Alternative search mechanism for web 2.0 resources∗

Laura Díaz1, Manuela Núñez-Redó1, David González1, José Gil1, Pau Aragó1, Edward Pultar1, Joaquín Huerta1

1University Jaume I, GeoTech/INIT, Castelló de la Plana, Spain. laura.diaz, nunezm, gonzaled, jose.gil, parago, pultar, huerta [@uji.es].

Abstract

Currently we are witnessing how ordinary citizens are willing to share geospatial information using the friendly and easy-to-use functionality provided by the web 2.0 platforms. These platforms act as social networks describing events with big social impact and reflecting what is now known as Citizen Science. In particular events, such as those close to human settlements and urban areas, social networks are filled with volunteered information. A great majority of the information is volunteered geographic information (VGI) and contains location information thus sharing knowledge about an event’s evolution and impact. For the time being, in order to discover and retrieve this VGI it is necessary to deal with the different search mechanisms provided by various web 2.0 services. This paper explores how to improve the interoperability of these platforms by providing a single service as a unique entry point with an interface that implements an open standard specification. This paper discusses and demonstrates the use of the Open Geospatial Consortium’s Open Search Geospatial and Time specification as an interface for a service which searches, retrieves and aggregates information available in different web 2.0 services. We discuss how this information could be useful in complementing other official and scientific information sources by providing an alternative, contemporary source. We demonstrate this with a proof of concept presented in a forest fire monitoring scenario. The intrinsic interoperability of the system is reflected in the collaborations shown with different information systems such as those placed at the biodiversity and forestry units in the Institute of Environment and Sustainability at the Joint Research Centre.

∗This work is licensed under the Creative Commons Attribution-Non commercial Works 3.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/ or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.
1. INTRODUCTION

Geographical Information Systems (GIS) have become indispensable tools for organizing and exploiting the large amount of geospatial content for environmental sciences. Additionally, GIS provide a framework for multidisciplinary analysis (Ramamurthy, 2006).

Along with the evolution of other information systems, GIS has moved towards distributed environments based on Web Services and Service Oriented Architectures (SOA) (Aalst, 2007). In order to increase interoperability in the geospatial domain, OGC provides multiple standard specifications for data encodings and service interfaces (Percival, 2008). The combination of these standards allows for establishing Geospatial Information Infrastructures (GIIs), also known as Spatial Data Infrastructures (SDIs) (Masser, 2005). However, GIIs suffer from a lack of user motivation regarding participation and content management (Coleman et al., 2009; Díaz et al., 2011). Recent natural disasters have demonstrated existing difficulties in accessing and efficiently exploiting geospatial resources in GIIs (Zlatanova et al., 2009). The difficulties stem from the absence of sufficient available resources and a lack of collaboration between different geospatial infrastructures and components.

On the other hand, in the present day we are witnessing the consolidation of a new generation of the World Wide Web in which the key features are bottom-up methodologies and user participation. The Web is now a collaborative environment that has turned users into active providers (Alameh, 2003) providing a massive amount of information (Belimpasakis et al., 2010). This information is mostly georeferenced according to a user’s location thus providing a large amount of georeferenced information available in a wide variety of domains.

The integration of this VGI provides a social view that complements scientific data. However, the integration of these data sources into GIIs poses new research challenges. For example, to retrieve VGI users must take into account the different capabilities and application programming interfaces (APIs) offered by each Web 2.0 service. To overcome this we propose an integrated, scalable solution that is based on standard specifications. The goal is to improve the interoperability between the many available heterogeneous Web 2.0 services. Hence this research addresses the following question:
How can we coalesce authoritative and volunteered citizen information to be utilized in environmental scenarios?

The goal is addressed with a middleware component that provides a single search interface. This aids in the discovery of information across different social networks and crowdsourced information platforms. The developed prototype is a discovery service that implements the OpenSearch Geo-Time standard interface specification.

The remainder of this paper is structured as follows: Section 2 defines the background and related work. The third section defines the Open Search specification. We present the software architecture of our approach in Section 4, and the prototype design in Section 5. Section 6 demonstrates the approach in a real world forest fire monitoring scenario and Section 7 demonstrates the improvements gained by utilizing interoperable techniques in the integration of various other system components. The paper closes with results and conclusions discussed in section 8.

2. BACKGROUND AND RELATED WORK

Geoscience research is an interdisciplinary field which benefits from the expertise of different specialists from disciplines such as remote sensing, biology, and technology (Goodchild, 2008). By means of Geospatial Information Infrastructures (GIIs), these experts share and manage data in order to run scientific models and produce useful information about our environment. On the other hand, Web 2.0 Services and crowdsourcing platforms used by ordinary citizens are populated with spatial information (in varying spatio-temporal resolutions) offering a complementary vision to monitor our environment. In this section we discuss the sharing of geospatial content using standards-based and Web 2.0 approaches then examine how to leverage both information sources.

2.1. Geospatial Information Infrastructures

GIIs enable users to share geospatial content in a distributed manner following the SOA approach. They are widely known as facilitators to coordinate geospatial information (Dessers et al, 2011). Additionally, GIIs play a key role in supporting users and providers by giving them the ability to discover, visualize, and evaluate geospatial data at regional, national and global scales (Masser, 2005).

International initiatives such as the Global Earth Observation System of Systems (Pearlman and Shibasaki, 2008 ) or the European Infrastructure (INSPIRE, 2007) describe the overall architecture and best practices for designing and implementing GIIs,where content is managed by means of regulated, standardized services. INSPIRE was adopted as a European directive in February 2007 and lays out a legal framework for the European GI regarding
policies and activities with environmental impact (INSPIRE, 2007). The technical level provides a range of interoperability standards available for integrating information systems (Mykkänen et al, 2008) and defining a network based on discovery, view, download, transformation, and invocation services. The inherent complexity of standardized SDI and the complex mechanisms of deployment get worse as GIIs grow (Béjar, et al., 2009). In this way, the publication of content, associated traditionally with the providers, is an arduous task provoking a lack of updated content (Díaz et al, 2011).

2.2. Web 2.0 Services and Volunteered Geographic Information

The emergence of Web 2.0 platforms and easy to use tools encourages ordinary citizens to produce and share Geographic Information (GI) on the Internet. This is a different technique compared to the top-down building methodologies and complex publication mechanisms of GIIs. Web 2.0-based activities show that users are willing to engage more actively in content creation. The progress in information technology and the evolution of the user into a provider have created the phenomenon resulting in concepts such as Web 2.0, Neogeography (Turner, 2006), Cybercartography (Tulloch, 2007) and volunteered geographic information (VGI) (Goodchild, 2007).

VGI, as an alternative source of information, to complement official information (Craglia et al, 2008), could be integrated within GIIs context in order to improve traditional geospatial analysis and decision support tasks (Flanagin and Metzger 2008; Pultar et al., 2009). Zook (2010) pointed out that VGI can provide “additional data at levels of granularity and timeliness that could not be matched by other means”.

Hybrid approaches for the integration of top-down and bottom-up methodologies to study the integration of Web 2.0 resources within the SDI context are founded on what has been coined as the reconceptualization of the SDI user role (Budhathoki et al., 2008; Omran and van Etten, 2007) (Craglia, 2008) and (Goodchild, 2010). Here the new SDI generation will be influenced by more active user participation. Other researchers have paid attention to the versatility of Web 2.0 systems in contrast to SDI maintenance and publication mechanisms, aiming to lower the barrier in SDI publication mechanism (Díaz et al, 2011). Recent research has also studied the retrieval of data directly from crowdsourcing services where a discovery service deployed on an INSPIRE-based infrastructure offers a standard, unique entry point for VGI retrieval (Nuñez et al, 2011).

2.3. Geospatial content discovery

Within GIIs, metadata and catalog services are key for properly discovering content (Craglia et al., 2007). The services support the ability to publish and search for metadata while also supporting information resource binding within
GII. In this context, Bigagli (2004) proposed GI-Cat, a SOAP-based web service providing basic functionalities of geospatial information cataloguing.

The EuroGEOSS Discovery Broker (Nativi et al, 2011) component based on GI-Cat is used in the context of the European Union project EuroGEOSS\(^1\). This allows for a unique access point to services provided by three disciplines covered in the project: biodiversity, forestry, and drought.

Web 2.0 services are immense online repositories with geo-referenced content. However, attempts at providing spatio-temporal search engines for VGI are relatively scarce (Tsai, 2011). The process of searching through multiple services becomes a tedious task because each service provides different data encodings, geo-referencing, and proprietary APIs. It is a goal of this work to provide an interoperable mechanism to search Web 2.0 content to be integrated in a GII context. Within the EuroGEOSS project the results of this research will be used to provide the EuroGEOSS Discovery Broker with Web 2.0 services as an additional source of content for the GII users.

3. THE OGC OPEN SEARCH GEO-TIME SPECIFICATION

Web 2.0 services provide public APIs to allow custom programs to interact with the services via specific encodings and functions. Since there is no standard set of methods or encodings, there is a technical barrier for discovering content from multiple sources in a homogeneous way.

The Open Geospatial Consortium (OGC) OpenSearch (OS) Geospatial and Time specification (Gonçalves, 2010) describes an interface based on minimal mandatory input which can be extended with spatial and temporal criteria. OS simplicity has made it rapidly become a successful search interface specification for Web repositories being adopted, for instance by Flickr and Wikipedia. This research has adopted the interface of the OpenSearch Geo-Time specification as further explained in the following sections.

3.1. Keyword-based discovery

OS defines a service interface for minimal search and retrieval capabilities. An OpenSearch-enabled service exposes an interface for client applications to send simple HTTP GET requests providing specific query parameters. As a result, responses are often encoded in lightweight data formats such as GeoRSS\(^2\), Atom\(^3\) or KML\(^4\). Each service must be described by its Description Document.

---

\(^1\) http://www.eurogeoss.eu
\(^2\) http://www.georss.org
\(^3\) http://tools.ietf.org/html/rfc4287
\(^4\) http://www.opengis.org/standards/kml
This document is a file that describes the search engine of the target service. The description may vary from one service to another, but there are several mandatory parameters:

- root node called `OpenSearchDescription`,
- `shortName` which contains a brief human-readable title to identify the search engine,
- `Description` which is a text description of the search engine
- URL with the location of where a search request may be executed.

The OpenSearch specification has only one mandatory query parameter called “searchTerms” allowing client applications to request information related to one or more keywords. Other query parameters such as those supporting pagination of results (i.e., "count", "startIndex", "startPage") are optional.

3.2. Spatial and Temporal-based discovery

In order to demonstrate the added value of VGI and our approach to leverage it within GII, section 6 will describe a particular use case where a particular user monitors a forest fire event and assesses its impact in the affected area and its biodiversity. With more level of detail this section will illustrate how more advanced search criteria are necessary in environmental-monitoring scenario, in order to add more constrains and refine results accuracy.

Specific search profiles are described by extending the OpenSearch specification. This way the OGC OpenSearch Geo Temporal specification defines a list of parameters to enable spatial and temporal filtering. This specificacion profile will allow the user to filter results by a particular place name, area, point and radius and by a time period.

4. SYSTEM ARCHITECTURE

In this section we elaborate on the architecture of the proposed approach. Our main goal is to extend traditional GII architecture with a middleware component that offers a standard interface to retrieve and integrate Web 2.0 content and information from GII sources.

Figure 1 shows a simplified overview of the proposed architecture based on INSPIRE. This architecture is composed (top-down) of the application layer, service layer, and content layer. GIIIs based on INSPIRE provide functionality to users by means of categorized services deployed on the service layer. Following these principles and the goal of having an interoperable way to access and integrate Web 2.0 content in the GII context, we propose to extend this architecture with a new Discovery Service. This new Discovery Service will provide the capacity of retrieving and integrating Web 2.0 content.
4.1. Content Layer

We focus on the integration of both official and non-official sources. Specifically, this work pertains to the retrieval of content provided by Web 2.0 services. These non-official resources are especially relevant due to the fact that users provide near real-time information and local knowledge that enriches official information.

Due to the substantial availability of resources in crowdsourcing platforms a large part of the retrieved results are not related to the target scenario and they present “noise” that has to be eliminated for an appropriated assessment.

4.2. Geospatial Networking Service Layer

A classic GII provides discovery, view, download and processing services that implement the standard-based interfaces in order to improve the interoperability of the system and their components. In our research we propose to extend this layer by adding a new discovery service called the Web 2.0 Broker (W2.0B). Following a brokering approach, the W2.0B provides the capacity to perform a spatio-temporal search within multiple Web 2.0 services and offers a homogeneous, unique entry point. To do so, we propose our W2.0B discovery service to implement the OGC OpenSearch Geo-Time interface. For a better integration of crowdsourced data with official sources, we have experimented with the integration of the W2.0B component as part of the EuroGEOSS Discovery Broker to create a search engine for accessing Web 2.0 content.
The results of the Web 2.0 search can then be integrated with official information available in the context of the GII. In this way the vast amount of VGI becomes a new source to complement scientific data.

4.3. Application Layer

In this layer users are presented with user interfaces, such as geoportals, providing the entry point to the functionality offered by the services. In this particular case, to demonstrate the functionality and added value of the Web2.0B, we deploy and test two types of Geoportals that access this service. The first is a dedicated Web 2.0 client application with an intuitive interface to perform queries according to keywords and spatio-temporal criteria. The W2.0B users may query to get results that will be transformed and presented in this layer using web mapping technology. Second, existing GIs and Geoportals (within the EuroGEOSS project) will make use of the functionality provided by the W2.0B both directly or indirectly through the EuroGEOSS Discovery Broker.

5. IMPROVING WEB 2.0 SERVICES INTEROPERABILITY: THE WEB 2.0 BROKER

One main goal is to improve interoperability between the multitude of different APIs published by Web 2.0 Services. The W2.0B follows a brokering approach and implements the Open Search Geo-Time specification to provide the ability to search and retrieve content from different Web 2.0 Services. A collection of social media services with geo-referencing capabilities have been analyzed (Table 1). Only those that support geospatial and temporal filtering functions through their public API have been selected as target repositories (Fonts et al, 2010). Although some Web 2.0 services implement the OpenSearch specification (e.g., Flickr, Wikipedia, Youtube), some of them do not offer the OpenSearch Geo-Time search interface. The W2.0B overcomes this limitation by offering spatial and temporal criteria queries to the services that provide this information.

Table 1 shows the current status of the W2.0B prototype. The colored cells show the service currently offered where the rows show both the GI resources and the Web 2.0 services. The columns show the operations available in the OS Geo-Time specification and implemented in order to query the Web 2.0 services. Results are available in Atom, KML and MIMETEXT KML (Abargues et al, 2010) formats.

5.1. Web 2.0 Broker –Design

Figure 2 shows the component diagram of the W2.0B. It illustrates its modular design and how the components are linked to each other. At the top of Figure 2 one can see how the W2.0B implements the OS Geo-Time specification.
The OS Core component deals with the interpretation of the query in the standard OS format. It retrieves the query and forwards it to the Search Engine component. Only the SearchTerms input parameter from the specification is mandatory, but other criteria can be specified. The accuracy of results is improved by adding spatial or temporal criteria.

The Search Engine component is in charge of transforming the query and the specified criteria to the concrete Web 2.0 services APIs. It plays a mediating role between service-specific APIs and the OpenSearch query by broadcasting the query to the requested Web 2.0 services. The capabilities supported by W2.0B are limited by the native functionality offered by each specific API.

**Table 1: Search Parameters implemented in the W2.0B applicable to Web 2.0 services.**

<table>
<thead>
<tr>
<th>Base OS params</th>
<th>Geo OS Extension</th>
<th>Data Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Term</td>
<td>Count</td>
<td>XLM</td>
</tr>
<tr>
<td>StartIndex</td>
<td>StartPage</td>
<td>Atom/RSS</td>
</tr>
<tr>
<td>Flickr</td>
<td>GoogleNews</td>
<td></td>
</tr>
<tr>
<td>GeoNames</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td>Meteo/Weather</td>
<td></td>
</tr>
<tr>
<td>OpenStreetMap</td>
<td>Youtube</td>
<td></td>
</tr>
<tr>
<td>Meteoclimatic</td>
<td>Youtube</td>
<td></td>
</tr>
<tr>
<td>Eurogloss</td>
<td>Discovery</td>
<td></td>
</tr>
<tr>
<td>Eurogloss</td>
<td>Broker</td>
<td></td>
</tr>
</tbody>
</table>

However, not all of the Web 2.0 services have an API for accessing data. For instance, weather information extracted from Meteo/Weather or environmental news from European media sites are provided as a data stream that is interpreted to transform it to a proper data encoding. This means that users can not perform custom queries. Currently the information retrieved from this sources is parsed to extract data related to Fire News in order to assess the value of this data in our use scenario for forest fire monitoring.
The Geo-reference Manager component deals with the management and improvement of the retrieved content geocoding in terms of spatial search accuracy and performance. This is because some results provided by the services do not contain any location. The core of this module, connected to geonames, concerns the improvement of this un-georeferenced data.

The OS Descriptor Manager dynamically generates the service description document\(^5\). This is a mandatory document specifying and detailing the service capabilities. It describes the valid query parameters and supported response formats. This allows the client application to understand the interfaces needed to build valid OpenSearch-style queries.

This component also allows for integrating a custom search engine within web browsers such as Internet Explorer, Firefox, and Opera. This custom search engine refers to the Multiquery search engine, which offers a multiple search in all the services provided by the W2.0B.

This broker architecture configuration is flexible enough to add new Web 2.0 services to the system without altering the broker’s discovery interface from the perspective of the client. Thus clients and modules are independent, loosely coupled components where each one evolves separately therefore enhancing the system scalability as a whole (Nuñez et al, 2011).

\(^5\) http://www.opensearch.org
5.2. **Web 2.0 Broker Implementation**

The W2.0B has been designed as a service with a standard interface to be re-used in different scenarios. In this section we illustrate how the W2.0B works when it is invoked. Figure 3 shows a sequence diagram illustrating the workflow of how the different components are invoked when the user sends a query. During the first step the client application can retrieve the required descriptors via the OS Descriptor Manager Component to see the services capacities in terms of search criteria and data encodings.

**Figure 3. W2.0B sequence diagram when doing a query.**

At this point, the client is able to build and send the OS query. When the W2.0B receives the query, the OS Core interprets it. The Search Engine contains a collection of Search engines or adapters that transform the query to a specific Web 2.0 API style. Next the W2.0B propagates the query to the different services by means of each search engine.

Since many of the retrieved results might not contain any location information, the next step is to improve georeferencing. The Geo-Reference component is in charge of parsing the result to find information as a toponym and using third party
services such as Geonames\textsuperscript{6} to extract coordinates from placenames to enhance the location of the result.

In the end the Search Engine aggregates the retrieved results and generates the response in the format requested by the client.

5.3. Integration of Web 2.0 Broker into EuroGEOSS Brokering Platform

W2.0B is integrated into the EuroGEOSS brokering platform as an accessor for the Discovery Broker. Its fundamental aim is retrieving multidisciplinary information from social networks which will be aggregated with the rest of the requested user content.

The EuroGEOSS broker provides multiple profilers (different interfaces to query data) and accessors (able to integrate multiple and heterogeneous sources) to increase interoperability. Since the W2.0B is integrated within the EuroGEOSS Discovery Broker, W2.0B acts as a mediator and enables Web 2.0 services to be accessed through other protocols due to the discovery broker’s profilers components. Figure 4 presents a screenshot of the EuroGEOSS Broker client demonstrating how this integration retrieves Web 2.0 data.

\textbf{Figure 4. Screenshot of W2.0B integration into EuroGEOSS Brokering Platform}

\textsuperscript{6} http://www.geonames.org/
6. MULTIDISCIPLINARY ENVIRONMENTAL MONITORING SCENARIO: FOREST FIRE MONITORING AND IMPACT ASSESSMENT

Environmental monitoring is a discipline that involves many possible scenarios with multiple variables and procedures. To illustrate the practical use of the W2.0B prototype we describe a multidisciplinary scenario involving forest fire monitoring. During and after the fire environmental experts must monitor the area in order to evaluate the fire damage and environmental (forestry resource, biodiversity loss and drought influences) and social impacts.

The monitoring during the post-fire phase will use both citizen science by collecting volunteered contributions through Web 2.0 services in addition to official environmental-related data. To illustrate with a practical example our scenario describes a fire event that occurred in Mijas, Málaga, Spain, in 2011.

In order to monitor the status of a detected fire, the user accesses the web client which provides a web map to visualize data coming from various sources. Additionally, the client shows the area burned by the fire retrieved from the European Forest Fire Information System (EFFIS) data services.

The web client offers simple and advanced interfaces (selectable 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.

For this example in Spain, the user restricts the area of interest to Mijas and limits the time period in order to retrieve Web 2.0 information concerning the specific data. At this moment the client application generates the OS-Geo-Time query and connects to the W2.0B to retrieve information about that area. Figure 5 shows the results of this query. The information retrieved from the system shows the integration of authoritative data with citizen-based Web 2.0 data. In particular, Figure 5 shows pictures taken from users who were near the forest fire when it happened. In the figure ones sees the added value of the citizen-based, Web 2.0 data and how it complements official data sources. As we can see the convex hull of the pictures provided by the citizens closely approaches the official burned area provided by EFFIS. In this case, in absence of official burned area data we could have roughly estimated it by using citizen data. In addition, the volunteered Web 2.0 data is available in near real time thus in most cases this information will be useful for time critical decisions and more human lives can be saved by acting on this data rather than waiting for official sources.

---

7 effis.jrc.ec.europa.eu
One of the main objectives of the W2.0B is helping studies by offering a citizen’s point of view. Citizens contribute an alternative source of information. Another example of this is shown in Figure 6 where VGI extracted from Flickr presents the Wildland-Urban Interface (WUI) (Stewart et al, 2007) between Ojén and Marbella in Spain. It has been demonstrated that identifying the area where houses meet or intermingle with undeveloped wild land vegetation will contribute to preventing and fighting forest fires (Haight 2004). Additionally, if this information is collected in real time it will aid in dynamic GIS calculations for when to evacuate a community in the event of a wildfire (Pultar et al, 2009).
VGI data is retrieved through the W2.0B client in different formats such as text, photo, or video. One formats for documenting fire behaviour is video provided by Youtube. In Figure 7 a result is shown where a Youtube video shows a plane taking part in extinguishing the wildfire in Mijas.

**Figure 6. VGI visual data of a Wildland-Urban Interface between Ojén and Marbella extracted from Flickr through W2.0B**

**Figure 7. Details of Youtube W2.0B video result in which a fire plane is extinguishing a forest fire.**
In addition, information regarding biodiversity may be extracted from VGI data. For example, Figure 8 presents the results of a query spatially restricted to the area of Mijas. Here flora species such as “Russelia equisetiformia” or “Delphinium gracile DC” have been geotagged by a Flickr user who uploaded the picture taken in the area near the forest fire.

Figure 8. Details of two W2.0B results retrieved from Flickr regarding biodiversity when querying for “flora”.

W2.0B is a tool to retrieve data to assist in multidisciplinary environmental monitoring based on citizen contributions. This is not a substitute for scientific data but could be a complementary source of information for natural disasters and hazards such as wildfires or hurricanes. Further steps are needed to analyze and model raw VGI in order to extract more accurate and relevant information out of the massive repository provided by social networks.

7. INTEGRATION OF W2.0B INTO OTHER SYSTEMS

The W2.0B component is based on standard specifications which improves its ability to work with other components and systems. This aspect of the component is illustrated in the following examples W2.0B’s added value is shown through the integration of the component into third party applications such as eHabitat.
(Dubois et al, 2011) and geo-wiki (Fritz et al, 2009). Figure 9 shows the use of the W2.0B component to retrieve information about a bird called “Circus Maurus” in South Africa through the eHabitat platform. This platform is designed for finding and assessing ecosystems with equal properties by the biodiversity department at the Institute of Environment and Sustainability (IES) at the Joint Research Center (JRC).

Figure 9. Screenshot of third partie eHabitat

Also, Figure 10 demonstrates how the forestry department at the IES of the JRC integrated a W2.0B client to access the features of this component for integrating scientific and citizen-based data into their system (Mcinerney et al, 2012).
Figure 10. Screenshot of third partie Eurogeoss Forestry web

Figure 11 shows results obtained from the W2.0B integrated in the geo-wiki project, which is a global network of volunteers aiming to improve the quality of different thematic datasets. 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).

Figure 11. Screenshot of integrating the W2.0B in the Geo-Wiki platform

8. DISCUSSION AND CONCLUSIONS
The massive amount of information provided by citizens in crowdsourcing platforms cannot be ignored. However, due mostly to their heterogeneity, leveraging this information presents many challenges in order to increase
interoperability and scalability to integrate them and consume them in scientific context such as GII.s.

Following the GII principles we have proposed a new Discovery Service to provide a standards-based, unique entry point to query multiple Web 2.0 services and retrieve citizen-based information to be integrated in GII contexts. This mechanism, the Web 2.0 Broker, interprets queries based on the standard-based Open Search Geo-Time specification.

The Web 2.0 Broker is able to interpret OS-Geo-Time queries and propagate them to a set of Web 2.0 Services then aggregate the results encoded in standard data formats such as GeoRSS, GeoJSON, KML, or ATOM.

One advantage of the proposed approach is that potential calibration mechanisms can be encapsulated in well-defined components that directly connect and use the specific Web 2.0 service APIs. However, the search criteria based on the OS need to be mapped into the specific Web 2.0 service APIs and this means that we could lose accuracy in certain parts of a query. This may have an impact on the number of VGI items retrieved and therefore further work and analysis is needed in this respect.

Our work indicates that VGI can complement GII data by providing high-scale, value-added information at a low cost. The W2.0B can be used to complement crisis management model inputs and to refine their output results. The particular scenario presented in this paper demonstrated the multidisciplinary capacity or our approach. The scenario illustrates how general environmental monitoring and, in particular a forest fire monitoring scenario can leverage the potential of the massive, multidisciplinary VGI available. In our use case we have consumed raw VGI having performed a preliminary visual analysis having a first glance of how it could complement the scientific workflow, however the validation of global models with this knowledge extracted by VGI analysis has yet to be fully exploited. There is a need for more sophisticated analysis to filter the massive amount of data, 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, that although can be less reusable can have more added value to be used in specific use cases and scenarios.

Further development of the Web 2.0 Broker is currently on-going. We aim to improve the means by which VGI is harnessed and integrated into GII thereby leveraging its full potential. First by increasing the number of Web 2.0 resources to be aggregated, and second to analyze the massive data flow to extract observations relevant to specific use cases. Current steps are to define a data model to describe forest fire observations and alarms to add a new source in
emergency response scenarios. The intrinsic multidisciplinary character of the component favors the plurality of use cases, therefore other tests in data mining VGI to infer and assess green routes in a natural environmental management scenario and or extracting meaningful places in a urban management scenario are being carried out.

ACKNOWLEDGEMENTS
This work has been partially supported by the European FP7 Project nr. 226487 called EuroGEOSS.

REFERENCES


Craglia, M. (2007). Volunteered Geographic Information and Spatial Data Infrastructures: when docparallel lines converge?, Proceedings the

Craglia, M., I. Kanellopoulos, P. Smits (2007). Metadata: where we are now, and where we should be going, Proceedings 10th AGILE International Conference on Geographic Information Science 2007. Aalborg University, Denmark.


