Web 2.0 Broker: A standards-based service for spatio-temporal search of crowd-sourced information

Laura Díaz [a,*], Carlos Granell [b,1], Joaquín Huerta [a,2], Michael Gould [a,3]

[a] University Jaume I, Institute of New Imaging Technologies, Av. Sos Baynat s/n, 12006 Castellón, Spain
[b] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, I-21027 Ispra (VA), Italy

ABSTRACT

Recent trends in information technology show that citizens are increasingly willing to share information using tools provided by Web 2.0 and crowdsourcing platforms to describe events that may have social impact. This is fuelled by the proliferation of location-aware devices such as smartphones and tablets; users are able to share information in these crowdsourcing platforms directly from the field at real time, augmenting this information with its location. Afterwards, to retrieve this information, users must deal with the different search mechanisms provided by the each Web 2.0 services. This paper explores how to improve on the interoperability of Web 2.0 services by providing a single service as a unique entry to search over several Web 2.0 services in a single step. This paper demonstrates the usefulness of the Open Geospatial Consortium's OpenSearch Geospatial and Time specification as an interface for a service that searches and retrieves information available in crowdsourcing services. We present how this information is valuable in complementing other authoritative information by providing an alternative, contemporary source. We demonstrate the intrinsic interoperability of the system showing the integration of crowd-sourced data in different scenarios.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

Geospatial Information Infrastructures (GIIs), also known as geospatial cyberinfrastructures (Yang, Raskin, Goodchild, & Gahegan, 2010) and Spatial Data Infrastructures (SDIs; Masser, 2005), have provided scientists and public sector organizations with instruments for organizing, sharing, access and exploiting the large amount of geospatial content for earth sciences decision-making. GIIs consist of a network of distributed nodes based on well-defined architectural styles and standards specifications. A given GII publicly exposes a set of geospatial web services, according to principles of Service Oriented Architectures (SOA), to allow software clients to discovery, access and retrieve geospatial data from them. To increase interoperability, the Open Geospatial Consortium (OGC) has promoted the creation of consensus specifications for data encodings and service interfaces to standardize communication protocols between these clients and services.

In parallel to the evolution of GIIs, we are witnessing the consolidation of a new generation of social networks and Web 2.0 services, which are characterized by greater levels of user participation and ease of content sharing. The Web is now a collaborative environment that has turned users into active providers (Coleman, Georgiadou, Labonte, 2009) capable of generating massive amounts of new content. This user-generated content often is georeferenced according to a user's location at the time of data capture. This means that a large amount of georeferenced user-generated content is increasingly available in a wide variety of topics.

A clear dichotomy has emerged between top-down and bottom-up approaches to georeferenced data collection and sharing. On one hand, GIIs rely on top-down methodologies, constrained by standards but which do not consider the participation of citizens outside the official system of creation and publication of authoritative geospatial data, collected and validated only by official institutions and mapping agencies. Moreover, the lack of assisted tools for data publication (Díaz, Granell, Gould, & Huerta, 2011) combined with the resource-demanding and time-consuming task of continuously updating these datasets (Zlatanova et al., 2009) have set real barriers in efficiently exploiting authoritative geospatial data. On the other hand, social networks are eminently user-centric and follow bottom-up methodologies whereby users are free to produce and share their data using a diversity of aggregation

* Corresponding author. Tel.: +34 964 729 078; fax: +34 964 728 730.
 E-mail addresses: laura.diaz@uji.es (L. Díaz), carlos.granell@jrc.ec.europa.eu (C. Granell), huerta@uji.es (J. Huerta), gould@uji.es (M. Gould).
 [1] Tel.: +39 0332 78 5758; fax: +39 0332 78 6325.
 [2] Tel.: +34 964 728 319; fax: +34 964 728 435.
 [3] Tel.: +34 964 728 317; fax: +34 964 728 435.

and broadcast technologies that become de facto standards at best. These data are in many cases more timely and, therefore, more useful during emergencies and time-critical scenarios (Goodchild, 2007). The diversity of solutions extends to service providers in terms of different publication capabilities and variability of application programming interfaces (APIs) offered by each Web 2.0 service. This makes ad hoc data access and retrieval difficult for client applications because, as opposed to GIIs, standard service interfaces and data encodings are not commonly followed.

The integration and interrelation of user-generated data with authoritative geospatial datasets may potentially bring together the best of both worlds for decision-making in many real-world scenarios such as the assessment of environmental issues (Granell, Díaz, & Gould, 2010). In this context, however, to assess the above hypothesis we need suitable mechanisms to seamlessly search and retrieve datasets from both worlds. Our ultimate goal is then to improve the interoperability among the available heterogeneous Web 2.0 services, so that distinct application scenarios can benefit from the integration of these search results and official datasets. To that end this paper addresses the following research questions:

- How can we retrieve user-generated content based on spatio-temporal criteria from multiple crowd-sourcing services?
- Under which circumstances can the integration of authoritative and non-authoritative data be valuable in scenarios beyond the mere visualization of these datasets such as for analysis and decision-making activities?

We propose an integrated, scalable software brokering solution based on standard specifications and well-defined software engineering patterns. We present a middleware component that *mediates* between client applications and backend Web 2.0 services and social networks. This component, called the Web 2.0 Broker (W2B), is implemented as a web service providing a single search and retrieval interface based on the OGC OpenSearch Geo-Time specification (Gonçalves, 2010). The component has been shown to improve interoperability in the discovery and retrieval of information across different crowd-sourcing services.

The remainder of this paper is structured as follows: Section 2 defines the foundations of this work. Section 3 highlights the most important aspect of the Web 2.0 Broker from a technological point of view. Section 4 introduces the reference architecture to make possible the integration of our solution in the proposed systems and the description of a set of scenarios. Section 5 presents a discussion on the described work. Section 6 ends the paper with some conclusions.

Foundations

This section describes in detail the foundational pillars on which the Web 2.0 Broker service rests.

Crowd-sourcing services: bottom-up approach

Due to the easiness to use data-creation tools available as Web 2.0 services, millions of users have become data producers. At the core of the Web 2.0 vision, simplicity allows users to produce and share their own data through "one-click" services. User-generated content is multidisciplinary and heterogeneous, making it an invaluable source of social data to be used by scientific and business communities. Nevertheless, at the same time, there is still a need for adequate tools for data search and retrieval (Munro, 2012).

An additional factor on the shift in the role of users and the exponential proliferation of user-generated data is the massive adoption of mobile devices. The inherent simplicity of sensor-
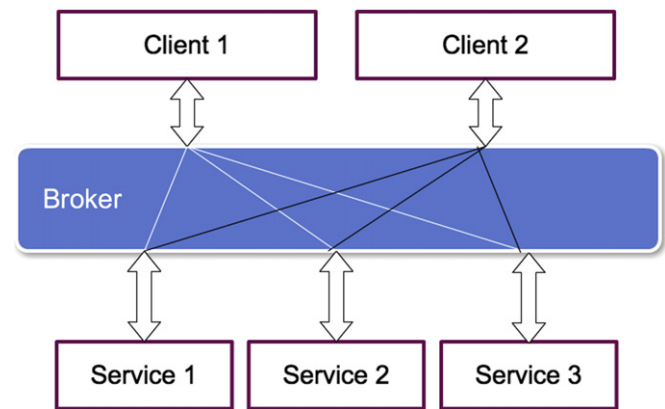

Fig. 1. The conceptual brokering approach.

enabled mobile devices to be "data-capturing tools" has turned users into real-time and ubiquitous creators of any kind of media content (e.g. text, videos, audio files, pictures). As such user-generated content refers to phenomena that are bound to a location, such a georeferenced content is acquiring a fundamental role in a wide range of applications. For example simple georeferenced messages from social networks such as Twitter[4] may play a major role in response actions to emergencies (Roche, Propeck-Zimmermann, Mericskay, 2012; Schade et al., 2012). Even when different types of data are combined, for example point of interests and pictures, new scenarios come out such as volunteered–based map creation (Neis, Zielstra & Zipf, 2012), and the collection of in situ biodiversity data (Newell, Pembroke & Boyd, 2012) and forestry data (Aragó, Tamayo, Viciano, Huerta & Díaz, 2011). This phenomenon has been tagged with different labels such as Neogeography (Turner, 2006), Cybercartography (Tulloch, 2007), and Volunteered Geographic Information (VGI) (Goodchild, 2007), being VGI the most widely accepted term in the geospatial community.

Although VGI data still represents a small percentage, its growth is being greatly accelerated largely by the expansion of sensor-enabled devices. It is thus reasonable to foresee that huge amounts of georeferenced data will be available in an immediate future. This poses the question whether VGI may be seen as an alternative source of information to complement authoritative data from GIIs (Craglia et al., 2008), so as to improve traditional geospatial analysis and decision support tasks (Flanagin & Metzger, 2008; Pultar, Raubal, Cova & Goodchild, 2009) since it provides data at high spatio temporal scales (Zook, Graham, Shelton, & Gorman, 2010). Experts and decision makers may even benefit from VGI data in multiple analysis situations such as epidemiology (Aanensen, Huntley, Feil, al-Own & Spratt, 2009) and geo-demographics (Singleton & Longley, 2009). In these cases the bottom-up approach exemplified by crowd-sourcing services may complement the top-down methodologies that articulate GIIs. In this paper we explore a mechanism to enhance the search and retrieval of data from crowd-sourcing services to be integrated among others in the GII context.

Geospatial information infrastructures: top-down approach

GIIs or SDIs, offer the capability of discovering, accessing and sharing a diversity of geospatial resources via standards web services among a wide range of actors at different scales (Masser, Rajabifard & Williamson, 2007; Rajabifard, Feeney & Williamson, 2002). They comprise a set of policies and standard activities promoting the creation of geospatial information services to assist
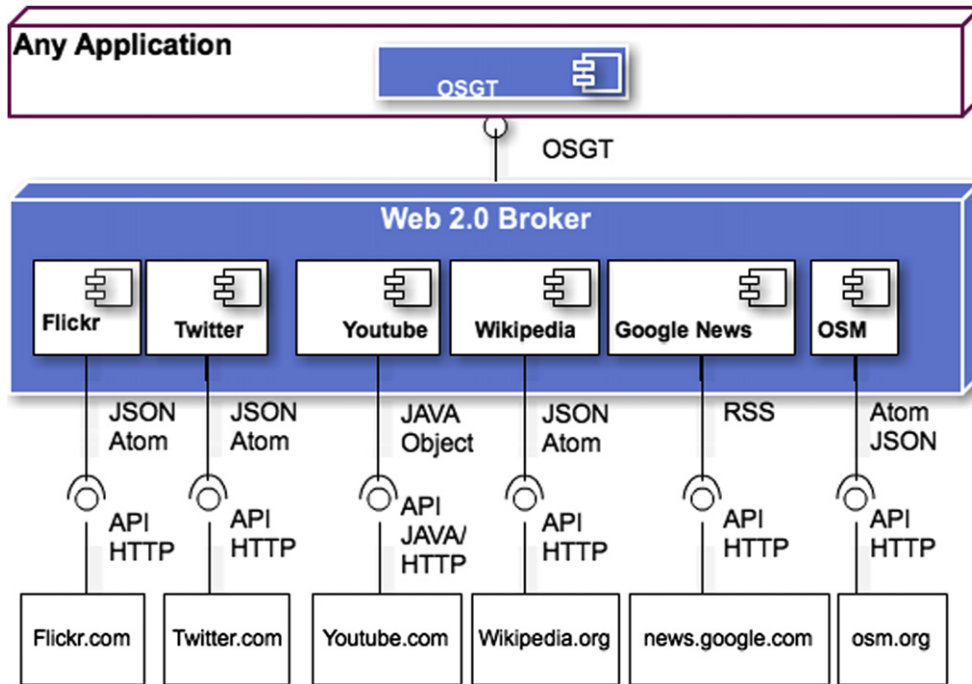
---

[4] https://twitter.com/.

Fig. 2. W2B service components diagram. Proposed uniform interface for clients.

diverse user communities in collecting, sharing and exploiting geospatial resources (Bishop et al., 2000; Davis, Fonseca & Camara, 2009; Masser, 2005; Nebert 2004). In this context, GIIs play a key role as facilitators and coordinators of geospatial data at regional, national and global scales (Dessers et al., 2012). International initiatives such as the Global Earth Observation System of Systems (GEOSS; Pearlman & Shibasaki, 2008) or the European INSPIRE directive (INSPIRE, 2007) describes the overall architecture and best practices for designing and implementing GIIs where geospatial resources are managed by means of regulated, standardized services.

Nevertheless, the inherent complexity of some service specifications and data standards used commonly in GIIs (Tamayo, Granell &

Huerta, 2012) and the complex mechanisms for spatial content publication (Díaz et al., 2011) may lead to management issues as GIIs grow in terms of data and services (Béjar, Latre, Nogueras-Iso, Muro-Medrano, & Zarazaga-Soria, 2009). Despite these issues crowdsourcing services are not displacing GIIs from the geospatial landscape. Indeed, hybrid approaches for the integration of top-down and bottom-up methodologies may co-exist promoting the integration of Web 2.0 resources within GIIs. Such new approaches are founded on what has been coined as the reconceptualization of the SDI user role (Budhathoki, Bertram, & Nedovic-Budic, 2008; Goodchild and Glennon, 2010; Omran & van Etten, 2007), where a more active user base will influence the new generation of GIIs. In this paper we assess these hybrid approaches in a set of real-world scenarios.
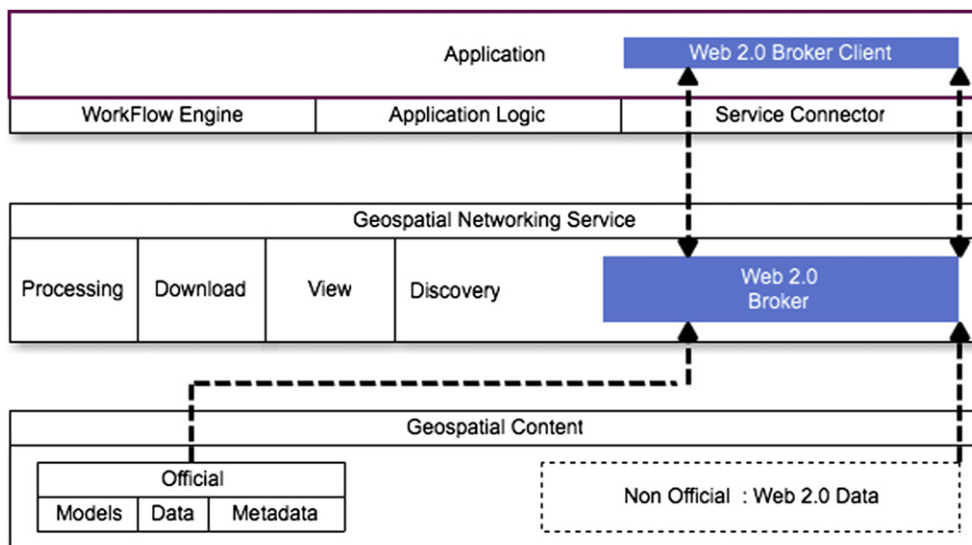

Fig. 3. Interoperability between W2B service and other infrastructures and communities (SDI, GII).
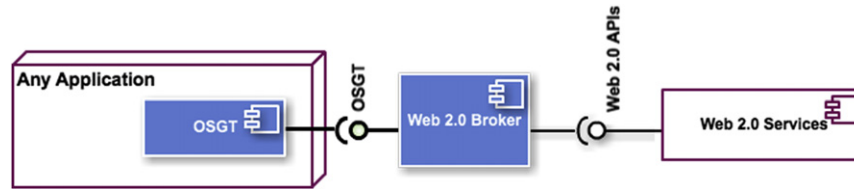
Fig. 4. Reference architecture.

A technological look inside W2B service

As seen above, GIIs and crowd-sourcing services follow different approaches but essentially pursue same objectives: facilitators of data. In order to explore the first research question mentioned earlier − the interplay of non-authoritative and authoritative data − we have developed the W2B service that enables to seamlessly search and retrieve datasets from both worlds. This section thus gives a brief summary of the main technological drivers that define the W2B service so that it can be an integral part of the scenarios described in Section 4. Díaz et al. (2012) complement this section from a technical perspective providing a deep discussion on the conceptual architecture and implementation details.

Brokering approach

Web 2.0 services provide public APIs to client applications to interact with them via specific encodings and methods. Although a few data formats are gaining popularity and wide acceptance (e.g. JSON), there is no standard consensus on the description of service interfaces and data encodings used. This situation clearly raises a technical barrier for discovering user generated content from multiple crowdsourcing platforms in a uniform manner, because service clients need to understand specific data formats and implement diverse APIs to be able to access and retrieve data from various Web 2.0 services. This issue has been long studied from the perspective of software engineering. Buschmann, Meunier, Rohnert, Sommerland and Stal (1996) proposed the architectural pattern Broker that may be easily understood as a middleware component connecting heterogeneous components and systems to improve scalability and interoperability.

The main goal of the W2B service is to facilitate the access and retrieval of data from diverse Web 2.0 services. Here, the broker pattern meets our requirements. Indeed, the W2B service follows a brokering approach (Nativi, Craglia, Vaccari & Santoro, 2011), which extends the basic broker pattern by transferring the required business logic such as coordinating requests and responses and handling with specific standards for data encoding to a brokering middleware. As Fig. 1 illustrated, a broker allows clients to connect with multiple services that follow their own syntax for requesting and delivering data. This is the case of the W20 service that enables smooth data discovery and access through a great variety of different APIs published by Web 2.0 services and platforms. This way the W2B service exposes one single interface (see next section) and makes it possible to transform a single user request into multiple, API-based requests addressed to varied Web 2.0 services.

Spatio-temporal search and retrieval

The OpenSearch (OS) specification[5] defines a set of search parameters and a communication protocol. For a client application, an OS-enabled service exposes a search interface based on HTTP GET requests with specific query parameters according to the specification. These parameters such as query term must be described using an OS Description Document file which is available to enable clients to build proper queries.

The OS specification can be extended for particular purposes and functionalities.[6] One of these extensions is the OpenSearch Geospatial and Time specification (OSGT) (Gonçalves, 2010), recently adopted by the Open Geospatial Consortium (OGC) as a standard specification. In short, OSGT extends the base OS specification with parameters to support spatio-temporal discovery. User can specify a place name, area, point and radius, and a time period as part of a spatio-temporal query (Díaz et al., 2012). The W2B service has adopted the OSGT interface to provide a single interface from the client perspective (See Fig. 2, top). This way, regardless of the nature of the backend Web 2.0 services supported by the W2B service, potential clients only need to understand a single search interface.

Client applications build valid search queries by consulting the OS Description Document file. Once received a query, the W2B service transforms it into specific queries (according to each particular API's encodings and methods) by means of the specialized components (e.g. Flickr, YouTube, etc. in Fig. 2), and broadcasts them to the requested Web 2.0 services. In the response phase, the results encoding format must be known to client applications. Fig. 2 (bottom) shows the array of Web 2.0 services mediated by the W2B and their interfaces and encoding formats. Although search responses are often encoded in lightweight data formats such as RSS,[7] Atom[8] and JSON[9] the OS specification in reality does not impose any particular response format. The W2B service transforms encoding formats used by the target Web 2.0 services into a middleware encoding format. From our experiments, we suggest that candidate encoding formats should at least support two basic constraints: they must support links and spatio-temporal features. Links are meant to provide direct access to actual data that client applications are seeking for. The latter condition refers to natively support and encode spatial and temporal information. To this respect the W2B service supports Atom and KML among others (Díaz et al., 2012).

Interoperability and compatibility

As earlier commented one of the key mechanisms to enable geospatial data sharing is the establishment of GIIs. Current trends in multilevel GII development enable end-users to share spatial data in decentralized structures where a top-down structure aims to achieve interoperability while the bottom-up structure aims to integrate user knowledge. When we talk about interoperability between two distinct services, we mean that they can successfully interact with each other. Take for example Twitter and Open Street Map services that do not interoperate a priori yet need to co-

---

[5] http://www.opensearch.org.

[6] http://www.opensearch.org/Specifications/OpenSearch/Extensions.

[7] http://www.rssboard.org/rss-specification/.

[8] http://tools.ietf.org/html/rfc4287.
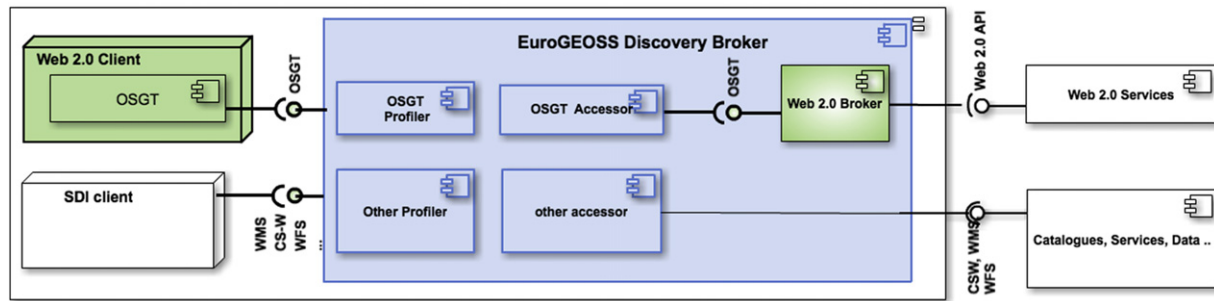
[9] http://www.json.org/.

Fig. 5. EuroGEOSS discovery broker integration architecture.

operate jointly in our scenario, permitting then users to search for geospatial data seamlessly from multiple Web 2.0 and crowd-sourcing services.

Fig. 3 shows a 3-tier architecture for GIIs where we can see how the W2B service and client applications are deployed. Any client application (as the Web 2.0 Broker client in the figure) that requires interacting with the W2B service will simply access to this service (see discontinuous line) as if it were any other standards-based geospatial services available in GIIs (see middleware layer in the figure), since the W2B service is regarded as a type of discovery service (Fig. 3). From the middleware layer the W2B service accesses the content layer for retrieving both official and crowd-sourced data.

Two determining factors are required to support the needed interoperability in the scenario described in the early paragraph. First, the W2B service increases the interoperability between geo-spatial users and the Web 2.0 services, since the brokering approach is flexible enough to make the access transparent to heterogeneous sources, capabilities and interfaces offered by the backend Web 2.0 services. Second, the W2B service promotes compatibility with existing standards and data formats already in use in current GIIs. In this sense, the W20 is able to integrate user generated data and official datasets in middleware formats (Section 3.2). This way search results through the W2B service can be easily consumed by traditional GII clients (e.g. map viewers, desktop applications). Not only is the ability to access and retrieve data from Web 2.0 services relevant in our case, but also important is to promote collaboration and integration with existing geospatial components and infra-structures to avoid isolated silos, so that the geospatial community may benefit from new synergies and integrated developments (Kates et al., 2001; Reid et al., 2010). That is, the W2B service acts as data discovery services available commonly in GIIs (Fig. 3). For doing so, the W2B service has been designed to be an easy-to-deploy service, which can be reusable in different environments and maximize thus its use in varied scenarios.

Flexible deployments

In a system, interoperability and reusability trend to correlate positively. This means that improvements in interoperability follow greater levels of reusability. However, some additional efforts have to be made to enhance reusability. Accordingly, the W2B service ach-ieves interoperability by adopting the brokering approach (Section 3.1). As regards reusability, the W2B service is designed as a standard web service, which makes it suitable to be smoothly deployed in service-oriented architectures such as GIIs (see Fig. 3). Thanks to the simplicity of the OSGT interface, different kinds of client applications can easily integrate a tiny OSGT-compliant component (see Section 3.2), which is able to access and interact with the W2B service. Fig. 4 illustrates the reference architecture that can be used to instantiate specific architectures when deploying the W2B in GIIs.

In the next section we describe extensions of the reference architecture which have been implemented, deployed and tested in varied scenarios. These extensions range from aggregating the W2B service with other brokers to the integration of tiny OSGT-compliant components into existing web applications. These scenarios demonstrate the key technological drivers and illustrate the value of the W2B service.

Scenarios with W2B service

The W2B service has been designed as an interoperable service implementing a standard interface to be re-used in different scenarios. To demonstrate the potential value of crowd-sourced information beyond the mere visualization on maps, in this section we describe how the W2B service has been deployed in a wide variety of scenarios. Section 5, however, discusses the applicability of crowd-sourced information, i.e. the output of our W2B service, in the context of these application scenarios.

Support for GEOSS components

One success story of the brokering approach mentioned earlier (Section 3.1) is the EuroGEOSS Discovery Broker (Nativi et al., 2011), an outcome of the European project EuroGEOSS[10] and an overall contribution to GEOSS (Global Earth Observation System of Systems) and its Common GEOSS Infrastructure (CGI) (Nativi, Craglia & Pearlman, 2012). The EuroGEOSS Discovery Broker provides a unique access point to services and data sources from three GEOSS societal benefit areas: biodiversity, forestry, and drought.

The W2B is also a result from the EuroGEOSS project and follows the same brokering approach. However, rather than focussing on data sources provided and published in GIIs and other authoritative sources, as the EuroGEOSS Discovery Broker does, the W2B service is designed to interact with crowd-sourcing platforms in which the spatio-temporal dimension is or may be relevant (Díaz et al., 2012). In this way the W2B service augments the functionality of the EuroGEOSS Discovery Broker with concepts emerging in the Web 2.0 communities with respect to user interactions and Web 2.0 resource discovery.

The scenario described here highlights the advantage of aggre-gating various brokers to work cooperatively to provide an integrated functionality that augments that of any single broker (Vaccari, Craglia, Fugazza, Nativi, & Santoro, 2012). Fig. 5 shows an extension to the reference architecture (Fig. 4), that integrates the W2B service into the EuroGEOSS Discovery Broker. The OSGT Accessor component in Fig. 5 is functionally identical to the OSGT-compliant component in the left part of Fig. 4. Both act as clients to interact with the W2B
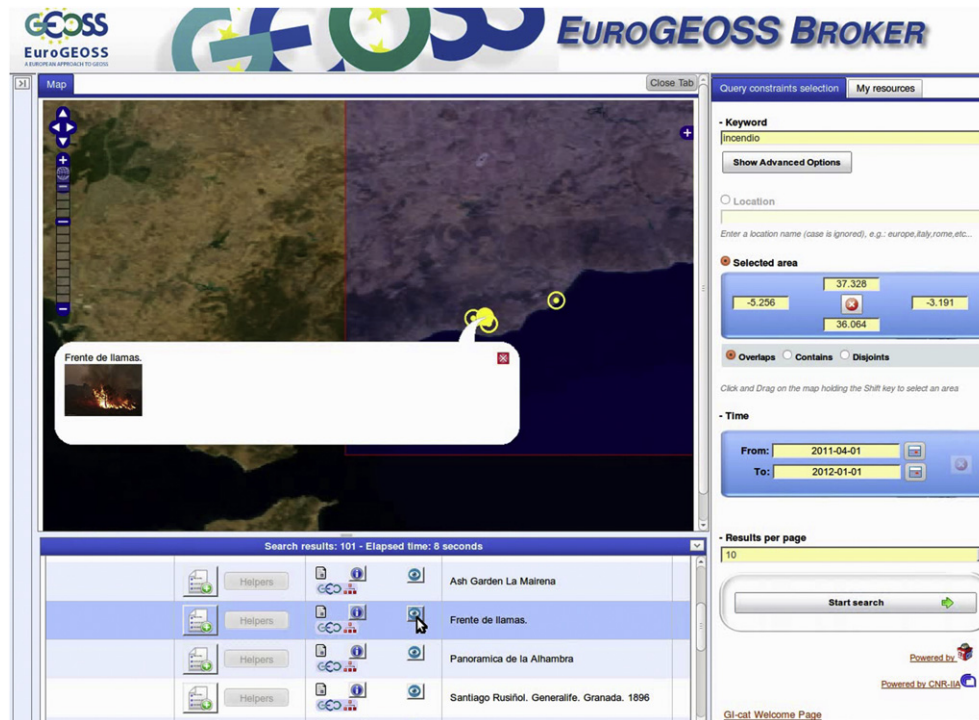

Fig. 6. Screenshot of W2B integration into EuroGEOSS brokering platform.

service. In the case of the aggregated EuroGEOSS Discovery Broker in Fig. 5, the Discovery Broker is aimed to search for authoritative, cross-thematic data (bottom part of Fig. 5), where the aim of the W2B service is to retrieve multidisciplinary user-generated data. Further details on implementation and service interfaces required to aggregate both brokers are described in Díaz et al. (2012).

Fig. 6 depicts a screenshot of a web client interacting with the EuroGEOSS Discovery Broker to retrieve Web 2.0 data from sources such as Flickr. In this scenario, the W2B emphasizes some of the technology drivers described in Section 3, especially the brokering approach (Section 3.1). The W2B service may be aggregated with other brokers to build on-demand brokering solutions while at the same time maintaining a simple service interface from the client perspective. That is, business logic is transferred to the broker components rather than to the client application.

Support for environmental monitoring applications

As we have seen in the previous example, following the reference architecture in Fig. 4, the W2B service can be aggregated into other brokers. Next we show how the W2B service can work in isolation to be accessed by operational environmental monitoring applications, in this case the European Forest Fire Information System (EFFIS) (McInerney et al., 2012) and the Habitat Assessment and Ecological Forecasting system (Dubois et al., 2011), both developed and hosted by the Institute of Environment and Sustainability (IES) at the Joint Research Centre (JRC). In both cases, a small OGCT-client component has been developed and integrated in these client applications.

The EFFIS[11] provides users with data and tools to monitor forest fires in Europe on a daily basis. Among others, EFFIS provides specific models and services to help experts monitor the spatial distribution of forest fires in order to evaluate fire damage as well as environmental (forestry resource, biodiversity loss and drought influences) and social impacts (McInerney et al., 2012). During the monitoring of a fire event, crowd-sourced information retrieved from Web 2.0 services may offer a more timely yet complementary view to the authoritative data (Curtis and Mills, 2012; Díaz et al., 2012). In this direction we have shown the practical use of the W2B service by using it from the EFFIS front-end application to allow EFFIS users to contrast or support official information by retrieving related information from crowd-sourcing platforms. The resulting architecture in Fig. 7 extends the reference architecture shown in Fig. 4 by including an OSGT-complaint client that allows keyword and area search from diverse crowdsourcing services.

A second scenario is concerned with environmental monitoring applications for assessment and forecasting of biodiversity habitats at worldwide scale: the e-Habitat[12] web application. This application is designed for locating and assessing ecosystems with similar environmental properties (Dubois et al., 2011). Like in the case of EFFIS, the e-Habitat technicians implemented the same solution illustrated in Fig. 7.

Fig. 8 shows the resulting e-Habitat client application that is able to interact with the W2B service to retrieve information about an avian species "Circus Maurus" in South Africa. In this scenario, the implementation of the OSGT-compliant client has a more or less fixed set of parameters such as the search term fixed to scientific species name. For instance, users can select the "Bird" region on the right hand side and search for a species by its scientific name (e.g. "Circus Maurus") from a dropdown list. The user is shown yellow squares indicating the regional results.

Support for ad-hoc applications

In the context of the EuroGEOSS project we developed an ad-hoc application called EuroGEOSS Web 2.0 client[13] to interact with the W2B service. Thanks to the simplicity of the access interface, the

---

[11] http://effis.jrc.ec.europa.eu.

[12] http://ehabitat.jrc.ec.europa.eu/content/web-services.
[13] It is currently available to the public at http://elcano.dlsi.uji.es/GF/.

Fig. 7. EFFIS Web 2.0 Broker integration architecture.

creation of ad-hoc client applications that exploit the potential of the W2B service may be a straightforward, alternative solution when the integration of such a client functionality is not possible in existing applications due for instance to software policy restrictions.

In the following scenario the EuroGEOSS Web 2.0 client is intended to monitor the status of a detected fire. The client provides a web map to visualize data coming from various sources. Additionally, it shows the burned areas retrieved from the EFFIS data services, which allow users to overlap and study official and crowd-sourced data reported in most cases by citizens affected by the disaster. This web client offers simple and advanced interfaces (customizable by the user) to specify search criteria and build a query. Users can add spatio-temporal criteria by selecting the area of interest in the form of a rectangle on the map or by providing point and radius information. Additionally, users looking for results within a certain time period may specify begin and end dates.

Next two examples show forest fires in Spain in different regions and periods of time. The first forest fire event occurred in Mijas (Málaga) in 2011. For this example, we performed queries restricted to the area of Mijas and time of the event. Fig. 9 shows georeferenced pictures taken by citizens who were near the fire event when it happened. In the figure one observes the added value of mixing authoritative and non-authoritative data. For instance, the convex hull of the pictures provided by the citizens closely approximates the official burned area provided by EFFIS. Here in the absence of official burn area the citizen data gives a worthy estimate. In addition, the crowdsourced data is available in near-real time, meaning that in most cases this information will be useful for time-critical decisions and potentially saving more human lives when responders act on this data immediately rather than waiting for official sources.

The second example refers to recent severe forest fires (as of June 2012) in the Valencia region (east coast of Spain), where almost 50,000 ha were burned. In Fig. 10 we see the burned areas reported by EFFIS and user-generated data published during the fire period. We can see fire-fighting labours being reported, pictures and videos reporting the status of the towns nearby completely covered with ash, and complaints and proposals by citizens living in the areas to avoid similar disasters in the future.

Event monitoring during the forest fire in a real world scenario demonstrates the use of the W2B service, where users are able to use both volunteered contributions through Web 2.0 services in addition to official environmental-related data. As we discuss in Section 5, these examples show how W2B make it easy to retrieve citizen contributions, which is not a substitute for scientific data but a complementary source of information that may assist decision makers in multidisciplinary monitoring scenarios such as in the case of natural disasters and hazards such as wildfires or hurricanes.

Apart from web-based applications we have also developed a mobile application to show the wide range of clients that can interact with the W20 service. Fig. 11 illustrates a screenshot of an iOS[14] application that connects to the W2B and shows user generated data regarding requested terms. This application has been designed to contain a base map (this case showing the topographic map served by ESRI[15]) and a module to translate user queries to OSGT and connect to the W2B, as the web-based client applications mentioned earlier do. Fig. 11, shows the results referring to recent forest fires (as of September 2012) in the municipality of Chulilla[16] again in the Valencia region; we can appreciate user-generated data reporting the event in the form of text messages published in Twitter, photographs retrieved from Flicker and video capturing the situation available in YouTube.

Support for VGI applications

In the context of environmental monitoring there are also volunteer–based applications. This is the case of Geo-wiki, a global network of volunteers aiming to improve the quality of different thematic datasets, with strong focus on land use (Fritz et al., 2009; Fritz et al., 2012). The Geo-wiki application shows authoritative data layers such as *CORINE Land cover* to be validated by users who have a great knowledge of their local surroundings and can validate such datasets. In this way the W2B service can search for related information published in crowdsourcing platforms to support the user who validates or ground-truths a specific area. Following the same principles the W2B service integration is implemented with the same architecture shown in Fig. 7. The Geo-wiki technicians integrated a simplistic OSGT client to allow users to visualize terms retrieved by a default query.

Fig. 12 shows results obtained from the W2B service integrated into the geo-wiki project. In this example the volunteered geographic information extracted from Web 2.0 services is related to Flickr content tagged as "nature" in Thailand (Schill et al., 2012). Note that in all these examples any publically accessible base map may be used; W2B is agnostic to that layer.

Support for Digital Earth applications

The vision of Digital Earth is widely regarded as a key milestone for the geospatial community (Craglia et al., 2008; Craglia et al., 2012; Goodchild et al., 2012). The digital replica or model of the entire planet as a virtual globe is gaining acceptance mostly within the scientific community. Recent virtual globes developments such as Google Earth, NASA World Wind, and ESRI ArcGIS Explorer have leveraged experts in studying environmental, climate and geological issues at global scale (Bayley, 2011). Nevertheless, although recent progress has been made towards virtual globe technologies to create Digital Earth applications to analyse and visualize authoritative data in a more intuitive manner than using traditional (2D) web map viewers and GIS desktops applications (Goodchild et al., 2012), the incorporation of crowd-sourcing data in such Digital Earth applications is still an open research issue (Craglia et al., 2012).

---

[14] Apple's operation system for mobile devices such as iPhone, http://www.apple.com/ios/.

[15] http://www.esri.com/.
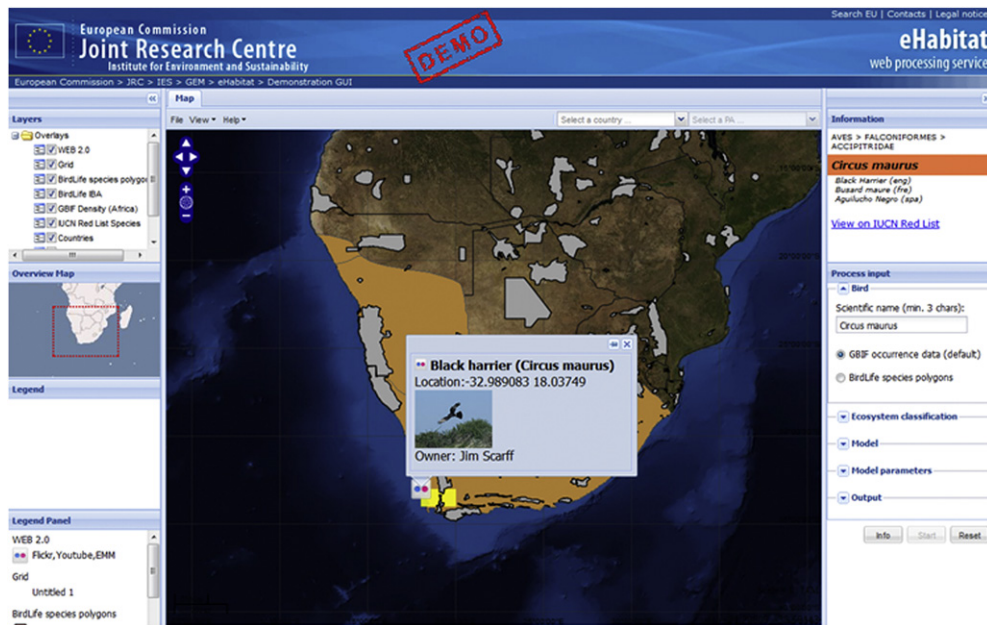[16] http://es.wikipedia.org/wiki/Chulilla.

Fig. 8. Screenshot of the third-party application e-Habitat.

In this context Beltran et al. (2012) presented a Virtual Globe application based on the NASA World Wind tool that exploits the capabilities of the W2B. The authors developed an OSGT client within the Virtual Globe so that users could perform spatial-temporal queries over Web 2.0 services. Georeferenced search results (e.g. video, pictures, tweets) were displayed over a virtual globe in function of the media type.

Again, the great range of client applications, from web mashup applications to specific web mapping clients to virtual global-based applications, demonstrate the flexibility and easy deployment of the W2B service to meet diverse requirements and needs.

Observations on the usage of crowd-sourced information

Users create and share massive amounts of data which are often timely and freely available in crowdsourced platforms, and thus it is becoming a major source of information. However, due to its heterogeneity, searching and integrating georeferenced crowd-sourced information with authoritative geospatial data still presents many challenges. In this section we discuss some advantages and disadvantages of the application of the W2B service to the context of the real-world scenario commented earlier, as well as open issues such as the interplay of crowdsourcing platforms and GIIs.

In previous sections we indicated clear differences between crowd-sourced and authoritative geospatial data collected by national mapping agencies and institutions. With respect to the nature of data, crowd-sourced data is often timely, of wide coverage, and comes with a variety of data types. In addition it is free and is subject to low-cost production means. On the other side, a clear limitation is its quality compared with authoritative data. Some authors have recently questioned the problem of assuring quality of
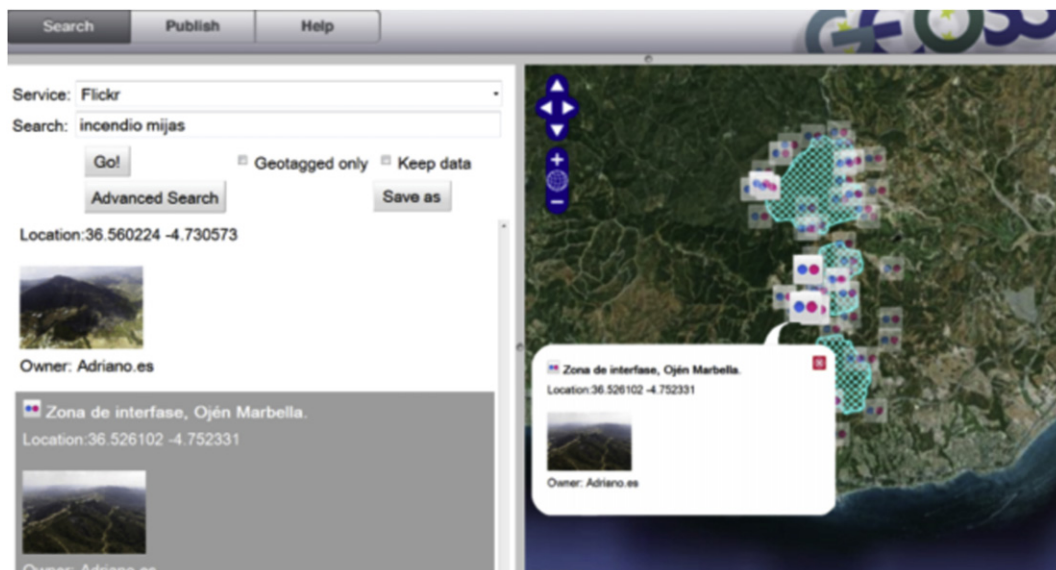

Fig. 9. Screenshot of W2B results retrieved both from Flickr and burned areas from EFFIS when doing a search for "incendio Mijas".
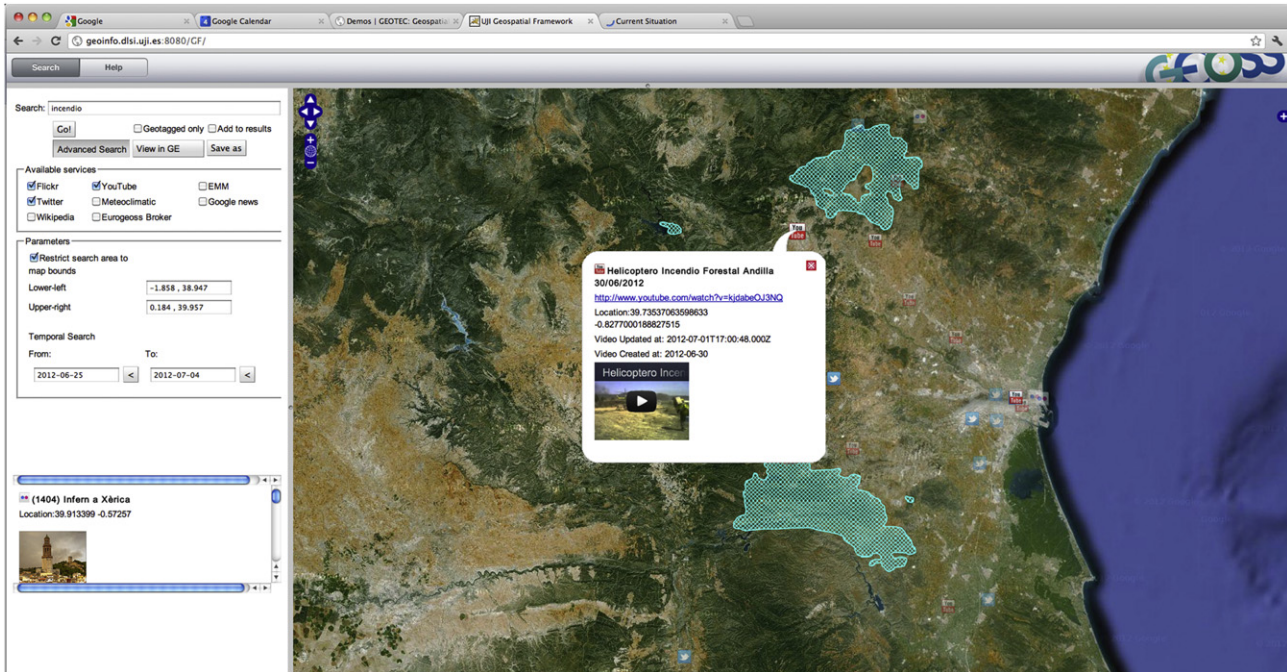
Fig. 10. Screenshot of W2B service results of a search for "incendio" (fire) in the Valencia region and burned areas from EFFIS provided via the EuroGEOSS discovery broker.

crowdsourced data during its acquisition (Goodchild & Li, 2012), since much research is needed to analyse and filter such data to extract more accurate information out of the massive repository provided by citizens. While authoritative data is collected and documented through well-established quality mechanisms by national mapping agencies (Goodchild and Glennon, 2010).

Apart from differences in terms of the nature of data, both kinds of data also play distinct but complementary roles when they are jointly used in the same scenarios. Given that crowd-sourced data is not tied to scientific procedures in its collection and management, it may be useful in the early, exploratory, and hypothesis-generating stages of scientific projects, while not so much as data required for scientific research activities such as decision-making processes, policies analysis, assessment modelling and simulation (Goodchild & Li, 2012; Díaz et al., 2012). This introduces the question whether crowd-sourced data provides real added value in scientific scenarios. This is difficult to assess as the benefits of data depend strongly on the scenario in which it is used. For instance, data collected by amateur biologists is useful when is combined with quality mechanisms in a centralized manner (Newell et al., 2012). Another example is the use of crowd-sourced data to monitor and participate in environment conservation (Bernard, Barbosa & Carvalho, 2011). However, the issue of filtering useful data from "noise" is still a real impediment as Schade et al. (2012) concluded after analysing millions of offline tweets (locally stored in a database) to extract those that referred to wildfires around Europe. Regarding quality, Ostermann and Spisanti (2011) performed a similar study using data mining techniques over millions of tweets to monitor forest fires and crisis management. They concluded that the added value and quality of VGI (georeferenced tweets) is difficult to quantify because this data is mostly subjective to the target use case, the end user (consumer), and the VGI creator (producer). Returning to our scenario in Section 4.2 about the forest fire in Mijas, a fire brigade was accompanied by a professional reporter who was in charge of capturing pictures in situ. In that case, VGI data was of high-quality and valuable for monitoring and decision-making processes because some requirements were met at the same time: An expert in producing georeferenced data (professional photographer), being at the right place and time to take pictures, and being

Fig. 11. Screenshot of user generated data reporting a fire event in the municipality of Chulilla (Valencia, Spain).
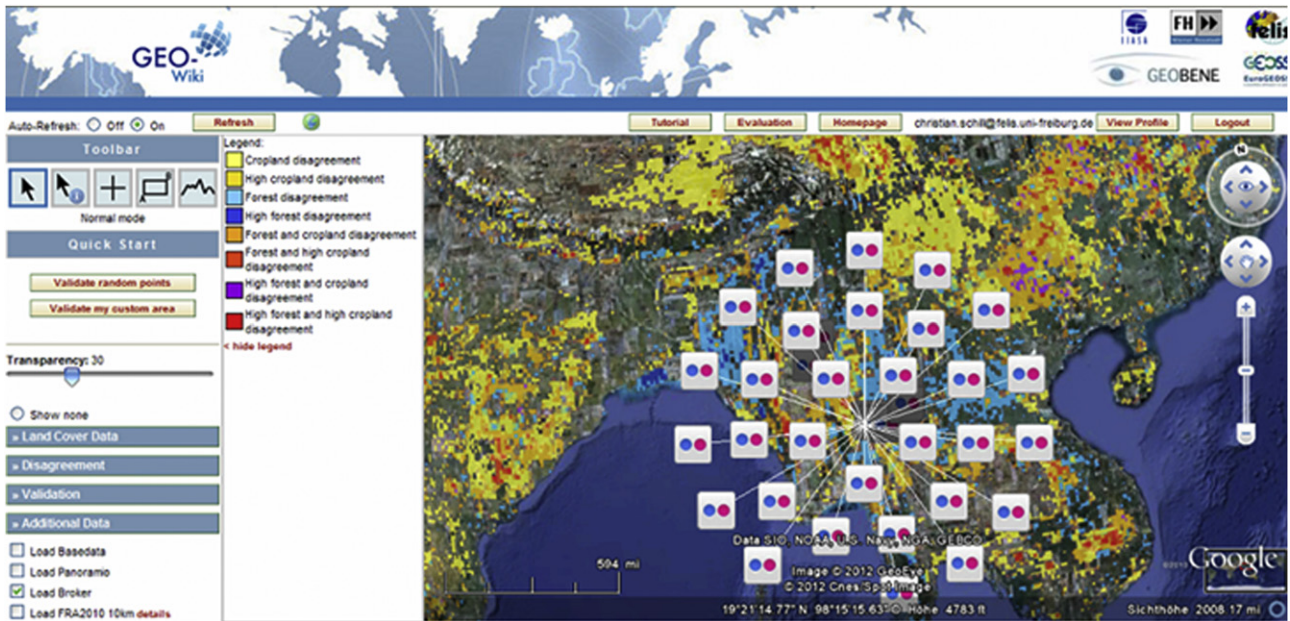
Fig. 12. Screenshot of embed results retrieved from the W2B into the geo-wiki platform.

advised by experts in the domain of wildfires (fire brigades). Unfortunately, these requirements are not likely met during the typical process of VGI acquisition and collection.

In the scenarios described earlier, the W2B attempts to maximize the use of crowd-sourced data in exploratory research and early stage of decision making support complementing existing scientific data. The aim of W2B is not to characterize or assure the quality of VGI, but discover and prepare it to be combined with other data for early decision making. Given this assumption, the W2B can be used to retrieve data to complement crisis management model inputs and to refine their output results. The scenarios presented in this paper demonstrated the interdisciplinary capacity and flexible deployment of the W2B service. The integrated modelling scenarios such as e-Habitat and EFFIS illustrate how general environmental monitoring can leverage the potential of massive amounts of this multidisciplinary data. In our use case we consumed raw data after having performed a preliminary visual analysis. This provides a first glance at how this data may add value to a scientific workflow. However, as commented earlier, making decisions from the validation of global models with this knowledge has yet to be fully exploited. For example, critical aspects in impact assessment and policy analyses are the verification of the source of and modifications performed in input data (e.g. data provenance) as well as to clarify the legal and ethical implications of its usage (e.g. licensing), clearly remain open issues in regard to crowd-sourced information. On the other hand, the Geo-wiki scenario in Section 4.2 allows us to experiment with the role of the W2B to help users in assuring the quality of coverage data by providing additional data for particular places. It offers a way to monitor editions and detect inconsistencies in a collaborative manner, that is, crowd-sourced information is regarded as a secondary source for addressing misunderstandings or ambiguities not resolved by means of authoritative data.

In addition to differences in the nature, quality, and target use, scale also plays a role in the value of crowd-sourced information. As commented earlier, crowd-sourced data is often of wide coverage which means that it may scale very well from local to regional and even national settings. In scenarios such as local and populated areas, crowd-sourced data may be preferred over official data. In the case of flash flooding, for example, in which heavy rains are commonly concentrated in small areas in a short time period (as is common on

the east coast of Spain), crowd-sourced information collected by citizens becomes practically the only source timely available. This happens because this data is captured by in situ residents in the absence of official meteorological stations. A brigade of amateur meteorologists periodically collects and shares meteorological–related data through the Meteoclimatic[17] web site, a Web 2.0 service accessible by the W2B service (Díaz et al., 2012). Thereby, in such cases where scale may be a limiting factor in the availability of authoritative data, crowd-sourced information may become a primary source of information for a preliminary analysis processes.

Regarding the open issues with respect to the use of crowd-source information, the current version of the W2B service may be improved to address some of them. In this direction, there is a need for more sophisticated analysis to filter the large crowd-sourced repositories, to extract more accurate information and avoid the inherent noise of consuming raw data. Modelling this data in order to detect specific patterns and changes can generate more relevant and accurate information. Data mining techniques have been proven effective in these contexts (Ostermann and Spisanti, 2011; Schade et al., 2012), which may lead to real benefits when used in specific use cases and scenarios as demonstrated in this paper.

Conclusion

Addressing the questions posed in the introduction section, we have presented the Web 2.0 Broker (W2B) as a new discovery and retrieval service to provide a standards-based, unique entry point to query multiple Web 2.0 services and crowdsourcing platforms to retrieve user-generated content (citizen-based information) to be prepared and integrated in GII contexts. This service interprets queries using the OpenSearch Geo-Time specification and smoothly propagates them to a set of Web 2.0 services.

Regarding how to coalesce authoritative and non-authoritative information in scenarios beyond mere visualization, the W2B service aims to improve the means by which VGI is integrated into GIIs thereby leveraging more of its potential. However, further

---

[17] http://www.meteoclimatic.com/.

research is necessary in this field. We will continue extending this solution by increasing the number of Web 2.0 resources to be aggregated as they become available. Additionally, we will continue to analyse the massive data flow in order to extract observations relevant to specific use cases. The next steps are to define a data model to describe environmental observations and alarms thereby adding a new information source for emergency response scenarios and exploiting the intrinsic multidisciplinary character of the W2B component which favours a wide range of use cases.

Acknowledgements

References

Aanensen, D. M., Huntley, D. M., Feil, E. J., al-Own, F., & Spratt, B. G. (2009). Epi-Collect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS ONE, 4*(9), e6968.

Aragó, P., Tamayo, A., Viciano, P., Huerta, J., & Díaz, L. (2011). Forest Fire Survey and Processing Tool for Android-Based Mobile Devices. In: *INSPIRE Conference 2011*. Edinburg.

Bailey, J. E. (2011). The role of virtual globes in geosciences. *Computers & Geosciences, 37*(1), 1–2.

Béjar, R., Latre, M.Á., Nogueras-Iso, J., Muro-Medrano, P. R., & Zarazaga-Soria, F. J. (2009). An architectural style for spatial data infrastructures. *International Journal of Geographical Information Science, 23*(3), 271–294.

Beltrán, A., Abargues, C., Granell, C., Nuñez, M., Díaz, L., & Huerta, J. (2012). A virtual globe tool for searching and visualizing geo-referenced media resources in social networks. *Multimedia Tools & Applications*, . http://dx.doi.org/10.1007/s11042-012-1025-0.

Bernard, E., Barbosa, L., & Carvalho, R. (2011). Participatory GIS in a sustainable use reserve in Brazilian Amazonia: implications for management and conservation. *Applied Geography, 31*(2), 564–572.

Bishop, I. D., Escobar, F. J., Karuppannan, S., Suwarnarat, K., Williamson, I. P., Yates, P. M., et al. (2000). Spatial data infrastructures for cities in developing countries: lessons from the Bangkok experience. *Cities, 17*(2), 85–96.

Budhathoki, N. R., Bertram, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *Geojournal, 72*(3–4), 149–160.

Buschmann, F., Meunier, R., Rohnert, H., Sommerland, P., & Stal, M. (1996). *Pattern-oriented software architecture volume 1: A system of patterns*. John Wiley & Sons Ltd.

Coleman, D. J., Georgiadou, P. Y., & Labonte, J. (2009). Volunteered geographic information: the nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research, 4*, 332–358.

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., et al. (2012). Digital Earth 2020: towards the vision for the next decade. *International Journal of Digital Earth, 5*(1), 4–21.

Craglia, M., Goodchild, M. F., Annoni, A., Câmara, G., Gould, M., Kuhn, W., et al. (2008). Next-generation Digital Earth: a position paper from the vespucci initiative for the advancement of geographic information science. *International Journal of Spatial Data Infrastructures Research, 3*, 146–167.

Curtis, A., & Mills, J. W. (2012). Spatial video data collection in a post-disaster landscape: the Tuscaloosa Tornado of April 27th 2011. *Applied Geography, 32*(2), 393–400.

Davis, C. A., Jr., Fonseca, F. T., & Camara, G. (2009). Beyond SDI: integrating science and communities to create environmental policies for the sustainability of the Amazon. *International Journal of Spatial Data Infrastructures Research, 4*, 156–174.

Dessers, E., Crompvoets, J., Janssen, K., Vancauwenberghe, G., Vandenbroucke, D., Vanhaverbeke, L., et al. (2012). Multidisciplinary research framework for analysing SDI in the context of business processes. *International Journal of Spatial Data Infrastructures Research, 7*, 125–150.

Díaz, L., Granell, C.,M., Gould, M., & Huerta, J. (2011). Managing user generated information in geospatial cyberinfrastructures. *Future Generation Computer Systems, 27*(3), 304–314.

Díaz, L., Nuñez, M., González, D., Gil, J., Aragó, P., Pultar, E., et al. (2012). Interoperable search mechanism for Web 2.0 resources. *International Journal of Spatial Data Infrastructure Research, 7*, 2012.

Dubois, G., Skoien, J., Mendes de Jesus, J., Peedell, S., Hartley, A., Nativi, S., et al. (2011). eHabitat: A contribution to the model web for habitat assessments and ecological forecasting. In *Proceedings of the 34th international symposium on remote sensing of environment* (pp. 1–4).

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *Geojournal, 72*, 137–148.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., et al. (2009). Geo-Wiki.Org: the use of crowdsourcing to improve land cover. *Remote Sensing, 1*(3), 345–354.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., et al. (2012). Geo-Wiki: an online platform for improving global land cover. *Environmental Modelling & Software, 31*, 110–123.

Gonçalves, P. (Ed.), (2010). *OpenGIS® opensearch geospatial extensions draft implementation standard. Version 0.0.2*. Open Geospatial Consortium Inc, Ref. OGC 09–084r3.

Goodchild, M. F. (2007). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research, 2*, 24–32.

Goodchild, M. F., Guo, H., Annoni, A., Bian, L., de Bie, K., Campbell, F., et al. (2012). Next-generation Digital Earth. *Proceedings of the National Academy of Sciences, 109*(28), 11088–11094.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth, 3*(3), 231–241.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics, 1*, 110–120.

Granell, C., Díaz, L., & Gould, M. (2010). Service-oriented applications for environmental models: reusable geospatial services. *Environmental Modelling & Software, 25*(2), 182–198.

INSPIRE EU Directive. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Official Journal of the European Union, L 108/1, 50*.

Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., et al. (2001). Sustainability science. *Science, 292*(5517), 641–642.

McInerney, D., Bastin, L., Díaz, L., Figueiredo, C., Barredo, J. I., & San-Miguel-Ayanz, J. (2012). Developing a forest data portal to support multi-scale decision making. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, . http://dx.doi.org/10.1109/JSTARS.2012.2194136.

Masser, I. (2005). *GIS Worlds: Creating spatial data infrastructures*. Redlands: ESRI Press.

Masser, I., Rajabifard, A., & Williamson, I. (2007). Spatially enabling governments through SDI implementation. *International Journal of Geographical Information Science, 22*(1), 5–20.

Munro, R. (2012). *Crowdsourcing and the crisis-affected community: Lessons learned and looking forward from mission 4636*. Information Retrieval. http://dx.doi.org/10.1007/s10791-012-9203-2.

Nativi, S., Craglia, M., & Pearlman, J. (2012). The brokering approach for multidisciplinary interoperability: a position paper. *International Journal of Spatial Data Infrastructure Research, 7*, 1–15.

Nativi, S., Craglia, M., Vaccari, L., Santoro, M. (2011). Searching the New Grail: inter-disciplinary interoperability. In *Proceedings of the 14th AGILE international conference on Geographic information science – advancing geoinformation science for a changing world, Utrecht, 2011*. AGILE. (AGILE 2011). Utrecht, The Netherlands, April 2011.

Nebert, D. (Ed.), (2004). *Developing spatial data infrastructures: The SDI Cookbook*. GSDI.

Neis, P., Zielstra, D., & Zipf, A. (2012). The street network evolution of crowdsourced maps: openstreetmap in Germany 2007–2011. *Future Internet, 4*(1), 1–21.

Newell, D. A., Pembroke, M. M., & Boyd, W. E. (2012). Crowd sourcing for conservation: Web 2.0 a powerful tool for biologists. *Future Internet, 4*(2), 551–562.

Omran, E. L. E., & van Etten, J. (2007). Spatial-data sharing: applying social-network analysis to study individual and collective behaviour. *International Journal of Geographical Information Science, 21*(6), 699–714.

Ostermann, F., & Spinsanti, L. A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. In *Proceedings of the 14th AGILE international conference on geographic information science – Advancing geoinformation science for a changing world, Utrecht*, 2011. AGILE. (AGILE 2011). Utrecht, The Netherlands, April 2011.

Pearlman, J., & Shibasaki, R. (2008). Global earth observation system of systems. *IEEE Systems Journal, 2*(3), 302–303.

Pultar, E., Raubal, M., Cova, T. J., & Goodchild, M. F. (2009). Dynamic GIS case studies: wildfire evacuation and volunteered geographic information. *Transactions in GIS, 13*(s1), 85–104.

Rajabifard, A., Feeney, M.-E. F., & Williamson, I. P. (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geo-information, 4*, 11–22.

Reid, W. V., Chen, D., Goldfarb, L., Hackmann, H., Lee, Y. T., Mokhele, K., et al. (2010). Earth system science for global sustainability: grand challenges. *Science, 330*(6006), 916–917.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2012). Geoweb and crisis management: issues and perspectives of volunteered geographic information. *Geojournal*, . http://dx.doi.org/10.1007/s10708-011-9423-9.

Schade, S., Díaz, L., Ostermann, F., Spinsanti, L., Luraschi, G., Cox, S., et al. (2012). Citizen-based sensing of crisis events: sensor Web enablement for volunteered geographic information. *Applied Geomatics*, . http://dx.doi.org/10.1007/s12518-011-0056-y.

Schill, C., Perger, C., Fritz, S., McAllum, I., Díaz, L., Nativi, S, et al. (2012). Web 2 tools to improve global land cover: linking the EUROGEOSS broker and geo-wiki. In *EuroGEOSS conference*. Madrid 25–27 January 2012.

Singleton, A. D., & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography, 29*(3), 289–298.

Tamayo, A., Granell, C., & Huerta, J. (2012). Measuring complexity in OGC web services XML schemas: pragmatic use and solutions. *International Journal of

Geographical Information Science, 26(6). http://dx.doi.org/10.1080/13658816.2011.626602, ISSN 1365–8816. Taylor & Francis.

Tulloch, D. L. (2007). Many many maps: empowerment and online participatory mapping. *First Monday, 12*(2), Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1620/1535.

Turner, A. (2006). *Introduction to neogeography.* Sebastopol: O'Reilly Media Inc.

Vaccari, L., Craglia, M., Fugazza, C., Nativi, S., & Santoro, M. (2012). Integrative research: the EuroGEOSS experience. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* . http://dx.doi.org/10.1109/JSTARS.2012.2190382.

Yang, C., Raskin, R., Goodchild, M. F., & Gahegan, M. (2010). Geospatial cyberinfrastructure: past, present and future. *Computers, Environment, and Urban Systems, 34*(4), 264–277.

Zlatanova, S., & Fabbri, A. G. (2009). Geo-ict for risk and disaster management. In Scholten., v/d Velde., & van Manen. (Eds.), *Geospatial technology and the role of locations in science* (pp. 239–266). Dordrecht: Springer.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy, 2*(2), 6–32.