Paper Number
UC1258

Title
Designing Service-Oriented Spatial Data Infrastructures

Author
Javier Morales

Abstract
Technology has made it easier to access, manipulate, and exploit spatial data. Consequently, a more sophisticated spatial awareness has developed among users, resulting in growing dependence on people and organizations for spatial data. This translates to diverse and changing user requirements. This presents a challenge for any supplier. To address this development, spatial data infrastructures need to change from being a data discovery and retrieval facility to becoming a service-oriented infrastructure in which users can rely for the provision of geo-services. Work then needs to flow across several companies, requiring not only the sharing of data but of resources, functions and processes. This is achieved with a unified framework architecture for the provision of geo-services across organizations and communities. It enables the integration of disparate systems, facilitating access and chaining of core services (sensor, visualization, processing and other services) to create customized user services in line with changing requirements.

Introduction
Today's geo-information business should not only focus on acquiring, storing and publishing data, attention has to be given to add-value and integrate spatial data. This will enable the development of information services which in tern could lead to the improvement of spatial data use and of the quality of its users' decisions.

Geo-Information (GI) Services
There exist three main phases in the value-chain of spatial data: generation, communication and use. In the last decade Spatial data infrastructures have played a significant role in the communication phase by facilitating discovery and access to previously generated data. Data use however has, to a certain extent, been neglected or left to chance. It is generally agreed that it is the task of a prospective user to make sense of large volumes and to use it correctly according to some specific needs. Such a passive approach has led to data being collected and advertised but never really used to its full potential. A more proactive strategy has to be put in place to increase data use. The development of information services is a suitable answer for just such a scenario. Developing services however presents its own set of problems. To be useful, a service has to be compliant with users' requirements. Requirements depend, to a large extent, on the way data is perceived, expected and used, which in time depend on the current forms of projects, markets and technology. Proper alignment of services with respect to such changing requirements requires a flexible approach to service realization. This can be achieved, for example, by identifying core (atomic) services that may be used in multiple non-persistent combinations to deliver different custom-based services. Inline with this description a geo-information service is defined as a non-persistent collection of elements, (e.g. data, operations, resources) organized in such a way that they exhibit behavior of value to a user. The elements of a collection and the way they interact come as a deliberate choice of a service provider in order to comply with a given set of requirements.

The modern spatial data infrastructure is changing from being a simple data discovery and retrieval facility to become an integrated system suitable for the provision of customized information and services. Services here are seen as the contribution of a system or part thereof to its users. This contribution can be defined in terms of data, operations, processes, resources, value-added products or any combinations of them. Normally GI providers address the issue of providing services by stringing together groups of functions in an ad-hoc manner. Such approach may satisfy a single need but doing this continuously and separately for different services hampers reusability. Moreover, lack of descriptions of the solutions obtained makes it hard to aggregate them to support more elaborated tasks.

GI Service Infrastructure

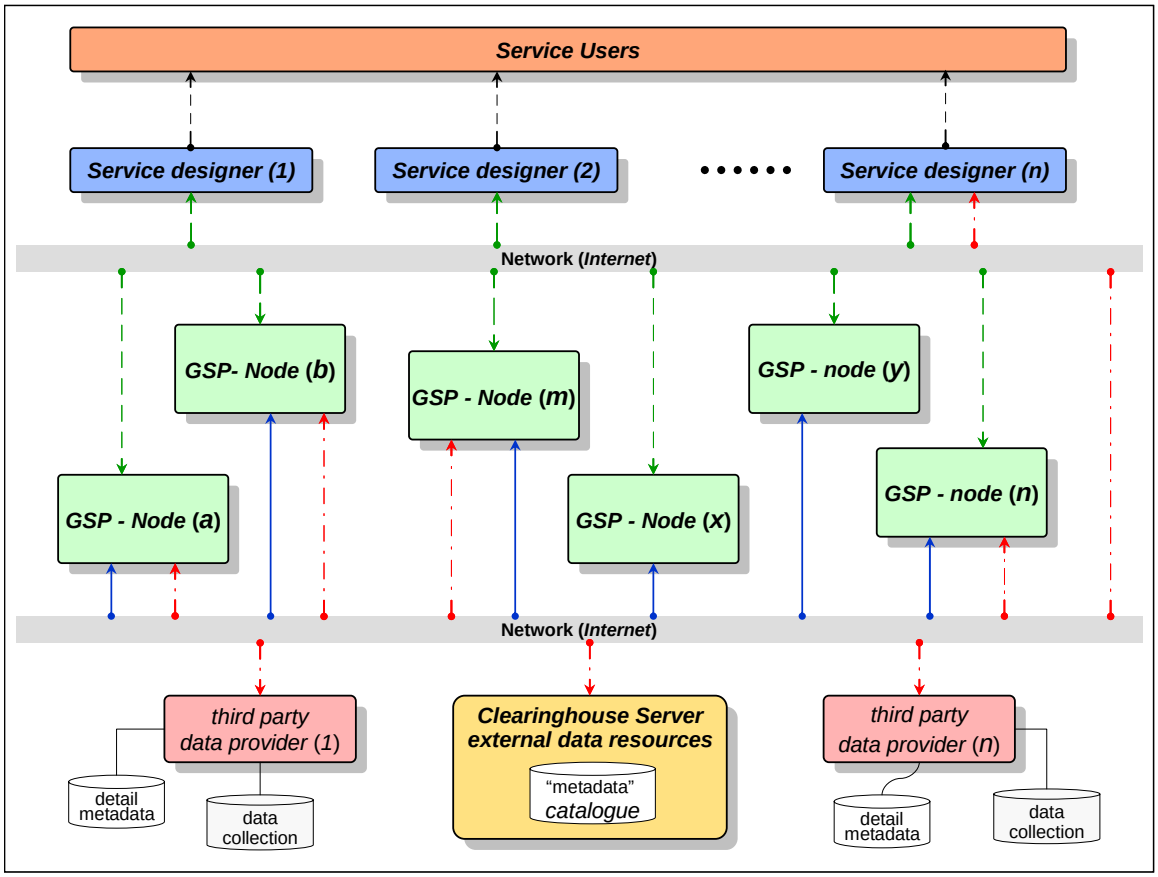

Figure 1. Roles within a gi-service architecture

Current research is focusing on the development of mechanisms to manage independent collections of core services in a manner that supports their combination or assembly and improves reusability and flexibility while maintaining the correctness of the resulting compositions. The so-called Geo-information Service Infrastructure, or GSI for short, aims at providing just such a facility. The underlying principle of the GSI is that independent gi-resources can be described, accessed and combined to create elaborated service chains which deliverables are compliant with a given set of requirements. Within the GSI, a common method is used to describe gi-resources and their interfaces, and then these gi-resources are made available for users to create service chains that perform complex geo-processing tasks. Such gi-resources are available along an infrastructure of interconnected nodes that include data repositories, data brokers, service providers, service brokers and clients. This service concept builds upon the layer of interoperability of

information as defined by the OpenGIS implementation specifications (OGC, 1999), therefore separating the actual implementation of services from their definitions and the perception of these services by the users.

Large geo-processing tasks are realized by combining or chaining sets of gi-resources (data, operations, processing units, sensors, etc.) located along the distributed nodes. In order to bind multiple gi-resources into a chain that accomplishes a large geo-processing task, a proper description of the participating available gi-resources is required. These descriptions focus on exposing the resource's behavior (function) and its interaction mechanisms or points of composition. These descriptions, which are presented as instances of well-defined models, make it possible to interchange and reuse gi-resources. We call these descriptions system metadata; they are stored and made accessible through a service repository.

The GSI system enables Geo-Service Providers (GSPs) to make use of functionality offered by others to supply a wide range of services and possibly to reach larger groups of service users. Figure 1 illustrates the interactions that take place as GSP-nodes provide services to their service designers. Service designers interact with the different GSP-nodes to request their specific services. Figure 1 shows these interactions as dashed-lines. GSP-nodes may make use of architectural elements available in other GSP-nodes in order to realize a particular service. These interactions between GSP-nodes are shown in Figure 1 as solid lines running from node to node. All connections are established through a network. At the bottom of Figure 1 we can see that additional data collections located at non-GSP-nodes may still be accessed, if needed, either by users or service providers. This is achieved by making use of the conventional data discovery functionality, of the clearinghouse server. These interactions appear in Figure 1 as dashed-dot-lines.

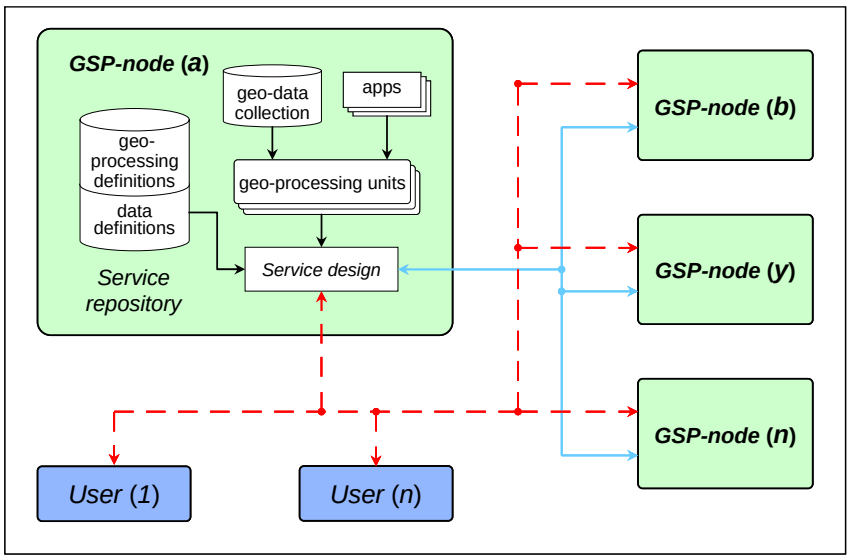

Figure 2. GSP-node internal architecture

Figure 2 shows the internal architecture of a GSP-node. The service repository component contains the descriptions of available gi-resources (service descriptions), either data definitions, process definitions, or previously assembled service chains. The geo-processing units are responsible for the execution of the various functions of the node. These units use data and applications during operation as specified in the definitions stored in the service repository. The service design component is in charge of defining how the different services are realized, It communicates with other gsp-nodes if their resources are needed for an specified service chain. The process of generating service

chains within the GSI can be broken down into three major activities: defining and registering gi-resources, assembling service chains (gi-services) and delivering the results. Three different roles can be identified from these activities: service providers, service consumers and end users.

- Service providers are responsible for describing and making their elementary services available for others to use. Service providers make use of a framework that enables the modeling of these gi-resources. These models act as descriptors that specify the function and the interaction point(s) of the gi-resource.
- Service consumers use available descriptions to design more complex services. Service consumers make use of the same framework used by the service providers to assemble individual gi-resources into chains. They define these chains by adding control elements that govern the relations between the gi-resources used in the chain. These control elements (mediators) help ensuring that the constraints and conditions defined at the interaction points of the gi-resources are satisfied. The resulting chain is realized by instantiating the behavior portrayed in the specification of the chain.
- End users trigger the definition and execution of service specifications by posting requests to the system.

The Service Repository

Our approach to gi-services design focuses on the use of conceptual models as an intermediate step in the development process, which sit in between requirements and the actual implementation. This is done solely to enable and facilitate reuse and to enhance flexibility. The main benefit of these models is to serve as the basis for the specification of complex services. If a model properly describes a gi-resource, that is, with the relevant information at the correct level of detail to enable one to determine what it does and how to access the function it provides, then this gi-resource can be easily reused. By reuse, we mean the inclusion of single gi-resources in multiple service definitions (gi-services). Since the models of gi-resources prescribe the behavior exhibited by individual elements, a gi-service model can be used to choreograph the realization of the service specified in the model. Additionally, once a service model is available it can itself be reused, as a non-atomic gi-resource, in another definition as a part of a yet more specialized service.

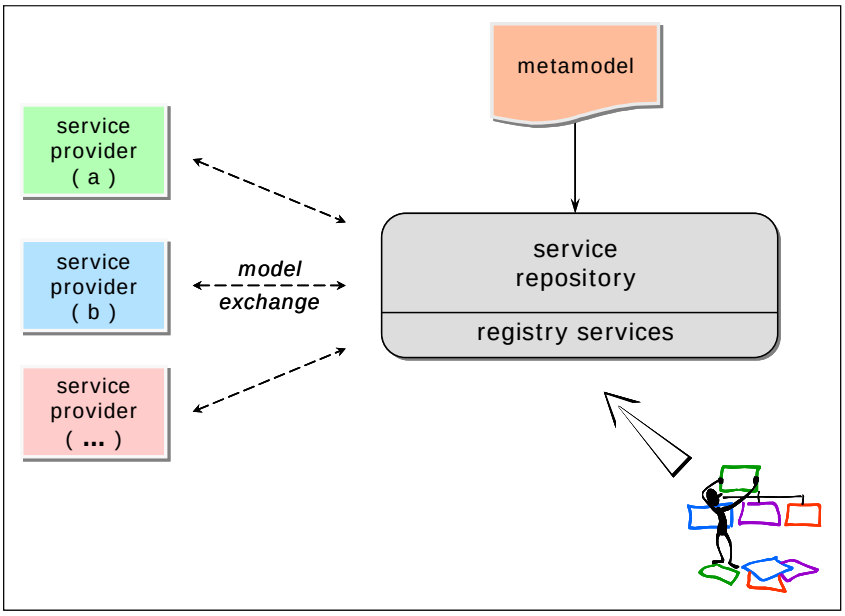

Figure 3. Relationship between the service repository and the metamodel

For this approach to work, models need not only to be interchanged between participants, but they also have to be understood by all parties involved. This can only be achieved if the models are based on the same metamodel. Such metamodel should therefore provide a rigorous abstract syntax for defining models. Figure 3 shows the role of the metamodel. The metamodel enables the implementation of a repository where compliant models of gi-services and gi-resources can be stored. Hence, the repository becomes the central component in a GSI system. The repository supports the exchange of models between different service providers, thereby facilitating the use of these models in combinations to form more complex service models that address specialized sets of requirements.

The GI-services Framework

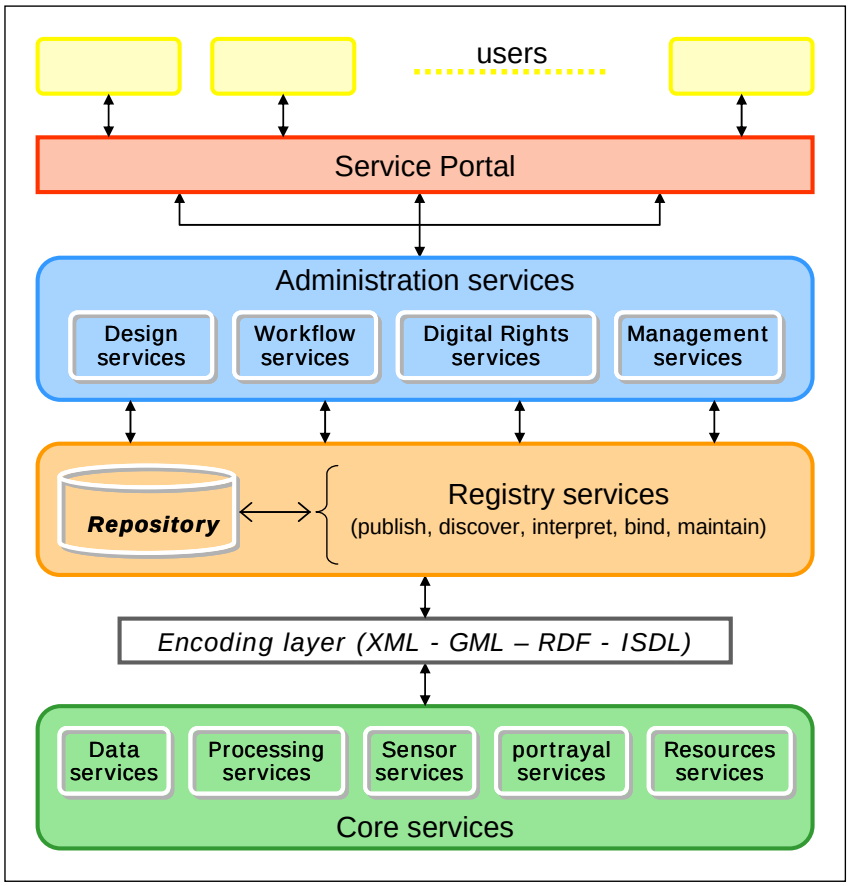

Figure 4. Service classification scheme

To really enable multiple geo-service providers to work in a collaborative manner, sharing there several resources requires also a number of functions provided by the supporting infrastructure. To this end a service classification scheme has been defined, and a number of administrative services to support the operation of the infrastructure have been established. Three main categories of infrastructure services are defined: Registry Services, Administration Services and Core Services. Registry services are those necessary to store, query and use descriptions of gi-resources. These descriptions are machine-readable and human-readable descriptions created with XML-based languages. Registry services provide a common mechanism to classify, register, describe, maintain, access and combine information about gi-resources. Registry services provide all the functionality needed for the use of the repository according to the principles of the GSI system. Core services are those provided by the gi-resources. The type of service provided by the

core services can be of the following types: Data Services provide access to collections of data available in data repositories and databases. Processing services operate on spatial data and `add value' to it. They are used to perform computations and to transform, combine, or create new data. Sensor Services: Provide access to sensor operations paths and almanacs and to raw images and raw data. Sensors may include traffic cameras, satellites, weather stations, etc. Portrayal Services provide visualization functions to be applied to information. HW-Resource services provide access to hardware resources for storage or processing. Administration services are those needed for the smooth running of the infrastructure. The various types of administration services are: Design Services that are used to define combinations of core services to create customized gi-services. Workflow Services allow to choreograph any of the service chains defined using 'design services'. Resource Management Services: Allow to control status and availability of hw-resources.

GI-service Composition example

The land information service, LI-service for short, is a service centered on the generation of information about a piece of land. This includes, a.o., the actual use, the ownership, location and size. The main objective of the service is to determine the price and taxes of specific areas of land. The LI-service is the responsibility of the Land Information Agency, a fictitious gi-service provider.

Multiple types of raw data are required for the execution of this service, for instance, timely imagery covering the areas of interest for the taxation period under consideration, geometric features describing parcel boundaries and nearby infrastructure, weather data, etc. In addition to this raw data, specialized processing tasks are needed to extract from the raw data, the necessary information for the tax calculation. Examples include, a.o., image processing and geo-referencing.

The Land Information Agency does not count with all the data and processing resources necessary to make the service possible. Furthermore, the Land Information Agency neither has the expertise required to implement and maintain the complete service, nor the responsibility for generating and maintaining much of the necessary data. Nonetheless, the Land Information Agency is responsible for the tax calculation, and therefore has to identify a mechanism to realize it. This is achieved by designing a collaborative system to allow the Land Information Agency to make use of specialized services from other geo-service providers whose expertise are in line with the particular needs of the LI-service.

The LI-service is initiated every time the Land Information Agency receives a request from the government to determine the taxes for a group of taxable parcels. Four major steps can be identified in the execution of the LI-service. The first step encompasses the activities that lead to obtaining the necessary data to start any processing. The second major step is to utilize the imagery and any other sources to extract the information about current use of the land for the areas of interest. The third step involves the collection of information regarding productivity, such as temperature and rainfall, and regarding infrastructure such as, proximity to cities, roads, railways, etc. In the fourth step, which is the real competence of the Land Information Agency, the results of previous steps are used in the calculation of the taxes.

Our methodology for service chaining prescribes a decomposition based on functionality in order to identify the core services required to assemble the LI-service. Using a behavior walkthrough, we identify the following set of required services.

1. A *Parcel Feature Service* is used to obtain boundaries of parcels. The Cadastral Numbers of the parcels are provided as input to the service and the result is a theme that contains the polygons that define the shape and location of the parcels.

2. An *Imagery Service* uses the coordinates from the geometric extent (limits) of the parcel's theme, and the dates of the taxable period to obtain the satellite images covering the areas where the parcels are located. These images correspond to the starting and ending dates of the taxable period.

3. An *Image Processing Service* is used to geo-reference the satellites images and obtain a mosaic of the area of interest. The input to this service is the set of satellite images, and the result is the geo-referenced image mosaic.

4. A *Change Detection Service* compares images to identify the areas within the parcels that have had a change in the land use.

5. A *Feature Extraction Service* is used to define the different types of land use within the taxable parcels.

6. A *Weather Data Service* is used to obtain the information regarding temperature, rainfall, etc. of the area of interest during the taxable period.

7. A *Feature Selection Service* is used to obtain information about nearby infrastructure.

8. A *Taxation Service* is used to determine each parcel's tax and deliver the invoices.

Figure 5 shows a service composition including all of the core services identified before and their corresponding interconnections. At the level of abstraction shown in Figure 5 we depict the relations between elements as single interactions. These interactions need to be refined in order to fully specify the relations they represent.

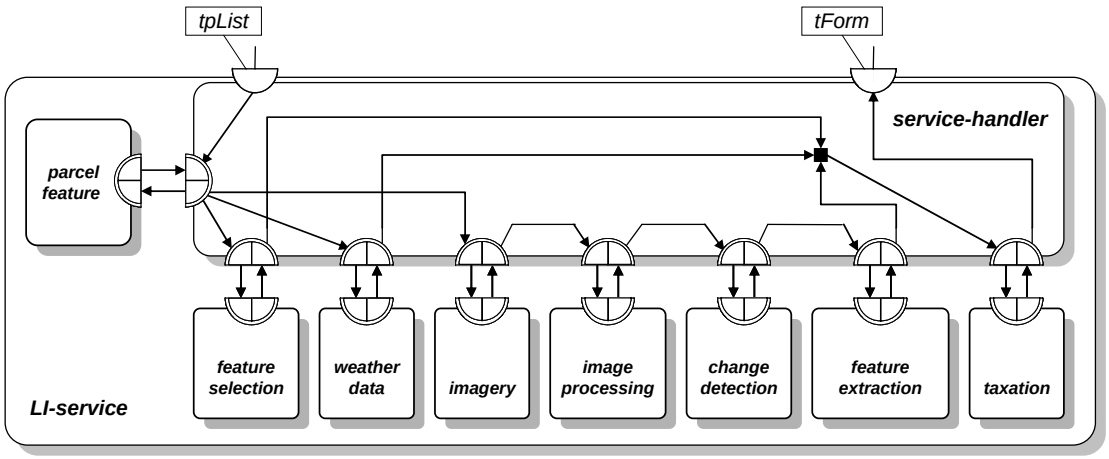

Figure 5. Initial internal perspective design

Figure 6 shows the refined definition of the interaction between the elements *weatherData* and *serviceHandler*. The Weather Data Service is an independent service that provides weather information for specific periods of time. The service requires as inputs two dates and an area of interest. These inputs are modeled in Figure 6 as *startDate*, *endDate*, and *region* respectively. The service also allows the provision of parameters to specify which type of weather data is being requested. If no parameter is provided the Weather Data Service generates only temperature data. The default option of the parameter, temperature, is shown underlined.

The *serviceHandler* element interacts with the *weatherData* element at interactions *wdReq* and *wdResp*. The *serviceHandler* element asks the *weatherData* element for the necessary weather data through the interaction wdReq. The information attribute of interaction *wdReq.serviceHandler* consists of three elements, *startDate*, *endDate*, and *region*, and the list of parameters, e.g., *temperature* and *rainfall*. These are used by the Weather Data Service to retrieve the corresponding weather data. The interaction *wdReq* is followed by the interaction *wdResp*, which delivers the requested data.

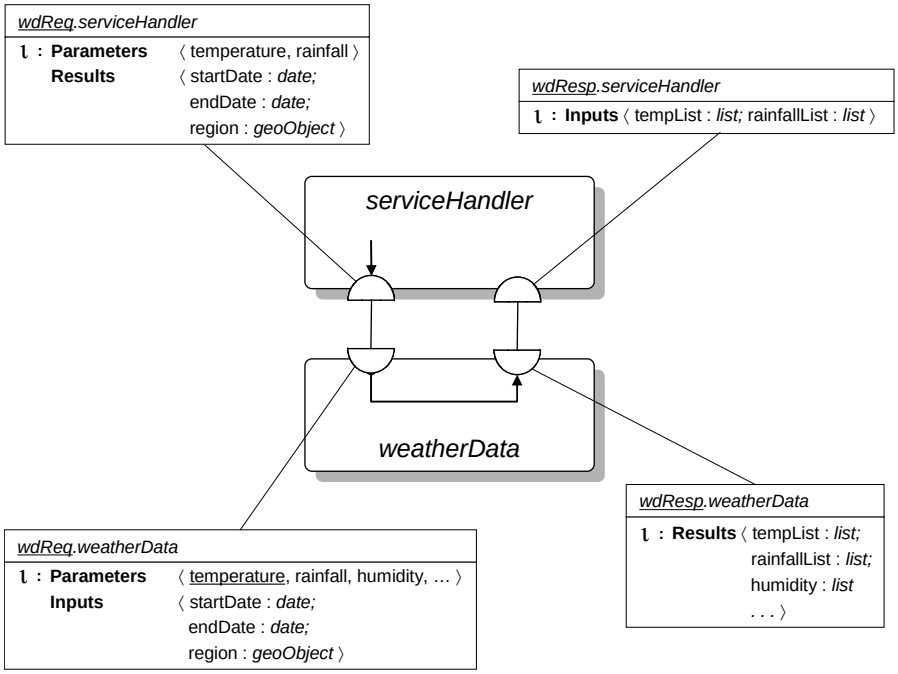

Figure 6. WeatherData service interaction diagram

The process of refining the interactions to define interaction contributions and inserting actions to carry out any additional task needed to enable the connections between two elements is repeated for all the other interactions defined for the service. For simplicity reasons we have only shown the refinement of the interaction between the *serviceHandler* and the *weatherData* elements.

References

Morales, J., Ferreira Pires, L. and van Sinderen, M., 2002. Model driven geo-information systems development. In: *Proceedings of the Sixth International Enterprise Distributed Object Computing Conference EDOC 2002*, IEEE Computer Society, Lausanne, Switzerland, pp. 155–166.

OGC, 1999. *The openGIS Abstract Specification Overview*. Version 4, Open GIS Consortium, Inc.

OGC, 2002. The openGIS Abstract Specification Topic 12: OpenGIS Service Architecture. Version 4.3, Open GIS Consortium, Inc. Editor: George Percivall.

OGC, 2003. *OpenGIS Reference Model*. Version: 0.1.2, No. OGC 03-040, Open GIS Consortium Inc. Editor: Kurt Buehler.

Szyperski, C., 2002. *Component Software: Beyond Object-Oriented Programming*. 2nd ed, Addison-Wesley Pub. Co., Essex, England.

Author Information

Name: Javier Morales Ph.D.
Title: Assistant Professor in Geo-information Systems and Services
Organization: International Institute for Geo-Information Science and Earth Observation
Telephone: +31 (0)53 4874299
Fax: +31 (0)53 4874575
Email: jmorales@itc.nl