How GEOSS Facilitates Interoperability through Identifying and Documenting Standards and Promoting their Adoption within User Communities

S.F. Browdy [a], D. Arctur [b], S.J.S. Khalsa [c, *]

[a] OMS Tech, Inc., 13506 Summerport Village Parkway, Suite 345, Windermere, FL 34786 – steveb@omstech.com
[b] Open Geospatial Consortium, 329 South Commons Ford Road, Austin, TX 78733-4006 – darctur@opengeospatial.org
[c] National Snow and Ice Data Center, University of Colorado, Boulder, CO 80309-0449 – sjsk@nsidc.org

Abstract − A principal element in the strategy for achieving the interoperability goals of GEOSS is the identification of, and convergence around, standards for all aspects of data collection and dissemination systems that are part of GEOSS. However, developments in GEOSS are increasingly concentrated in certain communities of practice (CoP), which are driving the implementation of data and services. These implementations can be largely detached from the GEOSS Common Infrastructure (GCI), which is the intended mechanism for achieving interoperability. The Standards and Interoperability Forum (SIF), which is responsible for the development and population of the Standards and Interoperability Registry (SIR) component of the GCI, has undertaken an assessment of progress towards achieving standards convergence and the interoperability of system components and services. We describe the work of the SIF, discuss initial findings of its interoperability assessment, and provide implications for the evolution of the GCI.

Keywords: GEOSS, interoperability, standards, Communities of Practice, SIF

1. INTRODUCTION

The Global Earth Observation System of Systems (GEOSS) exists to allow Earth observations, from various and disparate communities, to be shared as seamlessly as possible for societal benefit (GEO 2005). This goal is supported by the GEOSS Common Infrastructure (GCI), which is a set of functional software components that operate using Earth observation standards. In most cases, these standards have been developed and promulgated by the Earth observation and geospatial communities.

The European GEOSS INSPIRE and GMES Action in Support (GIGAS) project engaged in an interoperability assessment comparing INSPIRE, GMES, and GEOSS (GIGAS 2009) . This study looked at many areas, such as catalog interoperability, architectural support, application schemas, use of ontologies, data quality, etc. This study is being used by the SIF as it moves forward with its interoperability assessment.

This paper will discuss how GEOSS and the GCI, through support from the GEOSS Standards and Interoperability Forum (SIF), bring the standards to be used into the GEOSS architecture, and how the user communities are educated and assisted in adopting and using these standards. Additionally, the interoperability assessment of GEOSS, which the SIF is engaged in, will be discussed and preliminary findings detailed.

2. GEOSS ARCHITECTURE

2.1 The GCI

The GCI is composed of six core components that support and manage the interoperability of the GEOSS: GEO Web Portal (GWP), Clearinghouse (CL), Components and Services Registry (CSR), Standards and Interoperability Registry (SIR), Best Practices Wiki (BPW), User Requirements Registry (URR). These can all be accessed from http://www.earthobservations.org.

2.1.1 The Standards and Interoperability Registry

Of the core components, this paper is primarily concerned with the Components and Services Registry (CSR) and the Standards and Interoperability Registry (SIR). The SIR is overseen by the SIF, and is the registry that houses the standards that can be used by GEOSS providers and users to access and process Earth observation data. Where recognized standards are not used, many providers and communities can nominate special arrangements for inclusion in the SIR. Nominations to the SIR are provided through a web-based submission form.

The SIR entries can be used as standalone information for GEOSS users to discover what is expected in accessing data from GEOSS providers. The fewer standards that users need to implement, the greater the numbers of systems and datasets GEOSS users can interoperate with. This convergence of standards to a smaller core set is a goal of the SIF.

2.1.2 The Components and Services Registry

The CSR is the registry that houses the resources that are used to interoperate throughout the GEOSS. These resources are provided by organizations wishing to share Earth observation data, and are provided in two forms: components and services. Components are the higher level resources, such as websites, portals, catalogs, etc. that people can use to discover more details about the offered resource. Services are the interface points, in a Service-Oriented Architecture (SOA), where client applications can bind to request and receive data. In the CSR, services are contained by components, and can only be registered once an organization has registered a component.

Nominations to the CSR are provided through a web-based submission form. This form captures the metadata required for the components and services being submitted. During the process of registering a service, there is an opportunity to state what the associated standards are that the service uses. This is important for interoperability. Without being able to discover the standards being used by a service, the GEOSS user or client application cannot bind to the service easily.

Identifying the standards that can be associated with a service as it is being registered at the CSR is accomplished through the presentation of a pick list of existing SIR entries, expressed using the standards taxonomy categories implemented in the SIR. If a standard to be associated with a service being registered cannot be found, the submitter can choose to register that standard in the SIR in real-time as part of the CSR registration process.

2.2 The Broader Architecture

The broader GEOSS architecture is captured in the system of systems concept. GEOSS is based on a web-based SOA approach. The primary use case for GEOSS is publish-find-bind. The "publishing" aspect was discussed above with description of the SIR and CSR. The "finding" aspect involves the search and discovery process that takes place from the GWP, using the GEOSS Clearinghouse, once a GEOSS user initiates that functionality. Once the GEOSS user finds the metadata for the data desired, the "binding" process can proceed so that the user or client application can access the data.

Outside of the GCI, the GEOSS architecture is composed of all the systems, components, and services that take part in the sharing of Earth observation data. This can include community portals, other clearinghouses and catalogs, individual data sites, government agency sites, etc. An illustration depicting the broader GEOSS architecture is shown in Figure 1.

3. GEOSS INTEROPERABILITY ASSESSMENT

3.1 Purpose and Methodology of the Assessment

A primary goal of the GEOSS is to improve the interoperability between the observational, modeling, data assimilation, and prediction systems contributed by member countries and organizations participating in GEOSS. The GCI comprises the elements designed to enable discovery and access to these diverse data and information sources. The purpose of the interoperability assessment being undertaken by the SIF is to assess the progress towards achieving the goal of interoperability as envisioned by the GEOSS, to make recommendations regarding the evolution of the GEOSS architecture and overall data strategy to ensure fulfillment of the GEOSS vision, and to suggest metrics that can be reused to monitor interoperability going forward.

In performing this assessment, the SIF was looking for answers to the following main questions:
- To what degree can the mechanisms for accessing data, and the data themselves, be considered interoperable?
- Will the separate efforts by various Communities of Practice (CoPs 2011) within GEO, to build their own portals, lead to fragmentation or synergy?
- What communication, leadership, and outreach is needed with these and other communities to improve interoperability both within them and between them?

To proceed with the assessment, the SIF concentrated on GCI analysis, metrics, and community outreach. The GCI analysis looked at problems with, and recommended resolutions for, the ways in which the GCI functions. The work encompassing metrics sought consistent and meaningful ways to monitor and measure ongoing interoperability activity. Community outreach concentrated on gathering information from Earth observation communities regarding their own evolution with respect to

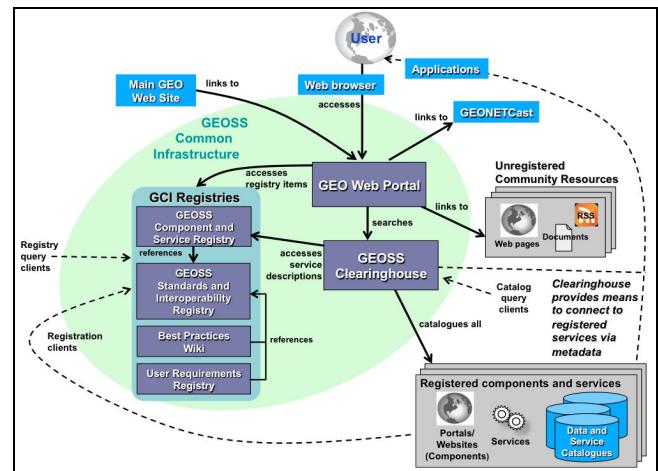interoperability, and how they were working, or intended to work, with the GEOSS.

Figure 1. GEOSS Architecture Overview

3.2 GCI Analysis

The analysis of the GCI was initially focused on collecting statistics from the SIR and CSR to look for any issues that could be identified as obstacles to interoperability.

Since interoperability within the GEOSS is based on the publish-find-bind use case, it is necessary for all registered components in the CSR to have at least one associated registered service in the CSR so that service binding can take place to access data. Figure 2 indicates how many services are associated with a component. The red bar in the figure shows that 185 registered components in the CSR have no associated registered services. Many of these are catalogs that contain large amounts of metadata records for available community data. This seems very high, but could also indicate that much more focus should be on communities rather than individual data providers.

In the process of binding with a service to access data, it is vitally important to discover what standards are being used to access the data and process it. If these associations cannot be found, then interoperability severely suffers. Figure 3 tallies the number of standards associated with registered services. Multiple associations can happen by associating a single primary standard with the service, and multiple secondary standards with the service. As can be seen, most services are only associated with a single standard. Although there is nothing inherently wrong with this, the limit of one associated primary standard prevents references to multiple versions and profiles of that standard being supported by the service, and, therefore, impedes interoperability for the data users.

Convergence of standards is a goal of the SIF, since it can improve interoperability. Convergence of standards makes it less complicated for data users to access and use data from multiple data providers. The more standards that are available for data providers to associate their services with, the more effort that can be imposed on data users to support those standards in order to achieve interoperability. Figure 4 shows how many standards are

registered in the SIR for the current standards taxonomy categories, and, as shown in red, how many standards in those categories are being referenced by services registered in the CSR.
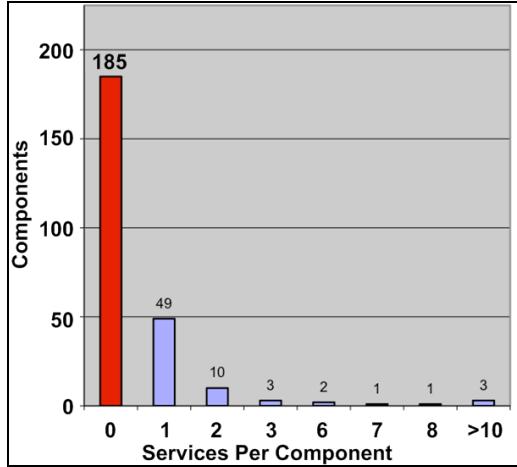

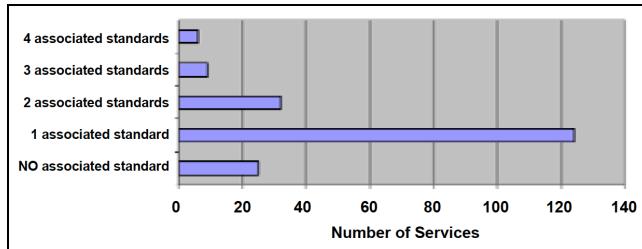Figure 2. Number of components with associated services

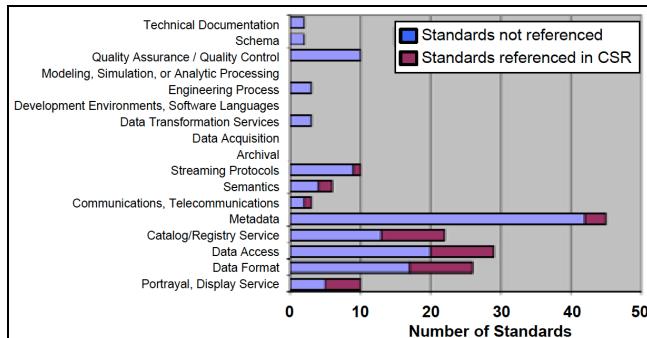Figure 3. Number of standards associated with each service

Figure 4. Standards registered in SIR, and referenced from CSR

Figure 5 shows how many references are being made to certain SIR entries by registered services in the CSR. Both Figures 4 and 5 show possible reasons why convergence of standards is desirable. Convergence will not result in the deletion of registered standards in the SIR, but will highlight the standards most used and most useful to the data sharing community at large.

3.3 Metrics
One direct source of metrics that serves the ongoing monitoring and measurement of GEOSS interoperability is the statistics gathered from the GCI analysis. The data that resides within the CSR and SIR can periodically be examined for completeness and correctness, and can be mined for patterns of associations and usage.
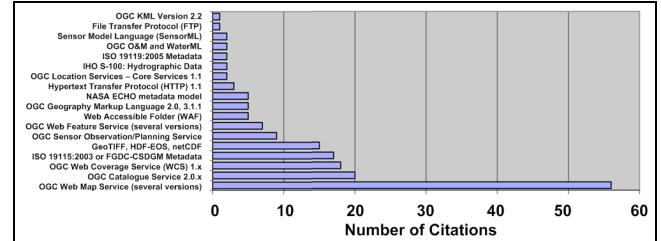

Figure 5. Specific standards cited in the CSR

Currently, the GCI registry statistics referred to in this paper only deal with syntactic interoperability. The SIF, however, has acknowledged that semantic interoperability and legal interoperability are also areas of future concern that need to be dealt with. Semantic interoperability can be measured, in some sense, by the number and types of taxonomies and ontologies being used. This translates into registered taxonomies and ontologies associated with services. The number of registered services engaged in semantic mediation and inference is also a measure of the state of evolution of GEOSS towards semantic interoperability.

Legal interoperability has to do with the manner in which data can be used in an aggregate way in keeping with the data access and use restrictions attached to the data. One such definition of legal interoperability, provided by Harlan Onsrud (Onsrud 2010), is:

"*A functional environment in which: a) differing use conditions imposed on datasets drawn from multiple disparate sources are readily determinable, typically through automated means, with confidence; b) use conditions imposed on datasets do not disallow creation of derivative products that incorporate data carrying different use conditions; c) users may legally access and use the data of others without seeking permission on a case-by-case basis.*"

Legal interoperability ensures that the use and access restrictions imposed on data can be determined unambiguously This suggests the use of licensing frameworks that are agreed to broadly, an example being Creative Commons. Currently, this cannot be assessed from entries in the CSR but may be possible in the future as the GCI evolves to handle licensing frameworks.

3.4 Community Outreach
The SIF has an ongoing mission of education and outreach. Until recently, this has been primarily focused on data providers contributing individual services to the GEOSS. Assisting data providers and users with standards and interoperability has been accomplished via e-mail and the Architecture Implementation Pilot (AIP) program. More currently, the SIF has come to understand that a focus on communities is essential going forward, and decided to include an outreach activity in the interoperability assessment.

The SIF performed community outreach for this interoperability assessment to gauge the level of interoperability that communities have collectively achieved, to identify the interoperability impediments that communities face, and to determine what the relationships are between the communities and the GEOSS. The SIF surveyed select communities to accomplish this outreach, beginning with a pilot survey involving just three communities. This pilot survey is being used to assist in formulating the final questions for the general survey (OPP 1992).

Once the pilot phase is completed, the survey will be disseminated more broadly. The strategy is to have leaders in the communities hand pick the respondents to ensure that completed surveys reflect the knowledge of those who have expertise in the various categories being surveyed. Legal interoperability was not surveyed, since very few examples of this currently exist.

4. CONCLUSIONS

As of the writing of this paper, the SIF is still actively engaged in the interoperability assessment, but there are certain conclusions that can already be drawn from the work. With regards to metrics, the GCI analysis has exposed very specific items to monitor. Components registered in the CSR must be able to be associated with standards. Otherwise, data users cannot learn which standards the data provider is using for data access and data use without interacting with the data provider directly. Currently, there is no way to associate standards with components, so a recommendation can be made to allow this to happen for improved interoperability, along with SIF monitoring to ensure that these associations are being made properly.

Components registered in the CSR often have no associated services registered. The number of services associated with a component is a metric that should be monitored carefully. Although, it is possible for a component to have no associated services, the registration process makes it possible for this to be overlooked, leaving a situation that impedes interoperability.

A metric exists related to registered services in the CSR. Every service must be associated with at least one standard. The SIF can monitor this metric and ensure that standards are associated properly. A related recommendation is to allow multiple standards from a primary standards taxonomy category to be associated with services. This will improve interoperability.

Standards convergence can be assisted by the metrics that track the number and types of standards cited by components and services registered in the CSR. In order to use these metrics, SIF monitoring must be in effect to ensure properly reported associations between components, services, and standards.

The Best Practices Wiki, currently, supports the publication of existing and evolving best practices as a means of conveying information to enhance interoperability. However, there is no way to integrate these practices with the other GEOSS registries. A recommendation from the SIF is to deploy a Best Practices Registry, which will be associated with the existing BPW, and can interoperate with the other registries in the GCI, in particular the SIR and CSR.

A recommendation can also be made to refocus efforts on community outreach. Informal results of the community survey, to this point, have shown that Earth observation communities are evolving their technology and interoperability according to their own needs and goals. Sometimes, this aligns with the interoperability goals of GEOSS, but not always. It is important to understand that there are different stages of evolution for communities, and, therefore, different mechanisms of interoperability are needed to relate to and work with these communities. A range of architectural nuances is necessary to engage, and maintain, involvement of communities with GEOSS. An example of an advanced architecture with communities is shown in Figure 6.
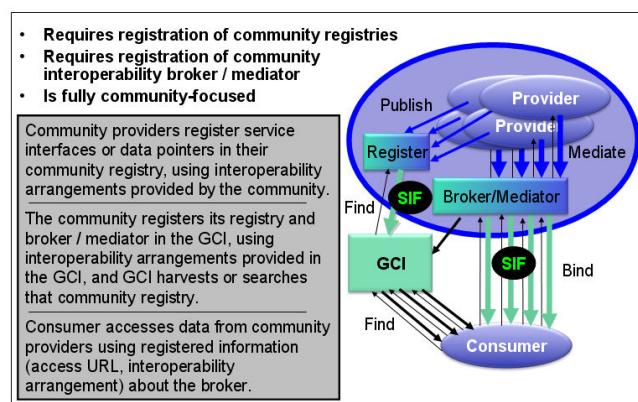

Figure 6. Proposed architecture, and roles for GCI and SIF, for advanced communities.

5. REFERENCES

EVAL 2010. "Mid-term Evaluation of GEOSS Implementation", http://www.earthobservations.org/docs_od_ple.shtml

OPP 1992. A. N. Oppenheim, "Questionnaire Design, Interviewing, and Attitude Measurement", ISBN 1855670437 (HBK) 0826451764, 1992

Onsrud 2010. Harlan J. Onsrud, "Legal Interoperability in Support of Spatially Enabling Society", http://www.gsdi.org/gsdiconf/gsdi12/papers/907.pdf

GIGAS 2010. A. Biancalana, S. Nativi, et al. "GIGAS Comparative Analysis Technical Note", GIGAS project deliverable D2.3b, Dec 2009

CoPs 2011. "GEO Communities of Practice" Website accessed 3 February 2011. http://www.earthobservations.org/cop.shtml

GEO 2005 "Global Earth Observation Systems of Systems – GEOSS – 10-Year Implementation Plan", GEO 1000 / ESA BR-240, February 2005, Published by ESA, The Netherlands