Semantic Interoperability of Volunteered Geographic Information based on Contextual Knowledge

G. Bordogna^{*}, M. Boschetti, P.A. Brivio, A. Crema, D. Stroppiana

CNR-IREA Via Bassini 15, 20133 Milano, Italy *bordogna.g@irea.cnr.it L. Frigerio, T. Kliment, S. Sterlacchini^{*}

CNR-IDPA Piazza della Scienza 1, 20126 Milano, Italy *simone.sterlacchini.g@idpa.cnr.it

Abstract

This contribution proposes a customizable SDI architecture to create, map, manage and share VGI in a full interoperable way by exploiting external knowledge on the context for which VGI is created, consisting of both domain ontologies and a representation of the entities of interest. The whole process of VGI creation by a smart app, VGI deploy on the Web, and finally VGI fruition is described and examples from the *Space4Agri* project case study are proposed.

Keywords: Spatial Data Infrastructure, Smart Application, Volunteered Geographic Information, Semantic Interoperability, Ontology.

1 Introduction

Geographic Information Systems (GIS), Web mapping systems, Spatial Data Infrastructures (SDI) and smart applications for creating, sharing, mapping and analysing georeferenced information, are becoming very popular among local authorities to support territorial planning.

While GISs are tools that require expert users so that still nowadays they are mainly utilized by geographers and scientists who need to analyse and process geographic data, since the early 1990s, SDIs have been introduced to increase the availability and accessibility of heterogeneous geographic information on the Internet in a standardized way through geoportals, by anyone connected to the Internet by a Web browser [1]. SDIs offer as main functionalities the possibility to discover distributed geospatial data and processes, accessing, viewing, and querying geospatial data. These operations are made possible thanks to the interoperability among Web services, acting together through standardized well-defined protocols.

Heterogeneous multisource information may comprise authoritative information, certified by local or national organizations in charge of providing a specific theme to the public, such as meteorological data provided by public environmental agencies, traffic data provided by transport companies, etc.

With the diffusion of smart mobile devices connected to the Internet, smart applications have been designed to create Volunteered Geographic Information (VGI) by citizens for different purposes: to tag places they visited, to review commercial resources, or to point out critical environmental conditions in need of interventions [2]. Such information, although not always recognized by the geospatial community for the high heterogeneity of its accuracy and questionable reliability [3], is becoming more and more used for both commercial and territorial planning due to several advantages it offers: it is provided free of charge and potentially by a continuously incrementing number of contributors; it is continuously updated in real time or near real time; it is created in situ, and thus is the result of a direct observation; and, finally, is characterised by high heterogeneity of contents which can offer complementary views at different levels of details of the same reality by enriching, correcting and completing official information available on the same territory. Its structure and format is very heterogeneous, consisting of free texts, pictures, videos, sounds, tags from a controlled vocabulary, associated with a geographic footprint that is either a point or an explicit extended area on Earth. The footprint is generally automatically assumed as the geographic coordinates of the place where the information was created, as detected by the GNNS sensor of the smart device, and this setting may contribute to the inaccuracy of VGI.

Until recently the exploitation of VGI by Web applications has been prevented for the difficulty of integrating VGI created by smart applications with authoritative geospatialdata. This is due to the lack of applications compliant with standards to publish on the Web the created VGI [4]. Applications platforms such as Ushahidi, or platforms for collecting and storying crowdsourced geo-referenced data within citizen science projects (see for instance inaturalist at http://www.inaturalist.org/, ebird at http://ebird.org/, "Did you feel it?" at http://earthquake.usgs.gov/data/dyfi/) are generally closed worlds [5]. To see and analyse VGI often there is no other alternative than accessing the application geoportal and downloading the data when made available as open data. This practice makes a burden task the correlation of VGI created within a project with both VGI by other applications, and authoritative geospatial-data.

On the other hand, SDI architecture allows distributing on distinct nodes both Web GIS servers, providing data sharing via Web services, and geoportals that allow accessing data, so that a single information source can serve different Web clients, and vice versa. This architecture is well suited to organize the access to VGI collected within distinct projects. Nevertheless, for a full interoperability, beyond the technical aspects of data sharing guaranteed by the use of standard Web services, applications creating and sharing VGI should provide support for a semantic interoperability [6], which is often missing. Applications using data schema related to different or even similar domains, should provide means to allow consumers of VGI to understand the meaning of the used terminology. To this aim, some emerging approaches like [7] enrich VGI calling upon the concept of the Semantic Web and Linked Data, i.e., the modality for sharing, and connecting RDF data via URIs on the Semantic Web [9]. Nevertheless, when creating VGI from a mobile device not connected to the Internet, Linked Data cannot be accessed and used.

In this paper we propose an SDI to manage and share VGI by providing semantic interoperability based on the use of external knowledge, stored both on the device of the VGI creator and in the database of the project¹. Since it adopts an SDI it allows sharing VGI on the Web by means of standard geospatial-services: one could then access distributed heterogeneous geospatial-data, comprising several different sources of VGI, through a single or multiple access points. Furthermore, both VGI creators and VGI consumers have the support of external knowledge to describe/analyse their observations: domain ontologies supply guided taxonomies to catalogue the observations, while a spatial data layer representing the entities of interest to the project allows conflating the observations footprints thus resolving potential imprecisions of the geo-localization. In the paper we will illustrate the architecture and describe its components by making examples drawn from the real study case of the Space4Agri project [10].

2 The Interoperability Problem of VGI

In order to increase interoperability of geospatial Web services, the Open Geospatial Consortium (OGC) has developed a number of open standards for geospatial data encodings and services. The use and integration of these standards make it possible to set up an SDI to share, discover, visualize, and evaluate geospatial data at different scales on the Internet in a distributed manner. The Global Earth Observation System [10] and the European Infrastructure [12] describe the architecture and components for designing and implementing SDIs. Within SDIs, metadata and catalogue services are key factors for enabling the discovery of spatial data [13]. Nevertheless, the adoption of a top-down approach to set up an SDI, limits the possible potential of bottom-up geographic data that may be contributed by volunteers, favoured by the widespread development of the geospatial Web and the increasing popularity of smart devices connected to the Internet. The management of VGI is characterized by the absence of standards for deployment and retrieval. This deficiency has caused the creation of ad hoc approaches and, consequently, of great heterogeneity of VGI - both in semantics, formats and contents. This heterogeneity can potentially be of great value to offer a collaborative view of the reality complementary to that provided by scientific and authoritative geospatial data, and thus VGI integration within SDIs would be of great usefulness [14][15]. This issue has been faced in the literature by proposing several solutions for VGI management within SDIs such as in [16] that defined a middleware component to provide a single search interface to various Web 2.0 repositories to the aim of discovering information across different social networks. They developed a prototypal discovery service that implements the OGC OpenSearch Geo-Time standard interface specification. Another approach proposed the retrieval of Web 2.0 data directly from their services by means of an INSPIREcompliant SDI discovery service that becomes a unique entry point for the data retrieval [17]. These approaches face the problem of syntactic interoperability of VGI sharing, integrating existing heterogeneous VGI with respect to the formats and services while disregard semantic interoperability of VGI from different sources. This last issue has been recently faced by two distinct approaches: either by adopting the Semantic Web and Linked Data framework [7][8] or by relying on domain ontologies [18]. Nevertheless, when creating VGI from a mobile device not connected to the Internet, Linked Data cannot be accessed and used. This is the main reason that motivates our approach to adopt domain ontologies for creating a crowdsourcing platform with related services that deploys full interoperable VGI both when creating and consuming VGI.

3 A Spatial Data Infrastructure for Full Interoperable VGI Management

The Cross-Community Interoperability initiative of the OGC [19] suggested recommendations to create, assemble, and disseminate VGI. Specifically, the initiative envisages a better understanding of both service contents and the relationships that exist between OGC services and the resources and their contents by the following tasks:

- the creation of ontologies that define the shared vocabulary used to model a domain with the definition of concepts, as well as their properties and relations;
- the extension of the use of OGC services and standards for VGI deploy and access;
- the expansion of gazetteer functionalities to include conflation and semantic linking.
- the management of the uncertainty and imprecision of VGI.

Our proposal to create and manage VGI retains these recommendations by fostering the use of domain ontologies to create and analyse VGI contents, OGC standard Web services to publish and access VGI on the Web, and conflation techniques of the geo-footprint of VGI in order to cope with ambiguities and imprecision of the geometric dimension.

Figure 1 depicts the layered architecture of our proposed framework. It comprises four main layers (processing environments) communicating only with the adjacent layers by standard JSON messages. Each layer has its own software and hardware. From left to right we have the "VGI creation environment", consisting of a smart application (smart app for short, that we have developed for Android) installed on the mobile devices of VGI authors. It allows both creating VGI and exchanging messages with the "geo database and VGI management environment" whose task is twofold: on the one side it sends/receives data to/from the smart app and stores it into the geo database; on the other side it answers SQL/Spatial queries received by the "geo Web deploying environment".

¹ The activity has been supported by both the Space4Agri project jointly funded by CNR and Regione Lombardia under contract n.18091/RCC, 05/08/2013, and the RITMARE Italian National Flagship Project (SP7 Subproject) funded by the Italian Ministry of University and Research.

This last layer comprises a Web GIS server, (in our implementation of the architecture we installed Geoserver). and a Catalogue server (we installed Geonetwork), which periodically deploy on the web the VGI and metadata respectively, by creating views of the geo database (in the implementation at every geo database updating). On the other side, these components answer OCG standard requests for both maps and time series of maps (WMS, WMS-T), features (WFS), coverages (WCS) and metadata (CSW) respectively, submitted by a user, in this case a VGI consumer, using a client geoportal of the "discovery, view and analysis environment". Indeed, the rightmost layer may be considered outside the proposed framework. In fact, it can be any web/desktop based client that allows accessing and using metadata and data by means of those OGC services (for example a desktop GIS like QGIS or a thin web client, like the S4A geoportal developed for the study case of the Space4Agri project² [10]).

Figure. 1 Architecture of the SDI for interoperable VGI creation and management.



Source: own processing.

This architecture supports both syntactic and semantic interoperability. While syntactic interoperability is guaranteed by the use of standard Web services to share VGI, semantic interoperability is supported by providing external knowledge on the context of the project: it comprises a "domain ontology", represented in both compact and extensive form, and a georeferenced representation of the entities of interest in the project, hereafter named "conflation data layer".

The smart app, when first installed on a mobile device, receives from the adjacent processing layer, together with registration data and control variables, a compact representation of the domain ontology that is locally stored for future use in order to support the volunteer in creating "understandable" VGI for the specific purpose. By storing the ontology on the mobile device card, the smart app can be used even when the internet connection is disabled or not available. On the other side, the domain ontology is displayed to the consumer of VGI when submitting WMS and WFS requests through the geoportal, so as to allow the correct interpretation of the VGI context: the ontology can be viewed in its compact form (as a legend) when analysing the contents of VGI items, or in its extensive form, by clicking on the URL to access a Web page where it is described. This SDI architecture can potentially be replicated at distinct nodes to set up multiple crowdsourcing platforms serving distinct citizen science projects. Each node has the same components but distinct domain ontologies and conflation data layers, i.e., different entities of interest, so as to enable the creation of VGI with the correct meaning for the context. On the rightmost laver, a consumer of VGI can access multiple SDI nodes in the same session of data analysis. When asking WMS or WFS requests to a node, he/she receives together with the VGI the correspondent domain ontology, thus enabling semantic interoperability of VGI at the consumer side.

3.1 VGI creation

The smart app is installed on a smart mobile device by which users can create VGI, supported by compact hierarchical domain ontology. Depending on the context for which the smart app is customized, users can tag entities by selecting concepts (from the domain ontology) describing some interesting features or classes of the observed entities, which are visualized on the mobile device screen by a multilevel hierarchical menu. This guided creation of VGI items serves the twofold objective of normalizing the VGI contents by guiding the volunteers towards a correct creation as well as of providing a consistent semantic contextual specification complementing the free textual annotations that the VGI creator can also provide to enrich the description of his/her observations. In fact, the administrator of the smart app can customize the menu to show helpful ontologies of terms, concepts and categories, whose selection depends on the purpose for which the smart app should be used. Figure 2 reports an extract from³ the extensive BBCH ontology used by agronomists to specify the growth stages of crops (table 1 reports the principal crop stages descriptions while Figure 2 displays an example of the second level, referred to rice growth stages). It can be seen that each code corresponds to a picture and a textual description. The smart app stores on the mobile device a compact version of the same ontology. For example, by using the Space4Agri smart app (free download from Google Play Store⁴) the user is guided to tag the crops observed on the field by a synthetic description of the crop stages: he/she can select from a multilevel menu the BBCH codes that best match the current observation (see Figure 3). Furthermore, the user can also add a picture and a short description in natural language to complement and enrich the tag selected from the ontology. The created VGI items are automatically geo-referenced by associating them with the geographic coordinates of the current GPS user position. The user can confirm this geo-reference or can move the pin to a different location just by clicking on the visualised map. This feature is a useful option that the volunteer can exploit either

² http://space4agri.irea.cnr.it/it

³https://www.politicheagricole.it/flex/AppData/WebLive/Agrometeo /MIEPFY800/BBCHengl2001.pdf

⁺ https://play.google.com/store/search?q=space4agri&c=apps

to correct imprecise geo-localization supplied by the GPS, thus increasing the positional accuracy of the created VGI, or to set the geo-reference of VGI when the observed entities are at a sight view from the current GPS user location. In order to use the smart app one must have a valid email address that uniquely identifies the volunteer as author of his/her created VGI. This contextual information is also important in order to allow the creation of suitable metadata of VGI items and possibly to set up quality assessment processes to identify malicious or trustful volunteers by analysing the characteristics of their created information.

Tał	ble	1:	Pri	inci	pal	gro	wth	sta	ges	of	cro	ps

Stage	Description
0	Germination / sprouting / bud development
1	Leaf development (main shoot)
2	Formation of side shoots / tillering
3	Stem elongation or rosette growth / shoot development (main shoot)
4	Development of harvestable vegetative plant parts or vegetatively propagated organs / booting (main shoot)
5	Inflorescence emergence (main shoot) / heading
6	Flowering (main shoot)
7	Development of fruit
8	Ripening or maturity of fruit and seed
9	Senescence, beginning of dormancy
-	

Source:https://www.politicheagricole.it/flex/AppData/WebLive/Agro meteo/MIEPFY800/BBCHengl2001.pdf.

Figure 2. BBCH extract for some rice growth stages. Rice



Rice Lancashire et al., 1991

Code	Description
Principa	al growth stage 4: Booting
41	Early boot stage: upper part of stem slightly thickened, sheath of flag leaf about 5 cm out of penultimate leaf sheath
43	Mid boot stage: sheath of flag leaf 5–10 cm out of the penultimate leaf sheath
45	Late boot stage: flag leaf sheath swollen, sheath of flag leaf more than 10 cm out of penultimate leaf sheath
47	Flag leaf sheath opening
49	Flag leaf sheath open

- 52 20% of panicle emerged
- 53 54 30% of panicle emerged 40% of panicle emerged

Source:https://www.politicheagricole.it/flex/AppData/WebLive/Agro meteo/MIEPFY800/BBCHengl2001.pdf.

Figure 3. Space4agri smart app screenshots showing the main menu for VGI creation and the first and second level entries of the compact BBCH agronomic ontology.



Source: own processing.

3.2 The geo database and VGI management

The smart app communicates via wireless or via mobile network with a back-end application installed and running on a Web server (second frame from left in figure 1) where a geo database stores the VGI items created by volunteers. The back-end application can send/receive JSON messages to/from the smart app, and can read/write into the geo database. When receiving a new VGI item from a smart device, the back-end application analyses the geo-reference and possibly conflates it with those of other VGI items already stored in the geo database, by exploiting the "conflation data layer" (when provided), representing the known geographic entities of interest to be considered the targets of volunteers observations. The administrator can change the conflation data layer to customize the framework to an application: all VGI items with geo-references included in the boundary of an entity in the conflation data layer are associated with that specific entity. When no entity boundary contains the VGI coordinates, the VGI item can be associated either with a null entity, or with more entities to distinct degrees inversely proportional to their distance from the VGI geo-reference. This conflation allows also disambiguating the VGI geo-reference and customizing the smart app to associate VGI to geographic entities of interest for a specific project. The "conflation data layer" makes sense when the footprints of the entities of interest are known, such as in the case of projects collecting reviews of POIs, or agronomic information on crop fields. Nevertheless, in the case the entities of interest do not have a known localization, such as in the case of birds in ebird project and flora or fauna in inaturalist project, the "conflation data layer" could still be useful to group the VGI items that have a geo-reference within the boundary of contextual entities of interest, such as the boundary of parks or natural ecosystems. On the other side, if VGI items are created far from the entities of interest, they can still be analysed and either validated by a moderator or filtered out as spam. For example, in inaturalist VGI of flora and fauna in unexpected ecosystems might be important information on the environment. For example, in Space4Agri project the entities of interest are cultivated fields, and thus the Space4Agri smart app associates agronomic categories of the BBCH ontology to the fields surveyed in the agronomic cadastral database, which is provided as the conflation data layer. VGI items outside the fields are dealt with as outliers and stored in a different layer (Figure 4 shows some fields tagged with VGI).

Figure 4: *Space4Agri* geoportal displaying crop fields tagged with VGI; the distinct colours of the fields represent distinct crop types. The table displays all the VGI items corresponding to all observations of the BBCH growth stages in the corn field marked with the blue pin.



Source: own processing.

3.3 The VGI deploy on the Web

This central component of the framework comprises a Web GIS server that deploys the VGI items stored into the geo database into distinct views on the Web, the conflation data layer, and possibly other multisource geospatial data of interest to the project via OGC Web Services. Specifically, the geo Web service provides distinct views of the VGI items created by the smart app, that are deployed in distinct vector layers depending on their type: a single layer is deployed for all VGI items containing pictures and free text; furthermore, each set of VGI items containing tags of the same category or subcategory is deployed as a single layer. A metadata template based on INSPIRE and National regulations, is created for each type of VGI layer with basic information describing the common characteristics of its VGI items, i.e., category type, link to the URL of the used extensive domain ontology, and common keywords to empower data discovery. These templates are automatically completed with the authors' name and creation date and updated (mainly the bounding box) whenever new VGI items are added to the corresponding layer. These metadata are periodically harvested from the WMS service endpoint deployed for the data collected by the smart app and indexed by the catalogue server to maintain updated metadata.

3.4 The discovery, view and analysis of VGI

The final processing layer depicted on the right of figure 1 can be any OGC standard application. Within *Space4Agri* project we developed an OGC geoportal that can access the created VGI together with other geo-data of the project. The geoportal can display the distinct views of the VGI items: as pins that one can click and open to see pictures and free text, as markers with distinct colours to distinguish the types of crops of the BBCH ontology described in the associated legend. One can also see in distinct colours the entities of the conflation data layer that have been tagged with a BBCH category by at least a VGI item (see Figure 4), and finally the map of VGI outliers, not associated with any surveyed field.

4 Conclusions

The paper describes the architecture of an SDI for creating, mapping and managing full interoperable VGI. Its main characteristic is the possibility to customize the creation of VGI and its fruition to meet distinct needs. The customization is achieved by ingesting into the SDI external knowledge on the context of the project in the form of domain ontology, specifying the concepts to be used when creating VGI, and a conflation geo-spatial data layer representing the target entities of VGI. These are used to help volunteers to limit both the ambiguities and imprecisions when creating VGI, as well as consumers when analyzing VGI. The evolution of this approach will be towards modeling the imprecision and uncertainty that might happen due to both deficiencies of the observation conditions that may limit the capacity of the observer to clearly identify the distinguishing characteristics of a target entity in order to select a tag from the ontology, and inadequacy of the volunteer knowledge of the problem, that might cause low quality VGI. To this aim we plan to extend the classic ontology approach towards the adoption of fuzzy ontologies.

References

- D. J. Maguire, P. A.Longley, The emergence of geoportals and their role in Spatial Data Infrastructures, *Comp., Env. and Urb. Syst.*, 29 (1), 3-14, 2005.
- [2]. M.F. Goodchild, Citizens as sensors: The world of volunteered geography., *GeoJournal*, 69, 211–221, 2007.
- [3]. M. Haklay, A. Singleton, C. Parker, Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, 2(6), 2011-2039, 2008.
- [4]. M. Craglia, Volunteered Geographic Information and Spatial Data Infrastructures: when do parallel lines converge?, *Position paper for the VGI Specialist Meeting*, 12/13-14/2007, Santa Barbara M., 2007.
- [5]. G. Bordogna, C. Carrara, L. Criscuolo, M. Pepe, On predicting and improving the Quality of VGI projects, *Int. J. of Digital Earth*, 9(2), 134-155, 2016.
- [6]. M. Bakillah, S.H.L. Liang, A. Zipf, J.J.Arsanjani, Semantic Interoperability of Sensor Data with Volunteered Geographic Information: A Unified Model, *ISPRS Int. J. Geo-Information.* 2, 766-796, 2013.
- [7]. C. Stadler, J.Lehmann, K.Höffner, and S.Auer, LinkedGeoData: A core for a web of spatial open data, *Semantic Web*, 3, 333–35, 2012.
- [8]. S. Ronzhin, Semantic enrichment of Volunteered Geographic Information using Linked Data: a use case scenario for disaster management, *Master of Science Thesis, University of Twente*, July 2015.
- [9]. C. Bizer, T. Heath, and T. Berners-Lee, Linked data-the story so far. *Int. J. on semantic web and information systems*, 5(3), 1-22, 2009.

- [10].G. Bordogna, T. Kliment, L. Frigerio, P.A. Brivio, A. Crema, D. Stroppiana, M. Boschetti, S. Sterlacchini, to appear in *Int. J. of Geo-Information*, open access, 5, doi:10.3390, www.mdpi.com/journal/ijgi, 2016.
- [11].M.N. Butterfield, J.S. Pearlman, and S.C. Vickroy. A System-of-Systems Engineering GEOSS: Architectural Approach, *IEEE Syst. J.*, 2(3), 321-332, 2008.
- [12].INSPIRE EU Directive: Directive 2007/2/EC of the European Parliament and of the Council of 14 March establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L 108/1, 50, 2007
- [13].Ming-Hsiang Tsou, An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet, In *Geographic Information Science* (GIScience 2002), M. Egenhofer and D. Mark (eds.), LNCS, 2478, Springer Verlag, Berlin, 313-332, 2002.
- [14].S. Nativi, M. Craglia, and J. Pearlman. The Brokering Approach for Multidisciplinary Interoperability: A Position Paper, Int. J. of Spatial Data Infrastructures Research, 7, 1-15, 2012.
- [15].P. Mooney, P. Corcoran, Can Volunteered Geographic Information be a participant in eEnvironment and SDI? . Hrebicek, G. Schimak, and R. Denzer (Eds.): in *ISESS* 2011, *IFIP AICT* 359, 115–122, 2011.
- [16].L. Díaz, M. Núñez-Redó, D. González, J. Gil, P. Aragó, E. Pultar, J. Huerta, Interoperable Search Mechanisms for Web 2.0 Resources, *Int. J. of Spatial Data Infrastructures Research*, 7, 277-299, 2012.
- [17].M. Núñez-Redó, L. Díaz, J. Gil, D. González, J. Huerta, Discovery and integration of web 2.0 content into geospatial information infrastructures: A use case in wild fire monitoring, in Availability, Reliability and Security for Business, Enterprise and Health Information Systems, LNCS, 6908, Springer Verlag, 50-68, 2011.
- [18].I. Martínez Ramos, A. Vandecasteele, Devillers, Semantic Integration of Authoritative and Volunteered Geographic Information (VGI) using Ontologies, *in Proceedings of AGILE* 2013.
- [19].CCI Cross Community Interoperability- Annex B-CCI Thread scope, Available at: http://www.opengeospatial.org/pub/www/ows10/rfq/ann exb-cci.html#cci-thread-scope [10 -2-2015] (2015)