The Brokering Approach for Multidisciplinary Interoperability: A Position Paper*

Stefano Nativi[1], Massimo Craglia[2], Jay Pearlman[3]

[1] National Research Council of Italy -Institute of Atmospheric Pollution Research (CNR-IIA) stefano.nativi@cnr.it

[2] European Commission Joint Research Center–Digital Earth and Reference Data Unit massimo.craglia@jrc.ec.europa.eu

[3] IEEE jay.pearlman@ieee.org

Abstract

Global sustainability research requires an integrated multi-disciplinary effort underpinned by a cyber infrastructure able to harness big data and heterogeneous information systems across disciplines. Two approaches are possible to achieve the interoperability desired across such systems and data: federating, and brokering. This position paper argues that the former is appropriate to single discipline or domain environments, but that brokering is more scalable and effective in complex multi-disciplinary domains. The paper identifies the principles of brokering, and gives examples of practical implementation relating to data discovery, semantic searching, and data access achieved in the EuroGEOSS project. The value of the EuroGEOSS brokering approach has been demonstrated in extending the data resources available through the Global Earth Observation System of Systems (GEOSS) from a few hundred to over 28 million in a matter of 3 months. Brokering offers therefore a real chance to facilitate truly multi-disciplinary big data science and address the scientific challenges of our time.

Keywords: interoperability, web services, global earth observation, multi-disciplinarity.

1.   INTRODUCTION

One of the key research challenges of our time is to increase our understanding of the complex relationships between environment and society, the risks associated with climate change, and the actions we can take to mitigate these risks or cope with those we cannot manage. The International Council for Science (ICSU) calls this field "global sustainability research" (ICSU, 2011) and argues that to meet this challenge it is necessary to have a stronger involvement and greater integration of the social sciences, health sciences, engineering and humanities, along with the natural sciences. In other words, there is a strong need to move from disciplinary research to multidisciplinary research. This effort needs to be underpinned by a computational and data cyber (e)-infrastructure (Nativi and Fox, 2010) harnessing the vast quantities of heterogeneous multi-domain data available, also referred as "Big Data" (Manyika et al. 2011; Hilber and Lopez, 2011). Such a cyber(e)-infrastructure requires a strong degree of interoperability among the multitude of different information systems and data infrastructures that have developed in each discipline over the last decades. This paper focuses on this multi-disciplinary interoperability challenge (Craglia et al. 2011; Mazzetti and Nativi, 2011; Vaccari et al. 2012), and argues that when multiple disciplines and communities of practice are involved, an approach based on mutual respect and mediation (Nativi and Bigagli, 2009), also referred to as brokering approach, is more feasible and fruitful than one based on standardization or federation.

Network-based systems and applications are commonly achieved by implementing a stateless Client-Server (C-S) architectural style. In this style, server components offering a set of services, listen for requests upon those services, while client components desiring a service send a request to the specific server via its interface (Fielding, 2000). The Service-Oriented Architecture (SOA) archetype is a well-used interpretation of this interoperability model. Another notable C-S architectural style is the Representational State Transfer (REST) one for distributed hypermedia systems. It implements a "Layered-Client-Cache-Stateless-Server" architectural style characterized by a uniform interface (Fielding, 2000). Resource/REST Oriented Architecture (ROA) technological guidelines are commonly used to implement architectures that are conformant to the REST style (i.e. RESTful). For the C-S architectures, an important role is played by the connecting interfaces. They can be "uniform" (like for the REST) or "component specific" (like for the Web Services architecture using WSDL (W3C, 2004)). Interface characteristics include message payload models and encodings in addition to message protocol types and bindings. They define all the necessary characteristics to exchange network messages as well as understanding and correctly using the data contained by those messages.

To enable interoperability C-S style, interfaces are objects of standardization at different levels, e.g. at the system, application, or infrastructure levels. The collection of these interface standards (i.e. interoperability arrangements) defines the "service bus" characterizing a network-based system, application, or infrastructure. A service bus may be defined as the middleware glue that holds a C-S architecture together and enables communication between network-based applications (Ortiz, 2007; Schmidt et al., 2005). By defining and using a "service bus", a given community realizes interoperability for enabling a domain application or a disciplinary infrastructure. The service bus philosophy is adopted by the federated systems introduced by several communities and initiatives. In fact, they use the "service bus" concept to mandate a limited set of standard specifications using a community (or federal) model to interconnect system components (i.e. clients and servers). In effect, federated systems address interface heterogeneity by pushing common standards adoption.

Interoperability has been achieved by several geosciences communities within their own domains by advancing a set of disciplinary (or "federal") service buses. As depicted in Figure 1, they are disciplinary infrastructures which implement real "enterprise platforms" based on common/federal specifications covering important interoperability aspects, such as: data and metadata models; encoding formats and languages; controlled vocabularies and ontologies; service interfaces and binding protocols; and data policies.

The INSPIRE Directive in Europe (EC 2007) is an example of such a federated approach based on standards and protocols agreed through a broad participative process involving representatives of the EU Member States. This approach is possible because it is backed by a legal framework that mandates the standards, protocols and interfaces agreed so that the infrastructures of the 27 Member States can work together as one (see INSPIRE Network Services Drafting Team 2008).

Different disciplinary cyber-infrastructures have defined different service buses, even if some popular solutions to achieve interoperability of services (e.g. the ISO TC211 and OGC standards for geospatial information) are often utilized. To overcome some of the difficulties associated with imposing a single common/federal model (or service bus), it was decided to integrate across such approaches by introducing mediation (Bigagli et al. 2006; Nativi and Bigagli, 2009; Nativi et al. 2009). Mediation was first used to map from existing (and well-adopted) specifications to the federal ones. This was useful to allow the "federation" of existing and legacy capacities trying to avoid high costs and to lower entry barriers represented by the need to implement difficult and often unknown standards.

Figure 1 - Ecosystem of infrastructures based on disciplinary service buses
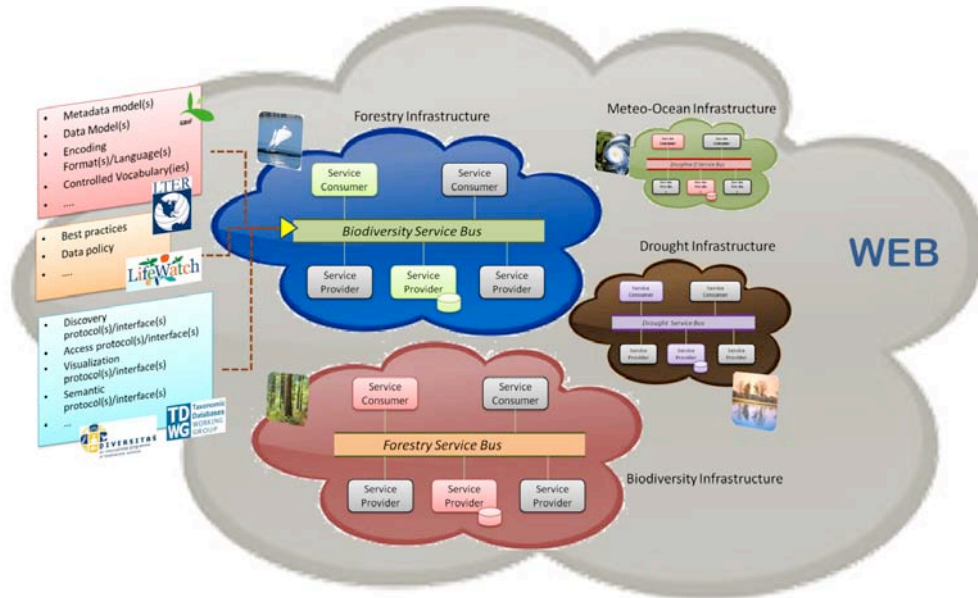

2. AN ECOSYSTEM OF FEDERATED SYSTEMS: THE MULTIDISCIPLINARY CHALLENGE

To implement interoperability for a complex (i.e. global and heterogeneous) domain like the Geosciences, it is necessary to address an ecosystem of information and service platforms (as shown in Figure 1).

Therefore, an application dealing with Earth System Science or Global Change that needs to access information and data managed by different communities is required to implement all their services buses as shown in Figure 2. This is also the case for all the other Communities/infrastructures that are not included in the figure, for the sake of space, but may contribute to Earth System Science studies and may help for applications such as the mitigation of the effects of Global Change.

As shown by Figure 2 in a disciplines ecosystem, entry barriers for multidisciplinary users (including software applications) are high. In fact, a user accessing data and information served by different community infrastructures needs to implement interface adaptors for the respective service buses and maintain these implementations up-to-date. In a global and heterogeneous context (like GEOSS for example), this might entail the implementation of tens of different specifications (i.e. standards and best practices).

Figure 2 - Entry level barriers for a multidisciplinary application in a traditional C-S architectures: a client application needs to support all the different disciplinary service buses to utilize datasets belonging to them.
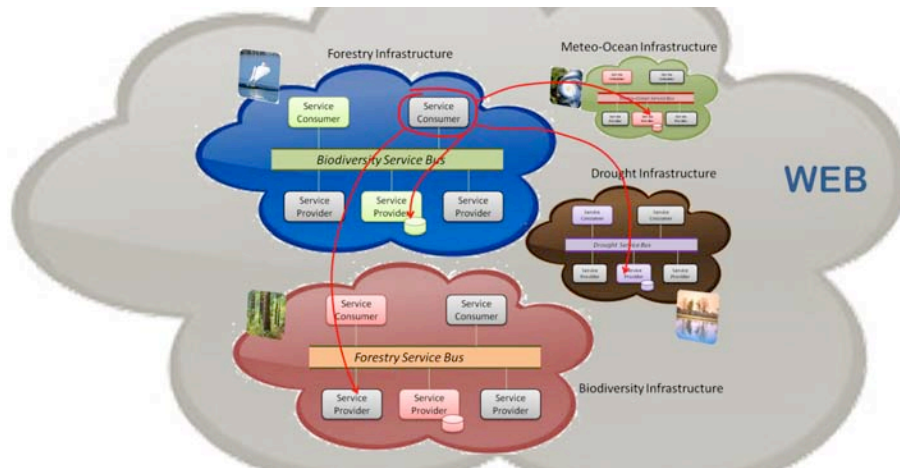

Another possible solution would be to require data providers to publish their data on all the different infrastructures by implementing the heterogeneous specifications applied by the disciplinary service buses, as depicted in Figure 3. Clearly, this would significantly raise the entry level barrier for data providers.

Figure 3 - Entry level barriers for data providers in a multi-disciplinary environment in a traditional C-S architectures: to make its datasets usable by different communities, a data provider needs to implement all the different disciplinary service buses.
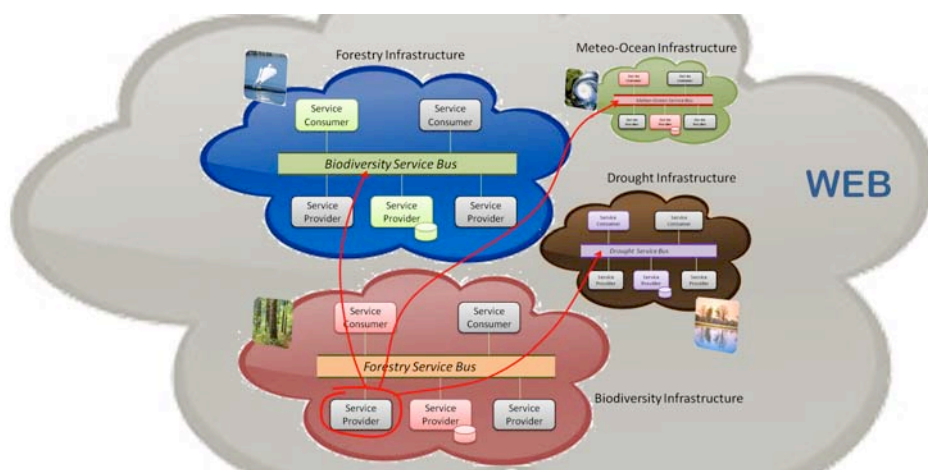
In a global and multidisciplinary context like the Global Earth Observation System of Systems (GEOSS)[2] or EarthCube[3], adopting a federation approach would mean introducing a global and multidisciplinary service bus incorporating a common set of model and protocol specifications to be implemented by all the disciplinary systems globally -as depicted in Figure 4.

Figure 4 - Entry level barriers for a multidisciplinary application in a traditional C-S architectures: a federated system approach
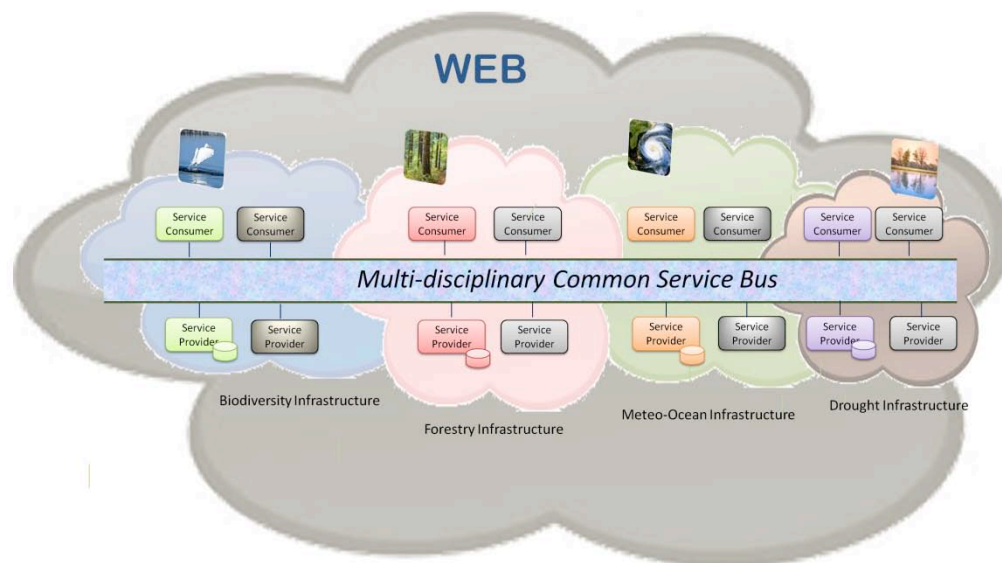
Overlaying a federated solution on the global Earth Science community raises many questions. Would this be feasible? Would it preserve Communities autonomy addressing their specific interoperability needs? Would it be flexible and scalable enough to implement effective multi-disciplinary resources sharing? Would this be too expensive for resource providers and users who are asked to implement hundreds of mediation services and to maintain them?
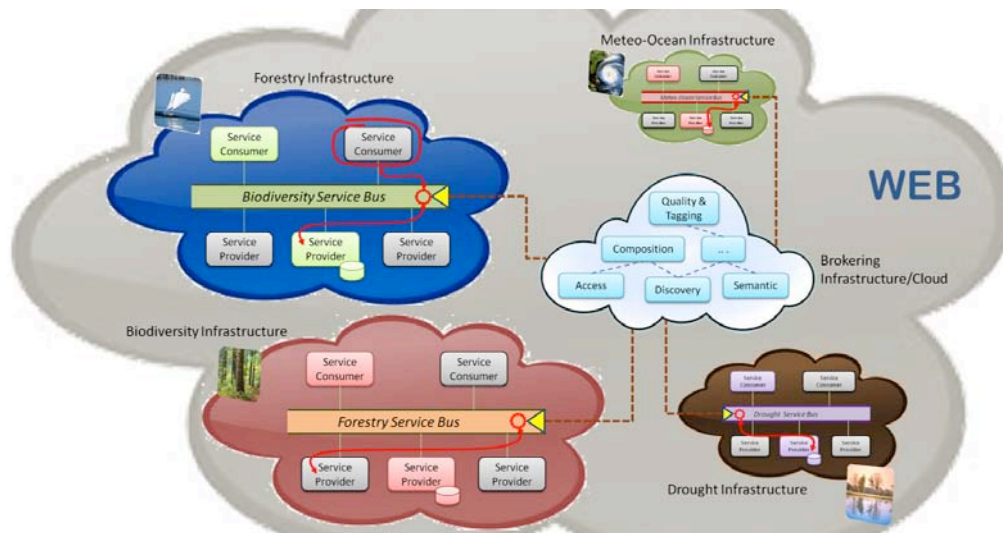
3.   THE BROKERING APPROACH

A solution to reduce the interoperability burden on data providers and applications is to introduce a third party to interconnect the different service buses, mediating their existing (and future) models and interface specifications.

---

[2] http://earthobservations.org/index.shtml
[3] http://www.nsf.gov/geo/earthcube/

In addition, this party must work out all the necessary distribution and virtualization capabilities to lower the entry barriers for multidisciplinary applications. According to Figure 5, this intermediary role is played by an infrastructure realising a Brokering approach; for this reason, it is called "Brokering Infrastructure" (Nativi et al. 2011). The Brokering Infrastructure may implement several functionalities at the multidisciplinary level, such as traditional discovery and access; in addition, it is the place to work out advanced and complex functionalities, like semantic finding, services composition, resources tagging, and quality-based ranking. The diffusion of cloud technology seems to be particularly suited to empower this kind of infrastructure.

Figure 5 - Lowering the entry level barriers by introducing a Brokering infrastructure: a multidisciplinary application implements only the service bus of the infrastructure it belongs to. The other disciplinary service buses are brokered; the other community datasets appear as virtual resources of the application infrastructure.

While the federation approach has gained popularity to implement disciplinary infrastructures, the brokering approach is fit to build multidisciplinary and complex systems -- like an ecosystem of infrastructures. Thus, the brokering approach applies a set of interoperability principles which reflect the aim of lowering entry barriers for multidisciplinary applications and accepting community heterogeneity:

1. To supplement, without supplanting, the individual systems' mandates and governance arrangements.

2. To keep the existing capacities as autonomous as possible by interconnecting and mediating between standard-based and non-standard-based capacities.
3. To be flexible enough so as to accommodate the existing systems as well as future ones.
4. To build in an incremental fashion upon the existing infrastructures (information systems) and incorporate heterogeneous resources by introducing distribution and mediation functionalities.
5. To specify interoperability arrangements focusing on the modularity of interdisciplinary concepts rather than just on the technical interoperability of systems.
6. To avoid to ask both the resource providers and the end users to implement new specifications and/or standards.
7. To avoid to ask both the resource providers and the end users to install specific components.

In a brokering approach, brokering components are supposed to provide the necessary advanced functionalities (i.e. mediation, adaptation, distribution, semantic mapping, quality checks, etc.) desired to address the complexity inherent in these principles. Thus, a new architectural style evolves for implementing the Brokering approach, which we call the "brokering style".

## 3.1 The Brokering-Oriented Architecture

In a layered architecture, each layer provides services to the layer above it and uses services of the layer below it such as seen in Figure 6. Two adjacent layers interact by implementing the same interface(s) -which constitute the service bus for C-S architectures. The brokering-oriented architecture constraints this architectural style by requiring the use of a specific intermediary layer: the brokering layer. This layer de-couples the server and the client layers advancing the C-S separation that is crucial for multi-disciplinary environments. In fact that moves the implementation of service buses mediation to the brokering layer, as showed by Figure 6.

This separation of functionalities simplifies the interoperability requirements for both the server and client components improving scalability and flexibility (see Figures 5 and 6). Client and server components can evolve independently, regardless their interface changes. The brokering layer, by adding broker services to the C-S style, facilitates greater architecture evolution and reuse.

A Broker component must implement mediating and distribution tasks while appearing as a well-characterized component for both clients and servers. Broker

components can also be asked to implement systemic features like quality control and ranking services to the system. Brokers must encapsulate legacy interoperability arrangements and service buses; on the other hand, they must shield new buses and interoperability arrangements from legacy components. This lowers entry barriers for client and servers. The Brokering-oriented architecture is a solution to implement interoperability for large scale and complex systems, where complete knowledge of all sub-systems and their service buses would be prohibitively expensive.

Figure 6 - Disciplinary Layered C-S architectures and the Brokering oriented architecture for multi-disciplinary environments.
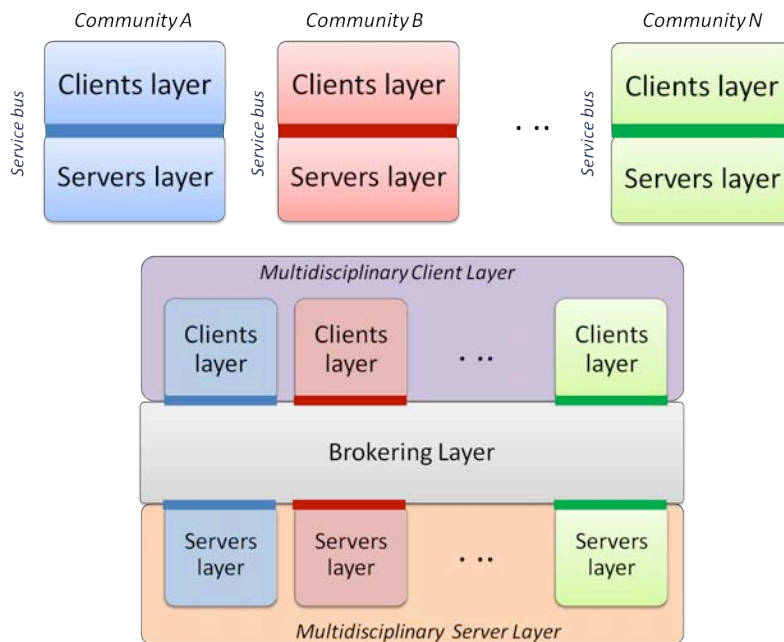
The primary disadvantage of layered systems is that they add overhead and latency to data processing, reducing user-perceived performance (Fielding, 2000). To address that, the Brokering style also applies shared caching at the brokering layers to replicate the result of an individual request such that it may be reused by later requests. Figure 7 shows the architectural styles applied by the Brokering-oriented architecture style.

Figure 7 - Brokering-oriented architecture style
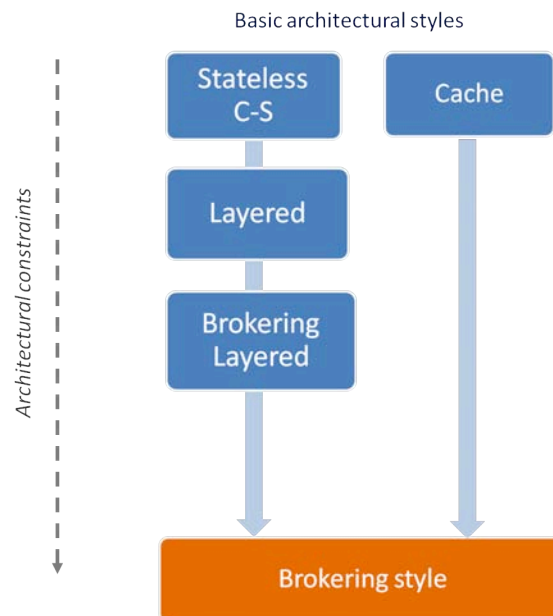
3.2 Broker Functionalities

A Broker is a middleware component which implements brokering services (i.e. mediation, distribution, added value, etc.) to facilitate the interconnection between client and server components in a Client-Server architecture.

Brokers are implemented by network-based systems which apply a Brokering architecture style. Therefore, a Broker implements the following functional modules:

- **Client request distribution** in an asynchronous way, providing consistent feedbacks to clients. This includes proxy functionalities. For example, a client can ask a broker to distribute requests for searching resources (data, services, documents, vocabularies, etc.) across one or more cyber(e)-infrastructures.
- **C-S interface protocols adaptation**. This includes gateway functionalities to match the many server protocols and the client ones, and vice-versa.
- **C-S interface models mediation**. This includes translation from the many server and client interface models to the broker internal model, and vice-versa.

10

- **Added-value functionalities at the C-S interconnection level**. This optional functionalities include added-value services which are specific to the broker type. As discussed, the introduction of the Brokering layer shifts the C-S architecture from a two-tier (Client-Server) to a three-tier (Client-Broker-Server) configuration. Hence, the broker can act as an "application server" and do more than just facilitate client and server interconnection: it may implement new capabilities, such as: advanced/semantic discovery, resource tagging and results clustering, resources ranking according to unified metrics, quality evaluation, load balancing, security checking, data pre-processing and transformation (to facilitate the download), work flow management, etc. These functionalities may complement analogous functionalities already implemented by the interconnected components.
- **Configurability at the C-S component level**. This is to allow the customization of the brokering services for each brokered client or service component. Generally, this is achieved by recognizing a set of brokering parameters to be configured on a case by case basis.

We recognize that with the brokering approach there is a single point of failure in the information system: the Broker. In general terms, this risk is present in any layered C-S architecture, where an "application server" layer is used. In fact, this risk is present in most of the "ultimate" and successful platforms, including: GoogleMap/Earth; Eye on Earth, Cloud Computing platforms, etc. They apply many of the principles recognized by the Brokering approach, such as to shift complexity from Users and Resource Providers to the infrastructure/platform. Their experiences show that this risk is addressed by developing better and more reliable, distributed infrastructures (e.g. adopting caching technologies, computational scaling solutions, Web optimizations, declarative approaches, modular software components, etc.).

4.   THE EUROGEOSS BROKERING FRAMEWORK

The EuroGEOSS project4 implemented the Brokering-oriented architecture to develop a multi-disciplinary cyber(e)-infrastructure and interconnect different disciplinary capacities, including: Forestry, Biodiversity, Drought, and Climatology. The multi-disciplinary cyber(e)-infrastructure developed is operational and allows users to discover, evaluate, access, and use resources provided by heterogeneous disciplinary systems.

The EuroGEOSS brokering architecture provides three brokers: the discovery, access, and semantic brokers: they are available at: www.eurogeoss-broker.eu. These brokers provide the following functionalities to facilitate client discovery, evaluation, access, and use of server resources:

---

4 www.eurogeoss.eu

Discovery:
- Discovery brokerage;
- Semantic augmentation of discovery request;
- Discovery of Web 2.0 resources;
- Results ranking according to a given "significance" metrics.

Evaluation
- Resource common description (using ISO 19115-part 1 and 2);
- Descriptions common encoding (using ISO 19139);
- Support of extensions for disciplinary descriptions (e.g. ebRIM, O&M, Darwin-core, CF, etc.).

Access
- Data access brokerage (for both Feature and Coverage based resources as well as pictorial Maps).
- User definition of a Common Environment for data access (i.e. access data according to a common: CRS, spatial & temporal resolution, sub-setting, data format).

Use
- Use of a Common Environment for data overlay;
- Support of multiple clients/tools for data discovery, access, processing, and visualization.

4.1   EuroGEOSS and the GEOSS Common Infrastructure (GCI)

Bridging between different communities makes it possible to meet the multidisciplinary needs of scientific research without assuming that everyone will converge on one selected standard. The advantage of adopting this more inclusive and scalable philosophy has been demonstrated in 2011, with the inclusion of the EuroGEOSS brokers in the GEOSS Common Infrastructure (GCI). The traditional C-S architecture adopted by the GCI in 2008-11, which effectively allowed only services registered and adopting OGC/ISO specifications to be searchable, resulted in only thousands datasets and services to be available in the infrastructure by the beginning of 2011. With the introduction of the EuroGEOSS broker it has been possible in just a matter of months to reach over 28 million data products, as demonstrated at the GEO Ministerial Plenary in November 2011. This achievement (see Figure 8) is an illustration of the benefits of brokering as a philosophy: "Do not expect different communities to come to you and adopt your standards, but go to them and build the bridges necessary so that they minimize their efforts".

Figure 8: Growth in number of data products discoverable in the GEOSS Common Infrastructure with the introduction of the EuroGEOSS broker in April 2011 (source: GEO Plenary, Istanbul Nov 2011)
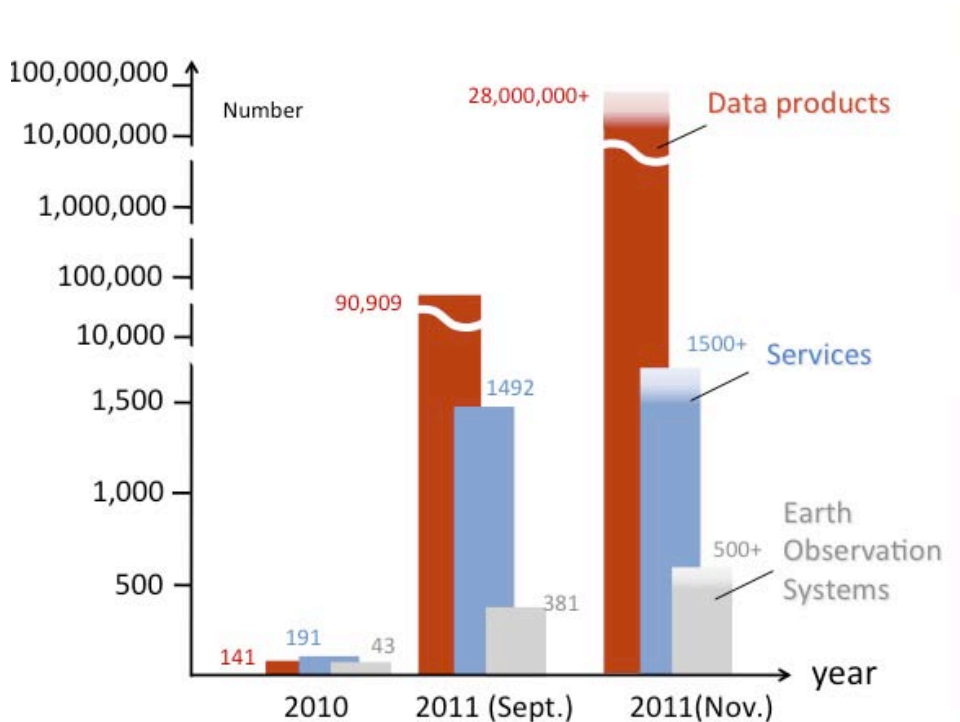
5.   CONCLUSIONS

The Brokering approach lowers User and Data Producers entry barriers considerably. This approach can offer the opportunity to support real bottoms-up infrastructure, building on existing capacities (supplementing and not supplanting them). Differently from other conceptual approaches, like federated systems, heterogeneity is addressed by focusing on mediation rather than standardization. This proved to be the best way to proceed in building complex cyber(e)-infrastructure where complexity stems from a multi-organization and multi-disciplinary nature of the infrastructure.

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. Finally, this facilitates cyber(e)-infrastructure sustainability.

The Brokering approach introduced a new architectural style: the Brokering-oriented architectural style; a brokering middleware layer is introduced to separate the client and server layers which implements brokering services (i.e. mediation, distribution, added value, etc.) to facilitate the interconnection between client and server components. The brokering-oriented style was successfully implemented and demonstrated by the EuroGEOSS project. This experience has shown that the Brokering approach is now mature: discovery Broker solutions are used by several communities (e.g. Meteo-Ocean, Hydrology, Climate, etc.) to discover, access, and use community components and services (like THREDDS[5] and GBIF[6] resources) through standard interfaces (like ISO/OGC compliant services).

Recently, the GEO Architecture and Data Committee decided to include the EuroGEOSS Broker to augment the capabilities of the GEOSS Common Infrastructure. This selection was because a brokering solution doesn't impose any common/federal model and thus facilitates broadening access to global data.

As any layered architecture, the brokering architecture needs to address broker latency and failure challenges. However, as illustrated in this paper, the advantages of the brokering approach far outweigh the potential risks as brokering has demonstrated to be able to open the door to truly integrative multidisciplinary research, a key to addressing the great research challenges of our time.

REFERENCES

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

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.

European Commission 2007. *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. Available at: http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML [accessed 10/3/2010].

Fielding, R. 2000 *Architectural styles and the design of network-based software architectures,* PhD thesis, University of California, Irvine.

---

[5] http://www.unidata.ucar.edu/projects/THREDDS/
[6] http://www.gbif.org/

International Council for Science (ICSU). 2011,. "Grand Challenges in Global Sustainability Research: A Systems Approach to Research Priorities for the Decade", ICSU: Paris.

INSPIRE Network Services Drafting Team. 2008 "Network Services Architecture", INSPIRE document D3_5_INSPIRE_NS_Architecture_v3-0.pdf, July 2008.

Manyika, J., Chui M., Bughin J., et al. 2011. *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

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

Nativi S., L. Bigagli, 2009. Discovery, Mediation, and Access Services for Earth Observation Data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(4):233 – 240.

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. *Proceedings of GEOWS '09. International Conference* DOI: 10.1109/GEOWS.2009.34.L.

Nativi, S. and Fox P. 2010. "Advocating for the use of informatics in the earth and Space Sciences", *EOS, Transactions, American Geophysical Union*. 91(8): 75-76.

Nativi S., S. J. Khalsa, B. Domenico, M. Craglia, J. Pearlman, P. Mazzetti, R. Rew, 2011, *The Brokering Approach for Earth Science Cyberinfrastructure*, NSF EarthCube White Paper.

Ortiz, S. 2007. Getting on Board the Enterprise Service Bus, *IEEE Computer*, 40(4):15-17. DOI: 10.1109/MC.2007.127.

Schmidt, M.-T.; Hutchison, B.; Lambros, P.; Phippen, R. 2005. The Enterprise Service Bus: Making service-oriented architecture real, *IBM Systems Journal*: 44(4) 781 - 797. DOI: 10.1147/sj.444.0781.

Vaccari L., Craglia M., Fugazza C. Nativi S. and Santoro M. 2012. Integrative Research: The EuroGEOSS Experience. *IEEE JSTARS*, forthcoming.

W3C, 2004. Web Services Architecture, W3C Working Group Note 11 February 2004, available at http://www.w3.org/TR/ws-arch/