

Article

A Novel Approach for Publishing Linked Open Geodata from National Registries with the Use of Semantically Annotated Context Dependent Web Pages

Adam Iwaniak^{1,2}, Marta Leszczuk^{1,2,*}, Marek Strzelecki^{1,2}, Francis Harvey³ and Iwona Kaczmarek^{1,2} 

¹ The Wrocław Institute of Spatial Information and Artificial Intelligence, C.K. Norwida 9/2, 50-374 Wrocław, Poland; adam.iwaniak@wizipisi.pl (A.I.); marek.strzelecki@wizipisi.pl (M.S.); iwona.kaczmarek@wizipisi.pl (I.K.)

² Wrocław University of Environmental and Life Sciences, ul. C. K. Norwida 25, 50-375 Wrocław, Poland

³ Leibniz Institute for Regional Geography, Schongauerstraße 9, 04328 Leipzig, Germany; f_harvey@ifl-leipzig.de

* Correspondence: marta.leszczuk@wizipisi.pl; Tel.: +48-717-297-819

Received: 31 May 2017; Accepted: 10 August 2017; Published: 15 August 2017

Abstract: Many of the standards used to build spatial data infrastructure (SDI), such as Web Map Service (WMS) or Web Feature Service (WFS), have become outdated. They do not follow current web technology development and do not fully exploit its capabilities. Spatial data often remains available only through application programming interfaces (APIs), reflecting the persistence of organizational silos. The potential of the web for discovering knowledge hidden in data and discoverable through integration and fusion remains very difficult. This article presents a strategy to take advantage of these newer semantic web technologies for SDI. We describe the implementation of a public registry in the age of Web 3.0. Our goal is to convert existing geographic information systems (GIS) data into explicit knowledge that can be easily used for a variety of purposes. This turns SDI into a framework to utilize the many advantages of the web. In this paper we present the working prototype system developed for the province of Mazowieckie in Poland and describes the underlying concepts. Further development of this approach comes from using linked data (LD) with expert systems to support analysis functions and tasks.

Keywords: linked data; semantics; national registries; web services

1. Introduction

In recent years, the concepts of Internet development work have changed. The initial model of the Internet was a static web, referred to as Web 1.0. At that time, internet users were passive recipients of information published on the Web. Hyperlinks created by page designers, writers and webmasters were commonly used on the websites to connect related material. Following the Web 1.0 era, Web 2.0 emerged. Internet users were no longer just passive recipients, but active participants in the development of the web, including the creation of new links to already published documents.

In Web 2.0, data contained in “silos”, is provided via APIs. There are no connections between servers and no options to exchange data. Although data is available in a structured form, it is very difficult to integrate, due to its syntactical and semantical heterogeneity. The next step in the development of the Internet, Web 3.0, is the semantic web, which aims to use artificial intelligence to create links not only between websites, but also between resources published on the web [1]. The semantic web is based on a set of standards and best practices for sharing data and the semantics between applications.

The approach we describe, built on semantic web technology, also involves linked data concepts and approaches. The foundation of linked data is a Resource Description Framework (RDF) model that represents the resource, URIs to identify it and Hypertext Transfer Protocol (HTTP) as a mechanism for looking it up [2]. The simplicity and versatility of the RDF model and the use of global Uniform Resource Identifiers (URIs) are the basis for solutions to the problem of integrating data, which had only been provided via APIs in Web 2.0. The practical application of semantic technology, despite its rapid development, is difficult. Information retrieval requires proficiency in RDF and SPARQL Protocol and RDF Query Language (SPARQL) standards [3]. An example of a longer term project aimed at integrating geospatial datasets and enriching them with additional information from different sources, using the concept of linked data, is LinkedGeoData. The project is based on converting data from OpenStreetMap to RDF and linking it to external resources, such as DBpedia or Geonames. Access to the data in the form of downloadable files is made possible through a SPARQL endpoint [4]. Although the project implements the linked data concept, its immediate practical use from the point of view of the standard, non-expert web user is difficult. However, it is excellent for automated transactions involving web applications, including user defined queries.

Several attempts have been made to develop a system for effective linked data exploration, based on linguistic analysis of queries posted in natural language [3,5,6] or on combining this approach with conversion of the queries into a SPARQL template [7]. However, the adoption of such systems, demonstrating that they are very useful, especially in term of SDI, remains limited.

The connection of semantic web technology to SDI technology is not obvious. Many concepts and assumptions concerning SDI were created before the rise of the semantic web [8,9]. The solutions are not up-to-date for the current technology. In existing SDI, the web is mainly used as a connection between servers and clients. Web infrastructure, including search engines, is used to a very limited extent. For example, the numerous metadata records available through CSW (Catalog Service for the Web) are hardly explored by search engine crawlers [10,11]. Using the SDI technology based on Open Geospatial Consortium (OGC) Web Services (WMS, WFS), which are not World Wide Web Consortium (W3C) standards, requires user expertise. The use of these services by a broader community is still limited [10]. Up to now, SDI services often do not meet the expectations of users because of inefficiency in accessing data sets [8]. Web users want to view the data as an image, as well as download it and customize data displays and analysis through socket-based adapted programming for their requirements. A number of problems are very apparent in the process of developing central data repositories and storing the data derived from the local SDI level, e.g., cadastral databases. A good example can be found in Poland, where the development of an Integrated Real Estate Information System (ZSIN Project) for the country has been unsuccessful for more than 15 years due to the great heterogeneity of local databases and the resulting challenges of integration [12,13]. Additionally, increased demand for cadastral data causes the scope of cadastral databases to be constantly growing. The number of descriptive attributes, that are to be derived from other registers, increases in every new version of the databases. As a result, most of the values of these attributes are empty and the existing ones have poorly defined thematic accuracy. Further, there is a lack of metadata about the source of the attribute value.

The relevance and significance of SDI technology however, remains due to developing technologies and the impact of political goals and expectations. In 2009, then US President Barack Obama signed a memorandum [14] on the introduction of changes in US public administration policy towards an open government policy and open, structured data. Following on this, the European Union developed a European Digital Agenda for open government policy [15]. The main goal of these activities is to stimulate the economy by using open data and, as a result, to create value added data and services, including using linked data approaches.

The question relevant for SDI development in conjunction to evolving policies is: do we need to change the approach to the creation of public registers in connection with newer SDI web development? Thus, an important issue is how to use Semantic Web technology to move from the implicit knowledge gathered in geoportals and turn it into explicit knowledge that is structured and easily searched

on the web. For example, the location of a specified area zoned for multi-family housing is now available through analysis of certain layers in geoportals, but it cannot be easily searched with Web 2.0 tools. Therefore, we need to consider how to change the WMS layers from geoportal data objects to web-based knowledge available to all potential users. How do we develop direct answers to user queries and support their applications? How do we design and maintain a system built on linked data?

In this article we address these questions through the presentation of a system concept we have implemented for publishing data and metadata from official or other registers or databases using the RDF data model, as linked data. The main objectives of the system are:

- Obtaining widespread knowledge currently only available in geoportals
- Facilitating information retrieval on the web
- Providing automatic update of data using ontology (up to date solution)
- Assuring a high quality system.

The system concept involves the representation of various user perspectives on cadastral data and some other useful features, especially SOoI—Spatial Object of Interest, and analogue to PoI—Point of Interest, derived from public registers on various perspectives. The desired result with this approach is a significant increase both in the ability to get a direct answer to users' questions and the usefulness of linked data compared to previous Web 1.0 and Web 2.0 data access technologies.

2. Methods

2.1. System Assumptions

The system we describe publishes data and metadata, derived among others from public registers, based on linked data and Semantic Web technology. The proposed concept publishes SOoI data as HTML documents with semantic annotations in RDFa format. Building websites for each SOoI allows them to be indexed by search engines and thus increases the ability for their retrieval on the web.

The key element in creating SDI 3.0 is the conversion of data to a RDF model and linking it with other spatial and non-spatial registers. The repository, created after data conversion, is used to automatically generate HTML documents with RDFa annotations for selected objects that are relevant to a given user perspective, or for whole classes of objects, such as parcels or buildings. HTML pages can be contextual. This means that depending on the purpose and user perspective, the same object has several representations of HTML and RDFa documents. HTML documents have added value, generated by the expert system, such as the results of spatial analysis, links to geoportals and also keywords, based on web user questions. The system concept provides capabilities for:

- Selection of Spatial Objects of Interest (SOoI)
- Development of domain ontology
- Conversion of resources to a RDF data model
- Semantic data enrichment
- Creation of semantically annotated HTML documents, linked e.g., with geoportals
- Indexing published HTML documents by search engines
- Retrieval of the resources on the web and finding the answers for user questions.

In the case of spatial data in a linked data-based system, it is essential to give distinct global and stable identifiers to all spatial objects so that they can be uniquely identified. According to the linked data principles such an identifier is a URI (Uniform Resource Identifier). It allows for the unambiguous identification of resources published on the web and their connection to external data sources, resulting in enriched data through a wide-range of linking and integration operations. The reference to the URI should provide useful information about the resource [2]. This means that when typing an URI in a web browser, the HTML document or the resource described in the RDF should be returned in response. Dereferencing of URI is possible by use of HTTP. The basic rule is that the URI should remain

as long as possible—“Cool URIs don’t change” [16]. For this reason, it is recommended that identifiers be as short as possible, and that they are neutral in meaning and language [17].

Today, geoportals also have their identifiers in the form of URLs. It is often possible to generate such an address for a particular map view, e.g., for a city boundary. However, such links are not indexed by search engines. They cannot be easily searched for so they are not available to the average web user. The approach proposed in the article is based on the assumption, that each object (SOoI) has its own URI and is represented as an HTML document, so it can be indexed by search engines.

2.2. RDF Model, RDFa Annotations, Domain Ontology

The standard model for data interchange on the Semantic Web is RDF, as recommended by the W3C [18]. It is a directed graph. Resources in linked data are represented by RDF triples which consists of a subject, a predicate, and an object. The subject is the resource identified by the URI, the object is another resource (also external) or the specific attribute value. The predicate defines the kind of relationship between subject and object, allowing for a proper interpretation of such a relationship [19]. The concept of the Semantic Web is to publish the data as both human-readable and machine-readable. This is possible by publishing resources as HTML documents with semantic RDFa annotations.

RDFa provides a set of HTML attributes that make it possible to combine human-readable text with machine-readable data without duplicating the content of the document in separate data formats. Data published in this way can be indexed by search engines and is readable for applications that consume and process RDF data. The RDF data model allows for representation of the semantics of objects, e.g., spatial and temporal relations, semantic similarity or trajectories [1]. This is possible with the use of domain ontology—a formal, explicit specification of a concept of a real-world model, designed for a particular purpose [20].

Ontologies allow for the standardization of resource description by providing terminology used to define classes, properties and relationships between them [19]. They decrease the chance for misinterpreting the information, that are “hidden” in the data, support interdisciplinarity and facilitate an integration of data from different sources, breaking down silos in the form of API [1]. The main idea supporting the use of ontology is the ability to re-use domain specific terms, which ensures interoperability of resources published on the Semantic Web. The most commonly used ontologies for describing data published on the Web as Linked Open Data (LOD), such as Dublin Core for describing metadata terms or WGS84 Basic Geo for representing the coordinates of points, are available on the Linked Open Vocabularies (LOV) project site. The use of common ontologies can also serve to build a mechanism (in an application for example) to automatically update HTML pages with RDFa annotations based on changes made to the source dataset.

2.3. System Architecture

The basic concept of the system derived from system assumptions is shown in Figure 1. It was further developed into the system architecture, with consideration of specific technical approaches and solutions. The components of the system are shown in Figure 2. Both spatial and non-spatial data are integrated and converted into a RDF model with the use of Semantic Components software, which opens the possibility of converting spatial data sources, in this case data from National Registries, into RDF representation according to commonly used spatial ontologies. Data in the RDF model is enriched with the result of spatial analysis, performed with the use of custom SPARQL extensions, such as distance to strategic points, neighborhood analysis, insolation analysis, and so on. The data is enriched with links to other RDF resources, such as Geonames, DBpedia, as well as with dereferenced links to the geoportals on which the resource can be found, which metadata was also converted into RDF for the purpose of linking. These operations are performed in spatial relationship and external links enrichment modules, which is strictly connected to a RDF repository and extends SPARQL engine capabilities with additional properties and SPARQL filter properties which can be

used in SPARQL queries. Spatial enrichment (linking spatial data) is performed by executing prepared SPARQL UPDATE query sets, with results recorded into a distinct subgraph.

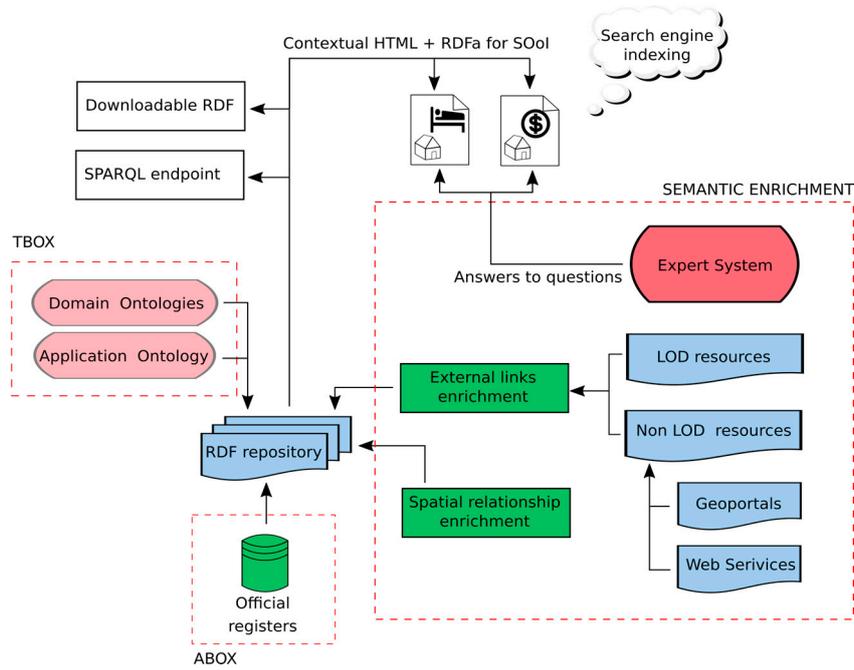


Figure 1. Concept of system architecture.

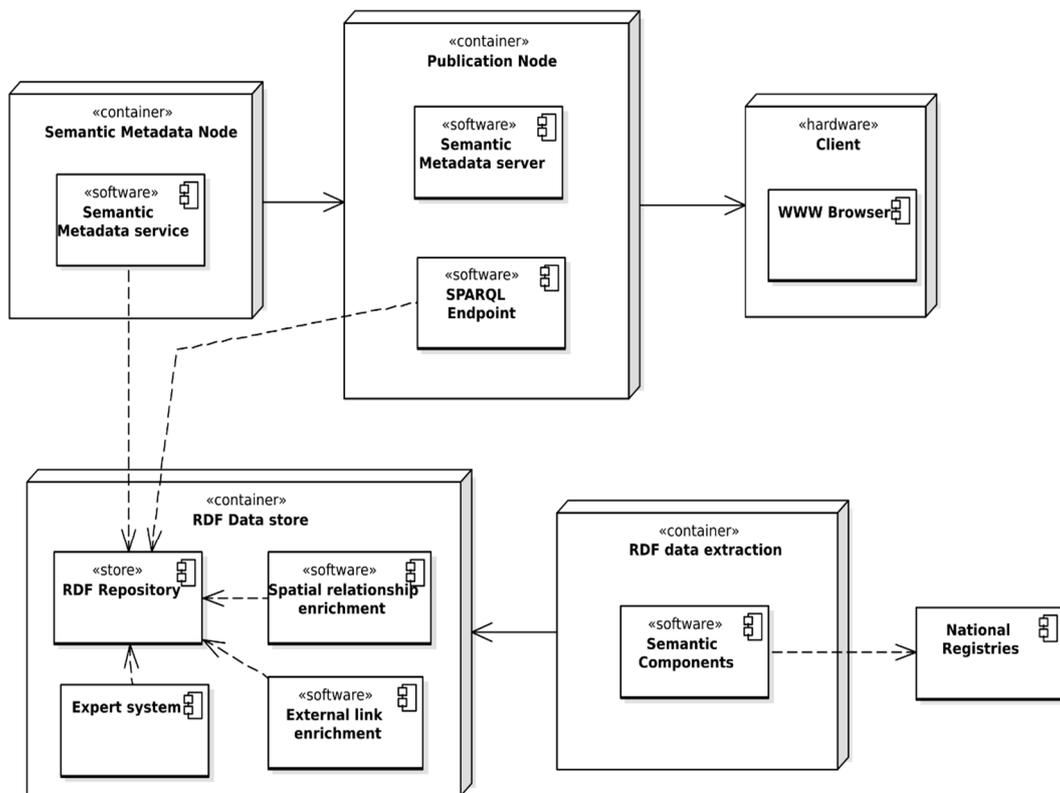


Figure 2. Component model.

For publication and updating the data in the RDF model, an ontology that uses common domain ontologies and defines its own classes and properties, has been developed. An ontology enables automatic updating of HTML + RDFa pages based on changes made on the institutions websites or in the Public Information Bulletin. The necessary condition for this is the inclusion of the appropriate classes, properties and relationships that allow the automatic generation of HTML document content using the set of attributes provided by the RDFa.

Based on RDF data for selected, real-world entities (SOoI), HTML documents with semantic RDFa annotations are created. The entities in the SOoI become data objects, that may be potentially searched on the web using a range of criteria, e.g., property for sale, for development, for recreational purposes, and so on. For publication of public registers, HTML documents can be created for all objects in a given class.

One object may have several representations as HTML + RDFa documents—pages are generated for specific user groups. For example, the same building can be treated as a property in one document or as accommodation for tourism purposes in another. An expert system generates keywords based on questions posted by users, such as “building near forest”, “parcel of area greater than 1000 square meters”, “for housing purpose”, “full utilities”, appropriately defined and then available for each representation of the object.

Documents containing both human and machine readable information in the form of HTML + RDFa documents are created in the Semantic Metadata module, implemented as a service, which takes the RDF data from repository along with application ontology which serves as a configuration for the process. The service pushes generated metadata documents into a publication service. The reason behind generating HTML + RDFa documents on the server side, instead of using modern and popular approaches based mostly on JavaScript solutions, is the need for web crawlers to process the static Document Object Model (DOM) tree of the page, with special consideration of RDFa annotations.

The publication service, which has the capabilities of a traditional static document HTTP server, also provides a SPARQL endpoint for the possibility of executing queries provided by external clients to share results created in the system. It is also possible to serve both HTML + RDFa metadata documents and RDF documents as downloadable files. HTML documents are then indexed by search engines and thus become visible and searchable on the web. With this approach, the user, typing his query in the web search engine, has the chance to find the direct, explicit answer for a question quickly.

2.4. Semantic Enrichment

Data enrichment can be defined as a method of utilizing the information units carried by data objects. Semantic enrichment is a process of improving the quality of the semantic data by adding new, semantically connected content to existing dataset in order to make the data more discoverable and useful for potential users. By utilizing specific methods of data matching and processing, it is possible to bring a new value to existing datasets, which in many cases are sets of “raw data”. In fact, data semantic enrichment is often the process of reasoning on semantically connected datasets, discovering facts, detecting dependencies and similarities between resources, thus extracting knowledge from data using various methods. The main purpose of providing semantically enriched data in the form of knowledge is to provide users with an answer to domain-specific questions, where the aggregation, merger and then interpretation of heterogeneous datasets is crucial.

2.4.1. Linked Open Data Cloud Integration

The fundamental and axiomatic principle of linked data is the provision of linkages between resources published on the Web with the use of specific technologies. It is a process of utilizing existing LD data sources to acquire external context for individual resources. The use of common ontologies and vocabularies provide the necessary common understanding of concepts and relations between them. In this case, the dataset is enriched with links to the resources available on the Web. In principle,

the process of linking two resources is simple, it can be accomplished by adding a RDF triple with two URIs of the resources and one URI of the property representing the meaning of the connection. However, the process of matching is never trivial. Resource matching can be done manually, which is an acceptable approach only for small datasets with few resources to encode. The more valid solution for this problem is to use a systematic and automated approach, which utilizes specific features of the resources, but in most cases matching depends largely on lexical similarity between resource labels.

The process of matching spatial objects can utilize one of the most important features of real world objects—spatial reference, which highly improves the quality of the connections. One of the possible strategies for matching spatial objects for the purpose of publishing Linked Open Data is presented in [21]. In this approach, we selected well-known LOD hubs—DBPedia and GeoNames. In the prototype implementation of this approach, for the purpose of external link enrichment another set of custom SPARQL functions were implemented, mostly for the purpose of measuring lexical similarity between two specified RDF resources using Damerau-Levenshtein distance and a bag-of-words approach. Predefined SPARQL queries are able to find connections between resources having similar `rdfs:label`, `dc:title` and `dc:keyword` values, which requires special preprocessing of external RDF resources. Choosing the best method for linking resources based not only on lexical similarity but also considering semantic similarity by using lexical databases such as WordNet, is a task that could still be improved.

2.4.2. Use of Non-LD External Web Services

Apart from linking LD resources with other published LD resources, it is also essential to provide links to other non-RDF resources published on the Web, in this case both standardized web geoservices such as WFS and custom geoportals or RESTful APIs, with the ability to display and provide information about specific spatial objects, with the use of prepared a HTTP GET query, which can make resources referable. With links to geoportals it is possible to provide new views of described and connected data for users, but in most cases it is not possible to extract new information. With links to structured information about resources such as GML, GeoJSON or custom JSON representation, it is possible to do so, however, unlike the extraction of data from RDF LOD resources, it is essential to support serialization or the format of resources provided by those web services. To be able to provide links to external geodata sources it is essential to process and interpret any kind of metadata. For the prototype, metadata about WFS and geoportals (which operates on WMS layers) services was obtained by processing GetCapabilities responses, extracting titles, keywords and spatial extent information. After converting such information into RDF triples it was possible to use SPARQL queries with combined spatial relationship and lexical similarity functions to link resources with specific external geodata sources.

2.4.3. Spatial Relationships

The third aspect of semantic enrichment presented in this approach is spatially linking data, which allows users to explore and query datasets based on topological relations connecting them. With such links it is possible to find and query neighboring, intersecting, inner and outer resources in terms of spatial reference and geometrical relations. According to GeoSPARQL standards, the support for basic OGC Simple Features topological relationships functions (`sfIntersects`, `sfOverlaps`, `sfTouches`, `sfWithin`, `sfContains`) as well as other non-topological functions such as distance and buffer has been implemented to perform basic spatial analysis at SPARQL level. Those relations are computed before publication of the datasets by executing predefined queries. After the materialization process, triples representing spatial relationships are used directly in HTML + RDFa document generation module. This way the linked data published as HTML documents are explorable with the use of simple links representing spatial relationship between resources, as opposed to the need for querying it manually by each user with spatial queries with GeoSPARQL or WFS, which requires additional knowledge.

2.4.4. Expert System

One of the possible approaches of extended semantic enrichment for the purpose of linked open data publication both in machine and human readable formats as HTML pages with RDFa annotations is the use of expert system modules. Publishing the data from national registries, such as cadastral or administrative unit registries with the use of RDFa standards allows agencies, organizations and regular users to publish information about resources in the form of human readable documents with machine readable annotations and links to other external or internal resources.

Linked Open Data is generally available as “raw” RDF datasets, transformed directly from registry databases with no particular application in mind. Publishing such information with HTML + RDFa is often designed for specified actors, such as citizens, businesspersons, or administration officials to answer a specific question which will use the system in different way. This approach requires the LOD resources to be semantically enriched and prepared to fit the requirements. The main idea of the approach that is presented is to implement semantic enrichment as a means of providing links not only to external resources in a LOD manner (cadastral parcels linked to specific zoning plan) and links to other spatially connected resources (parcel linked to neighbor parcels or road nearby) but also to provide direct answers for questions frequently asked by a specific actor, such as time of arrival to the city center, distance to the nearest parks or if the parcel is available to buy. Use cases of the system specified for various actors are shown in Figure 3.

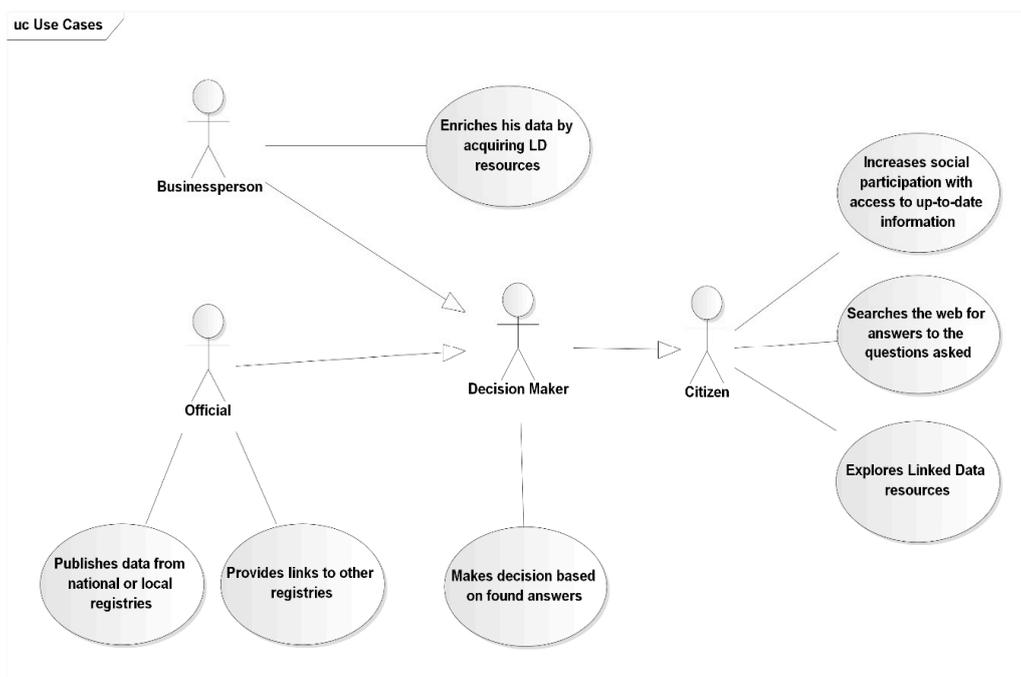


Figure 3. Use cases of the system.

In the approach we used for the Polish functional pilot project, frequently asked questions and answers were created semi-automatically with the supervision of a domain expert who manually constructed rules in the form of SPARQL Updated queries provide basic reasoning capabilities with the use of Closed World Assumptions (CWA), but in the current version of the prototype such process will be automated with the use of an expert system consisting of external modules implementing more advanced methods, which will require well defined, domain ontology. The expert system is still an area where improvements can be made and it could be the subject of separate research. The detailed description of the expert system will be published in later work.

2.4.5. The Metadata Hierarchy

An important aspect of linked data is the ability to create and use a metadata hierarchy. The metadata of spatial objects (SOoI) can inherit metadata from a specified dataset, allowing for enrichment of information about the published resource. The values of metadata elements include their origin or the time interval for which they are current. In other words, individual elements have their own metadata, which may be useful in many analysis and selection operations. The benefit of linked data concept is the ability to interrelate this information through reference to other datasets—“metadata for metadata”. Such information can be valuable to the user—it allows him or her to evaluate data reliability. An example of metadata elements for a cadastral parcel for sale, for which the values come from various sources is presented in Table 1.

A cadastral parcel in this system can be described by attributes such as number, area, destination, price, distance to strategic points in the field, time availability, and so on. Some of them, such as number or area, come from official information, gathered by the appropriate office, such as the Board of Surveying, Cartography and Municipal Cadastre, while others may come from field-specific documents (e.g., zoning plans) or may be the result of spatial analysis. The source of origin determines the reliability of the information. In addition to the source of origin, the time interval for which information is up-to-date is also important. For example, when giving time from the land parcel to a strategic point, such as the city center, it is important that this time is estimated for a particular hour as at other times of the day the travel time will be probably different. The linked data concept allows the presentation of this information by including relevant properties in the domain ontology and using these to create links between metadata elements.

Table 1. The metadata hierarchy in the prototype. The first column represents the name of sample metadata element, describing the cadastral parcel. The second column represents the sample source of the value of this metadata element, which is metadata for metadata indeed.

Name of Metadata Element	Example Metadata of the Element Value
Number of the cadastral parcel	Metadata from Board of Surveying, Cartography and Municipal Cadastre
Area	Metadata from Board of Surveying, Cartography and Municipal Cadastre or calculated on the basis of spatial analysis
Destination of parcel	Zoning Plan
Parcels price	Last notary transaction (date, name of notary, etc.)
Distance to the parcel from city center	Calculated on the basis of spatial analysis on Google Maps
Time availability from the parcel to a city center	Calculated on the basis of spatial analysis on Google Maps at 08:00

3. Case Study

3.1. Prototype of the System

We developed a working prototype of the system for evaluation. As part of the solution, the knowledge base has been created based on descriptive logic (Due to the fact that Polish data is published, the knowledge base has also been implemented in Polish. For the purpose of this paper, selected parts of the knowledge base and web pages have been translated into English). It allows an exploration of the knowledge of the data and thus, to convert implicit knowledge into explicit knowledge, accessible directly in the form of answers to user queries. It enables inference, detection of dependencies and similarities between resources and discovering new facts. The knowledge base consists of an assertion box (ABox) describing facts or instances (data) and a terminological box (TBox), containing definitions and descriptions of concepts in the form of domain ontologies. The system provides the opportunity to publish both spatial and non-spatial data from The Office of the Marshal of the Mazowieckie Voivodeship (Poland) as linked data. The source data includes:

- The National Register of Boundaries, containing the boundaries of administrative units in Mazowieckie Voivodeship, basic information about these units, such as name, code or type, and address points of the offices of local government units (shapefiles and GML file).
- The data, concerning the offices of local government units in Mazowieckie Voivodeship, including heads of those offices (an excel file). In the file, attributes such as name and type of the office, contact details including address, telephone number, e-mail name of the head of the office, name and type of the administrative unit, are recorded.
- Public e-services, provided by the offices of local government units (an excel file and data from official web pages). These include attributes such as the name of the e-service, classification, service recipients, fees, required documents or the responsible office.
- Metadata for spatial planning documents, complies with the standards ISO 19115 and ISO 19139 (XML files).
- Metadata for local geoportals (an excel file and XML metadata, concerning OGC web services, that are available through these geoportals—a response from a GetCapabilities request). The information such as name and URL of the geoportal, available OGC web services and layers, responsible office or bounding box of data, are recorded.

The data has been converted into an RDF model using the software package “Semantic Components”. The conversion is based on a mapping of the projects, which includes mapping between datasets and RDF types, and also between attributes and RDF properties. The geometry of spatial data is represented as Well-known Text (WKT) using GeoSPARQL standards and for points such as latitude and longitude coordinates using Basic Geo ontology.

Internal and external connections between data have been made using spatial analysis tools, available through GIS software as well as lexical and semantic analysis. The external connections include data sources such as:

- The National Register of Geographical Names (locality and physiographic objects), published in RDF model
- Geonames
- DBpedia
- The Public Information Bulletin

A class diagram presenting the main relationships between the published resources is shown in Figure 4. The links to the National Register of Geographical Names and Geonames was created based on spatial analysis. Objects from these resources, that are located in the area of the described resource are indicated and linked using properties: “seeInLocality(PRNG)”, “seeInPhysiographicObjects (PRNG)” and “seeInGeonames” properties. The links to the DBpedia are created based on lexical and semantic analysis, using the appropriate SPARQL query. They are represented as “seeInDBpedia” property. The connections to the Public Information Bulletin are based on lexical analysis and the URL (Uniform Resource Locator) pattern for this type of resource by substituting the part of the pattern with the proper element. The property “hasURL” derived from the vCard ontology [22] was used to represent this relationship.

For the purpose of data publication and extracting the knowledge from it, domain ontologies have been developed, that use other existing domain ontologies, such as vCard, Org [23], FOAF [24], GeoSPARQL [25], Basic Geo (WGS84 lat/lon) [26] or Dublin Core [27] to ensure interoperability and also define specific classes and properties. The ontologies have been created using an open source software, Protégé Desktop v. 5.2.0 and are represented in OWL 2 DL (Web Ontology Language). They contain the following main classes: administrative unit, address point, office of the local government unit (as subclass of FormalOrganization from Org ontology), head of the local government unit (as subclass of Person from FOAF ontology), metadata for the local geoportal, metadata for spatial planning documents, and public e-service. The relationships between classes are defined, so that it is

possible to perform inference. Part of the ontology for the National Register of Boundaries is shown in Figure 5.

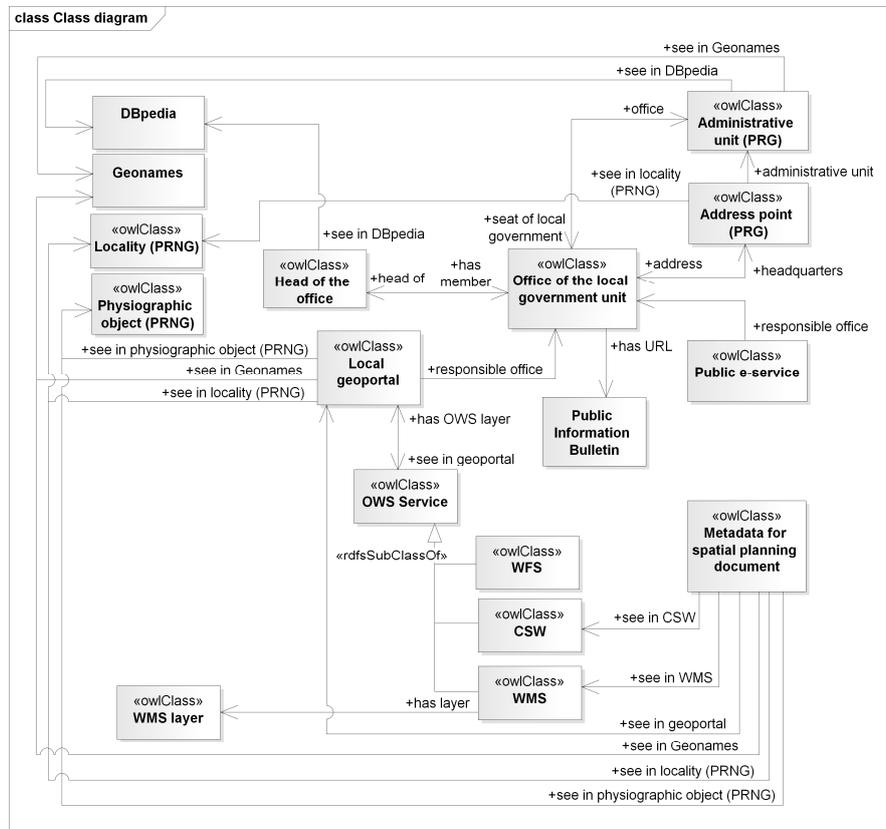


Figure 4. Class diagram showing relations between main ontology classes for prototype of the system.

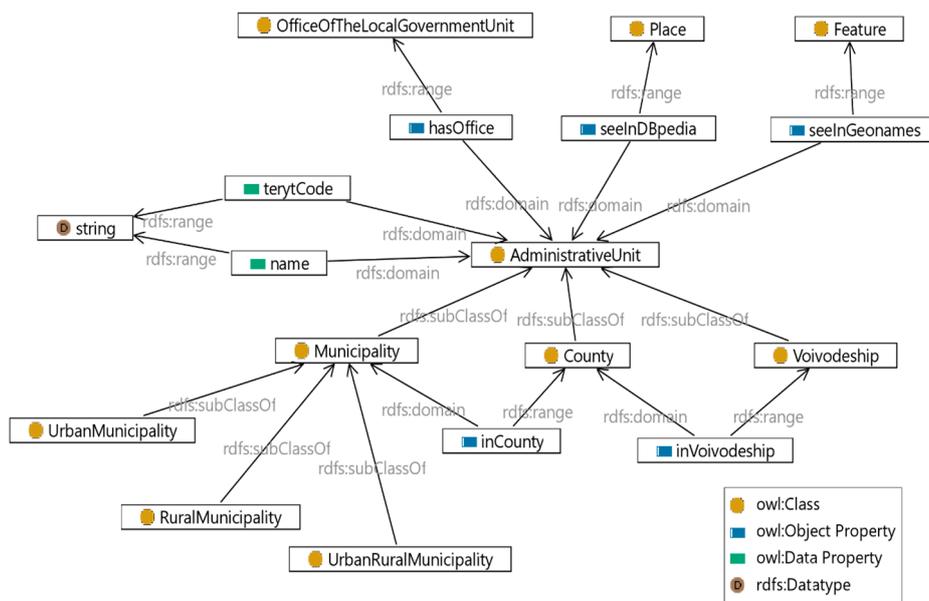


Figure 5. A part of the domain ontology for the National Register of Boundaries (visualization created with TopBraid Composer).

The data and domain ontologies, included in the knowledge base, are stored in Apache Jena Fuseki and are available through a SPARQL endpoint.

For selected sets of data in the RDF data model, HTML documents with semantic annotation in RDFa are generated and published on the web. These datasets are: the metadata for spatial planning documents, the metadata for local geoportals and the metadata for public e-services. The example of RDFa annotations included in HTML documents and concerning the title and the subject of selected resource is shown below:

```
<div class="form-row">
  <span class="form-label">Title: </span>
  <span class="form-value">
    <span class="sa_title" property="http://purl.org/dc/terms/title"> Local spatial development plan
of the city of Białobrzegi no. XV / 81/2004</span>
  </span>
</div>
<div class="form-row">
  <span class="form-label">Category: </span>
  <span class="form-value">
    <span class="lang_code_category">
      <a property="http://purl.org/dc/terms/subject"
href="http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/planningCadastre"
target="_blank">http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/planningCadastre </a>
    </span>
  </span>
</div>
```

The HTML documents contain all the information about resources available in the RDF repository, starting with the title of the resource or contact details and ending with the links to external resources such as Geonames or the National Register of Geographical Names. If the document concerns spatial data, a bounding box is also available and displayed on the map. The sample HTML document, containing metadata for local spatial development planning is shown on Figure 6. Based on keywords, the HTML documents are indexed by search engines. They retrieve keywords from the relevant elements of the HTML document and place them in a specially separated section of the search engine database, which increases the chances for finding a page by web user. The use of common ontologies and RDFa annotation also provides better searchability by enabling semantic search. This makes the information accessible through commonplace web browsers and for staff without special training.

3.2. Usability Assessment

The main advantage of this approach for users is the fact that it simplifies access to important information from search engine results as annotated web pages. It avoids reliance on SDI published resources available only through dedicated web application with advanced search techniques, which are often too complicated for regular users. The web pages produced by this approach contain two important pieces of information: a brief description of the resource, where the information scope is different for each use case and it contains links to other web pages—internal and external. The web page works as a hub of information about specific resources that the person is interested in. In terms of its usability, it involves a simple process of using commonly available search engines to find the resources published by public administration with the use of this service and then by using available links to access the information distributed across the Web. In this case the only thing to evaluate is the importance of the web pages, which can be measured by their position in the search results of search engines, although this process requires time to observe the popularity of individual resources.

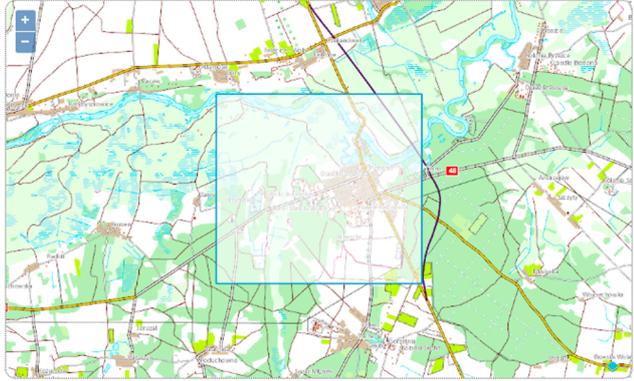
The other use of this service is the aggregation of information in a LOD manner with the use of RDFa annotations. The published resources can be consumed by web crawlers, which can convert published information and make use of it later, hopefully publishing the results of the process in a way

that can be used afterwards by our service. The use of a RDF model in the form of RDFa is the most generic approach to acquire structured information from the Web and the only methods of evaluation are to monitor how many connections (in a LOD manner) are available to our resources or how many applications were made using it. However, it also requires the time for users to find our resources, start trusting them and later use it in their own applications. Therefore, further study is needed to evaluate the usability and resource consumption based on data about usage beyond limited use during the prototype period.

(a)


www.mazovia.pl

Local spatial development plan of the city of Białobrzegi no. XV / 81/2004
 SEMANTIC METADATA



Abstract

Local Spatial Development Plan is the basis for spatial planning and is a planning document prepared for the municipality area in accordance with the Planning and Land Use Planning Act of 27 March 2003 (Journal of Laws 2003 No. 80 item 717, as amended). It establishes rules commonly applicable in the area, which are the basis for issuing administrative decisions. Legal basis - Resolution No. XV / 81/2004 of the Council of the Town and Commune of Białobrzegi.

Selected locations in the area

Białobrzegi , Pierzchnia , Białobrzegi

(b)

TOPOGRAPHIC OBJECTS INCLUDED IN THE RESOURCE AREA (GEONAMES):

Place / place / topographic object:	
Place / location identifier:	7533565
Name of the place / location:	Białobrzegi
Resource URI:	http://sws.geonames.org/7533565/
Place / place / topographic object:	
Site / location identifier:	762277
Name of location / location:	Pierzchnia
Resource URI:	http://sws.geonames.org/762277/
Place / place / topographic object:	
Site / location identifier:	776114
Name of the place / location:	Białobrzegi
Resource URI:	http://sws.geonames.org/776114/

Figure 6. Fragments of HTML document, containing metadata for local spatial development plan: (a) title, basic information and bounding box of data; (b) links to Geonames features, included in the resource area.

3.3. The Pros and Cons of the System

In assessing the working prototype we identified some advantages and disadvantages. The advantages of the system are:

- The system is complementary, it contains links to geoportals that can be dereferenced
- The added value of resources is created by combining them with other data sources
- Ease of data integration
- Easy to search the web, due to indexing by search engines

- Automatic updating of content on HTML pages, due to the use of ontology and artificial intelligence
- Consistency of information published on the web
- High scalability of the system
- Unified way of analysis using SPARQL query language
- Getting direct answers to specific user questions
- RDF supports native linking.

The disadvantages of the system are:

- Limited support for searches of occurrences of widely used keywords, e.g., “land registry”
- High dependency on search engine robots for indexing keywords.

4. Conclusions

The development of web technologies and politics support changing approaches to the construction of SDI through public registers and linked data technology. The project takes advantage of more recent semantic web technology developments to convert implicit knowledge to explicit knowledge. This follows linked data and Web 3.0 concepts and occurs through dereferencing and hierarchical metadata. The functionality of the website update and automatic generation of website content based on ontology supports high level ease of access and maintenance. It also allows for the creation of additional websites tailored to users’ needs and published in various contexts. In this sense, the development of this SDI technology can contribute to the strengthening of social participation in spatial planning. Citizens, for example, can report their comments at various stages of creating spatial planning acts in a web-based document that can automatically connect with other linked data objects. Publishing resources as HTML + RDFa documents makes them easier to find and provides up-to-date information related to them, especially their legal status, for example in a planning process. The ease of expansion and high scalability of the system will be tested in ongoing and future work. In the longer term, we aim to further develop AI approaches to support specific operational functions.

Acknowledgments: This research was supported by the National Science Centre, Poland—project number 2012/05/B/HS4/04197. The examples in the article were prepared using software “Semantic Components” created under the project “Development and Implementation of Innovative GeoMedia Enterprise Intelligence Technology Implementing Multi-Criteria Analysis with Spatial Data in the Desktop Environment and on the Web”, financed by the National Centre for Research and Development, Poland.

Author Contributions: Adam Iwaniak and Marek Strzelecki designed the system concept; Marek Strzelecki, Marta Leszczuk and Adam Iwaniak created an ontology and implemented a prototype of the system; Marta Leszczuk, Francis Harvey, Marek Strzelecki, Adam Iwaniak and Iwona Kaczmarek wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janowicz, K.; Scheider, S.; Pehle, T.; Hart, G. Geospatial semantics and linked spatiotemporal data—past, present, and future. *Semant. Web* **2012**, *3*, 321–332. [CrossRef]
2. Berners-Lee, T. Linked Data—Design Issues. Available online: <https://www.w3.org/DesignIssues/LinkedData.html> (accessed on 15 May 2017).
3. Shekarpour, S.; Marx, E.; Ngonga Ngomo, A.C.; Auer, S. SINA: Semantic interpretation of user queries for question answering on interlinked data. *J. Web Semant.* **2015**, *30*, 39–51. [CrossRef]
4. Stadler, C.; Lehmann, J.; Höffner, K.; Auer, S. LinkedGeoData: A core for a Web of spatial open data. *Semant. Web* **2012**, *3*, 333–354.
5. Ding, L.; Finin, T.; Joshi, A.; Pan, R.; Cost, R.S.; Peng, Y.; Reddivari, P.; Doshi, V.; Sachs, J. Swoogle: A semantic web search and metadata engine. In Proceedings of the 13th ACM International Conference on Information and Knowledge Management, Washington, DC, USA, 8–13 November 2004; pp. 652–659.
6. Lopez, V.; Fernandez, M.; Motta, E.; Stielor, N. PowerAqua: Supporting users in querying and exploring the Semantic Web. *Semant. Web* **2012**, *3*, 249–265. [CrossRef]

7. Unger, C.; Bühmann, L.; Lehmann, J.; Ngonga Ngomo, A.-C.; Gerber, D.; Cimiano, P. Template-based question answering over RDF data. In Proceedings of the 21st International Conference on World Wide Web—WWW '12, Lyon, France, 16–20 April 2012; pp. 639–648.
8. Sadeghi-Niaraki, A.; Rajabifard, A.; Kim, K.; Seo, J. Ontology based SDI to facilitate spatially enabled society. In Proceedings of the GSDI 12 World Conference, Singapore, 19–22 October 2010; pp. 19–22.
9. Patroumpas, K.; Georgomanolis, N.; Stratiotis, T.; Alexakis, M.; Athanasiou, S. Exposing INSPIRE on the Semantic Web. *Web Semant. Sci. Serv. Agents World Wide Web* **2015**, *35*, 53–62. [[CrossRef](#)]
10. Yue, P.; Guo, X.; Zhang, M.; Jiang, L.; Zhai, X. Linked data and SDI: The case on web geoprocessing workflows. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 245–257. [[CrossRef](#)]
11. Lopez-Pellicer, F.J.; Florczyk, A.J.; Nogueras-Iso, J.; Muro-Medrano, P.R.; Zarazaga-Soria, F.J. Exposing CSW catalogues as linked data. In *Geospatial Thinking*; Springer: Berlin/Heilderberg, Germany, 2010; pp. 183–200.
12. Kapuściński, A. Od integrującej platformy elektronicznej do zintegrowanego systemu informacji o nieruchomościach-podsumowanie projektów pilotażowych. *Rocz. Geomatyki* **2012**, *10*, 43–55.
13. Widawska, Z. Stan prac nad zintegrowanym systemem katastralnym w świetle potrzeb leśnictwa. *Rocz. Geomatyki-Ann. Geomat.* **2006**, *4*, 133–144.
14. Obama, B. *Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies*; The White House: Washington, DC, USA, 2009.
15. European Commission. *Communication from the Commission of 19 May 2010 to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—A Digital Agenda for Europe*; European Commission: Brussels, Belgium, 2010.
16. Berners-Lee, T. Cool URIs Don't Change. Available online: <https://www.w3.org/Provider/Style/URI> (accessed on 17 May 2017).
17. Hyvönen, E. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012; Volume 2.
18. RDF 1.1 Primer: W3C Working Group Note 25 February 2014. Available online: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/> (accessed on 25 May 2017).
19. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data—The story so far. *Int. J. Semant. Web Inf. Syst.* **2009**, *5*, 1–22. [[CrossRef](#)]
20. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [[CrossRef](#)]
21. Iwaniak, A.; Kaczmarek, I.; Strzelecki, M.; Lukowicz, J.; Jankowski, P. Enriching and improving the quality of linked data with GIS. *Open Geosci.* **2016**, *8*, 323–336. [[CrossRef](#)]
22. Iannella, R.; McKinney, J. vCard Ontology—For Describing People and Organizations: W3C Interest Group Note 22 May 2014. Available online: www.w3.org/TR/vcard-rdf/ (accessed on 4 August 2017).
23. Reynolds, D. The Organization Ontology: W3C Recommendation 16 January 2014. Available online: <https://www.w3.org/TR/vocab-org/> (accessed on 4 August 2017).
24. Brickley, D.; Miller, L. FOAF Vocabulary Specification 0.99: Namespace Document 14 January 2014—Paddington Edition. Available online: <http://xmlns.com/foaf/spec/> (accessed on 4 August 2017).
25. GeoSPARQL—A Geographic Query Language for RDF Data. Available online: http://schemas.opengis.net/geosparql/1.0/geosparql_vocab_all.rdf (accessed on 4 August 2017).
26. Basic Geo (WGS84 lat/long) Vocabulary: W3C Semantic Web Interest Group. Available online: <https://www.w3.org/2003/01/geo/> (accessed on 4 August 2017).
27. Dublin Core Metadata Initiative (DCMI) Recommendations. Available online: <http://dublincore.org/specifications/> (accessed on 4 August 2017).

