Chaining Data and Visualization Web Services for Decision Making in Information Systems

Ahmet Sayar, Marlon E. Pierce

Chaining Data and Visualization Web Services for Decision Making in Information Systems

Ahmet Sayar[1,*] and Marlon E. Pierce[2]

[1] Kocaeli University, Computer Engineering Department
Umuttepe Campus, 41380, Kocaeli-Turkey
[2] Community Grids Laboratory, Indiana University
2719 East 10th Street Bloomington, IN, 47408, USA

ahmet.sayar@kocaeli.edu.tr, mpierce@cs.indiana.edu

Abstract. Decision making in information systems increasingly relies on analyses of data in visual formats which are created from distributed heterogeneous data belonging to the separate organizations. This paper presents distributed service architecture for creating and managing the production of knowledge from distributed collections of data sources through integrated data-views. Web Services provide key low level capability but do not define an information or data architecture. These are left to domain specific capabilities metadata and domain specific common data model. Datasets are considered to be defined with domain specific spatial and non-spatial attributes for displaying and querying. We propose blueprint architecture and define the principles and requirements for general information systems domains.

1   Introduction

The World Wide Web and its associated Web programming models have revolutionized accessibility to data/information sources. At the same time, numerous incompatible data formats, data heterogeneity (both the data types and storage formats), and machine un-readability of data limit data integration and federation [1]. The seamless integration and sharing of data from distributed heterogeneous data sources have been the major challenges of information system communities and decision support systems [2]. In order to be able to integrate and share data/information, data sources need to be in interoperable formats and provide standard service interfaces interacted with standard message formats and transport protocols. The interoperability issues have

---

* Corresponding author. Tel.: +90-262-3033583
  E-mail address: ahmet.sayar@kocaeli.edu.tr

been studied by many public and private organizations over the last two decades at the data and service levels. Among these are the Web Service standards (WS-I) [3] for cross-language, platform and operating systems, and International Virtual Observatory Alliance (IVOA) [4] and Open Geospatial Consortium (OGC) [5] for defining domain specific data model and online service definitions in Astronomy and Geographic Information Systems (GIS), respectively.

Information systems are mostly used by decision makers coming from various kinds of domains and from various expert levels. Decision making increasingly relies on analyses of data in visual formats such as images, animations, etc. Most of the analyses include creation of integrated data views in which data sources provided by heterogeneous services. By this way decision makers query the heterogeneous data sources as a single resource. This removes the hassle of individually accessing each data source.

Decision making in Information systems mostly require integrating multiple heterogeneous data sets from various heterogeneous resources, and analyzing them from a single access point by visualization tools. There are many ad-hoc solution-frameworks developed for application specific purposes [6,7]. However, no one defines such a framework in terms of requirements and constraints. Katsis et al. [8] defines view-based integration issues from the point of global and local data format heterogeneities. They specify the relationship of sources with the global view. They determine that view-based integration systems follow global as view (GAV), Local As view (LAV) and Global and Local as View (GLAV) approaches. In GAV, the global database (schema) is expressed as a function of the local database (schema). LAV on the other hand follows the opposite direction. Finally GLAV is generalization of the two. Another closely related work is visual integration tool for heterogeneous data types by unified vectorization [9]. They propose a visual data integration technique to aid the end-users to identify semantically similar data attributes in a semi-automated fashion. This is achieved by using Self Organizing Map (SOM) to classify unfamiliar database entities. SOM is mostly used in text visualization related works, and applied in various projects.

We propose a framework for chaining data and visualization Web Services for decision making in information systems. The proposed framework is actually based on the idea of the Data-Information-Knowledge-Wisdom hierarchy (DIKW) [10,11]. In DIKW, wisdom is defined in terms of knowledge, knowledge in terms of information and information in terms of data. The system does not allow actual physical data integration. Physical integration results in both high data maintenance and storage costs. Therefore, it is better to keep data sets at their originating sources and maintain their internal structures. This enables large degree of autonomy. Individual data sets are integrated to the system through mediator services (see Fig. 1).

Web Services provide key low level capability but do not define an information or data architecture. These are left to domain specific capabilities metadata and domain specific common data model. Each-domain has different set of attributes for the data and its model is defined in core language. These are defined by standard bodies of the corresponding domain. Cross-domain interoperability issues are not handled in this paper.
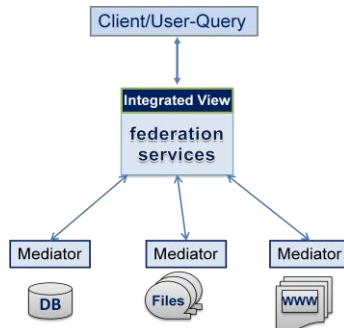
Fig. 1. Integrated data views

The remaining of the paper is structured as follows. Section 2 presents the proposed framework. Section 3 gives the details about the intercommunication and message exchange between the abstract components in the framework. Section 4 scrutinizes the three different information systems domains (GIS, Astronomy and Chemistry) and shows to what extent the proposed framework can be realized. Section 5 presents a sample ASIS framework for GIS domain, and Section 6 concludes the paper.

2   Application Specific Information Systems (ASIS)

This paper proposes blueprint architecture in terms of principles and requirements and calls it Application Specific Information System (ASIS). Developing such a framework requires first defining a core language expressing the primitives of the domain; second, key service components, service interfaces and message formats defining services interactions; and third, the capability file requirements (based on core-language) enabling inter-service communications to chain (link) the services for the federation (see Fig. 2).

The architecture proposed in this paper is a high level architecture that consists of abstract components and explains their data flow and components' interactions. In this section, we focus on the principles and requirements in terms of data and service interfaces in any information system domains. It should be noted that this abstract architecture is intended to be domain-specific. That is, it may be realized in chemistry or astronomy, for example, but we are not suggesting cross-domain interoperability.
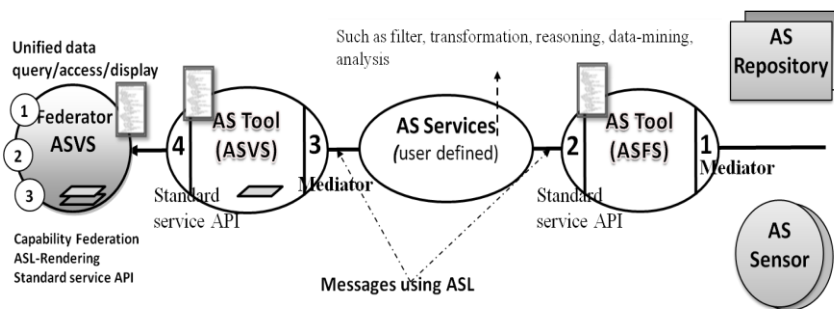

Fig. 2. Application Specific Information System (ASIS)

ASIS consists of two groups of filter-like Web Services. These are Application Specific Feature Service (ASFS), which is actually a data service, and Application Specific Visualization Service (ASVS), which is actually a display service. These services also behave as mediators. ASFS perform their mediation services with adaptors converting any data to common data model encoded in application specific language (ASL) (number 1 in Fig. 2). ASFS also mediate resource specific heterogeneous service interfaces with standard service interfaces (number 2 in Fig. 2). Standard service API enables querying on data (e.g. GetASL) and their metadata (e.g. GetCapabilities). On the other hand, ASVS mediate any data in common data model (ASL) and convert into a visual format (an image) with pre-defined attributes such as bounding boxes, projections and resolutions (number 3 in Fig. 2). ASVS also enables querying of ASL through visual image representations through standard service interfaces (number 4 in Fig. 2). Those interfaces enable performing attribute-based querying of ASL through displays. ASVS also provide service interfaces to list the available ASL data with their attributes, which enables clients to make appropriate queries.

ASIS is a proposed solution to heterogeneous data integration. This solution enables inter-service communication through well-defined service interfaces, message formats and capabilities metadata. Data and service integration is done through capability federation of these services. ASFS and ASVS are Web Services, and each service is described by corresponding generic metadata descriptions that can be queried through Web Service invocations. In addition to allowing service discovery, this approach also enables at least three important qualities of services. First, services of the same type that provide a subset of the request can be combined into a "super-service" that spans the query space and has the aggregate functionality of its member services. Second, the capability metadata can be used to determine how to combine services into filter chains with interconnected input-output ports. Third (and building on the previous two), capabilities of super-services can be broken into smaller, self-contained capabilities that can be associated with specific services. This enables performance gains through load-balancing.

ASFS and ASVS are filter-like Web Services and communicate with each other through capability metadata exchange interface. Capability metadata defines service and data attributes, and their constraints and limitations to enable clients to make valid queries and get expected results. Being a Web Service enables filter services to publish their interfaces, locate each other and chain together easily. Filters have inter-service capabilities and are chainable. If the filter is capable of communicating and obtaining data from other filters, and updates (or aggregates) its capability metadata with these data (after capability files exchange), then it can claim that it serves these data. Filter Services are information/data services that enable distributed data/information access, querying and transformation through their predictable input/output interfaces defined by capability document. Filter located in the same community network can update their capability metadata dynamically through the standard service interface of the filters (e.g. getCapabilities). In addition, dynamically updating capabilities of filters enable removal of obsolete data or down filters. Capabilities metadata and ASL are closely related to each other. One defines the domain-

specific data and other defines the query and response constraints over the service and data provided.

ASVS must visualize information and provide a way of navigating ASFS and their underlying database. ASVS must provide human readable information such as text and graphs (scalable vector graphic (SVG) [12] or portable network graphic (PNG) [13]) images. These are final outputs of the system and might be called as knowledge. ASVS defines the knowledge (formats, types etc.) in terms of what data needs to be rendered from what services. Services are basically ASFS and possibly other ASVS. This definition is done in ASVS's capability metadata. An ASFS is an annotation service providing heterogeneous data in common data model with an attribute-based query capability. ASFS serves data in ASL, which must be realized as a domain specific XML-encoded common data model containing content and representation tags. Heterogeneity in queries and data formats is handled through resource specific mediators. ASL is a domain specific encoding of common data defining the query and response constraints over the service and data provided.

A sample scenario to apply the presented framework can be given in GIS domain. Pattern Informatics (PI) [14] is a project on earthquake forecasting. In PI, data sets are earthquake seismic data records collected from sensors. ASFS provide data records in global data format with attribute and spatial based standard queries. ASFS overlay seismic data records and earthquake forecast values [15] on satellite map images and enable interactive visual querying. Depending on the magnitudes of expected earthquake seismicity, maps are also possibly overlaid with hot-spot plots in colored boxes.

3   Inter-service Communications

Inter-service communication is achieved through common service interfaces and capability metadata exchange. The standard service interfaces can be grouped into three types: a) capability metadata exchange: inter-service communication (set-up stage); b) interactive data display: selecting layer composition and bounding box regions; and c) querying of data itself over the display, getting further information about the data content and attributes.

As mentioned before, capability helps clients make valid requests for its successive queries. Capability basically provides information about the data sets and operations available on them with communication protocols, return types, attribute based constraints, etc. Each domain has different set of attributes for the data and it is defined in ASL common data model. For example, in GIS domain, attributes might be bounding box values (defining a range query for data sets falling in a rectangular region) and coordinate reference system.

Standard requests/query instances for the standard service interfaces are created according to the standard agreed-on request schemas. These are defined by open standards bodies in corresponding domains. The request instances contain format and attribute constraints related to the ASL common data model. For example in the GIS domain, getMap request defines a map images' return format (JPEG, PNG, SVG, etc.), height, width, bounding box values, and so on. Format, height and width are

related to display, but bounding box values are related to the attributes of the data defined in its ASL representation provided by ASFS. In this specific example of the getMap request, ASVS must both visualize information through the getMap service interface and provide a way of navigating ASFS services and their underlying database. ASVS make successive queries to the related ASVS to get the ASL data and render it to create final display for its clients.

In ASIS, the task of mediators is to translate requests to the standard service interfaces to those of the information/data sources', and transform the results provided by the information source back to the ASIS's standard formats. For ASFS, the returned data is ASL, and for ASVS the returned results can be any kind of display format such as images.

Acting as a proxy of information source, the mediators communicate with an information source in its native language and API. They communicate with ASIS in a commonly agreed language (ASL) and Web Service API calls. Because of the obvious heterogeneity between different science domains, each will need to extend and create its own service interfaces, as well as queries for those services. Using common data model-ASL and common services make autonomous resource to be integrated into the system in a manageable way.

The mediators-wrappers enable data sources integrated to the system conform to the global data model (ASL) but enable the data sources to maintain their internal structure. At the end, this whole mediator system provides a large degree of autonomy. Instead of actual physical data federation, system makes distributed querying and response composition on the fly. Mediators perform query and data format conversions.

4   Abstract Components and Matching to Sample Science Domains

GIS is a mature domain in terms of information system studies and experiences. It has standards bodies defining interoperable online service interfaces and data models such as OGC (Open Geospatial Consortium) and ISO/TC211, but many other fields do not have this. We have surveyed two science domains (Astronomy and Chemistry). Table 1 presents the results briefly in terms of service counterparts (ASIS vs. science domains).

Astronomy has a standards body, the International Virtual Observatory Alliance (IVOA), for defining data formats and online services that are somewhat analogous to the OGC standards. FITS (Flexible Image Transfer), Images and VOTable [16] are the data models. SkyNodes are database servers with an ADQL (Astronomy Distributed Query Language) based SOAP interfaces that return VOTable-encoded results. VOPlot and TopCat are two services to visualize the astronomy data in the format of VOTable, FITS and images. VOResource and UCD are the metadata definition and standards for the service descriptions [17].

Chemistry, although a vastly different field, does provide a common data model (CML[18]) that can be used to build up Web Services. Although many research

groups have investigated service architectures for chemistry and chemical informatics, the field has (to our knowledge) no Web Service standards-defining body equivalent to the OGC or IVOA.

Table 1. Components and common data models matching for sample science domains

Science Domains	Common Data Model (ASL)	Components		Metadata	Standard Bodies
		ASFS	ASVS		
GIS	GML	WFS	WMS	capability.xml schema	OGC and ISO/TC211
Astronomy	VOTable, FITS	SkyNode	VOPlot TopCat	VOResource	IVOA
Chemistry	CML, PubChem	None	NO standard JChemPaint, JMOL	None	None

5   ASIS Application to GIS Domain

Data in GIS are stored in various places using various formats. These formats may be closed (i.e., proprietary dependent) or opened, normalized or not. A consequence is that the analysis and/or decision-making may be difficult or time-consuming due to this heterogeneity of formats. To avoid these drawbacks and in order to facilitate accesses and uses of these data, the Open Geospatial Consortium (OGC) defined some propositions. These propositions allow accesses to geographical data without taking into account physical aspects of communications (i.e., with an URL). They also propose a set of parameters, methods and communication rules to simplify accesses and manipulations as soon as clients and servers respect them. The Web Map Service (WMS) and Web Feature Service (WFS) are two well-known web services defined by OGC standards. WMS deal with dynamic production of maps as images, built with geo-referenced data. WFS are designed to provide an access and a manipulation tool of geographical data within a map. WMS and WFS provide standard service interfaces and their operations provide data associated with a geographical map or object. Information may be alphanumerical such as a town name or graphical such as the segments defining a border of a town.

WMS correspond to ASVS in the proposed framework. WMS standard induces three main operations. Fig. 3 presents these operations: GetCapabilities, GetMap, GetFeatureInfo. The most important facility for WMS is to provide a map from its available layers. Parameters such as desired layers, representation styles, size, relevant geographical areas, the projection system or the output format must be provided to the GetMap operation. The obtained result is a map, which is an image that can be displayed by conventional web browsers. This operation corresponds to a visualization process.

The GetFeatureInfo operation corresponds to a query process. An end-user does not manipulate structured data but an image. To be able to access data, an end-user must

select an object on the image. The GetFeatureInfo operation provides this selection. Nevertheless, some limits appear in the sense that the provided schema is the set (or a sub-set) of available data for this/these object(s). No link is performed with the environment of this object. Furthermore, an end-user must be aware of the database schema in order to select only a sub-set of available data. This information is kept in Capabilities.xml document and provided by GetCapabilities service interface upon requests.

WFS in Fig. 3 correspond to ASFS in the proposed framework. WFS are OGC specifications for data access. It describes responses of a Web server to geographical data manipulation operations. These operations are based on the CRUD manipulations of geographic data based on alphanumeric/spatial constraints: creation (C), read (R), update (U) and deletion (D). The formalism used to model data exchanges for the WFS specification is GML (Geography Markup Language). GML correspond to ASL in the proposed framework. GML is a common data model for geographic data, and defined by OGC specifications. It is an XML dialect and is designed to encode, to manipulate and to exchange geographical data. The WFS specification defines five operations to send queries to a geographical data server and to get answers from it: GetCapabilities, DescribeFeatureType, GetFeature, Transaction and LockFeature. Fig. 3 presents the basic operations of the WFS specification.

The most important operation for a WFS server is the GetFeature operation. This operation delivers data instances typed by features, identifies properties that should be delivered and provides the results of spatial and non-spatial queries. Selection criteria are defined by using a filter. To get relevant data, a client specifies an object identifier in the filter. The WFS server receives the GetFeature query, determines the correct database, creates and sends a SQL statement to the database and formats the results. The filter is used to manage the "Where" clause of a SQL statement. Spatial data are handled with the spatial schema defined in the ISO 19107 norm. Before sending results to a client, the server transforms objects using a GML format.

In summary, Fig. 3 illustrates an application of ASIS in GIS domain. Such a framework has been used in some Geo-science projects [19-21] with application specific extensions and modifications.
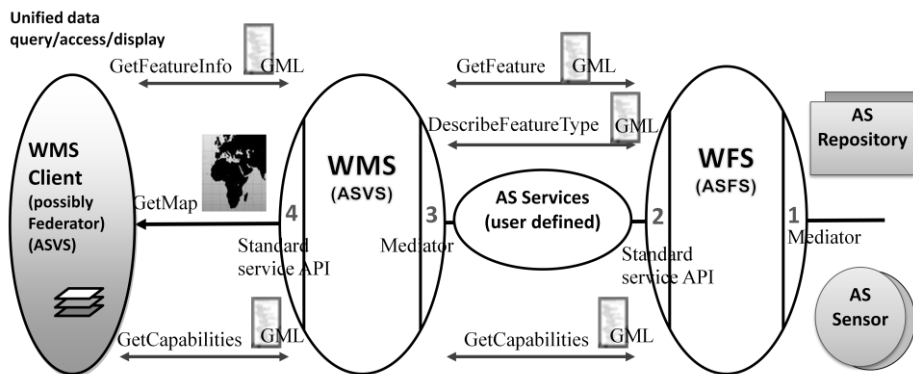

Fig. 3. Application of ASIS to GIS Domain

6   Conclusion

We have presented a service-oriented architecture for understanding and managing the production of knowledge from the distributed observation, simulation and analysis data through integrated data-views in the form of images. It enables unified data access/query and a display from a single access point. The infrastructure is based on a common data model, standard Web-Service components for data and visualization, and a federation service. The framework applies just-in-time (late-binding) federation in which the data is kept in its originating sources all the time. This enables autonomy and easy data maintenance.

The architecture proposed in this paper is a blueprint architecture defining the principles and requirements for application specific information systems for general science domains. We focus on the principles and requirements in terms of data and service interfaces in any information system domains. We have defined two general service types (ASFS and ASVS) with limited number of service interfaces. Service interfaces enable metadata exchange and data querying. Data flows from databases to users through ASFS and then ASVS. It should be noted that this abstract architecture is intended to be domain-specific. That is, it may be realized in chemistry or astronomy, for example, but we are not suggesting cross-domain interoperability. Due to the domain specific data heterogeneity, each domain should define its own ASL and corresponding queries.

References

1. Huang, Y., Cai, G., Wang, G.: Dynamic Service Composition Based on Federated Agents. Energy Procedia 13, 5536 – 5543 (2011).
2. Sayar, A., Pierce, M., Fox, G.C.: Unified Data Access/Query over Integrated Data-views for Decision Making in Geographic Information Systems. In: Bessis, N. (ed.) Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations. p. 360. IGI Global - Information Science Reference Bedfordshire, UK (2009)
3. WS-I: WS-I. http://www.ws-i.org/ (2012). Accessed 03/23/2012 2012
4. IVOA: International Virtual Observatory Alliance. http://www.ivoa.net/ (2012). Accessed January 12, 2013 2013
5. OGC Schema. http://schemas.opengis.net/ (2008). Accessed 09/14/2008 2008
6. Chan, B., Talbot, J., Wu, L., Sakunkoo, N., Cammarano, M.: Vispedia: Ondemand Data Integration for Interactive Visualization and Exploration. In: ACM SIGMOD International Conference on Management of data, Rhode Island USA, June 29-July 2, 2009 2009, pp. 1139-1142. ACM
7. Cammarano, M., Dong, X.L., Chan, B., Klingner, J., Talbot, J., Halevy, A., Hanrahan, P.: Visualization of Heterogeneous Data. IEEE Transactions on Visualization and Computer Graphics 14(6), 1213-1220 (2008).
8. Katsis, Y., Papakonstantinou, Y.: View-based Data Integration. Encyclopedia of Database Systems, 3332-3339 (2009).
9. Bourennani, F., Pu, K.Q., Zhu, Y.: Visual Integration Tool for Heterogeneous Data Type by Unied Vectorization. In: IEEE International Conference on Information Reuse & Integration (IRI 2009), Las Vegas, NV 2009, pp. 132 - 137. IEEE
10. Zins, C.: Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology **58**(1), 479-493 (2007).
11. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science 33(2), 163-180 (2007). doi: 10.1177/0165551506070706
12. Andersson, O., others: Scalable Vector Graphics (SVG) Specification Version 1.1. In. World Wide Web Consortium (W3C), (2003)
13. Adler, M., Boutell, T., Bowler, J., Brunschen, C., Costello, A.M., Crocker, L.D., Dilger, A., Fromme, O., Gailly, J.-l., Herborth, C.: Portable Network Graphics Specification (PNG). In., vol. REC-PNG-20031110 (2003)
14. Holliday, J.R., Chen, C.-c., Tiampo, K.F., Rundle, J.B., Turcotte, D.L., Donnellan, A.: A RELM Earthquake Forecast Based on Pattern Informatics. Paper presented at the American Geophysical Union (AGU) - fall meeting, San Francisco, California, 12/2005
15. Rundle, J.B., Turcotte, D.L., Shcherbakov, R., Klein, W., Sammis, C.: Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. Geophysics 41(4) (2003). doi:10.1029/2003RG000135

16. Williams, R., Ochsenbein, F., Davenhall, C., Durand, D., Fernique, P., Giaretta, D., Hanisch, R., McGlynn, T., Szalay, A., Wicenec, A.: VOTable: A Proposed XML Format for Astronomical Tables. In. US National Virtual Observatory, (2002)

17. Yasuda, N., Mizumoto, Y., Ohishi, M., O'Mullane, W., Budav´ari, T.a., Haridas, V., Nolan Li, T., Malik, A.S., Hill, M., Linde, T., Mann, B., Page, C.: Astronomical Data Query Language: Simple Query Protocol for the Virtual Observatory. Paper presented at the Astronomical Data Analysis Software and Systems XIII. ASP Conference Series, ASP Conf. Series, San Francisco, USA,

18. Holliday, G.L., Murray-Rust, P., Rzepa, H.S.: Chemical markup, XML, and the world wide web. 6. CMLReact, an XML vocabulary for chemical reactions. Journal of chemical information and modeling 46, 145-157 (2006).

19. Aydin, G., Sayar, A., Gadgil, H., Aktas, M.S., Fox, G.C., Ko, S., Bulut, H., Pierce, M.E.: Building and Applying Geographical Information Systems Grids. Concurrency and Computation: Practice and Experience 20(14), 1653-1695 (2008).

20. Sayar, A.: High Performance Federated Service Oriented Geographic Information Systems. http://gradworks.umi.com/33/44/3344771.html (2009). PhD

21. Pierce, M.E., Fox, G.C., Aktas, M.S., Aydin, G., Qi, Z., Sayar., A.: The QuakeSim Project: Web Services for Managing Geophysical Data and Applications. Pure and Applied Geophysics (PAGEOPH) 165(3-4), 635-651 (2008). doi:10.1007/s00024-008-0319-7