9 Exploiting Big VGI to Improve Routing and Navigation Services

Mohamed Bakillah, Johannes Lauer, Steve H.L. Liang, Alexander Zipf, Jamal Jokar Arsanjani, Amin Mobasheri, and Lukas Loos

CONTENTS

9.1 Introduction ................................................................. 175
9.2 What Is Big Data? .......................................................... 176
9.3 VGI as Big Data ............................................................ 178
9.4 Traditional Routing Services ............................................. 179
9.5 Routing Services Using Big VGI/Crowdsourced Data ............ 180
  9.5.1 Routing with Landmarks Extracted from Big VGI/
       Crowdsourced Data ....................................................... 180
  9.5.2 GPS Traces ................................................................ 181
  9.5.3 Social Media Reports ............................................... 182
9.6 Challenges for Exploiting Big VGI to Improve Routing Services .... 182
  9.6.1 Limitations of VGI and Crowdsourced Data .................. 183
  9.6.2 Impact on the Development of Routing and Navigation Services .... 183
    9.6.2.1 Interoperability ......................................................... 183
    9.6.2.2 Finding the Right Data ............................................ 183
    9.6.2.3 Analyzing and Interpreting Data ............................... 184
  9.6.3 Applicability of Big Data Solutions to Big VGI ................ 185
9.7 Summary ........................................................................ 186
References .................................................................................. 187

9.1 INTRODUCTION

The recent technological advances in data production and dissemination have enabled the generation of unprecedented volumes of geospatial data, giving rise to the paradigm of big data. Part of this trend is volunteered geographic information (VGI), that is, geographic information (GI) produced by volunteer contributors (Goodchild 2007), and crowdsourced data such as those obtained from social media. Whether big data refer to these huge volumes of data, which no longer fit traditional database structures (Dumbill 2013), or to the new technologies and techniques that
must be developed to deal with these massive data sets (Davenport et al. 2012), there is a common understanding that the full potential of such amount of data can only be exploited if information and knowledge with added value can be extracted from it.

As for many applications, the possibility to exploit massive data sets raises new opportunities to improve mobility services, including routing and navigation services. Traditionally, routing and navigation services relied on the digitization of the existing streets by commercial mapping agencies and mainly two companies (NAVTEQ and TomTom [formerly TeleAtlas]). The emergence of VGI and crowdsourcing-based applications is changing this perspective by offering a wide range of additional information that can be used to complement the official data produced by governments and mapping agencies. VGI and crowdsourcing applications have the potential to offer more diverse, more detailed, more local, and more contextualized data. Data from VGI and crowdsourcing applications with a potential of leveraging billion of contributors can fulfill the gaps found in official data that cannot be updated quickly at a low cost. For instance, if VGI and crowdsourcing applications could be exploited to some extent, developers could think of developing routing and navigation services that recommended routes personalized according to the travelers’ context. However, near real-time interaction with the changes detected on the road network through various sources could help reduce congestion, energy consumption, greenhouse gas and particle emissions, etc.

While VGI and crowdsourcing applications are clearly a sort of big geo-data, they feature some particularities that make their exploitation challenging at the moment. This chapter is dedicated to the analysis of the potential of VGI and crowdsourced data for improving routing and navigation services. The limitations of these types of data and how their exploitation is challenging are also discussed. Based on these limitations, we suggest some avenues for future research on the next generation of collaborative routing and navigation services.

This chapter begins with briefly reviewing the paradigms of big data and VGI. By providing a background to traditional routing and navigation services, we explain the types of big VGI data sources that could be exploited to upgrade routing and navigation services. Then, drawing from the limitations of VGI and crowdsourced data, we discuss the challenges for exploiting big VGI to improve routing and navigation services.

9.2 WHAT IS BIG DATA?

Recent technological advances in data production and dissemination have enabled the generation of unprecedented volumes of geospatial data. In 2012, it was estimated that the global volume of data was growing at a 50% rate each year (Lohr 2012), due to the increasing dissemination of digital sensors, smart phones, GPS-enabled devices, crowdsourcing applications, and social media, among other phenomena. While geospatial data have traditionally remained at the hand of experts (governments, mapping agencies), paradigms such as open data, social media, and collaborative mapping projects make it possible for an increasing proportion of these data to be virtually available to anyone, with the potential to benefit businesses, civil society, and individuals in general.
This huge amount of data has given rise to the term *big data*. Big data is a loosely used term that is employed to refer to two ideas. First, it is used to refer to the huge volume of data itself. For example, Dumbill (2013) states that “Big Data is data that exceeds the processing capacity of conventional data systems” and does not “fit the structure of your database architectures,” because of its size but also because of its dynamicity. Secondly, the term is also used, perhaps less frequently, to refer to the set of techniques that are being developed to deal with such volumes of data. For example, Davenport et al. (2012) report that the term is used to refer to “smarter, more insightful analysis” of large volumes of data, while Oracle defines it as “techniques and technologies that enable enterprises to effectively and economically analyze all of their data” (Oracle 2012). Indeed, big data in itself is of no great value unless we find means of managing and analyzing this less conventional data, which is not necessarily formatted according to the usual rows and columns of traditional databases. The question raised by big data is therefore how to extract information and knowledge from these raw data streams, since traditional approaches are not suitable for such amount and heterogeneity of data coming from various sources (Birkin 2012; Davenport et al. 2012).

In parallel with the aforementioned technological advances for data production and collection, storage technologies have also been significantly improved, making storage relatively cheap. In 2009, writing on the *Pathologies of Big Data*, Jacobs was already saying that “transaction processing and data storage are largely solved problems” (Jacobs 2009). Rather, major difficulties arise when it comes to extracting information and learning something from massive data sets. For example, Vatsavai et al. (2012) highlight the limitations of traditional spatial data mining algorithms such as mixture models for spatial clustering or the Markov random field classifiers for land cover analysis when confronted with massive data sets.

The range of techniques and technologies dedicated to big data is wide. Existing analytic techniques for extracting knowledge from the data are being improved to be able to deal with massive data sets. These techniques include SQL queries, data mining, statistical analysis, clustering, natural language processing, text analytics, and artificial intelligence, to name a few (Russom 2011). These analytic techniques can be deployed to improve performance of knowledge extraction algorithms such as social media analysis (Mathioudakis et al. 2010), change detection algorithms for high-resolution images (Pacifici and Del Frate 2010), and complex object recognition (Vatsavai et al. 2012), for example. The ability to deal with massive data sets is supported by underlying technologies such as Google’s MapReduce big data processing framework and its open-source implementation Hadoop, which is now considered by some as the de facto standard in industry and academia (Dittrich and Quiané-Ruiz 2012). MapReduce is a programming model that supports the development of scalable parallel applications for big data (Dean and Ghemawat 2004). It is based on a distributed file system where data, represented as (key, value) pairs, are initially partitioned in several machines before being processed. With MapReduce, data mining algorithms such as clustering, frequent pattern mining, classifiers, and graph analysis can be parallelized to be able to deal with massive data sets. Several ongoing researches are conducted to improve MapReduce, with, for example, enhanced join algorithms (Okcan and Riedewald 2011), query
Big Data: Techniques and Technologies in Geoinformatics

optimization techniques (Herodotou and Babu 2011), and indexing techniques (Jiang et al. 2010). While still being developed, these enabling technologies will certainly play an important role in the use of VGI.

9.3 VGI AS BIG DATA

Back in 1997, Goodchild pointed out that as networks became increasingly ubiquitous, the production of GI was moving from a centralized to a distributed process. Nowadays, users can produce GI via a variety of Internet applications; as a result, a global digital commons of geographic knowledge is created without having to rely solely on traditional geospatial data production processes (Hardy 2010). In 2007, Goodchild introduced the term volunteered geographic information to refer to the GI generated by users through Web 2.0 era applications. Later, Ballatore and Bertolotto (2011) stated that the VGI paradigm reflects the transformation of users from passive geospatial information consumers to active contributors. However, Coleman et al. (2009) argue that the concept of user-generated content is not new, referring for instance to public participation GIS where users can provide input and feedback to decision makers and involved communities through web-based applications. The novelty, they claim, lies in part in the community-based aspect of the users’ contribution to this digital commons of geographic knowledge (Coleman et al. 2009). VGI is often created out of the collaborative involvement of large communities of users in a common project, for example, OpenStreetMap (OSM) or Wikimapia, where individuals can produce GI that emanates from their own local knowledge of a geographic reality or edit information provided by other individuals. For example, in OSM, users can describe map features—such as roads, water bodies, and points of interest—using tags, providing information at a level that often goes beyond the level of detail that can be provided by traditional geospatial data producers (Goetz et al. 2012). As a result, and with the ever-increasing number of crowdsourcing applications, the volume of VGI is becoming huge, with no doubt that VGI is now an important component of big data.

Among the advantages associated with VGI, researchers highlight its use to enrich, update, or complete existing geospatial data sets (De Longueville et al. 2010; Goetz and Zipf 2011; Goodchild 2007; Gupta 2007; Tulloch 2008). This advantage is especially put forward in the context where traditional geospatial data producers, which are usually governments, may lack the capacity to generate data sets with comprehensive spatial and temporal coverage and level of detail (Gupta 2007; Tulloch 2008) such as those needed for routing services to be efficient. Furthermore, it was highlighted that VGI can be provided and disseminated in a timely, near real-time fashion, which is highly required in routing services. The advantages associated with VGI strongly suggest that this type of knowledge is highly valuable and is likely to help providing a dynamic picture of the environment (De Longueville et al. 2010; Mooney and Corcoran 2011). Nevertheless, the use of VGI for mobile applications such as routing and navigation is not yet fully achievable, as it is hampered by various obstacles related to large volumes, heterogeneity, and credibility.
9.4 TRADITIONAL ROUTING SERVICES

Vehicle routing is primarily based on digital geographic data. The off turning of selective availability of GPS by the order of the former US president Bill Clinton in the year 2000 (http://clinton3.nara.gov/WH/EOP/OSTP/html/0053_2.html) was the beginning of a new era of positioning on Earth. Since then, positioning has become more precise and mobile systems have been able to localize themselves precisely on the road. Consequently, routing is combined with navigation. The digitization of the existing streets by mainly two companies (NA VTEQ and TomTom, [formerly TeleAtlas]) provides the necessary data set for routing applications. In parallel, web services for routing were emerged. One of the most popular services is Google Maps that provides routes in many areas (http://googleblog.blogspot.de/2005/02/mapping-your-way.html). Since its launch in 2005, the service has been improved by the addition of public transport data and the integration of real-time traffic information. By developing new algorithms, especially those that incorporate hierarchical techniques, routing has become faster, particularly for long route calculations (Delling et al. 2009). Hierarchical techniques exploit the hierarchical structure of the road network to reduce the time required to compute queries (Bauer et al. 2008). Nowadays, vehicle routing can provide routes based on such criteria as distance, road type, and traffic. In terms of standards for interoperability, interfaces for location-based services (LBSs), and in particular for routing, were developed (Mabrouk et al. 2005).

With the arrival of VGI and, notably, of the OSM project, a second generation of route planning services is starting to emerge. VGI applications make routable data more easily available for free (Neis and Zipf 2007) and reduce information gaps. For example, the OpenRouteService.org uses OSM as a source of data (Figure 9.1). Users can add points of interest (POIs) as identified in OSM or search for POIs by name. Using other data sources (besides OSM) to search for POIs would enable users to

![OpenRouteService.org with selection of POIs.](image-url)
choose among a wider selection of POIs; however, extensive integration work would be required to merge the different heterogeneous data sets.

In addition, special routing services such as wheelchair routing, bike routing, and agricultural routing, to name only a few, are being designed. To provide the appropriate data required for the aforementioned types of routing services, new sources of data have been considered, for example, crowdsourced and sensor data. For instance, they help to consider weather conditions such as rain, ice, or snow, which can be measured by sensors and, therefore, the traffic conditions can be up-to-the-minute predicted. Undoubtedly, with the help of sensor data or crowdsourced information, it will be possible to consider traffic load, weather conditions, and physical road conditions to support near real-time route planning. For example, traffic jams caused by slippery roads can be avoided. But for realizing these visions, a closer look at the data is necessary.

9.5 ROUTING SERVICES USING BIG VGI/CROWDSOURCED DATA

Within a world where sensors are pervasive, we are now able to collect huge amounts of data. From weather stations, traffic information gathered by floating car data (FCD), data from vehicle telematics systems, mobile phone data, or sensor data of smart phones, the possibility for improving routing services through the use of such data is huge. Other data sources that can be of interest include social media like Twitter, Facebook, or Flickr. All of these data sets are voluminous (with size being dependent on the region). Consequently, these data sets fall under the big data paradigm because of their volume, dynamicity, and heterogeneity. As VGI services offer new types of data for example, photos, POIs, and events, VGI has the potential to substantially improve the performance and outcomes of the existing routing services. Some examples will be explained in the following section.

9.5.1 ROUTING WITH LANDMARKS EXTRACTED FROM BIG VGI/CROWDSOURCED DATA

Landmarks have been defined as any feature that can potentially serve as a point of reference (Lynch 1960). They play an important role when looking for the best route, and therefore, they help to increase the relevance and personalization of routing services (Duckham et al. 2010). Example of landmarks, identified in Zhu's classification of landmarks (2012), include social landmarks, which are places where people socialize and interact, such as parks and public places; economic landmarks, such as markets; and historical landmarks, such as monuments and historic buildings. It has been shown that such landmarks are important in forming mental representations of the environment (Couclelis et al. 1987).

The integration of landmarks in navigation and routing services involves two steps. The first step is the identification of features that could be considered as landmarks. The second step is the classification of landmarks, according to different parameters that indicate relevance, and the selection of the appropriate landmarks to be included in the routing preferences. There are a number of research projects focused on extracting and ranking of landmarks for routing purposes (Caduff and
Exploiting Big VGI to Improve Routing and Navigation Services

Timpf 2005; Duckham et al. 2010; Kippler and Winter 2005; Raubal and Winter 2002; Winter et al. 2008). Traditionally, landmarks are extracted from topological databases. However, other data sources have been used as well, such as web documents (Tezuka and Tanaka 2005).

Besides traditional spatial databases, VGI and crowdsourcing can be increasingly regarded as relevant sources for landmarks. Because VGI and crowdsourcing sources, in some regions, provide a more comprehensive spatial and temporal coverage and level of detail than sources provided by the governmental/commercial mapping agencies, they can complement these official sources where local features are identified from the point of view of users and local population. As an example, the Flickr photo portal is an interesting potential source for identifying landmarks. In Flickr, users can post photos, associate tags with these photos, and create groups of photos around particular themes. Photos, tags, and themes can be exploited to extract landmarks. The problem is that due to the very high volume of pictures and tags, as well as heterogeneity of tags, searching for landmarks in Flickr will result in a large number of irrelevant results. An appropriate landmark identification algorithm would need to cross-reference different information in Flickr (e.g., content-based image analysis, text, tags, social groups) in order to identify a relevant subset of landmarks. Similar problems arise when considering other potential sources of landmarks, such as OSM where objects are identified with (key, value) pairs that can use heterogeneous terms. Nevertheless, these sources have a huge potential to make more personalized routes where a large variety of landmarks can be selected according to users’ profiles and interests.

9.5.2 GPS Traces

Maps used for routing services are generally created from geographical databases. Such maps are not frequently updated because of the cost associated with these updates. Recently, interest in building road maps automatically from GPS traces as a complement to geographical databases has increased (Liu et al. 2012).

GPS traces are sequences of GPS coordinates generated by a moving device (e.g., instrumented probe vehicle). GPS traces can be obtained from cell phones, in-vehicle navigation devices, or GPS-enabled devices. Currently, most active road users, such as taxis, public transports, freight services, and agricultural vehicles, use navigation systems. Some of these vehicles are tracked by proprietary telematics systems (Lauer and Zipf 2013). These telematics systems can generate massive volumes of spatiotemporal data, from which relevant information that can improve routing services can be extracted. GPS traces allow extraction of vehicle trajectories and measurement of vehicle speed, idling, and traffic congestion. Such derived data can allow generating temporally detailed road maps, where the current speed of vehicles on road segments can be known in near real time to prevent other vehicles approaching congested areas. Using data on vehicle speed derived from GPS traces, we can also estimate fuel efficiency and greenhouse gas emissions. This would enable to make personalized route recommendations where greenhouse gas emissions and fuel consumption are minimized. However, deriving such data involves accessing and analyzing massive temporal GPS data.
9.5.3 Social Media Reports

Conventional traffic data collection methods are based on roadside inductive loop detectors, which are costly to deploy and maintain. As an alternative, social media can be leveraged to gather information on road conditions. A number of applications have been developed that allow sharing road network information. A popular example is Waze, which is a community-based traffic and navigation application (www.waze.com). The application can deliver alerts when road incidents such as accidents, road hazards, traffic jams, and meteorological events occur and are detected and shared by contributors. Waze integrates a traffic flow model derived from phone data in its routing engine to refine route calculation by adjusting edge weights. Another example is Traffic Pulse, a participatory mobile sensor web platform that relies on voluntary use of smart phones to report on road conditions (Li et al. 2012). This mobile application allows to querying and visualizing the city’s mobility information in real time. Other types of social platforms dedicated to more specific aspects of road networks are available, such as platforms to report on potholes (e.g., Pothole Info). Reporting of road incidents by several existing social media and web platforms presents a huge potential to keep drivers and travelers informed of the best route at any given time.

9.6 Challenges for Exploiting Big VGI to Improve Routing Services

9.6.1 Limitations of VGI and Crowdsourced Data

Credibility, reliability, and quality are among the main issues raised regarding VGI and crowdsourcing (Bishr and Mantelas 2008; Elwood 2008; Flanagan and Metzger 2008; Gouveia and Fonseca 2008; Jokar Arsanjani et al. 2013; Sieber 2007). VGI can be perceived as lacking credibility and reliability because it is produced by non-experts in a context that highly differs from the structured institution-initiated and expert-driven contexts (Elwood 2008). For example, while expert geospatial data producers are expected to generate data with a certain level of precision, users of VGI applications are not formally required to do so. As explained by De Longueville et al. (2009), users of VGI applications may have a vague memory of the geographic phenomenon they report on, or they could perceive only a certain portion of it. Another concern related to the quality of VGI is the fact that the profile and motivation of contributors are often unknown. As mentioned by De Longueville et al. (2010), the socioeconomic, sociological, and cultural aspects that characterize users can have an impact on VGI generation, by making it less reliable. Being aware of the relevant characteristics of the contributors can help properly interpret VGI and assess its quality and fitness for use, since evaluating the quality of geospatial data sets (e.g., VGI) in spatial data infrastructure is a challenging research problem (Mobashter 2013).

Besides the issue of VGI quality, the GI community still faces a number of obstacles regarding how VGI can be interpreted. VGI is often produced and stored using natural language, rather than agreed-upon terminologies and formal language usually employed in existing standardized geospatial database systems (Elwood 2008).
According to Scheider et al. (2011), the terms used by contributors to describe geographic phenomena lack \textit{unambiguous interpretation in terms of reproducible observations}. Nevertheless, some VGI applications, such as OSM, require contributors to employ a predetermined terminology (codified in an ontology). However, Scheider et al. (2011) indicate that it is difficult to reach a consensus regarding the terminology to use, while Mooney and Corcoran (2011) state that there is a lack of a mechanism for checking adherence to the agreed-upon ontology. As a result, the heterogeneity affecting VGI is likely to be more severe than the heterogeneity affecting traditional geospatial data (Grossner and Glennon 2007; Hyon 2007).

\section*{9.6.2 Impact on the Development of Routing and Navigation Services}

The aforementioned limitations associated with massive VGI and crowdsourced data sets have an impact on our ability to exploit them to improve routing and navigation services.

\subsection*{9.6.2.1 Interoperability}

In order to exploit VGI properly, various existing data sources must be conflated to routing and navigation services. However, this is not always possible due to interoperability issues. There exist a number of researches that have used VGI or crowdsourced data from a specific source for routing purposes; for example, Abassi et al. (2009) use Flickr tags and groups to identify landmarks. However, in order to improve routing services further, data fusion and data integration techniques must be applied to fuse diverse sourced data. For example, different characteristics of a landmark could be retrieved from different sources. This is currently hampered by heterogeneity of the formats, protocols, and semantics of these sources. A first avenue is therefore to use, adapt, or develop standards to support interoperability. Currently, there exist no standards to describe VGI or crowdsourcing applications. However, some research works are being conducted that suggest how to use existing standards to facilitate access to VGI and crowdsourced data. For instance, De Longueville et al. (2010) proposed a framework where OGC’s Sensor Web Enablement (SWE) standards, for example, sensor observation service (SOS), would support access to VGI sources, arguing that VGI can be described just as sensor observations. As another example, Auer and Zipf (2009) explore and demonstrate how OGC open standards can be used for user-generated content and, in particular, how the OpenRouteService uses data from OSM and WFS, WPS, WMS, etc.

\subsection*{9.6.2.2 Finding the Right Data}

Although having access to a large variety of sources creates a huge potential, it does not guarantee that the most appropriate data required for a given situation should be found easily. Certainly, VGI data contain redundant and noisy data as well, which demand for filtering and abstracting data. This is more necessary when it comes to providing more personalized routing services, where only the geographical objects of interest would be displayed on the map. Features that should be displayed on the routing and navigation map should be selected according to users’ context (his/her location, purpose of travel, mode of travel, POIs, etc.). This is not only because the
map should be personalized, however. Routing and navigation services are no longer limited to the 2D world but tend to increasingly include 3D representations of the environment. But experimental research reveals significant issues in using the third dimension to create effective data visualizations, thus the need to select only the elements that should be displayed, especially for mobile devices (Shepherd 2008). The need of displaying only a subset of relevant elements puts forward the need for modeling users’ context appropriately for mobile applications and gathering information on users’ context. The information on users’ context can also be obtained through the processing of VGI and crowdsourced data sets (e.g., logs, location, historic trajectories), which involves the use of efficient data mining algorithms of various types (text mining, clustering, etc.). Information on users’ context gathered from social media could also help gain insight into the drivers’ behavior, habits, and desires to recommend more suited routes.

Retrieving the relevant data can be achieved in different ways, for example, through pull-based or event-detection systems. Event detection systems should especially be considered for integrating recent changes in the road network that may affect the traffic, such as construction, meteorological events, or road accidents (Bakillah et al. 2013). This requires the coupling of various sources such as social media reports discussed earlier and the routing service. Change detection algorithms can also be leveraged too to identify changes in the road network. However, they must be able to deal with high-spatial and high temporal resolution images, which introduce important constraints on computational resources. Existing spatiotemporal data mining algorithms such as the spatial autoregressive model, Markov Random Field classifiers, and Gaussian process learning and mixture models need to be improved to be more scalable and able to achieve change detection with high spatiotemporal resolution images (Vatsavai et al. 2012).

9.6.2.3 Analyzing and Interpreting Data

Analyzing raw data from such variety of sources poses challenges due to high volume and heterogeneity of VGI and crowdsourced data. Heterogeneity, at the semantic level, means that different terminologies and representations are used to describe similar features of the real world. Because crowdsourcing and VGI applications do not necessarily adhere to standard terminologies, the heterogeneity problems are huge and far from being resolved. Heterogeneities hamper the ability to analyze and fuse heterogeneous data. Ontologies are used in some official spatial data sets to facilitate semantic interoperability; however, even for official data with controlled terminology, recent research in this domain demonstrates that it is difficult to establish consensus on a single ontology and difficult as well to bridge between the different ontologies used to describe different data sets. One solution that has been explored to deal with semantics of crowdsourced data is the use of folksonomies. The term folksonomy is a combination of the words folk and taxonomy (Wal 2007). It is a lightweight system to share resources and the user-created ad hoc labels associated with those resources. In a folksonomy, users are free to attach any term, called a tag in folksonomy nomenclature, to the resources in question. The control and design of a taxonomy are replaced in a folksonomy with the concepts of self-organization and emergence. There are several web-based applications
that demonstrate the usefulness of tags and folksonomies to gather user-generated information. These applications include Delicious for URL bookmarks, Flickr for pictures, and CiteULike for academic publications, which all use folksonomies. By enabling knowledge discovery and information filtering, folksonomies could be used to tackle the heterogeneity challenge of handling the large amount of user-generated data. However, folksonomies alone would only be one tool but could not be sufficient to resolve the heterogeneity problem, because folksonomies are heterogeneous themselves. More research is necessary to determine how folksonomies could be useful to support the fusion of different VGI sets.

Data obtained from VGI and crowdsourced applications also need to be analyzed for credibility. It is believed that such sources are not always credible because contributors are not professional and their motivation is unknown. Therefore, there is a strong need to develop methods to assess the credibility of VGI. However, it is not very likely that the motivation of individual users can be traced, since contributors are mostly anonymous. Therefore, if we cannot identify the contributors’ motivations, another avenue to assess the credibility of the contributed data is corroboration, that is, verifying if such information has also been provided by another source. This is linked to the aforementioned capacity to interoperate different sources with heterogeneous formats and semantics but also to the capacity to improve the efficiency of existing knowledge extraction and data mining algorithms with big data techniques to support data exploitation. Enabling technologies such as parallel computing, cloud computing, and virtual data marts is to be explored for this purpose.

The processing and analysis of VGI and crowdsourced data are also constrained by problems regarding the privacy of the contributors. For example, GPS trace data could allow tracking individual’s trajectory and traveling patterns, which is a potential threat to privacy and safety of the individual. To address this, one solution is to aggregate GPS trace data before making it available to applications.

**9.6.3 Applicability of Big Data Solutions to Big VGI**

The solutions that have been proposed to deal with big data have the potential to help enhance the interoperability, discoverability, analysis, and interpretation of large VGI data sets. However, solutions for big data are not sufficient alone, because, as explained in this chapter, VGI displays unique characteristics that differentiate it from other more conventional types of data. Table 9.1 examines some of the solutions that were put forward to deal with big data and explain their limitations or how they should be complemented with other techniques to be suitable for VGI.

Standardization (of information models and services) is one main solution to be able to deal with large amounts of data coming from different sources and ensure interoperability. One of the main drawbacks with respect to VGI is that it is difficult to impose standards to the VGI community. On the other hand, however, open standards being developed, for example, the OGC Web Processing Service, are more and more used by VGI applications because the open nature of these standards matches with the ideological objective of VGI, which is to make data open to anyone.

With respect to discovery and analysis capacity, the parallelization techniques such as MapReduce processing framework offer interesting potential to increase
processing capacity for performing these tasks. However, VGI data sets are intrin-
sically noisier than conventional data, so intensive data management quality must
still be developed for VGI in particular. A similar analysis can be conducted for
cloud computing, which allows to deal with volumes of data but does not address
the issue of quality.

Appropriate interpretation of data is one of the main keys to help discovery and
meaningful analysis of VGI data sets. Appropriate interpretation of data requires
explicit semantics associated with data. One of the solutions for giving meaning
to large data sets is Linked Data. Linked Data is a web of data coming from dif-
ferent sources, linked through Resource Description Framework (RDF) predicates
(Bizer et al. 2009). For example, in Linked Data, two entities (e.g., Department of
Geomatics Engineering and University of Calgary) can be identified by their unique
resource identifiers (URIs) and linked through the predicate within. As Linked Data
contains huge amount of data sets semantically linked to other data sets, it consti-
tutes a rich source for supporting the interpretation of data. However, to be appli-
cable to VGI data sets, there is a need for automated or semiautomated techniques
to generate semantically meaningful links between entities of VGI data sets and
entities in Linked Data. In this case again, the issues of noisiness and heterogeneity
are the main obstacles to establishing appropriate links.

9.7 SUMMARY

Routing and navigation services are moving from a relatively static, fit-for-all per-
spective to a highly personalized, near real-time routing and navigation experience.
The objective of routing services is no longer just to recommend the shortest route
but rather to recommend the route that takes into account a large variety of informa-
tion and preferences. These preferences include POIs, the very current road condi-
tions, and the optimization of various parameters such as energy consumption. In
addition, routing and navigation services are no longer exclusively considered as
personal services for individual consumers. They also present a collective interest.
Exploiting Big VGI to Improve Routing and Navigation Services

For example, while routing and navigation services can encourage drivers to choose routes that reduce energy consumption, they can also be considered as part of an integrated strategy by authorities and civil society to improve energy efficiency and to reduce atmospheric pollution.

This vision is partially driven by the emergence of VGI and crowdsourced applications, as well as by the increasing number of examples of successful use of these new types of data. However, as part of the big data paradigm, VGI and crowdsourced data are affected by characteristics that limit their use in routing and navigation services. These limitations, which were reviewed in this chapter, have an impact on the interoperability of such data, the ability to identify, filter, and retrieve relevant data for routing and navigation, and the ability to extract information from these data and to analyze and interpret them. Further research will be necessary to enable the use of standards for VGI and crowdsourcing applications that will facilitate interoperability, to extract from VGI and crowdsourced data the information on users’ context and their environment by exploiting big data techniques and technologies, to enable the production of temporally detailed roadmaps with current conditions and speed on road segments, to improve semantic descriptions and processing of VGI and crowdsourced data, and to protect privacy of contributors.

REFERENCES


Exploiting Big VGI to Improve Routing and Navigation Services


AUTHOR QUERIES

[AQ1] Kippler and Winter 2005 is cited in the text but not provided in the list. Please check.

[AQ2] Please check if edit to sentence starting “Existing spatiotemporal data…” is okay.

[AQ3] Please provide citation for the following references: D’Roza and Bilchev (2003), Klippel and Winter (2005), Lauer and Zipf (2009), Müller et al. (2010).

[AQ4] Please confirm the inserted publisher details for Abbasi et al. (2009), Scheider et al. (2011), Delling et al. (2009).


[AQ6] Please provide volume number and page range for De Longueville et al. (2009).


[AQ8] Please provide page range for Dumbill (2013).


[AQ10] Please provide accessed date for Mabrouk et al. (2005).
