Designing Service-Oriented Spatial Data Infrastructures

By Javier Morales, International Institute for Geo-Information Science and Earth Observation

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 on which users can rely for the provision of geographic information (gi-) services. Work then needs to flow across several companies, requiring the sharing of not only data but also resources, functions, and processes. This is achieved with a unified framework architecture for the provision of gi-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.

Today's geoinformation business should not only focus on acquiring, storing, and publishing data but also give attention to adding value and integrating spatial data to enable the development of information services that can lead to improved spatial data use and better decisions.

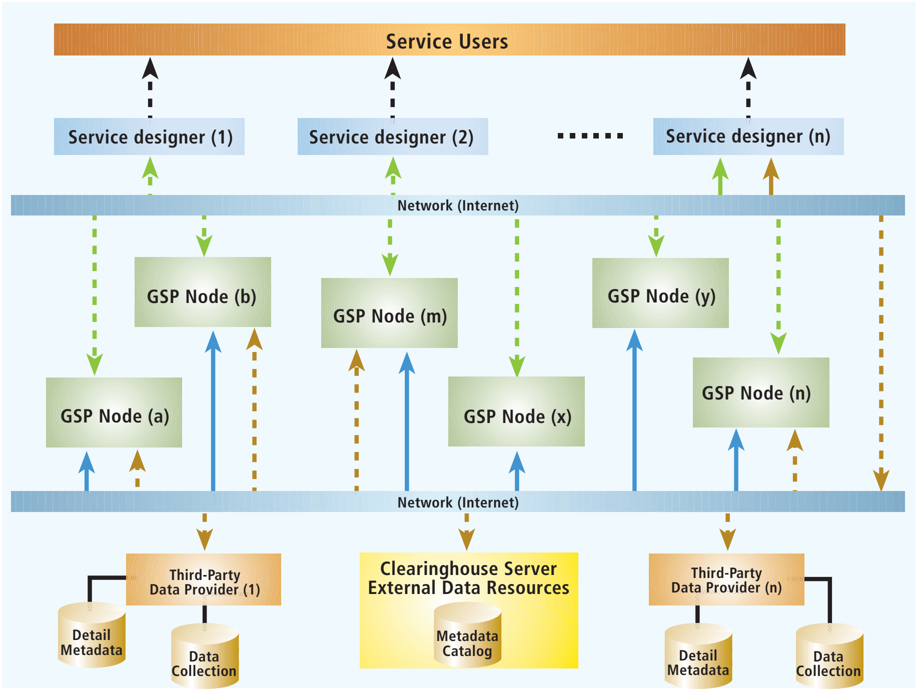

Figure 1: GSP nodes provide services to service designers.

Geoinformation 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.

However, data use 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 of data and 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. However, developing services presents its own set of problems. To be useful, a service has to be compliant with users' requirements. To a large extent, requirements depend on the way data is perceived, expected, and used, which in turn depends 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 nonpersistent combinations to deliver different custom-based services. In line with this description, a gi-service is defined as a nonpersistent collection of elements, (e.g., data, operations, resources) organized so that they exhibit behavior that is of value to a user. The elements of a collection and the way they interact are the results of deliberate choices by a service provider to comply with a given set of requirements.

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

GI Service Infrastructure

Current research is focused on developing mechanisms that 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 geoinformation 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 elaborate service chains to 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. These gi-resources are made available for users to create service chains that perform complex geoprocessing tasks. Such gi-resources are available along an infrastructure

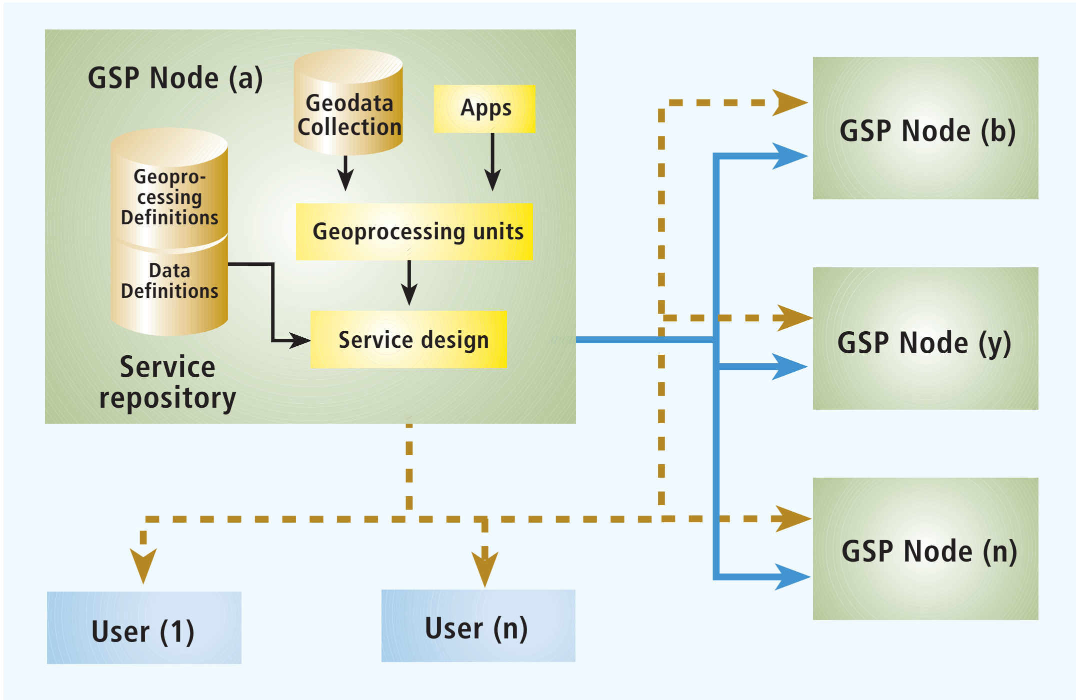

Figure 2: The internal architecture of a GSP node

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), which separates the actual implementation of services from their definitions and the perception of these services by the users.

Large geoprocessing tasks are realized by combining or chaining sets of gi-resources (data, operations, processing units, sensors, etc.) located along the distributed nodes. Binding multiple gi-resources into a chain that accomplishes a large geoprocessing task requires a proper description of the participating available gi-resources. These descriptions focus on exposing the resource's behavior (function) and its interaction mechanisms or points of composition. These descriptions, presented as instances of well-defined models, make it possible to interchange and reuse gi-resources. These descriptions, called system metadata, 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 reach larger groups of service users. Figure 1 illustrates the interactions that take place as GSP nodes provide services to service designers. Service designers interact with different GSP nodes to request specific services. Figure 1 shows these interactions as dashed lines. GSP nodes may make use of architectural elements available in other GSP nodes to realize a particular service. Interactions between GSP nodes are shown in Figure 1 as solid lines running from node to node. All connections are established through a network. The bottom of Figure 1 shows that additional data collections, located at non-GSP nodes, may still be accessed if needed either by users or service providers by using the conventional data discovery functionality provided by a clearinghouse server. These interactions appear in Figure 1 as dashed-dot lines.

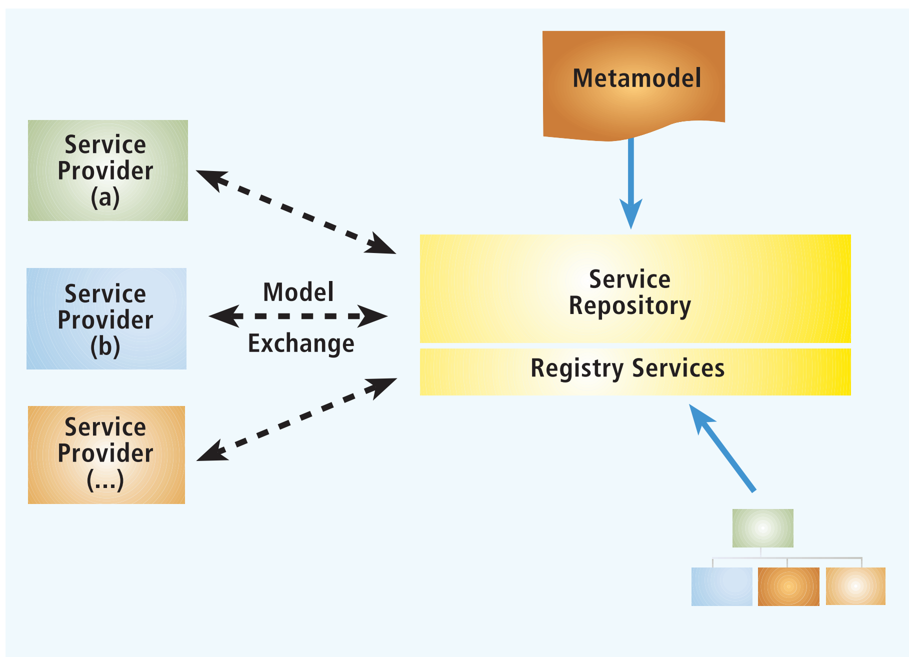

Figure 3: The role of the metamodel in enabling the implementation of a repository where compliant gi-service models and gi-resources can be stored

Figure 2 shows the internal architecture of a GSP node. The service repository component contains the descriptions of available gi-resources (service descriptions)—data definitions, process definitions, or previously assembled service chains. The geoprocessing 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 a 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. From these activities, three different roles can be identified: service providers, service consumers, and end users.

Continued on page 36

Designing Service-Oriented Spatial Data Infrastructures

Continued from page 35

- 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 relationship between the gi-resources used in the chain. These control elements (mediators) help ensure 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

This approach to gi-service design focuses on the use of conceptual models as an intermediate step in the development process, located between requirements and the actual implementation. This is done solely to enable and facilitate reuse and enhance flexibility. The main benefit of conceptual models is that they are the basis for the specification of complex services. If a model properly describes a gi-resource with the relevant information at the correct level of detail to enable one to determine what the gi-resource does and how to access the function it provides, the gi-resource can be easily reused. Reuse means the inclusion of single gi-resources in multiple service definitions (gi-services). Because gi-resource models 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 be reused as a nonatomic gi-resource in another definition as part of a yet more specialized service.

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

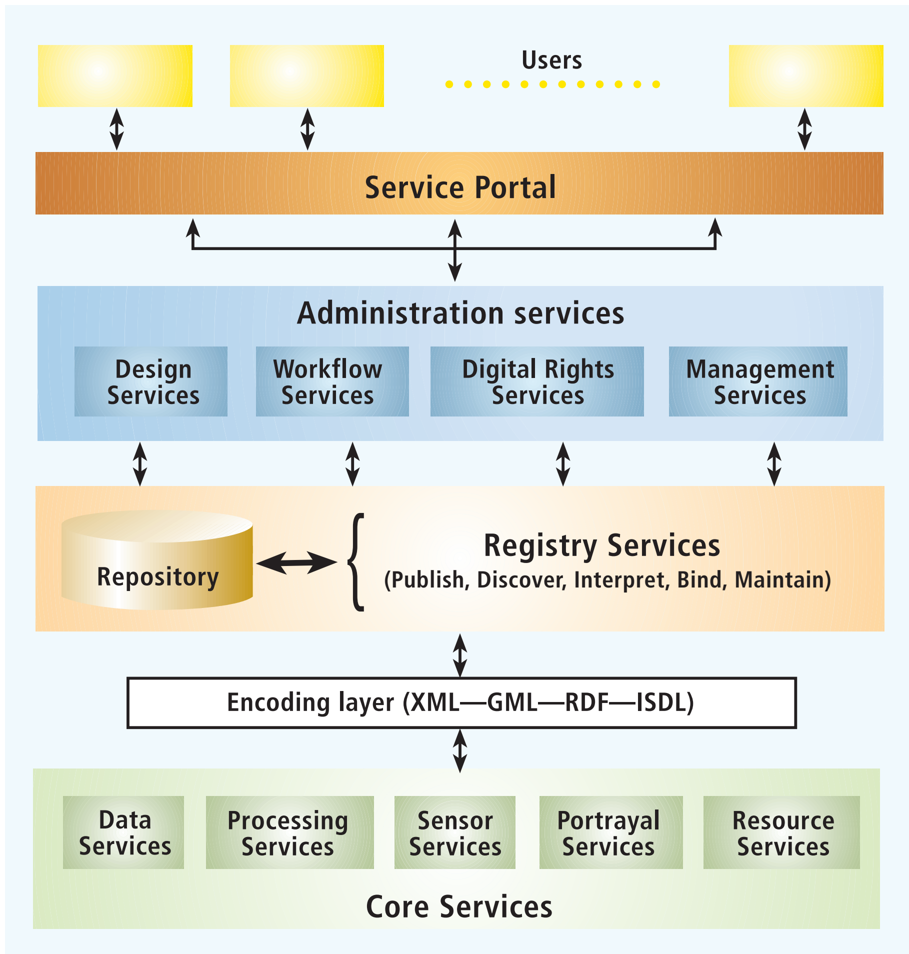

Figure 4: Service classification scheme

The GI-Services Framework

To really enable multiple geoservice providers to work in a collaborative manner, sharing several resources also requires 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, Core Services, and Administration 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. These 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. These 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 operation paths and almanacs and to raw images and raw data. Sensors may include traffic cameras, satellites, and weather stations.Portrayal Services provide visualization functions to be applied to information. Hardware (HW) Resource Services provide access to hardware resources for storage or processing.

Administration Services, needed for the smooth running of the infrastructure, can be of three types.

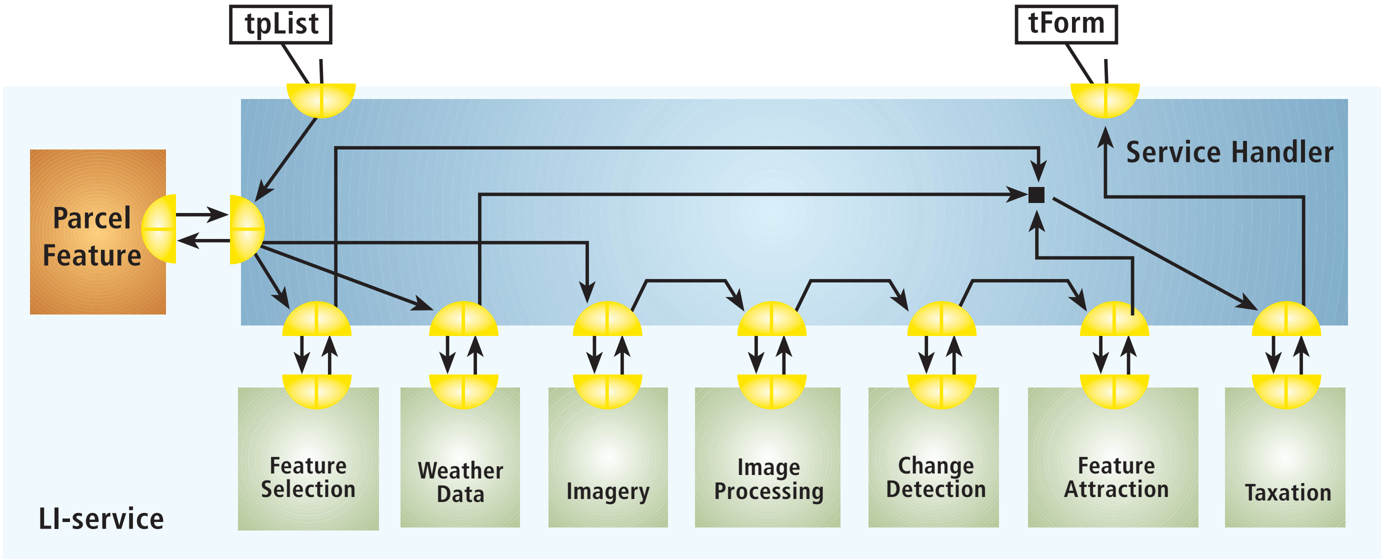

Figure 5: Initial internal perspective design for a service

- Design Services used to define combinations of core services to create customized gi-services.
- Workflow Services that choreograph any of the service chains defined using design services.
- Resource Management Services that control status and availability of HW Resource Services.

GI-Service Composition Example

The land information service, or LI-service, is a service centered on the generation of information about a piece of land. This includes information such as the actual use, 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 a fictitious gi-service provider, the Land Information Agency.

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

The Land Information Agency does not possess all the data and processing resources necessary to make the service possible, lacks the expertise required to implement and maintain the complete service, and is not responsibile for generating and maintaining much of the necessary data. Nonetheless, the Land Information Agency is responsible for the tax calculation and has to identify a mechanism to realize it. This is achieved by designing a collaborative system that allows the Land Information Agency to use specialized services from other geoservice providers that have expertise that meets 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

1. Performing the activities for obtaining the necessary data to start any processing
2. Utilizing imagery and any other sources to extract the information about current use of the land for the areas of interest
3. Collecting information regarding productivity, such as temperature and rainfall, and infrastructure such as proximity to cities, roads, and railways
4. Using the results of previous steps to calculate the taxes (by the Land Information Agency)

The methodology for service chaining prescribes a decomposition based on functionality to identify the core services required to assemble the LI-service. Using a behavior walk-through, the following set of required services can be identified.

1. A Parcel Feature Service is used to obtain boundaries of parcels. The cadastral numbers for 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 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 georeference the satellite 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 georeferenced image mosaic.
4. A Change Detection Service compares images to identify the areas within the parcels where land use has changed.
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 temperature, rainfall, and other information for 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 the core services previously identified and their corresponding interconnections. The level of abstraction shown in Figure 5 depicts the relationship between elements as single interactions. These interactions need to be refined to fully specify the relationships they represent.

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 two dates and an area of interest as inputs. These inputs are modeled in Figure 6 as startDate, endDate, and region, respectively. The service also allows the provision of

Continued on page 38

Designing Service-Oriented Spatial Data Infrastructures

Continued from page 37

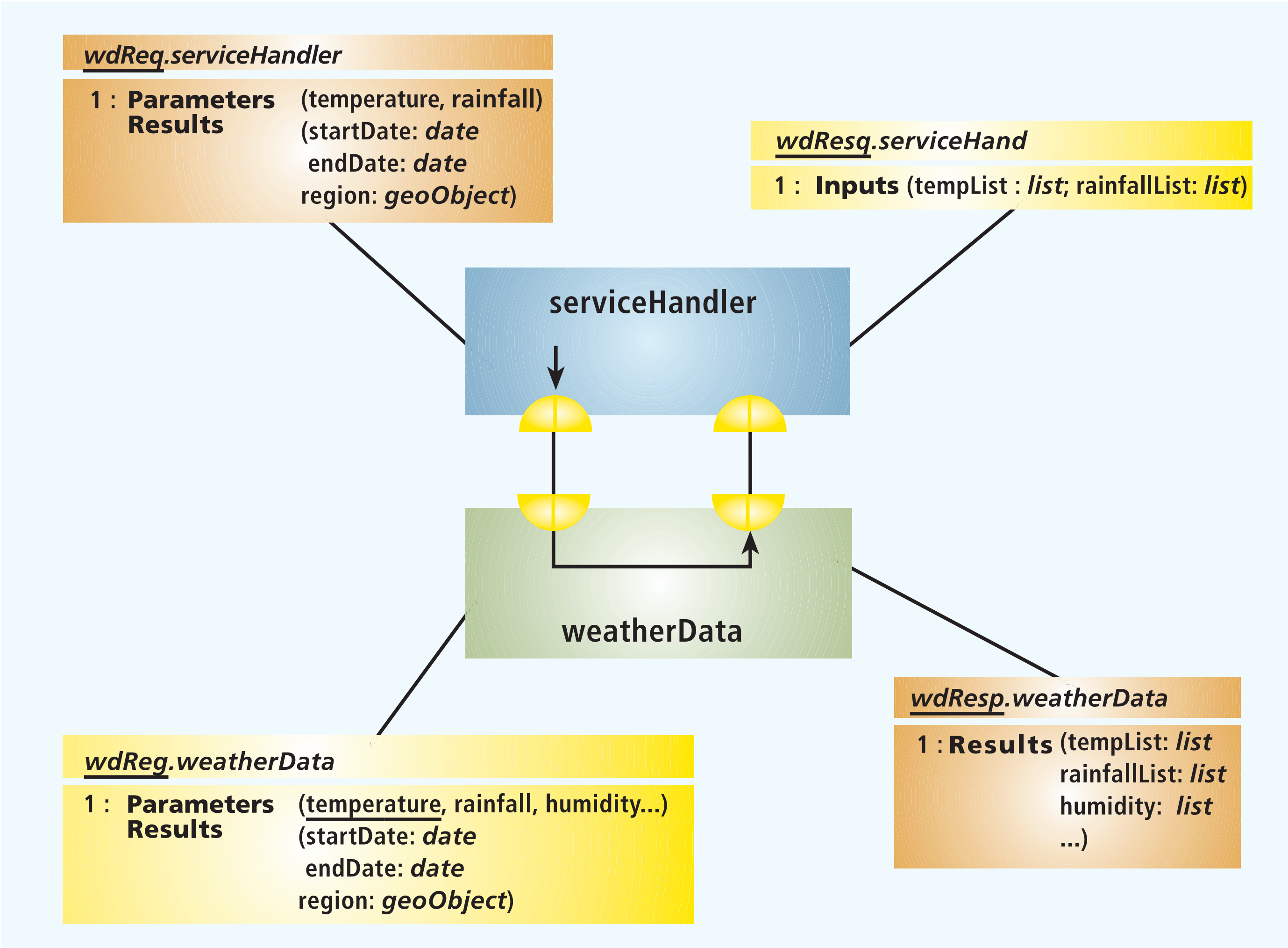

Figure 6: Weather Data Service interaction diagram

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 elements 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.

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, only the refinement of the interaction between the serviceHandler and the weatherData elements was shown.

For more information, please contact

Javier Morales, Ph.D.
Assistant Professor in Geo-information Systems and Services
International Institute for Geo-Information Science and Earth Observation
Tel.: +31-053-487-4299
E-mail: jmorales@itc.nl

References

Morales, J., L. Ferreira Pires, and M. van Sinderen, 2002. "Model Driven Geo-Information Systems Development," *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., 2002.

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

About the Author

Javier Morales works as assistant professor in geospatial data infrastructure at The International Institute for Geo-Information Science and Earth Observation (ITC) in Enschede, The Netherlands. He holds an MSc in geoinformatics and a Ph.D. in computer science from the University of Twente. His research interests are in data modeling, architectures for distributed (geoinformation) systems, Web technology, and spatial data infrastructures.