Semantic Geodemography and Urban Interoperability

Joaquín Borrego-Díaz, Antonia M. Chávez-González, Mónica A. Martín-Pérez, and José A. Zamora-Aguilera

Dept. of Computer Science and Artificial Intelligence - University of Seville, Avda. Reina Mercedes s/n. 41012 Seville, Spain

Abstract. Nowadays there exists an increasing interest on the use of the information collected by cities coming from different resources as data with dynamic nature like the one provided by sensor networks, as static data associated to the socio-technical system that the city performs. As well as the Semantic Sensor Web allows the standardization of data, it is essential to give an appropriate dealing to geo-demographic data. In this paper, an approach to the semantization of the geo-demographic information is presented, with the aim of achieving interoperability within other systems of the *geospatial cyberinfrastructure*. Furthermore, fundamental aspects of the creation of ontologies by starting from socio-demographical systems are discussed and the process is illustrated with a case study.

1 Introduction

Today, flows of information produced and extracted from cities are becoming increasingly interesting. The integration of these data for the improvement of processes and services is one of the mainstays on which the concept of *smartcity* is based. Data sources range from sensors to other with a more statical nature such as the ones provided by national census and other administrative and commercial databases or social and demographical surveys [4]. For an accurate combination of data we need semantic-level systems of operability. Although there exist initiatives such as *Semantic Sensor Web*, in the case of socio-technical systems (the ones which deal with socologic, demographic and cultural data) it has not been advanced so much as in that of the former. To provide metadata (and mainly ontologies) for geodemographic representation is essential to design innovative products, processes or services by semantic interoperability of two kinds of systems:

• Systems for Urban Computing: *Urban Computing* is a branch of Pervasive Computing that investigates urban settings and everyday lifestyles. A lot

^{*} Partially supported by TIN2009-09492 project (Spanish Ministry of Science and Innovation), cofinanced with FEDER founds, and Proyecto de excelencia TIC-6064 Conceptos emergentes en sistemas complejos. Aplicaciones en entornos urbanos y en complejidad cultural financed by Junta de Andalucía.

of information to develop pervasive applications for urban environments is often already available, even if scattered and not integrated: maps, points of interest, user locations, traffic, pollution, events are just a few examples of the digitalized information which we can access on the Web [8].

• Geodemographic information systems represent a kind of business tool for interpreting data that consists mainly of a demographic database, digitized maps, and software. Data are captured on the computer, updated, maintained and organized for effective use and manipulation. Locational and spatial variations of population features are analyzed. Although primary goal is marketing, such systems are widely used for several applications [7].

The achievement of a trustworthy semantic interoperability among both kind of systems is essential to the design of semantic sociotechnical systems which combine both kinds of information. Particularly in smartcities projects, because it would allow the design of taylored services and processes to specific neighborhoods or urban spaces. The interoperability will make easier the management and valuation of the socioeconomical consequences of regions. Thus, new methodologies should be designed for urban spaces and these methodologies represent an opportunity to take into positions in new cities growing at a fast pace in emerging regions [10].

The aim of this paper is to present an approach to the semantization of geodemographic information from the point of view of the Semantic Web framework (described in Sect. 2). The design of socio-geodemographic ontologies is the precious step to carry out the semantic interoperability between urban computing and geodemographic information systems associated to cities or regions (Sect. 3). Also we discuss the fundamental features of the transformation from geodemographic information to an ontology (Sect. 4), by showing the ideas with a case study (Sect. 5). Meta-descriptions of digital resources are represented, and the gap between sociodemographic characterizations and formal descriptions are shown. The paper ends with our conclusions about the approach as well as their relationships with other ontology-based solutions.

2 Semantic Web and (Urban) Ciberinfrastructure

The Semantic Web (SW) aims to extend the current WWW realm to trustworthy process the information by means the metadata representation, which is enriched (transformed in Knowledge) by means the use of its misunderstanding interpretation provided by ontologies. As an extension of the actual Web, the implementation of the WS must overcome big obstacles [3] from the point of view of the knowledge representation and reasoning. Its layer structure (RDF/ RDF(S)/ OWL/ Ontologies) sets several abstraction levels in which ontologies play a key role. The interest for applying typical WS techniques, as the ontologies are, comes from got the results, to a certain extent, by means of automatic reasoning, what grants them trust. WS technologies extend to important fields such as the *Semantic Sensor Web* [19] or the management of *Linked Data*.

Other aspect to bear in mind when considering the use of ontologies for urban surveys, is the analysis of the geospatial ciberinfrastucture (GCI) that the city

enjoys. A ciberinfrastructure (CI) combines data sources, computing platforms and services together to provide people information and computing tools in an information-driven world. Geospatial ciberinfrastructure adopts intrinsic principles and geospatial information to support processing abilities such as geospatial analysis and geospatial decision [23]. GCI analysis is essencial, not only for the establishment of a research and development agenda in the city, but to capture its intrinsic features and the influence and relationship between the city and physical, social and geographical elements of urban framework. Having in mind that the integration of every systems conforming a GCI is not feasible in practice, a GCI provides, by means of ontologies, a common semantic framework to enable semantic interoperability and shared understanding. In this context, a case of special interest is the one of the location based service, PDAs and iPhones, due to the increasing number of mobile devices, the further introduction of sensors and Geographical Information Systems (GIS) available devices.

Ontologies can play a key role to deal with and streamline the information that the city owns, by means of GCI and getting interoperability for heterogeneous urban information systems [13]. Nevertheless, the process of construction of ontologies for urban development or city management, faces the gap between pragmatic heterogeneity concerned with urban concepts and the difficulty about formalizing concepts within a technological framework like WS is (besides the representation in OWL-like standards). As an example, the concept "residence" of an individual is crucial, and its formalization should -in geodemographics- include features on the social role that the residence stamps [6]. There are other urban concepts for which not only there is not a precise definition but, in any case, the limits of a definition are fuzzy and, thus, capable to lead to inconsistencies.

Therefore, geodemographics should be included like one of the systems belonging to the GCI, thus information will be cut across within geospatial domain (social, environmental, etc.) by other GCI systems. The evolution of GCI will produce platforms where data are collected, analyzed and used by communities and, besides, the specialization and adaption of results to different sociodemographic realities provides the infrastructure with an added value.

3 Geodemographies as Knowledge Augmented Spaces

Geodemography deals with the study of the information about complex social realities which are the reason for emerging phenomena arising in cities and sensorbased information is not able to represent. The origins of geodemographics date from 19th century but the development of a software specifically designed to facilitate the design of classification systems to sort out people and places is which motivates the growth in social scientific interest in these classifications. That is how geodemographics play a key role in the analysis of the relation between places and society [6]. The relevance of the modernization of information processing and representation lies in the fact of it facilitates the study of networks, neighborhoods and communities making possible to analyze how these elements are perceived by citizen and which methods are more appropriate for understanding and promoting them. Although cities are complex systems in which a number of (physical, social, virtual, etc.) networks interweave and connect, and are based on scalability and urban-morphology principles [5], the social dimension linked to spaces strongly determines the rest of networks. The existence of a semantic geodemographics will facilitate the interoperability among the different systems at a semantic level. In that context, to make use of a geodemographical layer provides management systems the capability of connecting real-time extracted data with social, demographical and economical features leading to a suitable decision making with the best information and a better diagnosis. GCI can use this kind of information (with static nature) to advice and obtain connections between the behavior of the city (and residents) and the social framework of a concrete zone.

The opportunity of encourage the sustainability of sociotechnical systems is especially interesting. These systems are concerned with society, organizations, individuals, their behavior and the technological infrastructure that they use. Smartcities have become global sociotechnical systems, including the sustainability issue. Large and long-lived impacts on sustainability will require enabling, encouraging, and sustaining changes in behavior -on the part of individuals, organizations, and nation-states over the long term. Informatic technologies, and in particular real-time information and tools, can better equip individuals and organizations to make daily, ongoing, and significant changes in response to a constantly evolving set of circumstances [15]. For example, spatial formalizations are empirically linked to multiscale transport systems and a range of urban socioeconomic consequences in different urban cultures worldwide [18]. Therefore, sociogeodemography aids to understand the transport as a global phenomena of communities. It is worth to notice the opportunity of encourage the social benefits of connecting, by means open data, people who share the same neighborhoods or practices, reducing gas emissions, power consumption, as well as keeping citizen informed about local events [1].

Nevertheless, *Local* term can be controversial, although it is aimed that the information locate sectors and zones with a special interest. This term includes spatial categories such as "community", "neighborhoods" or area. With respect to physical place, it refers to the physical support of local communities. But it is possible that it refers to communities that persist in an area or a time [12]. Effectively, places and identities have gone hand in hand but social assignment of identities is a complex task created through practices of consumption across a range of spaces which are key in defining who we are.

4 Semantizing Geodemographies

From Urban Informatics' point of view, sociogeodemographic ontologies provides support for both metadata and interoperability in several levels of information ecosystem in the city (see Fig. 1). It allows to enhance information (metadata) from sensors with information about the population of the area, for example. Also provides information about the urban features associated to geodemographic indices. In this way the city model provided by the information can be enriched with knowledge that allows to argument decisions which influences city behavior.

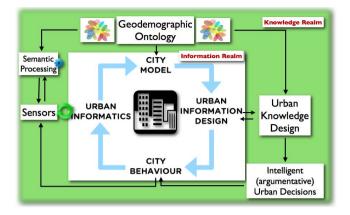


Fig. 1. Knowledge from the ontology to the urban information ecosystem

Intended use: The first step is to limit and specify the intended use of the ontology. As we commented, and thinking on smartcities, our aim is to fit together that information within the information flows of the city for providing it with quality from the social point of view. Of course, this goal must consider that the GCI is a *middleware* among different information sources and it integrates several functions. Therefore, the ontology has to reflect and standardize the socio-demographical information available to build systems for combining this information, in such a way that GCI provides us with (both spatial and digital) knowledge, etc. To suppose that the ontology is useful for other kind of tasks would mean that its scope of application would be unstable, fuzzy and therefore not usable. For example, to consider demographical geo-located sectors as communities de facto does not imply to consider virtual communities or the ones built under other criteria. In that case, social heterogeneity performed by semantic geo-demographics can be different from the one that the pre-existent GCI considers.

Requirements: Mainly two requirements have to be considered:

• Ontology must facilitate the high level information fusion that allows future social changes. The recollection of new information from residents is necessary to update the sociotechnical systems within GCI. On the one hand, a process for building formal ontologies to provide a conceptual framework for higher level fusion processing is necessary [14],. On the other hand, there are approaches [9] addressing issues related to the capability of generating and integrating user-generated information into the GCI, to be reused and shared. In [9] authors present semiautomatic mechanisms to augment the availability of user-generated information, improving the visibility of geospatial resources.

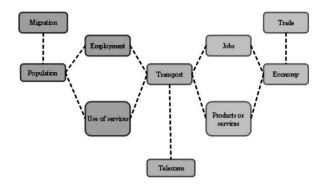


Fig. 2. The main subsystems of an urban and regional system [22]

• Ontology must provide knowledge about (and for) Urban Systems: Main urban subsystems (Fig. 2, extracted from [22]) depend from information that can be provided by geodemographic systems. Knowledge provided by geodemographic metadata aids to refine and analyze the modelization of such subsystems. The figure shows the elements in broad terms—the population and the economy as a framework with the addition of activities, interactions and infrastructure—all by location: people reside, work, shop and use services at a variety of locations. Indeed, representing spatial interaction is a key underpinning for many urban and regional models [22].

5 Extracting an Ontology. An Example

At this point, we face the matter of building an ontology from sociodemographical data. It is worth to remark that geodemographics systems can be considered under three points of view: as datasets apt to be treated statistically, as systems to be interpreted by especialistas and, also as a semiformal representation of a geodemographic conceptualization. The last one is the sound one to be considered for building an ontology, although above two considerations have to be taken in account. To illustrate the results of that process, we present some features of the ontology that we have extracted from Mosaic. Mosaic comprises a range of geodemographic segmentation systems which, by mean of statistical techniques of classification, leads to classify individuals into groups (subdivided into more specific profile types) under qualitative and quantitative criterion. Mosaic products have been created by Experian Group, and some of them are, for example, Mosaic UK, Mosaic Public Sector, Mosaic Global and Mosaic Commercial. Each one is oriented to a business type. The design of these tools includes the use of geographic information systems and software for database management. Mosaic Global is a tool used in marketing to get the segmentation of consumers focused to the analysis and assessment on customer research: prospective, recruitment and loyalty. Groups are defined by attending to demographic and socio-economic features such as age, ethnicity, level of affluence and accommodation among Variables - Mean% and Index

Charts are provided for each of the variables used to build and describe Mosaic Public Sector. The variables are grouped together by category. For each group/type, the charts show the Mean% and Index for each variable.

Understanding Mean% and Index Mean% show the percentage of this group/type with this characteristic. For example, consider car ownership for Group A: Mean% 11.47 access to car 37.67 1 car 2 cars 38.97 212 11.89 3+ cars 0 50 100 150 200 This shows that 11.47% of Group A households have no access to a car. 37.67% of Group A households have 1 car. 38.97% of Group A households have 2 cars, etc.

Fig. 3. Explaining Mosaic charts

others. These characteristics are, in general, similar for a large proportion of the population in a concrete zone.

The online Mosaic Interactive guide brings the consumer visual classification and descriptive and statistical information backed up by more detailed information in the eHandbook. It offers quick synopsis for describing groups and types, their features and behavior as well as graphics used to build groups (see Fig. 3).

There exist other segmentation systems as CAMEO (UK) and ACORN (UK). ACORN and Mosaic provide detailed descriptions of a range of sociodemographic environments, explaining the reasons and the scope of each one. From the ontological point of view, they are *ideal types* in which documentation explicit information is not included [6]. They are archetypes which can be explained by statistical data (interpreted by expert scientist), but they do not fully characterize each class. Finally, we must also take into account that original definitions in Mosaic present difficulties to transform them into metadata: variance across individuals into a class, variance of requirements for belonging to the class or lack of critical requirements [21].

5.1 The Semantic Gap between Semantics and Geodemographics

In the case of Mosaic, it uses about 400 data variables, 11 groups and 61 types. From the point of view of its utility, there are some strengths: it make feasible to carry out interventions and services in a more specific way, to locate social marketing and identify social inequalities. Several limitations are the following: lack of transparency in methods for compiling and processing data (considering the statistical robustness of results performed by the tool as well as the ecological side of this issue) [11]. For example, the proportion of the average with respect to a characteristic often do not match with the description of the group. It is very frequent, in particular with data such as net income of the family, one of the most significant features of certain groups and types. It is usual to find cases, in principle, featured by earning high (or very high) incomes, but percentages for them match with low incomes, as it occurs in type O63 (see Fig. 4).

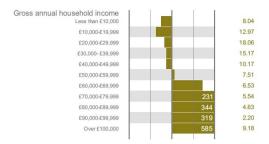


Fig. 4. Incomes in O63 Mosaic type

Other difficulty added is that there exist characteristics which are likely critical for the identification of groups (following their description), but these features have not been included in the definition because of sufficient statistical data have not been provided. This is what occurs, for example, with the (urban, rural, etc.) neighborhoods, the size of houses or tax rates. In other cases, even having enough statistical data, they look like void data, as the number of holidays. Finally, the information brought by the description is conflicting. As an example, we can cite the type O63. The age in this type is given by means of two ranges in the same document: 35-45 y 25-55 years. Other example can be found in type O61: on the one hand, we have "Well paid professional couples, often with children, choosing to live in diverse urban areas rather than the suburbs", and on the other hand, couples without children still (what matches with data).

These circumstances have made hard the classification and identification of the set of classes of the ontology, because many of the properties could not be included or have been included with low percentages. Further, we have to add the fact of several classes have been defined as conjunction of a range of properties and, therefore, the percentage of individuals belonging to these classes (fulfilling the properties) can be significantly reduced.

5.2 Methodology

The methodology to build the ontology consists of three stages (see Fig. 6):

1. Analysis of the geodemographic system:

- Analysis of geodemographic types used by the system
- Analysis of geodemographic data
- Analysis of expert's interpretation of geodemographic
- 2. Interpretation
 - Interpretation of types as (demographic) classes
 - Interpretation of data as (object or data) properties on classes
 - Interpretation of segmentation
- 3. Ontology engineering
 - Middle-out method for hierarchy construction
 - Axiomatization (actually necessary conditions) of classes by means the characterization of (object or data) properties
 - Interpretation of segmentation in axioms

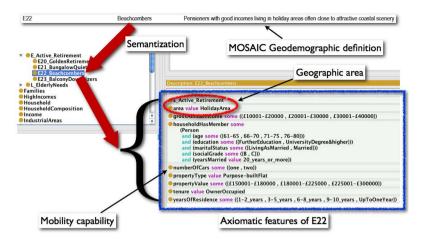


Fig. 5. Semantization of a Mosaic geodemographic class

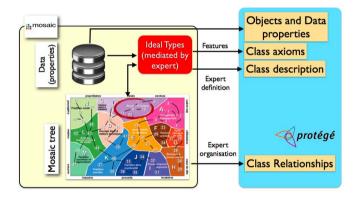


Fig. 6. Semantization of a geodemographic class

In each step the difficulties above commented have been faced. In general, there is a distance (separate by sociodemographics experts) between data presented in Mosaic and the description of every type, which is used to characterize (recognize) each class. From the point of view of the use of Mosaic in GCI, translated to an ontology, the definitions have to be profiled under weaker requisites but maintaining the richness of the information. For example, the class E22, as we show in Fig. 5, is described by including among the fundamental features, those ones which lead us to combine the information with mobility aspects, geodemographic zone, etc. Thus, that *axioms* can provide valuable information for the systems of the GCI. This option is the one we have choose by semantizing data and it provide us with a set of axioms (all of them are conditions or requisites for belonging to the class) which can be selected to get information about the Mosaic type we use (see Fig. 6).

6 Applications of Semantic Geodemography

In the context of the information ecosystem of a smarcity, the sphere of social knowledge added by a geodemographic ontology would influence all the processes of informational collect, interpretation and feedback, as much in the Urban Informatics scope as in the city management by leading the specialization of decisions and applications (see Fig. 1). The life cycle of knowledge in smartcities (including the acquisition, verification, documentation and decision) can be enriched with semantic processing of data, not only from sociodemographic ontologies. The value added by semantic technology allows us to mediate by (high level) reasoning with the processed knowledge. Of course, this aspect does not exclude the fact of data come from collaborative practices or *crowdsourcing*. Some of the main innovation lines in the field of applied semantic geodemography are related to smart cities (with their social features):

- Combined use with urban planning/landscape systems (as for example [16]). This combination facilitates knowledge to decide urban interventions. In emergent cities and regions which faces with the problem of their ground and developing [10], the reuse of this kind of knowledge could be possible.
- It facilitates the birth of knowledge-based markets for social products and services: location for new community centers, health service planning. It can estimate their social impact according available metadata.
- To increase urban resilience by means of the analysis of digital information [4] and the specialization of methods and process through metadata reasoning.
- It allows to interrelate the social dimension of distinct urban models [13]
- It facilitates the developing of *hyperlocal* social o community apps.

7 Conclusions and Related Work

In this paper we sketch the main lines of geodemographic ontology design and engineering. We also have pointed out the limitations, from Knowledge Engineering, to the develop of geodemographic systems. However, the opportunity and potential benefits of its application made the enterprise both interesting and necessary: ontologies provides knowledge to GCI. In general, geodemographic ontologies and metadata can enrich several urban subsystems (see Fig. 7 where a subset of properties of MOSAIC semantization are linked as knowledge source in the modelization of some urban subsystems).

There exist a number of works with aim to interrelate social and physical structure in cities and regions. Geodemographic ontologies have to be aligned with other semantic tools which shape geographic concepts as for example he Semantic Framework of the Universal Ontology of Geographical Space (UOGS) [17], mainly to a sound use of *location similarity* [17] concept in geodemographic field. In [2] introduces the *variograms* to determine binary similarities and their application on spatial data would allow the qualitative spatial reasoning with geodemographic zones.

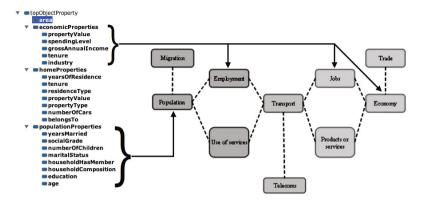


Fig. 7. Geodemographic knowledge for Urban Systems

With respect to urban ontologies, it is interesting to consider the relationship with $TownTology^1$ (see also [20]. Towntology project aims to develop ontologies for urban civil engineering, thus a geodemographic ontology can enhance the social dimension of the system, as well as to estimate the impact on the community of the urban intervention. The aligning of both ontologies is, however, problematic because there exist a gap among two disciplines and urbanist have to design ontologies which allow co-exploit metadata and knowledge.

There exist other geodemographic systems, as for example CAMEO², OAC³ by ONS/University of Leeds, ACORN⁴ or CLOUD CLIENT⁵ which can be semantized as well. In fact, the semantic interoperability among such systems is an interesting future research line.

Lastly, it is interesting to face the challenge of the revision of geodemographic ontologies. Urban ontologies have to be adaptive in some features because several urban subsystems have a dynamic nature [20]. Geodemography is static in essence, but its relationship with the urban dynamics can motivate ontology revision. The refinement (or reshaping) of demographic concepts can produce incoherences that ontology engineer can not explain. Thus, intelligent interfaces for ontology repairing have to be designed for this specific case which encompasses metadata and high level (rough) definitions.

References

1. Smarter Neighborhoods Smarter City. Solutions for a More Sustainable New York. Siemens Report, http://www.usa.siemens.com/sustainable-cities/

```
<sup>3</sup> http://areaclassification.org.uk/getting-started/
```

```
getting-started-what-is-the-output-area-classification/
```

```
<sup>4</sup> http://www.caci.co.uk/acorn-classification.aspx
```

¹ http://www.towntology.net/

² http://www.callcredit.co.uk/products-and-services/ consumer-marketing-data-and-segmentation/cameo-classifications/cameo-uk

⁵ http://cloud-client.co.uk/

- Ahlqvist, O., Shortridge, A.: Characterizing Land Cover Structure with Semantic Variograms. In: [17], pp. 401–415. Springer (2006)
- Alonso-Jiménez, J.A., Borrego-Díaz, J., Chávez-González, A.M., Martín-Mateos, F.J.: Foundational Challenges in Automated Semantic Web Data and Ontology Cleaning. IEEE Intelligent Systems 21(1), 42–52 (2006)
- 4. Aranda-Corral, G.A., Blanco-Escudero, A., Borrego-Díaz, J., Gomar-Acosta, M.: Data-in-the-cloud City. In: Proc. 7th Virtual Cities and Territories, Lisboa (2011)
- Batty, M.: Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies. In: Meyers, R.A. (ed.) Encyclopedia of Complexity and Systems Science, pp. 1041–1070. Springer (2012)
- Burrows, R., Gane, N.: Geodemographics, Software and Class. Sociology 40(5), 793–812 (2006)
- Burrows, R., Ellison, N., Woods, B.: Neighbourhoods on the net: The nature and impact of internet-based neighbourhood information systems. Joseph Rowntree Foundation (2005)
- Della Valle, E., Celino, I., Dell'Aglio, D.: The Experience of Realizing a Semantic Web Urban Computing Application. Transactions in GIS 14(2), 163–181 (2010)
- Díaz, L., Granell, C., Gould, M., Huerta, J.: Managing user-generated information in geospatial cyberinfrastructures. Fut. Gen. Comp. Syst. 27, 304–314 (2011)
- Dobbs, R., Remes, J., Manyika, J., Roxburgh, C., Smit, S., Schaer, F.: Urban world: Cities and the rise of the consuming class. Report of McKinsey Global Institute (June 2012)
- Dougan, S.: Using Mosaic en Islington, http://obesity.thehealthwell.info/node/26781
- Ellison, N., Burrows, R.: New Spaces of (Dis)engagement? Social Politics, Urban Technologies and the Rezoning of the City. Housing Studies 22(3), 295–312 (2007)
- Falquet, G., Métral, C., Teller, J., Tweed, C. (eds.): Ontologies in Urban Development Projects. Springer (2011)
- Little, E.G., Rogova, G.L.: Designing ontologies for higher level fusion. Information Fusion 10, 70–82 (2009)
- Millett, L.I., Estrin, D.L. (eds.): Computing Research for Sustainability. The National Academies Press (2012)
- Montenegro, N., Gomes, J.C., Urbano, P., Duarte, J.P.: A Land Use Planning Ontology: LBCS. Future Internet 4, 65–82 (2012)
- Riedl, A., Kainz, W., Elmes, G.A. (eds.): Progress in Spatial Data Handling. Springer (2006)
- Schumacher, P.: Parametricism. In: Leach, N. (ed.) Digital Cities, Architectural Design, pp. 14–45 (July/August 2009)
- Sheth, A., Henson, C., Sahoo, S.S.: Semantic Sensor Web. IEEE Internet Computing 12(4), 78–83 (2008)
- Teller, J., Cutting-Decelle, A.-F., Billen, R. (eds.): Future of Urban Ontologies. Proc. Final Conference of the COST Action C21 – Towntology: Urban Ontologies for an Improved Communication in Urban Development Projects (March 2009)
- Ure, J., Procter, R., Lin, Y., Hartswood, M., Ho, K.: Aligning Technical and Human Infrastructures in the Semantic Web: a socio-technical perspectiven. In: 3rd Int. Conf. on e-Social Science, Michigan, USA (2007)
- Wilson, A.: The Science of Cities and Regions. Springer Briefs in Geography. Springer (2012)
- Yang, C., Raskin, R., Goodchild, M., Gahegan, M.: Geospatial Cyberinfrastructure: Past, Present and Future. Computers, Environment and Urban Systems 34, 264– 277 (2010)