D4.1– THE aqua3S ONTOLOGY AND SEMANTIC REASONING SUPPORT

WP4 – Multi-sensor semantic data fusion for intelligent event detection

aqua3s project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 832876.

Document Information

GRANT AGREEMENT NUMBER	832876	ACRONYM	aqua3S
FULL TITLE	Enhancing standardisation strategies to integrate innovative technologies for Safety and Security in existing water networks.		
START DATE	1st September 2019	DURATION	36 months
PROJECT URL	www.aqua3s.eu		
DELIVERABLE	D4.1 – The aqua3S ontology and semantic reasoning support		
WORK PACKAGE	WP4 – Multi-sensor semantic data fusion for intelligent event detection		
DATE OF DELIVERY	CONTRACTUAL: October 2020	ACTUAL	November 2020
NATURE	Report	DISSEMINATION LEVEL	Public
LEAD BENEFICIARY	CERTH		
RESPONSIBLE AUTHOR	Anastasios Karakostas (CERTH)		
CONTRIBUTIONS FROM	Anastasia Moumtzidou (CERTH), Stefanos Vrochidis (CERTH)		
ABSTRACT	This deliverable presents the developed aqua3S ontology that models information pertinent for water distribution management systems. The aqua3S ontology was formalized in the Web Ontology Language version 2 (OWL2) and reused and extended numerous standardized resources. The ontology is capable of storing semantic information about the system and subsequently enables semantic reasoning to enhance the decision support services. A framework for semantic reasoning based on the ontology was also developed and acts as the basis for further decision support. Overall, the ontology models the core entities of the studied system, with accordance to international standards, and is used as part of the decision support services that the system offers.		

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	22/09/2020	ToC	Table of Contents	Anastasia Moumtzidou (CERTH)
0.2	2/11/2020	Working version	Version for internal review	Anastasios Karakostas (CERTH), Anastasia Moumtzidou (CERTH), Stefanos Vrochidis (CERTH)
0.3	12/11/2020	1st draft	Internal review	Eric Chauveheid (VVQ), David Pannocchia (CENTRIC)
0.4	16/11/2020	Working version	Addressing of the reviewers' comments	Anastasios Karakostas (CERTH)
1.0	17/11/2020	Final doc	Final quality check and issue of the	Anastasios Karakostas (CERTH),

			final document	Moumtzidou Anastasia (CERTH)
2.0	23/7/2021	Revised doc	Based on the received comments, the following adjustments were made: • Replace 'ISO 15975' with 'EN 15975' to the entirety of the document • Update of Section 5.2.1 regarding the use and extension of SAREF4WATR	Anastasios Karakostas (CERTH), Moumtzidou Anastasia (CERTH)

Disclaimer

Copyright message

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ACRONYMS

ABox	Assertion component
API	Application Programming Interface
CQ	Competency Question
DL	Description Logic
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
ESO	Event and implied Situation Ontology
EU	European Union
FOAF	Friend of a Friend
FR	First Responder
GIS	geographic information system
ISO	International Organization for Standardization
IoT	Internet of Things
JSON	Javascript Standard Object Notation
KB	Knowledge Base
KBS	Knowledge Base Service
LODE	Linking Open Descriptions of Events
OGC	Open Geospatial Consortium
OWL	Web Ontology Languge
RDF	Resource Description Framework
S4W	Smart Applications REFerence extension for Water
SAREF	Smart Applications REFerence
SIOC	Semantically-Interlinked Online Communities Project
SN	Social Network
SPARQL	SPARQL Protocol and RDF Query Language
W3C	World Wide Web Consortium

1. Executive summary

Task 4.1 that is entitled *Data harmonization, semantic representation and aqua3S ontology* aims to capitalize on semantic web technologies in order to harmonize data from heterogeneous sources. Additionally, the raw data created from the aqua3S system will be semantically annotated according to domain knowledge. Moreover, the semantic representation is compliant to corresponding standards.

This deliverable describes the developed ontology that aims to semantically represent knowledge from different tasks of WP3, Sensors, Crowdsourcing, Drones and Satellite data. Along with all the data sources, the semantic representation of the crisis itself is the backbone of the ontology and the connective link between the various sources. Towards the creation of an ontology that is semantically sound and adopts international standards, existing ontologies and resources were adapted and extended. In doing so, interoperability, extensibility and the sharing options of the ontology is significantly improved. The specification and the semantics of the *crisis* portion of the ontology are inspired mainly from the *EN 15975 - Security of drinking water supply* (EN-15975, 2013) that covers exactly the domain of interest of the aqua3S project. The modelling of sensor data was implemented by reusing the SAREF ontology (Daniele, den Hartog, & Roes, 2015) and its extension for the water domain, which is well explained and extensive. For crowdsourcing and specifically social media data and analysis, the developed ontology imports classes and properties from the SIOC (Bojars, et al., 2010) ontology, a W3C standard and used in a wide range of applications. The representation of visual data from Drones and from Satellite sources is still under consideration and not included in this document. In all cases, for temporal and spatial representation, the *Time OWL* (Hobbs & Pan, 2006) ontology (W3C standard) and the *GeoSPARQL* (Perry & Herring) ontology (compliant with OGC standards) are used respectively.

The semantic reasoning component ingests the data from the context broker and semantically transforms them according to the ontology. The incoming data are from the heterogeneous sources described above. After the transformation of the data, they are stored in a knowledge repository in the form of RDF triplets. This repository, also known as a Knowledge Base (KB), enables semantic operations and in particular semantic reasoning, which is used as part of the decision support services. The semantic reasoning follows a rule-based approach that combines data driven and knowledge driven reasoning. The semantic reasoning capabilities of the component are expected to enhance the decision support of the system and including but not limited to the provision of inferred knowledge about the crisis to the Crisis Classification component (T5.4).

2. Introduction

Water distribution systems are comprised of a variety of different components that must be monitored in order to combat crises as effectively as possible. In particular, the subsystems that monitor the different components are varied and diverse, and as a result, their produced data are heterogeneous and occupy different modalities. For example, a water distribution system might be monitored via a number of sensors that measure the properties of the water in different points, along with Social Media monitoring that detects changes in the public perception of the water quality. Consequently, in order to utilize them all it is imperative to build a uniform vocabulary that holds the pertinent information regardless its source. Doing so enables an improved understanding of the impacted system and improves the decision support capabilities.

The collected data provide the context to understand the situation that takes place, but beyond their face value, the meaning of the data might describe the situation even better. In order to tap into the deeper knowledge of the available data, it is necessary to understand the meaning they have in the context of the studied field. In order to do this, some background domain knowledge is necessary that allows for understanding of the meaning of the data, given the context. The meaning of the data is also called their semantics. In order to utilize said semantics in a machine processable way, it is required to define their representation in a uniform way. Moreover, the combination of semantics from different sources is important, because heterogeneous data cover a greater range of knowledge (this process is also called semantic enrichment). Finally, in order for the machine to actually provide some useful insights from the data and enhance the decision support, semantic reasoning is employed. Semantic reasoning is the process of inferring hidden knowledge based on explicit data, and according to the meaning of the data. The goal is to detect such knowledge that derives implicitly from existing data and is useful towards a better understanding of the system and the overall situation. The implementation of semantic representation, enrichment and reasoning is supported by tools developed for the semantic web.

Semantic web is the *web of data that can be processed directly and indirectly by machines* (Berners-Lee, 2001) and has multiple technologies for semantic enrichment and representation of data. In particular, ontologies are a powerful in representing semantic knowledge, and they promote interoperability and semantic reasoning. Ontologies are the specification of a vocabulary for semantically representing a shared domain of discourse (Gruber, 1993). In other words, an ontology can semantically model domain knowledge via the means of defining a set of classes (i.e. objects, concepts) of the domain and their properties (i.e. interconnections, attributes). Overall, defining the domain of interest in terms of ontologies according to semantic web technologies, allows for semantic representation of the available data in a standardized form and enables semantic reasoning capabilities for enhanced decision support.

In the context of the aqua3S project, semantic web technologies are a great asset towards data modelling and their enhancement with semantic capabilities. In particular, the definition of the aqua3S ontology allows for the adoption of existing standardized resources, which allows for both easier data consumption and sharing. Moreover, the ontology itself being developed in accordance with semantic web technologies can be easily shared and imported in other applications. Additionally, the adoption of semantic web technologies provides the means to automatically process the data. This includes the consumption of newly incoming data from aqua3S components and the internal reasoning mechanisms that aid decision support.

The remainder of this document is structured as follows: Section 3 presents the User Requirements that are pertinent to the development of the ontology, along with a catalogue of functionalities we identified as crucial for the overall usefulness of the ontology. Section 4 presents the State of the Art first for semantic representation and reasoning, and secondarily the most prominent ontologies and resources for the particular aspects of the ontology and data categories. In section 5, the developed ontology is presented, along with graphical representations of the most important aspects. Section 6 presents the mechanisms that enable decision the decision support functionalities of the component, along with some examples. Section 7 recapitulates the document and notes the next steps of the advancement of the component.

3. Modeling Requirements

The development of the ontology necessitates good understanding of the modeled system and the range/extent it is called to cover. In this section, a set of competency questions are defined in order to guide the ontology development process, and the relevant User Requirements are presented. Doing so, the relevant entities and the relationships between them are becoming more apparent. Overall, this section acts both as a guide for the development of the ontology and the corresponding semantic reasoning, and as means of validation.

3.1 Competency Questions

Competency questions are used as means to guide the development of the ontology in order to realize them as means of evaluations (Bezerra, Freitas, & Santana, 2013). They define queries that the ontology must be able to answer. The following list of competency questions was elicited through the User Requirements and discussions with technical partners. The user requirements were first defined in the deliverable *D2.1 – Use cases requirements V.1* but they have been updated since. The list is categorized by the major entities of interest.

Risks:

1. What are the risks of the system? (e.g. risks that might cause a crisis, if realized)
2. What is the likelihood of a risk?
3. What is the severity of a risk?

Crisis:

4. What is the type of crisis? (e.g. flood, mechanical failure, pollution)
5. What is the location of the crisis?
6. When has the crisis broken out?
7. What is impacted by the crisis?
8. What are the identified causes of the crisis?
9. What are the identified abnormalities within the crisis? (Incidents)

Incidents (deviation of normal operating conditions):

10. What is the location of the incident?
11. When was the incident detected?
12. What entities does the incident impact?
13. How were the incident detected? (what are the data that detected its existence)
14. What analysis task detected the incident?
15. What sensor measurement detected the incident?
16. Who are the first responder agents that are assigned to one incident?
17. What operational force does one first responder agent belong to?

Sensors and their measurements:

18. What is the location of a sensor?
19. What property does the sensor measures? (e.g. smell, ammonium concentration)
20. What is the basic category of a property? (acceptability/chemical/microbial property)
21. What entity within the system does the property belong to? (e.g. drinking water, sewage water)
22. What measurements has the sensor produced?
23. What is the value of a measurement?
24. What is the unit of a measurement?
25. When were the measurements made?
26. What are the measurements that are above the allowed limits?
27. What are all the sensor measurements that refer to a specific system entity?

Social media posts (mainly Twitter):

28. Who is the user that posted a social media post?
29. What is the first/last name of the user that posted something? (if available)
30. What is the username of the user that posted something? (if available)
31. What else has a user posted?
32. What is the text of a social media post?
33. What is the title of a social media post?
34. When was the social media post created?
35. In what language was it written?
36. What is the URL of the original post? (if available)
37. What is the subject of the post?
38. Where were the user when they made the post? (geolocation of the tweet, if available)
39. What are the referenced locations of the post?
40. How do other users interact with the post? (types of interactions, e.g. retweet, like, reply)
41. How many interactions does the post have?
42. What are the collections of posts (groupings of posts) that were produced by the social media analysis task?
43. What posts belong to a particular collection?
44. What does the collection evaluate? (i.e. what property)
45. What entity of the system does the collection refer to? (e.g. drinking water, sewage)
46. What is the sentiment or polarity of one evaluation?
47. What is the reliability score of one evaluation?

Geospatial and temporal specification:

48. Is the location an individual point?
49. Is the location an area defined by its boundary?
50. Which points from a set are within the boundaries of an area?
51. Do some areas overlap?
52. Between two-time instants, which one is the earlier?
53. Between two-time intervals, what is the overlapping period?
54. What measurements are made within an area of interest?
55. What social media posts that have as subject the drinking water were made after a point in time?
56. What incidents were detected during a time period?

3.2 User Requirements

This section presents the user requirements that involve the aqua3S ontology or the semantic reasoning support component. The user requirements drive the development of the ontology and the semantic reasoning and act as the means of validating their effectiveness. The following table shows for each user requirement the relevance of the ontology or semantic reasoning component, and how they will contribute towards meeting the requirements.

Use Requirement ID	Title	Description	Relevancy
UR_104	Provide warnings related to the sensor	Provide automatic warnings (e.g. through mail) related to the overtopping of some predefined alert thresholds in the measurements by the qualitative and quantitative sensors in the supply network, in the rivers that contribute to the water supply in Trieste and in the groundwater wells	Produce alerts based on predefined thresholds. The alerts will be forwarded to the context broker.
UR_105	Provide warning related to modeled scenarios	Provide automatic warnings (e.g. trough mail) related to specific scenarios that come from the models (i.e. extreme droughts, floods etc.)	Produce alerts based on semantic reasoning capabilities that will capture specific scenarios. The alerts will be forwarded to the central communication bus.
UR_107	Provide suitable metrics for the sensors and models data in case of predefined threshold overtopping	If a certain data (from models or sensors) has one or more thresholds associated and in case of threshold overtopping, provide some specific suitable metrics that summarizes the situation	The produced alerts on the overtopping of sensors will have included the relevant metrics that caused the alert.
UR_116	Report generation	After an event / crisis a report should be created including the data related to the event that lead to taking a specific decision	Provide the sum of semantic reasoning outputs within the crisis period. Initiated by report generator event received via the context broker.
UR_204	Provide warnings related to the sensor	Provide automatic warnings related to the overtopping of some predefined alert thresholds in the measurements by the sensors	see UR_104
UR_209	Backup/ Data archiving	Data format and data sets must be saved to allow an easy and fast retrieval for real time and historical data.	see UR_115
UR_210	Report generation	After an event / crisis a report should be created including the data related to the event that lead to taking a specific decision	see UR_116
UR_404	Visualize warnings on a dynamic GIS interface	Display on a GIS interface all information available. Real time Interactive map	Semantic reasoning output will be accompanied with geospatial information if available.

UR_406	Backup/ Data archiving	All data (raw sensor data, satellite images, social media posts and call centre call must be processed and archived for future reference and processing.	see UR_115
UR_407	Report generation	Report generation related to the water quality and the reported complains	see UR_116
UR_506	Backup/ Data archiving	Data format and data sets must be saved to allow an easy and fast retrieval for real time and historical data.	see UR_115
UR_708	Provide warnings related to the sensor for residual chlorine	Monitoring and sending alerts for water turbidity, for the levels of residual chlorine	Produce alerts when chlorine and turbidity is elevated.

Table 1. User Requirements relevant to the ontology and semantic reasoning

4. State of the Art

This section presents the State of the Art pertinent to the aqua3S ontology and the semantic reasoning it enables. In particular, this section presents the following: (1) the mechanism that enables semantic reasoning, (2) the languages that realize this mechanism, (3) relevant ontologies and resources that were considered for the development of the ontology.

4.1 Description Logics

The term Description Logics (DLs) (Baader, 2003) refers to the group of knowledge representation languages that implement well defined, logic-based languages that support reasoning services. The main building blocks of the DLs are the *concepts* that represent real world entities and roles that represent connections and relationships between the concepts. For example, a concept can refer to the species of *Dogs*, there can be specific instances of the Dogs concept (e.g. Max, being a specific dog), and also can have relationships between them (e.g. Max *has parent* Rex). Atomic concepts (or concept names) refer to base concepts (e.g. Dog) but restrictions on the atomic concepts are also supported. In detail, restrictions refer to the instances of an atomic concept that adheres to a specific constraint, for example, $\exists \texttt{hasParent.Dog}$ refers to the instances of the Dog concept that are related though the "has a Parent" role, in other words, it refers to the entities that are parents of a dog.

The knowledge repositories that are DL compliant, typically are defined via the TBox and ABox sets of statements. TBox statements (Terminological knowledge) describe the concepts and relationships they can have and defines the how the concepts can be associated. For example, $\texttt{Puppy} \sqsubseteq \texttt{Dog}$ defines that all instances of the concept *Puppy* are also instances of the concept Dog, all puppies are dogs. On the other hand, ABox statements (Asertional knowledge) use the TBox knowledge in order to define more complex real-world relationships of the entities, also called axioms. For example, the combination of the statements `Puppy (Max)` and `hasGender(Max, male)` shows that the Max the puppy is male. Table 2 presents the axioms used in Abox and TBox statements.

Axiom	Syntax	Semantics
Concept inclusion	$C \sqsubseteq D$	$C^I \subseteq D^I$
Concept equality	$C \equiv D$	$C^I = D^I$
Role Equality	$R \equiv S$	$R^I = S^I$
Role inclusion	$R \sqsubseteq S$	$R^I \subseteq S^I$
Concept assertion	$C(\alpha)$	$\alpha^I \in C^I$
Role assertion	$R(\alpha,b)$	$(\alpha^I, b^I) \in R^I$

Table 2. Terminological and assertional axioms

Beyond the basic knowledge statements, the DLs have potent reasoning services that check for consistency and make inferences based on the defined axioms. Such reasoning engines are remarkably efficient and reliable. Some implementations include Pellet (Sirin, Parsia, Grau, Kalyanpur, & Katz, 2007), HermiT (Shearer, Motik, & Horrocks, 2008), FaCT++ (Tsarkov & Horrocks, 2006) and RacerPro (Haarslev, Hidde, Möller, & Wessel, 2012). Some typical reasoning operation are presented in Table 3, given the Knowledge base implements a DL ($K = (T, A)$).

Reasoning operation	Definition
Subsumption	A concept C is subsumed by D in T (equivalently $\mathrm{T} \vDash C \sqsubseteq D$) iff $C^I \subseteq D^I$ for any interpretation I
Equivalence	Two concepts C and D are equivalent in T (written $T \vDash C \equiv D$) iff $C^I \subseteq D^I$ and $D^I \subseteq C^I$ for all interpretations I
Disjointness	A concept C is disjoint to a concept D iff $C^I \cap \mathrm{D}^{\mathrm{I}} \neq \varnothing$ in every interpretation I
Consistency	An ABox A is consistent with respect to T iff there is an interpretation that is a model of both A and T
Instance checking	An individual a is an instance of C with respect to K (written $K \vDash \mathrm{C(a)}$) iff $a^I \in C^I$ is true for all interpretations I of K
Realisation	The realization of an instance a with respect to K includes finding the most specific concepts C for which $a^I \in C^I$ is true for all interpretations I of K

Table 3. DL reasoning operations

Intuitively, the **subsumption** operation defines if some instances of a concept belong to subclass or subcategories of the original concept (D). Subsumption is crucial in the definition of taxonomy between concepts. For example, if the axiom $Parent \equiv \mathrm{Person} \sqcap \exists\, \mathrm{hasChild.Person}$ holds, then it is inferred that the concept Parent subsumes the concept Person, or is a particular subcategory of a Person. **Equivalence** is the operation that checks that some instances refer to the same entity. **Disjointness** is the property that defines that if two concepts are disjoint, the instances of the one concept cannot be instances of the other concept. **Consistency** entails that the axioms defined in an ABox must be sound with respect to the TBox axioms. For example, if the concepts Cat and Dog are disjoint concepts, then the existence of the statements $Cat(Max)$ and $Dog(Max)$ leads to inconsistency, as the same instance cannot belong to both of the disjoint concepts. **Instance checking** refers to the control of whether an instance belongs to a particular concept. Instances can belong to a single concept or to multiple.

One important assumption that most reasoning engines follow, is the open-world assumption. It refers to the assumption that the presence of some information does not imply that only the known information is true. Sometimes this assumption fuels the tendency of the reasoning engine to justify everything, and it is most apparent in loosely defined models. For example, the statement $\boldsymbol{hasChild(Rex, Max)}$ does not conclude that Max is the only child of Rex. On the contrary, the closed world assumption implies the opposite, if something is not explicitly defined, then is assumed to be false. Open world or closed world assumptions dictate how the semantic reasoning needs to be implemented.

4.2 Web Ontology Language

The Web Ontology Language or OWL (McGuinness, Van Harmelen, & others, 2004) is a language for knowledge representation and the authoring of ontologies. OWL is endorsed by the World Wide Web Consortium (W3C) and is the most widely used language for ontology development. The W3C is the main international standards organization for the World Wide Web. Ontologies are the formal specification of objects and the relationships between them. Ontologies are designed so that they promote combinations between different ontologies and their expansion. This design encourages the reuse of existing ontologies and consequently the homogeneity of semantic representations across different applications. OWL has three sublanguage variants that have different levels of expressiveness. With ascending order of expressiveness, the sublanguages are OWL Lite, OWL DL and OWL Full, each one is more expressive than the previous. However, the expressivity of the language inhibits the decidability, which in turn renders the implementations of the language and semantic reasoning engines more difficult. For this reason, OWL Lite and OWL DL are used in practice as they are more decidable, and especially OWL DL strikes the best balance between decidability and expressivity. Even so, the OWL DL's expressibility and the "tree model property" (Vardi, 1997) style it adopts often was deemed insufficient in common use cases. In particular, only tree-like hierarchies are supported in OWL DL.

In order to resolve the limitations of OWL Lite and OWL DL, the next iteration of the language was developed, called OWL 2 (Hitzler, et al., 2009). It follows the same principals of the original languages but with some additions and some redesigns. Cardinality restrictions were added on OWL 2 and they enable quantitative relationships and axioms. For example, a family can be defined as a large family if they have at least 5 members, or $LargeFamily \equiv Family \sqcap \geq 5\ hasMember.Person$. Another advance made with OWL 2 is the introduction of a set of datatypes native to OWL that cover more of the typical applications.

4.3 Spatial information resources

Basic Geo (WGS84 lat/long) Vocabulary[1] (Brickley, 2006): Is a lean vocabulary, suitable for representing the spatial position of single entities. The position is represented according to the World Geodesic System (revision 1984) which is a well-established OGC standard for spatial reference system with worldwide applications. Notably it is used in cartography and satellite navigation, including Global Positioning System (GPS). The vocabulary itself supports the spatial representation of single points according to the WGS84 as a function of their latitude, longitude, and altitude if necessary. Moreover, it includes the generalizable class *SpatialThing* that represents anything with spatial extent, which can be used to describe complex shapes and areas as a combination of individual points. However, the vocabulary does not support natively commonly used shapes (e.g. square, circle, polygon), thus making it less than optimal for representing areas in real world scenarios.

GEO OWL[2] (Lieberman, Singh, & Goad, 2007): Is a proposed extension of the **Basic Geo Vocabulary** by the W3C incubator group that aims to introduce predefined shapes and geospatial entities. In particular, rectangles, points, lines, and polygons are the proposed spatial entities of the extension. The group proposed the new entities based on the OpenGIS Simple Features[3] (OGC implementation specification) and reused entities from the GeoRSS[4] (OGC standard). Overall, this extension proposes enriched catalogue of possible spatial objects that covers most common use cases, thus, enabling a more direct and easier modelling.

[1] https://www.w3.org/2003/01/geo/
[2] https://www.w3.org/2005/Incubator/geo/XGR-geo/
[3] https://www.ogc.org/standards/sfs
[4] https://www.ogc.org/standards/georss

GeoSPARQL[5] (Perry & Herring): Is an OGC standard vocabulary distributed in OWL and RDF suitable for representing and querying geospatial information. The geospatial representation makes use of former and well-established standards and implementation (Geography Markup Language[6] and Well-known text[7]). The vocabulary supports geometrical primitives (Point, LineString and Polygon) and multiple extensions and specifications that cover most use cases. In addition to the richer representation of geographical entities, this vocabulary also introduces a set of qualitative and quantitative relationships between them. Finally, the ontology is designed so that it can be queried via the SPARQL query language and some RDF databases (e.g. RDF4J, GraphDB, Strabon) have already implemented the defined relationships.

4.4 Temporal information resources

Time Ontology in OWL[8] (Hobbs & Pan, 2006): Is an OGC and W3C standard for expressing temporal properties and their relationships. In particular, the main entities are the Instant and the Interval, where the first is defined as a particular point in time, while the second represents a period of time with a particular start, end, and duration. Additionally, the time intervals have temporal relationships between them as defined by the standard interval calculus and shown by Figure 1. The adopted encoding of date/time representation follows the ISO 8601 (e.g. “2020-08-26T13:32:58+00:00”).

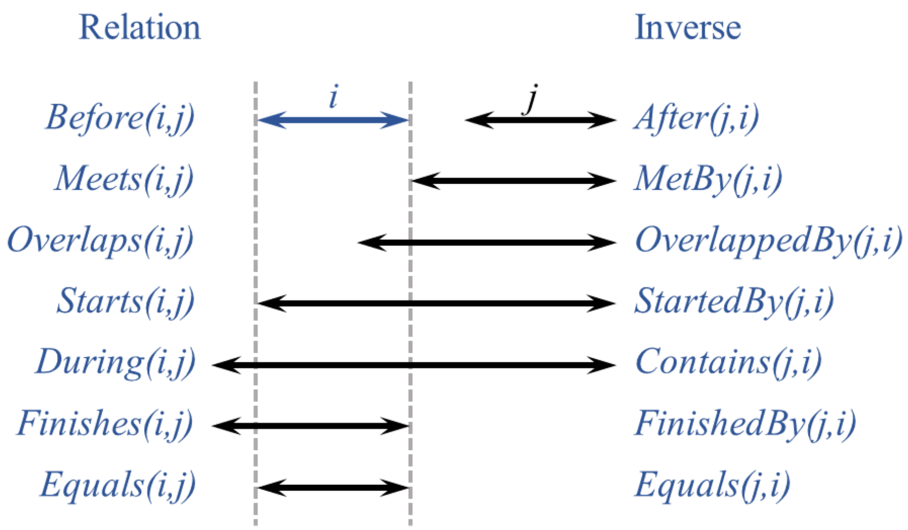

Figure 1. Time interval relationships

Timeline ontology[9] (Raimond, Abdallah, Sandler, & Giasson, 2007): Is an ontology that emphasizes the distinction of individual timelines that organize relevant temporal events together. It supports time instants and intervals that are placed onto a timeline either in absolute terms (timestamp) or in relational terms (before, after, etc.). Temporal entities can be placed in particular timelines, or in one

[5] https://www.ogc.org/standards/geosparql
[6] https://www.ogc.org/standards/gml
[7] https://www.ogc.org/standards/wkt-crs
[8] https://www.w3.org/TR/owl-time/
[9] http://motools.sourceforge.net/timeline/timeline.html

timeline that contains everything (global timeline). Timelines are used to group relevant events together, providing a more organized facet of the known temporal events. Additionally, timelines can be associated with other timelines so that the temporal relationships between events of different timelines can be compared.

Reusable Time Ontology (Zhou & Fikes, 2002): Is another work that tackles the issue of temporal representation in a cohesive and interoperable manner. It models both time instants and time intervals that are the main focus, as they offer more complexity. In this case the focus is on the fundamentals of interval types. The intervals are categorized as open or closed, and convex or non-convex. Openness of an interval depends on whether the extremities are included in the interval or not (typical notation of closed and open intervals: [a, b], (a, b)). Non-convex time intervals are those that are not connected and noncontinuous, while the convex are continuous periods of time. For example, non-convex time interval could refer to “every Tuesday in September”, and convex time intervals are particular weeks, months, years, etc. Additionally, it adheres to the requirements of Allen relations, thus, complex relationships between intervals are possible. Finally, the ontology has predefined classes that represent time in different granularities so that it covers a wide range of potential applications ranging from microseconds to years.

4.5 Crisis ontologies and resources

beAWARE ontology (Kontopoulos, et al., 2018): The ontology that were used for the Horizon 2020 project beAWARE that handled heterogeneous data in climate crisis situations in order to provide decision support. In particular, the ontology is directed towards natural climate crises and disasters, for example flood, earthquake, forest fire etc. The ontology supports data both from sensors and from human agents. Typical sensors for such use cases are photo/video cameras, temperature sensors, location sensors etc. that can shed light on the condition of the crisis. Additionally, humans can also give important information about a crisis and its effects, thus, input from humans was also included in the ontology (mainly via social media). Moreover, the ontology has some first responder assignment capabilities along with operational missions in order to aid the management of the crisis. Finally, the ontology breaks down the crisis into particular incidents for more fine-grained management. One incident has attached the relevant sensor measurements, social media feedback along with their respected analyses.

ISyCri ontology (Bénaben, 2008): This ontology was used on the ISyCri project and provides a general crisis management ontology that is suitable for a wide range of crises (e.g. technical, political, legal, natural, etc.). The focus is on the description of the crisis and the related entities in a somewhat abstract manner, considering mainly the higher-level information about the ontology. In particular, the crisis is characterized by the gravity factors (conditions that may change the severity of the crisis), and the complexity factors (conditions that may change the type of the crisis). For example, in case of forest fire, high wind speed is a condition that may increase the severity of the situation (gravity factor), while the existence of wildlife in a polluted lake may change the crisis to an environmental disaster (complexity factor). Moreover, the ontology introduces risks as an ambient factor of the system that is always present and can cause or exacerbate a crisis. Regarding the effects of the crisis, ISyCri ontology categorizes the possible affected entities in four categories (People, Natural sites, Civilian Society and Goods) with the possibility to expand them in particular instances.

The ISyCri ontology provides the means of describing a higher level of the crisis but lacks in expressing sub-instances within the crisis. In particular, does not offer more fine-grained facets for organizing smaller incidents of interest within the whole crisis. A second shortcoming of the ontology is the absence of connection between the crisis and the relevant data. Finally, the First Responder (FR) management is on a high level in terms of services actors.

EN 15975 Security of drinking water supply – Guidelines for risk and crisis management (EN-15975, 2013): Is a European standard approved by CEN followed by European nations for the management of their drinking water supplies. It focuses on the procedures necessary to ensure the safety of crisis management for risk mitigation. However, in doing so it defines the relevant terms in an ontology-like manner. The ISO defines the issues mainly in terms of Crisis and the associated Hazards and Risks and notes the importance of preparing for all potential types of crises (natural, technical, malicious etc.). In order for effective handling of potential crises, common understanding between the cooperating parties is crucial, in order first to assess quickly and effectively the situation, and then in order to communicate the situation to other potential involved parties (typically First Responders in case of emergency). Prevention of potential harmful situation is the real goal in risk management, thus, preparing for individual incidents or combinations that may cause the crisis also important to be identified. The ISO promotes the definition of procedures that work towards risk mitigation, in all time frames; before, during and after the crisis (preparation, operative and follow-up stages respectively). Overall, the ISO stretches the proactive stance that is necessary, and the clear definition of potential crisis in order for effective mitigation and communication to take place.

The Hazardous Situation Ontology Design Pattern (Lawrynowicz & Lawniczak, 2015): Is a design pattern in formalized in OWL that aims to model Hazardous Situations and was used in the EU project Musing[10]. The term Hazardous Situation is closely related to the term Crisis, in particular is defined as a state that some entities are exposed to hazard, thus, they are potentially subject to harm. The Design Pattern emphasizes on causality, consequences, and affected entities (see Figure 2). As a design pattern, aims to define the basis of a risk management system in an application agnostic manner and it does not go in much detail. The pattern is associated with other widely used ontologies (DOLCE, ESO) in order to facilitate interoperability, and it considers the ISO 31000 about risk management.

[10] https://cordis.europa.eu/project/id/027097

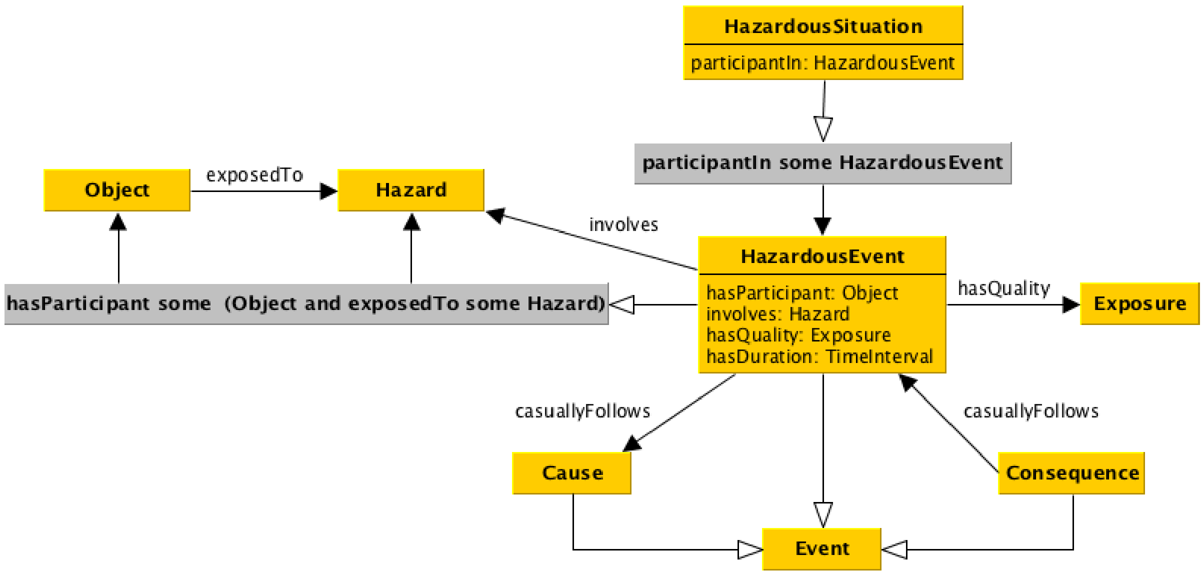

Figure 2. The Hazardous Situation Ontology Design Pattern

4.6 Sensors and IoT devices

SSN/SOSA ontologies (Janowicz, Haller, Cox, Le Phuoc, & Lefrançois, 2019)**:** SSN is a joint OWL2 ontology of W3C and OGC that models sensors along with their characteristics, observations, procedures, features of interest and more. SOSA is the core ontology of SSN that contains the information for Sensors, Observations, Samples and Actuators, and is self-contained. For example, a turbidity sensor can be characterized by its location in a water pipe, its units of measurement, the particular turbidity readings, the feature of interest (the water inside the pipe), and additionally, the actions that the sensor might incite. Overall, the ontologies provide a basis for representing sensors, their readings and the corresponding actions, in order to improve semantic interoperability and integration.

Smart Appliances REFerence (SAREF) (Daniele, den Hartog, & Roes, 2015): Is an ontology that was created with the support of European commission that aims to promote Internet of Things in the context of smart appliances and devices. The core ontology (SAREF) models smart appliances and devices along with their functionalities and the transmitted commands. Moreover, sensors are also included in the core ontology as a subcategory of devices, and they perform measurements of relevant features of interest (e.g. water). For geospatial representation, the GeoSPARQL vocabulary is reused and the entities that require it they inherit from the geo:Feature class. Overall, core ontology includes the expressibility necessary for a wide range of smart devices without focusing on particular instances, but the SAREF project is continuously expanded to cover more use cases and different domains.

SAREF4WATR[11] is a recent SAREF extension that specializes for the domain of water management. In particular, it introduces classes and hierarchies that are highly relevant to the domain and expands on

[11] https://saref.etsi.org/saref4watr/v1.1.1/

procedures that are typical of the use case. For example, the Feature of Interest class was subdivided into particular water forms (raw, storm, waste, and drinking water) that correspond to usual types of water management. Another important addition is the categorization of potential water properties that sensors measure, assigned to three major classes of water properties, acceptability, chemical and microbial properties. These three categories include the main concerns of water quality in a water management situation. Additionally, the S4W ontology includes numerous instances of relevant water properties that belong to the main three categories. Other additions include a water tariff system that could be utilized by water distribution services and the introduction of water assets and infrastructure entities.

4.7 Social Media

Friend of a Friend – FOAF[12] (Brickley & Miller, FOAF vocabulary specification 0.91, 2007): Is a staple ontology for describing persons, their activities and their relations to other people and object. In the context of Social Networks (SN), the ontology includes arguably the most important classes (Person, Online Account, Document/Creative Work). However, the ontology captures the entities in an abstract way that is not sufficient for detailed representation of SN data, as it fails to represent the intricate relationships and hierarchies. Overall, FOAF is a widely used ontology that can function well as a basis for the SN representation.

Semantically Interlinked Online Communities – SIOC[13] (Bojars, et al., 2010): This ontology reuses FOAF and expands more on collections of User Accounts (that correspond to Social Networks) and on layouts and contributions on web resources. The ontology provides more detailed facet of the User Accounts as it links avatars, specific roles (e.g. owner, admin) and interactions (e.g. subscriptions). Communities are groups of User Accounts that might have various roles within the community. Moreover, the SIOC ontology includes the class Post that represents the online post of User Accounts and might be any of numerous modalities. Additionally, the ontology focuses on the representation of different types of online Containers that might host the activity of users, for example, Forums, and conversation Threads.

[12] http://www.foaf-project.org/
[13] http://rdfs.org/sioc/spec/

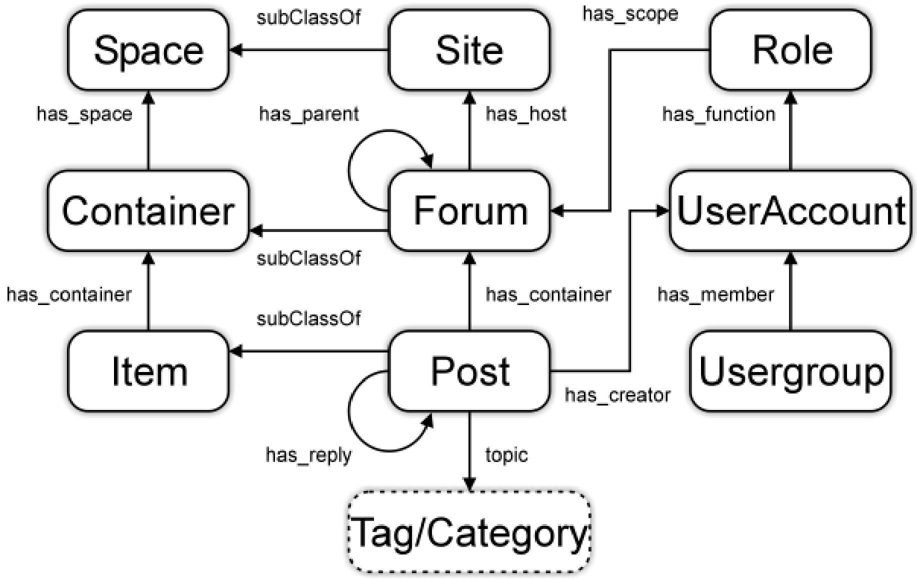

Figure 3. The main entities of the SIOC ontology and their interactions

SocIoS: A Social Media Application Ontology (Tserpes, et al., 2012): This ontology aims to device a unified Social Network model that is directly mapped to the most prominent social networks APIs (OpenSocial, Facebook, Twitter, FlickR and Youtube). It provides associations with other notable ontologies (FOAF, SIOC, LODE) in order to promote interoperability. The ontology itself includes different facets of Social Media: Users, their Activity, Groups, Reliability, Multimedia Items, Events, Locations and Messages. Overall, it models the basic entities of SNs in a comprehensive manner, while it retains consistency between different SN APIs.

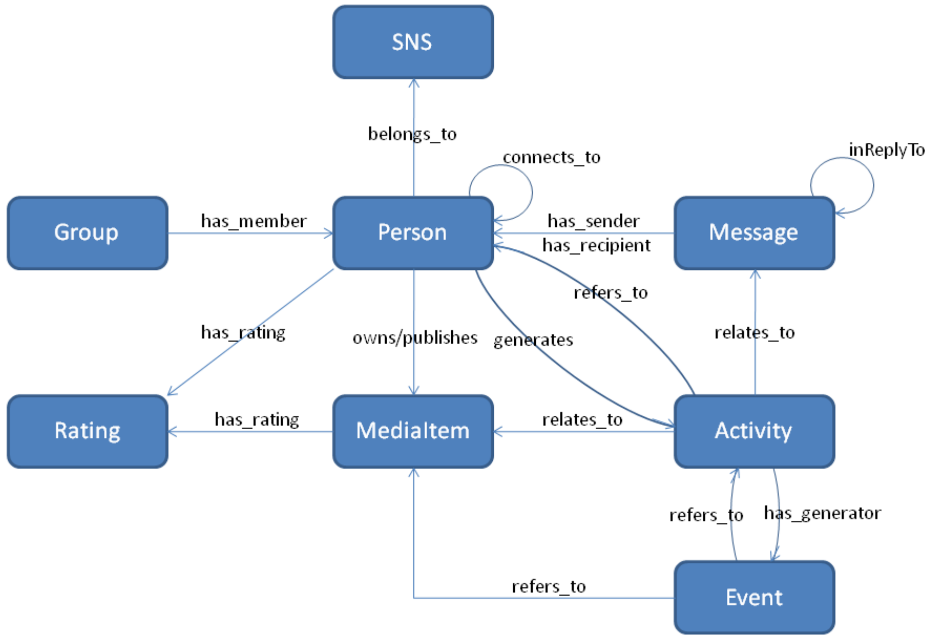

Figure 4. The main entities of the Soclos ontology and their relationships

5. Ontology conceptualization

5.1 The core ontology for crises

The core ontology represents the crisis itself along with some high-level crisis information, such as causes and risks. It is important for the decision support processes because the characteristics of the crisis affect everything that happens within its boundaries. The crisis was modeled based on the definitions of the EN 15975 (EN-15975, 2013), as it was determined by members of the consortium as appropriate basis for the aqua3S project. The ISO studies the risk and crisis management of water supply systems in a time horizon that covers the pre-crisis, during and post-crisis phases. Each of the three phases have different goals, in particular, the pro-crisis phase (preparatory crisis management) aims to detect the risks and prepare mitigation mechanisms. During the crisis (operative crisis management), the aim is to reduce the negative impacts as soon and effectively as possible. Finally, after the crisis has passed (follow-up crisis management), the objective is to restore the operational order and to debrief the situation in order to improve the mitigation mechanism for future crises. Within the context of aqua3S the decision support focuses on the phase when the crisis is ongoing (operative crisis management) because it has the highest potential impact. A secondary focus is to provide support for debriefing purposes after the crisis event, although this is done by means of exporting the semantic reasoning that took place during the event.

In the rest of this section, the core of the crisis ontology is presented in order to form the basis that can be applied in numerous different crises of the water distribution domain. The goal of the core is to define an extensible model that can be specialized according to particular needs of different applications. In particular, the main entities (also referred to as classes according to the OWL2 terminology) and the relationships between them are presented, along with some graphical representations of the model.

Starting with the most important classes that define a crisis, as illustrated on Figure 5, the Crisis entity corresponds to the event or situation that has the potential to affect seriously a drinking water supply chain. A Hazard is a condition or agent in the water with the potential of causing harm to public health. One or more Hazards may trigger a Crisis, if the Hazards are not contained. The Hazards are classified into five main categories (Pollution, Geological, Climatic, Environmental and Technological) depending on their nature. The five categories are indicative some major different types, and they stand to be specialized further whenever meaningful distinctions are determined. For example, Pollution hazards can be subdivided into chemical and microbiological pollutions.

Hazardous Events are events that introduce Hazards (or hazardous situation) to the system, based on existing System Risks. Such risks exist inherently in the system and have the potential to be realized by an event and subsequently introduce a Hazard to the system. System Risks are characterized by their Likelihood (defines how likely the event is to occur, not necessarily given as a probability) and their Severity (defines the extent of the expected impact if the Risk comes to pass). For example, the risk of flood in a particular area might be exceptionally low (low likelihood) and the severity of the consequences exceedingly high. Overall, the importance of the risks is the product of the likelihood and the severity. However, the numerical representations of the likelihood and the severity lack the corresponding standards, so their product is not straightforward in many cases. Similarly, the EN 15975 does not define a specific formula that combines the likelihood and severity; thus, we elected to follow the same approach and specialize it according to specific needs in the future. Finally, having the likelihood and severity of a risk is useful in order to plan for them, and to adjust the decision support functionalities.

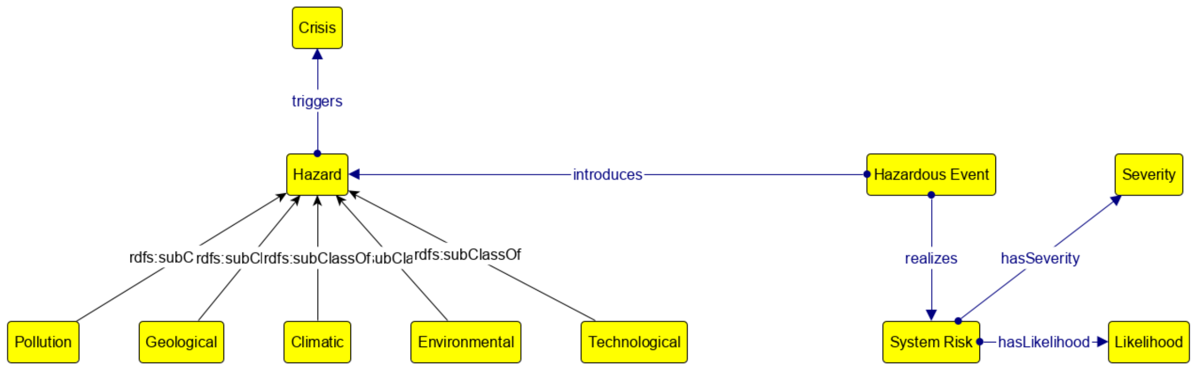

Figure 5. Hazards, Hazardous Events and System Risks

The crisis management aims to avoid or ameliorate the negative impact of a crisis on the associated assets. Thus, the inclusion of the components that are liable to be harmed is important in order to provide better understanding of the crisis. Illustrated in the Figure 6 the System Components are potentially impacted by the Crisis or by particular Hazardous Events within the crisis. They are subdivided primarily into four categories (Social, Resources, Environmental, and Economic), and subsequently can be further categorized. Finally, a Disaster is classified as a Crisis in case it has already caused widespread losses on some System Components.

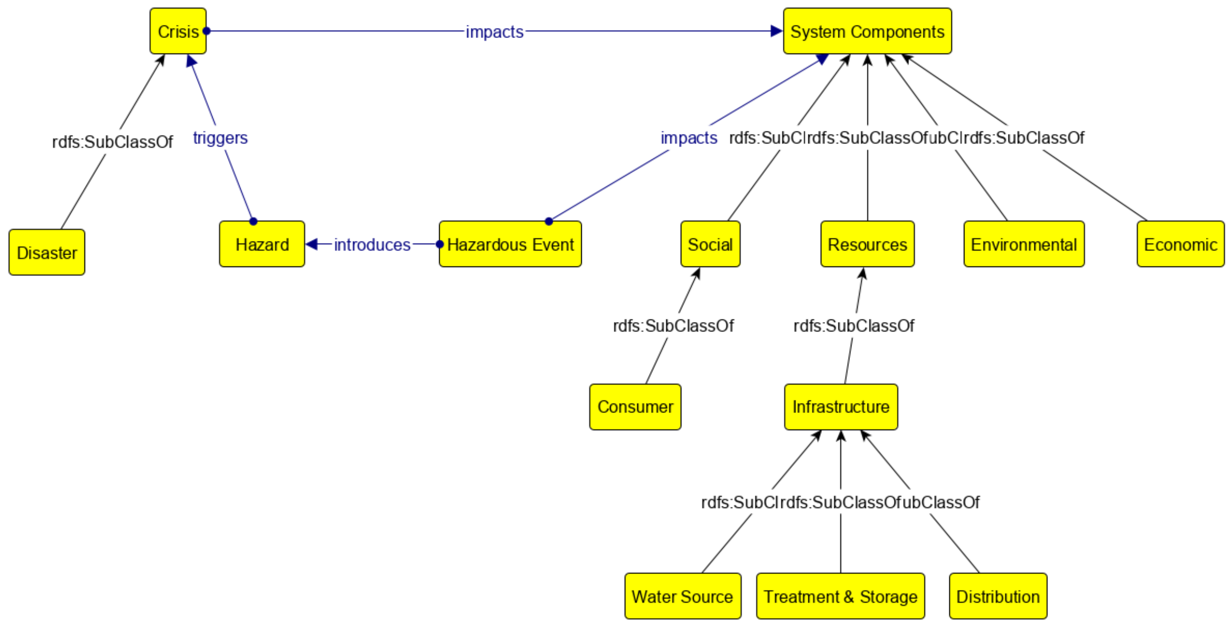

Figure 6. The impacted components of the system

Different types of crises will impact differently the system components; thus, it is meaningful to distinguish them. The category of the crisis is important for the specification of the situation, for example the first responders that need to be involved and the protocols of actions to be followed are heavily reliant to the characteristics of the crisis. Moreover, it might affect the semantic reasoning operations and the decision support specifications. The Figure 7 shows the first level of identified subclasses of crisis that are likely to be included in the aqua3S project and its use cases.

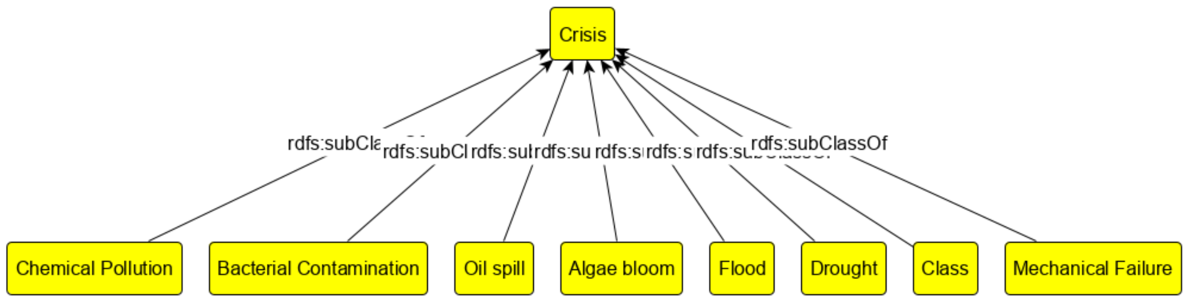

Figure 7. Subcategories of Crisis

In crises, often it is necessary to involve external forces in order to mitigate the losses. Such forces are typically first responders from various fields of expertise. In particular, the FR Agent class represents entities that belong to some FR Operational Force, and mainly are either FR Teams, or individual FRs (Figure 8). The FR Operational Force is used to categorize the expertise of the team and to provide better view of the Crisis. One important detail from the perspective of decision support is that the FR Agents are assigned to resolve particular actionable Incidents (including Hazardous Events) within the crisis, and not loosely assigned to the whole crisis. Finally, the core classes of the FR Agents with respect to their structure allows for the definition of teams (potentially with elaborate structure) that include (via the *'has member'* property) multiple individuals and other sub-teams. Further specialization can be achieved through subclasses of FR, depending on the role within their organization.

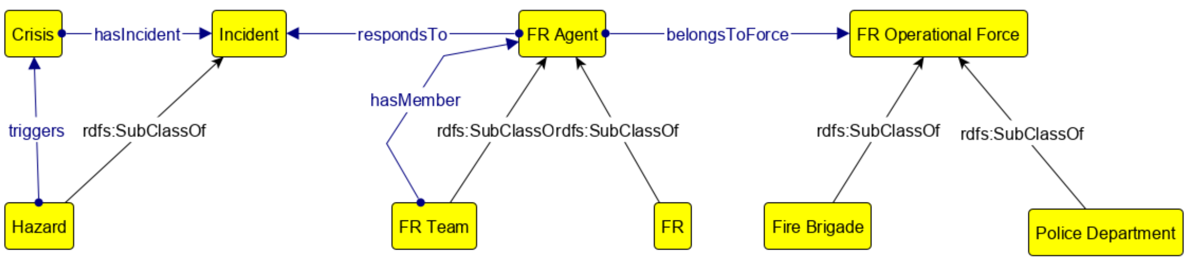

Figure 8. Relationships between First Responders and the crisis

Moving to a finer granularity of the crisis, the Incidents (Figure 9) are defined as "deviations from normal operating conditions" (EN-15975, 2013). They represent particular happenings that hold some interest for the specification of the crisis. Incidents may also be Hazards that are manifestations of underlying Risks (Hazard is a subclass of Incident).

Within the aqua3S project, it is important to have a connection between the data generators and the crisis. In particular, the objective is to connect the various abnormal sensor measurements and the detections from analysis components with the crisis itself. This is done via the incident class that denotes an irregularity within the crisis, which also might be a Hazardous Event. Doing so allows for the provenance of the irregularities to be transparent and available for further decision support. Additionally, the impact of the Incident is a secondary way to connect to the crisis, as it shows which System Components are impacted by the Incident.

Figure 9 shows the structure of the incidents that is adopted, and it is based on the beAWARE project that was in a similar but not equivalent domain to aqua3S. The class Incident is the blanket entity that correlates the abnormalities to the crisis, regardless of their provenance. For the spatial extend of the

Incident the GeoSPARQL Geometry class is used that allows for a variety of different spatial geometrical shapes. The temporal characteristics of the incident are defined according to the specification of the owl Time ontology. In relation to the analyses components the association is made through the Media Items that they produce. In particular, the Media Items are the results of analysis Tasks (e.g. object detection analysis and social media analysis) and they justify the existence of the Incident. The Media Item class can be subdivided into particular categories according to the nature of the Media, with the most prominent cases being Images, and Textual Entities. In regard to the sensor measurements, they do not pass through an analysis task, thus, the association with the incident is directly via the sensor measurements. The representation of the sensor measurements is adopted according to the SAREF ontology and its extension SAREF for Water (S4W) and is detailed in subsection 5.2.1.

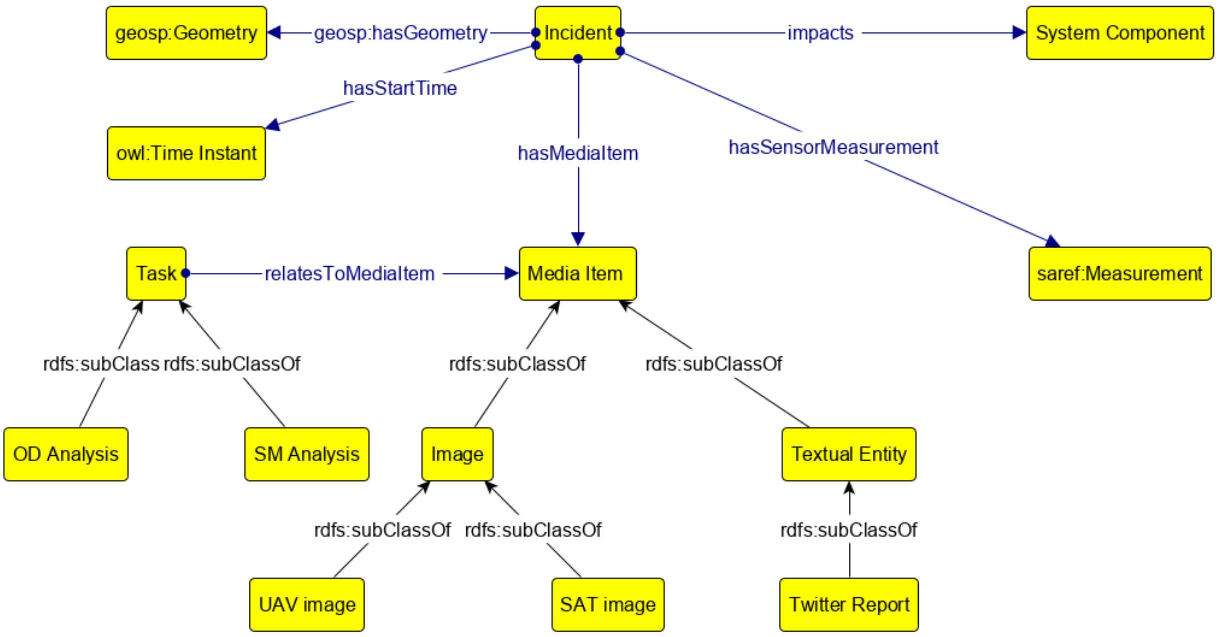

Figure 9. The Incident, Media Item and the Task classes

5.1.1 Geospatial Information

In order to represent spatial information, the GeoSPARQL ontology was elected, which is an OGC standard and complies with the spatial representation within the aqua3S project, including the data model used by the aqua3S context broker. GeoSPARQL consists of a lightweight ontology for representing spatial information and an ontology that defines the SPARQL interface that allows qualitative and quantitative relationships to be expressed via a SPARQL endpoint. The most prominent RDF stores (RDF4J and GraphDB) support natively the GeoSPARQL relationship schema, thus, allowing for complex spatial semantic reasoning. This representation is also used by the SAREF ontology, thus, achieving a uniform spatial representation within the semantic reasoning component.

In more detail, the GeoSPARQL ontology allows for representation of spatial entities in various modalities. The simplest representation are the Points (defined by their latitude and longitude) and they offer the basis for more complex shapes to be developed. Line Strings, Curves and Polygons are defined using set of coordinates and offer more complex spatial representation capabilities. The other important feature of the GeoSPARQL ontology is the support of complex relationships between spatial entities. In particular, supports region connection calculus (Figure 10) that includes relationships like disconnected, externally connected, and partially overlapping areas.

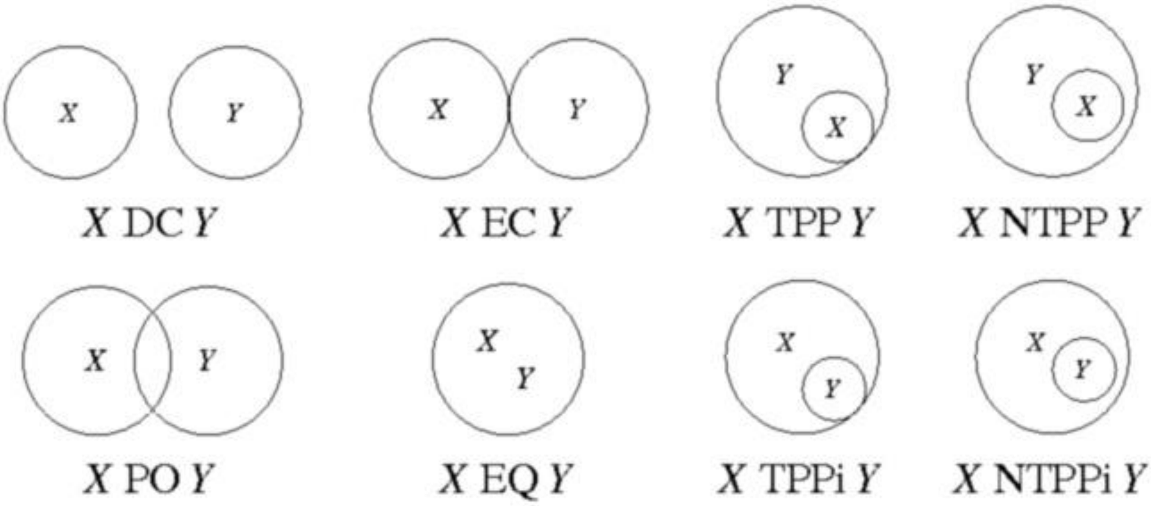

Figure 10. Region Connection Calculus[14]

The decision support functionalities of the aqua3S project heavily rely on spatial configurations in various forms. In order to utilize the functionalities of the GeoSPARQL ontology, all the entities that hold geospatial information inherent the GeoSPARQL Feature class (Figure 11). This class is connected with the GeoSPARQL Geometry class that allows the spatial specification. The most useful geometries within the aqua3S project are the Points and the Polygons, and secondarily the Line Strings. The Points are used to represent position of various entities (sensors, individuals, tweets etc.), while the Polygons represent specific areas (algae blooms, floods, etc.). Line Strings can be used to represent trajectories of entities, e.g. the movement of a First Responder.

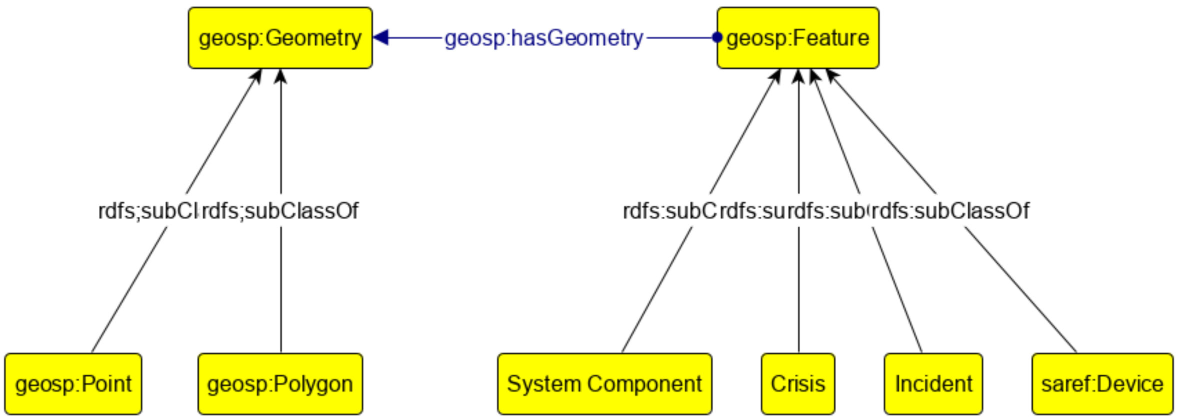

Figure 11. Geospatial representation of relevant classes

[14] https://en.wikipedia.org/wiki/Region_connection_calculus#

5.1.2 Temporal Info

For temporal information, the OWL Time ontology is used that is a W3C candidate recommendation, and a widely used ontology. It represents time in terms of time Instants and time Intervals. The timestamps within the aqua3S project follow the format of the ISO 8601 (Houston, 1988) for Date/Time Representations. An important set of features that the ontology offers is the relationships between time entities. Such relationships include but not limited in; before, after, and more complex such as overlapping, and contains. Overall, temporal relationships are crucial in the characterization of the crisis, for example, the temporal sequence of hazardous events and incidents affects the crisis in a major way.

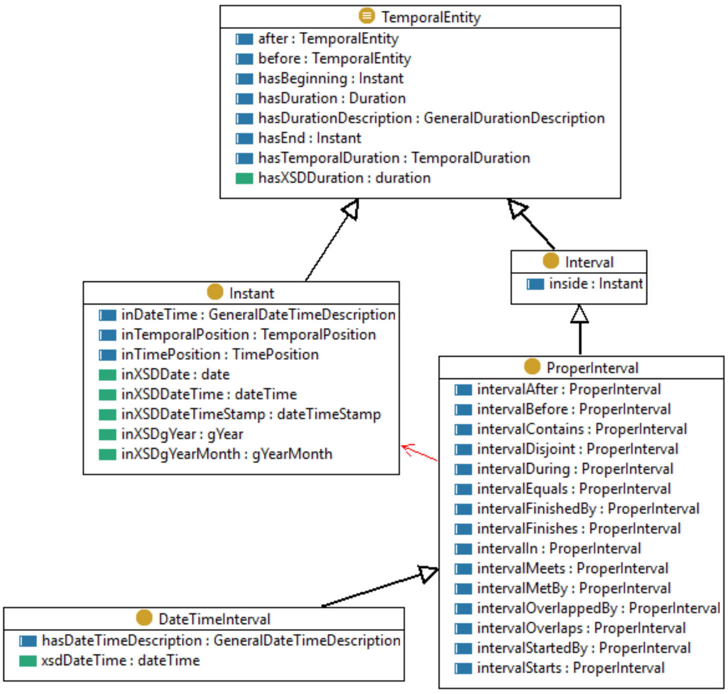

Figure 12. The main classes of the Time OWL ontology

5.2 Extended ontology

The core ontology lays the groundwork for the representation of information about the crisis along with temporal and spatial metadata. However, in order to fully utilize the data from different sources, some adjustment is necessary. In particular, the ontology needs to be specialized to specific use cases and the corresponding data. The varying nature of the data requires the storing of different information, thus, additions of the classes and properties of the ontology. This section presents the extensions of the ontology that are specific for the data from sensors and data from social media. In the future, the ontology will be extended and specialized even more according to the available data.

5.2.1 Sensors

The first extension of the core ontology is on the dimension of sensors that will provide raw data for the project. The data produced by the sensors are associated with the Incident class that were described on the previous section. Figure 13 shows the representation of the sensors according to the SAREF ontology. The main entity is the Device that is subdivided into two subclasses, the Sensor, and the Water Device. The Device produces some Measurements that are defined by their temporal specification, the numerical value of the measurement and the Unit of Measurement. The spatial specification is inherited by the location of the device itself and is modelled as GeoSPARQL Geography class (not pictured). In aqua3S, we reuse the SAREF4WATR extension of SAREF in order to ensure that our ontology is interoperable and is aligned with the commonly accepted conceptual models and vocabularies. With regard to SAREF4WATR ontology, we have made some updates to it including the capturing of thresholds for the values measured by the sensors. Moreover, a liaison is established with ETSI and based on the outcome of the evaluation realized during the pilots, it will be decided whether these updates are essential to the completion of the SAREF4WATR ontology.

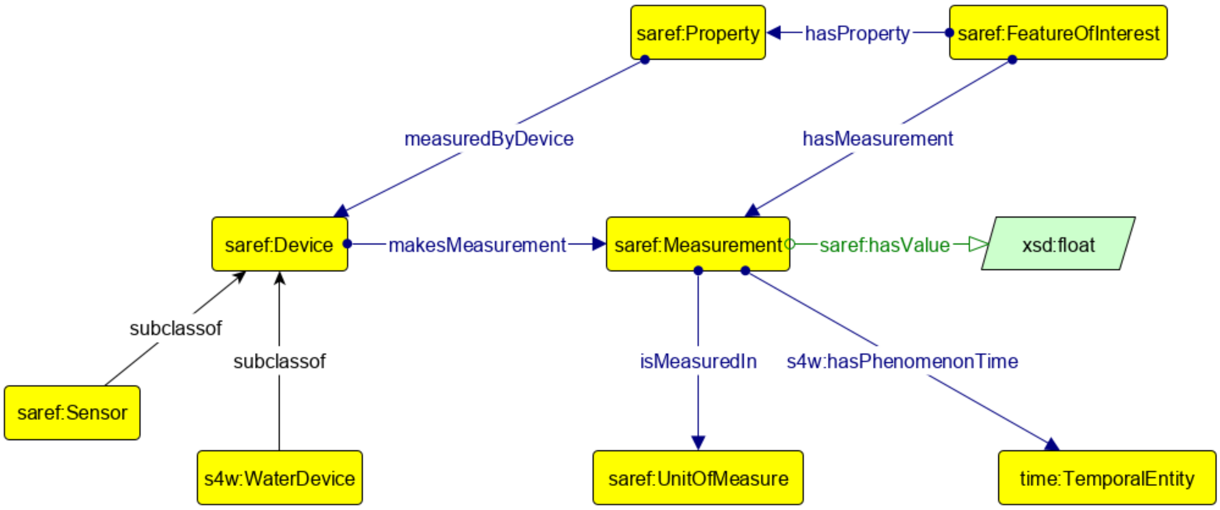

Figure 13. The extension for devices, sensors, and their measurements

Another important characteristic of the sensors and their measurements are the Property that is measured by the device, explaining why the measurement is relevant. As shown in Figure 14 the property is a characteristic of one feature of interest. In the context of aqua3S the feature of interest is the Water or more specifically, Raw Water, Waste Water and Drinking Water. The possible characteristics of the water are categorized into three classes, the Acceptability, Chemical and Microbial Properties. Acceptability properties refer to properties that inhibit the user experience and also might indicate some other more important issue. For example, abnormal water odour, taste and smell are acceptability properties. Chemical Properties examine the concentration of chemical substances, e.g. chlorine and hydrocarbons. Finally, the Microbial properties examine the concentration of microbes in the water, usually harmful ones like E. coli.

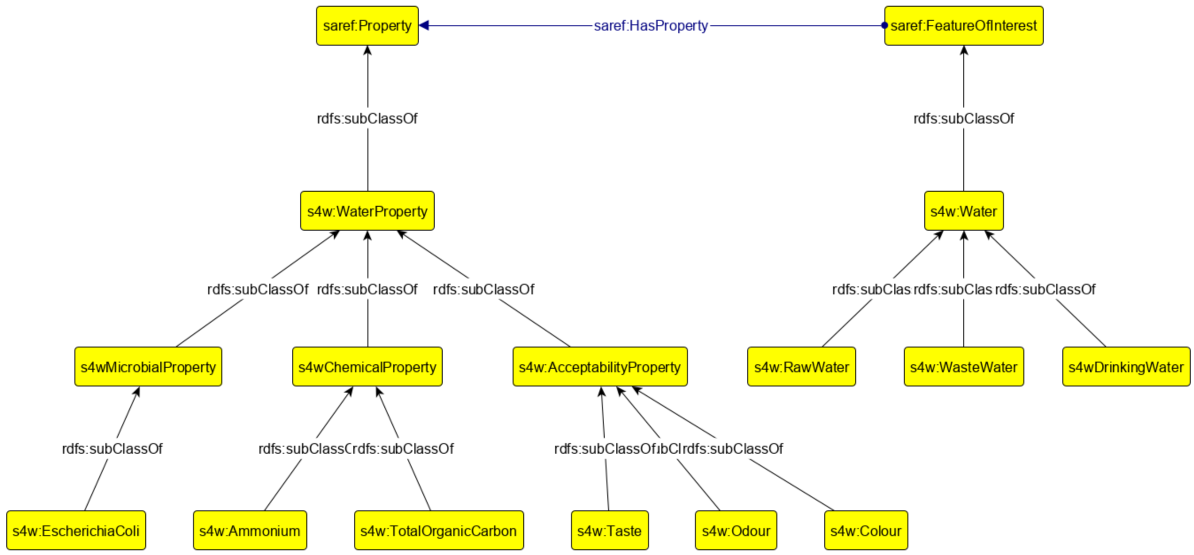

Figure 14. Properties and Features of Interest based on the S4WTR ontology

5.2.2 Social Media

Social media analysis is an important part of the aqua3S platform, as it studies the public perception of the water distribution system, and the water services. Especially with the advent of microblogging services (like Twitter), the monitoring of Social media can unveil abnormalities on the water distribution system in a timely fashion. Twitter was deemed the most suitable platform for such monitoring, thus, the ontology representation is focused on it, while providing a scheme that is extensible to other Social networks. In order to promote interoperability, the entities used for the representation of social media knowledge were mapped as closely as possible to the SIOC[15] ontology. SIOC is widely used and accepted ontology (W3C member submission), which aims to represent online users, their activity, and their relationships in the context of communities. In aqua3S, the social media analysis component considers mainly the activity of the users in terms of submitted Tweets.

[15] Semantically Interconnected Online Communities

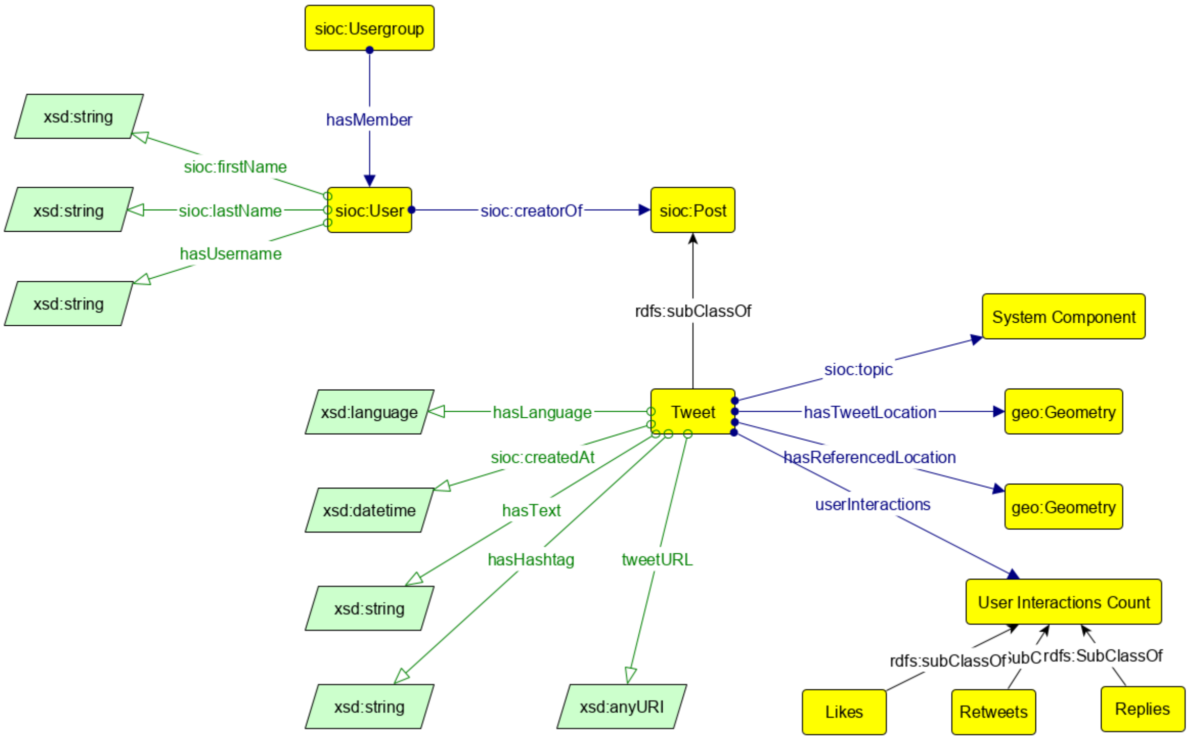

Figure 15. Properties of the Tweet class and the relationship with the User

Figure 15 shows the relevant part of the ontology that models the social media activity. The prefix "sioc" denotes that the class or property is imported from the SIOC ontology. The main class shown in the figure is the Tweet that is a subclass of the sioc:Post class, meaning is a more specific category of an online post. The Tweet class represents an online post made by a user in the Twitter microblogging platform. Tweets (or sioc:Posts in general) are created by online users (sioc:User) that are characterized by their first and last name and their username. The users can also be organized into communities or Usergroups that have individual users as their members. Overall, in the aqua3S the focus is on the posts (tweets) themselves rather than the users that created them or the communities they belong to, thus, the sioc:User and sioc:Usergroup classes were not expanded.

Tweets are described by their attributes (or properties); where some represent inherent fields of the tweet while others represent metadata or higher-level information about the content. In the first category are the properties sioc:createdAt, hasText, hasHashtag, tweetLanguage, and the URI of the tweet. On the contrary, the property userInteractions shows how many times the tweet has been retweeted or replied to. The sioc:topic property declares the subject of the tweet, which within the context of the aqua3S project is a relevant System Component. Regarding spatial information, the hasTweetLocation property defines the location of the user at the moment they posted the tweet, if available. The hasReferencedLocation defines the location or locations that are included within the tweet and might be different from the location of the tweet.

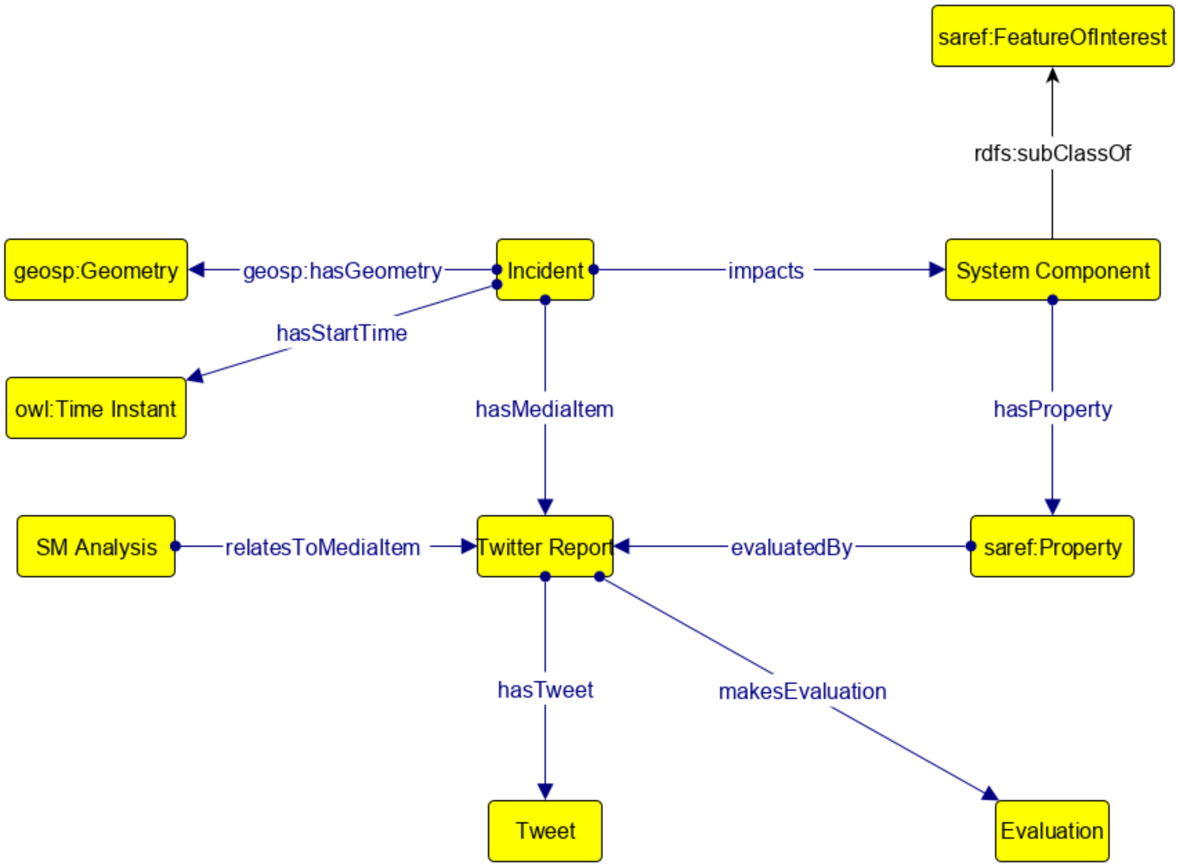

Figure 16. The relationships between the Incident, the Twitter Report and the Evaluation

The Figure 16 illustrates the connection between the Twitter data and the crisis, which is through specific incidents that correspond to collections of tweets (Twitter Report). The transition from individual tweets to collections is done because the social media analysis component will analyse collectively multiple tweets in order to evaluate the situation. The results of the analysis will be attached to or initialize one Incident that defines a deviation from normal conditions. Additionally, the Twitter Report evaluates a property of a particular System Component, which also is a Feature of Interest. For example, it might evaluate the Acceptability Property that examines the odour of a Feature of Interest that is the drinking water of a city. Overall, a deviation from normal conditions that is detected through the social media analysis will be populated to the ontology in order to provide a fuller view of the whole situation and, thus, enable effective decision support.

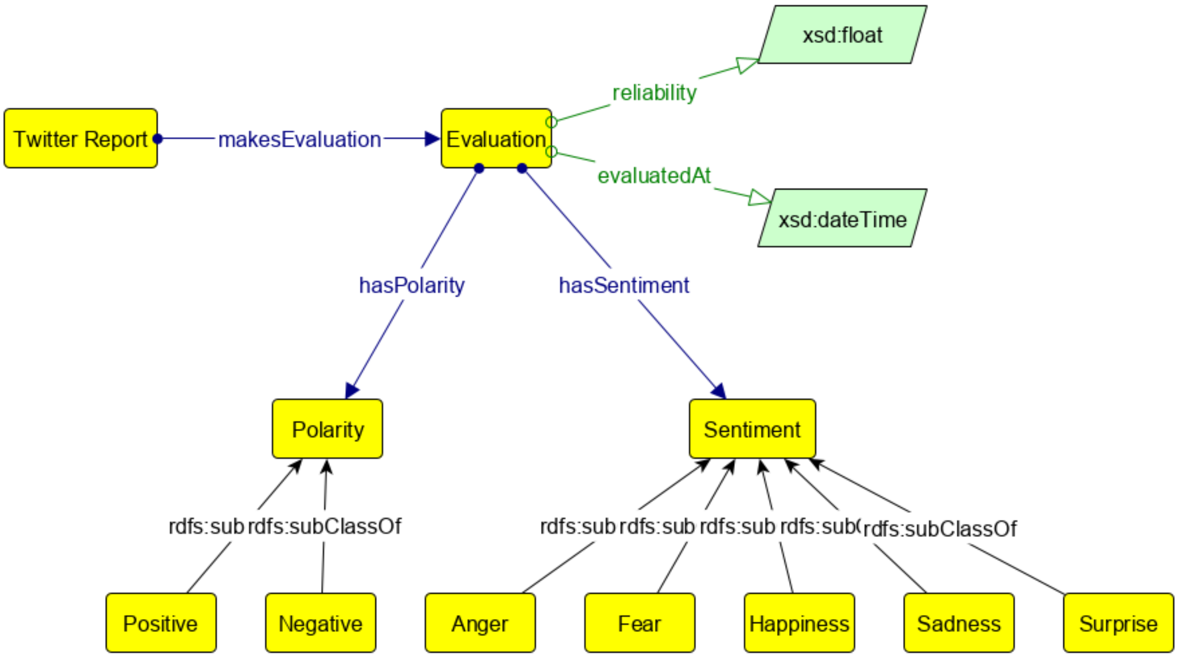

Figure 17. The Evaluation class

In more detail, the social media analysis component will analyse the collected Twitter data through some axes of interest that will characterise the Feature of Interest (the drinking water). In Figure 17, the Evaluation class is illustrated with some important properties that characterize it. The reliability property assigns a numerical value to the evaluation that declares the margin of error. The evaluateAt property defines the time of the analysis. Most importantly, the hasPolarity and hasSentiment properties define what the result of the evaluation is. For example, if the Twitter Report evaluates the taste of the water and the polarity is negative, this shows that the public perception is negative towards the taste of the water and likely indicates some abnormality.

5.3 Application of ontology model

The ontology models the schema and the structure of the stored information, in order to add semantics according to domain knowledge and subsequently enhance the understanding of the situation. So far, the ontology schema was presented in terms of classes, hierarchies, and properties. However, in applications that utilize ontologies, the data are transformed as instances of the ontology in order to adhere to the specification. In detail, instances are the real-world entities that belong to some class, or alternatively, instances are concrete entities that correspond to some abstract entities (the classes). In this subsection an example of the representation of some knowledge that adheres to the ontology schema is shown. For more details on how the incoming data of the aqua3S system are stored according to the ontology, see section 6.

Figure 18 illustrates the instance graph that corresponds to some knowledge regarding a crisis and is modeled according to the ontology. In particular, the crisis instance is of *Chemical Pollution* subclass and as triggered by a Hazard that defines that the Chlorine concentration is elevated. The *hazard* itself was realized by a Hazardous event that also has a starting time associated, and a Risk instance. Another aspect of the crisis is the impacted system component, in particular, the *lake* instance. The particular system component is bounded by a *polygon* instance of the corresponding GeoSPARQL class. Moreover, the lake has the *chlorine* instance of the class saref:Property that is measured by a device, and corresponds to the chlorine concentration of chlorine in the lake.

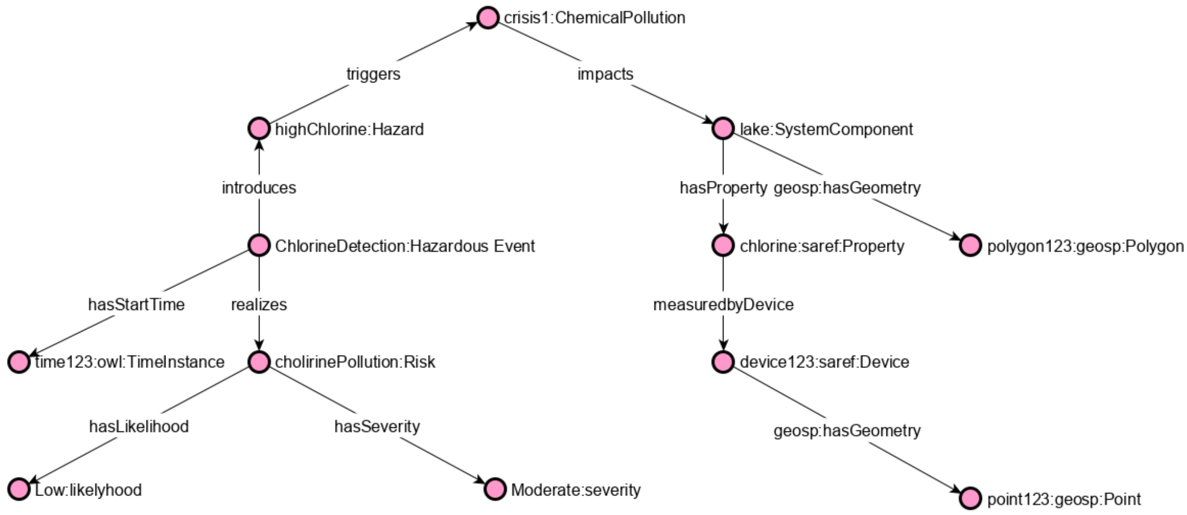

Figure 18. Example of instance graph about a crisis

6. Semantic Reasoning Framework

6.1 Component description

The component will provide decision support through rule-based semantic reasoning to the system based on raw data and other analytics components. The decision support mechanism utilizes semantic reasoning capabilities that are implemented on a Knowledge repository called Knowledge Base (KB). The KB is a data store that stores RDF triplets (or triplestore) that supports semantic reasoning operations. Semantic reasoning (or rules engine) is a piece of software that infers logical consequences based on a set of asserted facts. In practice, semantic reasoning examines the available data and adds new knowledge to the system that might not be directly apparent but is a result of domain knowledge. The component is expected to support mainly the Crisis Classification component, while having the option to notify other components if notable inferences are produced. Inversely, there can be the option for other components to request some reasoning results, for example for report generation purposes.

An overview of the architecture of the component is illustrated in Figure 19. The interaction between the Semantic Reasoning Component and the rest of the system is exclusively through the Context Broker. Firstly, the inputs of the component are raw data (e.g. from sensors/social media) and results of analysis components (e.g. social media analysis), which are both read from the context broker. Inversely, the results of the component are also published through the Context Broker, and they are directed towards the correct recipients. Additionally, potential semantic reasoning requests will also be read through the Context Broker. Overall, the component will read new data and requests (if applicable) from the Context Broker and will publish the reasoning inferences also on the Context Broker.

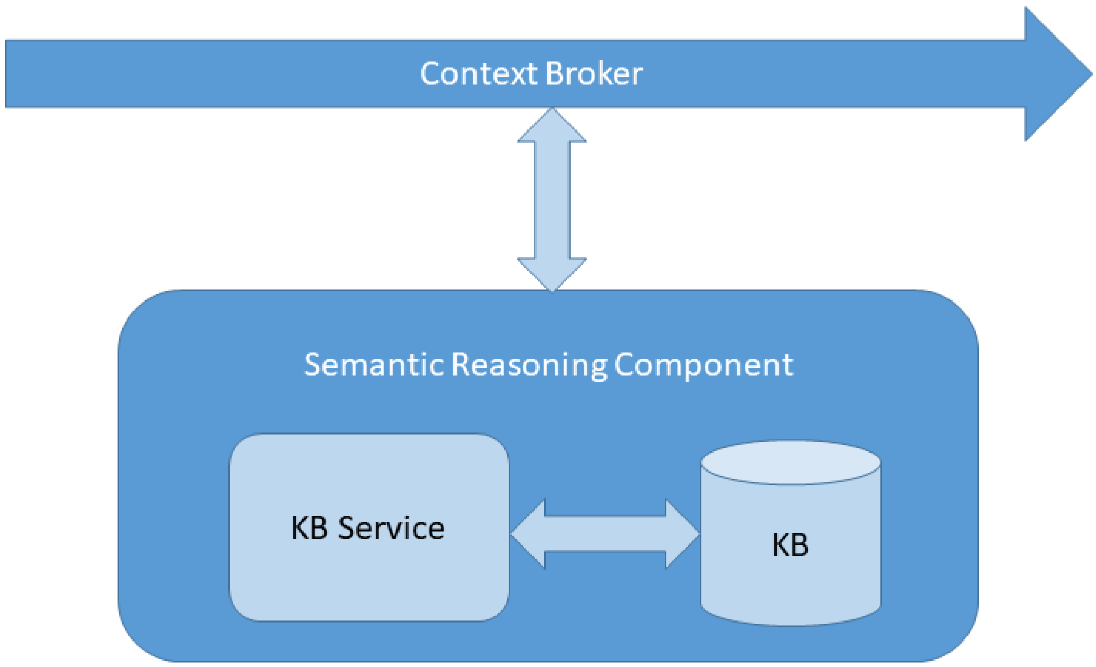

Figure 19. The architecture around the Semantic Reasoning Component

Internally, the component is divided into two subcomponents, first the Knowledge Base is the repository that stores the RDF triples and holds the available knowledge of the system. The KB supports natively semantic reasoning thus enabling decision support by inferring the consequences of a crisis bases on data from the KB. However, the KB is self-contained and has limited means of interaction, thus, necessitates the development of a secondary subcomponent that handles the interactions with the KB, called KB Service (or KBS). KBS is the interface between the semantic reasoning engine and the rest of the project. On the inwards direction, it reads and parses the incoming data, semantically transforms them according to the ontology and stores them to the KB. On the outwards direction, the KBS is responsible for scheduling the semantic reasoning operations and notifies the system for notable semantic reasoning results.

Regarding the interactions between the KB and its Service, they are implemented via a SPARQL endpoint that allows for both storing and retrieving data. For storing data, the KBS parses the incoming data and creates the corresponding "insert" SPARQL query that updates the content of the KB. The retrieval of the data is mainly in the context of semantic reasoning where the KBS creates the corresponding "select" SPARQL query to initialize the reasoning within the KB and retrieve the relevant results. Another use case for data retrieval is the request for report generation that simply reads data without requiring semantic operations to take place. However, in all cases the interaction is through the SPARQL endpoint and the corresponding queries. This modular architecture allows for flexible scheduling schemes for reasoning, as the reasoning frequency can be regulated. Typically, reasoning starts either periodically or whenever some specific semantic integration operations take place, e.g. whenever a new sensor measurement is populated. Between the two options, the first is better if the KB population is very frequent, as it reduces the overhead for each new insertion. The second option has the benefit of having more immediate reasoning results, as they are produced as fast as the new data are inserted.

6.2 Semantic Integration

Semantic integration is the process of parsing structured data, translating them to semantic knowledge and storing it to the Knowledge Base (KB). This is done by using the produced data along with domain knowledge in order to translate the raw data to knowledge and subsequently store them to the KB. The process of storing semantic knowledge to the KB is also called population of the KB. Semantic integration is a prerequisite of Semantic Reasoning as it renders the available data usable within the KB. In this section, the currently supported semantic integration is presented. In particular, it is shown how raw sensor measurement data and social media data are transformed to semantic knowledge and how they are represented in graph form.

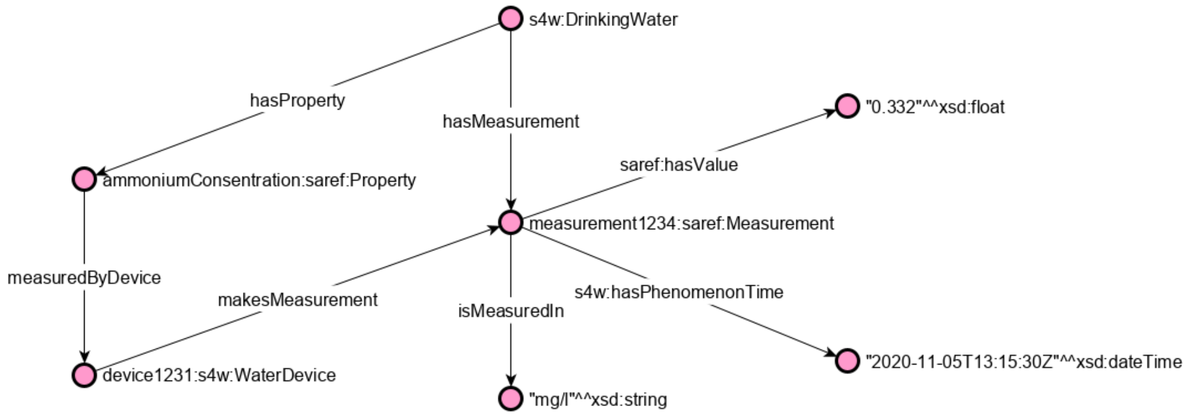

Figure 20. The portion of the Knowledge Graph about sensor measurement data

In particular, when having data from sensor measurements the Figure 20 illustrates the resulted section of knowledge graph. The figure assumes the incoming data (typically formalized as a JSON file) contains information about the **device** that produced the reading, the **time**, the **value**, and the **units** of the measurement. Information about the subject of the measurement (Drinking Water) and the measured Property is also expected as an input from the system, possibly as an attachment of the device. In detail, the Figure 20 shows the knowledge graph after the semantic integration of a measurement of Ammonium concentration in Drinking Water that has value 0.332 mg/l and was measured at a specific time and position. As new measurements are produced, the knowledge graph is updated with more measurement instances, according the same scheme.

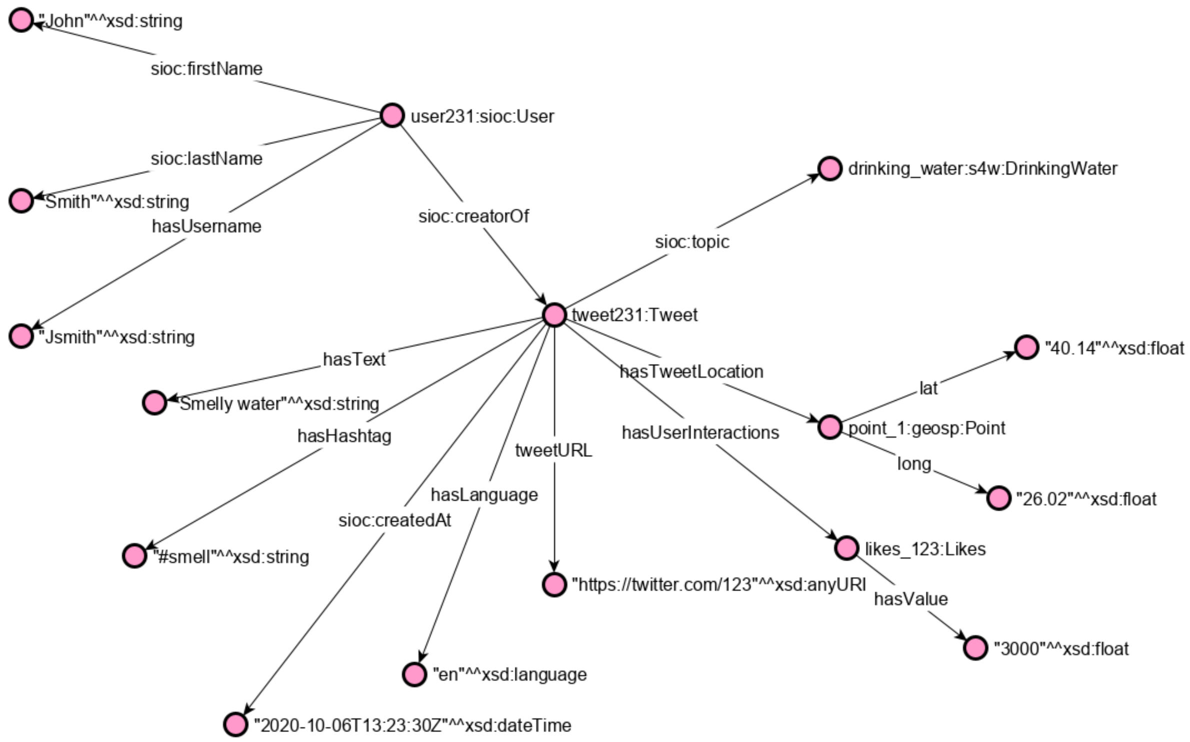

Figure 21. The portion of the Knowledge Graph with information about a tweet

Figure 21 shows the relevant part of the knowledge graph after it has been populated with social media data. In particular the incoming information includes the Twitter user that posted the tweet, data about the tweet itself and some metadata about the location and interactions and the subject of the tweet. In particular, the graph shows that a User with username "Jsmith" has posted a tweet with the text "Smelly water", and the hashtag "#smell". Such tweet could be relevant for the detection of abnormalities of the drinking water. The location of the tweet is also important in order to help the localization of the issue. In this case, the topic of the tweet is assumed to be a result of the social media analysis, as it usually not apparent (for software components) in raw Twitter data.

6.3 Semantic reasoning

Semantic reasoning refers to the operation that extracts implicit information based on the available data stored in the KB. Semantic reasoning is performed according to a set of predefined rules that describe the preconditions that are necessary for the inferences to be realized. The results of the reasoning typically update the contents of the KB with the inferred knowledge and if applicable they are notified to the system. In this section, two examples of implemented semantic reasoning are shown, along with the effects the reasoning has to the Knowledge Graph.

6.3.1 Hazardous situation of Elevated Chemical Substance

If the concentration of Ammonium is above 0.5 mg/l in the drinking water,

then this constitutes a hazardous situation.

This example of semantic reasoning shows how an elevated concentration of a chemical element causes the instantiation of a Hazard (hazardous situation). In particular, the rule is for the chemical substance of Ammonium, but the same logic is followed for other substances. The reasoning defines that if a sensor measures the concentration of ammonium to be above a limit, then this constitutes a Hazard that is created and inserted to the Knowledge Base. The inserted Hazard can be associated with a Hazardous Event that introduces the Hazard if it is known. The allowed concentration limits are from the *Drinking Water Directive 98/83*, a drinking water supply and sanitation regulation by the European Union.

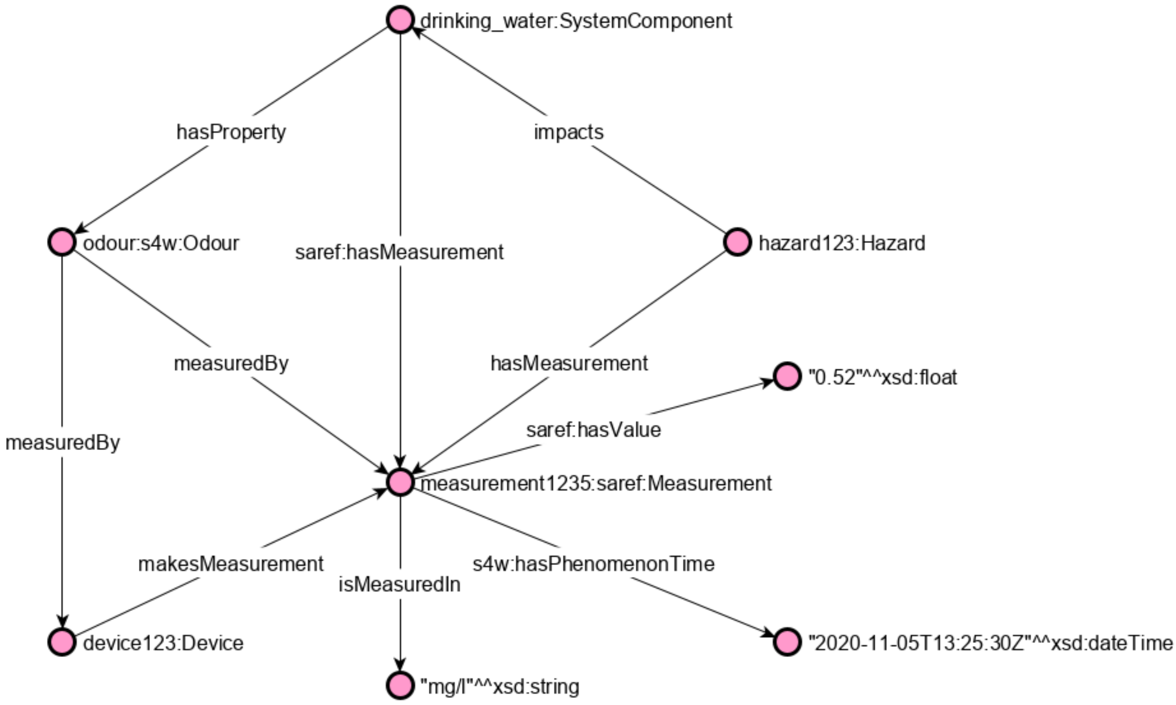

Figure 22. The instance graph for the created Hazard based on sensor measurements

Table 4 shows the SPARQL code that is executed in order to realize the semantic reasoning rule. In particular, the "where" portion of the query checks if the preconditions of the rule are realized, while the "insert" makes the changes to the KB. In detail, the preconditions define that there must be a measurement of Ammonium concentration in Drinking Water that is measured in mg/l and has value above 0.5 (defined by the FILTER statement). The "insert" portion of the SPARQL query defines that a new hazardous situation instance must be created that impacts the Drinking Water and is associated with the particular measurement. Additionally, if the rule is realized, a corresponding alert is produced and forwarded to the context broker to notify the system.

```
insert{
    :hazard_123 rdf:type owl:NamedIndividual.
    :hazard_123 rdf:type :Hazard.
    :hazard_123 :impacts ?foi.
    :hazard_123 :hasMeasurement ?measurement.
}
where{
    ?measurement rdf:type saref:Measurement.
    ?measurement saref:hasValue ?val.
    ?measurement saref:hasUnits ?units.
    ?device saref:makesMeasurement ?measurement.
    ?device saref:measures ?property.
    ?property rdf:type s4watr:Ammonium.

    ?foi rdf:type s4watr:DrinkingWater.
    ?foi saref:hasMeasurement ?measurement.

    FILTER (?val>=0.5 && ?units="mg/l")
}
```

Table 4. SPARQL reasoning query for the first reasoning use case

6.3.2 Hazardous situation of Elevated Chemical Substance

If the public sentiment of the odour of drinking water is negative, then this constitutes a hazardous situation.

This rule determines that if the public sentiment of drinking water odour is negative, then this signifies an abnormality of the drinking water distribution system. The public sentiment is evaluated based on the sum of collected tweets that are relevant to the property of interest. The class Twitter Report groups a number of tweets and has an evaluation associated that refers to a particular property. The evaluation is made by the social media analysis component. In particular, Figure 23 shows that the Twitter Report instance evaluates the saref:Odour property of the drinking water. Additionally, the social media analysis component judged that the public sentiment expressed towards the odour of the drinking water is negative. This sentiment (or polarity) of drinking water signifies that there is an abnormality, thus, the semantic reasoning creates the corresponding Incident to increase the awareness of the crisis. The Social Media analysis component evaluates whether the public perception is strong enough to declare an abnormality to the system. The abnormality might be mapped to particular thresholds in the future in cooperation with the Social Media analysis component.

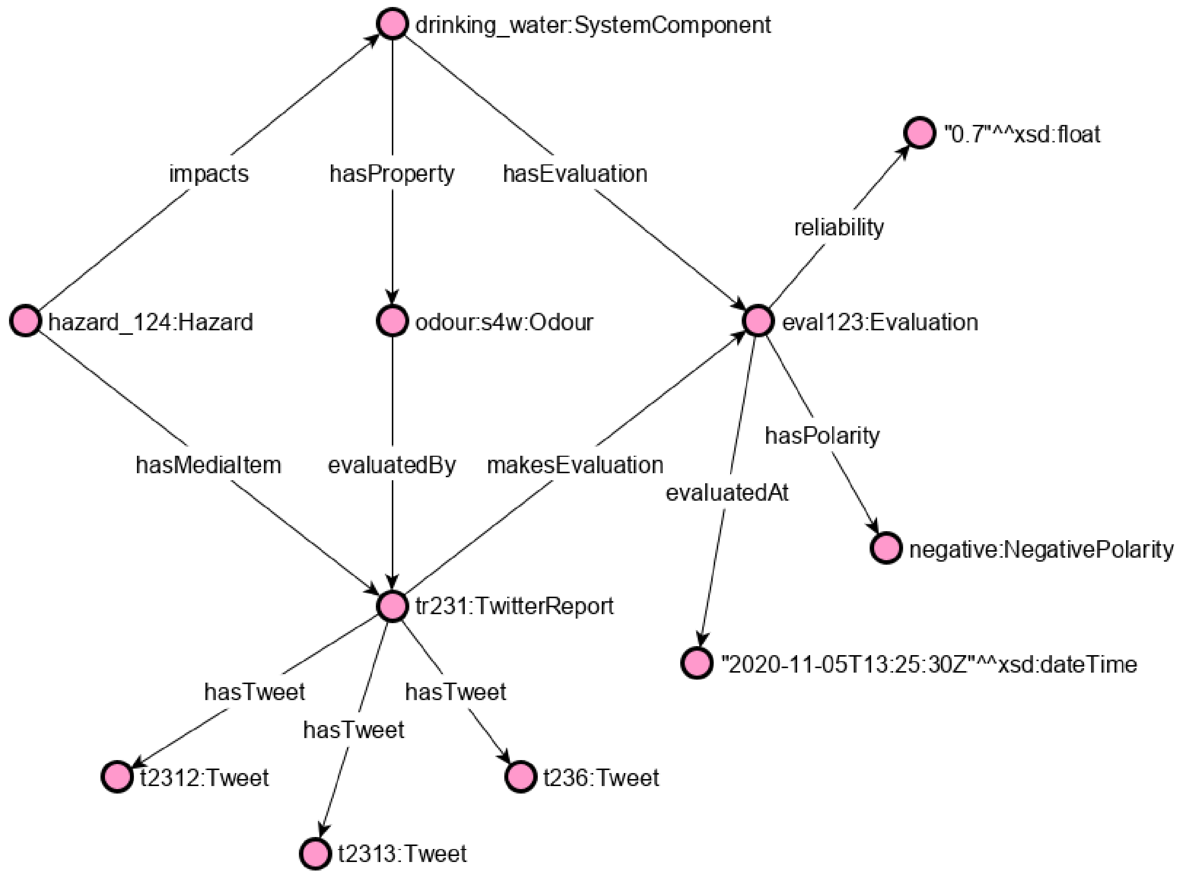

Figure 23. The instance graph for the Incident of odorous drinking water

Table 5 shows the corresponding SPARQL query that implements the semantic reasoning rule. In particular, it defines that if there is a Twitter Report with negative polarity towards the drinking water, then a new Hazard instance is created and is associated with the media that declared the negative polarity. Whenever the rule is successfully executed, its results are also stored in the KB, in order to enrich and improve the overall situational awareness and support the decision-making processes.

```
insert{
    :hazard_124 rdf:type owl:NamedIndividual.
    :hazard_124 rdf:type :Hazard.
    :hazard_124 :hasMediaItem ?media.
    :hazard_124 :impacts ?water.
}
where {
    ?eval rdf:type :Evaluation.
    ?eval :hasPolarity ?polarity.
    ?polarity rdf:type :NegativePolarity.
    ?media :makesEvaluation ?eval.
    ?media rdf:type :TwitterReport.
    ?property :evaluatedBy ?media.
    ?water rdf:type s4watr:DrinkingWater.
    ?water saref:hasEvaluation ?eval.
}
```

Table 5. SPARQL reasoning query for the second reasoning use case

7. Conclusions and future work

This document presented the development of the initial version of the aqua3S ontology and the initial decision support functionalities through the means of semantic reasoning. The core of the ontology that was developed is aimed towards the representation of crises, their high-level characteristics, and offers an initial framework for expansion towards finer granularity of incidents within the crisis. The ontology models the crises according to the specification of EN 15975 (Security of drinking water supply), which was determined to cover the needs of project. The core ontology offers the basis for representation of crisis of the water domain but is flexible enough to cover all use cases of the aqua3S project. Additionally, standardized ontologies were adopted and imported for the specification of temporal and spatial information.

Moreover, the ontology was extended in order to cover knowledge originating from sensor measurements and from social media. In particular, for sensor measurements, the adopted structure is mapped as closely as possible to the SAREF ontology and its new extension for water, in order to ensure interoperability. Regarding the social media knowledge representation, the focus was on both individual posts and collections of multiple posts. Collections of posts are expected to be analyzed and produce better insights than individual posts on the public perception. For the social media representation, entities, and relationships from the SIOC ontology were adopted, and new additions were made when necessary.

The ontology provides the means that allow for the semantic reasoning and decision support to be implemented. Thus far, some initial semantic reasoning is implemented that provides a complete and valid view to the crisis classification component and to the rest of the aqua3S system. In particular, the semantic reasoning is implemented through SPARQL rules, and works on data from sensor measurements and social media.

Next steps of the development of the aqua3S ontology include the continuous expansion of the ontology to accommodate for satellite and drone data and adjust to the maturing analytics components that will produce insightful knowledge. The semantic reasoning also stands to be expanded with more and more complex rules to infer more advanced knowledge and thus better support the decision-making process. Moreover, both the ontology and the semantic reasoning capabilities will continue to be updated and validated as the available data approach the actual real-world data. On the operational side, the development of the Knowledge Base Service is pending, along with its integration with the context broker. Finally, some report generating functionalities will be implemented when the system becomes more mature.

8. References

Baader, F. (2003). Terminological cycles in a description logic with existential restrictions. *IJCAI*, 325-330.

Bénaben, F. e. (2008). A metamodel and its ontology to guide crisis characterization and its collaborative management. *ISCRAM*, (σ. Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management). Washington, DC, USA.

Berners-Lee, T. a. (2001). The semantic web. *Scientific american*, 34-43.

Bezerra, C., Freitas, F., & Santana, F. (2013). Evaluating ontologies with competency questions. *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 3*, σσ. 284–285.

Bojars, U., Breslin, J. G., Berrueta, D., Brickley, D., Decker, S., Fernández, S., . . . others. (2010). SIOC core ontology specification. *WC Member*.

Brickley, D. (2006). Basic geo (WGS84 lat/long) vocabulary. *Documento informal escrito en colaboración*.

Brickley, D., & Miller, L. (2007). FOAF vocabulary specification 0.91. *FOAF vocabulary specification 0.91*. Citeseer.

Daniele, L., den Hartog, F., & Roes, J. (2015). Created in close interaction with the industry: the smart appliances reference (SAREF) ontology. *International Workshop Formal Ontologies Meet Industries*, (σσ. 100–112).

EN-15975. (2013). *Security of drinking water supply – Guidelines for risk and crisis management.* International Organization for Standardization.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 199-220.

Haarslev, V., Hidde, K., Möller, R., & Wessel, M. (2012). The RacerPro knowledge representation and reasoning system. *Semantic Web, 3*, 267–277.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., Rudolph, S., & others. (2009). OWL 2 web ontology language primer. *W3C recommendation, 27*, 123.

Hobbs, J. R., & Pan, F. (2006). Time ontology in OWL. *W3C working draft, 27*, 133.

Houston, G. (1988). *ISO 8601: 1988 Date/Time Representations.* International Organization for Standardization.

Janowicz, K., Haller, A., Cox, S. J., Le Phuoc, D., & Lefrançois, M. (2019). SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics, 56*, 1–10.

Kontopoulos, E., Mitzias, P., Moßgraber, J., Hertweck, P., van der Schaaf, H., Hilbring, D., . . . others. (2018). Ontology-based Representation of Crisis Management Procedures for Climate Events. *ISCRAM.*

Lawrynowicz, A., & Lawniczak, I. (2015). The Hazardous Situation Ontology Design Pattern. *WOP.*

Lieberman, J., Singh, R., & Goad, C. (2007). W3c geospatial vocabulary. *W3C Incubator Group*.

Liu, S., Brewster, C., & Shaw, D. (2013). Ontologies for crisis management: A review of state of the art in ontology design and usability. *ISCRAM*, (σσ. 1–10).

McGuinness, D. L., Van Harmelen, F., & others. (2004). OWL web ontology language overview. *W3C recommendation, 10*, 2004.

Perry, M., & Herring, J. (n.d.). GeoSPARQL—A Geographic Query Language for RDF Data; Proposal for an OGC Draft Candidate Standard 11-052r1, Open Geospatial Consortium Inc., 2011. *Access restricted to OGC members*.

Raimond, Y., Abdallah, S. A., Sandler, M. B., & Giasson, F. (2007). The Music Ontology. *ISMIR*, *2007*, σ. 8th.

Shearer, R., Motik, B., & Horrocks, I. (2008). HermiT: A Highly-Efficient OWL Reasoner. *Owled*, *432*, σ. 91.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Journal of Web Semantics, 5*, 51–53.

Tsarkov, D., & Horrocks, I. (2006). FaCT++ description logic reasoner: System description. *International joint conference on automated reasoning*, (σσ. 292–297).

Tserpes, K., Papadakis, G., Kardara, M., Papaoikonomou, A., Aisopos, F., Sardis, E., & Varvarigou, T. (2012). SocIoS: A Social Media Application Ontology. *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, (σσ. 574–584).

Vardi, M. Y. (1997). *Why is modal logic so robustly decidable?* Tech. rep.

Zhou, Q., & Fikes, R. (2002). A reusable time ontology. *Proceeding of the AAAI Workshop on Ontologies for the Semantic Web.*