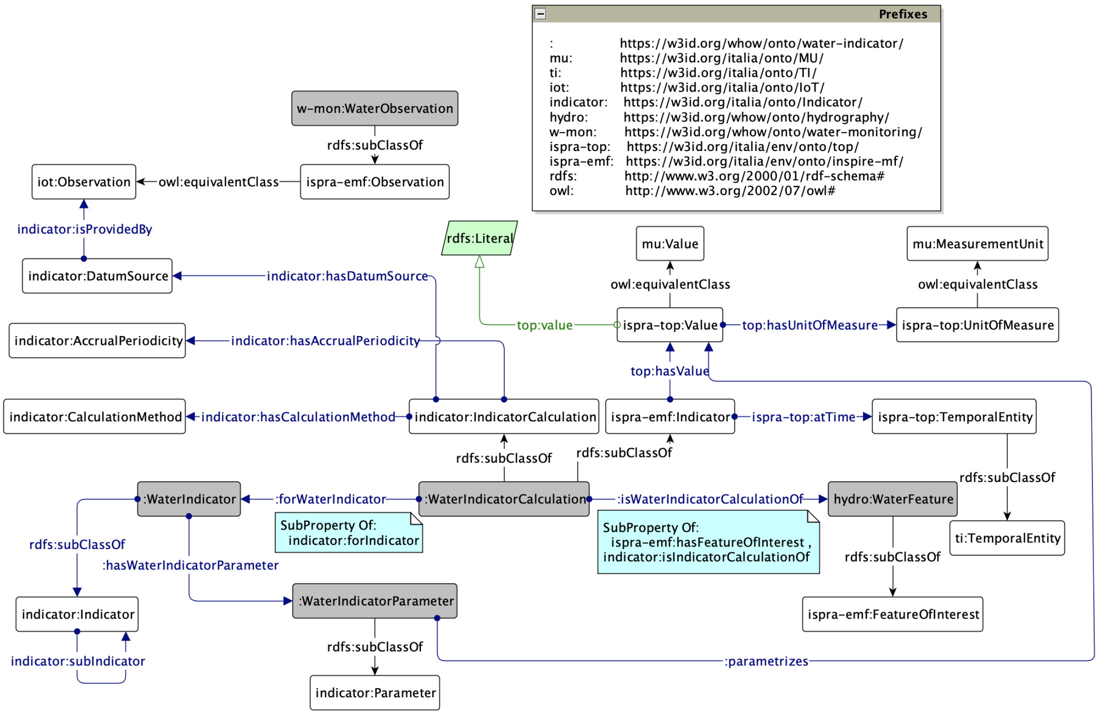

Figure 5: The Water Indicator ontology.

Weather Monitoring module. Similarly to the Water Monitoring module, the *Weather Monitoring* ontology, with prefix `wh-mon:`[29] (cf. Figure 6), has its focus on a `wh-mon:WeatherObservation` related to a `wh-mon:WeatherFeatureOfInterest` (either ground-level soil, air, wind, snow or rainfall), `wh-mon:WeatherObservableProperty` and `wh-mon:WeatherSensor` hosted by a `wh-mon:WeatherStation`. It reuses the ISPRA ontology network to model observations and related properties. This model is meant to address the need to represent weather observations that could serve as a basis to derive information on extreme events monitoring and prediction, such as rainfalls and snow levels.

[26] The prefix `w-ind:` stands for the namespace https://w3id.org/whow/onto/water-indicator.
[27] https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie/Indicator/latest.
[28] https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master.
[29] The prefix `wh-mon:` stands for the namespace https://w3id.org/whow/onto/weather-monitoring.

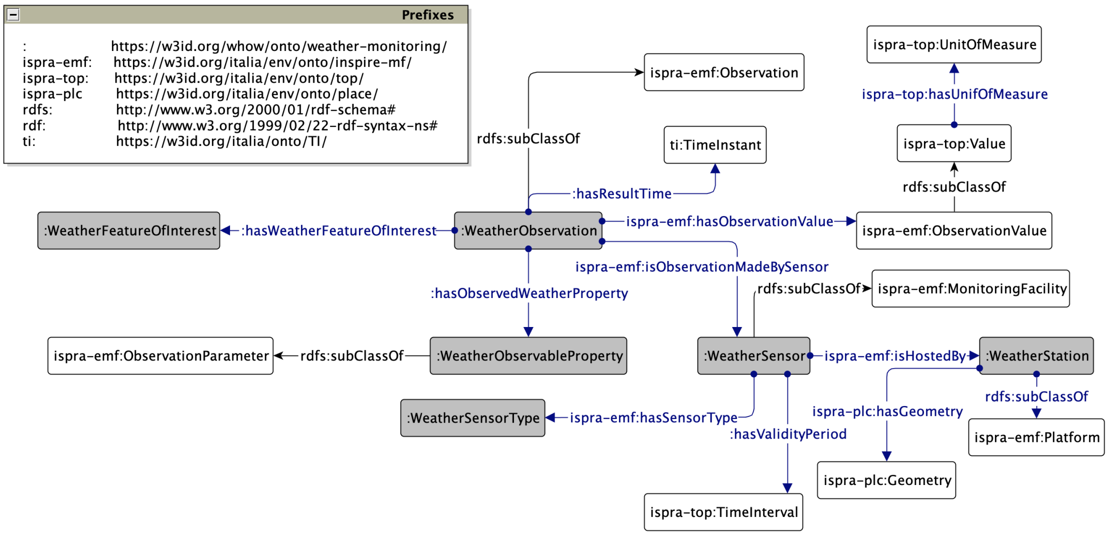

Figure 6: The Weather Monitoring ontology.

Health Monitoring module. Finally, the *Health Monitoring* ontology, whose prefix is `hm:`[30] reuses the OntoPiA Indicator ontology and focuses on the representation of health indicators coming from regional healthcare facilities. Examples include drug distribution rates and hospital accesses according to disease code and facility involved (cf. Figure 7). Different types of `hm:HealthcareIndicatorCalculation` are defined, based on the typology of indicator they describe, i.e. infectious disease rate, death rates related to diagnosis, average hospital stay and drug distribution. The indicator calculation also refers to a statistical dimension class, `hm:ClinicalCohort`, which specifies the population referred to as defined by a number of criteria, that is `hm:CohortCriteriaDescription`, such as age and gender. By reusing the `ispra-top:` ontology, it is also possible to model the health agency that supervises a specific area.

5. Conclusions and future work

In this paper, we have introduced the ontology network of the Water Health Open knoWledge project (WHOW) that links water quality observations with health parameters (e.g. infectious disease rates), thus implementing the well-known connection of water quality effects on people's health. The WHOW ontology network is (i) modular, (ii) open to maximise re-use, (iii) multilingual in that labels and comments are provided in both Italian and English, when possible, and (iv) built according to FAIR principles. As part of our ongoing and future work we plan to construct a knowledge graph, i.e. WHOW-KG, by producing Linked Open Data from the data providers involved in the WHOW project. Currently two data providers, i.e. the Italian

[30]The prefix `wh-mon:`, stands for the namespace https://w3id.org/whow/onto/weather-monitoring.

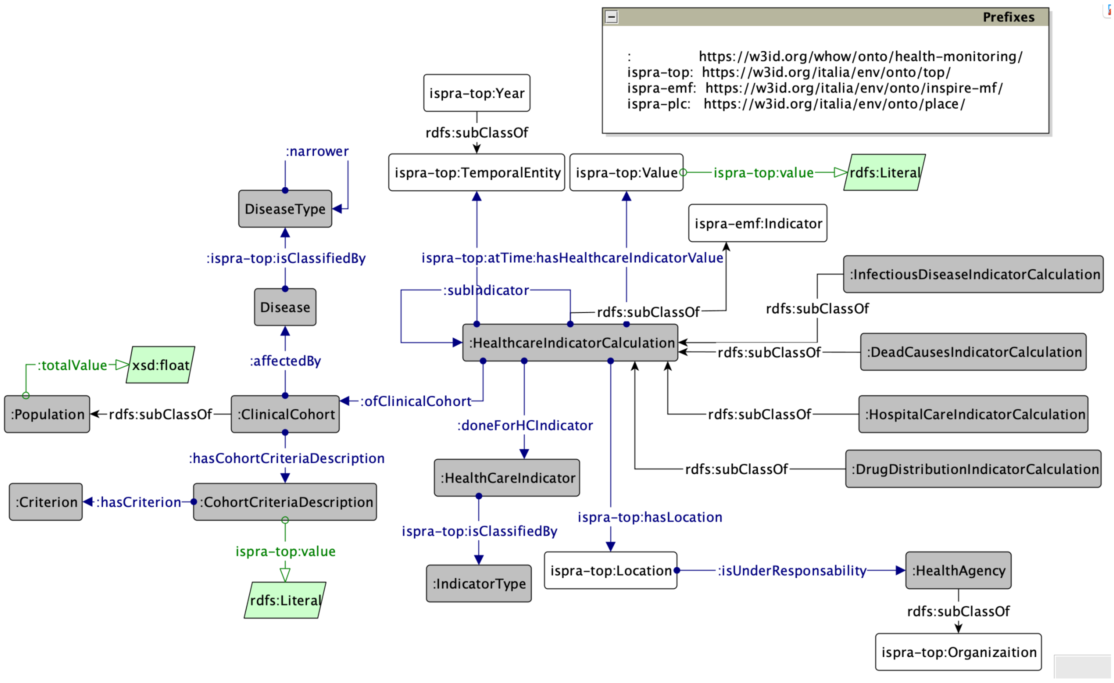

Figure 7: The Health Monitoring ontology.

National Institute for Environmental Protection and Research[31] (ISPRA) and ARIA Spa[32], are involved in the knowledge graph construction process. The resulting knowledge graph follows a decentralised and distributed paradigm. In this scenario new data providers might publish their data as linked open data compliant with the WHOW ontology network by using their preferred persistent URIs and setting up their own SPARQL endpoint, thus maximising the sustainability of the WHOW-KG.

Acknowledgements

This work has been supported by the Water Health Open knoWledge (WHOW) project co-financed by the Connecting European Facility programme of the European Union under grant agreement INEA/CEF/ICT/A2019/206322.

References

[1] E. Blomqvist et al. “Experimenting with eXtreme design”. In: *Proceedings of the 17th International Conference on Knowledge Engineering and Knowledge Management.* Edited by P. Cimiano et al. Volume 6317. Lecture Notes in Computer Science. Springer. 2010, pages 120–134. DOI: 10.1007/978-3-642-16438-5_9.

[31] https://www.isprambiente.gov.it/en.

[32] https://www.ariaspa.it/wps/portal/Aria/Home.

[2] Eva Blomqvist et al. “Engineering Ontologies with Patterns - The eXtreme Design Methodology”. In: *Ontology Engineering with Ontology Design Patterns*. Edited by Pascal Hitzler et al. Volume 25. Studies on the Semantic Web. IOS Press, 2016, pages 23–50. ISBN: 978-1-61499-676-7. DOI: 10.3233/978-1-61499-676-7-23.

[3] Valentina Anita Carriero et al. “Agile Knowledge Graph Testing with TESTaLOD”. In: *ISWC Satellites*. Edited by Mari Carmen SuÃ¡rez-Figueroa et al. Volume 2456. CEUR Workshop Proceedings. CEUR-WS.org, 2019, pages 221–224.

[4] Valentina Anita Carriero et al. “The Landscape of Ontology Reuse Approaches”. In: *Applications and Practices in Ontology Design, Extraction, and Reasoning*. Edited by Giuseppe Cota et al. Volume 49. Studies on the Semantic Web. IOS Press, 2020, pages 21–38. DOI: 10.3233/SSW200033.

[5] Valentina Anita Carriero et al. “Pattern-based design applied to cultural heritage knowledge graphs”. In: *Semantic Web* 12.2 (2021). Publisher: IOS Press, pages 313–357. ISSN: 2210-4968. DOI: 10.3233/SW-200422.

[6] Michael Compton et al. “The SSN ontology of the W3C semantic sensor network incubator group”. In: *Journal of Web Semantics* 17 (2012), pages 25–32. DOI: 10.1016/j.websem.2012.05.003.

[7] Simon Cox. *Guidelines for the use of Observations Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development*. INSPIRE Maintenance and Implementation Group (MIG), Jan. 2011, pages 1–32.

[8] Juan D. Rondón Díaz et al. “Characterizing water quality datasets through multi-dimensional knowledge graphs: a case study of the Bogota river basin”. In: *Journal of Hydroinformatics* (2022). DOI: 10.2166/hydro.2022.070.

[9] Riccardo Falco et al. “Modelling OWL Ontologies with Graffoo”. In: *ESWC (Satellite Events)*. Edited by Valentina Presutti et al. Volume 8798. Lecture Notes in Computer Science. Springer, 2014, pages 320–325. ISBN: 978-3-319-11954-0. DOI: 10.1007/978-3-319-11955-7_42.

[10] Aldo Gangemi et al. “Ontology design patterns”. In: *Handbook on Ontologies*. Springer, 2009, pages 221–243. DOI: 10.1007/978-3-540-92673-3_10.

[11] Michael Grüninger et al. “The role of competency questions in enterprise engineering”. In: *BenchmarkingâTheory and practice*. Springer, 1995, pages 22–31.

[12] Armin Haller et al. “The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation.” In: *Semantic Web* 10.1 (2019), pages 9–32. DOI: 10.3233/SW-180320.

[13] Krzysztof Janowicz et al. “The Stimulus-Sensor-Observation Ontology Design Pattern and its Integration into the Semantic Sensor Network Ontology”. In: *SSN 2010*. Edited by Kerry Taylor et al. Volume 668. CEUR Workshop Proceedings. CEUR-WS.org, 2010.

[14] Valentina Presutti et al. "The role of ontology design patterns in linked data projects". In: *Proceedings of the 35th International Conference on Conceptual Modeling*. Edited by Isabelle Comyn-Wattiau et al. Volume 9974. Lecture Notes in Computer Science. Springer. 2016, pages 113–121. DOI: 10.1007/978-3-319-46397-1_9.

[15] Xiaolei Wang et al. "An Observational Process Ontology-Based Modeling Approach for Water Quality Monitoring". In: *Water* 12 (Mar. 2020), page 715. DOI: 10.3390/w12030715.