Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

GI Systems for Public Health with an Ontology Based Approach

Nurefşan Gür
Universitat Jaume I – INIT, Institute of New Imaging Technologies, Avda. Sos Baynat, s/n
Castellón, Spain
nurefsan.gur@uji.es

Laura Diaz Sanchez
Universitat Jaume I – INIT, Institute of New Imaging Technologies, Avda. Sos Baynat, s/n
Castellón, Spain
diazl@uji.es

Tomi Kauppinen
University of Münster – IFGI, Institute for Geoinformatics
Weseler Strasse, 253
Münster, Germany
tomi.kauppinen@uni-muenster.de

Abstract

Public health research brings substantial benefit to society. Finding data in an efficient way is necessary to carry out relevant space time analysis to address particular studies such as mortality rates and their causes for instance due to environmental exposure. However, health-related information systems remain isolated from other systems such as those managing geospatial and environmental information making it difficult and time-consuming to study relations and patterns in multidisciplinary scenarios. In this paper we employ Linked Open Data technologies to publish health and related data. We report results of the approach with a case study to expose mortality atlas data of the Valencia Community in Spain. The results show how to overcome the lack of semantic relations between resources and published data, and how to reduce disparity and redundancy.

Keywords: Linked Open Data, Health, Geo-spatial data

1 Introduction

Health is an indispensable attribute of human life. To improve our understanding of relations between human health and their environment within the context of genetic diversity, causation and variation, complex methodologies and multidisciplinary studies are required. Users and experts need appropriate information infrastructures to exploit the potential of cross-disciplinary topics and questions, for example finding links between health and pollution. To achieve this, methods and techniques are needed to interconnect data from different disciplines such as health, environmental, socio-economic, statistical, and geospatial technologies and semantic web communities. Linked Data[1] accelerates the pace of discovery with the main aim of integrating these resources in a transparent and communicable way so that they can be used to identify trends and relevant events.

In this paper, we report the creation of a semantics-based health information system by creation of Linked Data from existing raw data sets.

First in Section 2 we describe our plan starting with the theoretical background how the study is inspired through underlining the problems and contribution. We describe envisioned methodological background in Section 3 in order to draw a model for converting and interlinking relevant data sets with the attributes of statistical, spatial and temporal health related data. This is done by defining the needed semantic relations with corresponding ontologies in Section 4. We finalize the paper with concluding remarks.

[1] Linked Data refers to a set of best practices for publishing and connecting structured data on the web, see http://linkeddata.org.

2 Motivation

In many research fields – from genetics and molecular biology to social sciences – data sharing is already ingrained in how researchers work [12]. The early example of using GI Science by John Snow to show the relation between water supply and cholera outbreaks in London, 1854 was achieved by the use of public data to link between contaminated water and disease [16]. Human genome project is also completed by global cooperation based on sharing data [7]. In many countries, especially in US, UK, and in some other EU Countries, formally sharing data on governmental or non-governmental platforms is becoming a common practice [19].

Noticeable importance and prevalence of public data and open data, lead us to underline the significance of open data integration with an ontology-based approach in context of the spatio-temporal health statistics.

2.1 Problem

Hans Rosling, health scientist-turned-statistical-public speaker talks about a rhetoric disease called DbHd: Database Hugging Disorder in his open data speech at World Bank (May, 2010). Formerly to this talk World Bank cleared itself from this disease by making World Bank data publicly available (April, 2010) [5]. Despite the common practices, culture of open data is not widely embraced yet. Much of the infrastructures, technical standards, and incentives that are needed to support data sharing are lacking, and these data can hold particular sensitivities [11]. In the scope of this study we will elaborate technical constraints and solutions, referencing the fact that well-established repositories and tools enable researchers to access and interrogate shared data resources, and build on one another’s work [4].

Disparate data sources, unstructured data and several data format are available on the web, as data silos require a common structured hierarchy for a global database in order to communicate each other. This necessity brought the idea of designing a virtual documentation system for data sharing on the sky and applying to the web of data by Linked Data Principles[2] outlined by Berners-Lee (2006) as follows:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

2.2 Contribution

Location-based information means information that is *immediately relevant*, which is the essence of the next generation of WWW- Semantic Web[3]. "Particular diseases are prone to occur in certain places" with respect to geographic, economic, wealth and environmental profiles of the location [9]. Some of the location related health information and data across other disciplines is listed in Table 1. Hence location matters for health; the capability of GIS for visualizing and analyzing spatial data is used, with ontology based approach and a semantic flavor with regard to Linked Data Principles[2].

Table 1: Examples of location specific health information.

Location Specific Health/Medical Information
- Local disease rates
- Life expectancy according to social-economical profiles
- Mortality rates at certain locations with certain profiles
- Addresses of local health care facilities
- Local weather, pollen and air quality alerts
- Local health risks and hazards
- Targeted health education
- Travelers' health information
- Local health news
- Local drugs/drug trade names and prices
- Health sciences articles published as a case study for a specific location

Source: Extended table from International Health Geographics 2003, 2 [9].

Compared to traditional relational databases and tables, linked data has the potential to provide the infrastructure for the Web of Data on a large-scale, distributed and accessible manner. In order to encourage data providers, research agencies, governmental organizations to use this technology pilot models, examples and tutorials[4] are needed and yet being developed and published by academic institutions, governments, private organizations and individual entrepreneurs [10]. Particular case studies for a certain scope of topics are more favourable in order to draw a path for similar cases. This study takes as a case study the mortality rate data observed for 10 years in region of Valencia for different districts, classified by gender and cause of death as diseases. Statistical data is aggregated with different trends based on time and space and combinations of these two variables.

This study's contribution is to display the process of data handling, ontology designation for geographical, statistical and health disciplines; and deployment of a model for multidisciplinary reuse of information.

3 Methodological Background

The Linking Open Data project[5] is a visible example of the integration of diverse data that covers significant portions of health sciences; genes, proteins, drugs and clinical trials [10,1] as well as statistical and geographical data [17]. Use case data requires to be prepared by following Linked Data Principles to able to communicate with other data sources. URI - Unified Resource Identifiers are used for identifying the use-case data (attribute names, attribute values) that gives more generic means to identify existing entities. These entities are constructed with `http://` scheme, which can be dereferenced on the Web by the HTTP protocol. HTTP URIs carries importance for establishing the single data model as RDF (Resource Description Framework) for publishing structured data on the web. Whilst HTML as a dominant document format linking documents on the web, RDF utilizes linking data on web. It allows to formulate statements about resources where each statement is consisting of subject-object-predicate. Subject and object are both URIs describing an object and predicate is also a URI defining the relations between subject and object. Predication is to say something about the subject and a good practice for choosing predicates is considered to reuse existing ontologies and vocabularies like RDF, RDFS (Resource Description Framework/Schema), OWL (Ontology Web Language), SIOC (Semantically Interlinked Online Communities), FOAF (Friend of a Friend), Dublin Core[6] etc. for defining interoperable relations [17].

Gruber has given the widely cited definition of ontology: "ontology is an explicit specification of conceptualization" [18]. Ontologies, schemas and vocabularies, all mean roughly the same thing, which are defining RDF information about RDF information [8]. Study topic covers spatio-temporal health statistics as mentioned earlier in Section 2. In order to convert the data into RDF triples, existing ontologies and vocabularies are elaborated, grouped and listed in Table 2. Choosing ontologies is a deliberate process to define the data objects correctly. For instance coordinates, statistically aggregated values or a date are expressed by numbers. Though they require more elucidation than a numeric type like decimal, integer or float to be communicable with other data

[2] http://www.w3.org/DesignIssues/LinkedData.html
[3] http://www.w3.org/2001/sw/
[4] http://linkedscience.org/
http://data-gov.tw.rpi.edu/wiki/
http://learnlinkeddata.com/
http://data.southampton.ac.uk/

[5] http://www.w3.org/CommunityProjects/LinkingOpenData
[6] http://www.w3.org/RDF/
http://www.w3.org/TR/rdf-schema/
http://www.w3.org/TR/owl-features/
http://rdfs.org/sioc/spec/
http://www.foaf-project.org/
http://dublincore.org/

Table 2: Domain Specific Ontologies

Domain	Ontology	Predicates
Spatial	prefix gn: <http://www.geonames.org/ontology#>	Interlinking with `<owl:sameAs>`
	prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>	geo:lat, geo:long
	prefix dcterm: <http://purl.org/dc/terms/>	dcterm:location
	prefix tisc: <http://observedchange.com/tisc/ns/#>	tisc:areasize
Time	prefix time: <http://www.w3.org/2006/time#>	time:year
Health	prefix MeSH: <http://www.nlm.nih.gov/cgi/mesh/2012/ >	Interlinking with `<owl:sameAs>`
	prefix Diseasome: <http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseases/>	
	prefix dbpedia :<http://dbpedia.org/resource/>	
Statistical	prefix qb: <http://purl.org/linked-data/cube#>	qb:dimension qb:slice qb:item
	prefix scovo: <http://purl.org/NET/scovo#>	

sets. Another considerate process is interlinking data objects similarly to unique name assumption (UNA) approach. In logics with the unique name assumption, different names always refer to different entities in the world [15]. Instead of making this assumption OWL constructs a predicate to express two names refers the same things in a mutually compatible way. For this purpose `owl:sameAs`[7] is the OWL predicate that asserts two URIs refer to same entity. This linkage maintains to reference for the same data object from different information providers.

4 Data Management

Data management for establishing a semantic system is the backbone of the study [3]. Efficient management of RDF data is an important factor in realizing the semantic vision [3]. This section firstly introduces the relational data model and explains the progress of RDB to RDF mapping. Second part describes linking the different datasets on a conceptual level and underlines semantics between those datasets.

4.1 Structural Data Management

Initial data is received as several data formats; first, several files aggregated with statistical methods on a time span for certain diseases are provided in raw R-data[8], second, geospatial data in shape-file format[9]. R-data files are an output of statistical calculations. Files were organized into groups according to some key attributes like gender, disease and encoded with place IDs. What is emphasized here is that the data was syntactically structured but not semantically. This type of data doesn't allow to be consumed by others due to lack of metadata and is not suitable for interlinking with other datasets. The data is clear just for the ones who produce and analyze it however limited for cross-domain analyses and usage. The only way to consume this data is to visualize on graphs, charts and maps dynamically without the ability to be upheld. First step of data management is to carry the data up to advanced levels and show the bidirectional functioning between those. Due to the pervasiveness of relational databases a logical schema is drawn and data transferred into a relational database. Sample fields of the tables are shown in Figure 1. The problem domain is expressed in logical schema within relation to entities through primary key and foreign keys as the observations in spatio-temporal trends table are for certain cities, for specific diseases. Observations denotes mortality trends (spatio-temporally aggregated statistical values) for each disease with variables; time, gender and place. Metadata can be asserted to this logical data model though semantic description of entities in problem space will remain absent.

Figure 1: Logical Data Model

CITIES TABLE

C_ID	Name	ZIP_Code	Population	Area	GeonamesID	Latitude	Longtitude	Polygon
77	Benicassim	12028	9037	36039979.8	6356985	40.0577061	0.04280985	"0.08,40.08|...

DISEASES TABLE

DiseaseID	Label	DiseaseName_ES	DiseaseName_EN	MeshID	DBPedia_Name	DiseasomeID
16	Colon	Tumor maligno de colon	Colonic Neoplasms	D003110	Colorectal_Cancer	1914

SPATIO-TEMPORAL TRENDS TABLE

Obs_ID	C_ID	DiseaseID	Year	Gender	SpatioTemporal_Probablity	SpatioTemporal_Mean
14	77	16	2001	W	0.49401197605	99.91804655165

This relational schema can be rendered to RDF by assigning primary keys as unique identifiers, which will serve as subject of RDF triples whereas when they are foreign keys of a table they serve as the object of the triple. Subjects and objects are linked through meaningful URIs called predicates given in Table 2, raise the semantic description.

Each entity in the tables is identified by URI's instead of their names or primary keys. Sample data in Figure 1 can be observed as URIs in Figure 2 grouped under directories according to the range of table names; diseases, cities and spatial trends.

Openlink Virtuoso assigned as triple store that provides built-in data loader utilities from existing data sources (relational databases, CSV files, XML etc.) for generating linked data [11]. RDF Views created from the imported structural data, outlines a path for linked data wrapping. Ontologies can be altered and metadata is enhanced through built-in wizard during this process.

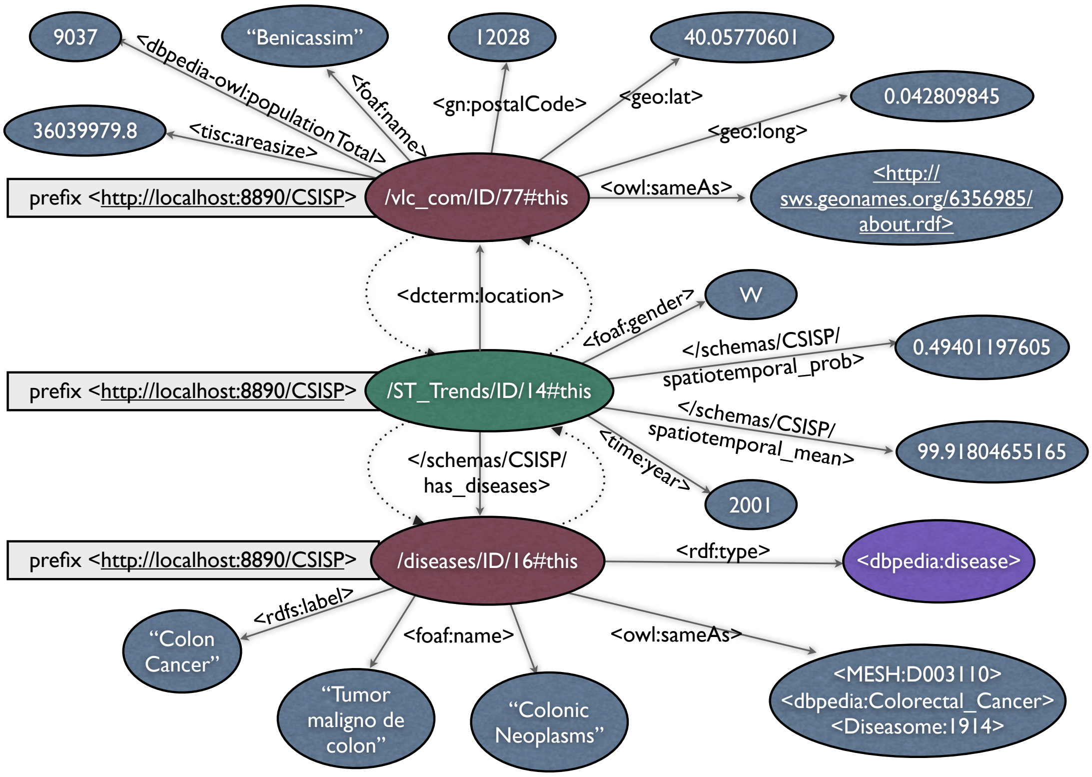
[7] `owl:sameAs` statement indicates that two URI references actually refer to the same thing. http://www.w3.org/TR/owl-ref/

[8] Rdata is a lazy load database data that provides an object related information dynamically when it's actually needed by instantiating the related objects and its properties when it's requested. http://www.r-project.org/

[9] http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 2: Conceptual Data Model



4.2 Content Based Data Management

Relationships between two entities are not explicit in the model at Figure 1. Foreign keys relating two tables don't precisely express the nature of relationship.

The necessity of vocabularies in domain of health can be exemplified by referring to the interpretation of a disease name in medicine terminology and different languages. A disease has various names within the medicine terminology and for different languages. It also exists as a linkable data object on the web of data in MeSH[10] vocabulary, DBPedia[11] and Diseasome[12] network among many others. Linking the data to these existing ontologies with the predicate `owl:sameAs` enriches the local data. By usage of this predicate two data, which doesn't know each other with a link reference will implicitly known by the rest of the data, linked to the subject.

Heritable disorders of the disease, related diseases and medicine, causes and treatments, associated genes are some of the data exists in these three ontologies linked to our data. Attribute values ID's and unique names that are stored in the fields of the diseases table at Figure 1, for health related ontologies are appended to their prefixes, which is listed under the health domain at Table 2. The same process is done for the cities table entities' matching with Geonames[13]. Extent of the data covers Valencia, an autonomous community defined as a first level administrative division, with its 541 cities/municipalities, which are third level administrative divisions. Place name (toponym) information differ in the quality of their associated data, such as the feature types (e.g. city vs. municipality, community vs. state) and the spatial footprints (i.e. reference points, coordinates), there is need for a form of conflation [6]. For this reason multiple sources are merged so that the different aspects of each source can be exploited. Extraction of the rdf dump from Geonames ontology is limited to the extent of Valencia region and then string matching algorithms are used for generating RDF links which resulted as almost fully match of the entities, for 540 cities of 541. Statistical data is of paramount importance in this data integration study. Semantics of the statistical data or in other means RDF representation is fairly different than usual methods. A single statistical data item is described in a multi-dimensional way by using statistical vocabularies; *cube* and *scovo* as mentioned in Table 2. However there exist studies and tools [2] are based on scovo vocabulary it's noted on the homepage[14] as it's depreciated and directs to use RDF Cube Vocabulary [14].

The cube model consists of three main components; dimensions, measures and attributes. Dimensions identify statistical observations in terms of concepts like gender, region, time and such. Measures represent the phenomena

[10] http://www.nlm.nih.gov/mesh/
[11] http://dbpedia.org/
[12] http://www4.wiwiss.fu-berlin.de/diseasome/
[13] http://www.geonames.org/
[14] http://vocab.deri.ie/scovo

observed and attributes qualify and interpret the observed values in terms of the units of measures or scaling factors. In the context of the use-case sample of statistical data components are mapped as follows:

- Dimensions; Cities of Valencia region, 10 year time span and gender {F, M}
- Measure: Mortality rates
- Attributes: Probability distribution and mean calculation methods or unit measures of spatio-temporal trends. (Optional)

Creating data structure definitions for the spatio-temporal trends table shown at Figure 1 is sorted in the following listings. SDMX- Statistical Data and Metadata Exchange[15] standard provides content oriented guidelines, which define a set of common statistical concepts and associated code lists that are intended to be reusable across data sets as cube vocabulary re-use in the examples. Listing 1 is an example of how to define dimensions as RDF links. The abbreviation used for cube vocabulary is `qb`. New predicates defined in the domain of CSISP are represented within the range of relations previously depicted in Figure 2. The second listing describes components; measure gives the value of each individual observation attached by different observed values for different calculation methods (probability distribution and mean) for the represented phenomena i.e. Mortality rates.

Listing 1: Dimensions

```
<CSISP:vlc_com>
  rdf:Property qb:DimensionProperty;
  rdfs:subPropertyOf sdmx-dimension:RefArea;
  qb:concept sdmx-concept:RefArea;
  rdfs:range dcterm:Location.
<CSISP:Gender>
   rdf:Property qb:DimensionProperty;
   rdfs:subPropertyOf sdmx-dimension:sex;
   rdfs:range foaf:gender.
<CSISP:Time>
   rdf:Property qb:DimensionProperty;
   rdfs:subPropertyOf sdmx-
dimension:refTime;
   qb:concept sdmx-concept:RefTime;
   rdfs:range owl-time:year.
```

Listing 2: Measures

```
<CSISP:MortalityRate>
   rdf:Property qb:MeasureProperty;
   rdfs:subPropertyOf sdmx-measure:obsValue;
   rdfs:range xsd:decimal;
   qb:measure CSISP-measure:stProbablity;
   qb:measure CSISP-measure:stMean.
```

Optionally, attributes, which clarify the unit measures or statistical methods for measures, are given as an example in Listing 3. These components are used to re-define the dataset with stronger semantic relations which are likely consistent with other statistical domains. Finally in Listing 4, sample data is shown as a `qb:observation` by its dimensions and measures.

[15] http://sdmx.org/

Listing 3: Attributes

```
<CSISP-measure:stProbablity>
rdf:Property qb:AttributeProperty;
qb:attribute
dbpedia:Probability_distribution.
<CSISP-measure:stMean>
  rdf:Property qb:AttributeProperty;
qb:attribute dbpedia:Mean.
```

Listing 4: Observation

```
</ST_Trends/ID/14/this#>
  rdf:Property qb:Observation;
  qb:Dataset  CSISP:ST_Trends;
  CSISP:vlc_com </vlc_com/ID/77/this#>;
  CSISP:Gender sdmx-code:sex-W;
  CSISP:Time 2001;
  CSISP-measure:stProbablity 0.49401197605;
  CSISP-measure:stMean 99.91804655165.
```

Cube vocabulary supports the creation of slices by grouping observations with fixed dimensions to reduce the verbosity of the dataset and guide consumer applications [14]. An example of creating slices would be by creating time series from ST_Trends table. Another dimension of the slice can be chosen as regions. This way of designation of the slice allows to note a change in measurement process, which affects a particular time or region[16].

5 User Access Methods

Users, developers, researchers, anyone who would like to use this data can access it through various interfaces. SPARQL endpoint serves the data like a proxy service to Linked Data frontends. SPARQL endpoint can be accessed directly through a HTML browser and complex queries can be sent in order to receive comprehensive replies with various formats. Additionally, SPARQL endpoint allows to be accessed from third party applications or frameworks for statistical analysis and can be used in developing applications with map, timeline and graphic visualizations. (i.e.: R-Project SPARQL package, Pivotviewer, Openlink AJAX Toolkit, SIMILE-EXHIBIT Framework[17].)

6 Conclusion

Relational database and Linked Data integration carries an important value for enterprises. It alleviates of heterogonous data integration challenges and allows the public to discover linked data as well as create data mash-ups. Overall interoperability and extensibility is assured as RDF allows

[16] Especially time dimension is important in the case of infectious diseases like influenza, the study of their geographic distribution frequently involves examining the diffusion of the disease through space over a given period of *time* (spatio-temporal mapping and analysis) [9].

[17] http://cran.r-project.org/web/packages/SPARQL/SPARQL.pdf
http://www.openlinksw.com/PivotViewer
http://oat.openlinksw.com/
http://www.simile-widgets.org/exhibit/

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

integration of data in a platform-independent manner. Added value of Linked Data is achieved by building an RDF schema over the traditional way of storing data.

The absence of semantic relation for the logical data model is compensated by an RDF based conceptual model that utilizes the semantic expressivity of RDF, RDFS, OWL, FOAF and domain related ontologies.

Statistical data is shaped in a form that represents more than just numbers but also advanced meta-data with space and time dimensions of the observation through this ontological approach. Spatial health data for mortality rates is carried over into a system with an ontological based approach that can be used in several visualization technologies and integrated with the web of data for complex queries. Our approach for the model retains relational databases and web services for application integration, visualization and deploys linked data for disparate data mashing, discovery and drill down analysis.

7 Future Work

In this final section we round the paper off with directions for future implementations and studies. Linking the health data with environmental data is stated as a promising inducement in Section 1.Etiology of mortality and morbidity based on environment can be caused from following examples of environmental data and sources; toxic release inventories, air pollutant data like ozone, sulfur dioxide, CO2, pesticide exposures, water reservoirs and safe drinking water information sources [20]. After studying the relations between health and environmental data geographic scale of the study needs to be specified. Following this pattern aggregated environmental hazard or exposure data can be correlated with aggregated health data for each unit of observation by geographic boundaries and extents covering a contamination region.

Remained part of the infrastructure to implement Geo-Web services is one of the directions for establishing a Linked SDI – Spatial Data Infrastructure. WFS – Web Feature Services as a part of SDI implementation could connect Linked Data and OGC. With the proposed Linked SDI architecture INSPIRE[18] – Infrastructure for Spatial Information in the European Community and GEOSS[19] – Group on Earth Observation System of Systems are promising vendors by their background and studies.

References

[1] A. Jentzsch, O. Hassanzadeh,C. Bizer, B. Andersson,S. Stephens. Enabling Tailored Therapeutics with Linked Data. In *Proceedings of the 2nd Workshop on Linked Data on the Web*, 2009.

[2] A. Zaveri, R. Pietrobon, T. Ermilov, M. Martin and N. Heino. Evaluating the disparity between active areas of biomedical research and the global burden of disease employing Linked Data and data-driven discovery. Tuberculosis IMISE Repo, 1-7, 2010.

[18] http://inspire.jrc.ec.europa.eu/

[19] http://www.earthobservations.org/geoss.shtml

[3] Daniel J. Abadi, A. Marcus, S. R. Madden and K. Hollenbach. Scalable Semantic Web Data Management Using Vertical Partitioning. In *Proceedings of the 33rd international conference on Very large data bases* - VLDB '07, 2007.

[4] Guy R. Cochrane and Michael Y. Galperin. The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. Nucl. Acids Res, 38 (supple 1):D1-D4, 2010.

[5] Hans Rosling at World Bank: Open Data, 2010. http://www.youtube.com/Hans_Rosling_at_World_Bank Last Visit: 2012, January 20.

[6] Hastings, J. (2008). Automated Conflation of Digital Gazetteer Data. *International Journal of Geographical Information Science* (22), 1109-1127.

[7] Human Genome Project Information. http://www.ornl.gov/sci/Human_Genome Last Visit: 2012, January 20.

[8] Joshua Tauberer. Using RDF for Distributed Information, 3/2/2005. http://www.govtrack.us/articles/20050302rdf.xpdh

[9] Kamel Boulos MN.Location-based health information services: a new paradigm in personalised information delivery. *International Journal of Health Geographics*, 2:2, 2003.

[10] M.Belleau, N. Tourigny,P. Rigault, and J. Morissette. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, 41(15):706-16, 2008.

[11] Mapping Relational Data to RDF with Virtuoso's RDF Views - http://openlinksw.com/rdb2rdf_mapping, Last Visit: 2012, January 20.

[12] Mark Walport and Paul Brest. Sharing research data to improve public health. *Lancet*, 377 (9765): 537-9,(2011,February 12)

[13] McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet, 9: 356-69, 2008

[14] Richard Cyganiak, Dave Reynolds and Jeni Tennison. The RDF Data Cube vocabulary, 2010-07-14.

[15] Russell, S. J., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd Edition ed.). New Jersey: Prentice Hall.

[16] Steven Johnson. The Ghost Map: *The Story of London's Most Terrifying Epidemic and How it Changed Science, Cities and the Modern World*, page 206.. Riverhead Books, 2006.

[17] T. Heath, M. Hepp and C. Bizer. Special Issue on Linked Data *International Journal on Semantic Web and Information Systems* (IJSWIS).

[18] Thomas R. Gruber. *A Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition, 5(2): 199-220, 1993.

[19] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space.*1:1, 1-136. Morgan & Claypool, 2011.

[20] Mather, F. J., Ellis, W. L., Langlois, E. C., Shorter, C. F., Swalm, C. M., Shaffer, J., et al. (2004). Statistical Methods for Linking Health, Exposure, and Hazards. *Environ Health Perspect , 112*, 1440-1445.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke