Future Internet enablers for VGI applications

Denis Havlik[1], Javier Soriano[2], Carlos Granell[3], Stuart E. Middleton[4]; Hylke van der Schaaf[5], Arne J. Berre[6], Jasmin Pielorz[7]

Abstract

This paper presents the authors experiences with the development of mobile Volunteered Geographic Information (VGI) applications in the context of the ENVIROFI project and Future Internet Public Private Partnership (FI-PPP) FP7 research programme.

FI-PPP has an ambitious goal of developing a set of Generic FI Enablers (GEs) - software and hardware tools that will simplify development of thematic future internet applications. Our role in the programme was to provide requirements and assess the usability of the GEs from the point of view of the environmental usage area, In addition, we specified and developed three proof of concept implementations of environmental FI applications, and a set of specific environmental enablers (SEs) complementing the functionality offered by GEs. Rather than trying to rebuild the whole infrastructure of the Environmental Information Space (EIS), we concentrated on two aspects: (1) how to assure the existing and future EIS services and applications can be integrated and reused in FI context; and (2) how to profit from the GEs in future environmental applications.

This paper concentrates on the GEs and SEs which were used in two of the ENVIROFI pilots which are representative for the emerging class of Volunteered Geographic Information (VGI) use-cases: one of them is pertinent to biodiversity and another to influence of weather and airborne pollution on users' wellbeing. In VGI applications, the EIS and SensorWeb overlap with the Social web and potentially huge amounts of information from mobile citizens needs to be assessed and fused with the observations from official sources. On the whole, the authors are confident that the FI-PPP programme will greatly influence the EIS, but the paper also warns of the shortcomings in the current GE implementations and provides recommendations for further developments.

1. Motivation

The designers, developers and owners of the legacy services and applications of the Environmental Information Space (EIS) face the challenge of a technology and paradigm shift caused by the emerging Future Internet (FI). Examples of the trends which are incompatible with the traditional EIS systems, security and business models include: (i) "everything as a service" approach backed up by the underlying cloud services; (ii) blurring of the differences between data users and data providers backed up by Volunteered Geographic Information (VGI) applications and the rise of the SensorWeb and Internet of Things (IoT); (iii) the power shift resulting from the ability of citizens to analyse the available information, visualize the re-

[1] AIT Austrian Institute of Technology GmbH., Donau-City Strasse 1, A-1220 Wien Email: denis.havlik@ait.ac.at, Internet: http://www.ait.ac.at

[2] Universidad Politécnica de Madrid, Madrid, Spain. Email: jsoriano@fi.upm.es, Internet: http://conwet.fi.upm.es/

[3] European Commission, Joint Research Center, I-21027 Ispra, Italy. Email: carlos.granell@jrc.ec.europa.eu, Internet: http://ec.europa.eu/dgs/jrc/

[4] University of Southampton IT Innovation Centre, Gamma House, Enterprise Road, Southampton SO16 7NS, UK Email:sem@it-innovation.soton.ac.uk, Internet: http://www.it-innovation.soton.ac.uk/

[5] Fraunhofer IOSB, Karlsruhe, Germany. Email: hylke.vanderschaaf@iosb.fraunhofer.de , Internet: http://www.iosb.fraunhofer.de

[6] SINTEF, Oslo, Norway. Email: arne.j.berre@sintef.no, Internet: http://www.sintef.no

[7] UBIMET GmbH, Dresdner Straße 82, A-1200 Wien. Email: jpielorz@ubimet.com, Internet: http:///www.ubimet.com

sults in various ways (including in particular interactive maps), and develop their own applications in ad-hoc manner from the high level building blocks that are freely available online.

The Future Internet Public Private Partnership FP7 Programme (FI-PPP) has set a stage allowing the ICT and thematic experts to interact, co-develop and experiment with the emerging European FI platform and the thematic applications using this platform. The core platform (FI-Ware) is developed by the project of the same name. The ENVIROFI project participated in the programme as a representative of the environmental usage area. The overall approach taken by FI-PPP, the FI-Ware architecture and the ENVIROFI approach for linking the FI-Ware GEs and the applications and services of the EIS have been explained in (Havlik et al. 2011) and (Usländer et al. 2013) and will not be discussed in detail in this paper. Instead, we shall concentrate on concrete experiences with FI application developments pertinent to two proof-of-concept VGI applications developed within ENVIROFI.

Section 2 discusses the functionalities and challenges common to many mobile VGI applications. Section 3 presents the functional architectures of the two pilots and discusses the roles of key functional building blocks. Finally, section 4 presents the paper conclusions and the outlook for future developments.

2. Functionalities and challenges of mobile VGI applications

Three main Use Cases (UCs) pertinent to mobile VGI applications are presented in Figure 1 below:

Figure 1
Three main use cases pertinent to mobile VGI applications

1. The "View existing knowledge" UC refers to passive use of information provided by an application, for improved situation awareness. In mobile applications the main area of interest is often the users' vicinity (local situation awareness).
2. The "Report observation" UC makes a difference between a simple "viewer" and the participative sensing functionality which is inherent to VGI applications. The users and their sensors can submit new objects of interest (OOIs) as well as new observations on existing OOIs.
3. The "Receive information/events" UC has a dual use. First, the application can inform the user about interesting or dangerous objects and events in his/her vicinity. Second, the application can motivate the users by explicitly asking for observations (crowd-tasking).

Each of these use cases presents various technical, organizational and ethical challenges. In particular, VGI applications have to handle issues related to: (1) an inaccurate, conflicting and incomplete observation heap; (2) unreliable networks; (3) users' motivation; and (4) users' privacy. Detailed discussion on our approaches for answering these challenges can be found in Havlik et al. (2013).

In addition to challenges that are specific to VGI applications, ENVIROFI also considered the more general challenges of efficient application development, scaling and delegation of administrative responsibilities that are inherent to FI technology.

3. FI enablers in ENVIROFI pilot applications

In order to test the enablers and services developed within the ENVIROFI and FI-Ware projects, we applied them in three pilot uses cases from the biodiversity, atmospheric and marine domain. Two of these use VGI functionalities: "Bringing Biodiversity into the Future Internet" (BIO) and "Personal Information System for Air Pollutants, Allergens and meteorological condition" (PEIS). The main goals and implementation architectures of these two applications are discussed in section 3.1, followed by presentation of the generic and specific enablers used in these applications.

3.1 Implementation architectures of the ENVIROFI VGI pilots

The BIO pilot focuses on the use of FI technology for the survey, analysis, quality assurance, persistence and dissemination of biodiversity data. The Biodiversity Survey Application (BIO App) enables users from a wide variety of backgrounds to (1) access data on interesting biodiversity occurrences at their Areas of Interest (AOIs), (2) provide data on biodiversity occurrences they are currently observing, and (3) provide feedback on the correctness of existing biodiversity occurrence records in the system. Users can either add information to existing entries on individual plants, or provide information on specimen currently not contained within the system. BIO App can operate in any geographical region. It has been tested with large sample data sets based on open data tree layers in the cities of Vienna and Florence.

The PEIS pilot aims at enhancing the users' interaction with their environment and their understanding of the individual's exposure to air pollutants and allergens as well as meteorological conditions. There are three main foci in this pilot: (1) the provision of air quality and meteorological information from nearby stations, (2) the provision of warning notifications in case an air quality index is above a certain threshold and, (3) the possibility to provide own environmental reports that can be stored and ultimately shared with other users. Current prototype offers air quality data for NOx, PM10, PM2.5 and O3 in Oslo and meteorological parameters, i.e. temperature, pressure and wind layers available in Karlsruhe and Vienna regions.

The implementation architectures of the two pilots are similar and related to the ENVIROFI architecture as shown in figure 2. The parts of the generic architecture and services that are in particular focused in the VGI applications are marked in red. The ENVIROFI architecture is based on application support through three levels of enablers. Environmental Specific Enablers (SEs) that provide environmental specific services, Geospatial Specific Enablers based on standard OGC geospatial services and FI-WARE Generic enablers (GEs) from the FI-Ware.

The Environmental SEs are divided into: VGI, Semantic TAGging, MEDiation, Fusion and Notification SE groups. Fusion services (section 3.4) are capable of interpolating and extrapolating the data. Intelligent analytical services (section 3.3) support assessment and quality of observations.

The Geospatial SEs contain standard geospatial services from the OGC standard set of services, such as WFS, WMS, WPS, SOS, SPS, and also data access management through OpeNDAP.

The GEs are divided into 6 groups ("chapters"): (1) Cloud Hosting GEs provide generic enablers for VM allocation, object storage and cloud proxies; (2) Data/Context Management GEs provide support for data processing , event processing, event brokering, big data analytics and semantic annotation; (3) App Services GEs for building of service repositories, marketplaces and mashup applications; (4) Internet of Things (IoT) GEs for handling of sensors and actuators; (5) Interfaces to Network and Devices (I2ND) for network and device management; and (6) Security for identity management, security monitoring and policy enforcement.

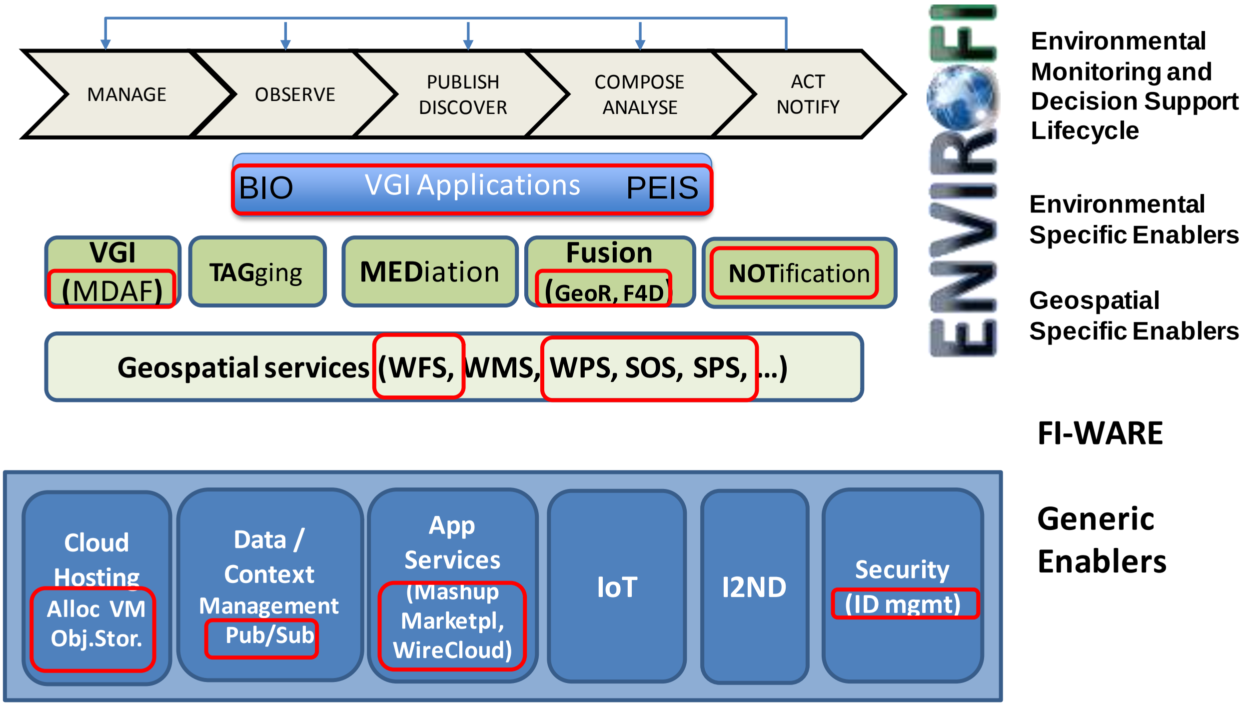

Figure 2
ENVIROFI architecture related to the VGI applications

Further information on ENVIROFI pilots, as well as on the SEs and GEs which were used to build these pilots can be found at http://catalogue.envirofi.eu. Full list of currently available FI-Ware GEs can be found at http://catalogue.fi-ware.eu. Following sections introduce selected SEs and GEs that are particularly interesting for the VGI applications

3.2 Mobile Data Acquisition Framework

The Mobile Data Acquisition Framework (MDAF) consists of three SEs which work together to provide the central functionality required by VGI applications:

1. The Environmental Georeferenced Observation App SE is a generic and easily extensible mobile App providing the view/report/alert functionality mentioned in section 2.
2. The Environmental Georeferenced Observation Service SE handles the observation heap generated by users and automated services. It generates the application specific views on the data, notifies the users of new OOIs, observations and requests for additional actions and coordinates the activities and data flows among backend services.

3. The Environmental Georeferenced Observation Proxy Service SE can replicate a subset of data relevant for the user[8] to a remote database and implements the same service interfaces as the Observation service SE. The proxy is used to improve the reliability of the service over unreliable networks.

Much of the MDAF functionality is inherited from the CouchDB database (http://couchdb.apache.org/). This document-oriented database offers easily extendible RESTful API, natively supports GeoJSON and implements a sophisticated multi-master replication mechanism assuring the eventual consistency of the data. MDAFs' data model is similar to OGC O&M, but the service interface is RESTful and the data is encoded as GeoJSON rather than in XML. Feasibility of encoding this data in O&M and accessing it over OGC Sensor Observation Service has been demonstrated in the PEIS pilot.

As illustrated in the figure 3, MDAF also features an internal event handler and processing modules. These modules have one major advantage over external services: they can easily access the users' information which is not disclosed to the rest of the world to protect users' privacy. This is mainly used to decide which information should be pushed to MDAF users, depending on their previously declared AOIs

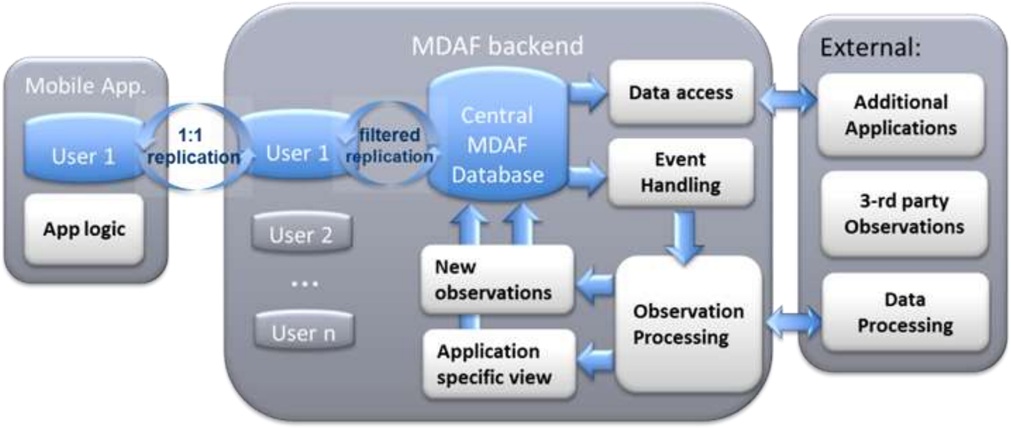

Figure 3
Use of MDAF in VGI applications

We are currently implementing additional client-side logic which will assure the user is reminded of this information whenever he or she is in the vicinity of the OOI referred in the observation and therefore either likely to be affected by it (alerts) or in a position to assess the validity of the information (crowd-tasking). Thanks to pre-filtering based on users' AOIs, MDAF can provide this functionality in a robust and efficient way, while at the same time protecting the users' privacy.

3.3 Assessing the data meaning and quality of the information

The meaning and quality of acquired VGI can vary considerably depending on the area, species and data provider. VGI authors vary in terms of their domain expertise, interpretation of domain concepts, ability to provide useful metadata and their commitment / ability to provide high quality observation data. Assessing and improving of the data quality in ENVIROFI VGI applications relies on a generalization of the consensus building principle shown in figure 4.

[8] The "data relevant for the user" is defined as a subset of the data pertinent to user-defined Areas of Interest (AOIs).

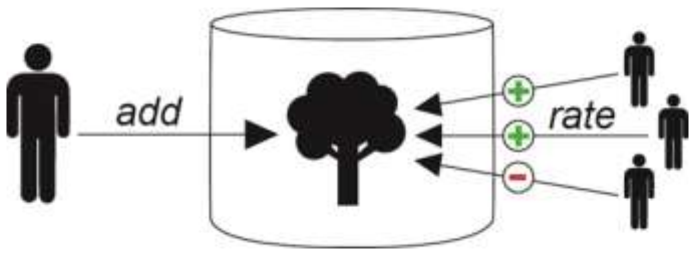

Figure 4
Consensus building in VGI applications

Whenever a new observation is made, other users interested in the area are notified and asked for confirmation. The observation is considered confirmed if a majority of users agree, but the exact algorithm is considered application-specific and can be tuned over time.

These "other users" can also be automated processes. Examples of such processes are the Leaf Image Classification SE and the Image Quality Assessment SE in the BIO pilot. Each observation of the "leaf image" type is automatically checked for image quality (e.g. too much blur, bad colour saturation etc.) and then classified against a library of known leaf images. Automatic labelling can help human experts in reviewing of the observation collections, e.g. by identifying the poor images and unusual outliers. Figure 5 illustrates the integration of the Image Classification SE, Quality Assessment SE and Image Archive SE in ENVIROFI VGI applications.

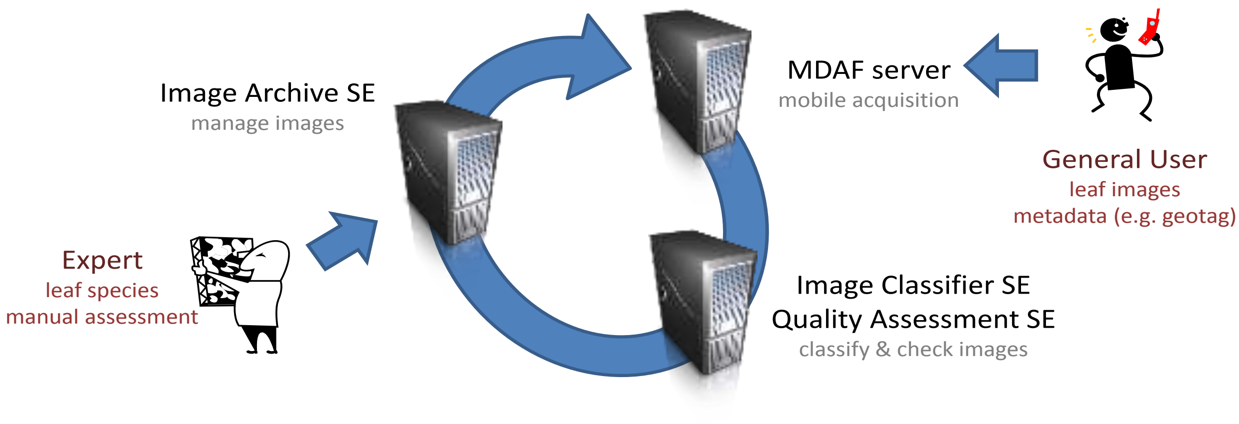

Figure 5.
Use of Image Classification SE in VGI applications

Our image classifier prototype has a training set of 54 tree leaf species, each with about 30 leaf images making a total of about 1600 expert labelled leaf images. The image archive service allows users to do leaf image and metadata management. The leaf metadata labels are mapped to standard biodiversity domain ontology concepts, such as the TaxMeOn ontology, and the user interface includes multi-lingual support (English and Italian at present) to help end users navigate leaf images in their own language using common species names as opposed to just scientific species names (i.e. Latin names). Login to the image archive is performed using either an OpenID or the FI-WARE Identity Management GE and the image archive can be queried using the FI-WARE Media Enhanced Query Broker GE.

3.4 Filling the data gaps

The available information on the environment is inevitably incomplete. Filling the gaps in the data that is inherently discontinuous such as the missing information on three positions, heights, species, etc. is to a certain extent feasible[9], but the achievable quality of the results is not on pair with requirements of the biodiversity usage area. The situation is completely different when it comes to continuous phenomena such as temperature, pressure or concentration of the particular matter in the air. In this case, the VGI users are often interested in values at their positions, rather than at the positions of the existing sensors. The required information can be estimated in a variety of ways, ranging from a simple linear interpolation of the values from the nearest available observations to complex models taking into account the meteorological conditions like wind direction and rainfall as well as the positions of buildings, roads and known pollution sources in the vicinity.

These models are often written in specialised modelling languages, like MATLAB or R, and are usually not developed with a service-oriented architecture in mind. The Fusion4Desicion (F4D) SE (van der Schaaf et al. 2013) provides a convenient way to access these models using the standardised OGC services Sensor Observation Service (SOS) and Sensor Planning Service (SPS). SPS is used to specify parameters start and stop model runs and monitor the progress, and the results are made available using SOS. In this framework the model is viewed as a sensor and the model results are viewed as observations. This allows us to run the integrated models on demand, for the region the user is interested in, and also to feed the models with the data from different providers distributed over several third party servers. Within ENVIROFI, the Fusion4Decision SE was used for interpolation of the air quality data in PEIS application.

The F4D system allows domain experts to develop their models in the modelling environment they are familiar with, while still having the option of making those models available as a web service with a standardised interface.

3.5 Extending the VGI application functionality with web mashups

MDAF framework covers the core functionality required in all VGI applications, but the customization for specific applications cannot be easily done by end-users. In addition, the framework does not offer a web- or desktop-GUI. This task is delegated to third-party applications and web mashup platforms.

Web mashup applications integrate heterogeneous data, application logic (exposed as services in general), and UI components (a.k.a. widgets) sourced from the Web to create new, value-adding composite applications (Yu et al. 2008, Gebhardt et al. 2012). They typically serve a specific situational (i.e. immediate, short-lived, customized) need, albeit with high potential for reuse. This "situational" character precludes them from being offered as 'off-the-self' functionality by solution providers: their high potential for reuse and the fact that they target end users (as prosumers) put web mashup applications and their constituent mashable application components in a unique position to enable the crowd-sourcing of Web applications and to leverage the "long tail" (Anderson 2006) of web services/APIs.

Web mashup tools and platforms allow end users (business staff, customers or citizens with no programming skills) to build ad-hoc applications from existing web widgets. They help to leverage innovation through experimentation and rapid prototyping by allowing their users (i) to discover the best suited mashable components (widgets, operators and prefab mashup-lets[10]) for their devised mashup from a vast, ever-growing distributed catalogue, (ii) to visually mash them up to compose the application, and (iii) to

[9] For example, we could combine the information extracted from aerial photographs could be with the known positions, sizes and species of the trees which were previously reported in the area of interest for a good first estimate on the positions sizes and species of the trees which were not reported so far.

[10] Mashups, once they are built, can be published in a catalogue and used as building blocks to create other mashups.

share them with other users. Wirecloud (http://conwet.fi.upm.es/wirecloud/) represents a prominent example of such a Web mashup platform that builds on cutting-edge end-user development and semantic technologies to offer a next-generation end-user centred mashup platform. Wirecloud is the reference implementation of the FI-WARE application mashup GE, which, in turn, is based on a composition model for end-user development described in (Lizcano et al. 2011).

Figure 4 shows a screenshot of a joint ENVIROFI/FI-WARE web mashup application demo. The demo complements the functionality of the ENVIROFI Bio app and provides the users with an alternative web GUI for assessing the quality of the observations?

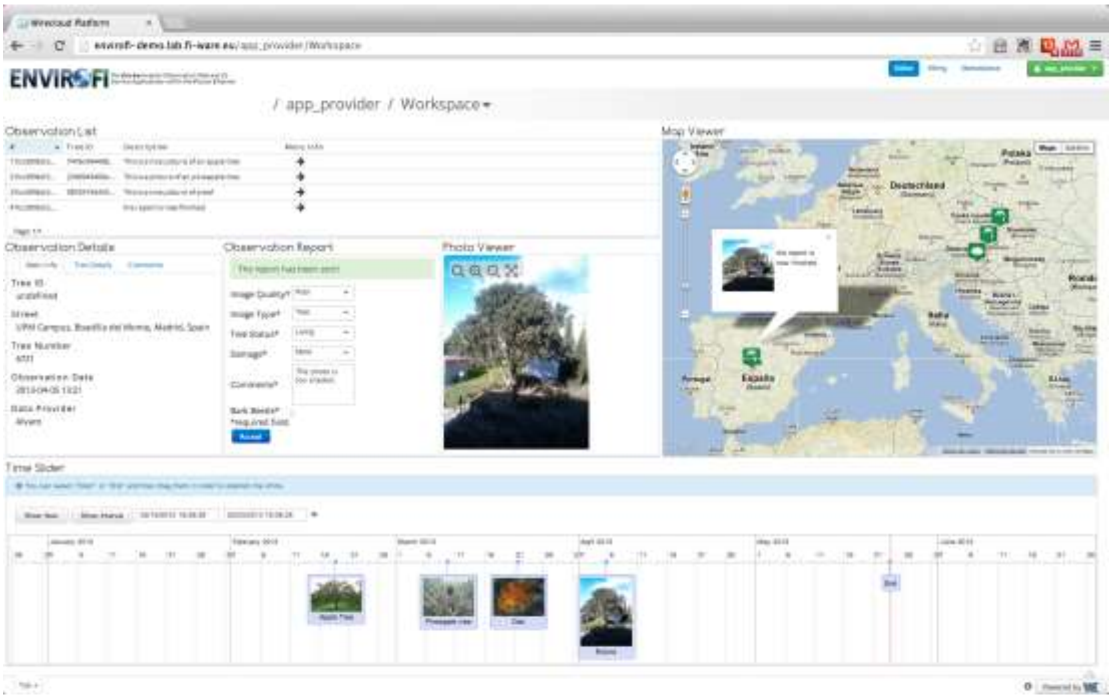

Figure 4
Web mashup extension of the ENVIROFI BIO application

In addition to WireCloud, this demo uses several other GEs: the combination of the Marketplace, Registry and Repository GEs assure the widgets and mashup applications can be found; the Pub/Sub Context Broker GE regulates the exchange of observations between MDAF and the WireCloud; the Object Storage GE is used to exchange large binary objects (photos); and the Identity Management GE to federate the user accounts between MDAF, Object Storage GE and the Wirecloud.

The results are quite impressive concerning the modest effort required for the development. The development of the new ENVIROFI-specific widgets took less than 0,5 man-month in effort, whereas the more complex and elaborated widgets such as the Map Viewer, the Photo Viewer and the Time Slider were taken from the Wirecloud catalogue and slightly customized.

4. Outlook and conclusions

As one of the “Phase 1” FI-PPP projects, the main task of the ENVIROFI project was to articulate the requirements on FI-WARE GEs, and illustrate how the generic FI enablers and specific environmental enablers can be combined in environmentally enabled FI applications. Our architectural analysis which was reported in (Usländer et al. 2013), as well as the experiences with development of the prototype applications show that such applications are indeed feasible and able to profit from the generic functions, scalability and interoperability of the cloud-enabled FI-WARE framework. However, we also experienced some constraints.

Main limitation of FI-WARE GEs, from the point of view of the environmental usage area, is the lack of explicit support for geospatial data and processing. While such support can be easily introduced through SEs or through extensions of the data processing capabilities of the relevant GEs, the potential users should be aware of this limitation and design the applications accordingly.

In addition, the combination of the low maturity of the FI-WARE GEs which were available at a time we performed most of the testing (Q4 2012 and Q1 2013) and the limited time and resources at our disposal has prevented us from integrating some very promising generic enablers in our applications. This includes the various data/context processing GEs other than pub/sub broker and query broker; all the GEs from the IoT and I2ND chapter; as well as the GEs from security chapter which eliminate the need for application-specific policy enforcement and secure data storage. Finally the Marketplace-related GEs could become a valuable channel for advertising and selling the environmental data and services in cross-domain business environments.

5. Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements n◦ 284898 (ENVIROFI) and 285248 (FI-WARE).

Bibliography

Anderson, C. (2006): The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion

Gebhardt, H. et al. (2012): From Mashups to Telco Mashups: A Survey, in IEEE Internet Computing Vol. 16, Nr. 3, pp. 70-76

Havlik, D., Schade, S., Sabeur, Z. A., Mazzetti, P., Watson, K., Berre, A. J., Lorenzo. J. M. (2011): From Sensor to Observation Web with Environmental Enablers in the Future Internet, in: Sensors Nr. 11, pp. 3874-3907, doi:10.3390/s110403874

Havlik, D., Egly, M., Huber, H., Kutschera, P., Falgenhauer, M., Cizek, M. (2013): Robust and trusted crowd-sourcing and crowd-tasking in the Future Internet, to be published in Proceedings of ISESs 2013

Lizcano, D., Alonso, F., Soriano, J., and López, G. (2011): A new end user composition model to empower knowledge workers to develop rich Internet applications, in: Journal of Web Engineering Vol. 10, Nr. 3, pp. 197-233

Usländer, T., Berre, A., Granell, C., Havlik, D., Lorenzo, J. M., Sabeur, Z. A., Modafferi S. (2013): The Future Internet Enablement of the Environment Information Space, to be published in proceedings of ISESs 2013

Van der Schaaf, H., Kobernus, M., Falgenhauer, M., Pielorz, J., Watson, K. (2013): Data fusion in the environmental domain, to be published in Proceedings of ISESs 2013

Yu, J. et al. (2008): Understanding Mashup Development, in: IEEE Internet Computing Vol. 12, Nr. 5, pp. 44–52