J2.5

STRATEGIES TO ENSURE THAT SATELLITE OBSERVATIONS CAN BE FULLY EXPLOITED WITHIN EMERGING ENVIRONMENTAL SYSTEMS OF SYSTEMS

Carroll A. Hood*
Raytheon Company

1. Introduction

The value proposition for emerging environmental systems of systems (SoS) such as the Global Earth Observation System of Systems (GEOSS) is predicated on enabling data, information, products and services to be easily discovered, accessed, and exploited in ways that extend beyond their original intent. Satellite derived products are voluminous, becoming more ubiquitous and have the potential to be applied to a wide range of application domains. Thus, in order for the value of these products to be fully realized by a diverse user community (some of whom may be distant from the traditional set of primary and secondary remote sensing consumers), new and innovative strategies for discovery, access, and integration will need to be created and socialized. A service-oriented approach to overcoming syntactic and semantic barriers to interoperability will be discussed.

2. The GEOSS Value Proposition

The primary value proposition for environmental systems of systems such as GEOSS lies in its capability to enable both *new* and *existing* environmental observations, data, products and services to be used for applications and activities beyond their original intent. This is best described by recognizing that there are two distinct yet interconnected Value Streams within this enterprise (Figure 1).

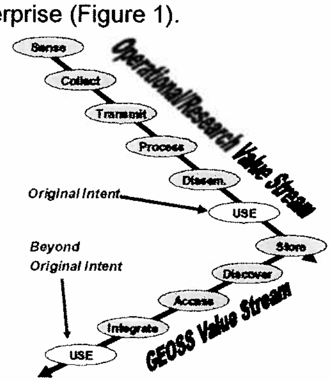

Figure 1 GEOSS Value Streams

The first Value Stream represents the activities required to collect, process and deliver environmental observations, products or services for some intended purpose (e.g., operational weather forecasting, research objective, etc.) The second Value Steam represents the discovery, access, and transport activities that are required to apply these observations, products or services towards some other application or end-use. In this simple view, the two Value Steams intersect at a data store; in reality, the connection may occur further up the original value stream.

Environmental satellites present a significant challenge because the secondary uses of the resulting data, information, products and services are almost unbounded. This situation will be exacerbated by the impending new generation of remote sensing satellites (e.g., NPOESS, GOES-R) that will improve the spatial, temporal and spectral resolution of the "raw material" that can be exploited. For example, the NPOESS C1 will generate approximately 30 Environmental Data Records (EDRs) spanning atmospheric, climate, oceanic, land, and space environmental domains. (NPOESS, 2008) Because satellite systems tend to be relatively expensive to develop, build and operate, the need for secondary uses to extend the value proportion is even more pronounced,

Once the notion of "beyond original intent" is understood and appreciated, the next step is to decompose that idea into specific activities that can actually generate quantifiable social or economic value. For environmental SoS, value-generation activities include:

- Enabling environmental and/or geospatial perspective to be incorporated into decision making for selected social and economic issues whenever and wherever it makes sense. The implication here is the inclusion of an environmental perspective will enable *smarter, better-informed decisions providing tangible value to both society and our economy.* This is the most commonly documented business rationale for environmental SoS.
- Encouraging the development of a value-added market for Earth observations data and information products and services. The Internet and GPS are examples of how systems built for a specific original intent have evolved to become economic engines within domains far removed from their original audience. In other words, it is likely the bulk of environmental SoS-enabled products and services available 10 years

from now *have not even been imagined today*

- Supporting capacity building in developing countries. This is a critical aspect that cannot be overlooked. Many environmental issues have global implications, and oftentimes, even local solutions require a global perspective. This approach not only makes us wise stewards and savvy entrepreneurs, but also altruistic benefactors for the emerging global economy.

From an architectural perspective, the tenets of this value proposition need to be included within the environmental SoS design. A common method of capturing these design drivers is to articulate them in terms of quality attributes (Table 1). We will use these attributes to evaluate potential environmental SoS designs in later sections.

Value Component	Related Quality Attributes
Enable environmental and/or geospatial perspective to be incorporated into decision making	• ease of discovery, transport, access • tailorable interface • ease in integration and analysis
Encourage the development of a value-added market for Earth observations data and information products and services	• extensible to new data types and applications • ability to support diverse business models
Support capacity building in developing countries	• low barrier to entry, participation • maintainability

Table 1. Environmental SoS Quality Attributes

3. GEOSS Compliance

Another consideration for the development of a design for an environmental SoS is that it should be compliant with the emerging architecture for GEOSS. Fortunately, GEOSS compliance is based on a few simple constructs:

- Improved interoperability through adoption and application of relevant standards that apply primarily to the interfaces of services and components.
- A searchable registry of services, components, and approved standards.
- A portal that connects the GEOSS registries and extent catalogs/registries through a clearinghouse (Figure 2) (ICEO, 2008)

So, the good news is, developing an environmental SoS design that is GEOSS compliant will be easy. The bad news is the way in which the design deals with interoperability has significant implications on how well the design addresses the key quality attributes defined above.

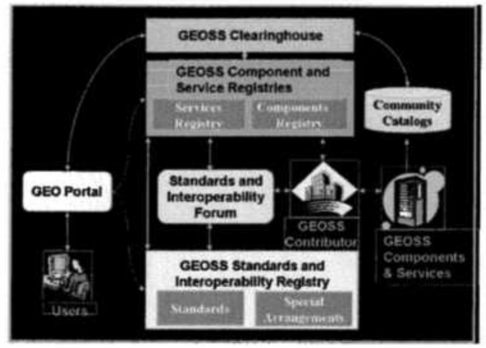

Figure 2 GEOSS Architecture

4. Interoperability and System Design

The cornerstone for GEOSS is interoperability; as a result, much of the effort in environmental SoS design is focused on overcoming the barriers to interoperability. In general, these barriers can be subdivided into two categories

- Those that are intrinsic to the data or information themselves [e.g, the format/structure of the data (syntax), and the meaning of the data (semantics)]
- Those that are intrinsic to the environment or infrastructure in which the data reside or operate (e.g., related to transport/access mechanisms, networks, etc.)

The strategy for overcoming these barriers is typically tied to either the adoption of a relevant standard (e.g., netCDF, CF conventions, etc) or normalization of the barriers into a common baseline (which may or may not be tied to some underlying standard.)

The ability to leverage effectively a barrier mitigation technique across an enterprise is related to the extent and effectiveness of governance that exists within or across that enterprise. This point is particularly relevant to environmental systems of systems because they maximize their value potential when data, products and services are being used across any number of disparate domains. Establishing effective governance within a single domain can be a challenge; doing it across multiple geographically dispersed, and/or culturally distinct domains may be next to impossible.

Thus, if we focus on the data-intrinsic aspect of interoperability and the level of governance within an enterprise, we can define a two-dimensional space depicted on Figure 3. The x-axis defines the extent of centralized or federated governance within the enterprise, and the y-axis defines the degree to which the syntax and semantics are normalized as a method to improve interoperability These two axes are not really orthogonal since normalization across domains would be much easier within an enterprise that has an effective and highly centralized governance. Thus, most of the potentially viable systems designs within this space would occur within the blue oval.

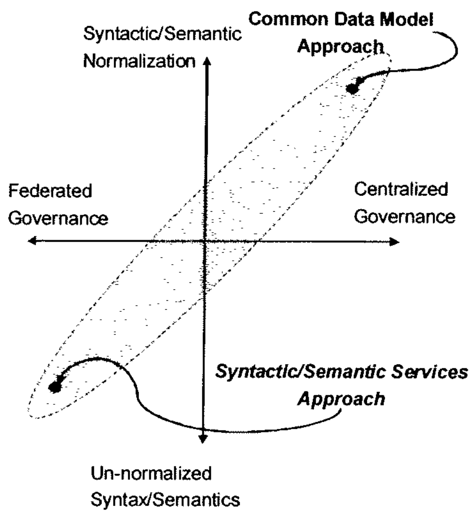

Figure 3 Viable System Design within Governance/Data Normalization Space

Within the oval, we will now choose to examine two potential environmental SoS designs:

- One design is characterized by a strong normalization of the data syntax and semantics. In this case, data from a variety of data providers are brought together, normalized and placed into a common storage architecture that is characterized by a common schema. If a user understands the syntax and semantics of the common data store, all of the data contained within is accessible and can be used. We call this method the **Common Data Model** (CDM) approach.
- In the second design, data providers expose their data and services with their native syntax and semantics preserved in accordance with their local governance. Services are used to mediate the discovery and access processes. Data are delivered along with a machine readable characterization of their native syntax and semantics. It is left up to the user to navigate these barriers, although additional services can be created to overcome them. We call this method the **Semantic/Syntactic Services** (SSS) approach.

Note that the CDM and SSS are viable system design solutions at opposite ends of the blue oval. Other viable approaches (including hybrids of the two) are possible.

5. Discussion

5.1 CDM Approach

The CDM approach is elegant, logical and forms the foundation of many environmental SoS in operation today. A couple of notable examples include

- NOAA/NWS Advanced Weather Interactive Processing Systems (AWIPS) (AWIPS, 2007)
- DoD Distributed Common Ground Systems (DCGS) Integration Backbone (DIB) (includes environmental data) (Goodman, 2007)
- Southeast Coastal Ocean Observing Regional Association (SECOORA) Data Management and Communication (DMAC) infrastructure. (SECOORA, 2008)

A nominal architecture for a CDM approach is given in Figure 4.

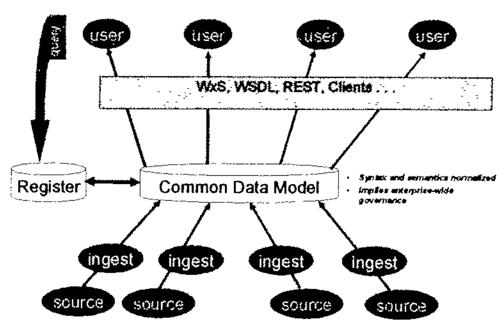

Figure 4 Nominal CDM Architecture

Each unique data source requires a distinct ingest processor that transforms the data into something that can easily be loaded into the common data store. Information about the holdings is stored in a registry. Users access the information via web services or other clients.

Some advantages of the CDM approach include:

- The heavy lifting occurs at the ingest side of the Value Stream. As a result, the front-end can be optimized for low latency, interactive, gaming-style access to the data. This type of access is especially important for applications such as weather forecasting where rapid decision making can have a large impact on life and property.

- QA/QC can be performed during the ingest process such that every item in the data store has a consistent, characterized level of quality.
- An appropriate service level agreement (SLA) for access to the Common Data Store can be implemented, although the content may be depend on individual SLAs with the data providers.
- The underlying schema may facilitate downstream integration and analysis.
- Existing implementations can provide a design roadmap for future use in other domains.

Some disadvantages of the CDM approach include:

- The inclusion of new data types usually requires the development of a new ingest adapter and may require modification to the underlying data schema. As a result, the barriers to participation may be high
- Enterprise-wide governance may be required to achieve consensus on the data schema and acceptable translation methods
- Normalization may result in the loss of some syntactic or semantic richness.

5.2 SSS Approach

The SSS is less well understood and, as of yet, has never been fully implemented within an environmental SoS (although the OOSTethys initiative ((OOSTethys, 2008) incorporates many of the concepts.)

A nominal architecture for the SSS approach is given in Figure 5.

The SSS approach begins, like the CDM approach, with a set of distributed data providers. However, in this case, each provider maintains its own local data store and registry. Each data provider is encouraged to provide both a semantic characterization (e.g., an XML schema of the native product syntax) and a semantic characterization (e.g., OWL or RDF-based ontology) of their holdings. These metadata are delivered along with the data products.

The image in Figure 5 presents four different scenarios that can be enabled via the SSS approach. These scenarios are meant to serve as exemplars only. They depict a range of discovery and access capabilities that might be provided to users, but many more scenarios could be envisioned. From left to right they include the following:

- The simplest scenario is one that does not include wire access to the data holdings. In this case, the data provider simply loads the suggested metadata (including the syntactic and semantics characterizations described above) in a local registry. Users can discover the existence of the holdings (via some type of discovery service that performs semantic mediation), but may have to use non-automated means to actually access the data.
- In the second scenario, the data are available on line, and a parsing service utilizes the syntactic characterization to parse and deliver the desired data elements directly from the data store.
- In the third scenario, selected data elements are refactored from the original data store into an on-line accessible data store. This may be to isolate secondary access from an operational system or perhaps to enable a certain type of web service access to the data store that otherwise would not be available.
- The final scenario depicts the creation of a new system with the SoS. In this case,

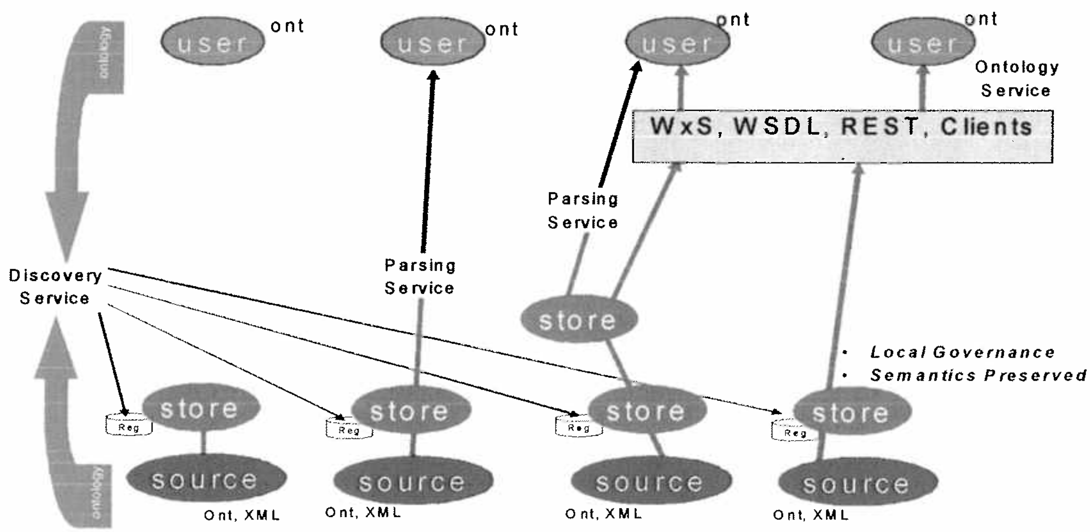

Figure 5 SSS Nominal Architecture

discovery and access are designed from the beginning to be service-oriented.

The advantages of the SSS approach include the ability:

- to integrate systems across disparate domains, each maintaining its own local governance;
- to support a variety of implementation strategies. Some of these strategies have very low barriers to participation; others support highly sophisticated interfaces
- to embrace multiple business models that could be applied at different points within the enterprise simultaneously.
- to distribute system operation and maintenance distributed (often a cost driver) fairly to all enterprise participants.

Some disadvantages of the SSS approach include:

- Effective implementation will require the maturation of some semantic and syntactic services;.
- System level attributes (e,g., system availability, end-to-end latency) may be difficult to optimize.

Since the SSS approach has never been fielded in its entirety, Figure 6 pictorially depicts a high-level concept of operations showing how this approach might be implemented.

The proposed services (represented by the ovals) currently do not exist as envisioned in Figure 6, yet elements of these services have already been created.

- The Marine Metadata Interoperability (MMI) initiative (http://marinemetadata.org/) has created a tool called VOC2RDF which creates a simple RDF-based ontology based on user input of fundamental semantic triples. This is a primary element of the Ontology Service envisioned above.
- MMI has also created a tool called VINE that enables users to establish equivalencies between ontologies. This is the first step in creating a Discovery Service as envisioned above.
- The Parsing Service as envisioned above should be able to leverage XSLT, a language for transforming XML documents into other XML documents.

It is important to point out that both the CDM and the SSS approach are highly dependent on the rigorous collection and maintenance of life cycle provenance/metadata. The procurement strategy for many environmental remote sensing systems (particularly environmental satellites) tends for focus on the initial Value Steam only. Thus, some of the metadata or provenance that would be useful for the second Value Steam is never identified as a requirement for the observing system. This is a short-sighted deficiency that needs to be corrected. It is very

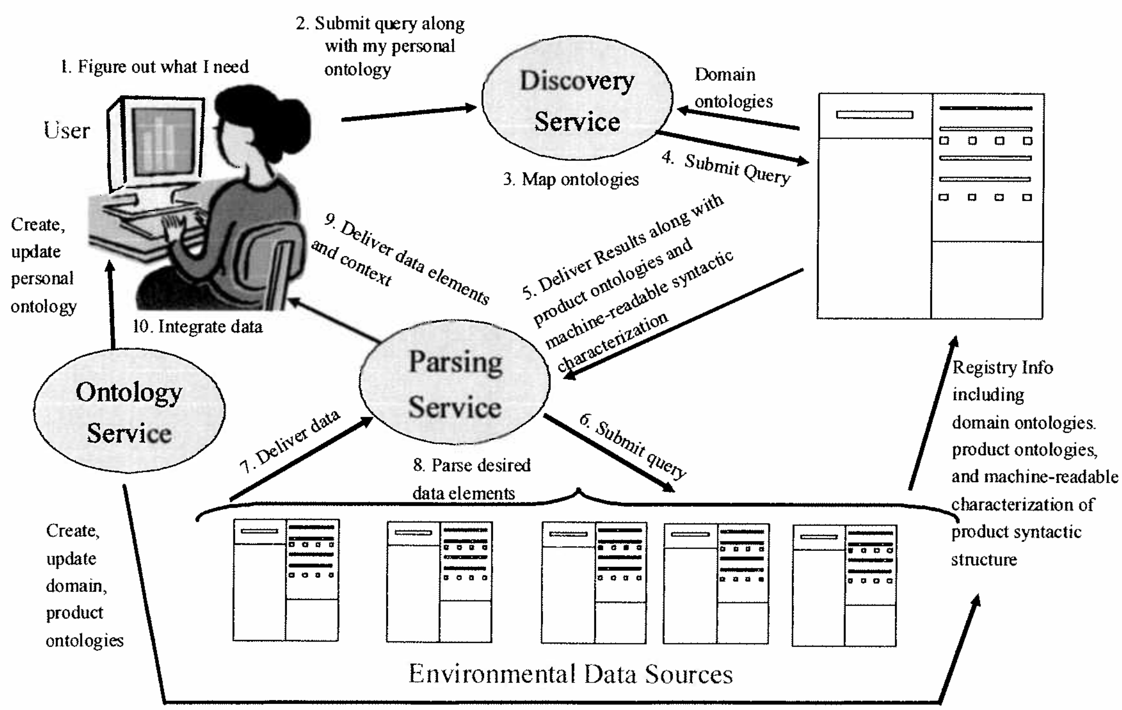

Figure 6 SSS Concept of Operations

likely that the marginal return measured in downstream value would far exceed the marginal cost of including such a requirement within the observing system specification.

6.0 Conclusion

Environmental observing systems in general and satellite-based systems in particular, can be expensive to build, operate and maintain. The return on the national investment in these systems can be greatly enhanced if they are designed and built with GEOSS-level interoperability in mind. The metadata required to enable a system that meets this design specification needs to be captured at the time of inception (It typically cannot be regenerated after the fact).

We have considered two different environ mental SoS design paradigms defined by the extent of enterprise-wide governance and the degree to which a key barrier to interoperability has been normalized. Both approaches can be GEOSS-compliant (if GEOSS-approved standards are employed); however they differ in their ability to fully support the GEOSS Value Proposition.

If we evaluate these two designs against the quality attributes defined earlier, it is clear that both approaches have merit (Table 2). The CDM approach may serve as a useful reference architecture for inter-domain environmental SoS, while the SSS approach may be more useful as a high-level cross domain environmental SoS reference architecture.

Quality Attribute	CDM	SSS
Ease of discovery, transport access	X	
Tailorable interface	X	X
Ease of integration, analysis	X	
Extensible to new data types and applications		X
Ability to support diverse business models		X
Low barrier to entry, participation		X
Maintainability		X

Table 2: SoS Design Mapping to Quality Attributes

7.0 References/Endnotes

AWIPS, 2007. A brief overview of the AWIPS architecture can be found in the AWIPS Application Integration Framework Manual at http://www.weather.gov/mdl/awips/

Goodman, GW. 2004. *Intel Internet.* ISR Journal, August 2004.

IEEE Committee on Earth Observations (ICEO). 2008. Figure 2 taken from http://www.ieee-earth.org/adc/ADCArchitecture

NPOESS, 2008. The list of NPOESS C1 EDRs can be found at http://www.ipo.noaa.gov/index.php?pg=edrs

OOSTethys, 2008. A description of the OOS Tethys initiative can be found at http://www.oostethys.org/

Southeast Coastal Ocean Observing Regional Association (SECOORA), 2008. A schematic of the "as-is" Data Management and Commutations (DMAC) architecture can be found at http://secoora.org/documents/technical/system-architecture-diagram

---

* *Corresponding author address:*
Carroll A. Hood,
Raytheon Intelligence and Information Systems,
16800 E. CentreTech Parkway,
DN/S79/MS1306
Aurora, CO 80011-9046
Email: cahood@raytheon.com