Discovery and Integration of Web 2.0 Content into Geospatial Information Infrastructures: A Use Case in Wild Fire Monitoring

Manuela Núñez-Redó, Laura Díaz, José Gil, David González, Joaquín Huerta

Discovery and integration of Web 2.0 content into Geospatial Information Infrastructures: A use case in wild fire monitoring

Manuela Núñez-Redó, Laura Díaz, José Gil, David González, Joaquín Huerta
Institute of New Imaging Technologies, University Jaume I, Castellón, Spain
{nunezm, laura.diaz, jose.gil, gonzaled, huerta}@uji.es

Abstract—

Efficient environment monitoring has become a major concern for society to guarantee sustainable development. For instance, forest fire detection and analysis is important to provide early warning systems and identify impact. In this environmental context, availability of up-to-date information is very important for reducing damages caused. Environmental applications are deployed on top of Geospatial Information Infrastructures (GIIs) to manage information pertaining to our environment. Such infrastructures are traditionally top-down infrastructures that do not consider user participation. This provokes a bottleneck in content publication and therefore a lack of content availability. On the contrary mainstream IT systems and in particular the emerging Web 2.0 Services allow active user participation that is becoming a massive source of dynamic geospatial resources. In this paper, we present a web service, that implements a standard interface, offers a unique entry point for spatial data discovery, both in GII services and web 2.0 services. We introduce a prototype as proof of concept in a forest fire scenario, where we illustrate how to leverage scientific data and web 2.0 content.

Keywords—VGI, SDI, Geospatial Information Infrastructures, Web 2.0, Open Search

I. INTRODUCTION

Analyzing the Earth's behavior requires a multidisciplinary approach and the assessment of a broad range of thematic areas [1] [2]. Geospatial information is essential for addressing related challenges. The amount of scientific geospatial data collected has increased significantly due to advances in data-capturing technologies. Geospatial Information Systems (GIS) have become indispensable tools for organizing and exploiting this geospatial content for environmental sciences, and providing a framework for multidisciplinary analysis [3].

In this domain, we find multiple geospatial standards for data encodings and service interfaces. The combination of these standards allows for establishing Geospatial Information Infrastructures (GIIs), also known as Spatial Data Infrastructures (SDIs) [4]. These multi-participatory infrastructures allow administration and other official providers to publish environmental information. However, GIIs are dynamic and require continuous maintenance. Still, GII complex deployment mechanisms limit the possible contributions of expert users suffering from a low-rate of user motivation regarding participation and content management [6][7]. Recent natural disasters such as the Indian and Chilean tsunamis, forest fires in Greece and California or the earthquake in Haití have demonstrated that difficulties still exist inefficiently exploiting geospatial resources in GII. The difficulties stem from the absence of sufficient available resources and a lack of collaboration and interrelation between different geospatial infrastructures and components.

In contrast, we are witnessing the consolidation of a new generation of the World Wide Web, in which the main feature is user participation. Tim O'Reilly (2005) popularized the evolving nature of the web by introducing the term 'Web 2.0.' The Web is now a collaborative environment where the increasing number of web-based social networks has turned users into active providers [11] [31], motivated to provide a massive amount of information [30]. This user-generated content [16] is mostly geo-georeferenced to the user location, leading to huge amounts of geo-referenced information available in practically any domain [32]. For example, the photos uploaded by the public on Flickr during the California wildfires in 2007. These photos provided a quicker overview of the situation than information coming from official channels, such as mapping agencies.

This information provides a complementary view to the scientific data and shows the social impact of environmental events like forest fires. Furthermore, due to advanced devices that allow users to capture and share data from the field, there is a massive source of geo-information available at near real time. This humans-as-sensors paradigm [17] provides a new means of providing data in its context, which is fundamental to the vision of a spatially enabled society [14]. To leverage this new source of information, has to be integrated, in the context of the GIIs, to enrich scientific information with social and local knowledge.

However, the integration of these sources of information in GII poses new research challenges. Environmental applications need to retrieve this valuable information. This means to deal with the different search interfaces of each web 2.0 service and its heterogeneous capacities.

To address these issues, we propose a more scalable solution, which aims at improving the interoperability of the heterogeneous nature of the multiple Web 2.0 services available.

Our proposal materializes in a middleware component that provides a homogeneous search interface to improve the discover ability over social networks and crowd sourcing platforms. This middleware is a discovery service that implements the OpenSearch Geo-Time standard interface specification. Although some of the Web 2.0 services already expose an OpenSearch interface to increase interoperability, they do not offer the spatio-temporal query capacity, which is crucial for most of the environmental scenarios. The realization of our proposal, called Web 2.0 Broker (W2.0B), offers a common entry point to retrieve and aggregate web 2.0 content according to spatial and temporal criteria.

The remainder of this paper is structured as follows: Section II defines the overall context of geospatial information infrastructures, volunteered geographic information and geospatial discovery. Section III defines the Open Search specification. We present the architecture in Section IV, the prototype (Section V), and conclude the paper with a discussion and an outlook towards future work (Section VI).

II. BACKGROUND AND RELATED WORK

Geoscience research is a multidisciplinary field that demands heterogeneous data and a multitude of expert profiles such as technologists and remote sensing specialists [1] [2] [8]. These experts collect and manage data to run scientific models and produce information. On the other hand web 2.0 Services and crowdsourcing platforms have demonstrated how ordinary citizens, encouraged by technological advances, are also able to generate and publish high scale spatial information at near real time offering a complementary vision to monitor our environment. In this section, we briefly reflect on standard based approaches to share geospatial content and approaches to leverage both sources of information to enrich environmental monitoring.

A. Geospatial Information Infrastructures

A trend involves deploying geospatial and environmental applications on service-oriented architectures (SOA) [9]. One of the goals of SOA is to enable interoperability among existing technologies and provide an interoperable environment based on reusability and standardized components.

GIIs enable end-users to share geospatial content in a distributed manner following an SOA approach. They play a key role in supporting users and providers in decision-making where they can discover, visualize, and evaluate geospatial data at regional, national and global levels [4].

International Initiatives describe the overall architecture and best practices to design and implement GIIs. Content is managed by means of regulated and standardized service types. Adopted as a European directive in February 2007, INSPIRE sets out a legal framework for the European GII, regarding policies and activities with environmental impact [5]. It defines a network based on discovery, view, download, transformation and invocation services. The technical level provides a range of interoperability standards available for the integration of information systems [12].

In this context many authors address questions concerning the increasing number of people participating in crowdsourcing platforms while GIIs traditionally face problems to attract users. GII researchers have called for a user-driven model [60] [61] [62], which relates to the hybrid GII that incorporates user generated content. Combining scientific knowledge and public information is not new, according to [58] . [59] develope the "citizen panels" in the 1970's involving experts and citizens to allow everybody to participate. In the context of municipal activities, [57] also proposed to capture and utilize the "city knowledge" from those close to a particular phenomenon with the richest geospatial knowledge. Another example is the management of natural resources in the Amazon where there is a need for user participation to integrate their local knowledge [63].

B. Web 2.0 Services and volunteered geospatial information

With the emergence of Web 2.0, ordinary citizens have begun to produce and share Geographic Information (GI) on the Internet. These Web 2.0-based activities show that users are willing to engage more actively in the production and provision of contents. This gives rise to a new phenomenon, which has been referred to as "neogeography" [15] [18], "cybercartography" [30], or "voluntary geographic information" (VGI) [17] [21].

VGI provides a massive source of information that cannot be ignored. This information can complete gaps in official data including cheap and big scale up to date information. Research has related these collaborative services, paying special attention to the trust and credibility [23] [25], quality, and reliability as compared to official data, [22] [26] constraints and user motivation [14] [24].

In our scope we consider VGI as a complement to official data. Scientists will be provided with discovery mechanisms

to retrieve appropriate VGI that will complement the social aspect with their scientific information.

The use of a hybrid approach that integrates bottom-up and top-down methodologies has already been demonstrated [27], with the purpose of integrating user generated information, scientific tools and official information in the same geospatial infrastructure. In this context merging the top-down SDI model with VGI infrastructures has already been described [13] [17] [28] [29].0 [7] [20] describe a publication service to hide complex standards and assist users in publishing content directly into standard geospatial data services. This direct user publication raises issues about data consistency and quality in GII. A Second approach is the retrieval of data directly from crowdsourcing services. In this context we propose a Discovery Service that deploys on an INSPIRE-based infrastructure that offers a standard and unique entry point to retrieve Web 2.0 resources according to spatial and temporal criteria that can be integrated with official environmental information to produce a richer and more up to date system [14].

We have performed a survey of the multiple Web 2.0 Services available, with the aim of selecting those where spatial and temporal queries were reliable. The service selection is the following: Twitter a social networking and micro-blogging service. Its users can send and read text-based posts of up to 140 characters, so-called "tweets", which are publicly visible by default. Flickr is an online application that allows uploading, storing and organizing photographs. It enables the creation and retrieval of the pictures. YouTube allows sharing videos that can be geo-referenced. OpenStreetMap and Geonames are both explicit VGI platforms. Meteoclimatic is a weather resource and Wikipedia is an open free web encyclopedia.

C. Geospatial content discovery

Within GIIs, metadata and catalogue services are key to discover content properly [54] [55]. In this context most of the issues arise because metadata creation and publication is a complex and arduous task that has to be done manually [53]. On the contrary, the ease of content production and publication in Web 2.0 Services makes vast amounts of VGI available. As a result, social networks are immense online repositories with geo-referenced content. However, attempts at providing spatial and temporal-based search engines over VGI are relatively scarce [33]. Despite the popularity of Web 2.0 services, there has not been many integrated and interoperable approaches that allow users to search for content regardless of the nature of the underlying services [37]. One of the reasons may be found in the diversity and heterogeneity of these types of services and their interfaces.

The process of searching over multiple services becomes a tedious task because they provide different data encodings, geo-referencing and proprietary application programming interface (API).

In order to overcome this heterogeneity the use of a common interface, following a standard specification, would increase interoperability. Walsh [35] points out the need to pay attention to discovery interfaces widely spread in other information communities different from the established catalogues services in the geospatial domain [36]. Several Web 2.0 services, like Flickr, expose basic discovery capacities through the Open Search specification allowing for a common technique to run term-based queries. In this sense, our approach extends this to the design a mechanism that allows users to search using the OpenSearch (OS) specification [38] by also adding spatial-temporal criteria to retrieve content over different social networks and services.

III. THE OPEN SEARCH SPECIFICATION

Web 2.0 services expose their own API to be accessed, using specific encodings formats and schemas. This constitutes a technical barrier for discovering content in a homogeneous way.

To overcome this problem our proposal is materialized in a discovery web service, the so-called Web 2.0 Broker. The W2.0B implements the OpenSearch specification with the Geo and Time extension [51], adopted by the Open Geospatial Consortium and the geospatial community as a standard de facto.

OS specification offers an interface based on minimal input, which can be extended, among others, with spatial or temporal criteria. OS has rapidly become a successful search specification over web repositories, which are increasingly adapting it to demonstrate their search interfaces in a standard and simple way. In this section we describe the OpenSearch Geo-Time specification and how we have adopted it as the interface for our discovery service to allow users to perform spatio-temporal queries over social networks.

A. Keyword-based discovery

OS defines a service interface for minimal search and retrieval capabilities. The simplicity of OS for search fits into the basic search interfaces that identified many Web 2.0 services. 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 [39], Atom [40] or KML [41].

The OpenSearch specification has only one mandatory query parameter called "searchTerms" allowing client applications to retrieve information related to one or more keywords. Other query parameters like those supporting results pagination ("count", "startIndex", "startPage") are optional. The W2.0B implements this interface to broadcast keywords-based search over a selected pool of Web 2.0 services. Each service must be described by its Description Document, a file whose aim is, to describe the search engine of the target service. This description could vary from one to another, but there are several mandatory parameters like the root node called *OpenSearchDescription*, *shortName* that contain a brief human-readable title to identify this search engine, *Description* which is a text explanation of the

search engine and the URL with the location to execute a search request.

B. Spatial and Temporal-based discovery

More advanced search criteria are necessary in our environment-monitoring scenario. Specific search profiles are described by extending the OpenSearch specification. The OGC OpenSearch Geo Temporal specification [51], defines a list of query parameters to enable spatial and temporal filtering. This extension is built upon the basic OpenSearch specification, so all mandatory and optional query parameters previously mentioned are also available. The spatial and temporal extensions define spatial and time specific, optional query parameters. The "geo:box" parameter filters are the result of a rectangular area. The "geo:lat", "geo:lon" and "geo:radius" parameter filters results from a circular area around a point. The "geo:geometry" parameter defines a geographic filter by means of an arbitrary geometry. The "geo:name" parameter allows filtering by place name. The "time:start" and "time:end" allow the definition of a temporal range for valid results.

The W2.0B implements the OS Geo-Time specification. Therefore, we increase the interoperability of the selected crowdsourcing platforms because now client applications can run spatio-temporal queries in these services by using OS-Geo-Time as unique interfaces. Table 1 shows the current status of the W2.0B prototype. The colored cells show the service currently offered where the rows show the web 2.0 services where the query is propagated and the columns shows the operations, available in the OS Geo-Time specification, implemented to query the Web 2.0 services. Regarding the response formats, Atom, the format recommended by [51], KML and MIMETEXT KML[43] are supported.

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

		Base OS params				Geo Extension				Time Ext		Data Formats		
		Search Terms	Count	Start Index	Start Page	bbox	Lon,lat,radius	geometry	Name/location	start	end	KML	KML/MIMEXT	ATOM (+GeoRSS)
Web 2.0 Service	Twitter													
	OpenStreetMap													
	Meteoclimatic													
	YouTube													
	Flickr													
	Geonames													
	RSS (Geonames)													
	Wikipedia													
	Multiquery													

IV. 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.

The W2.0B offers the functionality to perform a spatio-temporal search of VGI in multiple Web 2.0 services for its integration with official environmental information available in the infrastructure. In this way the vast amount of VGI becomes another data source in GIIs to complement scientific data. The integration is performed at the client side since the W2.0B provides common data encoding.

Figure 1 shows a simplified overview of the INSPIRE technical architecture which basically extends a classical three-layered SOA. This architecture is composed (top-down) of the application layer, service layer and content layer. The next subsections will describe each of these layers and the components that describe our contribution.
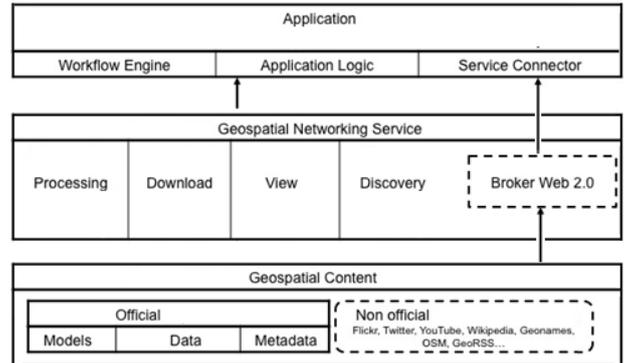

Figure 1: Classical three-tier GII architecture (extracted from [10]) extended with a new channel connecting to VGI sources.

A. Application Layer

Whereas traditional applications offer access to the content via the Geospatial Networking Service layer (Figure 1), this architecture offers an additional entry point to the W2.0B. This functionality is part of the Service Connector module, since it is the component in charge of connecting to the available middleware.

In this layer users are presented with a friendly interface to perform queries according to keyword and spatial-temporal criteria. Both standard catalogues and the W2.0B can be used to retrieve results. In the case of the W2.0B, this query will be transmitted and the results transformed to a well-known common data encoder to be presented in this layer using geospatial Web Mapping technology.

B. Geospatial Networking Service Layer

A classical GII provides discovery, access and processing services that, implement the OGC standards, such as Catalogue Services (CS-W), Web Mapping Service (WMS), Web Feature Service (WFS), and Web Processing Service (WPS), providing access to the geospatial content. In our research we propose to extend this layer with a new discovery service that, by implementing the OpenSearch Geo-Time interface, acts as a search engine to access Web 2.0 content.

W2.0B allows searching for VGI data through its standard OS Geo-Time interface and through a specific API. Client that have two ways to connect on the one hand, through its

Java API which facilitates the development of client applications implemented in JAVA or accessing the OS standard interface.

C. Content Layer

We focus on the integration of both official and non official sources. Specifically, we focus on the retrieval of the content provided by Web 2.0 Services. These non-official resources are especially relevant due to the fact that users provide real time information, local knowledge and social impact to enrich the official environmental information.

Table 1 shows the set of crowdsourcing platforms being queried by the W2.0B. The content provided by these platforms differs in nature, for instance, geo-tagged photographs are shared through Flickr [45], short text messages shared via Twitter [48], or videos uploaded to Youtube [50], encyclopedic descriptions in Wikipedia [49], place names using Geonames [46], tagged vector geometries in OpenStreetMap [47] or information about weather stations provided by Meteoclimatic [56]. Due to the substantial availability of resources in crowdsourcing platforms a big part of the retrieved results are not related to the target scenario and they represent somehow "noise" that has to be eliminated for an appropriated assessment. Although it is out of the scope of this current investigation to assess data quality, section VII overviews a preliminary discussion and outlines open questions that remain for future research.

V. WEB 2.0 BROKER: OPEN SEARCH SERVICE

W2.0B implements the function to search different social networks and Web 2.0 Services. A collection of social media services, with geo-referencing capabilities, has been analyzed (Table I), and only those that support, to some degree, geospatial and temporal filtering functions through their public API have been selected as target repositories [44].

Although some Web 2.0 services implement the OpenSearch specification (Flickr, Wikipedia, Youtube OpenStreetMap), some of them do not offer the OpenSearch-Geo-Time search interface. The W20B overcomes this limitation by offering spatial and temporal criteria queries to these services.

A. Web 2.0 Broker –Design

Figure 2 shows the component diagram of the W20B. It illustrates its modular design and how the components are linked to each other. Below, we will elaborate more on each component and its functionality.
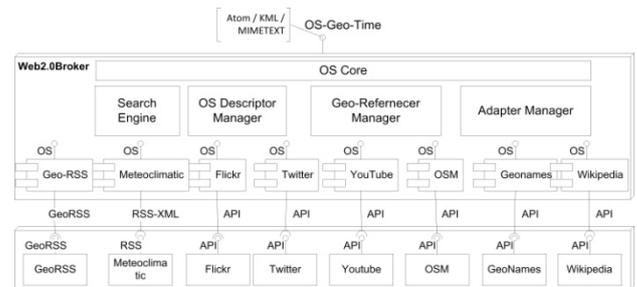

Figure 2: W2.0B components diagram.

At the top of Figure 2 we can see how the W2.0B implements the OS Geo-Time specification to provide discovery capabilities over heterogeneous VGI resources.

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 either by adding spatial filtering with the bbox, location or lon, lat and radius parameters, or adding time criteria by setting up the start and end parameters as we will see in the discussion section.

The *Search Engine* component contains the logic to map the query and the specified search criteria to then perform the concrete operations offered by the different web 2.0 services. It will broadcast the query to the different search engines and adapters that the client has selected. The supported query can add spatial constraints to the queries sent to the Web 2.0 services that natively support geographic search capabilities through their own API. For instance, users can search for resources that are restricted to a given area of interest represented as a polygonal geometry.

The *OS Descriptor Manager* dynamically generates the service description document [38] specifying how the different services must be queried. This document, mandatory by the specification, is generated by each service adapter. Its function is to advertise the set of accepted query parameters and supported response formats. This allows both calls from outside and within a client application to understand the discovery interfaces supported by services and how to build valid OpenSearch-styled queries.

The *Geo-reference Manager* component deals with the management of geospatial content. In terms of spatial search accuracy and performance, the W20B relies on the spatial search capabilities provided by the services queried and the content available, for instance, only a few tweets are actually geo-referenced. In this case, the Geo-reference Manager follows a methodology to extract the location from the user profile by using the Geonames service to extract the location of a place name found in the content.

The *Adapter Manager* component is the manager of a set of specific adapters for each service. It plays a mediating role between service-specific APIs and the OpenSearch query.

The W2.0B clients control the search procedure by selectively activating one or more services. The Adapter

Manager delegates on the selected adapters which adapt the query to the specific APIs. The manager will aggregate the results. Furthermore, since these social networks and media services offer specific discovery interfaces, they also provide different response formats.

Table 1 shows the specific capabilities of each adapter in terms of the OpenSearch Geo-Time features. Capabilities supported by W2.0B were shaded in the corresponding cell(s). These capabilities are limited by the functionality offered natively by each specific API. For example, some services allow filtering by bounding box and others by centre and radius. In all cases, KML and Atom extended with GeoRSS are provided as standard geographic formats in the response. In addition, the nature of each service leads to different constraints and requirements in terms of discovery. For instance, Flickr's resources can be queried over time while Twitter's resources are only discoverable during a narrow time frame. Indeed, these open issues pose new challenges in the field of social mining.

However, not all the web 2.0 services have an API from which the data is accessed. For instance wheather information extracted from Meteoclimatic or environmental news from European media sites are analyzed and interpreted to add them information sources. For the time being, the RSS information that is retrieved is related to Fire News and MODIS Hotspots of the last seven days, and it is gathered from European Forest Fire Information System (EFFIS)[52] sources through the European Commission's Joint Research Centre (JRC).

This component also allows the integration of the custom search engine with the most popular web browsers such as Internet Explorer, Firefox and Opera. This custom search engine refers to the Multiquery adaptor which offers a multiple search of each service provided by the W2.0B (Table 1). To do so, we add auto-discovery by adding to the search client a HTML tag which points to the corresponding OpenSearch description document. This tag activates the "search engine manager" of the browsers to offer the possibility of adding Web 2.0 broker as a new Search Engine as we can see in Figure 3.
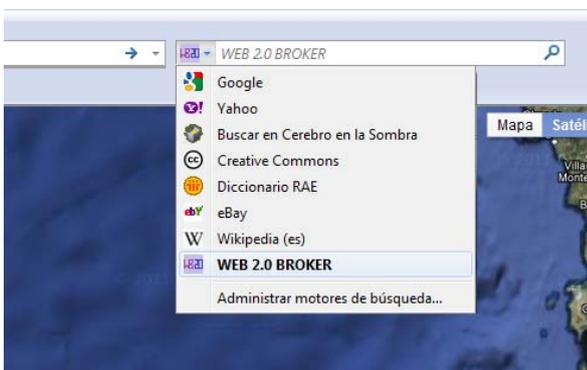

Figure 3: Web 2.0 Broker search engine added in a common browser.

This broker architecture configuration is flexible as new adapters may be added without altering the broker's discovery interface from the client perspective. In doing so, clients and adapters are independent, loosely coupled components where each one evolves separately, enhancing the system scalability as a whole [34].

B. Web 2.0 Broker implementation

The W2.0B has been designed as a middleware component with a standard interface to be re-used in different scenarios. OpenSearch Geo-Time is the standard of choice to implement the W2.0B in order to increase interoperability when accessing multiple Web 2.0 services. In this section we illustrate how the W2.0B works when it is invoked. Figure 4 shows a sequence diagram illustrating the workflow of how the different components are invoked when running a query. The W2.0B receives a OS GeoTime styled query such as the following:

http://elcano.dlsi.uji.es:8082/broker.jsp?service=service &q={searchTerms}&per_page={count?}&page={startPage}&format={responseFormat}&name={geo:name}&lon={geo:lon?}&lat={geo:lat?}&radius={geo:radius?}&bbox={geo:box?}&start={time:start}&end{time:end}&format=dat format

This adapter injects the query parameters into the specific Flickr API discovery methods and carries out the query. Optional query parameters may be encoded in the URL itself for results pagination, language selection, and character encoding.

First of all, as shown in Figure 4, search clients retrieve the required descriptors via OS Descriptor Manager component.
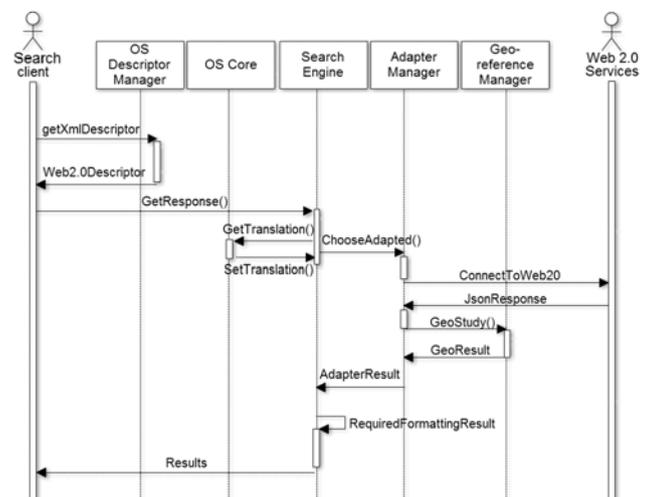

Figure 4: Web 2.0 Broker sequence diagram.

Once the descriptor is loaded, the client is able to build and send the OS query. When the query is received, the OS Core interprets it and the Search Engine propagates the query to the Adapter Manager. Afterwards, the adapter

6

manager connects and queries each service by means of the specific adapter.

At this point, the Geo-Reference component becomes important as it is in charge of several geospatial features such as extracting coordinates from a placenames, as well as the inverse functionality, or getting the center coordinates and radius from a bounding box based on the Haversine formula, implemented in this module.

The Adapter Manager aggregates the retrieved results and the Search Engine component generates the response to be derived to the client.

VI. PROTOTYPE: USE CASE IN FOREST FIRE MONITORING

Forest fire disasters are increasingly frequent events around the globe. The growing severity of fire disasters is a consequence of increased vulnerability of the natural environment. Forest fires are not only an environmental problem; as a social concern a forest fire is reflected in the social networks. People use social networks in Internet to reveal their perception and feeling. Therefore, it is possible to find pictures, videos, real-time information, NGO reports, and scientific papers in Internet that describe fires, the post-fire consequences and even events regarding vegetation recovery. Our purpose is to include this web 2.0 content to complement data coming from the "official sources" at global, national or regional level.

To demonstrate the added value of our solution to the scientific workflow we will describe a scenario and how the solution is integrated into it. We focus on assisting scientific users in the data collection step so that the user can compare the official information with Web 2.0 content to refine the output and help in the decision making. The central functioning of the W2.0B as middleware can be best illustrated with a practical example.

For demonstration purposes we have designed and developed a web client application to access the W2.0B. This client, developed with Google Web Toolkit technology, offers a user-friendly interface to facilitate users' access to the functionality of the W2.0B in a simple and visual way.

A. Forest Fire monitoring scenario

Forest fire monitoring is a complex scenario that involves many phases and procedures. To illustrate our prototype we focus on the post fire monitoring phase, i.e., once the fire has taken place, how it will be monitored to evaluate its environmental and social impact. In our scenario the chosen geographical area is the region of Ibiza, one of the Balearic Islands in Spain. In order to monitor the status of the detected fire, the user accesses the web client that provides a map viewer. This map viewer is able to visualize data coming from SDI services. In this context we add some layers to the map coming from the Data Services of EFFIS to overview some official information in the fire warning index or HotSpots as is shown in Figure 5.
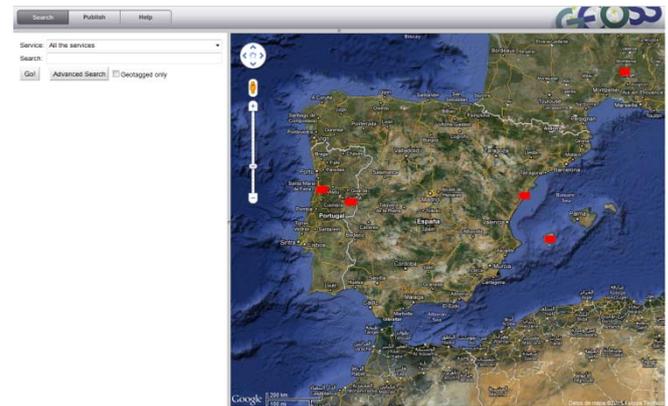

Figure 5: Screenshot of the EFFIS layer with hotspots of the last seven days.

For the first prototype, W2.0B encompasses the adapters for a selection of services: Twitter, Flickr, YouTube, OpenStreetMap, Wikipedia and Geonames. Different levels of expertise and quality can be found in crowd sourcing platforms. In our scenario different queries demonstrated that the most relevant information was retrieved from Youtube, Twitter and Flickr services.

The web client offers a simple and an advanced user interface to the user to specify the search criteria and build the query. Users can add spatial temporal criteria by setting the area of interest by selecting a rectangle in the map or by point and radius information; users can also set the time period by what the results are valid.
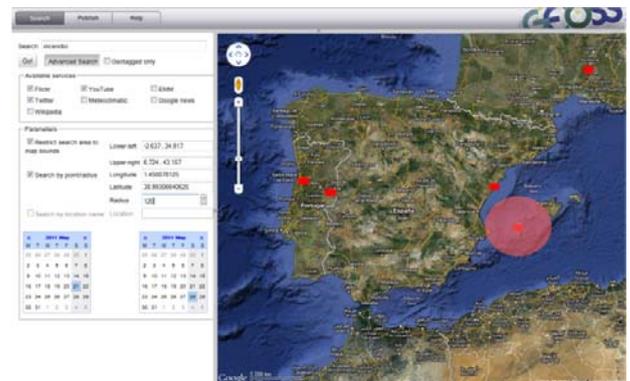

Figure 6: Screenshot of the area selected by the user.

In our scenario the user restricts the area of interest to the island of Ibiza, Spain and fills out the time constraint in order to retrieve information about fires during a one week time period beginning on May 21st.

According to the parameters available in the OS Geo and Time specification, the OS query set to the W20B is as follows:

http://elcano.dlsi.uji.es:8082/broker.jsp?service=fck,twi,ytb&q=incendio&format=kml&bbox=2.637,34.917,6.724,1.450&lon=1.450&lat=38.993&radius=120&start=2011-05-21&end=2011-05-28

When the W2.0B receives the query sent by the search client, it is broadcasted to the web 2.0 services that were specified in the query. Users are able to select different web 2.0 services at once to be queried. This increases the amount of retrieved information. In our case we specify "service=fck,twi,ytb" which means that the W2.0B queries the services Flickr, Twitter and Youtube. Figure 7 shows some of the retrieved results, which in this case are returned in KML format to be visualized in the map.
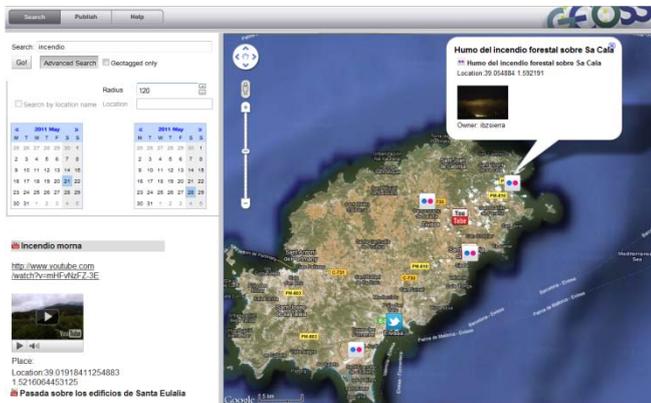

Figure 7: Screenshot of W20B results.

Figure 8 shows an example of a photograph of the fire in Ibiza. Some other retrieved results are videos about the fire from YouTube, and user comments expressing current concerns or describing other events at near real time from Twitter.

Figure 8: VGI Data extracted from Flickr through W2.0B.

VII. Discussion and Conclusions

The implemented prototype illustrates the current work on the W2.0B. Practically speaking, we have developed a mechanism to collect and aggregate VGI from multiple platforms.

To increase interoperability and scalability this mechanism offers a standard entry point to query multiple web 2.0 services to be integrated [44] and implements a mechanism to use a simple query interface to integrate a set

of Web 2.0 Services and improve data accessibility. Based on this previous research we propose a web service, the Web 2.0 Broker. The functionality of this service is to access the Web 2.0 Services functions (search interface, geographic content data type) through a unique entry point implementing a common simple query interface: OpenSearch Geo and Time extension.

OpenSearch (OS) defines a minimal interface to query a search engine that is extensible by adding extra parameters to define other filtering criteria. Such extensions include the time and geo-extension [42] [51] allowing the use of spatial and temporal filters: bounding box, circle, polygon, place name and valid period of time.

OS and its geo and time extensions are proposed as the query interface to access spatial content, both for Web 2.0 Services and SDI services. The Web 2.0 Broker is able to receive OS queries, propagate them to a set of Web 2.0 Services and return the results encoded in standard data formats such as GeoRSS, GeoJSON, KML or ATOM.

As a first assessment of the results retrieved in the previous section, using spatial and time criteria, we have performed a preliminary study that may be used as a starting point for future work. Table 2 shows some statistics of the results retrieved in three of the services. Three queries have been built and sent; the first one (identified as OS row) contains only the term criteria (fire) with the basic Opens Search parameters, the second one (OS GEO) using spatial criteria and the third one (OS GEO TEM) also adds time criteria. Regarding the location information, the retrieved results are handled by the Geocoding Manager of the W2.0B to extract the coordinates when missing, especially in the case of Twitter.

Relevance and accuracy in the content is achieved thanks to the addition of spatial and time criteria to the query. This is also reflected in Table 2.

Table 2: Findings of the W2.0B comparative study of Open Search Extension in three services.

		Content Information			Location		
		Non relevant	Related to other fire	Related to use case fire	Given by Web 2.0 service	Extracted from W2.0B	None
Flickr[1]	OS	97%	3%	0%	100%	0%	0%
	OS GEO	92%	7%	1%	100%	0%	0%
	OS GEO TEM	8.69%	0%	91,3%	100%	0%	0%
Youtube[2]	OS	34%	64%	2%	92%	6%	2%
	OS GEO	2%	96%	2%	2%	98%	0%
	OS GEO TEM	0%	1%	99%	50%	50%	0%
Twitter[3]	OS	29%	70%	1%	0%	0%	100%
	OS GEO	14%	53%	33%	3%	83%	14%
	OS GEO TEM	12%	30%	54%	3%	93%	4%
[1] 100 results [2] 50 results [3] 100 results							

The Web 2.0 Broker aggregates Web 2.0 adapters that translate an original OS query to the concrete syntax of each Web 2.0 Service API. This approach implies the development of OS adapters for each Web 2.0 Service instead of using the proprietary search tools of each Web 2.0 resources. Its advantage is that potential calibration

mechanisms can be encapsulated in well-defined components, which directly connect and use the specific Web 2.0 service APIs. However, the search criteria based on the OS needs to be mapped into the specific Web 2.0 service API and this means that we could lose accuracy in certain parts of a query. This may have an impact on the numbers of VGI items retrieved and therefore further research has to be performed.

The wild fire examples illustrate the complexity of VGI data integration where VGI publication patterns differ depending on the considered types of phenomena and the associated geospatial, temporal and event criteria.

The potential of the massive VGI availability monitoring, which is the state of the environment to validate global models with local knowledge, has yet to be fully exploited.

Our work indicates that VGI can be complementary and can provide high-scale value-added information at low cost. Furthermore, this approach could be used to enrich crisis management models inputs or to refine their output results. Further developments of the Web 2.0 Broker are ongoing to improve the means to harness VGI to integrate it into GII and to leverage its full potential.

REFERENCES

[1] M. Craglia, M.F. Goodchild, A. Annoni, G. Câmara, M. Gould, W. Kuhn, D.M. Mark, I. Masser, D.J. Maguire, S. Liang, E. Parsons. "Next generation Digital Earth". A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. International Journal of Spatial Data Infrastructure Research, vol. 3, 2008, pp. 146–167.

[2] M.F. Goodchild. "Geographic information science: the grand challenges". In J.P. Wilson and A.S. Fotheringham, editors, The Handbook of Geographic Information Science. Malden, MA: Blackwell, 2008, pp. 596–608

[3] M.K. Ramamurthy, "A new generation of cyberinfrastructure and data services for earth science education and research". Advances in Geosciences, pp. 69-78, June 6, 2006.

[4] I. Masser. GIS Worlds: Creating Spatial Data Infrastructures. Redlands: ESRI Press. 2005.

[5] 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, Volume 50, 25 April 2007.

[6] D.J. Coleman, P.Y. Georgiadou, and J. Labonte. "Volunteered geographic information: the nature and motivation of produsers". In: International journal of spatial data infrastructures research : IJSDIR, vol. 4, 2009, pp. 332-358.

[7] L., Díaz, C., Granell, M., Gould, J., Huerta. Managing user generated information in geospatial cyberinfrastructures. Future Generation Computer Systems, vol. 27, no.3, pp. 304-314, 2011.

[8] S. Schade, and M. Craglia. "A Future Sensor Web for the Environment in Europe". EnviroInfo Conference, Berlin, Germany, 2010.

[9] M. Papazoglou, W-J. Van den Heuvel. "Service oriented architectures: approaches, technologies and research issues". The VLDB Journal, vol. 16, issue 3, pp. 389- 415, 2007.

[10] W.M.P. Aalst, M. van der Beisiegel, K.M. van der Hee, D. Konig, C. Stahl. "An SOA-based architecture framework". International Journal of Business Process Integration and Management, vol. 2, issue 2, pp. 91-101, 2007.

[11] N. Alameh. "Chaining Geographic Information Web Services". IEEE Internet Computing, vol. 7, issue 5, pp. 22-29, 2003.

[12] J. Mykkänen, and M. Tuomainen. "An evaluation and selection framework for interoperability standards". Information and Software Technology, vol 50, issue 3, pp. 176-197, 2008.

[13] N.R. Budhathoki, B. Bertram, Z. Nedovic-Budic. "Reconceptualizing the role of the user of spatial data infrastructure". GeoJournal, vol, 72, pp. 149-160, 2008.

[14] D.J., Coleman. Volunteered Geographic Information in Spatial Data Infrastructure: An Early Look At Opportnities And Constraints. GSDI 12 World Conference. Singapore. October 2010

[15] A. Turner. Introduction to Neogeography (O'Reilly Short Cuts series). O'Reilly Media. 2006

[16] T. O'Reilly. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. 2005

[17] M.F Goodchild, M.F. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. International Journal of Spatial Data Infrastructures Research 2: 24–32. (437). 2007

[18] M.F Goodchild. NeoGeography and the nature of geographic expertise. Journal of Location Based Services, Vol. 3, No. 2, p. 82-96. 2009

[19] M. Bishr and W. Kuhn. Geospatial information bottom-up: A matter of trust and semantics. In S. Fabrikant & M. Wachowicz (Eds.), The European information society: Leading the way with geo-information (pp. 365–387). Berlin: Springer. 2007.

[20] L. Díaz and S. Schade. GEOSS Service Factory: Assisted Publication of Geospatial Content. Accepted to the 14th AGILE International Conference on Geographic Information Science (AGILE 2011). Utrecht, The Netherlands, April 2011

[21] Tulloch, D.L. (2007). Many many maps: Empowerment and online participatory mapping. First Monday.

[22] Coote, A. and Rackham, L. (2008), "Neogeographic Data Quality – Is it an issue?" Paper presented at the 2008 Annual Conference of the Association for Geographic Information (AGI).

[23] Harvey, F. (2003), "Developing geographic information infrastructures for local government: the role of trust". The Canadian Geographer. Vol. 47, No. 1, pp. 28-36

[24] Elwood, S. (2008), "Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS". GeoJournal, vol. 72, no. 3-4, pp. 173-183.

[25] Flanagin, A.J. & Metzger, M.J. (2008), "The credibility of volunteered geographic information", GeoJournal, vol. 72, no. 3-4, pp. 137-148.

[26] S. Shade, G. Luraschi, B. De Longueville, S. Cox, Laura Díaz. Citizens as sensors for forest fires: Sensor Web Enablement for Volunteered Geographic Information. M.A. Brovelli, S. Dragicevic, S. Li, B. Veenendaal (Eds): ISPRS Workshop on Pervasive Web

Mapping, Geoprocessing and Services XXXVIII-4/W13 (WebMGS 2010). Como, Italy, August 2010, ISSN 1682-1777

[27] Jankowski, P. 2009. Towards participatory geographic information systems for community-based environmental decision making. Journal of Environmental Management 90. 1966–1971

[28] Craglia, M. (2007). Volunteered Geographic Information and Spatial Data Infrastructures: when docparallel lines converge? Specialist Meeting on Volunteered Geographic Information. 13-14 December 2007. Santa Barbara

[29] Gould, M. (2007). Vertically interoperable geo-infrastructures and scalability. Specialist Meeting on Volunteered Geographic Information. 13-14 December 2007. Santa Barbara

[30] Belimpasakis P, Saaranen A (2010) Sharing with people: a system for user-centric content sharing. Multimed Syst 16:399-4216[

[31] Boll S (2007) MultiTube–Where Multimedia and Web 2.0 Could Meet. IEEE Multimed 14(1): 9-13

[32] Nature Editorial (2008) A place for everything. Nature 453(2):2

[33] Tsai FS (2011) Web-based geographic search engine for location-aware search in Singapore. Expert Systems with Applications 38:1011-1016

[34] Nuñez M, Díaz L, Granell C, Huerta J (2011) Web 2.0 Broker: a tool for massive collection of user information. Accepted to European Geosciences Union (EGU) General Assembly 2011 (EGU 2011). Vienna, Austria, Apr 2011

[35] Walsh J (2007) On Spatial Data Search. Terradue White Paper. http://www.terradue.com/images/T2-Research-07-003-OnSearch.pdf

[36] Nogueras-Iso J, Zarazaga-Soria FJ, Béjar R, Álvarez PJ, Muro-Medrano PR (2005) OGC Catalog Services: a key element for the development of Spatial Data Infrastructures. Computers & Geosciences, 31(2): 199-209

[37] Naaman M (in press) Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. Multimed Tools Appl. DOI 10.1007s/11042-010-0538-7

[38] Clinton, D. (2010) OpenSearch 1.1 Draft 4 specification. http://www.opensearch.org

[39] GeoRSS (2011) http://www.georss.org/Main_Page

[40] Nottingham M, Syare R (2005) The Atom Syndication Format. RFC 4287, http://tools.ietf.org/html/rfc4287

[41] OGC KML (2008) OpenGIS Keyhole Markup Language (KML) Implementation Specification, Version 2.2.0. Open Geospatial Consortium Inc (Open GIS Consortium Inc). http://www.opengeospatial.org/standards/kml

[42] Turner, A (2010) The OpenSearch Geo extension (draft 2)

[43] Abargues C, Beltrán A, Granell C (2010) MIMEXT: a KML extension for georeferencing and easy share MIME type resources. In: Geospatial Thinking. Springer, pp. 315-334.

[44] Fonts O., Huerta, J. Díaz, L, Granell C. (2009).OpenSearch-geo: The simple standard for geographic search engines. Proceedings IV Jornadas SIG Libre

[45] Flickr.com (2011) http://www.flickr.com

[46] Geonames (2011) http://www.geonames.org

[47] Open Street Maps (2011) http://www.openstreetmaps.org

[48] Twitter (2011) http://twitter.com

[49] Wikipedia (2011) http://www.wikipedia.org

[50] YouTube (2011) http://youtube.com

[51] Gonçalves (2010) OpenSearch Geo –Time extension. OpenGIS.

[52] http://effis.jrc.ec.europa.eu/data

[53] Díaz L, Martín C, Gould M, Granell C, Manso M.A,(2007) Semi-automatic Metadata Extraction from Imagery and Cartographic data, International Geoscience and Remote Sensing Symposium (IGARSS 2007). Barcelona,. IEEE CS Press, pp. 3051-3052.

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

[55] Nogueras-Iso, J. Zarazaga-Soria, F.J. Béjar, R. Álvarez, P.J. . Muro-Medrano, P.R. . (2005) OGC Catalog Services: a Key element for the development of Spatial Data Infrastructures, Computers and Geosciences, vol. 31/2, 199-209.

[56] Meteoclimatic: http://www.meteoclimatic.com/

[57] Carrera, F. Ferreira, J. (2007) The Future of Spatial Data Infrastructures: Capacitybuilding for the Emergence of Municipal SDI. International Journal of Spatial Data Infrastructures Research, vol. 2,49-68

[58] Jankowski, P. (2009) Towards participatory geographic information systems for community-based environmental decision making. Journal of environmental management, vol 90, 1966-1971

[59] Dienel, P. (1989) Contributing to social decision methodology: citizen reports on technological projects. Social Decision Methodology for Technological Projects (Theory and Decision Library, Series A), Kluwer Academic Publishers.

[60] Williamson, I.(2003) SDIs—setting the scene. Developing spatial data infrastructures: From concept to reality.CRC Press. (pp. 3–16).

[61] Masser, I. (2005). GIS worlds: Creating spatial data infrastructures (First ed.). ESRI Press.

[62] Budhathoki, N.R. Nedovic-Budic, Z.(2007). Expanding the SDI knowledge base. In H. Onsrud (Ed.), Research and theory in advancing spatial data infrastructure (pp. 7–32).

[63] Fonseca, C.R., Ganade, G., Baldissera, R., Becker, C.G., Boelter, C.R., Brescovit, A.D., Campos, L.M., Fleck, T., Fonseca, V.S., Hartz, S.M., Joner, F., Käffer, M.I., Leal-Zanchet, A.M., Marcelli, M.P., Mesquita, A.S., Mondin, C.A., Paz, C.P., Petry, M.V., Piovezan, F.N., Putzke, J., Stranz, A., Vergara, M., Vieira, E.M., (2009). Towards an ecologically sustainable forestry in the Atlantic Forest. Biological Conservation vol 142, 1144–1154.