Towards Integrated Web Processing Services for Spatio-temporal Analysis of Geo-data, The Case of Lloyd's Lists

A. Mobasheri [1,2,*], M. Bakillah[1,3], K. Schultes[2], A. Zipf [1]

1 Chair of GIScience, University of Heidelberg, Berliner Str. 48, 69120, Heidelberg, Germany
2 Cluster of Excellence Asia and Europe in a Global Context, University of Heidelberg, Germany
3 Department of Geomatics Engineering, University of Calgary, Alberta, Canada

ABSTRACT

In this paper we present the research objectives and methodology for a running project of distributed web processing geo-services for spatio-temporal analysis of geo-data, the case of Lloyd's lists. We introduce the current problems and discuss the solutions in means of theoretical and practical perspectives. Initial results from geo-data modeling using Unified Modeling Language (UML) is presented and discussed. The paper ends with a discussion and conclusion of the next steps that has to be taken in the future.

Keywords: Web Processing Service, Spatio-temporal analysis, Lloyd's lists

1. Introduction:

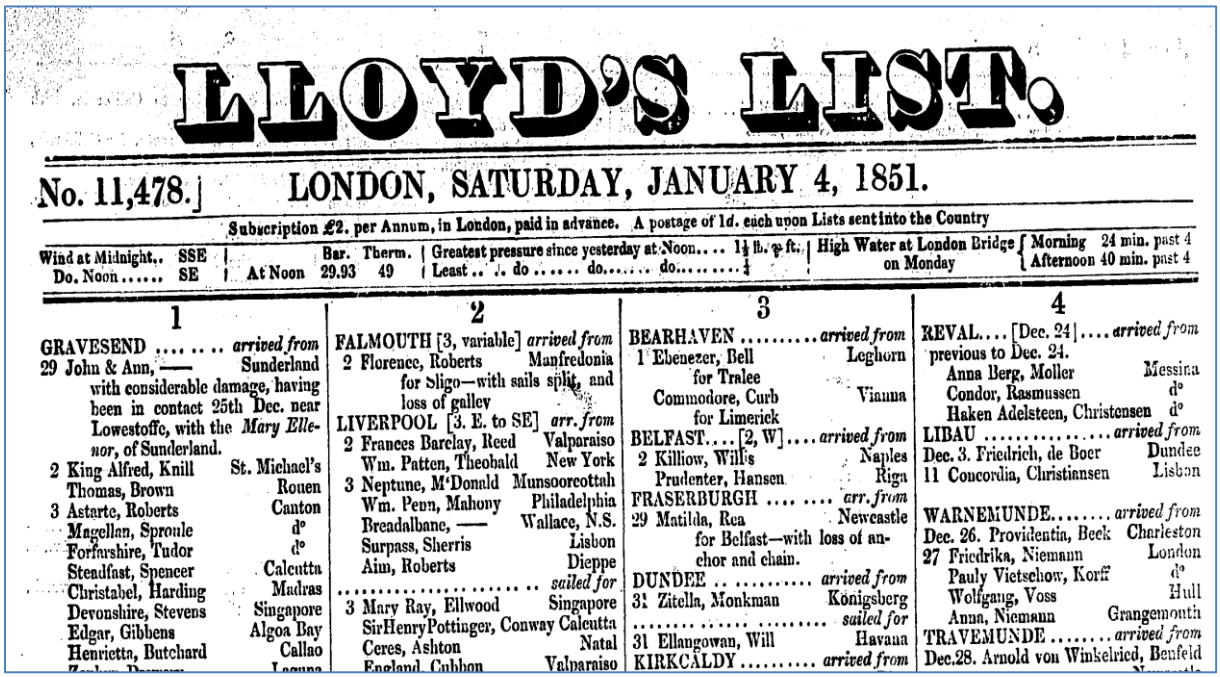
Since the late seventeenth century, the shipping newspaper Lloyd's List (Figure 1) and its direct predecessors contain weekly and later daily information on global shipping. The core of the Lists' mostly tabular contents is formed by the categories *"Shipping Intelligence"*, *"Speakings"*, *"Foreign Mail"*, *"Casualities"*, and *"War"*. Especially, the first two categories are essential in our case. "Shipping Intelligence" consists of exhaustive lists of the arrivals, departures and other nautical activities of civilian ships in practically all important ports of the world. The *"Speakings"* list sightings of ships at the high seas and give both the sighted and the reporting ship with name and geographical coordinates.

So far, only very few researchers have recognized the analytical potential of the Lloyd's lists. Yrjö Kaukiainen [1] and later Roland Wenzlhuemer [2] have looked at the shrinking of global communication times (first due to improved naval technology, later thanks to telegraphy) that become visible in the "Shipping Intelligence". But the "Intelligence" holds much more information that – with a primarily quantitative approach – will allow us to analyze, for instance, the shifting patterns of shipping routes; the time it took to get information, goods, and humans from one harbor to another depending on the year and season; the "black spots" and interruptions of service due to natural disasters or wars; the shifts in trade intensity between specific regions; the constantly changing patterns of transcontinental/international trade and migration; or, in short, the transformation of a variety of "global spaces" during times of rapid globalization.

Therefore, the hypothesis of this research is that geo-processing the spatiotemporal Lloyd's Lists data on the movement of people, goods and information onboard of

* Corresponding author: Amin Mobasheri, GIScience research group, University of Heidelberg, Berliner Str.48, 69120, Heidelberg, Germany. Email: amin.mobasheri@geog.uni-heidelberg.de

nineteenth-century ships vividly brings to life the migrational, commercial and, thus, cultural circuits and networks that powered contemporary processes of globalization.



LLOYD'S LIST.

No. 11,478.] LONDON, SATURDAY, JANUARY 4, 1851.

Subscription £2. per Annum, in London, paid in advance. A postage of 1d. each upon Lists sent into the Country

Wind at Midnight.. SSE | Bar. Therm. | Greatest pressure since yesterday at Noon.... 1½ lb. ? ft. | High Water at London Bridge on Monday { Morning 24 min. past 4 / Afternoon 40 min. past 4
Do. Noon..... SE | At Noon 29.93 49 | Least .. .. do ..... do..... do.....

1

GRAVESEND ..... *arrived from*
29 John & Ann, —— Sunderland
with considerable damage, having been in contact 25th Dec. near Lowestoffe, with the *Mary Ellenor*, of Sunderland.
2 King Alfred, Knill St. Michael's
Thomas, Brown Rouen
3 Astarte, Roberts Canton
Magellan, Sproule d°
Forfarshire, Tudor d°
Steadfast, Spencer Calcutta
Christabel, Harding Madras
Devonshire, Stevens Singapore
Edgar, Gibbens Algoa Bay
Henrietta, Butchard Callao

2

FALMOUTH [3, variable] *arrived from*
2 Florence, Roberts Manfredonia
for Sligo—with sails split, and loss of galley
LIVERPOOL [3. E. to SE] *arr. from*
2 Frances Barclay, Reed Valparaiso
Wm. Patten, Theobald New York
3 Neptune, M'Donald Munsoorcottah
Wm. Penn, Mahony Philadelphia
Breadalbane, —— Wallace, N.S.
Surpass, Sherris Lisbon
Aim, Roberts Dieppe
..... *sailed for*
3 Mary Ray, Ellwood Singapore
SirHenryPottinger, Conway Calcutta
Ceres, Ashton Natal
England, Cubbon Valparaiso

3

BEARHAVEN ..... *arrived from*
1 Ebenezer, Bell Leghorn
for Tralee
Commodore, Curb Vianna
for Limerick
BELFAST....[2, W].... *arrived from*
2 Killiow, Willis Naples
Prudenter, Hansen Riga
FRASERBURGH .... *arr. from*
29 Matilda, Rea Newcastle
for Belfast—with loss of anchor and chain.
DUNDEE .. .. *arrived from*
31 Zitella, Monkman Königsberg
.... *sailed for*
31 Ellangowan, Will Havana
KIRKCALDY ..... *arrived from*

4

REVAL....[Dec. 24]....*arrived from* previous to Dec. 24.
Anna Berg, Moller Messina
Condor, Rasmussen d°
Haken Adelsteen, Christensen d°
LIBAU ..... *arrived from*
Dec. 3. Friedrich, de Boer Dundee
11 Concordia, Christiansen Lisbon

WARNEMUNDE..... *arrived from*
Dec. 26. Providentia, Beck Charleston
27 Friedrika, Niemann London
Pauly Vietschow, Korff d°
Wolfgang, Voss Hull
Anna, Niemann Grangemouth
TRAVEMUNDE ..... *arrived from*
Dec.28. Arnold von Winkelried, Benfeld

Figure 1: Part of a page of Lloyd's list published on Jan 4th, 1851.

As an extra data source, we will also use the digitized maps of shipping routes downloaded from David Rumsey's map collection (Figure 2) and the CLIWOC database [3] which provide weather data (wind speed and direction, air temperature, etc.). The aim would be to integrate these datasets with the Lloyd's data in order to improve the normal trajectories (digitized from the maps) to a higher abstraction level, leading to semantic trajectories.

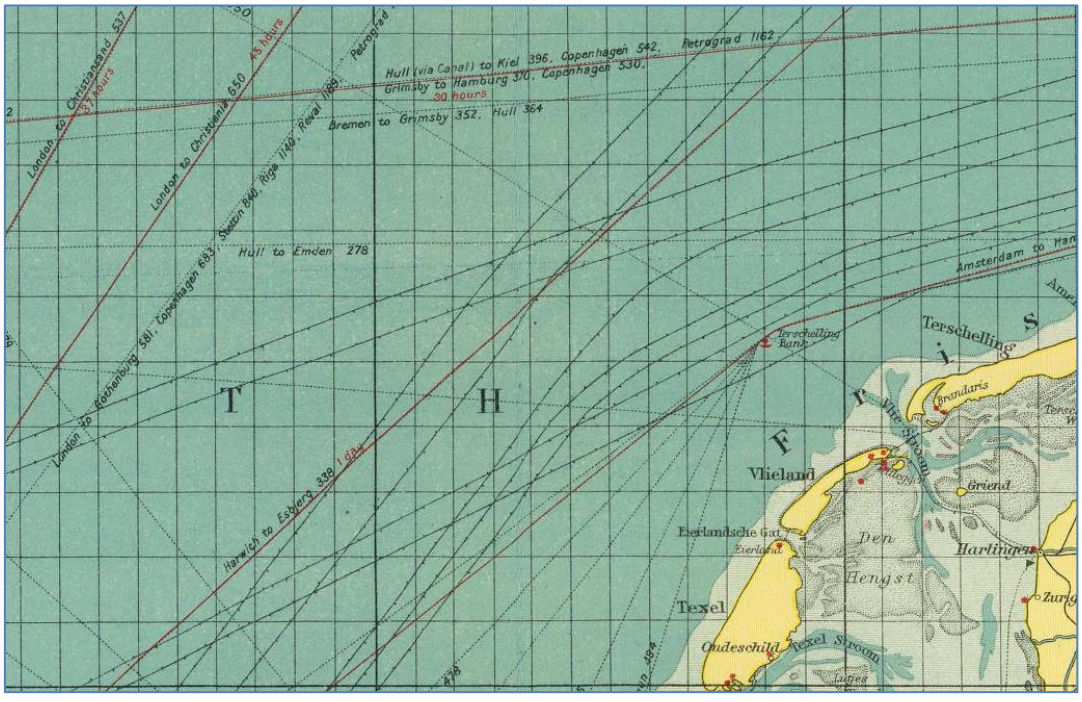


Figure 2: Part of map of Southern North Sea shipping routes with travel duration [4]

2. Research objectives:

The envisaged research will establish an exemplary interactive system that allows for the exploration of various historical sources within the spatiotemporal context dealing with arising ambiguities and vagueness (possible pasts). To achieve this it is necessary to develop specific tools for linking, managing, visualizing, and analyzing historical information simultaneously via the World Wide Web, which goes beyond simple static mapping and allows for the identification of previously unknown patterns.

The research mainly deals with semantic interoperability of geo-services in means of access, filtering and integration as well as performing spatial analysis of geo-data. The main objective of this research is to adapt/design formal semantics (e.g. ontologies) in order to make geo-information and geo-services interoperable for means of access, filtering, integration and analysis of geo-information (e.g. Lloyd's data). The following are sub-objectives related to the main objective:

- To define different needs and interests of historians in Lloyd's data.
- To study different geo-statistic methods that can be applied to Lloyd's data.
- To study about available standards of geo-information and geo-services (design, storage, and retrieval).
- To analyze the appropriate tools and methodologies for defining formal semantics (e.g. ontologies).
- To design Web Processing Services for different geo-statistic methods and algorithms based on OGC standards [5-7].
- To design necessary ontologies for WPS chaining and orchestration (local, application, and data ontologies).
- To design a prototype Web GIS that integrates various WPSs and ontologies/taxonomies in order to provide spatio-temporal analysis of Lloyd's data in a user friendly manner.
- To evaluate and validate the designed web GIS as well as the quality of output [8].

3. Research Methodology:

This research is a technological research and is broken down into five phases. Each phase has its own research type/activity, and goal. Hereby we elaborate the plan of works that have to be done in each phase, and later in the next section present the initial results of the work that has been done for the first phase.

3.1. Data preparation and modeling

As mentioned before, the Lloyd's list is a shipping newspaper published since 17th century. In the first, these data should be digitized and prepared in a geo-database for later use. A specific time period would be selected and the *"Speakings"* and *"Shipping Intelligence"* would be extracted from the newspapers and concerted into excel spreadsheets in a formal tabular format. In the next step the information would be used in

order to model the dataset. Unified Modeling Language (UML) is the most powerful and common language used for this purpose since it is a general-purpose modeling language and includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems. The UML classes and properties defined within the model could later be used in order to create a geo-database, and the available information for the selected time period would be transferred to the geo-database.

3.2. Method selection and Data analysis

In this stage, a comprehensive literature review would be conducted in order to select the most relevant geo-statistics methods (such as kernel density estimation, K(d) function, Spatial Scan Statistics, etc.) that could be applied on Lloyd's data for means of clustering data, finding correlations, finding trends and patterns, detecting hot spots, trajectories etc. These will use the *"Speakings"* and *"Shipping Intelligence"* data as case studies to build an interactive system that allows the user to explore manifold historical data sources (beyond the test- and showcase of Lloyd's Lists) within a spatiotemporal context. The methods would be selected, and applied on the dataset in order to justify their relevancy.

3.3. Web GIS design and implementation

In this stage the OGC standards for Web Mapping Service (WMS) [5], Web Feature Service (WFS) [6] and Web Processing Service (WPS) [7] would be used in order to design and develop a Web GIS for visualization and analysis of Lloyds' data. The WFS would be responsible of reading and storing data from/to geo-database. The WMS is connected to WFS and renders and visualizes the data on the web browser. In addition, for each method to be applied on the Lloyds list a WPS would be developed and various services are chained together in order to respond to users' queries. OpenLayers, JavaScript, and HTML 5.0 are the technologies that would be employed for creating the user interface and its connection with the services.

3.4. Semantic Interoperability and aggregation of web services

This phase mainly deals with designing and using the ontologies/taxonomies [9, 10], and the results gained from previous phase in order to integrate various WPSs together (service chaining), thus creating an interoperable collection of services that can talk together in order to respond to users' queries [11]. The result of this phase contributes to the main research problem for chaining and orchestration of GI-services [12, 13].

3.5. Evaluation of Web GIS prototype and refinement of possible problems

This phase includes the steps that are going to be taken for testing and validating the Web GIS prototype. For this issue, by use of web-service, the user (historian experts) would try to filter and analyze the Lloyd's dataset by using different spatio-temporal functionalities which are provided via User Interface. The strengths and weaknesses of the designed Web GIS are found and new approaches for improvement would be developed.

The service would be evaluated in two different means: First, the functionality of Web GIS would be tested. Speed of loading the data and working with the service (user

friendliness) are the main parameters to be evaluated. Second, evaluating the results of spatio-temporal analysis is another main point of interest. We would set up an experiment and ask historians to work with the service: How useful are the results for developing new questions? How do results fit into established knowledge about shipping? What other source material may be of historical interest? etc. The results of evaluation would be used to improve the Web GIS in the second cycle of Web GIS design. Evaluating the quality of geo-datasets and their influence on the results (error propagation issues) are also the final steps in this research [8].

4. Data preparation and modeling

As mentioned before, the Lloyd's list is a shipping newspaper which scanned copies of them for specific periods were ordered by the group for research purposes. As an initial attempt towards digitization, a large amount of published information for a specific time period which was of interest for historians; specifically years 1851 and 1871 were read and transferred into Excel spreadsheets in a formal tabular format. In the next step, in order to design a geo-database for the Lloyd's data, a geo-data modeling task was performed. From all the necessary information that needed to be recorded, a total amount of 6 tables were designed, each of which carry special attributes. Figure 3 illustrates the geo-database schema containing the name of designed tables. Note that the first four tables are normal tables containing several necessary information (e.g. Ship name, Captain name, Event date, Journal Date, etc.), yet the two last tables are considered as feature types since they have spatial components, thus can be treated as point features.

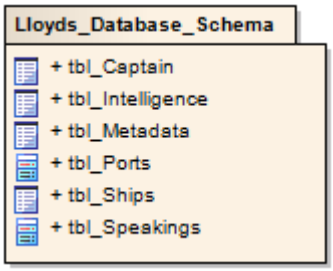


Figure 3. A schema package containing grouping of tables in geo-database

As it can be seen two main tables are *tbl_Speakings* and *tbl_Intelligence* which contain the shipping information. Figure 4 illustrates the data model for each table in a class diagram fashion. Furthermore, *tbl_Speakings* is the most special table in this model which contains coordinate values in forms of latitude and longitudes where the actual speaking has happened (somewhere in the sea) as well as information about the plan of the other ship (the sighted ship) such as its port of origin and destination. The information from this table could later be mapped in order to understand trends and patterns such as trajectories and hotspots where the *speakings* were likely to happen more often, if any. The other possibility for use of this information could be to correct the generic shipping routes that were already digitized based on collected maps from David Rumsey [4].

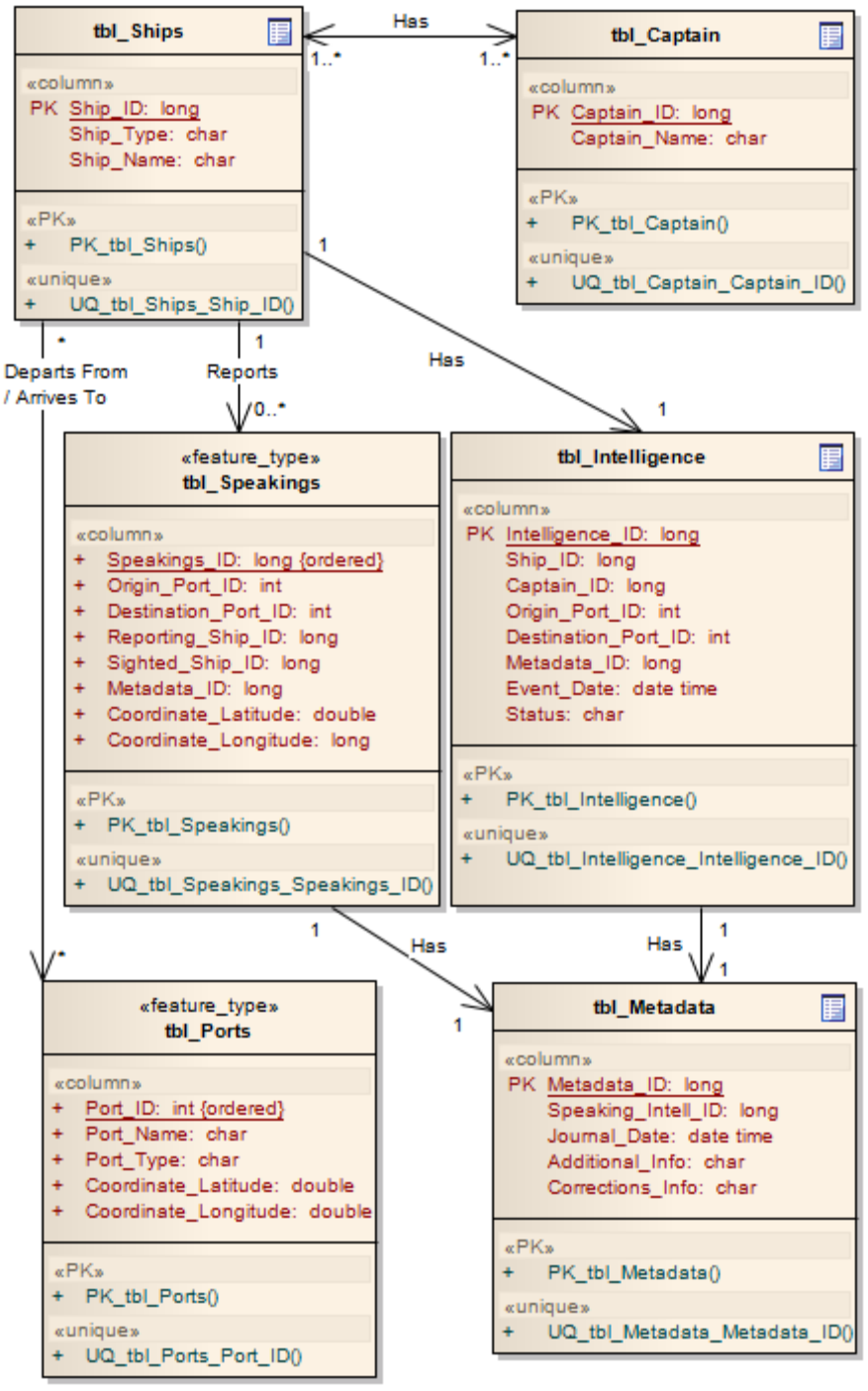


Figure 4: Main part of the data model in Unified Modeling Language

5. Discussion and Conclusion

In this paper we presented the research objectives and methodology for a running project of distributed web processing geo-services for analysis of spatial data, the case of Lloyd's lists. In the first stage, a data model for the Lloyd's lists is designed in order to provide unique format for the whole dataset. Afterwards, the data model was be used in order to create a geo-database, which will be populated with Lloyd's data.

Later, for future outlook, the data would be read from the geo-database through a Web Mapping Service (WMS) based on the region of interest defined by the user. Various Web Processing Services (WPS) would also be developed each of which perform special geo-statistic functionality on Lloyd's data.

Furthermore, using simple sample points would not be the only case in this research, we will also use a higher level of abstraction in order to introduce semantic trajectories of data. The reason is that more application-oriented ways of analyzing segments of movement suitable for specific purposes of the application domain is needed. This trend has promoted semantically rich trajectories, rather than raw movement, as the core object of interest in mobility studies. Therefore, in the future our research study will provide the definitions of the basic concepts about mobility data, as well as the approaches and techniques for constructing trajectories from movement tracks, enriching trajectories with semantic information to enable the desired interpretations of movements, and using data mining to analyze semantic trajectories and extract knowledge about their characteristics.

Acknowledgement

This research is made possible by a grant from the Cluster of Excellence "Asia and Europe in a Global Context" as well as GIScience research group of University of Heidelberg. Also, the authors would like to thank the hard work of students in the cluster as well as the department of history (HGIS Club) for extracting the data samples of the Lloyd's lists.

References

[1] Kaukiainen, Yrjö. 2001. Shrinking the World: Improvements in the Speed of Information Transmission, c. 1820-1870. European Review of Economic History 5 (1):1-28.

[2] Wenzlhuemer, Roland. 2010. Transforming Global Spaces: The Telegraph, Communication and Globalization in the Nineteenth and Early Twentieth Century, Habilitationsschrift eingereicht an der Philosophischen Fakultät der Universität Heidelberg.

[3] Wheeler, D., Garcia-Herrera, R., Koek, F.B., Wilkinson, C., Können, G.P., Prieto, M.R., Jones, P.D. and R. Casale. 2006: CLIWOC, Climatological database for the world's oceans: 1750 to 1850; Results of a research project EVK1-CT-2000-00090. European Commission, Brussels, ISBN 92-894-8279-6.

[4] David Rumsey Map Collection, available at http://www.davidrumsey.com/view, accessed in August 2013.

[5] de La Beaujardiere, J. 2002. Web map service implementation specification.*Open GIS Consortium*, *82*.

[6] Vretanos, P. 2005. Web feature service implementation specification. *Open Geospatial Consortium Specification*, 04-094.

[7] Stollberg, B., & Zipf, A. (2007). OGC web processing service interface for web service orchestration aggregating geo-processing services in a bomb threat scenario. In *Web and Wireless Geographical Information Systems* (pp. 239-251). Springer Berlin Heidelberg.

[8] Mobasheri, A., Bakillah, M., Zipf, A., & Liang, S., 2013. QualEvS4Geo: A peer-to-peer system architecture for semi-automated quality evaluation of geo-data in SDI. The IEEE 3rd International Conference on Innovative Computing Technologies (INTECH'13), London, United Kingdom.

[9] Gruninger, M., Fox, M.S., 1995. Methodology for the Design and Evaluation of Ontologies. In: IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995.

[10] Mobasheri, A., van Oosterom, P., Zlatanova, S., Bakillah, M., 2013. Semantic annotation of existing geo-datasets, a case study of disaster response in Netherlands. The 29th Urban Data Management Symposium (UDMS), London, United Kingdom.

[11] Bakillah, M., Mostafavi, M. A., Bédard, Y., 2006. Development of an Integrated Geosemantic Approach for Spatiotemporal Queries in Evolving Geospatial Multidimensional Databases for the Forestry and Land Management Domain. Colloque Géomatique 2006 – Au Coeur des processus, 25-26 October, Montreal, Canada.

[12] Walenciak, G. & Zipf, A., 2011. Interoperable Analysen in Geodateninfrastrukturen - Untersuchungen zur Entwicklung von Taxonomien von GIS-Operationen für die Definition von WPS-Profilen, AGIT 2011, Symposium Angewandte Geoinformatik, Salzburg, Österreich.

[13] Bakillah, M., Mostafavi, M.A., 2012. Semantic Interoperability in Ad Hoc Networks of Geospatial Databases: Challenges, Achievements and Perspectives. In: XXII International Society for Photogrammetry & Remote Sensing Congress.