Citation
Tanasescu, Vlad; Gugliotta, Alessio; Domingue, John; Gutiérrez Villarías, Leticia; Davies, Rob; Rowlatt, Mary; Richardson, Marc and Stini, Sandra (2007). Geospatial data integration with Semantic Web services: the eMerges approach. In: Scharl, Arno and Tochtermann, Klaus eds. The Geospatial Web: how geobrowsers, social software and the Web 2.0 are shaping the network society. Advanced Information & Knowledge Processing. London, UK: Springer, pp. 247–256.

URL
https://oro.open.ac.uk/23125/

License
None Specified

Policy
This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding
Geospatial Data Integration with Semantic Web Services: the eMerges Approach


Abstract. Geographic Space still lacks the semantics allowing a unified view of spatial data. Indeed, as a unique but all encompassing domain, it presents specificities that Geospatial Applications are still unable to handle. Moreover, to be useful, new spatial applications need to match to human cognitive abilities of spatial representation and reasoning. In this context, eMerges, an approach to geospatial data integration based on Semantic Web Services (SWS), allows the unified representation and manipulation of heterogeneous spatial data sources. eMerges provides this integration by mediating legacy spatial data sources to high level spatial ontologies through SWS and by presenting for each object context dependent affordances. This generic approach is applied here in the context of an emergency management use case developed in collaboration with emergency planners of public agencies.

1 Introduction

Web2.0 applications, by offering large amounts of resources to users for small fees, by weaving social networks where only forests of text based hyperlinks existed, and providing desktop like applications to the browser, are changing the way we interact on the Web. Part of this evolution is a renewal of the available mapping applications; closed, static and symbolic traditional web map applications are progressively replaced by web2.0 maps employing new means to achieve a map reality effect, which is the ongoing effort of rooting the maps into the cognitive reality by giving more natural looking insights into the geography covered by it. Also, by freely distributing APIs, new web2.0 maps lead to an explosion of mashups, minimal applications developed by independent technically skilled users which aggregate data in a spatial context in order to fulfill a specific goal.

The popularity of web2.0 maps and mashup applications\(^1\) shows the interest and the appeal of the geographic environment for web users; mashups are used for such a wide variety of goals, that it seems that space, mediated through realistic web maps, may provide the terrain for data integration rooted into human cognition that the more abstract textual web has not yet succeeded to achieve.

However mashups, as isolated attempts at data integration, do not have to cope with the semantic complexity of multiple heterogeneous data sources; usually the service providing the data is integrated by the developer as a single and isolated map layer, making the related semantics clear.

However, to allow large scale integration, semantic descriptions are needed (Egenhofer 2002). Semantic Web Services (SWS) are the result of an acknowl-
edgement that Web Service technology (WS), even in its standardized form, cannot achieve a satisfying level of interoperability without appropriate high-level semantics. Indeed, WS based on ad hoc REST APIs or on standards such as UDDI\(^2\) for discovery, WSDL\(^3\) for interface description, and SOAP\(^4\) for message passing, simplify the task of the developer but without dismissing his or her knowledgeable intervention. Particularly, when new services are to be integrated to an application, developers need to study the WS descriptions to match inputs, outputs and invocation workflows with the existing systems.

By using SWS, if the vision of fully automatic interaction and composition is still a research question, the following tasks are already greatly alleviated:

- **Discovery** of useful services is achieved by matching a formal task description against SWS’ semantic descriptions.
- **Mediation** between heterogeneous services can be specified at the level of data format, message protocol and business processes.
- **Composition** of services provides a means of creating a new service by aggregating existing components.

IRS-III (Cabral et al. 2006), a platform and broker for developing and executing semantic Web services, adopts a semantic Web approach based on ontological descriptions, expressed formally in OCML (Motta 1999). In particular IRS-III incorporates and extends the Web Services Modeling Ontology (WSMO) (Roman et al. 2004). **Goals**, a concept existing in WSMO to describe user’s needs as distinct from specific WS functionalities, can be invoked in this extension, which ensures a more intuitive way of interacting with clients in a Semantic Web (SW) context.

The eMerges approach applies SWS technologies to the Geospatial Web, which has been designed as an e-Government use case domain in the context of the DIP project (funded under the European Union’s IST programme FP6). eMerges illustrates the way in which spatially related data delivered through SWS can ease the management of specific use cases by aggregating data originating from different sources, and presenting it in a way which is both consistent and task relevant.

We first describe how SWS applications are build using IRS-III, giving an example of how to use SWS, then briefly present the specificities of Geographic Space as well as the eMerges generic approach to handling spatial objects in context, and finally, before concluding, discuss functionalities of the eMerges prototype implementation.

2 The IRS-III Approach to SWS Applications

Applications using IRS-III follow a layered approach (cf. Figure 1. Generic architecture used when creating IRS-III based applications) in which (micro-) functionalities of legacy systems are exposed through Web Services – based on standards or on REST – and described with ontologies. These Semantic Web Services can then be invoked from the (web) presentation layer, by using a provided API, SOAP messages, or the REST protocol.
The Web Service Modelling Ontology (WSMO) (Roman et al. 2004) is a formal ontology for describing the various aspects of services to enable the automation of WS discovery, composition, mediation and invocation. The meta-model of WSMO defines four top level elements: Ontologies, Goals, Web Services, and Mediators.

**Ontologies** (Gruber 1993) provide the foundation for describing domains semantically. They are used by the three other WSMO components. **Goals** define the tasks that a service requester expects WSs to fulfill. In this sense they tend to reflect the service user’s intent. **Web Service** descriptions represent, in terms of capabilities (what the service can do) and interface (how to use it), the behaviour of a deployed Web Service. The description also indicates how WS communicate (choreography) and how they are composed (orchestration). **Mediators** handle issues of data and process interoperability that arise between heterogeneous systems. One of the characterizing features of WSMO is that all components – Ontologies, Goals and Web Services – are linked by Mediators. In particular, WSMO provides four kinds of mediators:

- **oo-mediators** for mediating between heterogeneous ontologies;
- **ww-mediators** connect WS to WS;
- **wg-mediators** connect WS with Goals;
- **gg-mediators** link different Goals, solving input conflicts and transforming processes.

By extending WSMO’s Goal and Web Service concepts, clients of IRS-III can invoke web services via goals. That is, IRS-III supports so called capability-, or goal-driven service invocation which allows the user to use only generic inputs, hiding the possible complexity of a chain of heterogeneous WS invocations. The decoupling of the actual user vision of a task and its execution allows us to get closer to the user’s cognition of the situation and task. Mediators link goal and web services, solving existing mismatches, and allowing complex composition of services to be constructed.
The implementation use case was designed with the Essex County Council (ECC). The ECC is a large local authority in South-East England (UK). Following several interviews with spatial data holders in the ECC it was decided to focus the scenario on the ECC Emergency Planning department, and precisely, on a previous emergency situation: the snowstorm which occurred in the vicinity of Stansted airport on the 31st of January 2003. Because of the snow, drivers were trapped several hours in their cars on the M11, a motorway in the UK; as a result, access to Stansted Airport was difficult, and individuals required transport to nearby shelters, or to hospitals in some cases.

eMerges was used as the underlying conceptual framework to implement a decision support system assisting the Emergency Officer in handling the dynamics of the emergency situation and gather information related to a certain type of event, faster and with increased precision.

Data was integrated from three different sources. UK’s Meteorological Office providing snow level information, ViewEssex, a centralized database maintained by British Telecommunications (BT) managing spatial-data for the ECC, and BuddySpace, an Instant Messaging client built on top of the Jabber protocol and providing lightweight communication and collaboration means (Eisenstadt et al. 2003). Services were described by using domain ontologies which were mapped to integration ontologies. This process involved building goals and mediators to provide added value to the services for example through composition.

To illustrate such a composition we describe in the following the structure of the WSMO description associated with an example goal, Get-Polygon-GIS-data-with-Filter-Goal (cf. Figure 2. Structure of the WSMO description of the eMerges prototype. To avoid cluttering the diagram, wgM and Web Services balloons were omitted). This goal describes the request of a class of shelter (hospital, inn, hotel, etc.) in a delimited query area. The user selects a class of shelter while the polygon query...
area is provided by the context. However, the only WS available returns a specific class of shelter in a circular query area. Moreover, results also have to be filtered in order to return only shelters relevant to the task (in our case, the management of a snowstorm emergency). Therefore problems for invocation are: (1) selection of the adequate WS; (2) mediation of the different area representations (polygon vs. circular); (3) orchestration of the retrieve and filter data operations. IRS-III offers different approaches to deal with these issues:

- **WS Selection**: each WSMO description of WS defines, in its capability, the specific class of shelter that the service provides. All descriptions are linked to Get-Circle-GIS-Data-Goal by means of a unique wg-mediator (wgM). The goal expects as input a class of shelter, and a circular query area. At invocation time IRS-III discovers through the wgM the WS associated to it. Then it selects one amongst them according to the specific class of shelter described in WS capabilities.

- **Area mediation and orchestration**: Get-Polygon-GIS-data-with-Filter-Goal is associated to a unique web service that orchestrates – here, invokes in sequence – three sub-goals. The first one simply gets the list of polygon edges from the input; the second is the above mentioned Get-Circle-GIS-Data-Goal; and finally the third invokes the smart service that filters the list of GIS data. The first two sub-goals are linked by means of three gg-mediators (ggM) that convert the list of polygon edges provided by the first sub-goal to the centre (latitude and longitude) and radius of the circle that circumscribes that polygon. To accomplish this, we created three mediation services invoked through Polygon-to-Circle-Lat-Goal, Polygon-to-Circle-Lon-Goal, and Polygon-to-Circle-Rad-Goal. The results of the mediation services and the class of shelter are the inputs of the second sub-goal. A unique ggM connects the output of the second to the input of the third sub-goal. No mediation service is necessary here.

Other improvements upon WS are made possible by IRS, such as: describing complex orchestrations through a full workflow model expressed in OCML; supporting dataflow and solving mismatches through mediators; and defining how to interact with a single deployed WS (e.g. policies) on the basis of a set of forward-chaining rules (Cabral et al. 2006).

3 The eMerges Approach

3.1 Semantics for the Geographic Space

It is well acknowledged that the spatial domain is somehow special (Peuquet 2002). Indeed, Geographic Space encompasses objects quite different from the ones we usually manipulate or are used to describing in knowledge bases; here scale, orientation, boundaries, and cultural conceptions, amongst other elements, seem to matter to a greater extent (Smith and Mark 1999).

If a full review of the specificity of the geographic domain is beyond the scope of our work, three aspects of this specificity particularly oriented our research:
• **Object/Field Divide:** it has been recognized that objects and fields – the assignment of values to spatial locations – have to coexist in geographic applications (Couclelis 1992). However, this distinction still constitutes a problem for the object representation tradition. Indeed, *why* is an object such as a mountain a field or an object, or, better, *when* do we want it to be a field or an object? What about fields composed of other fields (e.g. land coverage)? If answering these questions in a generic manner is hard, human cognition never fails in choosing the best representation, object or field or composition of both, according to a context.

• **Cognitive Imperative:** space is experienced before being understood, as shown by Naïve Geography (Egenhofer and Mark 1995), which demonstrates to what extent useful representations of space are to be rooted into human cognition. This is highlighted in yet another way by web2.0 maps, in which multiple reality effects are embedded, such as seamless continuity in map browsing instead of image by image retrieval, satellite imagery, road level or oblique photography, 2.5 or even 3D features. These representations are appealing since they allow the transition from the world of symbolic representation toward iconic models of reality used commonly in daily life, and therefore allowing applying cognitive models. These glimpses of a world behind the map provide us with new *affordances* (Gibson 1986) (what an element of the external world *allows me to do* as more essential than its other characteristics; symbols direct and focus the perception of affordances while the vision of realistic images allows the full range of them), *image schemata* (Mark 1989) (an element can be further reduced to simple concepts which are self-understandable; we are used to such kind of abstractions from perceptive inputs), or *conceptual spaces* (Gärdenfors 2000) (a concept as a point in a multi dimensional space of simpler representations; the meaning of a real object is somehow defined by one’s perspective on it).

• **Multi-Representation:** at the intersection of the object field divide problem, and the need of cognitive approach to object representation, spatial applications need to represent spatial objects, objects which representation simply *changes* not only according to the level of detail needed or requested (*generalization*), but also depending on the task at hand. For example an airport such as Stansted will be a node in a flights graph from an international point of view, then become an independent region in a land cover study, or a simple traffic node, or a complex environment itself containing a road network and buildings, or a group of 3D structures with emergency access path in a fire escape scenario, etc. The multitude of contexts and corresponding relevant representations raises the question of the possible uniqueness of geographic object representation; indeed, if many representations are useful how can they be linked and accessed in a timely manner, according to contextual information?

**eMerges** is an ongoing effort to address these concerns, by linking them via the notion of context. Indeed, in order to ultimately (a) alternate object and field representations, (b) provide cognitively relevant information, and (c) choose between
multiple representations of the same element, the representation of spatial objects becomes context dependent. We are going to define both notions in turn.

### 3.2 Spatial Objects

Firstly, in order to describe and to reason about Spatial Objects in all their generality, a simple yet precise definition is needed. Our model is based on Galton’s theory of objects and fields (Galton 2001) which defines a spatial-object as belonging to a given type and having a location component (some “whereness”) as well as attributes (also called features).

Mapping of arbitrary domain entities to spatial objects can be automatic or manual. In automatic mapping, a procedure collects each object’s attribute value and transforms it into an attribute name/value pair of a spatial object, with a special treatment for id and location. In manual mapping, arbitrary transformations are possible. Once spatial objects are gathered, further mappings are needed to achieve independence between objects and their actual use.

For example generalization is achieved in eMerges by using an Archetypes ontology providing generic abstractions (e.g. container, house, agent, etc.) to which entities have to be mapped. In this way even if the client application does not understand the type of element that is to be represented, a choice of representations and affordances is still possible by reasoning on the attached archetypes, which clients are requested to be aware of. A hospital for example, can be represented as a house, the attached archetype, with affordances including how to get there, which is in any case sounder than other archetypal representations such as agent or link, which are distinct archetypes.

Moreover, to adapt the representation of a spatial object to a particular interface, the HCI ontology maps an object to a particular HCI representation. For example some interfaces need “pretty names” selecting a feature to privileged display (e.g. on hovering on the object); an attribute of an adapted HCI concept allows us to specify which information, by automatic mapping (e.g. a procedure choosing any slot containing the string “id” or “name”) or with a manual one.

These ontologies, together with the attached mapping mechanisms, are called integration ontologies since they allow the integration of spatially related data sources ranging over very different domains. Alone, they allow to integrate spatial data sources in a generic way, however, as the number of data sources increases, the task of presenting objects and possible queries according to the context, in order for the user not to be overwhelmed by the amount of information, becomes essential. Hence, the notion of spatial context becomes important in order to provide only relevant information and services.

### 3.3 Spatial Context

In order to alternate cognitively sound representations and actions it is acknowledged that some extent of context-awareness (Dey and Abowd 2000) is needed. In eMerges, the main components of context are related to user role, task, location, and focus of interest. Indeed a user identifies him- or herself as having a particular role, such as firemen responsible of transportation in a snow storm emergency, or
police forces responsible for victim’s accommodation. Moreover, weather information is available only in the region covered by the service, and the option of asking for it must be presented only in relation to objects related to weather investigation or emergency planning.

Object representations differ according to the context; e.g. emergency planners view shelters as points independently of scale, while the fire brigade responsible for transport need precise access plans at a greater proximity. Secondly, to spatial objects are linked possibilities of action which allow getting more information in a precise context. For example an area defined as an evacuation zone may offer goals allowing finding the nearest supermarkets – providing food – or hotels – providing accommodation –, etc. This links the SWS notion of goal to the cognitive notion of affordances attached to an object. Therefore, when involved into a context, a spatial object receives specific affordances, linked to WSMO goals. Affordances allow navigation through the Geographic Space by successive and uniform information retrieval steps, i.e., as hyperlinks allow to navigate from web pages to web pages, affordances are attached to an object depending on the context and allow retrieving additional spatial objects. For example, in a given context, a town object will afford retrieving nearby hospitals, while in another will allow retrieving its administrative subdivisions. All retrieved objects are also captured within a similar context and present relevant affordances.

To achieve this, the question of whether a specific context reasoning engine has to be used is open. However, we believe that in the context of SWS, a more scalable solution may be achieved by distributing the task of context handling amongst smart services which also implement reasoning in our architecture. Indeed context pervades the elements of a SWS application, and can be represented (a) at an affordances level, i.e. by offering very specific goals only, according to the context, e.g. a get-heated-shelters affordance will be presented in an emergency case involving low temperatures, or (b) at a composition level, i.e. generic affordances are presented but smart composition between goals ensures context relevance, e.g. the generic affordance get-shelters is presented to the user but will highlight heated shelters according to the snow storm task. The first solution has the advantage of being more explicit, whilst the second is easier to implement since it requires fewer goal definitions. Being able to handle context at every level makes both solutions possible in SWS based applications.

4 Interaction in eMerges

The prototype implementation is a web interface using Google Maps for the spatial representation part of the application. The interface is built using the Google Web Toolkit, using AJAX techniques on the client to communicate with a Java servlet, which itself connects to IRS-III through its Java API. The most significant component of the interface is a central map, supporting spatial objects. A spatial object can have an area based location, in which case it is displayed as a polygon, or a point based one, in which case it is displayed as a symbol. All objects present the same interface, with affordances and features, displayed in a pop up window or in a hovering transparent region above it (cf. Figure 3. Screenshot of the eMerges im-
plementation; a field (bottom left), point elements and a polygon (middle), affordances and features dialog (top right)).

As an example of practical usage, we describe how an Emergency Officer (EO) gathers information regarding a possible emergency situation, and the affordances which are made available depending on the context. The procedure is as follows:

1. Based on external information about the possibility of a weather emergency the EO defines an area of interest on the map.
2. A popup window containing a tree view appears showing elements of the available ontologies which location is an area and which are therefore candidates to be the type of the region. The choice in this example is weather-investigation-area.
3. As defined in the ontology, the new instance has attached features and affordances. One affordance allows to login into BuddySpace to ask field agents for more information, while another one accesses meteorological information.
4. The EO requests snow information in the area.
5. The result is a field of snow-value objects attached to regions. Although essentially representing scalar values each polygon constitutive of the field is itself an object, presenting affordances and features.
6. Depending of the results the original investigation area becomes a low-snow-hazard-emergency or a high-snow-hazard-emergency, as described in the ontology.
7. A high-snow-hazard-emergency provides affordances which allow the user to ask for more information, for example to request all rest-centres inside the region.
8. Rest centres are retrieved with features and affordances.
9. And can be used to get more context relevant information, i.e. other resources nearby such as hospitals.
10. The EO can also choose to log into BuddySpace to contact the relevant persons to request action or information.

A screencast of similar interactions as well as a live version are available online\textsuperscript{7}, to be used preferably with the Firefox Web browser\textsuperscript{8}.

**Discussion**

Two main aspects of eMerges can be related to other approaches: data integration and context based navigation of data.

Integration of new data sources is relatively simple in eMerges, although not entirely trivial. Indeed IRS-III SWS integration allows the description of any XML data source available on the web. From an expert point of view the data source integration approach presents notable advantages compared to approaches based on standards such as the one demonstrated in the OWS-3 Initiative\textsuperscript{9}. These advantages are framework openness (i.e. standards make integration easier but are not mandatory) and high level service support (i.e. all the benefits of the underlying SWS platform, such as discovery, composition, etc. are immediately available). The steps involved in the process of adding a new data source, as well as the ability to automate each step, are described in the following:

- **Ontological description of service**: the service, composed of the data types involved as well as its interface, can be described in a low level ontology, i.e. at a level to remain close to the data. This step can be automated in many cases based on information contained in the schema of the service.

- **Lifting definition**: the lifting operation allows the passage of data type instances from a syntactic level (XML) defined in the data schema to an ontological one (OCML) specified in the ontology definition. This process can be automated every time the previous step can be.

- **Goal description**: a new goal has to be defined which represents the newly integrated web service.

- **Mediator description**: the goal has to be linked to the WS with a mediator, which is often a trivial operation.

- **Lowering definition**: the lowering operation transforms instances of aggregation ontologies into syntactic documents to be used by the server and client applications. It is automatic since integration ontologies do not change.

- **Mapping to integration ontologies**: this process is achieved by the knowledge engineer who modifies an ontology, defining which affordances are relevant to which context, with immediate effect.

This last step, which links affordances to a context rather than to a map, a system or simply to an object allows meaning to emerge into an otherwise overwhelming amount of geographic data. This is absent from automatic syntactic mashup build-
ers\textsuperscript{10}, or even from semantic ones such as Geo-Names\textsuperscript{11}, which gather feeds on a map without taking context into account.

Other similar approaches which seem context aware, such as the use of tasks in the recent – and mostly undisclosed at the time of writing – ESRI ArcGIS Explorer product\textsuperscript{12} are, to the best of our knowledge, actions attached to maps and which return heterogeneous results. This results do not seem to provide new tasks in an uniform and meaningful way.

An alternative method of adding meaning to spatial data can be found in the AKTive.Response\textsuperscript{13} approach, where the Compendium tool is used for collective sensemaking, i.e. while gathering information from multiple (spatial) data sources collectively building context relevant concept maps on the fly, with the help of other various ontology aware tools (Tate et al. 2006). However, context, task relevance, and the choice of affordances are still mostly left to the emergency planner, and data source integration seem to include essentially information messages.

**Conclusion**

The eMerges approach to spatial data integration presents advantages for the end user as well as for the data integration expert. Indeed it allows the end user to handle tasks in a data rich environment without being overwhelmed by the amount of information or by the complexity of the queries, and to the expert an easier approach to data integration.

In 2006 the eMerges prototype has won a prize for the integration of web scripting technologies with Semantic Web ones\textsuperscript{14}, and has been selected amongst the five finalists of the Semantic Web Challenge\textsuperscript{15}. Future developments will include an increase in the complexity of the integration ontologies (spatial, HCI and archetypes) in order to allow multi representation and an improved management of context to offer more cognitively sound features. Also, making the integration of new data sources even easier constitutes a long term goal for the IRS SWS execution platform.

**References**


2 UDDI. http://www.uddi.org/.
3 WSDL. http://www.w3.org/TR/wSDL.
4 SOAP. http://www.w3.org/TR/soap12-part1/.
9 OGC. http://www.opengeospatial.org/committees/?id=162.
10 e.g. GEORSS. http://mapufacture.com/georss/.