Article

An Ontology-Based Framework for Publishing and Exploiting Linked Open Data: A Use Case on Water Resources Management

Pilar Escobar [1], María del Mar Roldán-García [2]⊙, Jesús Peral [1],*, Gustavo Candela [1]⊙ and José García-Nieto [2]⊙

[1] Lucentia Research Group, Department of Software and Computing Systems, University of Alicante, 03690 Alicante, Spain; mpilar.escobar@ua.es (P.E.); gcandela@ua.es (G.C.)

[2] Departamento de Lenguajes y Ciencias de la Computación, ITIS Software, University of Málaga, E.T.S. de Ingeniería Informática, Campus de Teatinos, 29071 Málaga, Spain; mmar@lcc.uma.es (M.d.M.R.-G.); jnieto@lcc.uma.es (J.G.-N.)

* Correspondence: jperal@dlsi.ua.es; Tel.: +34-965-90-37-72

check for updates

Abstract: Nowadays, the increasing demand of water for electricity production, agricultural and industrial uses are directly affecting the reduction of available quality water for human consumption in the world. Efficient and sustainable maintenance of water reservoirs and supply networks implies a holistic strategy that takes into account, as much as possible, information from the stages of water usage. Next,-generation decision-making software tools, for supporting water management, require the integration of multiple and heterogeneous data sources of different knowledge domains. In this regard, Linked Data and Semantic Web technologies enable harmonization of different data sources, as well as the efficient querying for feeding upper-level Business Intelligence processes. This work investigates the design, implementation and usage of a semantic approach driven by ontology to capture, store, integrate and exploit real-world data concerning water supply networks management. As a main contribution, the proposal helps with obtaining semantically enriched linked data, enhancing the analysis of water network performance. For validation purposes, in the use case, a series of data sources from different measures have been considered, in the scope of an actual water management system of the Mediterranean region of Valencia (Spain), throughout several years of activity. The obtained experience shows the benefits of using the proposed approach to identify possible correlations between the measures such as the supplied water, the water leaks or the population.

Keywords: ontology engineering; linked data; business intelligence; decision-making; water resources sustainability

1. Introduction

According to the BCC Research Report "Novel Water Sustainability Technologies: Key Projects and Opportunities, Financing, and Venture Capital, Transactions and Trends" [1], natural resources and ecology are fundamental to local economies. The growth in demand for natural resources has created a global market based on sustainable technology to harvest maximum supply. The nexus between water and sustainability can be connected via a thorough understanding of water needs, the socio-economic effects of water projects and transfers, and economic and financially efficient water management technologies. With the population explosion (current United Nations estimates forecast 8.5 billion by the year 2030), the society is in dire need of natural resource management. The main goal of sustainable water technology is to make water affordable, accessible and abundant for all human

beings in the coming decades. In this new scenario, novel water technologies and tools play a leading role to ensure water policies, at both national and local levels. In this context, an open challenge in modern cities lies in the efficient planning and management of water supply networks, which is evolving from the mere execution of hydrologic engineering projects, to the generation of data-driven approaches for the integrated water management involving multiple sources of data.

Emerging Big Data and Linked Open Data (LOD) technologies enable the development of integrated data lakes oriented to the rational management of water resources [2]. The challenge is to integrate multi-source data concerning water management holistically, then leading new and existing tools to present contextual and significant information in an intuitive and interactive way to users, hence aiding in the decision-making process. In general, this is an emerging topic in sustainable development, as it is being applied to multiple related projects such as sustainable tourism [3], educational annotation [4] and green supplier selection [5].

Therefore, from the early design of water monitoring services, a key task is to gather and integrate knowledge from different domains of application, including reservoir strategic levels, environmental and meteorological information, quality measures and registrations, climate change issues, urban planning, end user feedback, energy consumption, etc. [6]. Unfortunately, there are still serious barriers in terms of interoperability that restrict data interchange between these information silos, and limit the analytic power of up-level Business Intelligence (BI) tools.

Linked data and Semantic Web technologies are conceived upon protocols and W3C international standards for sharing and annotating structured data on the web. In this context, conceptual modeling is conducted by means of ontologies, which define classes, relationships, instances and axioms of a particular domain of knowledge [7,8]. In addition, ontology objects refer to concepts in terms of entities and events in the real world, including the relationships between these entities that represent the semantic links. In this context, linking water data to enable semantic knowledge graphs of contextual information is a critical task, since representing water distribution and usage data in the form of linked data repositories would make them open. This will indeed allow for easily combining with external (linked) data from heterogeneous, although relevant, domains.

The main motivation of this study is then to design and develop an ontology model to allow knowledge consolidation and extraction in real-world scenarios, where water supply networks are managed with data-driven BI solutions. This paper describes the fundamentals of this proposal, detailing the main components of the initial envisioned architecture, as well as piloting implementation of a water management data workflow, as a proof of concept tool for decision-making strategies.

For testing purposes, the proposed ontology model is materialized by means of an RDF (Resource Description Framework) repository (RDF in W3C https://www.w3.org/RDF/ (visited on 24 October 2019)) , which is gradually populated throughout mapping procedures of multiple data sources. These data are collected from heterogeneous sensorized data and computed measurements from different geographic zones, in a real-world water management network. The resulting RDF repository can be accessed by high level queries using federated SPARQL sentences.

The main contributions of this paper are outlined as follows:

- A water supply network oriented ontology is proposed, which allows for modelling, generating, integrating, publishing and exploiting a dataset, enabling general users to interact with the data. This ontology has been developed in OWL 2 and considers a large set of concepts, attributes and relationships to contextualize water management supply networks field.

- Our approach is tested on real-world data from a water management supply network in the Mediterranean region of Valencia (Valencian Community, Spain). It is a southeastern zone of Spain where autumn storm episodes are quite common, with flooding of urban areas, but with usual annual droughts. Different cities of the Valencian region such as Alicante or Valencia have developed an integral and sustainable water management plan, including flood prevention and supply network deep management among their priorities. Reported results allow us to support domain experts in the decision-making process.

- A semantic model has been implemented for materialization of all the involved concepts and measures from the data sources, as well as those processes and components required. The concepts are integrated according to the ontology scheme and integrated in the RDF repository. On top of this, a series of SPARQL queries have been formulated for federated querying.
- In this regard, the links to external repositories have been used to enrich the original data in order to facilitate data reuse and interoperability. Thus, in our use case, the links to GeoNames and Wikidata have allowed to add contextual information to original data.

The remainder of this article is organized as follows: Section 2 contains background concepts and overview of related literature. In Section 3, the general framework is described with details of the RDF repository and mapping functions. Section 4 presents a real-world use case of the water supply network for validation describing the proposed ontology and discussion extracted from the experimentation. The main conclusions and future lines of research are drawn in Section 5.

2. Background and Related Work

In this section, the main background concepts in the area of Semantic Web are briefly explained with the purpose of a better understanding of this paper. A series of related works in the state of the art are also revised in order to point out the main differences with regards to the proposed approach.

2.1. Background Concepts

- Ontology. Ontologies offer a formal model of concepts of interest (classes), features and attributes of each concept (properties) and property restrictions, involving a specific knowledge domain in the real world [9,10]. Ontologies are a layer of the W3C standard stack (https://www.w3.org/standards/semanticweb/ (visited on 24 October 2019)). A knowledge base is made up of an ontology and its instances (set of class and property individuals). A knowledge base provides services to make heterogeneous systems and databases interoperability easier.
- RDF. Resource Description Framework [11] is a W3C recommendation for describing resources in terms of triples. An RDF triple comprises a subject, a predicate and an object. RDFS (RDF Schema) [12] provides a language for describing vocabularies that are used in RDF statements.
- OWL. The Ontology Web Language (OWL) is an extension of RDF and RFFS for defining machine understandable ontologies on the Web. From a formal point of view, an OWL ontology corresponds to a TBox in the context of a very expressive DL (description logic) [13]. Thanks to the equivalence of OWL with DL, OWL-DL provides maximum expressiveness while keeping computational completeness and decidability [14].
- SPARQL [15] is the W3C query language for querying RDF repositories. SPARQL works with RDF graphs [16] where data sources are identified by URIs.

2.2. Related Work

Providing information about water consumption and water availability will increase end-user awareness and improve the quality of water management decisions and water government [6]. In addition, improving access to data and fostering open exchange of water information are crucial to solving water resources issues [17].

With respect to the massive volumes of data, as it occurs with water information, they are already present in modern cities and still rapidly growing as a result of diverse data sources, including all types of smart devices and sensors (Internet of Things, IoT) and social networks. This fact has led to an increasing interest in incorporating these huge amounts of external and unstructured data into traditional applications. However, the potential opportunity offered by the gathered information is often not exploited due to advanced techniques of data analysis, visualization and services to enable data exploration have to be developed. In this regard, the use of specific ontologies in smart

city domains, such as Smart City Ontology (SCO) [18,19], allows a semantics-enabled exploration of urban data.

The publication of ontologies related to water management has aroused great interest in the research community. Water management in smart cities is often focused on providing the adequate water supply to the citizens [20]. Several approaches are focused on ontology frameworks for effective management of water, taking into account aspects such as quality, efficiency of water reuse, or detection of failures [21,22]. In addition, a model is introduced with the objective of integrating water distribution networks, including a review of the most well-known related ontologies [23].

Nowadays, many public and governmental institutions are investigating different ways to make visible their data and publish them as LOD. Furthermore, they are developing new interfaces to facilitate the interaction with general users for exploring data. However, many aspects must be taken into account in order to exploit the full potential of the data.

With regard to the enrichment process, several open Knowledge Graphs (KGs) have been created based on LOD concepts, such as DBpedia [24], Wikidata [25] and YAGO [26]. KGs are a rich source of information covering general knowledge to enrich original datasets. They provide structured information which allows: (i) the connections with other external repositories (e.g., GeoNames), and (ii) multilingual aspects through the access to descriptions and properties in different languages. In addition, for a dataset to be useful, it is crucial that the data be valid and consistent [27]. On the other hand, The European Data Portal (https://www.europeandataportal.eu (visited on 19 December 2019)) harvests the data collections available on public data portals across European countries. According to the Reusing Open Data report [28], geographic data are the second data category most reused by institutions in the member states of the European Union.

Moreover, the usage of reliable standards promoted by known organizations for data publication enable easy reuse. The combination of different datasets facilitates the creation of multidimensional models in which different evaluation techniques can be applied. It is important to mention that W3C recommends the RDF Data Cube Vocabulary [29] for the publication of multidimensional data.

We can conclude this section after reviewing previous work that, to the best of our knowledge, our general framework is the first one which allows for modelling, generating, publishing and exploiting a dataset related to water management domain. Furthermore, general users can visualize and analyse the enriched data with the links to external repositories.

3. Proposed Approach

With the growing interest in publishing open data on the Web, different practices and guidelines have been published. Several works, such as the approaches of Hyland [30] and Villazón-Terrazas et al. [31], present lifecycle models including different activities focusing on publishing data in standard open Web formats. Regarding the access and reuse of open government data, the W3C Government Linked Data Working Group has proposed a guide to aid in this task [32].

In our approach, we have followed the lifecycle of Villazón-Terrazas et al. [31], from the specification to the exploitation. We have used it since its methodological guidelines are oriented towards exploiting data in a similar manner to that of our proposal.

The framework (Figure 1) consists of four phases: (i) data mapping and pre-processing of sources, (ii) data model generation, (iii) data storage and (iv) data exploitation. Next, we will describe in detail each step of this process:

1. Data mapping and pre-processing of sources: The way in which the dataset is published is vital to enhance its management, exploitation and reuse. The dataset format defines the structure of the published data which will be used by both human and machines. Different formats are used by institutions to publish their data: (1) CSV (Comma Separated Files) is the most used format due to its simplicity, it is highly reusable and machine-readable; (2) XLS allows the use of macros and formulas, which may be challenging to handle, to obtain advanced calculations in a readable

format; (3) XML, RDF and JSON provide a more detailed information of source data including the semantics [33].

However, due to the heterogeneity of data sources regarding data formats and vocabularies, a pre-processing step of the data is compulsory. This process allows data from different institutions and organizations to be processed in a similar way. The pre-processing consists of ETL (Extract-Transformation-Load) tools [34] and parsers in order to obtain normalized information from source data.

2. Data model generation: The guide *Best practices for publishing Linked Data* proposed by the W3C Government Linked Data Working Group recommends the use of standardized vocabularies to improve the published Linked Data facilitating its usage and expansion [32]. However, in some cases, it is necessary to build an ontology reusing existing ontologies or from scratch. Lately, Protégé [35] have attracted strong interest from the research community to construct a large number of diverse intelligent systems, in particular ontologies covering different domains, such as biomedicine, e-commerce or organizational modeling.

   This step also includes the definition of a method to transform the source data into RDF, a machine readable language. RDF facilitates the interoperability and the definition of connections or links to other repositories. The process of conversion to RDF may be done either in batch mode or in an interactive way (for example using graphic applications). We can mention two representative tools: (1) Jena (https://jena.apache.org/documentation/rdf/index.html (visited on 4 February 2019)) is an open source Semantic Web framework for Java. It allows the definition and manipulation of RDF graphs. The graph is represented as an abstract "model" in which classes are used to represent resources, properties and literals. (2) OpenRefine (https://github.com/OpenRefine/OpenRefine (visited on 4 February 2019)) is an open source desktop application to transform raw data into a machine-readable format. The transformations (actions) to be made are defined by the user and stored in a project. Subsequently, a graphical mapping from the project to an RDF skeleton is carried out. Finally, it is exported in RDF format.
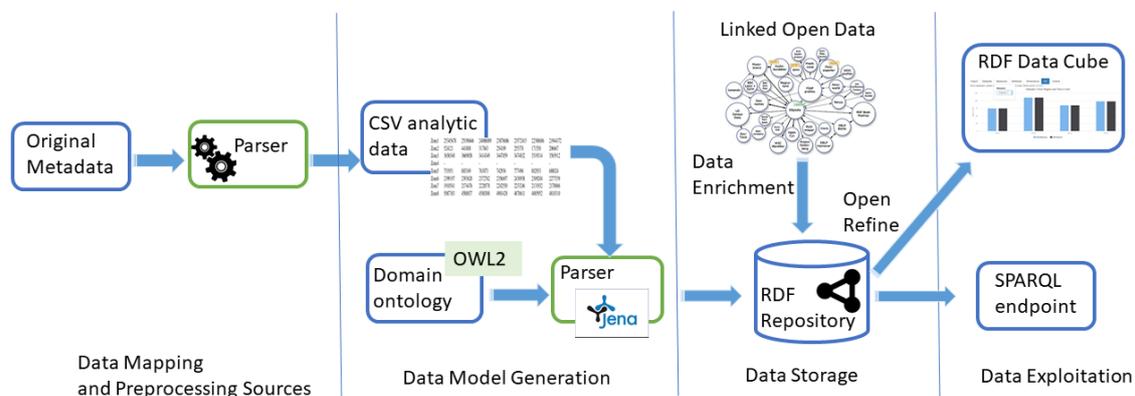

Figure 1. Framework proposal to publish a dataset.

3. Data storage: The Semantic Web is an extension of the Web through standards by the W3C. The standards promote common data formats and exchange protocols on the Web, most fundamentally the RDF. This has led to a considerable increase of RDF data on the web. Consequently, a set of techniques have been proposed for storing RDF data. Different works have previously studied the RDF data storage in an efficient way [36,37] allowing inference, update, scalability, distribution, or SPARQL endpoint.

   In addition, datasets can be enriched by means of external links in order to add information related to the context. In general, the connection process to external repositories consists of two stages: (i) automatic parsing of source information in order to unveil possible links to external sources; (ii) manual validation of the candidate links carried out by experts in data curation.

In this stage, different tools (See, for example, https://tools.wmflabs.org/mix-n-match/ (visited on 4 February 2019)) can be used to facilitate data curation. The selected links will be defined through the *owl:sameAs* relationship.

Nowadays, we can mention different representative repositories used as external source data: (1) GeoNames is a geographical database available and accessible through various web services; it allows the linking of source textual information to geographical locations and currently is one of the most used external repositories [38]. (2) DBpedia is a project aiming to extract structured content from the information created in various Wikimedia projects; it is a Knowledge Graph which stores knowledge in a machine-readable format. (3) Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation and is also a Knowledge Graph; it is a document-oriented database, focused on items, which represent topics, concepts, or objects. (4) Many institutions rely on VIAF (http://viaf.org/ (visited on 4 February 2019)) to connect authority data.

The tools to validate and check data integrity help to enhance its correctness and consistency. For example, constraints provide one method of implementing business rules. Other tools are based on test driven data-debugging frameworks that can run automatically generated (based on a schema) and manually generated test cases against an SPARQL endpoint [39].

4. Data exploitation: The publication of data as LOD allows data reuse. The use of standard vocabularies based on RDF enhances the interoperability, the reuse and the exploitation by other institutions. SPARQL endpoints not only facilitate the access to the data, but also enable federated queries run on other SPARQL endpoints.

   Linked data also enhance the inference of new knowledge by discovering new relationships and automatically analyzing the content of the data, such as identifying possible inconsistencies [40]. Many experiments have been conducted regarding this area [41,42]. In general, inference takes into account the transitivity of predicates such as *rdfs:subClassOf* and *rdfs:subPropertyOf*.

4. Use Case

The approach described in Section 3 has been evaluated using the data from a water supply company in the Mediterranean region of Valencia (Valencian Community, Spain) that manages all processes related to the complete water cycle: catchment, drinking water treatment, transport and distribution for human consumption with full health guarantees.

Currently, the water crisis is already a reality in many Mediterranean countries, threatening their economic growth and the livelihoods of their peoples. The thought is that the problems will only accelerate. Given international forecasts, multiple thumbscrews are operating on these vulnerable fresh water resources and problems will accelerate. The driving forces are strong: population growth and urbanization; tourism and industrialization; globalization; and climate variability and change, decreasing precipitation and increasing the frequency of droughts [43].

The Mediterranean is home to 60% of the world's population classed as living in *water poverty*, less than 1000 m^3 of water available per capita per year. Nearly 20 million Mediterranean people have no direct access to drinking water, especially in the rural areas of the south and east. This growing water scarcity and the uncertainties which climate change may bring only reinforce the need to adapt both water policies and land planning policies that impact water management [43].

In this context, our case study is based on the Mediterranean country of Spain, since is one of the most inhabited communities in Spain. Currently, it has about 5 million inhabitants and is the fourth most populated community in Spain. Explicitly, the data are from the Valencian Community of the period from 2008 to 2014. In addition, the data have been enriched with external repositories in order to add contextual information.

With respect to water consumption in the cities of the Valencian Community, 476 cubic hectometres (hm^3) of water were approximately supplied to the urban water supply networks in 2014. Approximately, 25% of this water volume (94 hm^3) was unregistered water originated by: (a) physical

problems in the supply network (leaks, breaks and breakdowns in the network), and (b) measurement problems (measurement errors and frauds carried out by customers). The remaining 75% (382 hm^3) corresponds to the registered water, i.e., measured on the users' meters [44]. In addition, we have studied water consumption in Spanish households during the analysed years (2008–2014). It is important to highlight that the Valencian Community, although it is not the most populated Spanish region, is the region of Spain that had the highest average water consumption with a value of 164.43 L per inhabitant per day during the studied period.

Taking into account these data, the aim of our work is to achieve a sustainable management of natural water resources (minimising, for example, unregistered water caused by physical problems in the supply network as it involves a large volume of water that is annually wasted). To this end, in our use case, we will analyse the impact of the different measures or available data (such as leaks or breaks in the network and the number of inhabitants) and their relationship with water consumption with the purpose of reducing it.

For instance, decision makers and users might be interested in information about the water leaks in the city of Alicante in 2014 or the possible correlation between population and water consumption in the city of Valencia during a specific period of time. However, sometimes the information provided by water companies is not structured (e.g., textual information), or partially registered and inaccurate, making the information retrieval process more difficult.

In the next three subsections, we will explain the proposed ontology, the application of our proposal to the use case, and the conclusions extracted from the experimentation.

4.1. Water Supply Network Ontology

One of the main aims of this work is to capture, consolidate and integrate data from different water-related sources. For this reason, we opted to design a semantic approach for data sharing and reconciliation, whereby an agreed ontology model is used to archive a common understanding of the domain in which the system operates. In concrete, we have developed an OWL ontology to describe the indicators used by a water supply company. Important terms in the ontology were obtained directly from the company decision makers. The proposed ontology has a total of 31 classes (groups of individuals sharing the same attributes), 23 object properties (binary relationships between individuals), 2 data properties (individual attributes), 156 logical axioms and 12 individuals.

Figure 2 shows the main set of classes (*Zone*, *Indicator* and *Unit*) in the hierarchy starting from the top class *Thing* ($\top$). These main classes are related to other classes and some of them have subclasses (such as *Zone1*, *Zone2*, *NumberOfWaterSupplies*). The class *Zone* models the zones where the indicators were measured. The ontology contains eight zones (from *Zone1* to *Zone8*). Each zone is divided into sub-zones. For example, *Zone3* has nine sub-zones, from *zone3.1* to *zone3.9*. In our ontology, sub-zones are modeled as individuals (*zone3.1*, *zone3.2* ... *zone3.8* are individuals of the Class *Zone1*), which allows querying aggregated data of a specific zone. The class *Indicator* represent formulas for measuring business activities. In our ontology, each *Indicator* is related to the sub-zone where the value is measured and it has a value with a unit of measure in a corresponding year. The class *Unit* contains the different unit of measure as individuals (km, m^3, kWh, etc.). Finally, the class *Year* contains the different years.
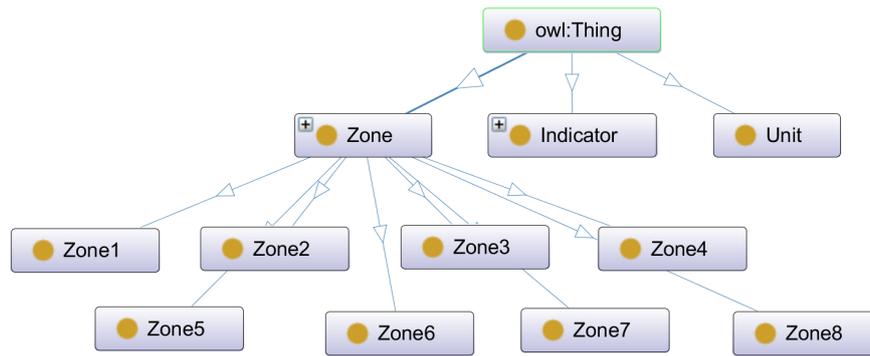
Figure 2. Main classes of the water Ontology. Arrows refer to subClassOf.

Technical Details of Water Ontology

In the following, we describe the main classes, *Indicator* and *zone*, including some of their most interesting properties.

- ***Indicator***. Those attributes provided by the water supply company. In this case, indicators represent measures, i.e., known formulas for measuring business activities with no known targets or thresholds. Each Indicator has a value with a unit of measure in a corresponding year (see object and data properties in Table 1). Subclasses of Indicator in the ontology ($\sqsubseteq$ Indicator) are, among others (see Figure 3), *NumberOfWaterSupplies*, *LeaksTansportNetwork/100kmNetwork*, *AverageAgeDistribution*, *LengthRenovatedNetwork*, etc. The class *Unit* contains the different units of measure as individuals, for example: *Km*, m^3 or *number/1000WaterSupplies*. The class *Year* contains the different years.
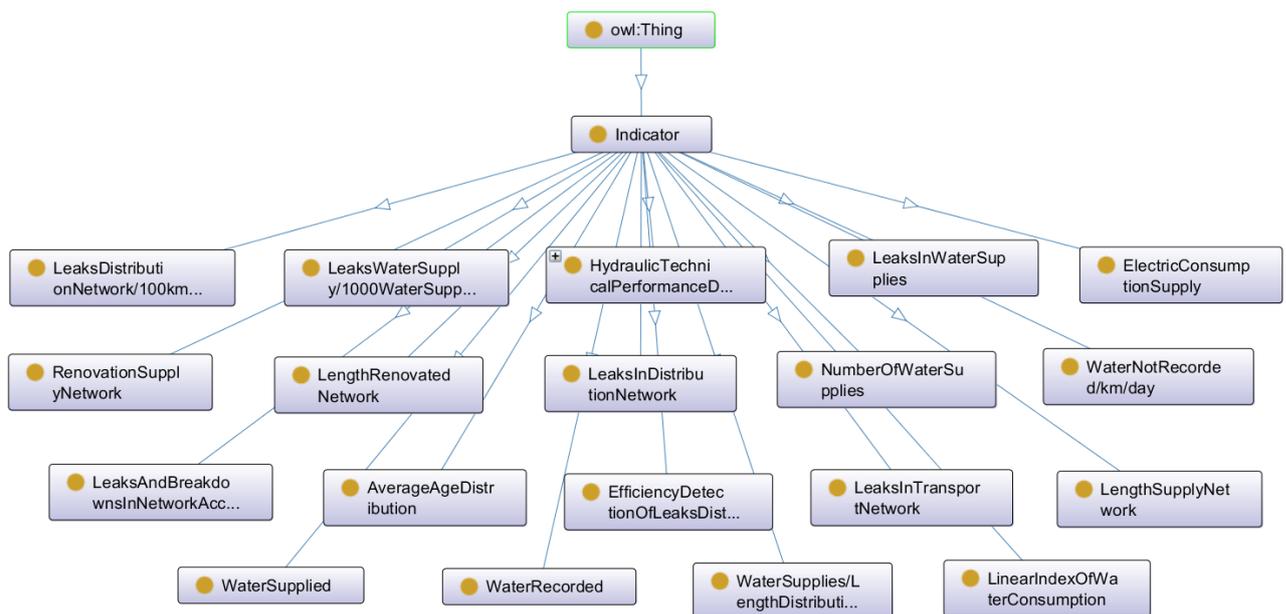

Figure 3. Water Ontology: Subclasses of Indicator.

Table 1. Indicator: object and data properties.

Object Properties	Description Logic
unit	$\exists$ unit.Thing $\sqsubseteq$ Indicator
	$\top \sqsubseteq \forall$ unit.Unit
Data Properties	Description Logic
value	$\exists$ value.Datatype Literal $\sqsubseteq$ Indicator
	$\top \sqsubseteq \forall$ value.Datatype double
inYear	$\exists$ inYear.Datatype Literal $\sqsubseteq$ Indicator
	$\top \sqsubseteq \forall$ inYear.Datatype integer

- *Zone*. This class models the zones where the indicators were measured. The ontology contains eight zones (from *Zone1* to *Zone8*). Each zone is divided into sub-zones. For example, *Zone2* has 14 sub-zones, from *zone2.1* to *zone2.14*. In our ontology, sub-zones are modeled as individuals (*zone2.1*, *zone2.2* ... *zone2.14* are individuals of the Class *Zone2*), which allows querying aggregated data of a specific zone. Each indicator is related to the sub-zone where the value is measured. This is modelled by the*hasIndicator* object property and their subclasses (one for each subclass of *Indicator*, i.e., *hasNumberOfWaterSupplies*, *hasLeaksTansportNetwork/100kmNetwork*, *hasHydraulicTechnicalPerformanceDistribution*, etc.). Therefore, a concrete zone is related to a subclass of *Indicator* by means of the corresponding subproperty of *hasIndicator* and its value, unit of measure and year are specified. Figure 4 shows a graphical representation of these relations for the indicator "Hydraulic technical performance distribution", while Table 2 presents the logical axioms used in the ontology. Figure 5 depicts an example of how the measure of the indicator "Hydraulic technical performance distribution" for sub-zone 2.4 in 2009 is modeled following the ontology axioms. The rest of indicators are modeled in a similar way.

Table 2. Modeling of "Hydraulic technical performance distribution".

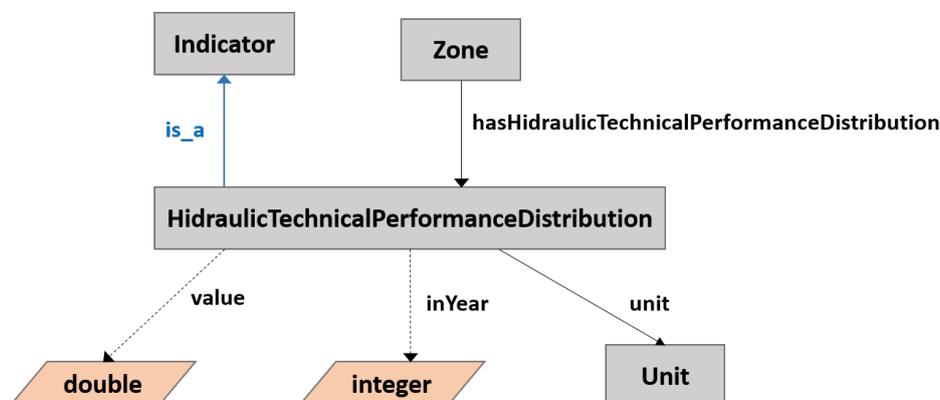| Object Property	Description Logic
hasHydraulicTechnicalPerformanceDistribution	$\sqsubseteq$ hasIndicator
	$\exists$ hasHydraulicTechnicalPerformanceDistribution.Thing
	$\sqsubseteq$ Zone
	$\top \sqsubseteq \forall$ hasHydraulicTechnicalPerformanceDistribution
	HydraulicTechnicalPerformanceDistribution
classes	Description Logic
HydraulicTechnicalPerformanceDistribution	$\sqsubseteq$ Indicator

Figure 4. Relations between Zones and Indicators. Example of "HydraulicTechnicalPerformance Distribution" modeling.
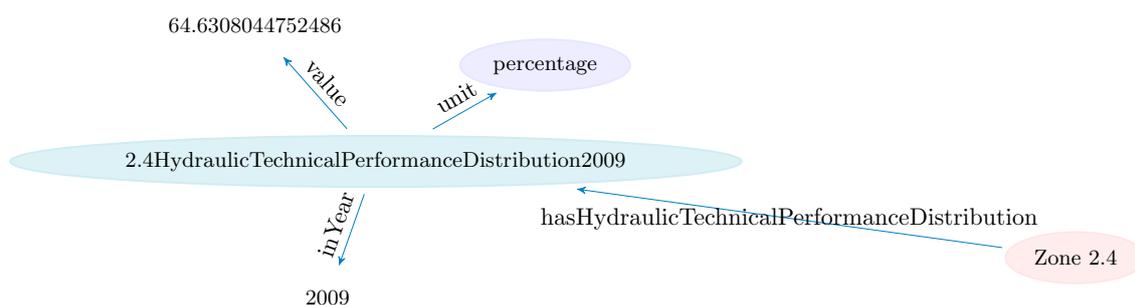
Figure 5. Example of graph representation according to the water Ontology.

4.2. Application to Water Supply Networks Management

Following the strategy and use case specifications commented before, a series of steps are now detailed to materialize the semantic model accordingly.

Step 1. Data mapping and pre-processing of sources: The specification of data sources is not intended to be exhaustive, but it just describes the most important points. The original XLS data format describes hydro-graphic zones, sub-zones and values, including indicators such as water leakages or structural breaks. The output is a machine-readable structured data file, using non proprietary formats, like CSV (Table 3). In our experiment, most of the data were anonymized due to privacy reasons.

Table 3. Data source information as plain text.

Zone	Measure1	Measure2	Measure3	Measure4	Measure5	Measure6	Measure7
Zone1.1	25,345,678	25,190,646	24,006,089	23,878,686	23,572,415	22,308,686	21,944,372
Zone1.2	524,121	441,808	317,843	254,109	255,378	171,530	206,667
Zone1.3	3,656,540	3,669,858	3,414,349	3,447,059	3,474,832	3,519,314	3,565,912
Zone1.4	-	-	-	-	-	-	-
Zone1.5	751,951	803,349	763,073	742,936	777,496	802,933	688,024
Zone1.6	2,399,197	2,393,628	2,372,762	2,356,697	2,430,938	2,309,204	2,277,539
Zone1.7	1,910,541	2,174,476	2,228,578	2,242,550	2,233,246	2,131,932	2,178,066
Zone1.8	5,087,383	4,588,837	4,580,388	4,901,428	4,870,611	4,885,952	4,818,310

Step 2. Data model generation: A parser has been implemented in Java, using Apache Jena (https://jena.apache.org/ (visited on 4 February 2019)) as a framework that applies mapping rules between the original dataset and our proposed ontology (classes, properties and relationships) (see the code at https://github.com/smartdataua/rdfwater (visited on 4 February 2019)). As a result, the RDF file is obtained containing all the original data mapped to the domain ontology specifying the classes, properties, and relationships.

The output dataset has been obtained following the pattern catalogue for modelling Linked Data [45]. Table 4 shows how the path to the resources is defined using explicit descriptions of the entities. The dots represent the common water prefix, while the asterisks represent a particular value. Table 5 describes the prefixes and the used namespaces in the dataset.

Table 4. Design patterns for the entities of the water supply company domain.

Entity	Pattern
Zone	. . . /zone/*
Water supply	. . . /hasNumberOfWaterSupplies/*
Water leaks	. . . /hasLeaksDistributionNetwork/*
Water not recorded	. . . /hasWaterNotRecorded/*
Length supply network	. . . /hasLengthSupplyNetwork/*

Table 5. Prefixes for namespaces.

Prefix	URI
void	http://www.w3.org/TR/void#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
owl	http://www.w3.org/2002/07/owl#
rdfs	http://www.w3.org/2000/01/rdf-schema#
water	http://www.khaos.es/ontologies/lucentia/ontologies/water.owl#
wd	http://www.wikidata.org/entity/
gn	http://www.geonames.org/ontology#

Step 3. Data storage: Several options to provide SPARQL access to the RDF storage were evaluated, including OpenLink Virtuoso (http://virtuoso.openlinksw.com (visited on 4 February 2019)), Eclipse RDF4J (http://www.rdf4j.org/ (visited on 4 February 2019)) and Stardog (https://www.stardog.com/ (visited on 4 February 2019)). The last one was selected in order to implement the access to the data, since it satisfied the requirements by supporting integrity constraint validation, batch indexing and reasoning.

As previously mentioned, to facilitate data reuse and interoperability, links to external resources have been defined. LOD repositories based specifically on the Valencian Community are limited to some extent, but some approaches such as datos.ign.es (datos.ign.es is an initiative accessible at the European Data Portal to provide the original geographic information from the Spanish Instituto Geográfico Nacional as LOD. http://datos.ign.es/ (visited on 19 December 2019)) provide geographic information as LOD. However, cross domain repositories and international geographic datasets can be used in this sense. Therefore, the zones and subzones were manually linked to Wikidata using the *owl:sameAs* relationship (see, for example, the example in Listing 1). These links contribute to the rich connectivity promoted by LOD. Finally, the data exportation to RDF (for instance, N3) is carried out. Wikidata acts as a hub, providing links to other datasets such as GeoNames. Thanks to the enrichment, additional properties can be used in order to provide contextual information in the SPARQL queries that were unavailable in the original data source, such as population, coordinate locations and administrative territorial entities.

Listing 1. External links generated in RDF/XML.

```
water:zone1
  owl:sameAs
  <http://www.wikidata.org/entity/Q935589> .
```

The RDF dataset has been evaluated using several methods: (i) nearly 40 constraints were defined and the data were validated against them using Stardog's Integrity Constraint Validation [46]; (ii) RDF data were validated by the W3C RDF validator (see http://www.w3.org/RDF/Validator (visited on 4 February 2019)); (iii) acceptance sampling and manual revision was performed on several hundreds of records; and (iv) a procedure was implemented testing that the number of zones and subzones matches the numbers in the original data source.

Step 4. Data exploitation: LOD technologies allow the consumption of data in a different way to traditional systems by using federated queries that include external indicators from knowledge bases such as Wikidata and GeoNames. SPARQL allows the creation of queries such as the example in Listing 2 to retrieve the length of the supply network by zone and year.

Based on the same dataset we have used in our experimentation, our study was focused on the analysis of potential relationships between indicators and enabling the discovery of goals that may be hidden [47]. Conversely, an initially expected relationship between the measure leaks and breaks in the network and water lost is not supported by the data. After the experimentation, the authors explained that it is necessary that the company reviews the way it is monitoring their goal. i.e., how the company is measuring breakdowns, or reviewing the suitability of the relationship, i.e., breakdowns do not

cause severe water loss. Other approaches can help in this matter, using external information in order to facilitate the decision-making process.

Listing 2. SPARQL query to list the length of the supply network by zone and year.

```
SELECT ?year ?value ?unit
WHERE {
 water:zona6.3 water:hasLengthSupplyNetwork ?length .
 ?length water:inYear ?year .
 ?length water:unit ?unit .
 ?length water:value ?value}
ORDER BY ?year
}
```

In this context, Listing 3 shows a SPARQL sentence that retrieves, as a result, the water supplied and the population from Wikidata (using the SERVICE instruction) in the particular zone 4.1. This sentence can be used in order to identify a possible relation between the water supplied and the population.

Listing 3. SPARQL federated query to retrieve the water supplied by year in a specific zone and the population from Wikidata.

```
SELECT ?waterSupplied ?population
WHERE {
 water:Zone4.1 water:hasWaterSupplied ?waterSupplied .
 water:Zone4.1 owl:sameas ?wikidataLink .
 ?waterSupplied water:inYear ?year .
 BIND(concat("", ?year) as ?yearTr)
 ?waterSupplied water:value ?value~.

 FILTER(regex(str(?wikidataLink), "wikidata"))

 SERVICE <https://query.wikidata.org/sparql> {
    ?wikidataLink p:P1082 ?populationStatement.
    ?populationStatement ps:P1082 ?population;
                         pq:P585 ?date .
    FILTER(xsd:integer(YEAR(?date)) = xsd:integer(?yearTr))
 }

}
```

In Figure 6, the results obtained with the SPARQL federated query can be shown. We studied the correlation between the water supplied and the population although few data were available in our case study. We used the Pearson's correlation [48] to measure the degree of variation between both variables. The correlation coefficient $R(5)$ was -0.6262, which means a moderate negative correlation. We analyzed if the correlation was significant, the value being $p = 0.1326$ (greater than the statistical significance levels of 0.05 and 0.10), and we concluded that the result was not statistically significant. Furthermore, we calculated the Spearman's correlation [49] obtaining a correlation coefficient $R_s = -0.5371$ and $p = 0.2152$ concluding that the relation between the two variables would not be considered statistically significant [50].
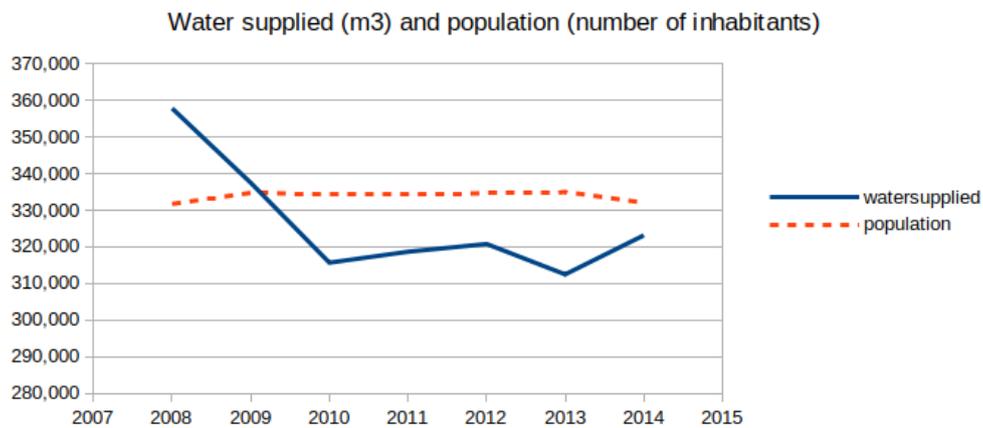
Figure 6. Supplied water and population of the zone 4.1. from 2008 to 2014.

4.3. Discussion

Linked Data facilitate the reuse and enrichment of data by using external sources, hence enabling the extension of the data models. As a result, data integration and exploration through complex relationships become more efficient and easier. Federated queries allow the execution of queries distributed over different SPARQL endpoints. In addition, these queries allow the inclusion of new non-existent attributes in the original data source. In our experiments, we have enriched the original data with external information from different repositories, such as Wikidata and GeoNames. The enrichment allows for filtering the results according to the new attributes of a specific zone (i.e., the population or the per capita income). Based on this new information, we will be able to extract new indicators, as well as new indicator correlations that will improve the decision-making process. For instance, a new correlation between the water supplied and the population, or water leaks of a specific zone could be discovered.

In this context, given that the information provided by the water company is accurate in terms of completeness, consistency and quality, the ontology model could help in identifying correlations between internal indicators and external information. For example, an increase of the temperature may cause an increase in water demand while an increase of per capita incomes in a specific zone may imply a higher demand of water. In the use case, water usage versus population change has been presented. Similarly, additional examples such as water usage versus temperature or water usage versus per capita incomes could be explored to find out hidden correlations. In this regard, the decision-making process can be enhanced by means of additional metrics, facts, or figures gathered from external datasets providing valuable insights. Furthermore, this information enrichment takes into account the context for making decisions. That is, for each zone, the manager could retrieve related external information in order to discover insights for each specific area. For example, the water consumption of *Zone 2.1* could be influenced by temperature because it is extremely hot; otherwise, in *Zone 8.2*, it could happen that temperature did not have a direct impact on water consumption since it is not an extreme temperature. In conclusion, the granularity allows us to customize each subzone with the relevant factors that influence water consumption.

With respect to data access and integration, many LOD repositories have recently appeared in a broad sense. However, there is still limited capacity regarding specific sectors when reusing the datasets due to different reasons such as the use of traditional formats (e.g., PDF) that are not accessible in a easy way. In this context, the gradually increasing use of LOD may help with the integration of different datasets. In addition, it is important to notice that, while the use of public SPARQL endpoints enable the integration of the internal information facilitated by the company with information provided by external repositories, its use requires IT skills.

It is worth noting that a possible comparison between the data extracted using the ontology-based model versus model without ontology, lies in the fact that the former provides a common framework in which the data are integrated and able to provide machine-readable queries (that can be easily adapted for each case), whereas in the latter case it always requires the reimplementation ad-hoc adaptations, since there is no a standard and fast methodology for data access and integration. Therefore, comparisons in this sense should focus on the time spent for data integration, which thanks to our ontology-based model is possible to avoid the manual reimplementation of these processes, just by adapting new SPARQL queries. This is advantageous from the point of view of the acquisition and availability of quality data.

We can conclude that LOD provides an innovative approach to identify factors that affect water consumption. The adoption and progress of water sustainability policies are crucial in order to meet the needs of the present without compromising those of future generations.

5. Conclusions

The ongoing growth of human population has a greater impact on water availability. Other aspects, such as climate change, are also in close relation with water scarcity. In this sense, the use of technology can help in order to identify key priority areas to strengthen action and possible strategies to reduce their impact.

In this context, we have proposed a novel approach focused on an efficient water management. This paper presents an ontology-based framework which allows for modelling, generating, integrating, publishing and exploiting data concerning real-world data from water supply networks management. The framework works in four steps: (1) data mapping and pre-processing of sources, (2) data model generation, (3) data storage and (4) data exploitation. Our framework is general and can be applied in diverse domains; here, the water management domain has been used.

The decision-making process in water management requires the integration of multiple and heterogeneous data sources and of different data domains. Linked Data is a set of web technologies based on the Semantic Web that enables consolidation of different data sources, as well as the efficient querying for feeding Business Intelligence processes. In this paper, we have applied our approach on a case study with real data from water supply networks. The main objective is to help the decision makers in order to obtain better results with the integrated and enriched information. In addition, the use case revealed that there is no correlation between the water supplied and the population by executing a SPARQL sentence that uses Wikidata as an additional repository.

The main novelties presented are the following: (1) the design and implementation of the domain ontology (water ontology) to capture, consolidate and integrate data from different water-related sources; (2) The application of the proposed architecture in the context of a real-world use case of a water supply network in the Mediterranean region of Valencia in Spain (the Spanish region had the highest average water consumption with a value of 164.43 liters per inhabitant per day during the years studied in the use case); (3) the concepts from the different data sources were integrated and stored in a common RDF repository, which has been subsequently validated; and (4) the links to external resources were used to enrich the original data in order to facilitate data reuse and interoperability. In our use case, the links to GeoNames and Wikidata by means of the *owl:sameAs* relationships allowed for adding contextual information to original data.

In the future, it is expected that more data providers will appear, thus enhancing the interlinking process and being helping in crucial tasks such as automatic disambiguation. It is also expected that the visibility of datasets will increase by means of the creation of new tools and applications using their metadata.

As future work, we foresee several opportunities to improve our work, such as including new knowledge bases with which to enrich the dataset. We also plan to integrate and reuse ontologies in order to include new concepts and relationships. In addition, developing new scoring methods with which to evaluate the quality of the datasets is crucial in order to promote their reuse by the community.

References

1. Joshi, B. *Novel Water Sustainability Technologies: Key Projects and Opportunities, Financing, and Venture Capital, Transactions and Trends*; BCC Research Report, MST074A; BCC Research: Wellesley, MA, USA, 2019.
2. Ponce Romero, J.M.; Hallett, S.H.; Jude, S. Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector. *Sustainability* **2017**, *9*, 2160. [CrossRef]
3. Vázquez Loaiza, J.P.; Pérez-Torres, A.; Díaz Contreras, K.M. Semantic Icons: A Sentiment Analysis as a Contribution to Sustainable Tourism. *Sustainability* **2019**, *11*, 4655. [CrossRef]
4. Gayoso-Cabada, J.; Goicoechea-de Jorge, M.; Gómez-Albarrán, M.; Sanz-Cabrerizo, A.; Sarasa-Cabezuelo, A.; Sierra, J.L. Ontology-Enhanced Educational Annotation Activities. *Sustainability* **2019**, *11*, 4455. [CrossRef]
5. Konys, A. Green Supplier Selection Criteria: From a Literature Review to a Comprehensive Knowledge Base. *Sustainability* **2019**, *11*, 4208. [CrossRef]
6. Curry, E.; Degeler, V.; Clifford, E.; Coakley, D. Linked Water Data For Water Information Management. In Proceedings of the 11th International Conference on Hydroinformatics (HIC 2014), New York, NY, USA, 17–21 August 2014.
7. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]
8. Barba-González, C.; García-Nieto, J.; del Mar Roldán-García, M.; Navas-Delgado, I.; Nebro, A.J.; Aldana-Montes, J.F. BIGOWL: Knowledge centered Big Data analytics. *Expert Syst. Appl.* **2019**, *115*, 543–556. [CrossRef]
9. Noy, N.F.; McGuinness, D.L. *Ontology Development 101: A Guide to Creating Your First, Ontology*; Stanford University: Stanford, CA, USA, 2001.
10. Guarino, N. Formal ontology and information systems. In Proceedings of the FOIS, Trento, Italy, 6–8 June 1998; Volume 98, pp. 81–97.
11. McBride, B. The resource description framework (RDF) and its vocabulary description language RDFS. In *Handbook on Ontologies*; Springer: Berlin, Germany, 2004; pp. 51–65.
12. Staab, S.; Studer, R. *Handbook on Ontologies*; Springer Science & Business Media: Berlin, Germany, 2013.
13. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]
14. McGuinness, D.L.; Van Harmelen, F. OWL web ontology language overview. *W3C Recomm.* **2004**, *10*, 2004.
15. Harris, S.; Seaborne, A.; Prud'hommeaux, E. SPARQL 1.1 query language. *W3C Recomm.* **2013**, *21*, 1–6.
16. Prud, E.; Seaborne, A. SPARQL query language for RDF. Available online: https://www.w3.org/2001/sw/DataAccess/rq23/ (accessed on 3 December 2019).
17. David, B.; Read, E.; Lucido, J.; Slawecki, T.; Young, T. An analysis of water data systems to inform the Open Water Data Initiative. *J. Am. Water Resour. Assoc.* **2016**, *52*, 845–858. [CrossRef]
18. Bianchini, D.; Antonellis, V.D.; Garda, M.; Melchiori, M. Exploiting Smart City Ontology and Citizens' Profiles for Urban Data Exploration. In *On the Move to Meaningful Internet Systems, Proceedings of the OTM 2018 Conferences—Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, 22–26 October 2018*; Proceedings, Part I; Lecture Notes in Computer Science; Panetto, H., Debruyne, C., Proper, H.A., Ardagna, C.A., Roman, D., Meersman, R., Eds.; Springer: Berlin, Germany, 2018; Volume 11229, pp. 372–389. [CrossRef]

19. Rani, M.; Alekh, S.; Bhardwaj, A.; Gupta, A.; Vyas, O.P. Ontology-based Classification and Analysis of non-emergency Smart-city Events. In Proceedings of the 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), New Delhi , India, 11–13 March 2016.

20. Goel, D.; Chaudhury, S.; Ghosh, H. Smart Water Management: An Ontology-Driven Context-Aware IoT Application. In *Pattern Recognition and Machine Intelligence, Proceedings of the 7th International Conference, PReMI 2017, Kolkata, India, 5–8 December 2017*; Lecture Notes in Computer Science; Shankar, B.U., Ghosh, K., Mandal, D.P., Ray, S.S., Zhang, D., Pal, S.K., Eds.; Springer: Berlin, Germany, 2017; Volume 10597, pp. 639–646. [CrossRef]

21. Ahmedi, L.; Jajaga, E.; Ahmedi, F. An Ontology Framework for Water Quality Management. In Proceedings of the 6th International Workshop on Semantic Sensor Networks co-located with the 12th International Semantic Web Conference (ISWC 2013), Sydney, Australia, 22 October 2013; Corcho, Ó., Henson, C.A., Barnaghi, P.M., Eds.; CEUR Workshop Proceedings; CEUR-WS: Aachen, Germany, 2013; Volume 1063, pp. 35–50.

22. Katsiri, E.; Makropoulos, C. An ontology framework for decentralized water management and analytics using wireless sensor networks. *Desalin. Water Treat.* **2016**, *57*, 26355–26368. [CrossRef]

23. de Rivera, D.S.; Robles, T.; López, J.A.; de Miguel, A.S.; de la Cruz, M.N.; Martínez, J.A.; Skarmeta, A.F.; Gómez, M.S.I. Adaptation of Ontology Sets for Water Related Scenarios Management with IoT Systems for a More Productive and Sustainable Agriculture Systems. In Proceedings of the SEMANTiCS 2017 Workshops co-Located with the 13th International Conference on Semantic Systems (SEMANTiCS 2017), Amsterdam, Netherlands, 11–14 September 2017; Fensel, A., Daniele, L., Aroyo, L., de Boer, V., Darányi, S., Elloumi, O., García-Castro, R., Hollink, L., Inel, O., Kuys, G., et al., Eds.; CEUR Workshop Proceedings; CEUR-WS: Aachen, Germany, 2017; Volume 2063.

24. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z.G. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, 6th International Semantic Web Conference, Proceedings of the2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, 11–15 November 2007*; Aberer, K., Choi, K., Noy, N.F., Allemang, D., Lee, K., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Germany, 2007; Volume 4825, pp. 722–735. [CrossRef]

25. Tanon, T.P.; Vrandecic, D.; Schaffert, S.; Steiner, T.; Pintscher, L. From Freebase to Wikidata: The Great Migration. In Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, 11–15 April 2016; pp. 1419–1428. [CrossRef]

26. Rebele, T.; Suchanek, F.M.; Hoffart, J.; Biega, J.; Kuzey, E.; Weikum, G. YAGO: A Multilingual Knowledge Base from Wikipedia, Wordnet, and Geonames. In Proceedings of the Semantic Web—ISWC 2016—15th International Semantic Web Conference, Kobe, Japan, 17–21 October 2016; Proceedings, Part II; Lecture Notes in Computer Science; Groth, P.T., Simperl, E., Gray, A.J.G., Sabou, M., Krötzsch, M., Lécué, F., Flöck, F., Gil, Y., Eds.; Springer: Berlin, Germany, 2016; Volume 9982, pp. 177–185. [CrossRef]

27. Färber, M.; Bartscherer, F.; Menne, C.; Rettinger, A. Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semant. Web* **2018**, *9*, 77–129. [CrossRef]

28. European Data Portal. Re-Using Open Data. 2017. Available online: https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf (accessed on 3 December 2019).

29. Cyganiak, R.; Reynolds, D. The RDF Data Cube Vocabulary. 2014. Available online: https://www.w3.org/TR/vocab-data-cube/ (accessed on 8 April 2018).

30. Bernadette Hyland, D.W. The Joy of Data—Cookbook for Publishing Linked Government Data on the Web. 2011. Available online: http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook (accessed on 3 December 2019).

31. Villazón-Terrazas, B.; Vilches-Blázquez, L.M.; Corcho, O.; Gómez-Pérez, A., Methodological Guidelines for Publishing Government Linked Data. In *Linking Government Data*; Wood, D., Ed.; Springer: New York, NY, USA, 2011; pp. 27–49. [CrossRef]

32. Hyland, B.; Atemezing, G.; Villazón-Terrazas, B. Best Practices for Publishing Linked Data. W3C Working Group Note. 9 January 2014. Available online: https://www.w3.org/TR/ld-bp/ (accessed on 20 May 2018).

33. International, O.K. Open Data Handbook: File Formats. 2012. Available online: http://opendatahandbook.org/guide/en/appendices/file-formats/ (accessed on 7 April 2018).

34. Bansal, S.K. Towards a Semantic Extract-Transform-Load (ETL) Framework for Big Data Integration. In Proceedings of the 2014 IEEE International Congress on Big Data, Anchorage, AK, USA, 27 June–2 July 2014; pp. 522–529. [CrossRef]

35. Musen, M.A. The protégé project: A look back and a look forward. *AI Matters* **2015**, *1*, 4–12. [CrossRef] [PubMed]

36. Pan, Z.; Zhu, T.; Liu, H.; Ning, H. A survey of RDF management technologies and benchmark datasets. *J. Ambient Intell. Humaniz. Comput.* **2018**, *9*, 1693–1704. [CrossRef]

37. Faye, D.; Curé, O.; Blin, G. A survey of RDF storage approaches. *ARIMA J.* **2012**, *15*, 11–35.

38. Acheson, E.; Sabbata, S.D.; Purves, R.S. A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Comput. Environ. Urban Syst.* **2017**, *64*, 309–320. [CrossRef]

39. Kontokostas, D.; Westphal, P.; Auer, S.; Hellmann, S.; Lehmann, J.; Cornelissen, R. Databugger: A test-driven framework for debugging the web of data. In Proceedings of the 23rd International World Wide Web Conference, WWW '14, Seoul, Korea, 7–11 April 2014; Companion Volume; Chung, C., Broder, A.Z., Shim, K., Suel, T., Eds.; pp. 115–118. [CrossRef]

40. W3C. Inference. Available online: https://www.w3.org/standards/semanticweb/inference (accessed on 20 November 2018).

41. Ren, X.; Curé, O.; Ke, L.; Lhez, J.; Belabbess, B.; Randriamalala, T.; Zheng, Y.; Képéklian, G. Strider: An Adaptive, Inference-enabled Distributed RDF Stream Processing Engine. *PVLDB* **2017**, *10*, 1905–1908. [CrossRef]

42. Colucci, S.; Donini, F.M.; Sciascio, E.D. Reasoning over RDF Knowledge Bases: Where We Are. In *AI*IA 2017 Advances in Artificial Intelligence, Proceesdings of the XVIth International Conference of the Italian Association for Artificial Intelligence, Bari, Italy, 14–17 November 2017*; Esposito, F., Basili, R., Ferilli, S., Lisi, F.A., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Germany, 2017; Volume 10640, pp. 243–255. [CrossRef]

43. Thivet, G.; Fernandez, S. *Water Demand Management: The Mediterranean Experience*; Technical Focus Paper; Global Water Partnership (GWP): Stockholm, Sweden, 2012.

44. INE. Estadística sobre el Suministro y Saneamiento del Agua, Año 2016. Notas de Prensa. Instituto Nacional de Estadística. 2018. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176834&menu=ultiDatos&idp=1254735976602 (accessed on 3 December 2019).

45. Dodds, L.; Davis, I. A Pattern Catalogue for Modelling, Publishing, and Consuming Linked Data. 2012. Available online: http://patterns.dataincubator.org (accessed on 6 November 2018).

46. Pérez-Urbina, H.; Sirin, E.; Clark, K. Validating RDF with OWL Integrity Constraints. 2012. Available online: https://www.stardog.com/docs/4.1.3/icv/icv-specification (accessed on 29 October 2018).

47. Maté, A.; Trujillo, J.; Mylopoulos, J. Key Performance Indicator Elicitation and Selection Through Conceptual Modelling. In Proceedings of the Conceptual Modeling—35th International Conference, ER 2016, Gifu, Japan, 14–17 November 2016; pp. 73–80. [CrossRef]

48. Pearson, K. VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. *Philos. Trans. R. Soc. Lond. Ser. A* **1896**, *187*, 253–318.

49. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, 72–101. [CrossRef]

50. Hauke, J.; Kossowski, T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaest. Geogr.* **2011**, *30*, 87–93. [CrossRef]