Open Geospatial Consortium Inc.

Date: 2010-02-16

Reference number of this document: **09-182r1**

Version: 0.4.0

Category: OpenGIS® Engineering Report

Editors: Josh Lieberman

End to End Discovery and Access Engineering Report GEO Architecture Implementation Pilot, Phase 2: Engineering Report

Warning

This document is not an OGC Standard. This document is an OGC Public Engineering Report created as a deliverable in an OGC Interoperability Initiative and is not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, any OGC Engineering Report should not be referenced as required or mandatory technology in procurements.

Document type:	OpenGIS® Engineering Report
Document subtype:	Approved
Document stage:	Engineering Report version 0.4.0
Document language:	English

GROUP ON EARTH OBSERVATIONS

End to End Discovery and Access Engineering Report GEO Architecture Implementation Pilot, Phase 2

Version 1.0
OGC Document # 09-182r1

Revision History

Version	Date	Editor and Content providers	Comments
0.1 Outline	2 June 2009	Joshua Lieberman	Outline
0.1 Draft	9 June 2009	Hervé Caumont	Annotated Outline
0.1 Draft	28 July 2009	Joshua Lieberman	
0.1 Draft	27 August 2009	Marten Hogeweg	Expanded on several aspects. Included some suggestions
0.5 Draft	1 September 2009	Joshua Lieberman	Completer draft
0.6 Draft	1 September 2009	Steven F. Browdy	Contributed 5.2
0.7 Draft	8 September	Joshua Lieberman	Completion and response to comments
1.0	23 September	Joshua Lieberman	Final Version
1.0	13 October	Joshua Lieberman	Revision from comments by Marten Hogeweg

Document Contact Information

If you have questions or comments regarding this document, you can contact:

Name	Organization	Contact Information
Joshua Lieberman	OGC IP Team	jlieberman@traveresetechnologies.com tel:+1 617 297 2113
Hervé Caumont	ERDAS	Herve.Caumont@erdas.com
Marten Hogeweg	ESRI	mhogeweg@esri.com tel: +1 909 793 2853 x1129
Doug Nebert	FGDC	ddnebert@usgs.gov
Ted Habermann	NOAA	Ted.Habermann@noaa.gov
George Percivall	OGC	gpercivall@opengeospatial.org

Table of Contents

Table of Figures

Table of Tables

1. Scope

This Architecture Implementation Pilot, Phase 2 Engineering Report (AIP-2 ER) describes the practice of deploying, documenting, and registering contributed resources from the point of view of classes of GEOSS users who rely on GEOSS to support discovery and access to those resources. It emphasizes two paradigms for the GEOSS Common Infrastructure: 1) Service-oriented infrastructure for development of service-based community applications by technically advanced users; and 2) Content-oriented search facility and Web-based access mechanisms for end-users with a range of technical skills and domain knowledge. "End-to-end" here refers to the bidirectional connection between desired discovery practices and goals on the user end; and the required resource interfaces and documentation on the provider end.

2. Introduction

This ER is a key result of the second phase of AIP. AIP-2 was conducted from July 2008 to June 2009. A separate AIP-2 ER describes the overall process and results of AIP-2 and thereby provides a context for this E2EDA ER.[1] Another ER describes the general use cases which this report seeks to weave together. Yet another set of ER's covers the ways in which discovery and access have been implemented in specific SBA (Societal Benefit Area) community scenarios in the course of AIP-2.

2.1 Definition of End to End Discovery and Access

The term 'end-to-end discovery and access' (E2EDA) is coined to describe the relationships discerned during the course of AIP-2 between practices for publishing and providing earth observations, and the paradigms employed by GEOSS users for finding and getting to those observations. It comprises a variety of GEOSS practices:

- Providers choose to deploy resources (data, services, applications) based on earth observations.
- Providers then document and register those resources so they can be found and accessed.
- Consumers (users) pursue an objective and choose questions related to societal benefits and/or earth observations
- Consumers seek both relevant resources and the tools to work with them, such as client software.

These practices have to be connected if they are to be successful. Providers wish to deploy resources so that they can be put to good use. Consumers wish to find available resources to meet their needs. Certain threads dominate this end-to-end connection:

- Users can only find resources that are described by the metadata and indexed by the properties that matter to them and with which they are familiar. In the realm of earth observations, this typically includes at least the specific increments of space and time which have been observed, and the specific phenomena which have been measured in the course of the observation. Other properties and classifications may also be important
- Earth observation information can only be utilized effectively when it is offered through services for which clients are readily available for access, comparative evaluation, and exploitation. Typically, metadata are necessary but not sufficient to make final decisions about suitability–some direct experience of information resources is also required.

E2EDA emphasizes above all the need for clear communication of requirements across GEOSS between all stakeholders, so that each link in the discovery – access chain is both functional and essential.

[1] A listing of all AIP-2 Engineering Reports: http://www.ogcnetwork.net/AIP2ERs

2.2 Initial conditions: CFP, SOW's, Kickoff meeting, Working groups

Initial plans for discovery and access were established in the AIP-2 Call for Participation, particularly in the GEOSS Architecture (CFP Annex B), and the scope of activity was further refined based on statements of work from the participant. An AIP-2 kickoff meeting held in September 2008 established an important distinction between working groups focusing on particular knowledge communities, and 'transverse' working groups focusing on general aspects of the development and use of GEOSS.

Transverse technology working groups:

- Catalogues, Clearinghouse, Registries and Metadata (CCRM) WG: Doug Nebert, USGS; Josh Lieberman OGC/Traverse; Ted Haberman, NOAA
- Workflow and Processing WG: Greg Yetman, CIESIN; Eugene Yu, GMU; Satoshi Sekiguchi, AIST;
- Test Facility for service registration WG: Gianni Sotis, Mauro Semerano, ESA
- Portals and application clients WG: Nadine Alameh, OGC/Mobilaps; Herve' Caumont, OGC/ERDAS

Particularly in the CCRM working group, the connection between registration processing, metadata publication, and the ability of users to find and access the right resources has been developed.

2.3 E2EDA and GEOSS Common Infrastructure

GEOSS as a 'system of systems' is composed to a considerable extent of the constituent systems established or being established at both governmental and community levels. Nonetheless, there is intended to be a core system organized around the GEOSS Common Infrastructure (GCI) serving to connect the resources of the constituent systems. As a form of system 'glue', the GCI has been the primary focus for work on E2EDA

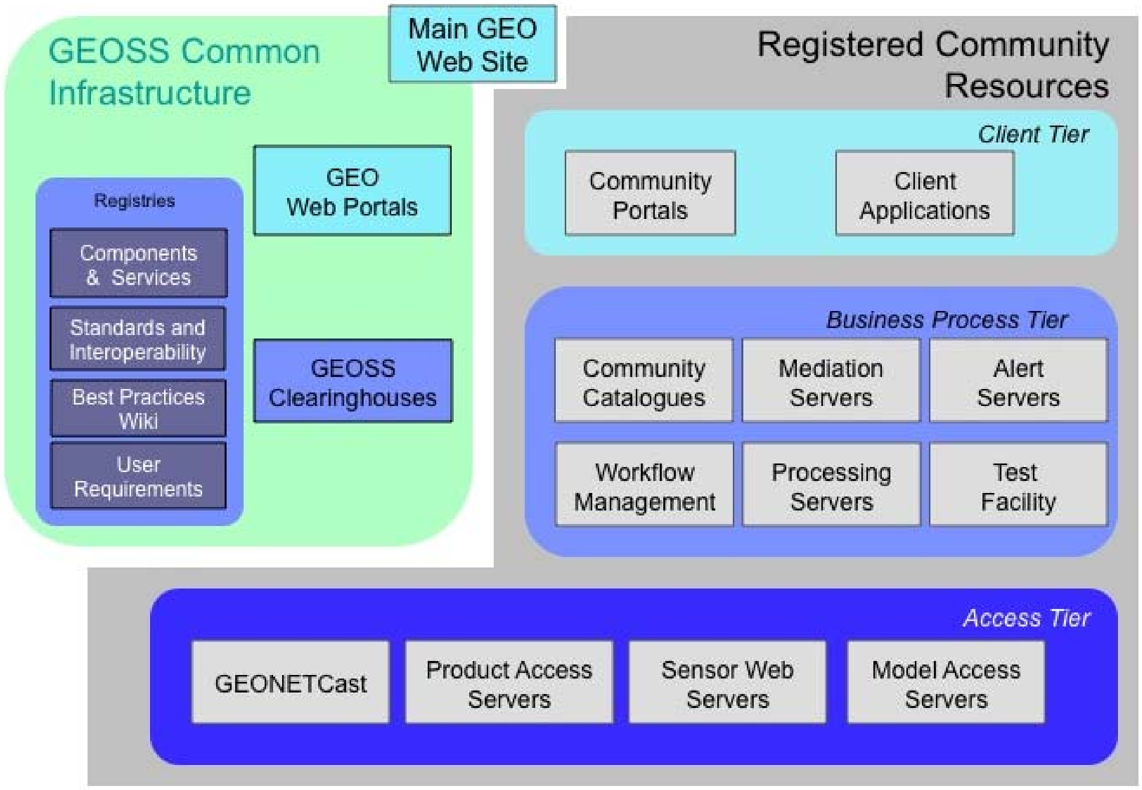

Figure 1 – AIP-2 Engineering Components Augment the GCI IOC

Figure 1 illustrates the various types of engineering viewpoint components considered to make up GEOSS as a whole. GCI is composed of a particular subset of these components, namely the GEOSS Registries, GEOSS Clearinghouses and GEO Web Portals. There has been discussion from the beginning of AIP-2 as to whether Web Portals should be considered part of the GCI, but work on user needs in AIP-2 has made clear that the availability of a consistent user interface and experience, at least for discovery and minimal access, is an essential common element of GEOSS.

While GCI may be the common ground for enabling E2EDA, it is the registered community resources that are the ultimate source of the resource metadata to be discovered and the interfaces to be accessed.

2.4 GCI Resources and Roles

The basis for the functional paradigm of the GCI consists of the distinct roles played by its three main systems:

- GEOSS Registry – GEOSS Providers register their *Components* and *Services* in the **Component and Service Registry** according to *Interoperability Arrangements* registered in the **Standards Registry**. Providers either register their resources directly or register metadata services that in turn describe their resources.
- GEOSS Clearinghouse – A **Clearinghouse** is a catalog server providing uniform query interfaces to all of the diverse resource metadata and metadata services registered directly or indirectly with the GEOSS Registry. The Clearinghouse harvests the registration records from the Registry and then either harvests the registered metadata services in turn, or relays queries from Clearinghouse Clients to those services.
- GEO Web Portal – One type of Clearinghouse Client is a **GEO Web Portal**. It provides a catalog user

interface for building Clearinghouse queries and also provides a Web Map Service client for those GEOSS resources provided through such a service. A GEO Web Portal may provide additional organization of GEOSS resources, for example into *Societal Benefit Areas*.

The types of resources that can be registered in GEOSS are discussed further along in this report.

2.5 GCI and the outside world

GEOSS exists in an environment where information sharing has become common practice. In various parts of the world, spatial data infrastructures (SDI) are being developed or are evolving (INSPIRE, US NSDI, GSDI, etc.). As a result of these developments many organizations have started to host services and components that may be of use for the GEOSS user community. In many cases the organizations participating in those efforts already participate in GEOSS.

"Here Be Dragons"

The development of these resources external to the GCI results in a wealth of existing content potentially being available to the GEOSS user community. It also means, however, that if the GCI wants to leverage these external developments it will have to deal with a heterogeneous offering of components and services and their descriptions in metadata; it will also have to deal with diverse prior user experiences when it comes to the way that discovery and access work in the GCI.

3. User Requirements

While AIP-1 focused more on the initial capabilities of the GCI and on the initial provision of resources to GEOSS, AIP-2 was able to examine in more depth how users and communities might make best use of GEOSS. This has led to an examination of both the types of users (defined for example in the GEOSS Concept of Operations[2]) and their typical activities (expressed in use cases) as guides to further requirements for GCI functionality.

3.1 User types

A set of user types is listed in Table 1.

2 http://www.earthobservations.com/documents/excom/ec14/09_Concept%20of%20Operations%20Document%20GEOSS%20Common%20Infrastructure.pdf

Table 1 – GEOSS User Types

Type of User	Description	Examples of Interoperability Arrangements
Publisher	Individual(s) authorized by Member and Participating Organizations to commit GEOSS Components and/or Services	ISO19115, GML, SensorML
Operator	The agency/organization responsible for the operation and maintenance of a committed service and related data	HTTP, CSW, WMS, KML, GeoRSS
Approver	Acts to approve or disapprove an entry or update in one of the GEOSS Registries and the GEO Web Portals	ISO19115, CSW, ebXML, GEOSS Record
Integrator	A class of user typically engaged in support of one or more application areas who is able to use GEOSS to locate suitable services, data, and related resources, and to develop and deploy integrating software solutions (e.g. applications) that cater to a specific context or subject area	CSW, SRU, SOS, NetCDF, OpenDaP, WCS, WPS
Experienced User	Users who understand the concepts of GEOSS and seek registered resources through the GEO Web Portal interface or desktop applications	O&M, NetCDF, WMS, WCS, KML, GeoRSS
Issue-oriented User	Researchers and science-to-policy analysts who work on specific issues that fall within one or more Societal Benefit Areas.	HTTP, HTML, KML, GeoRSS

3.2 Discovery and Access Use Cases

The connection between user types and discovery activities is described by discovery use cases. The AIP-2 Use Cases Engineering Report[3] describes a fundamental set of GEOSS use cases in detail. The report contains generalized use cases intended to be applicable across a spectrum of GEOSS users and applications. Specialized use cases derived from the general ones are then defined in separate AIP-2 SBA Scenario ER's. In this report, the relationship of the discovery and access use cases to E2EDA is particularly considered.

Figure 2 illustrates some of the relationships among the generalized use cases. Those three to the left involve a user or consumer of GEOSS resources, while the four to the right involve a publisher or provider of those resources. The three use cases in the middle primarily involve machines and software, for example client applications and Web services, as actors. While E2EDA focuses on the user experiences and provider responsibilities, successful implementation of all of the use cases, including machine-centric ones, is necessary to achieve the desired end-to-end flow of information.

[3] http://www.ogcnetwork.net/AIP2ERs#UseCases

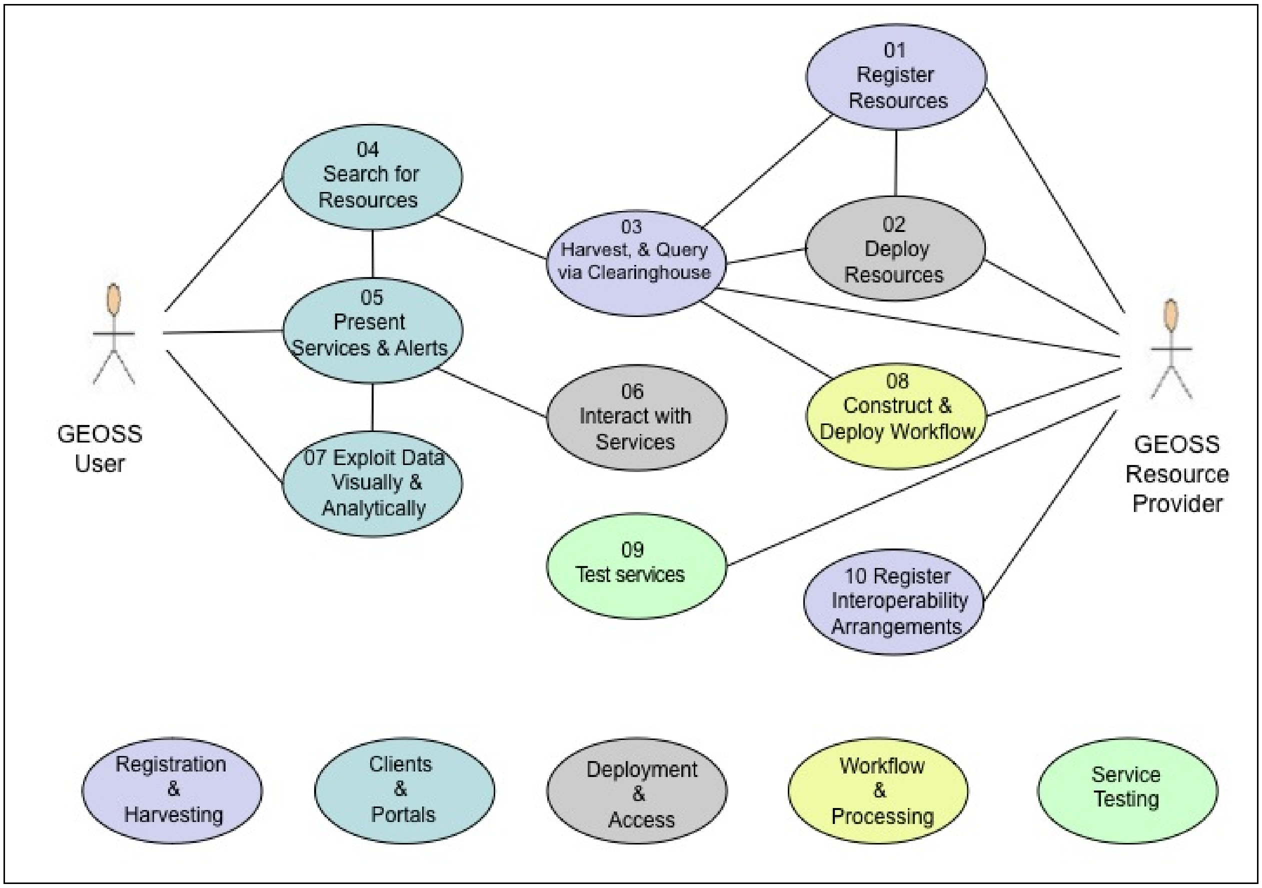

Figure 2 – Use Cases for the AIP Service-Oriented Architecture

Figure 2 presents the principal "transverse" use cases. The use cases are keyed by color to the type of activity they describe. Lines indicate the main actors in each use case. For example, Use Case 5 involves both a GEOSS user and a client application. Use Case 6 involves the interaction of that client application with one or more deployed service resources (latter connection is not shown for clarity). Use Case 02 in turn involves the publishing of the service resource by a GEOSS service provider.

3.2.1 Discovery use cases

Discovery use cases primarily involve interactions with metadata such as creating it, deploying it to metadata services, registering it with the GEOSS Registry, querying and presenting it. Successful discovery requires metadata to be a reliable description of the resources being sought. Discovery workflow typically also involves approval that the published metadata is complete and correct. The availability and functionality of described services also need to be verified through some sort of testing.

Table 2 describes which principal discovery use cases shown in Figure 2 might involve particular types of (human) users.

Table 2 – Discovery Use Cases by User Type

Type of User	**Description**	**Use Case Involvement**
Publisher	Individual(s) authorized by Member and Participating Organizations to commit GEOSS Components and/or Services	Register Resources, Register Interoperability Arrangements
Operator	The agency/organization responsible for the operation and maintenance of a committed service and related data	Deploy Resources
Approver	Acts to approve or disapprove an entry or update in one of the GEOSS Registries and the GEO Web Portals. Approval may be formal or informal and take place in any of the three GCI elements, including Clearinghouses that access indirectly registered metadata and Portals that access external clearinghouses.	Register Resources, Test Services
Integrator	Discovers resources for the purpose of developing and deploying integrated software solutions that cater to a specific context or subject area.	Search for Resources, Present Services and Alerts, Construct and Deploy Workflow
Experienced User	Capable of navigating service-oriented computing protocols and components, but may prefer to focus on specific knowledge domain activities.	Search for Resources, Present Services and Alerts
Issues-oriented User	User looking for geospatial resources to support addressing a specific issue within their own GIS/analysis environment.	Search for Resources, Present Services and Alerts

3.2.2 Access use cases

Access use cases involve both the availability (ability to be accessed) and usability (ability to be interacted with) of resources through interoperability arrangements that can be implemented by providers and utilized by consumers. For example, a user of a Web service for data access typically needs to find both the service endpoint where resources of interest are available and a suitable client for that type of service that interacts successfully with that particular service and provides functionality to the user.

Table 3 lists the principal access use cases shown in Figure 2 that would typically involve particular types of users.

Table 3 – Access Use Cases by User Type

Type of User	Description	Use Case Involvement
Operator	Individual(s) authorized by Member and Participating Organizations to commit GEOSS Components and/or Services	Deploy Resources Test Services
Experienced User	Individual connecting to a component or service registered with GEOSS, utilizing a variety of lightweight and heavyweight clients	Interact with services
Issue-oriented User	User wanting to utilize geospatial resources to support addressing a specific issue. More likely to work with a browser-based client application following links provided directly in discovery results.	Interact with services

3.3 Roles of E2EDA in community scenarios

Figure 2 shows that end-to-end discovery and also access depend on threads of use cases being implemented and actuated in concert. The transverse use cases also stand as general / conceptual activities to be specialized and instantiated for community scenarios. Support for E2EDA should then be visible as connected steps that derive from the use case threads. As Table 4 shows, this visibility is spotty within the documented community scenarios. There appear to be two reasons for this. To be fair, articulation of the E2EDA thread concept developed at a late stage of AIP-2. The second reason is a continued focus in the community scenarios on solving community problems with collected and known resources, rather than on increasing the visibility of the resources themselves within GEOSS.

Table 4 – E2EDA in Community Scenarios

Scenario	E2EDA "thread"	Relevant Generalized Use Cases
Air Quality	A Web-accessible folder is registered as a "Community Catalog" in GEOSS Registry	01 – Register resources
Renewable Energy	JSR-168 portlet is registered as a "service" in GEOSS Registry	01 – Register resources
Pika Distribution (CC & Biodiversity)	A brokering and mediation component performs searches for resources on behalf of a client application	04 - Search for resources
Polar Ecosystems (CC & Biodiversity)	Scientist discovers IP3 Client Application in GEOSS Registry	04 - Search for resources
Disaster Management and Response	Specific RSS feeds and resources are searched for in the ESA EO Clearinghouse as well as the GEOSS Clearinghouse	04 - Search for resources

4. Metadata

4.1 Roles for metadata in discovery and access

Discovery and access of resources is supported by metadata. The GEOSS components and services are described by various metadata standards and profiles, typically driven by the context in which they were created or organization that created them.

4.2 Implications of user expectations for provider documentation

There is a difference between describing something so that it can be discovered and describing something so that it can be fully understood. When looking at current trends on the Web where people share videos, photos, and such, there is a trend to provide very brief information: a title, description, and a reference to the resource. This information is often sufficient for general discovery purposes. To fully understand the resource in order to assess its usefulness or applicability to GEOSS actors like the Experienced User or the Issue-oriented User, additional information is necessary.

Based on these general Web experiences, Issue-oriented Users expect from GEO Portals to find relevant resources based on a simple search interface that is similar to what they encounter on the Web. This search interface may support keyword as well as spatial searches. The users will also expect fast responses. The combination allows them to iterate through searches, modify searches by adding more specific words, limiting results to a specific area and such.

4.3 Metadata standards and crosswalks

With the advent of ISO 19115/19119/19139 there is a framework for describing these resources more fully for the purpose of understanding. However, these international standards allow for variations through the creation of profiles. The large SDI initiatives (North America and INSPIRE particularly) already have created profiles of the ISO standard that are different in various aspects. At the core (metadata to support discovery) there is generally sufficient overlap but considerable work in AIP-2 was needed to even define this overlap and document resulting gaps[4].

4.4 GeossRecord as a common denominator

A “common” common denominator between the various metadata standards in use may be defined as the following items for search:

- Unique identifier
- Title
- Abstract
- Spatial extent of the resource
- Reference to resource
- Type of resource
- Last date of modification of the resource

This corresponds reasonably well to the csw:Record structure as defined by OGC in the CSW specification, which consists largely of the more common Dublin Core elements. There have been a number of indications in both AIP-1

[4] http://spreadsheets.google.com/ccc?key=pG0cD35SB_A-4LRkQTesePQ&hl=en&pli=1

and AIP-2, however, that certain additional or more precise metadata elements play a critical role in discovery of earth observation resources. In csw:Record, for example, an ows:BoundingBox is included for greater specificity and ease of processing than what might be provided for by dct:Spatial. Three The following additional / more specific parameters have been proposed:

- Observed footprint – particularly in the case of satellite observations, a bounding box may be too coarse to be sufficiently discriminating, and there are many cases where the location an observation was made is not in fact identical to the footprint of that part of the earth to which the observation applies ("feature of interest").
- Observed time period – this is another case where any number of temporal markers may be described in metadata (publication, update, processing, etc.), but for earth observations the critical element is the time instant or duration that the observation event actually discerned.
- Phenomenon – while this may roughly correspond to dc:subject, a precise indication of what the observation served to estimate (not necessarily the same as what was measured) is the critical property for most earth observation applications. This may or may not include the other indicator of what was examined, medium.
- Medium – the type of earth material that exhibits the observed phenomenon (e.g. water versus air temperature). This often is the critical property that distinguishes observation data of interest to a particular community.
- Model – this term is used broadly to describe either the process / sensor used to derive an observation from measurements, or the simulation model used to predict or interpret an observation from older or indirect measurements. This is another parameter that is critical for both discovery and selection of EO resources, and is only vaguely accounted for in more general metadata elements (e.g. dc:source)

The combination of csw:Record elements and the above additional EO parameters has been suggested as a mandatory "geoss:Record" for contributed GEOSS resource descriptions.

This discussion describes the common catalog service queryables, however, with respect to the returned metadata, it is recommended that catalog services ensure inclusion in the response fields of URL's both to the full metadata, and to the data/service endpoint. Existing csw:Record fields could hold this information, but clarification of the format and content mapping should be considered both in GEOSS and in CSW. Another recommendation from practice is to encourage use of more robust, standard metadata content – with a recommendation of ISO 19115/19139 and specific service profiles. This would greatly assist in defining minimum interoperability at an actionable level for discovery.

4.5 Data-><-Service and other resource associations

Some of the specifications in use allow for associations between metadata. These associations typically signify associations between resources (a service exposes a dataset, service metadata references dataset metadata, a client application binds to a service). For many discovery and access scenarios, they form essential queryable parameters, but are not yet well supported or managed in either the generally used metadata records or in the present registry and catalog implementations.

5. Registration

5.1 Registration types, roles, and goals

Table 5 – GEOSS Registration Component Types

Component	Description	Example Service-based Interoperability Arrangements
Observing System or Sensor Network	System or network capable of generating observation data	Sensor Planning Service
Exchange and Dissemination System	System for disseminating earth observations	DVD production service
Modeling and Data Processing Center	Computing resource	Globus grid services
Dataset or Database	Earth Observation data product provided through one or more service interfaces	Web Map Service
Catalog, Registry, Metadata Collection	Metadata holding	OpenSearch
Portal or website	User client software able to bind to one or more service interfaces	Catalog Service for the Web
Software or application	User client software able to bind to one or more service interfaces	Web Coverage Service
Computational model	Model used to analyze, interpret, derive, simulate observations	Web Processing Service
Initiative or Programme	Undefined	HTML
Information feed, RSS, or alert	XML document	Atom Publishing Protocol
Training or educational resources	Undefined	SCORM
Web-accessible document, file, or graphic	Anything which a URL can locate?	JPEG

5.2 Standards and Interoperability Registry

The Standards and Interoperability Registry (SIR) supports the registration and discovery of the standards and special arrangements that can be used by GEOSS in support of interoperability. Standards are formally recognized specifications that are published and maintained by a standards development organization, while special arrangements are specifications not formally adopted as a standard, but widely used within one or more communities as if it were a standard. Although a standard or special arrangement proposed for inclusion in the SIR is typically associated with services registered in the Components and Services Registry, it need not be.

Registration of a standard or special arrangement can occur two ways. It can be registered in the first way by navigating to the SIR site and choosing to propose an entry. This function requires that the person wishing to submit the standard or special arrangement be a registered user at the SIR. Once user registration is complete, and the submitter has logged in at the SIR, then the act of registering the standard or special arrangement may take place. The form displayed for the submitter collects information about the proposed entry, as well as contact information.

The fields and their descriptions, other than contact information, are described in Table 6.

Table 6 – SIR Registration Fields

Field Name	Required	Description
Entry Type	Yes	Identifies the proposed entry as a standard or special arrangement.
Name	No	A compact designation to denote the proposed standard or special arrangement.
Version	Yes	The formal version number of the proposed standard or special arrangement.
Title	Yes	The formal title by which the proposed standard or special arrangement should be known. It should include the version and/or the publication date.
Description	Yes	A description of the proposed standard or special arrangement, stating its nature, properties, scope, and other qualities.
Author	No	The name of the body or organization that originally created the proposed standard or special arrangement. This is a pick list of standard development organizations. If OTHER is chosen, then the submitter can enter the author.
Publisher	Yes	The name of the body or organization responsible for the publication and maintenance of the proposed standard or special arrangement. This is a pick list of standard development organizations. If OTHER is chosen, then the submitter can enter the publisher.
Primary Taxonomy Category	Yes	A category from the Earth Observations Standards Taxonomy. This is a pick list showing the current taxonomy supported by the SIR. If a new category is needed, choose the closest category now, and explain in the Comments field the reason for the new category, remembering to supply the new category's proposed name and parent.
Secondary Taxonomy Category	No	A category from the Earth Observations Standards Taxonomy. This is a pick list showing the current taxonomy supported by the SIR. A secondary category is one that also fits the standard or special arrangement, but not as strongly as the primary taxonomy category chosen.
URL	No	The URL that points to where the specifications for, or additional information about, the proposed standard or special arrangement may be obtained.
URN	No	A globally unique, persistent identifier used for recognition or access to the proposed standard or special arrangement, and typically assigned by the standards development organization responsible for the standard or special arrangement.
Superseded URN	No	A globally unique, persistent identifier used for recognition or access to the standard or special arrangement being superseded. This must be chosen from the already registered standards and special arrangements.
Comments	No	Any additional information to help in the evaluation of the proposed standard or special arrangement, such as special circumstances, disciplines served, references to implementations or test harnesses of the interoperability arrangement, etc.

The submitter can either submit the proposed standard or special arrangement, or save a draft of it so it can be completed at a later time. At any time while completing the form, if the submitter encounters a usability or functionality problem, a form can be filled in to report the details of the problem to the SIR administration team. This form can be accessed by clicking on the feedback link in the page.

The second way in which a standard or special arrangement can be proposed for inclusion in the SIR is when a service is being registered at the Components and Services Registry. See Section 5.3 for details. The fields to be supplied are the same, but there is no opportunity to save a draft. Once a standard or special arrangement is submitted from the Components and Services Registry, it is automatically associated with the service being registered.

The SIR supports discovery of registered standards and special arrangements through a search capability. Searches can be basic or advanced. Advanced searches allow searches based upon specific field values, as well as combinations of field values. Searching the SIR does not require logging in, but drafts will not be found through a public search.

5.3 Components and Services Registry

The Components and Services Registry (CSR) supports the registration and discovery of the components and services that are associated with the data sharing goals of GEOSS. The registered components are viewed as GEOSS resources that fall broadly into the categories of an offered system, program, or initiative. Some of the specific choices include observing systems, catalogs, portals, datasets, and information feeds. The registered services describe the means to access separate functional aspects of a registered component. Services can only be registered in association with a component; therefore, components must be registered prior to their associated services. Additionally, each service interface can be associated with one or more GEOSS-registered standards to promote interoperability and accessibility by end users and their software.

In order to register components and services at the CSR, the contributor must be a registered user at the CSR, and be logged in at the CSR. The registration order is components, followed by associated services, and for each service a set of associated standards or special arrangements. These do not all need to be accomplished simultaneously. A CSR user can register a component, and then return to the CSR at a later time to register associated services. Similarly, the set of associated standards and special arrangements for a service can be registered at a later time than the service. As stated in Section 5.2, the registration of standards and special arrangements can take place at the CSR or the SIR.

To register a component or service at the CSR, one navigates to the CSR and chooses the link to contribute a resource to GEOSS. Then, a page is displayed with the choice of registering a component or a service. A service can be registered for an existing component only if the user who created the component is logged in to register the service. Before registering a service, the existing component that the service is to be associated with will need to be provided; then a link will be made available to access the service registration form. There is a separate registration form for components and services, as well as the standards or special arrangements associated with the services. However, the form for standards and special arrangements has exactly the same fields as the SIR form. Except for contact information, the fields for component registration are described in Table 7 and the fields for service registration are described in Table 8.

Table 7 – CSR Component Registration Fields

Field Name	Required	Description
Resource Name	Yes	A long name or title identifying the component.
Abbreviation	No	A compact designation for the component.
Description	Yes	A long description of the component being registered, including its purpose and the functionality offered.
GEO Affiliation	Yes	The Member or Participating Organization of GEO offering this component. This is a pick list with the opportunity to choose more than one affiliation, if necessary.
Responsible Organization	No	The name of the organizational entity that is operating the component, if different than the GEO Sponsor.
URL to Resource	Yes	A URL that describes the component offering in greater detail.
Resource Category	Yes	A set of categories that classifies the component, from which the submitter can choose one or more.
Societal Benefit Areas	Yes	The set of GEOSS Societal Benefit Areas, from which the submitter can choose all that apply to the component.
Resource Availability	Yes	A selection that specifies the operational status for the component.

Table 8 – CSR Service Registration Fields

Field Name	Required	Description
Service Instance Name	Yes	A long name or title identifying the service instance.
Abbreviation	No	A compact designation for the service.
Description	Yes	A long description of the service being registered, including its purpose and standards or special arrangements used.
Service Information URL	Yes	A human-readable service description URL referring to the service implementation, e.g. HTML documentation or metadata file for the service interface.
Service Interface URL	Yes	A URL that is used by software to invoke the service, e.g. WSDL, OGC GetCapabilities, data download/access URL, CGI script, SOAP service endpoint, etc.
Service Geographic Extent	No	The northernmost latitude value, southernmost latitude value, easternmost longitude value, and westernmost longitude value. The values are entered as decimal numbers following the EPSG:4326 coordinate reference system.
Service Time Period of Information Content	No	Dates signifying when the information provided by the service begins and ends. The begin date can be indefinite, and the end date can be ongoing.
Standards / Special Arrangements Reference Information	No	Two lists composed of taxonomy categories from the SIR, with registered standards and special arrangements under their respective categories. Choosing an existing standard or special arrangement immediately associates it with the service being registered. Choosing a category allows the completion of the standard or special arrangement entry form, and associates the proposed standard or special arrangement with the service being registered.

The submitter can either submit the proposed component or service, or save a draft of it so it can be completed at a later time. At any time while completing the form, if the submitter encounters a usability or functionality problem, a form can be filled in to report the details of the problem to the CSR administration team. This form can be accessed by clicking the feedback link on the page.

The CSR supports discovery of registered components and services through a user interface and a programmatic interface. The user interface allows basic and advanced searches. Advanced searches allow searches based upon specific field values, as well as combinations of field values. Programmatic searches are facilitated via an exposed OGC CSW interface. Searching the CSR does not require logging in, but neither drafts nor pending submissions will be found through a public search.

5.4 Best Practices Wiki

Registration of recommended "best" practices is supported by means of the Best Practices Wiki. This was not particularly exercised during AIP-2 but there was rough consensus that focusing interoperability arrangements by means of recorded practice will be necessary for an effective GCI.

6. Clearinghouse

6.1 Overview

Clearinghouses provide federated and uniform access to distributed metadata and also include some facilities to manage the content of the clearinghouse:

- Publish metadata
- Validate metadata
- Remove metadata
- Approve metadata
- Manage harvesting process – ingest, update, map, integrate
- Manage cascading query generation

6.2 Clearinghouse capabilities and examples

During AIP-2, three GEOSS Clearinghouses were in operation:

- ESRI
- Compusult
- FGDC

6.2.1 ESRI GEOSS Clearinghouse

The ESRI Clearinghouse and ESRI GEO Web Portal are built using the ESRI ArcGIS Server Geoportal Extension, a complete solution for discovery and access to geospatial resources, supporting:

- Searching with text and spatial ranking and thesaurus support.
- Discovery and publishing of content through OGC CSW 2.0.2 (KVP + SOAP) and REST interfaces as well as OpenSearch support.
- Clearinghouse management functionality.
- Various workflows for publishing and validating clearinghouse content, including online forms, registering services of any kind, and harvesting remote clearinghouses.
- Extensible support for metadata schemas and profiles.

6.2.2 Compusult Clearinghouse

The Compusult Clearinghouse and GEO Web Portal are based on WES Catalog, a component included with Compusult's Web Enterprise Suite series of products, supporting

- Automatic harvest and periodic re-harvest of OGC Web Services and other standards-based metadata
- Automatic registration and export of metadata in multiple formats (FGDC, ISO, DDMS, Dublin Core)
- Supports access to metadata content from Z39.50 catalog servers
- Deployment as an OGC CSW 2.0.2 Web Service
- Query capabilities including spatial regions, keywords, temporal, service types, etc..
- Distributed search of remote registered CSW catalogs.
- Automatic generation of RSS feeds for latest information published

6.2.3 FGDC Clearinghouse

The FGDC Simple Clearinghouse is a prototype GEOSS Clearinghouse instance consisting of:

- A small database schema of fields common to most metadata records An instantiation of the database schema (currently in MySQL)
- A mapping between the database fields and metadata elements in various metadata formats (e.g., FGDC, ISO10119)
- A series of perl scripts to harvest and ingest metadata records

- A CGI perl script to serve as a demonstration web-based interface for searching and displaying records
- A web-based administrative interface

The Clearinghouse queries the GEOSS Service Registry for relevant services (catalogs primarily but not exclusively) and stores the service entry points from those services to use for later harvesting.

6.3 Clearinghouse-><-Registry interactions

The GEOSS Clearinghouses connect to the GEOSS Components and Services Registry (CSR) to discover registered components and services. Some of the components are themselves metadata catalogs. The Clearinghouses index the components/services and may harvest the referenced metadata catalogs. The interaction between the GEO Clearinghouses and the CSR is automated through a periodic harvesting mechanism.

The CSR implements an ebRIM type information model, exposed through an OGC CSW interface. The GEOSS Clearinghouses may themselves implement different information models, some based on OGCCORE, others based on the ISO Application profile of CSW. This may result in differences in handling the information retrieved from the GEOSS Registry. As long as this does not negatively affect the findability of the components and services registered in the GEOSS Registry, this diversity may not be a significant problem.

6.3.1 Registration and Clearinghouse policy

In order to make harvesting metadata records from community catalogs into Clearinghouses both useful to Portals and relatively non-invasive to the remote site, a number of harvest-related parameters and constraints should be made explicit. These constraints should generally be stated by the catalog sites to be harvested, since their content and bandwidth will be affected the most by the harvesting process. The natural place for these constraints to be specified is in the Service Records held in the Component and Service Registry (CSR). Doing so would make the constraints an intrinsic part of the service discovery mechanism for Clearinghouses and also would make specification of harvesting constraints part of the registration process. Following is a preliminary list of considerations for developing a harvesting strategy that the Clearinghouses could utilize:

- Harvest or Distributed Search?
  - Should the Clearinghouse harvest records from the catalog site (thereby allowing the Clearinghouse to search the records locally), or should queries to the Clearinghouse be relayed to the catalog site?
- If Harvest:
  - How often can the remote catalog site be harvested or updated?
  - Can all records be harvested at one time? Over a period of time? Not at all?
  - If not all, how many records can be harvested at one time and how many total records can be harvested?
  - If not all, what is a query (or queries) which would return a representative subset of records to be made searchable by the Clearinghouse?
  - Is there some sort of record aggregation that can be used to group records which are similar in content (perhaps, differing only in spatial and temporal footprint)?
- If Distributed Search:
  - Does the remote catalog publish a searchable interface (Web Accessible Folders cannot be search by themselves)?
  - Do the remote catalogs support a common set of queryables and returnables to allow for consistent presentation of search results?

- o Are the ranking algorithms used in the remote catalogs compatible? This will affect the ability or usefulness of merging results from more than one remote catalog.
- o Is the remote catalog available and what is its performance? The clearinghouse operator is not in control of these two factors when implementing distributed search.

Note that if harvesting is not permitted, one can sometimes but not generally infer what searches and protocol are permitted from the Service record itself or, where appropriate, from a Capabilities document pointed to by the Service record.

6.4 Harvesting / cascading practice

As mentioned above, the GEO Clearinghouses may harvest metadata holdings from catalogs/clearinghouses that have been registered with the GEOSS Registry. This cascaded harvesting may have some unanticipated side effects, such as indexing an exponentially growing number of resources or indexing resources multiple times that have been registered in more than one catalog/clearinghouse.

The number of resources will affect system performance, but may be addressed by proper sizing of the hardware environment used to host the GEO Clearinghouse. It has to be noted that performance of the GCI is determined not just by the performance of the GEO Clearinghouse, but largely depends on performance of the components and services registered. It is suggested that service level agreements be put in place supported by the proper resources to sustain the desired level of performance and availability.

The issue if resources recur in multiple catalogs may be addressed by ensuring that every metadata document for the resource has a global unique identifier that is part of the metadata document and is persistent when contained in multiple catalogs. The metadata standards that are in use provide some mechanisms to assign such a globally unique identifier, but in common practice these are treated as optional and not forcibly unique. It is suggested that GEOSS engage in discussions with the various standardization bodies to address this issue fundamentally.

6.5 User queries

User queries to the GEO Clearinghouses typically occur through a front-end application as provided by the GEO Portals. Integrators and application developers may choose to interact with the GEO Clearinghouses directly through their client applications. In this case, the application uses the service interface provided by the GEO Clearinghouses.

All of the GEO Clearinghouses that are part of the AIP support the OGC CSW 2.0.2 interface, providing a common denominator in the interaction through use of the OGCCORE mandatory queryable parameters of CSW. The result is that although individual GEO Clearinghouses may implement different information models (ebRIM, ISO, OGCCORE), integrators need not be concerned when accessing the GEO Clearinghouses through the OGCCORE implementation of the CSW 2.0.2 service.

In addition to the CSW interface, some of the GEOSS Clearinghouses (for example the ESRI GEO Clearinghouse) implement support for OpenSearch (http://www.opensearch.org) with a URL-based RESTful interface. This interface supports performing searches of the Clearinghouse from within lightweight Web clients.

7. Geoportals and geo-applications

7.1 Portal functions

GEO Web Portal functions include:

- Discovery and access of resources
- Interaction with the services (e.g. WMS preview in support of determining if service/component is useful for the purposes of the Experienced User or Issue-oriented user)
- Access to community catalogs/portals

- Access to general GCI information
- Organization of GEOSS Resources, e.g. along the lines of Societal Benefit Areas.

Typically a GEO Portal will include some functions that require security in the form of authentication (personalization, contribution of content, administration…). It is suggested that GEOSS investigate the use of common authentication mechanisms or standards (such as OpenID).

7.2 Portal – clearinghouse interactions

GEO Portals may interact with one or more GEO Clearinghouses. The user uses the GEO Portal to formulate queries. The GEO Portal then submits the query to one or more GEO Clearinghouses. Responses from the GEO Clearinghouses are presented to the user in the GEO Portal.

There are several models to consider:

- Work with 1 clearinghouse that contains harvested content (geodata.gov)
- Work with many clearinghouses but direct searches to one clearinghouse at a time (ESRI GEO Portal)
- Work with many clearinghouses and federate searches to more than one clearinghouse at a time (not yet exercised)
- All of the above

7.3 Portal information management

Portals may include a Content Management System (CMS) as a tool to maintain the portal content. This would not maintain content that is obtained from a Clearinghouse, only information, metadata, or documents being made available exclusively through the GEO Web Portal. GEO Web Portals may integrate with these CMS to provide a seamless user experience. Single sign-on support is a key element of this seamless user experience, although this sort of management of unique Portal content, particularly across Portals, is not yet an explicit part of the GCI.

7.4 GEO Web Portals, Community portals, and domain applications

Given Web service interfaces for both the GEOSS CSR and the Clearinghouses, it is of course possible and even desirable for a variety of client applications to be to discover resources within them. This does introduce a variety of challenges, however, to accomplishment of a uniform and successful user discovery and access experience. There are at least three issues here that have been encountered in AIP-2:

1. Different client applications will typically implement different user interfaces for the same service, such as a Clearinghouse. The queries they generate may also not be consistent from application to application, leading to differences in what resources a particular user will find. To be fair, this is also encountered to some extent when utilizing different Clearinghouses.

2. Organizing or categorizing metadata which is unique to a GEO Web or Community Portal will not be available for use in other client applications even when those applications are used against the same Clearinghouse.

3. A domain-specific application will presumably provide a more efficient and capable interface for users from a particular community, but there may be no easy way for that application to then generate Clearinghouse queries that focus on resources of interest to that community. An example from AIP-2 was the AQ community looking for resources related to “air” without any assurance that all possibly relevant resources were actually tagged or described in this way.

8. Interoperability testing of discovery functions

8.1 Interchange tests

When Web services and clients may come from different vendors, operate on different platforms, and exchange different content resources, validation of interoperability among these software packages is critical. Validation is typically carried out through interchange experiments in which, for example, a catalog service is accessed by several different catalog clients, and conversely several different catalog services are accessed by a particular client. Such interchanges were initiated during AIP-2 but have not yet been rigorously completed[5].

8.2 Conformance tests

When interoperability between system components is enhanced by use of standard arrangements, testing of conformance with those standards is a critical validation step. The Test Facility Working Group examined many of the issues involved in such testing[6].

8.3 Status tests

The first step in successful access to a provided Web service, of course, is for that service to be available and responsive. Tests were made with a service status checker provided by USGS[7] against registered GEOSS services to examine the feasibility of this type of testing.

9. Resource provider experiences and practices in AIP-2

This report section discusses some of the participant experiences and developed / piloted practices in AIP-2 as they relate to the theme of end-to-end discovery and access.

9.1 Registration process

The GEOSS registration process was a subject of much discussion and experimentation since it serves something of a dual role. One role is to point the Clearinghouses to metadata for query / harvest. In this role, any sort of path to the metadata that a Clearinghouse can follow will do. Another role, though, is to let a provider declare a contribution to GEOSS. To this end, particular information is useful not only of what is being published but exactly how it is being provided for use in GEOSS.

9.1.1 Direct versus indirect registration

Just as component and service registration can serve two purposes, registration can also be performed in two ways. Using direct registration, a provider registers each service resource (for example, a WMS) directly and then provides a link to somewhere that a service description is provided. Indirect registration involves registering a metadata service (e.g. a catalog) describing provided resources instead of the components and services themselves. This in turn provides metadata about service resources such that their contribution to GEOSS is implied.
Reasonable assumptions can be made that each contribution should show up only once in a Clearinghouse search, and that metadata attributes concerning either the component (e.g. dataset) and service interface (e.g. WMS) might be important for discovery. In this case, there are drawbacks to each registration approach. In the direct method, only one component can be registered “per” service, so registration of more than one dataset exposed through a WMS can result in the WMS appearing multiple times. In the indirect approach, the holdings of a registered metadata service such as a community catalog may not really be intended for GEOSS and generally lack metadata elements included in direct registrations.

[5] https://sites.google.com/site/geosspilot2/Home/testreports

[6] https://sites.google.com/site/geosspilot2/Home/test-facility-working-group/wg-test-facility-best-practices/20081120AIPTestFacilityEngineeringReport.doc

[7] http://registry.fgdc.gov/statuschecker/services/rest/post.php

9.1.2 Component – service and other relationships

In general, the present limitation in the CSR on resource associations (only one-component-to-one-or-more-services) results in some confusion and difficulty in describing provider contributions. This is particularly the case with the different types of components that can be registered. While some additional relationships can be expressed within provided metadata (one example is to provide valid metadata document links for each layer in a WMS), this is not often done in metadata and even more rarely utilized for discovery. In a more general sense, relationships and the links that represent them are the cornerstones of Web search, so being able to register or otherwise make them available for discovery would have additional benefits.

9.1.3 Metadata construction and registration

A number of AIP-2 participants, particularly those developing Community Catalogs for the first time (e.g. as Web-accessible Folder types of metadata service provision) worked on development of metadata documents for their services and other resource contributions. A particular challenge for them, with guidance from Habermann and others, was to balance selection of the metadata elements for a variety of purposes, from Clearinghouse harvest and user discovery methods to dataset evaluation and selection of most appropriate service clients. One advantage of registering metadata created specifically for GEOSS, at least, is that metadata usage particular to GEOSS discovery and access scenarios can be accounted for in the process. This was noted particularly in regards to Web services, where OGC-style capabilities documents and WSDL documents generally do not have all of the metadata elements important to user discovery needs.

9.1.4 Clearinghouse preference expression

Since exposure of registered resources through a GEOSS Clearinghouse is somewhat unique to GEOSS, there are preferences and guidance to be expressed as to how this should be done and maintained. Such information is not normally included in metadata descriptions and is not yet part of the attributes of the registration process. Examples include the frequency with which metadata changes and should be re-harvested, whether a metadata service should be harvested versus accessed via distributed query, and whether a service registered for one component is in fact the same service instance as one registered for another component.

9.2 Findability

Issues of "findability" relate both to whether a user query shows up the resources it intends to (e.g. all air quality-related resources), whether only the resources that in fact match the query show up in the results (i.e. do all the needles show up without the haystack) and whether enough information is presented to be able to access relevant resources.

9.2.1 Queryable parameters

The most important factor in structured discovery is whether the right queryable parameters are available and have been registered for the resources of interest. An example discussed earlier concerns whether a set of observations has been described accurately by the phenomena that were observed. A second concern for queryable parameters is whether the domain of valid values for each queryable parameter is available (e.g. in the form of hints) to guide the user in forming successful searches. In the End-to-end sense, this guidance also needs to be arranged effectively for the user of a search client application such as a GEO Portal and optimally filter information resources to correspond to a user's available means of access (e.g. client applications). Where possible, parameter domains should be also conform to interoperability arrangements such as standard taxonomies in order to further simplify and facilitate successful query strategies.

9.2.2 General versus GEOSS-specific publication.

As discussed above, many instances were uncovered in AIP-2 where GEOSS-specific metadata and methods of publication contribute significantly to successful E2EDA. For example, given GEO Portal capabilities to support WMS access to discovered resources, links to such an interface in the metadata, and provision of WMS interfaces to resources at a the desired granularity (e.g. sensor or group of sensors) have particular value. Creating and maintaining metadata separately for GEOSS and non-GEOSS publication, however, can be a sizeable burden for some participants.

9.2.3 Global identifiers

Under the category of minimizing haystacks, ambiguities in the GEOSS registration process as well as multiple community catalog registrations, challenges in asserting resource relationships, and version or lineage issues can lead to too many query responses, unless an agreed global identifier scheme allows true uniqueness of resources to be determined. One particular manifestation of this concerned provider interest in registering both a resource and a metadata service holding for that resource without double entries showing up in Clearinghouse queries.

9.2.4 Drill-down capabilities

A common requirement expressed by a number of participants involved users being able to discover high-level resource collections (e.g. entire sensor nets) or entities (e.g. organization involved in meteorology) of interest, and then drill-down to sub-collections and even individual data elements through iterative discovery and access. This constitutes as well another sense of E2EDA, describing levels of exploration running from high-level discovery to fine-grained access and evaluation. Successful implementation of this capability requires not only the right metadata and links at multiple collection summary levels but also widespread publication of services to facilitate the fine-grained access end of the drill-down spectrum, such as a WMS implementing time and/or sample dimension access to single observation layers.

9.3 Sustainability

One of the goals of GCI is to support discovery and access to earth observation resources in as “live” a manner as is feasible. Several issues worked during the Pilot pertained to maintaining the currency of data published through GEOSS as well facilitating up-to-date discovery of the holdings exposed by other services.

9.3.1 Harvesting and query distribution options

One question as to *resource currency* is how often a particular component or service resource is updated and how extensive a metadata resource might be. There was interest in expressing for benefit of the Clearinghouses how often the metadata for a resource should be re-harvested and/or whether instead queries to the Clearinghouses should be distributed dynamically to metadata services such as Community Catalogs. There have been arguments in favor of different locations for denoting these preferences. Alternatives include the metadata records themselves, the CSR registry records, or with directly with each Clearinghouse. It makes sense from the point of view of clarity and responsibility to do this in the CSR, but does introduce additional overhead to keep the preferences themselves up to date.

9.3.2 Resource update and versioning / lineage

In the case of both dynamic and derivative earth observation information discovered and accessed through GEOSS, there is a need for rather more metadata than might ordinarily be the case, so that in the end a user can query or determine the currency and validity of available data. For example, time descriptions may need to indicate:

1. When an observation was made
2. When or what version of information (model result, processed feature of interest) was derived from the observation.
3. When the description of the above information was last updated.
4. When a metadata service resource, e.g. Clearinghouse holding, was last updated to check whether 3) had been updated.
5. And so on…

In the larger theme of versioning and lineage, while there are general metadata elements such as in ISO 19114 to describe these properties, but it was clear during AIP-2 that best practices for usage of these elements should be agreed and followed. Issues of version notification were raised in the course of the pilot, e.g. changes to processed observation results triggered by subsequent inter-calibration adjustments, but this was not explored in any detail.

9.3.3 Testing

As discussed earlier in this report, several types of testing were recognized as relevant to users being able to rely on service resources published to GEOSS. In the case of testing interoperability, there is clearly a tradeoff between welcoming diverse resource contributions and being able to exercise and document a useful range of interchanges between data sources, service instances, client applications, and types of users. This is another area within which narrower best practices were recognized as critical to effective E2EDA, but not a lot of progress could be made in this direction during AIP-2.

9.3.4 Community metadata and social networking

It is clear that formal metadata and registration is only a part of what most users rely on for finding and getting the most out of information resources. Word of mouth and expert recommendations can be even more important. This was recognized, even for expert users themselves and certainly for facilitating cross-community and "unexpected" resource discovery. Facilities for this inside or outside of GCI were not, however, stood up or exercised to any significant extent during AIP-2.

9.4 Meeting user needs

During the last phases of AIP-2 in the context of hashing out the E2EDA concepts, new ad hoc and anecdotal information on user needs provoked extensive discussion.

9.4.1 Structured query

Service-oriented architecture and traditional record-oriented catalog practices tend toward queries in which values for a set of metadata parameters needs to be specified by the user in order to filter and find resources. This process sounds good in concept, but becomes increasingly problematic for many users as the number and diversity of resources increases and even expert / technical users become accustomed to Web search methods. Three issues at least have been raised in this vein:

- Diverse resources are typically described by diverse parameters, so some but not all potentially relevant resources can be found using a particular query strategy.
- "Hints" or "Facets" as to valid parameter values and available holdings for a parameter value important for successful discovery and winnowing down of results, but not widely supported in structured query facilities
- Full-text search is important for "unexpected" discoveries, but needs to be supported by effective ranking methods, which methods are usually supported by analysis of links between resources and informal metadata about them (e.g. tags, recommendations, prior usage)

An interesting case in point was developed by the AQ-Health working group. They attempted to find their own and other registered resources by looking for anything relevant to "air". It was unclear which search field could be used uniformly for this purpose, even keywords, and text search did not generally support both effective expansion (e.g. *atmosphere* as well as *air*) and filtering in order to "find the needles".

9.4.2 Application development / integration

Much of the activity in the SBA working groups centered around development of scenario "applications" comprising relevant data or data options, services, clients to access those services, and workflow to bring these elements together into a beneficial outcome. This arena is where structured query is generally most beneficial as the application developers know what they are looking for, how it is typed and described, and what clients can be used or developed to exploit the utilized services. The role of discovery, however, in the AIP-2 scenario applications remained limited, although more extensive than in AIP-1. To some extent, this is a reflection of the expected abundance of resources. As such applications are reused and discoverable GEOSS resources grow, the usefulness of discovery should grow as well. It is not yet clear given this lack of exercise whether the scenario application work to date has developed a practice that will be able to find and integrate new resources effectively in the future.

9.4.3 Unexpected finds

An important objective of E2EDA through the GCI is the "unexpected" discovery and utilization of data in domains and applications for which it was not created or intended. Of course it is difficult to plan for the unexpected but "fortune favors the prepared". Most reports of unexpected use describe this arising from personal and informal communications. It would be valuable to look at what information these communications are based on and what patterns of discovery may be involved, so any relevant preparation can be made. This might even be as simple as a WMS "map of the day" offered up by GEO portals which is related in some way to the text of recent queries, or it might involve ad hoc tagging. Indexing of GCI holdings by Google Custom Search was explored by Nebert during AIP-2 and might provide another approach to this sort of preparation for serendipity.

9.5 Other issues and discussion

9.5.1 Is registration relevant to Web search?

As GEO looks at Web search paradigms to enhance discovery, it is worth considering whether a registration / clearinghouse / query process really can provide much of the hypermedia richness which is required for Web search to be effective. Much of what lies behind the REST principles of information architecture has to do not with formal publication of metadata but with exposing as much as possible of the relational riches that are contained within content resources by way of Web protocols, namely HTTP. This issue was discussed during AIP-2 but not explored in detail and should form part of subsequent work.

9.5.2 Determining and communicating "best standards"

There has clearly been a progression in the development of GCI from a "bring whatever you have" approach to recognizing that a limited set of best standard interoperability arrangements are the way to improve the effectiveness of all aspects of GEOSS discovery, access, and exploitation. In particular, this is the shortest route to developing trust that if scenario applications support specific arrangements, then future resources such as new services will likely be compatible as well.

9.5.3 Framework data and standard data types

It is no accident that by far the most common form of data integration, particularly in what are termed "mashups" involves combining diverse datasets on a map. While it is clear that the geospatial aspect of earth observation data is critical to address for GEOSS, a reliable basis for global geospatial / geodetic interoperability has so far not been developed. There seemed to be a good consensus at the end of AIP-2 that remedying this gap with framework data and standard concrete data types (e.g. not just GML, but "Level-2 political boundaries") would be a worthwhile and important next step.

9.5.4 Role and feasibility of mediation between GEOSS communities

Several SBA scenarios, particularly the biodiversity ones, utilize explicit mediation components to help users discovery and access data and services which they may not have the vocabulary or protocol capabilities to work with on their (or their client's) own. To a less formal extent, both the Clearinghouses and the GEO Portals are intended to provide some mediation by organizing resources according to more familiar categories. What was not achieved was explicit representation of the mappings and inferences in order for mediation itself to become an interoperable resource. This is another area for further productive work.

10. Recommendations

Some other recommendations for future work on E2EDA and GCI are listed below.

1. GEOSS should anticipate heterogeneous metadata but promote minimum documentation for specific purposes, e.g. GEOSSRecord for discovery.
2. GEOSS should aim to use existing standards/specifications and work through the proper channels where modifications to these standards/specifications are deemed appropriate.
3. GEOSS should perform an analysis of the expected system load on the GCI in terms of numbers of users, numbers of resources. GEOSS should define (realistic) performance requirements based on the usability testing experiences.
4. It is suggested that service level agreements be put in place supported by the proper resources to sustain the desired level of performance and availability.
5. It is suggested that GEOSS engage in discussions with the appropriate standardization bodies to address the issue of including a globally unique metadata identifier as a mandatory item in the metadata specifications.
6. It is suggested that GEOSS investigate the use of common authentication mechanisms or standards (such as OpenID)