Integration and Visualization of Tourism Data

Otakar Čerba*, Karel Jedlička*, Karel Janečka*, Tomáš Mildorf*, Josef Fryml, Pavel Vlach*, Dmitrij Kozuch***, Karel Charvát****

* University of West Bohemia, Plzeň, Czech Republic
** Czech Centre for Science and Society, Praha, Czech Republic
*** Help Service – Remote Sensing, Benešov, Czech Republic

Abstract. The goal of this paper is to define specifics of tourism data and to propose techniques and methods to harmonize, integrate, process and visualize this type of data. The aim is to reduce data heterogeneity and to improve its quality and attractiveness. The following text offers particular ideas, solutions and best practices of harmonization processes of spatial data related to tourism. They have arisen from various projects of applied research and similar international activities (for example SDI4Apps – Uptake of Open Geographic Information Through Innovative Services Based on Linked Data, Smart Open Data, CentraLab, Exliz – Excellence in Human Resources as a Source of Competitiveness, Habitats, Humboldt, Plan4all, Plan4business, Foodie, OpenTransportNetwork, GEPAM.EU – 3D Modelling of Terezín Memorial and Castle Kozel – research cooperation with National Heritage Institute of the Czech Republic on multidimensional register of cultural heritage) realized by authors.

Keywords: tourism, spatial data, data visualization, harmonization, integration.

1. Introduction

Tourism activities represent an important part of the world economy. According to the web page of the World Tourism Organization (UNWTO)[1] tourism covers 9% of global gross domestic product (GDP), 1/11 of jobs and 6% of the world's export.

The importance of tourism and related activities come through spatial data and geoinformation technology sphere. It is evident from a huge number of

[1] http://www2.unwto.org/content/why-tourism

applications focused on tourism on Google Play or from the popularity of websites and services such as Trip Advisor, Booking.com or various map portals.

The existence of too many applications is a bit uncomfortable for users. Tourists, sightseers or travellers do not want to come through many applications, compare their results and interconnect information from various domains. They are not able to find all relevant information (for example accommodation, transport, sights and guide services) in one place (application or web service)[2]. This problem does not arise from insufficient number of suitable data and information, but from their heterogeneity (detailed information on heterogeneity of spatial data in Vaccari et al., 2009, Čerba et al., 2012, Čerba, 2013). The heterogeneity covers many aspects of data including for example data format, data model, data provision, licences, frequency of updates and data quality. The heterogeneity of data and information represents a significant barrier of data integration and development of applications based on interconnected data resources.

The goal of this paper is to define specifics of tourist data and to propose techniques and methods to harmonize, integrate, process and visualize this type of data. The aim is to reduce data heterogeneity and to improve its quality and attractiveness. The following text offers particular ideas, solutions and best practices of harmonization processes of spatial data related to tourism. They have arisen from various projects of applied research and similar international activities realized at the Geomatics section[3] at the University of West Bohemia[4] (Czech Republic), non-profit organization Czech Centre for Science and Society[5], private company Help Service Remote Sensing[6] (Czech Republic) and other partners.

Chapter Tourism Data briefly describes specifics of tourism data. Particular solutions developed in selected projects are summarised in Chapter Integration & Visualization. It includes not only a description of the final results, but also methodologies, references, designes and implementations.

---

2 There are several exceptions, but usually with local coverage.

3 http://gis.zcu.cz

4 http://www.zcu.cz

5 http://www.ccss.cz

6 http://www.bnhelp.cz

2. Tourism Data

This paper deals with spatial data related to tourism, tourist industry and travelling. It might seem that there are no specifics of such data in comparison to other spatial data sets. However, there are several common characteristics, which are able to influence processing and visualization of these data sets. The following list summarizes the common properties of tourism data sets, which are taken into consideration in the next section focused on data harmonization and integration.

1. There are a lot of data sets used in various applications and services related to tourism and travelling.

2. Tourism data are usually localized (it contains a spatial component).

3. Tourism data sets are composed of vector features, except satellite images used as base maps.

4. The data are not very often collected, processed, visualized, provided and used by geomatic or cartographic experts.

5. The data sets as well as particular objects could be (and usually they are) very heterogeneous. The heterogeneities consist of many nonuniform aspects of data such as coverage (from global to local), granularity (level of detail), frequency of updates (it could be based on seasons and types of tourism, but there are several types of real-time data) or topics (accommodation, services, guiding etc.).

6. There are many geospatial analyses applied to the data (for example routing, searching in neighbourhood).

7. There are a lot of queries to these data sets. Therefore it is necessary to optimized data structures as well as storages to support selected query language.

8. A lot of tourism data is connected with volunteered geographic information, crowdsourcing and social media.

9. Presentation of tourism data is usually realized by maps and other similar cartographic products.

10. Data sets are presented mainly by mobile applications. This fact is connected with limited size of displays, memory, storage as well as changeable speed of the Internet connection.

3. Integration & Visualization

This section presents several solutions and approaches to harmonization and integration (including cartographic visualization) of tourism data. The following solutions are based on several international projects and activities realized by authors. These projects include SDI4Apps – Uptake of Open Geographic Information Through Innovative Services Based on Linked Data, Smart Open Data, CentraLab, Exliz – Excellence in Human Resources as a Source of Competitiveness, Habitats, Humboldt, Plan4all, Plan4business, Foodie, OpenTransportNetwork, GEPAM.EU – 3D Modelling of Terezín Memorial and Castle Kozel – research cooperation with National Heritage Institute of the Czech Republic on multidimensional register of cultural heritage. These projects focus not only on tourism data processing, but also on pilot activities related to tourist applications or services. The following proposed methods take into consideration properties of tourism data concluded in previous section.

3.1. Standards

The following section evidences the key role of standardization in the process of data harmonization and integration regardless of topic or domain. From the technical point of view there are six basic types of standards for – data and metadata formats (for example Geography Markup Language), data models (for example the INSPIRE /Infrastructure for Spatial Information in Europe/ specifications), communication protocols (for example various APIs /Application Programming Interface/), classification systems, hierarchies and nomenclatures including semantics (for example European Tourism Indicators System[7]), methodologies (for example the 5-steps harmonization framework – see Janečka et al. 2013) and visualization rules (for example for various INSPIRE data specifications).

Even though the standardization seems to be very uniform process providing unambiguous results, there are many open questions:

- One thing (data format, data model, vocabulary item etc.) can be described by various standards. The standards are sometimes in contradiction and not harmonised in terms of definitions, descriptions and classifications of geographic concepts in thesauri and controlled vocabularies (details in Čerba & Jedlička, 2014).

---

7 http://ec.europa.eu/enterprise/sectors/tourism/sustainable-tourism/indicators/index_en.htm

- There are conflicts between standards advanced by commercial sphere, academic and research institutes, legislation and communities (for example various ways of transport network modelling in commercial software, INSPIRE and OpenStreetMap).
- It is necessary to mention, that standardization is related to openness, independence and accessibility of standards and their licences.

3.2. 5-step Harmonization Framework

Data harmonization is necessary task if data from heterogeneous sources (for example regional or cross-border datasets, data collected by volunteers) should be combined together into integrated, consistent and unambiguous information products. Such datasets can then be viewed, queried and analysed in a unified manner. Janečka et al. (2013) proposed a general framework how to harmonize data coming from the domain of spatial planning. Nevertheless, the proposed 5-step harmonization framework (see Figure 1) is also applicable on tourist data and could be used mainly by data providers and stakeholders. The first three steps are the common steps for all scenarios. The understanding of both the source and the target data is based mainly on particular data specifications, documentation and metadata. It is quite necessary to correctly understand the source and the target data to define all the necessary harmonization steps.

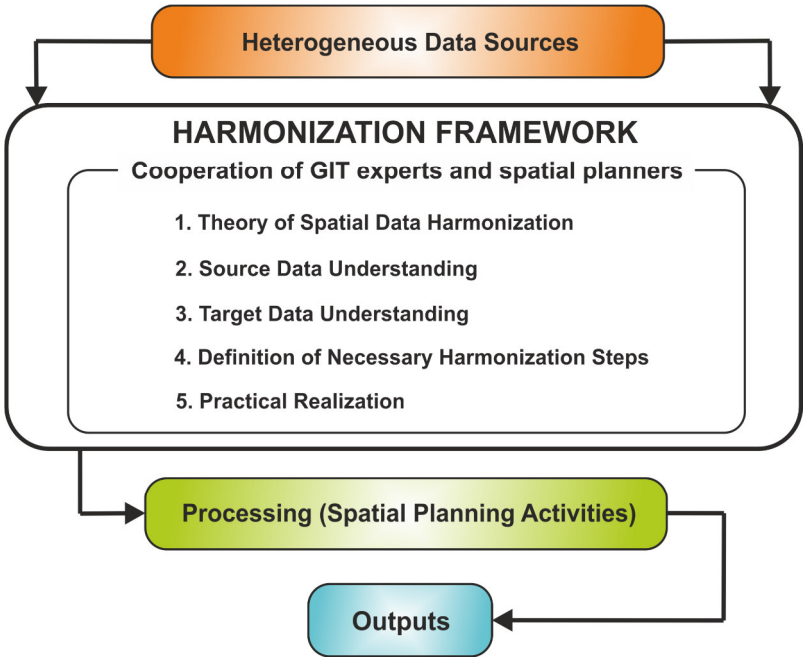

Figure 1. The proposed 5-step harmonization framework for harmonization of spatial planning data (Janečka et al. 2013) that is applicable also on tourist data.

Regarding the practical realization, there are several ways how to handle data harmonization. According to Janecka et al. (2013), it is possible to use extract-transform-load (ETL) tools or apply the current capabilities provided by modern relational database management systems. Eventually, a classic geographic information system (GIS) does the job.

3.3. Linked Data

The Linked Data approach enables efficient sharing and combination of data from various resources. It is very important in case of tourism. As it was mentioned above, data for tourism cross traditional borders between local and global or between general and detailed. Users or developers of a tourism application need to transform data to a common model and store them in one storage or implement principles of Linked Data and to use external data resources.

The principles of Linked Data were published in the article Design issues: Linked data (Berners-Lee, 2006). There is mentioned the five-star system designed for Linked Data rating. Berners-Lee (2006) describes two crucial technologies enabling development links between data sets and objects – Uniform Resource Identifier (URI), which guarantees a mechanism of the

unique identifiers, and Resource Description Framework (RDF), that enables formalized description of each information based on triple (subject – predicate – object). Linked Data (fitting for tourism) are provided by many applications such as Dbpedia or GeoNames.org.

The Linked Data principles have been applied in two ontologies designed for tourism purposes – ontology of important sights in Rome and persons related to these sights and knowledge of application focused on ski resorts in Europe (Let's Do Some Skiing – LDSS knowledge base). The Linked Data principles were implemented in the following four main ways:

1. Standard or respected Linked Data component – data producers do not have to develop new vocabularies, data types or relations, but reuse existing components (for example the Rome ontology deals with many DBpedia relations). It guarantees better interoperability, querying and visualization.

2. Links to equivalent or similar objects in Linked Data resources – all languages used for Linked Data encoding (for example the Web Ontology Language) have specific relations for equivalent or similar concepts (for example owl:sameAs). Such relation can interconnect two objects "Tignes" (ski resort in France) in LDSS knowledge base (which contain specific information for skiers) and in DBpedia (general information). Publication of such link means that user can link to other data resource and find new information.

3. Transformation of selected parts Linked Data resources to data – this way represents an extension of the previous approach. Data from external resource can be not only linked, but also added to local knowledge base physically. The XSLT (Exctensible Stylesheet Language – Transformation) script generating web page of particular ski resort downloads data from DBpedia and publishes it as a part of the HTML (HyperText Markup Language) code.

4. Online loading of result of SPARQL (SPARQL Protocol and RDF Query Language) queries (an online version of the previous way) – the description of particular ski resorts is online gained from the original resource. It is allowed by a SPARQL query that is integrated to the URL /Uniform Resource Locator/ of the Javasctipt element of the web page. The following code is a part of the XSLT style generating SPARQL query for each resort.

```
<script          src="http://dbpedia.org/sparql?default-graph-
uri=http%3A%2F%2Fdbpedia.org&amp;query=prefix+dbpedia-
owl%3A+%3Chttp%3A%2F%2Fdbpedia.org%2Fontology%2F%3E%0D%0Asele
ct+%3Fabstract%0D%0Awhere+%7B%3C{$h_abstrakt}%3E+%3Chttp%3A%2
```

F%2Fdbpedia.org%2Fontology%2Fabstract%3E+%3Fabstract+.%0D%0Afilter+%28LANG%28%3Fabstract%29+%3D+%27en%27%29%0D%0A%7D&format=application%2Fjavascript&timeout=30000&debug=on"/>

3.4. Multilinguality

Tourism data and applications belong to exemplary cases of necessity of multilinguality. Users usually require information on an unknown foreign area, but they want to have information in their mother tongue or in some "lingua franca".

There are two main ways how to support more languages in tourism data. On the level of the data code in an XML-based format there are so-called language tags (above all xml:lang). They are able to interconnect a data record and the language used for its description. The language tags re-use RFC 5646, Tags for the Identification of Languages (Phillips & Davis, 2006), which provides standardized codes for describing any language.

The tools focused on geographical names and their expression in various languages are the second way to support the implementation of more languages in data. Many data formats (for example RDF) enable to use more than one label for a data record. For example Praha as the capitol of the Czech Republic can have many labels such as Prague, Prag or Praga. As sources of geographical names there are gazetteers (geographical dictionaries) such as GeoNames.org (providing linked data as well), NGA GEOnet Names Server, Getty Thesaurus of Geographic Names, JRC Fuzzy Gazetteer or many local gazetteers. Questions related to geographical names are also accented by the United Nations Group of Experts on Geographical Names (UNGEGN, 2006) or by the joint group of the International Cartographic Association and International Geographical Union – ICA/IGU Working Group/Commission on Toponymy.

3.5. Third Dimension

Data for tourist must presented in an attractive form, contrariwise for example to analytical and research applications of GIS, where complex data representations are often used. And there the third dimension can help to perceive the information in natural way. The authors of this paper have a long term interest (since 2005) in extending GIS and cartographic visualization to 3D. The key issue of 3D GIS is that not every 3D visualization is sufficient to registering and analytical tasks which are common in classical 3D GIS. Pure visualisation of a digital terrain model and 3D buildings with tourism data doesn't offer any new functionality or added value. To do so, two more things have to happen:

- The 3D models have to be geometrically, topologically and semantically interconnected with the other above mentioned 2,5D data (the registering function of GIS).
- The GIS has to offer analytical tools working in three dimensions (the analytical function of GIS).

Comparison of the above text to definitions of the Virtual Geographic Environments (VGE) shows that taking advantage of VGE concept can lead to successful implementation of 3D GIS:

VGE is understood as an integrated virtual space which is used to simulate and analyze complex phenomena and process of geography. It supports activities of research on geography as collaborative work, knowledge sharing and group decision-making. It can become one of the scientific methods and advanced technologies of modern experimental geography study, Lin & Huang (2009). Virtual Geographic Environments (VGEs) are built on the foundations of geographic information systems (GIS) and geographic information science (GI Science) in which considerable attention is paid to the user in terms of the manner in which they interact with the software, Liu & Zhu (2005). But there is a strong stress given to a user interaction with such a system. The user interaction can vary from very generic exploration of a common user to research analysis made by professionals. Because there is also a strong stress to a realistic representation of the world, VGEs use 3D representation of the world instead a 2D representation which is typically used in GIS, adapted from Lin & Batty (2011).

In summary, the VGE concept enriches the 3D GIS. On the one hand, following that concept prevent a designer from creating a non-semantic visualization. On the other hand, it has to be mentioned that there exist many more theoretical articles about VGEs, but only few of them are describing particular working solutions. See for example Pavelka (2011). Therefore, the geomatics team at the University of West Bohemia focuses on the best practice examples showing the 3D GIS potential, which are inspired by the broader understanding of 3D, adopted from VGE[8]. A comprehensive description of a practical 3D GIS development can be found in Hájek et al. (2013) and Jedlička & Hájek (2014). Whereas the outputs from the research cooperation with the National Heritage Institute of the Czech Republic at the Kozel castle are focused on both castellans who could use the outputs for castle management and tourist who enjoy the attractive form of presentation in 3D, the outputs from the GEPAM project (3D model of Terezín Memorial) are primarily focused on tourists.

---

8 http://gis.zcu.cz/projekty/3DGIS/

3.6. Responsive design

Nowadays, in the time of rapidly increasing number of devices which can be used for browsing the Internet, the requirements for the quality of websites and web applications have been rising. The applications focused on tourism and travelling are not the exception. Thus, it is important to take this fact into account in a process of designing applications (Rampton 2014). The developers should strive to create applications which fit different types of devices. In fact, interacting with websites for instance on smartphones or tablets is not the same as on desktop computer monitors. Many factors such as clicking versus touching, screen size, pixel resolution have become crucial while designing web applications.

Generally, the design has a huge influence on usability and accessibility of the application. The developer should aim for creating user-friendly and obstacle-free application, which satisfies the expectations of users and allows them to carry out what they need. The design should contribute to the quality of using, effective organization and simple navigation (Schobesberger 2012). Other significant aspects of a successful design are attractiveness, aesthetic, appropriateness, simplicity or consistency.

As it was mentioned above, the websites designers have to respond to the rapid development of devices. They need to deal with a range of viewport sizes (Firtman 2014) as it is seen on the W3Counter website[9]. Responsive design is an effective way how to adjust the websites to various screen resolutions. The term responsive design was firstly used by Ethan Marcotte (2010). It is an approach of laying-out and coding website such that the website provides an optimal viewing experience across a wide range of devices (Rampton 2014). According to Marcotte (2010), there are three essentials for responsive web design – fluid grids, flexible images and CSS (Cascading Style Sheets) media queries (the rules which enable to change styling according to screen size).

On the one hand, the reasons for designing responsively are obvious. Content can be created once and delivered to multiple devices (Firtman 2014). Thus, the user's experience is the same no matter which device accesses the application. This is important also in the case of tourism map apps where the user often examines the map at home on a desktop and then only checks the current situation on the way via a smartphone. In addition, all of the content is automatically adapted to the screen. It is more convenient for user to read and navigate on the website (Rampton 2014).

---

9 http://www.w3counter.com/globalstats.php

Moreover, user does not have to bother with zooming and moving horizontally the map or text on the screen.

On the other hand, responsive design shouldn't be the sole alternative for mobile devices. The best solution is not delivering the same data amount regardless of screen size. There still exist slow mobile network connections and not all users will wait for the application to load (Firtman 2014, Marcotte 2010).

To sum it up, it is crucial these days to adjust the tourism map applications to a wide range of devices, because they are intended to be used on a desktop as well as on a mobile. It enhances users experience and allows them to use the application promptly and efficiently, always in the same way.

4. Conclusion

The freedom of movement, including travelling, is one of the fundamental human rights. It seems that tourism is quite simple and undemanding with respect to spatial data and geoinformation technologies. But the needs of tourism from a view of geomatics and cartography consist in the following specifics:

- A huge number of very heterogeneous data and information.
- A large group of potential users with various requirements and experience.
- Many data and information providers with different scopes, expectations and targets.

For inventive developers there is a tempting opportunity to reuse existing data, information, tools and services for creating new complex applications for tourism. But it necessary to keep in mind the above mentioned specifics and to reuse not only data or pieces of software, but also methodologies and approaches designed in the above-mentioned projects. The methodologies, approaches and other best practices include interesting and helpful resources, but also processes of data harmonization or integration, attractive forms of data presentation as well as essential rules supporting accessibility and comprehensibility.

Therefore, there are project activities such as SDI4apps, which try to interconnected existing solutions and provide to non-expert developers and users (from the view of geomatics and cartography) a useful set of existing

tools and principles to make a development of tourism applications and services easier.

The authors would like to point out three key facts connected with the integration and visualization of tourism data and design of applications based on tourism data:

1. The target group (users as well as developers) represents the basic limiting factor influencing a selection of standards, methodologies and tools. It is necessary to use, process and present data in a way supporting strengths of the target group and not to allow any ambiguities or doubts, which could mean potential problems and outflow of customers. But it does not mean a total adaptation to requirements of the target group and rejection of traditional, respected and correct approaches and solutions.

2. It necessary to involve other domain experts in the development process. Computer programmers, web developers, designers as well as cartographers and GIT (geoinformation technologies) experts tend to "do it yourself" solutions. But a large group of people from various spheres improve data processing and application development. The products focused on tourism should be created in a close cooperation of spatial data experts, cartographers, designers, web page developers, programmers, tourism experts, providers of particular services, local people, experts on nature, history etc. There is also a key role of users in providing a feedback.

3. The summary of methods provided in this document is not complete. The whole research of processing of spatial data focused on tourism is much broader. It touches questions of eye-tracking (Brychtová et al., 2012a, Brychtová et al., 2012b), processing of big-data (Graham & Shelton, 2013, Crampton et al., 2014), semantics (Vohnout et al., 2014, Čerba & Jedlička 2014) or particular technological solutions and standards.

References

Berners-Lee, T. (2006). Design issues: Linked data. World Wide Web Consortium.

Brychtová, A., Popelka, S., & Dobesova, Z. (2012). Eye-tracking methods for Investigation of Cartographic Principles. Proceedings of the12th International Multidisciplinary Scientific GeoConference SGEM, 2, 1041-1048.

Brychtová, A., Popelka, S., & Voženílek, V. (2012) The Analysis Of Eye Movement as a Tool For Evaluation of Maps and Graphical Outputs from GIS. Geography and Geoinformatics: Challenge for Practise and Education, 154.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data'and leveraging the

potential of the geoweb. Cartography and Geographic Information Science, 40(2), 130-139.

Čerba, O. (2013). Information Around Us: Questions Connected to Information and Data Heterogeneities in Planning Activities. REAL CORP 2013: Planning Times, Acquario Romano (House of Architecture), Rome, Italy.

Čerba, O. & Jedlička, K. (2014). Geomatic Concepts in Agriculture Thesauri. In IS-AF 2014, Jelgava, Latvia.

Čerba, O., Jedlička, K., Charvát, K., Ježek, J., Janečka, K. & Mildorf, T. (2012). The Overview Of Spatial Data Harmonisation Approaches And Tools. In Proceedings Vol. 1. 4th International Conference on Cartography and GIS, Albena, 2012. ISSN: 1314-0604.

Firtman, M. (2014). You May Be Losing Users If Responsive Web Design Is Your Only Mobile Strategy. In: Smashing Magazine. Online: http://www.smashingmagazine.com/2014/07/22/responsive-web-design-should-not-be-your-only-mobile-strategy/.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. Dialogues in Human Geography, 3(3), 255-261.

Hájek, P., Jedlička, K., Vichrová, M., Fiala, R. (2013) Conceptual approach of information rich 3D model about the Terezín Memorial, Geoinformatics FCE CTU. 02/2013. On-line: http://geoinformatics.fsv.cvut.cz/pdf/geoinformatics-fce-ctu-2013-11.pdf.

Janečka K, Čerba O, Jedlička K, Ježek J (2013) Towards Interoperability Of Spatial Planning Data: 5-Steps Harmonization Framework. In: Proceedings of 13th SGEM GeoConference on Informatics, Geoinformatics And Remote Sensing, ISBN 978-954-91818-9-0 / ISSN 1314-2704, June 16-22, 2013, Vol. 1, 1005 - 1016 pp. DOI:10.5593/SGEM2013/BB2.V1/S11.051.

Jedlička, K., Hájek, P. (2014) Large scale Virtual Geographic Environment of the castle Kozel – best practice example. In Proceedings of International Conference on Cartography and GIS 2014, Varna, Bulgaria.

Lin, H., Batty, M. 2011. Virtual Geographic Environments: A Primer. In: Lin, H., Batty, M (Editors). Virtual Geographic Environments. ISBN: 9781589483187. 2011. 364 pages.

Lin, H., Huang F., Lu, G. 2009. Development of Virtual Geographic Environments and the New Initiative in Experimental Geography. Acta Geographica Sinica. 2009-01. ISSN: 0375-5444. CN: 11-1856/P. YP: 1934.

Lin, H and Zhu Q. 2005. Virtual Geographic Environments. In: Zlatanova S and Prosperi D (Editors). Large-scale 3D Data.

Marcotte, E. (2010). Responsive Web Design. In: A List Apart. Online: http://alistapart.com/article/responsive-web-design.

Pavelka, K. 2011. XXIII CIPA Symposium - Prague, Czech Republic - 12/16 September 2011 – Proceedings. Edited by Pavelka, K. Online: http://cipa.icomos.org/index.php?id=69.

Phillips, A., & Davis, M. (2006). Tags for identifying languages. BCP 47, RFC 4646.

Rampton , J. (2014). How Important Is Responsive Design? In: HuffPost Business. Online: http://www.huffingtonpost.com/john-rampton/business-mobile-responsive-design_b_5267077.html.

Schobesberger, D. (2012). Towards a Framework for Improving the Usability of Web-mapping Products. Vienna. PhD Thesis, University of Vienna.

United Nations Group of Experts on Geographical Names, & United Nations. Statistical Division. (2006). Manual for the national standardization of geographical names (Vol. 88). United Nations Publications.

Vaccari, L., Shvaiko, P., & Marchese, M. (2009). A geo-service semantic integration in spatial data infrastructures. International Journal of Spatial Data Infrastructures Research, 4, 24-51.

Vohnout, P., Cerba, O., Kafka, S., Fryml, J., Krivanek, Z., & Holy, S. (2014). SmartTouristData approach for connecting local and global tourist information systems. In IST-Africa Conference Proceedings, 2014 (pp. 1-6). IEEE.