The Brokering Approach for Earth Science Cyberinfrastructure

Stefano Nativi[1], SiriJodha Khalsa[2], Ben Domenico[3], Max Craglia[4], Jay Pearlman[5], Paolo Mazzetti[1], Russ Rew[3]

[1] National Research Council of Italy -Institute of Atmospheric Pollution Research (CNR-IIA)

[2] National Snow and Ice Data Center (NSIDC), University of Colorado

[3] Unidata, University Corporation for Atmospheric Research (UCAR)

[4] EC-Joint Research Center (JRC) -SDI Unit

[5] IEEE

Abstract

There are significant benefits in relieving individual information systems from the complexities of implementing interoperability in the environment where interdisciplinary cooperation is key to scientific understanding. These include increased information access and use, sustainability of cross-domain discovery and facilitating future technology insertion in a consistent manner. The Brokering approach to interoperability uses a central mechanism to converge disparate vocabularies, promote translatable standards and enable uniformity of search and access in divergent operating environments. The broker approach and implementation is maturing rapidly with embedded capabilities to include new technologies. Already, the Brokering approach can address web 2.0 interfaces, support for model webs and advanced information flows. As a proposed element of an Earth Cube Data Infrastructure, we argue that the Broking approach supports the current and future needs of the science and applications communities.

Introduction

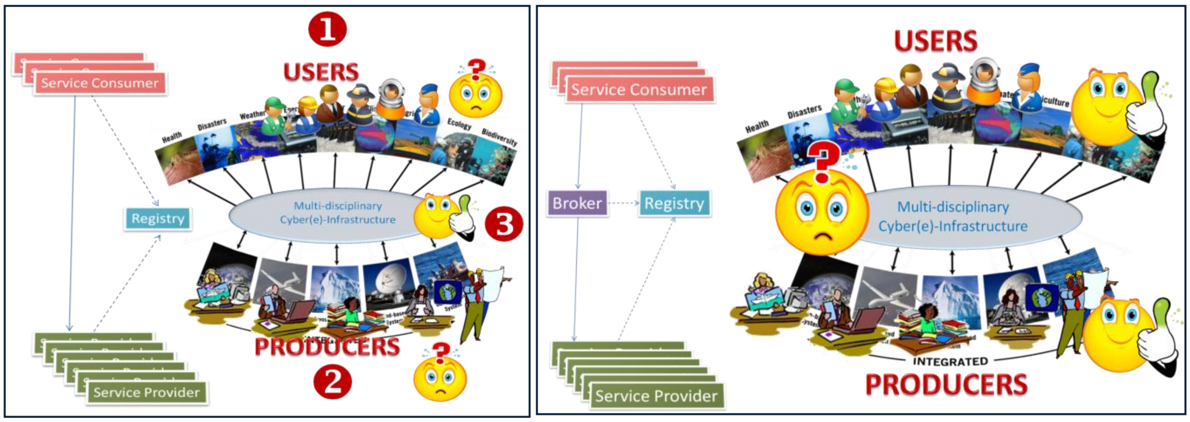
Applying the SOA approach

The development of multi-disciplinary cyberinfrastructures that build on existing capacities is becoming more and more essential to studying the Earth system, addressing global change, global multi-disciplinary analyses and supporting decision makers. In recent years, many important programs and initiatives have tried to address these ambitious objectives. These programs recognize the needs of three main stakeholder types within the science and broader communities:

1. Users;
2. Data/Information Producers;
3. Cyber(e)-Infrastructure Architects and Managers.

The goal of Infrastructure Architects is to interconnect Users and Data/Information Producers in diverse communities and contexts. Almost invariably these architectures resemble the Service Oriented Architecture (SOA) archetype: Users play the role of *service consumers*, Data Producers of *service* *providers*, while the cyberinfrastructure acts as a *service clearinghouse* or *service registry* (see Figure 1 a). In keeping with SOA, this approach applies the following principles:

- Keeping the cyberinfrastructure level as simple as possible;
- Use of a common data model;
- Using a few select international standards for interoperability protocols (i.e. Web services protocols).



(a) (b)

Figure 1. Implementation/maintenance load for current Multi-disciplinary CybeInfrastructure (a) and Brokering approach (b)

According to this approach, a multi-disciplinary cyberinfrastructure consists of a set of common standards and rules (e.g. best practices) adopted by both Data Producers and Users to publish available resources, discover them and bind to them. Infrastructure registry services and a general portal acting as user interface are commonly implemented at this infrastructure level.

Architectures implementing this pattern are known as Federated Architectures. In fact, they are based on the concept of "common/federal data model" and "common federal interoperability protocol(s)". Most Spatial Data Infrastructures (SDIs) around the world are based on this model (see for example the National Spatial Data Infrastructure in the USA[1], and INSPIRE[2] in Europe. Examples in the geosciences are GEON[3], CUAHSI[4] and OneGeology[5]).

SOA shortcomings

The SOA approach works well in consistent and controlled frameworks (e.g. enterprise environments), or where the approach is embedded in a strong legal framework that makes it mandatory for stakeholders to adopt the agreed standards and protocols (e.g. the NSDI for US federal government, and INSPIRE in Europe). Even in these strongly controlled environments, the transition from the discovery of information resources to their access and use (particularly if in multi-national, and multi-cultural contexts) increases the importance of semantic interoperability, along with the challenge of establishing agreed policy frameworks for data access and use. Another impediment stems from the nature of SOA: a rather complex technology mainly focused on machine-to-machine connections, lacking intuitive human-guided service interactions and compositions. An additional significant

[1] http://geo.data.gov
[2] http://www.inspire-geoportal.eu/
[3] http://www.geongrid.org
[4] http://www.cuahsi.org
[5] http://portal.onegeology.org/

challenge is the level of effort required by both data producers and consumers to participate in SOA and upgrade their capabilities with evolving technologies and needs. These challenges apply both to generic SDIs and to special-purpose systems designed to serve a particular scientific community.

Entry Barriers

It has become apparent that regional and global Earth Science/Observation frameworks demand complex infrastructures that cannot be managed as enterprise environments. Important lessons have been learned by cross-disciplinary initiatives like GEOSS (Global Earth Observation System of Systems) [6] which has faced difficulties in getting participants to utilize the infrastructure that has been put in place.

Keeping the cyberinfrastructure level as simple as possible, for example by advocating a common data model and few interoperability protocols, entails loading the infrastructure complexity almost entirely on Data Producers and Users . Such burdens are known as Data Producers and User entry barriers and reduce participation. For most of the existing multi-disciplinary cyberinfrastructures, these barriers are quite high; both Producers and Users do not seem to be willing to invest precious resources to become compliant. On the contrary, they are focused on developing disciplinary and thematic capacities that increase the need for cross-organizational, multi-disciplinary infrastructures.

The Brokering approach

To lower the entry barriers to both Data Producers' and Users, it is important to address the complexity at the cyberinfrastructure level. This approach requires extending the SOA archetype by introducing a new component: the Broker (see Figure 1 b). Discovery and binding complexity due to heterogeneity are addressed by the Broker component, thus easing Users' and Producers' burden.

Principles

The Brokering approach applies the following principles:

- Keep the existing capacities as autonomous as possible by interconnecting and mediating standard and non-standard capacities.
- Supplement but not supplant systems mandates and governance arrangements.
- Assure a low entry barrier for both resource users and producers.
- Be flexible enough to accommodate existing and future information systems and information technologies.
- Build incrementally on existing infrastructures (information systems) and incorporate heterogeneous resources by introducing distribution and mediation functionalities to interconnect heterogeneous resources.

Thus, Brokering components are needed to realize all the necessary mediation, adaptation, distribution, semantic mapping, and even quality checks required to address the complexity of the cyberinfrastructure.

[6] http://www.earthobservations.org/geoss.shtml

Brokering service types

The introduction of the Broker component shifts the SOA binding architecture from a two-tier (User-Provider) to a three-tier (User-Broker-Producer) solution. Like an "application server," Brokers can do much more than just facilitate discovery and access of available resources: they may implement new capabilities for Users and Data/Information Providers such as:

- Advanced/semantic discovery;
- Resource tagging and clustering of results;
- Quality control;
- Data pre-processing and transformation (not only download);
- Work flow management.

In a cyberinfrastructure, different Brokers can be developed to facilitate different capabilities such as: Discovery, Access, Semantic interoperability, Service Composition, etc. This facilitates initial implementation, future upgrades and long-term evolution.

The EuroGEOSS project[7] has investigated in depth the Brokering approach by implementing: a Discovery Broker, an Access Broker, and a Semantic Discovery Broker. A service Composition Broker is under development in the framework of another project: UncertWeb[8]. Incorporating both OGC and other international discovery standards, the Brokering approach has been used by several communities such as Meteo-Ocean, hydrology, climate, etc, to access THREDDS and other resources through a standard OGC CSW interface. Continued application development in Europe is being funded by the GEOWOW project.

Broker Functionalities

A Broker is a service component that implements the following functional modules:

a. User's request Distribution (e.g. searches and Composition of services) in an asynchronous way (providing consistent feedbacks to Users);
b. Mediation (e.g. translation from the many Producer and Users service models to the broker internal model, and vice-versa);
c. Adaptation (e.g. between the many Producer service protocols and the Users' ones, and vice-versa);
d. Specific Added-value functionalities.

A Broker implements added-value functionalities related to its specific scope: Discovery, Access, Semantic expansions, etc. For instance, the EuroGEOSS Discovery Broker achieves tagging and clustering of all the heterogeneous results returned by the many brokered data providers. In the GeoViQua project[9], the Discovery Broker will implement quality check and ranking. These are application-level services are provided by the infrastructure to enrich the basic brokering functionalities. This is a unique capability offered by the Brokering approach, in comparison with other architectural solutions.

[7] www.eurogeoss.eu
[8] www.uncertweb.eu
[9] http://www.geoviqua.org/

Additionally, a broker must also implement functional modules for:

e. Sessions management;
f. Configuration (brokering parameters configuration at the level of single provided resource);

Through these functionalities, the Broker makes it possible to connect information resources from multiple domains, building bridges among scientific communities and supporting the multi-disciplinary scientific enquiry, which is key to the progress of science.

Discussion

This white paper proposes that Earth Cube focus on translating and adapting these brokering services for the geoscience research community to yield capabilities and benefits beyond what traditional SOA architectures can deliver. For example, brokering services can be an important bridge with Web 2.0 resources. In the EuroGEOSS the discovery broker implements well accepted profiles of OpenSearch (the Web 2.0 discovery protocol); thus, Web 2.0 users may utilize the Broker to access "traditional" Web products (e.g. ISO/OGC resources) and vice-versa. From the work in UncertWeb, brokering solutions are advancing the capabilities that address an important challenge for Geosciences: environmental model access and interoperability. What is most important in moving toward a Broker-based infrastructure is that it offers a greater level of flexibility than other architectural solutions. This adds not only an avenue for innovation with technology evolution, but also provides the potential for interoperability with cultural, social and economic information that will ultimately play a role in decision making and transferring the impacts of science and research.

The Brokering approach lowers User and Data Producers entry barriers considerably. This approach can offer the opportunity to support real button-up infrastructure, building on existing capacities (supplementing and not supplanting them). The Brokering approach is mature: discovery Broker solutions are used by several Communities (e.g. Meteo-Ocean, Hydrology, Climate, etc.) to access THREDDS[10] resources through standard ISO/OGC interfaces. Recently, the GEO Architecture and Data Committee decided to include the EuroGEOSS Broker to augment the capabilities of the GEOSS Common Infrastructure. In fact, a brokering solution doesn't impose any common/federal model but is able to implement different federal/common solutions and mediate between them. Hence, the brokering solution may provide a clear advantage over federated and other SOA solutions.

We recognize that with the brokering approach there is a single point of failure in the information system. However, this risk is present in most of the "ultimate" and successful platforms, including: GoogleMap/Earth; Eye on Earth, Cloud Computing platforms, etc. In fact, they apply almost the same principles recognized by the Brokering approach: to shift complexity from Users and Resource Providers to the infrastructure/platform. Their experiences showed that this risk is addressed by developing better and more reliable infrastructures (e.g. adopting caching technologies, computational scaling solutions, Web optimizations, declarative approaches, modular software components).

Bibliography

Craglia,M., Nativi, S., Santoro, M., Vaccari, L., Fugazza, C., 2011. Inter-disciplinary Interoperability for Global Sustainability Research. Proceedings of Geos 2011 Conference, Lecture Notes in Computer Science, Springer edition, 1-15.

[10] http://www.unidata.ucar.edu/projects/THREDDS/

Mazzetti, P., Nativi, S., 2011. Multi-disciplinary interoperability for Earth observation sensor web, Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, 978-1-4577-1005-6/11.

S. Nativi, M. Santoro, L. Vaccari, M. Craglia, M. Grasso, 2011. Specification of EUROGEOSS Initial Operating Capacity, EuroGEOSS Deliverable D.2.2.2.b, March 2011.

J. Pearlman, M. Craglia, F. Bertrand, G. Dubois, S. Fritz, G. Gaigalas, S. Nativi, and S. Niemeyer, 2011, Interdisciplinary Approaches for Decision-making Information Across Forestry, Biodiversity and Drought, Proceedings of EGU General Assembly 2011, Geophysical Research Abstracts Vol. 13, EGU2011.

M. Santoro, G.Dubois, J. O. Skoien, S. Peedell, C. Fugazza, S. Nativi, and M. Craglia, 2011, Progress towards the setting up of a multidisciplinary infrastructure for ecological modeling with eHabitat, Proceedings of EGU General Assembly 2011, Geophysical Research Abstracts Vol. 13, EGU2011-3532-4

M. Santoro, P. Mazzetti, C. Fugazza, S. Nativi , and M. Craglia, 2011, Enhancing Discovery Capabilities of SDIs with Semantic Queries, Proceedings of EGU General Assembly 2011, Geophysical Research Abstracts Vol. 13, EGU2011-5963-4.

S. Nativi, P. Fox, 2010. "Advocating for the use of informatics in the earth and Space Sciences", EOS, Transactions, American Geophysical Union, Vol. 91, No 8, Page(s): 75 – 76.

L. Vaccari, S. Nativi, M. Santoro, 2010. Report on requirements for multidisciplinary interoperability, EuroGEOSS Deliverable D.2.1.1, Oct 2010.

Ramapriyan, H.K.; Rochon, G.L.; Duerr, R.; Rank, R.; Nativi, S.; Stocker, E.F.; 2010. Advances in spatial data infrastructure, acquisition, analysis, archiving & dissemination, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, Digital Object Identifier: 10.1109/IGARSS.2010.5653074.

M. Santoro, E. Boldrini, F. Papeschi, A. Baldini, S. Nativi, 2010. Specification of EUROGEOSS Initial Operating Capacity, EuroGEOSS Deliverable D.2.2.2.a, Feb 2010.

Maidment, D R, Salas, F, Domenico, B, Nativi, S. (2010), Crossing the Digital Divide: Connecting GIS, Time Series and Space-Time Arrays (Invited), Abstract IN13A-1095 presented at 2010 Fall Meeting, AGU, San Francisco, Calif., 13-17 Dec.

S. Nativi, L. Bigagli, 2009."Discovery, Mediation, and Access Services for Earth Observation Data", Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of Volume 2, No 4, Page(s):233 – 240.

Khalsa, S.J., Nativi, S., Geller, G., 2009. The GEOSS Interoperability Process Pilot Project (IP3), IEEE Transactions on Geoscience and Remote Sensing special issue on data archiving and distribution, vol. 47, num. 1. January 2009, pp. 80-91.

Nativi, S.; Bigagli, L.; Mazzetti, P.; Boldrini, E.; Papeschi, F., 2009. GI-Cat: A Mediation Solution for Building a Clearinghouse Catalog Service. Advanced Geographic Information Systems & Web Services, 2009. GEOWS '09. International Conference on Digital Object Identifier: 10.1109/GEOWS.2009.34.

L. Bigagli, S. Nativi, P. Mazzetti, 2006. "Mediation to deal with information heterogeneity", Advances in Geosciences, Journal of the European Geosciences Union, Page(s) 3-9. SRef-ID: 1680-7359/adgeo/2006-8-3. June 2006.