The GEOSS Interoperability Process Pilot Project (IP3)

Siri Jodha Singh Khalsa, Senior Member, IEEE, Stefano Nativi, Member, IEEE, and Gary N. Geller, Member, IEEE

Abstract—There is an ever-increasing need to integrate knowledge from the diverse disciplines engaged in studying the constituent parts of the complex Earth system. With the emergence of the Global Earth Observation System of Systems (GEOSS), which is bringing together thousands of previously isolated Earth observing systems, the necessity of establishing methods that will help in the integration of varied discipline information systems becomes even more urgent. The Group on Earth Observations (GEO) was established to oversee the creation of GEOSS which seeks to advance the convergence of Earth observing systems based on interoperability arrangements agreed to by consensus. We describe the specific approaches that GEO has proposed for achieving interoperability among its component systems and give an overview of the GEOSS Interoperability Process Pilot Project (IP3). The IP3 is helping to develop an advanced information infrastructure that supports the formation and operation of Earth System Science communities based on cross-disciplinary information exchange. This means moving from discipline-specific monolithic data-centric systems toward modular service-oriented information systems. GEOSS seeks to provide scientists, researchers, and decision makers with a persistent set of independent but interoperable information services that can be applied to address many pressing societal issues. The IP3 is furthering this cause by piloting a framework for multidisciplinary knowledge integration.

Index Terms—Data processing, informatics, interoperability, modeling.

I. GEOSS INITIATIVE

Environmental change, population stress, susceptibility to natural disasters, and increasingly interdependent economies are driving the need for comprehensive sustained global environmental monitoring. To address this need, governments are collaborating through the Group on Earth Observations (GEO) to link together thousands of observing systems that until now have operated in isolation. GEO is creating a Global Earth Observation System of Systems (GEOSS) to bring societal benefits through improved access to, and synergy among, current and future Earth observing systems.

GEO was established by a series of three ministerial-level summits. It currently includes 74 member countries, the European Commission, and 51 intergovernmental, international, and regional organizations with a mandate in Earth observation or related issues. GEOSS is being built initially from existing systems and initiatives, with an emphasis on the creation of synergies among GEOSS components that provide increased benefits to society. The goal is to leverage existing programs and established standards wherever possible and to broaden convergence of systems based on interoperability arrangements which have been agreed to by consensus.

A. GEOSS Architectural Principles

Components provided by GEO members and participating organizations include observing, processing, modeling, and data dissemination capabilities. These systems are built to serve particular needs and operate within their own mandates. However, those systems should also be designed or adapted so that their interfaces support interoperability with other systems. The GEOSS architecture principles describe how contributed components fit together to produce an overall system that is capable of providing data and information to meet societal needs better than the individual components operating in isolation.

In planning for GEOSS, the impracticality of building a monolithic data system through which all Earth observation data would be managed was recognized. Instead, the GEOSS 10-Year Implementation Plan [1] outlines a service-oriented architecture (SOA) for GEOSS, in which contributed components interact by passing structured messages over network communication services. Such interactions will take place according to agreed-to “ interoperability arrangements” that should be based on nonproprietary open standards. Rather than attempting to define new specifications, GEOSS seeks to recognize standard specifications agreed to by consensus, with preference given to formal international standards such as ISO.

B. SOA Approach

SOA is based on the notion that it is beneficial to decompose a large problem into a collection of smaller related pieces. This helps to establish a high form of abstraction that encapsulates both application and process logic. Services act as “black boxes,” hiding their details from the outside world [2].

For the Earth system domain, an SOA offers considerable flexibility in aligning information technology functions and
processes. SOA is a flexible extensible architectural framework that reduces cost, increases revenue, and enables rapid application delivery and integration across organizations and “siloed” applications [3].

In SOA, the service provider publishes a description of the service(s) it offers via the service registry (Fig. 1). The service consumer, which may be either a person or process, searches the service registry to find a service that meets a particular need. Therefore, a client might not know which procedure it will call until it starts, searches the registry, and identifies a suitable candidate. The goal is total modularization of the distributed computing environment as opposed to recreating the large monolithic solutions of more traditional platforms [4].

This paradigm is useful for efficiently organizing and utilizing distributed capabilities that may be under the control of different ownership domains, as is the case in GEOSS. It provides a uniform means to offer, discover, interact with, and use heterogeneous resources. In fact, a computer system or software is componentized as a service; interoperability between services operating on diverse platforms and between applications implemented in different languages is provided by adopting open standards and self-describing models such as XML. Thus, services are accessed from other services easily. This paradigm supports diverse processing requirements, enables cross-platform communication, and adapts dynamically to meet changing needs [5].

For global Earth observations in particular, interoperability will be achieved through SOA international standards and Earth system science multidisciplinary best practices. Earth system information is encoded using representations (models) agreed to by each particular community, while SOA standards are built using a combination of industry specifications. Based on these, diverse but interoperable Earth science disciplinary information systems can interoperate, thereby enhancing the quality of global Earth information delivered to society.

C. Interoperability as the Basis for GEOSS

All GEOSS components are bound by the requirements on contributed systems as stated in the GEOSS 10-Year Implementation Plan [1] and its companion reference document [6]. Quoting the latter, “The success of GEOSS will depend on data and information providers accepting and implementing a set of interoperability arrangements, including technical specifications for collecting, processing, storing, and disseminating shared data, metadata, and products.” As a consequence, the initial work of the GEO Architecture and Data Committee (ADC) focused on interoperability situations that were surfaced by actual requirements to connect contributed systems through GEOSS “interoperability arrangements.”

Utilization of Earth observation data for cross-disciplinary research and applications requires infrastructure that supports diverse data sources. The key to GEOSS utility and effectiveness is interoperability of data and information systems. Interoperability must be addressed not only through conceptual solutions but also through prototypes and demonstration of functionality.

D. GEOSS Core Components

Fundamental to the operation of GEOSS is access to information about its components. GEOSS uses a “Components Registry” and a “Services Registry” to store information about components and the services these components offer (Fig. 2). The interoperability arrangements that are used by GEOSS services are described in the “Standards and Interoperability Registry.” Fig. 2 also shows additional GEOSS Core Infrastructure components that were not yet implemented at the time of the initial “Interoperability Process Pilot Project” (IP3). These are the Requirements Registry, where user requirements for GEOSS are captured, and the Best Practices Wiki, an open resource for the aggregation and community review of best practices in all fields of Earth observation. All GEOSS registries are interoperable and accessible via the GEOSS Clearinghouse. The GEOSS Standards Registry and the Web-based interfaces to it are being hosted by the IEEE.

II. IP3

The architectural principles of GEOSS provide the basis for interoperability. Building from these principles, the GEO ADC outlined specific approaches that GEOSS should implement for achieving interoperability among its component systems. Recognizing the need to begin implementing the GEOSS infrastructure and testing this process for reaching interoperability, the ADC initiated the IP3 in July 2006 and instructed GEO Task AR-06-01 to develop a series of projects involving different societal benefit areas (SBAs), leading up to a suite of demonstrations suitable for presentation at the GEO Cape Town Ministerial Summit in November 2007.

In this paper, we describe the phases and status of the IP3, identify the system components, and discuss the standard contracts (interface protocols), interoperability agreements, and best practices currently used by these systems. This information is being captured in Web-accessible components which implement the infrastructure discovery functionalities: catalogs and registries that are part of the core GEOSS architecture.

A. IP3 Development Phases

The IP3 approached its goals in phases. These phases will be repeated with new use cases and components, or with new
interfaces to existing components, to encompass other disciplines and SBAs.

Phase I—populate the GEOSS registers. The goal of this phase is to register the participating systems in the GEOSS registries and to identify the standards, interface protocols, and other interoperability arrangements they use. The main actions of this phase are to identify de jure and de facto standards in use for data and metadata models and service protocols. Another aspect of Phase I is analysis of the data integration activities of the projects represented by the components being registered. All of the candidate GEOSS components participating in the initial IP3 use cases were themselves composed of heterogeneous distributed systems. The IP3 studies why and how interoperability arrangements were made and maintained and how the interoperability arrangements worked under the computer and network environment of the time. Some of this information will be entered into the GEOSS Best Practices Wiki.

Phase II—cross-system interoperability use cases. In this phase, use cases are developed that require the exchange of data and information between disparate systems which have not yet established a mechanism for such an exchange. The use cases must address needs identified in one or more of the SBAs and be relevant and realistic. These use cases demonstrate how the GEOSS architecture, data, and services make it possible to realize benefits that would not have been easily achieved otherwise. In addition, the interdisciplinary nature of the use cases ensures that the GEOSS Standards and Interoperability Forum (SIF) [7] is invoked. The SIF was created by the GEO ADC to oversee interoperability issues for GEOSS and is an essential element of the GEOSS interoperability process.

Phase III—demonstration. Where the component systems supporting a use case have the resources available to create implementations of the use cases, these will be prepared, and a demonstration will be presented to the GEO ADC and eventually made available to all GEO members and contributing organizations. A live implementation of data and information exchange via the defined arrangements will be demonstrated in conjunction with those GEOSS infrastructure components that have been implemented.

Phase IV—higher levels of interoperability. In Phase IV, the IP3 seeks to develop higher levels of interoperability. This could involve exercising more completely the capabilities of the protocols and interfaces identified. It also addresses semantic interoperability—whereby the meaning of the data and information exchanged through the interfaces is automatically intelligible to the recipient systems.

B. Description of IP3 Use Cases

The IP3 identified four discipline-specific data systems as sources for the pilot project. These were: Earth’s water cycle, seismology, climate, and biodiversity. This selection was based on the desire to have participation from diverse disciplines, although commitments of representatives from the disciplines to actively support the process were also important. The IP3 then developed five interoperability use cases involving these systems. These were as follows.

1) Species response to climate change: The goal was to apply ecological niche modeling to study the adaptation of different species (e.g., butterflies in Canada and Alaska and pikas in Western North America) to various climate change scenarios.
2) Flood risk: The goal was to support a distributed surface run-off model to provide data for flood warning and dam operation systems. This required interoperability between the systems comprising the coordinated enhanced observation period and components of the World Meteorological Organization Information System (WIS).

3) Landslide risk: The goal was to determine whether rainfall and seismic data can be used to predict locations where hillsides are susceptible to collapse. This required interoperability between the International Federation of Digital Seismographic Networks (FDSN), the WIS, and a landslide database network in Australia.

4) Fault lubrication: The goal was to determine whether high rainfall rates or excess groundwater produces increased earthquake activity in areas of known faults. This required interoperability between FDSN and WIS components.

5) Seismic events from glacier/ice sheet disintegration: The goal was to combine analysis of seismic signals, remote sensing data, and in situ data to determine whether global seismic networks can be used to monitor ice sheets and glacial activity and thus remotely sense effects of warming on a global scale. Required interoperability between FDSN and databases at the National Snow and Ice Data Center.

All five use cases were carried through Phase II of the IP3 process. The Species Response to Climate Change use case had the resources to be developed through Phase III and is the primary focus of the work described here.

III. INITIAL IP3 FRAMEWORK

A. Earth Science System-Level Approach

Earth scientists are engaged in integrating knowledge from different disciplines about the constituent parts of the complex Earth system with the objective of understanding its properties as a whole. Such system analysis is a challenge for scientists as well as for the information technology supporting them. In fact, the scope and complexity of Earth system investigations demand the formation of distributed multidisciplinary collaborative teams. This requires the integration of different discipline information systems, characterized by heterogeneous and distributed data and metadata models, different semantics and expertise, diverse protocols and interfaces, and different data policies and security levels [8]. Advanced infrastructures like GEOSS will support the formation and operation of an Earth system science community, based on multidisciplinary knowledge integration. As outlined by the GEOSS architecture specification, this implies scaling from specific, monolithic, and data-centric information systems toward independent, modular, and service-oriented systems. GEOSS infrastructure must provide scientists, researchers, and decision makers with a persistent set of independent services and information that scientists can integrate into a range of more complex analyses. The extended IP3, discussed in Section IV, will further this cause by providing a framework for such multidisciplinary interactions.

B. Challenge for Information Technology

Many Earth and space sciences communities (e.g., oceanography, geology, ecology, atmospheric science, Earth observation, etc.) have developed tailored data and metadata models and service protocols for enabling online data discovery, inventory, and access. These communities conceived their specifications by doing either of the following: 1) developing application “views” of international standards for geospatial information or 2) enriching existing and well-accepted community conventions to capture and describe the additional semantics required to address interoperability needs. The first approach, called standard profiling, is a suitable strategy for improving data completeness (by adding all relevant information) and accuracy (by removing irrelevant information). Valuable examples are the application metadata profiles based on the international reference standard for geographic metadata modeling: the ISO 19115 standard [9], the user community catalog profiles implementing the abstract specification for discovery services: the Catalog Service for Web (CS-W) [10] by Open Geospatial Consortium (OGC), and the application languages using the grammar defined by the Geography Markup Language [11]. Examples of the second approach include the CF-netCDF/OPeNDAP/THREDDS [12]–[14] data and protocol framework for the fluid Earth sciences, and the Taxonomic Database Working Group (TDWG) standards [15] for the biodiversity and ecology communities.

Where communities have not converged on a unified approach to data representation, description, and discovery, an independent “harmonization” process may be required. This problem is exemplified by the heterogeneity in land cover maps which results from the different methods and standards used in creating such maps. The harmonization mechanism proposed in [16] addresses syntactic, schematic, and semantic aspects of the problem through standards, software, and practices that allow retention of existing data and conventions while providing a mechanism for cross-referencing through registered classifiers.

Geosciences programs and projects have been applying these heterogeneous data discovery, inventory, and access mechanisms based on specific data and metadata models, thereby providing access to a huge amount of valuable information and data sets. Hence, GEOSS has a clear need to conceive and put in place a multidisciplinary framework that is able to fully implement distribution and mediation capabilities.

C. Species Response to Climate Change

The Species Response to Climate Change use case was selected to reflect the threat of climate change, which is expected to cause the extinction of 15%–37% of species by 2050, accelerating the mass extinction already precipitated by widespread land use changes [17]. This use case demonstrates interoperability between the Global Biodiversity Information Facility (GBIF) and components of the WIS. This required “special arrangements” to be registered through the GEOSS SIF, an important objective of the IP3 [18]. A functional demonstration fed species occurrence data and climate model output data into an ecological niche model (ENM) running in
an open modeling framework. The result was a new product showing the impacts of climate change on the geographic distribution of the selected species. ENM is a widely used approach for predicting the distribution of species based on environmental data [19], [20] and is now employed for a range of global change and macroecological applications [21]. GBIF has promoted this scientific approach for using primary biodiversity data for studying adaptation to various climate change scenarios.

Two scenarios were demonstrated. The first dealt with the Canadian common roadside skipper butterfly (*Amblyscirtes vialis*), because they are known to respond rapidly to climate change [22], [23]. The second analyzed the impact of future higher temperatures on the range of the American pika (*Ochotona princeps*) considering the research conducted at the University of Colorado [24], [25] for the western U.S. The first scenario showed that the range of *Amblyscirtes vialis* will shift north under the most likely climate change scenario for 2030–2050, as shown in Fig. 3. The second scenario showed a large loss of habitat for *Ochotona princeps*. Both results must be validated by *in situ* observations. These demonstrations were presented to the GEO ADC and also displayed at the GEO Ministerial Meeting in Cape Town, South Africa, in November 2007 [18].

### D. IP3 Service-Oriented Infrastructure

The development of such an ambitious use case required us to address several important challenges: 1) For biodiversity data, these were online discovery and selection of species occurrences by fixed criteria (e.g., availability, representativeness, expected susceptibility, and importance) and online accessibility; 2) for climatological data, these included online discovery and generation of useful parameters (e.g., daily/monthly average/maximum temperature, rainfall, vegetation, land use cover, etc.), temporal and spatial coverage (e.g., to span at least 30 years and cover the useful areas), useful georectification, spatial resolution, and online accessibility; and 3) for ENM, these included access to online functionalities and heterogeneous modeling algorithm support; and 4) for user client, these included online accessibility, acceptable performances, simplicity, and portability (e.g., Web browser).

### E. Demonstrated Solution

The infrastructure we developed implemented discovery, access, and management of heterogeneous data resources (i.e., biodiversity, climatological, and environmental resources). In addition, simple workflow services were required to interface the ENM. For this scope, IP3 realized a general Web service architecture infrastructure supporting predictions of species response to climate change.

A simplified schema of the infrastructure architecture is shown in Fig. 4, which shows the role played by the different IP3 components according to the SOA paradigm (i.e., provider, consumer, and registry). The main component interactions are depicted using colored dashed arrows.

When implementing an SOA, an organization’s computing environment is represented as a set of distributed software components. A component is a software building block that exposes one or more services to potential requesters. A component’s specification requires the definitions of its interaction capabilities with other components being the basis for the component’s composability: This type of specification is called the component’s contract [26]. In order to support interoperability, component contracts are required to implement well-known artifacts supporting third-party access: either universally adopted standards or special interoperability arrangements.

The GEOSS Implementation Plan introduced few interoperability standards; rather, it addressed how contracts are defined. For this aim, GEOSS promoted the SIF to discuss and recognize the supported standard contracts. Standards are widely available to support the most basic form of contracts. International standard organizations, such as ISO TC211 [27] and OGC [28], have been working on developing standard contracts for the geospatial information domain. Finally, many Earth science disciplinary communities introduced contracts suited for their specific components. Several of these international standards and disciplinary conventions are already becoming commonplace in scientific computing.

IP3 modular components are interconnected through a common shared bus (i.e., the Web platform). This acts as the communication platform that allows the components to communicate. Components are written in a plug and play fashion to
subscribe to data of interest on the bus, and publish them and provide services for use by other components.

The IP3 infrastructure for species response to climate change comprises the following main components (Fig. 4).

1) **Biodiversity Data Provider:** This component publishes biodiversity occurrences which are discovered and accessed through specific Web services. These services are implemented and published by the GBIF data portal [29] utilizing the TDWG Darwin Core as the standard metadata model. The Darwin Core is a standard designed to facilitate the exchange of information about the geographic occurrence of species and the existence of specimens in collections [15]. The GBIF component contracts (i.e., data models and functional interfaces) were registered to GEOSS SIF as special interoperability arrangements.

2) **Climatological Data Provider:** This component publishes climatological data; they are discovered and obtained from the National Center for Atmospheric Research (NCAR) GIS portal [30] which provides Web access to free global data sets of climate change scenarios. These data (spanning 50 years from 2000 to 2050) have been generated for the 4th Assessment Report of the Intergovernmental Panel on Climate Change [31] by the Community Climate System Model. NCAR processed these data sets to generate grid coverages which were inventoried and served through an OGC WCS 1.0 server [32]. IP3 planned to discover and access these data from the TIGGE (the THORPEX Interactive Grand Global Ensemble) database, but that interface was not yet available. The future IP3 infrastructure will tie back into the WIS.

3) **ENM Provider:** The component used for processing collected data and generating future projections is the openModeller [33], [34], an open-source ENM framework. It publishes an online Web interface to access its functionalities, implementing some of the standard basic contracts for SOA. This component is able to run an ENM on the selected biodiversity and climatological data sets.

4) **Infrastructure Distributed Catalogue:** Supporting interactions among disciplinary component services requires third parties or intermediate services to provide service discovery. Moreover, in multidisciplinary environments, mediation is required to harmonize the heterogeneous standard contracts (i.e., data models and protocol interfaces) that characterize the diverse disciplinary components. Discoverability and mediation are not natively supported by Web services; hence, they must be implemented by the architecture. Consequently, a central aspect of this demonstration was the creation of a catalog service that is able to discover and distribute queries to heterogeneous data providers (i.e., the discovery of and access to the available biodiversity, climatological, and environmental data sets). In addition, this catalog component implements mediation capabilities to harmonize the international standards and the “special arrangements” that characterize the heterogeneous communities contributing to the use case.

The IP3 experimentation exposed the need for mediation services in addition to distributed discovery. Mediation services prove to be a valuable and flexible approach to harmonizing component contracts, enabling a holistic view of infrastructure-provided data sets. Mediation services offer a uniform way to access the information stored in the data sources. The uniformity is reached by adopting a common conceptual schema, a uniform metaquery language and a common set of metadata for describing both the providers and their data content [35].

The infrastructure distributed catalog allows users to browse and search registered services based on full-text and field-based queries, including spatial and temporal constraints. This
component was realized by applying the GI-cat technology [36], [37]. GI-cat is an open solution for developing catalog components which implement distributed discovery, data model mediation, and access services. GI-cat provides a consistent interface for querying heterogeneous catalogs and data providers that implement international geospatial standards and special arrangements, making it possible to federate heterogeneous data sources by specifying mediation rules for interoperability. Both GBIF and NCAR—Intergovernmental Panel on Climate Change (IPCC) portals were registered at GI-cat.

5) User Client: This is a Web browser application capable of running on different platforms (e.g., Windows, MacOS, and Linux). The application implements a workflow control service to manage and interact with the ENM and the infrastructure distributed catalogue. In fact, the user is guided through the process of discovering data by submitting queries to the distributed catalog (red arrow in Fig. 4), accessing selected data through the biodiversity and climate prediction providers (blue arrows), and running the ENM projections (green arrow). Finally, the client displays results. The AJAX [38] technology was used to implement this component, ensuring its online availability and multiplatform characteristic. AJAX stands for Asynchronous JavaScript And XML; it consists of a group of interrelated Web development techniques used for creating interactive Web applications. An important characteristic is the increased responsiveness and interactivity of Web pages.

IV. DEVELOPING AN EXTENDED IP3 FRAMEWORK

The IP3 will extend its initial framework by incorporating additional disciplines and data sources, as well as helping to develop a framework for model interoperability. Through this, the IP3 is furthering the stated goals of addressing interdisciplinary interoperability, standards implementations, and the need to enable disparate user communities to work together.

A. Challenge: Extensibility of Distributed Catalog

One of the challenges of the extended IP3 framework is that it requires a sophisticated and extensible distributed catalog service that implements mediation capabilities. This service must be able to federate the community catalogs serving the different communities that will be participating in the IP3 interdisciplinary use cases.

According to the general definition, geospatial data catalogs are discovery and access systems that use metadata as the target for queries on geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among communities [39]. For the IP3 cross-domain use cases, the catalog service must be able to address metadata heterogeneity as well as component distribution. This is possible by applying mediation approaches and protocol adaptation for implementing query distribution and virtual resource aggregation. Such a distributed catalog may be referred to as extensible because it is able to be “easily” extended by federating new resource component types, as long as they are characterized by documented special arrangements for interoperability. In addition to flexibility, important requirements concern the catalog usability (i.e., one-stop-shop server) and performance-enabling caching and/or parallelism. On the other hand, extended distributed catalogs present some issues to be addressed, including cyclic queries, data identity, and resource mediation maintainability.

An abstract model for distributed catalog service is formalized by OGC CS-W [10]. Starting from this specification, the distribution aspects must be extended with functionality to add and remove federated resources. Moreover, distribution cannot be fully addressed without considering mediation capabilities. In fact, catalog service heterogeneity (e.g., the different CS-W application profiles and the catalog specifications of heterogeneous communities) is a matter of fact. An extensible distributed catalog must implement both these functions (i.e., extended distribution and mediation capabilities). The initial IP3 infrastructure demonstrated that it is feasible to operate such a catalog component, realizing a flexible framework to federate well-accepted catalog, inventory and access standard services as well as community standards.

B. Model Resources: Growing a Model Web

Computer models are an important component of the IP3 use cases and will be a critical part of GEOSS. However, models are generally built in isolation and do not easily interoperate with others, resulting in lost opportunities to address important questions. Barriers to interoperability are both technical and cultural. The Model Web [40], a concept being developed by the NASA Ecological Forecasting program, addresses this problem. The Model Web would be an open-ended system of interoperable computer models and databases communicating via SOA. It would consist of a distributed multidisciplinary network of independent interoperating models, plus related data sets and perhaps sensors. Like the World Wide Web, it would grow organically, without central control, but within a framework of broad goals and data exchange standards or guidelines that emerge naturally from the modeling community. Models and data sets would be maintained, operated, and served independently by a voluntary and dynamic network of participants.

A Model Web could not realistically be planned and built; rather, it should be encouraged, facilitated, and gradually converged upon. Because adapting existing models requires significant effort, and lowering barriers to interoperability is a gradual process, it should be viewed as a long-term infrastructure-building process. In both these respects, the Model Web is much like GEOSS itself and perhaps can be considered as the “model” portion of GEOSS.

The Model Web demonstration system that is underway will make use of the IP3 extended framework. Regionalized climate scenarios will be provided to the IP3 by a NASA model called Terrestrial Observation and Prediction System (TOPS). This demonstration system will act as a core onto which additional model components can be added so that growth of the Model Web can begin. Fire and phenology models are being considered as some of the initial expansion areas for the prototype system. As both the Model Web and the number of online data
sources grow, the Model Web will increasingly depend upon standardized discovery and access capabilities such as those being developed for the IP3 infrastructure.

C. Addressing the Challenge

The extended IP3 framework is addressing the aforementioned challenges by developing the following.

1) Effective distributed discovery and access capabilities involving heterogeneous community component contracts (i.e., data models and protocol interfaces) and conventions, such as those relating to catalog and data access services: Utilization of interoperability special arrangements and data encoding reconciliation are included in this activity; data and metadata models are harmonized at the semantic, structural, and syntactic levels among different Earth and space sciences communities. Ontologies, taxonomies, and multilingual issues are included.

2) Model interoperability, which is limited in all disciplines that rely on computer models (e.g., ecology, climatology, hydrology, air quality, and oceanography): IP3 will develop and register model interoperability arrangements along with specific mediation solutions and possibly recommendations for new standards. As discussed earlier, the extended IP3 use cases are implementing the first elements of an interoperable model framework called the Model Web [40].

3) Coordination with the proposed GEO-portal solutions in the multidisciplinary IP3 context: The GEO Web portal (aka GEO-portal) is a key architectural component for GEOSS (see Fig. 2). It must implement the interaction with the discovery and access components providing portrayal functionalities. Presently, there are some candidate solutions to realize the GEO-portal. The interaction between GEO-portal component and model servers (either directly or through a “control” component) is an important issue to be investigated.

The extended IP3 framework helps support the GEO ADC SIF by providing information necessary to refine SIF processes and support science communities in establishing their interoperability arrangements. In addition, lessons learned in bringing GEOSS components online will provide valuable material for the GEOSS Best Practices Wiki.

As in the past, there will be close coordination with the GEOSS Architecture Interoperability Project (AIP) initiative [41]. The IP3 and AIP are complementary, with both activities supporting long-term infrastructure implementation. The IP3 and AIP activities provide valuable input on the implementation of the GEOSS architectural components shown in Fig. 2.

D. Realizing the Extended IP3 Framework

The extended IP3 framework is built upon components developed under the Species Response to Climate Change use case. Particular attention is being devoted to exploring use cases involving information repositories relating to landslides, air quality, and human health. The processes for contributing and linking systems in GEOSS will be refined and extended. This is essential to establish a multidisciplinary framework, underpinning a cyberinfrastructure for an Earth science system. This is, in effect, carrying the IP3 into Phase IV, higher levels of interoperability, as described in Section II-A.

The schema shown in Fig. 5 shows the primary components of the extended framework.

1) GEO-portal (resource consumer): This will be instantiated as one or more of the current GEO-portal candidates depending on the resources available from the portal teams for coordination and interoperability experimentation.

2) Distributed catalog server: This component implements a fully distributed catalog service that is able to implement interoperability among geosciences resources, in the context of international standards as well as community interoperability arrangements. This component consolidates and extends the catalog services that were part of the initial IP3 infrastructure by implementing distribution and mediation capabilities for new Earth science information communities. These communities will be able to register modeling resources as well as data provider components. Additional well-accepted user community catalog profiles (based on the CS-W standard) will be supported, and their interoperability with GEO-portal candidates will be explored. The catalog server is based on the GI-cat solution which already supports two well-adopted CS-W profiles (i.e., the CS-W ebRIM-CIM and ISO profiles [42], [43], [44]). GI-cat itself implements and publishes these two catalog standard profile interfaces to allow third party components to access it. The distributed catalog is able to access and query most of the resources harvested from one or more GEOSS registries.

3) Data set resource providers: In addition to the already contributed GEOSS components, the extended framework will work with diverse Earth and space science discipline resources, helping them to register to the GEOSS structure. These resources will become discoverable and accessible via international standards as well as community standards registered as special arrangements (i.e., well-known component contracts). The additional data set resources will be enabled by extending the catalog distribution and mediation capabilities on the basis of the new component contracts (i.e., data models and service interfaces).

4) Model resource providers: A special type of resource is represented by model servers. These are developed and operated by different researchers from a variety of disciplines. For example, the ENM, developed through the collaboration of GBIF with openModeller, is already part of the IP3 framework. As mentioned earlier, the NASA TOPS model [45], [46] will soon add local and regional environmental data to the ENM. Interoperability of the various components of the ENM and TOPS requires development of interoperability arrangements and workflow solutions. Modeling services are like any other GEOSS registered service, but there is no well-accepted
international specification for interfacing them. It is likely that the IP3 experimentation will provide new requirements for the GEOSS component and service registries and possibly for new standards. The IP3 framework will work to federate other Web-enabled model resources developed by heterogeneous information communities. This is one of the most significant extensions to be implemented by the new IP3 infrastructure.

5) Workflow/control providers: Where model components do not implement standard control or workflow services, the interaction between them and portal components must be implemented by means of specific solutions. One of these components has already been developed for the interaction with the ENM experiment using the open-Modeller Web interface. The extended framework will systematically analyze these kinds of services introducing “standard” components which implement them. This is another of the most significant extensions to be implemented by the new IP3 infrastructure.

Step 2) GEO-portal passes the query to the catalog component which, in turn, distributes the query to components identified as relevant based on information harvested from GEOSS and other registries. Presently, GI-cat uses an internal list of queryable components. In the final implementation, this list of components will be generated using the GEOSS registries. The query will be distributed to varied resource types being mapped into the different data models implemented by them. These include international standards like the OGC specifications for accessing mapping resources (e.g., the Web map service) and imagery resources (e.g., the Web coverage service). Resources implementing community standards are also supported, mapping the query into the data models characterizing the THREDDS/OpenDAP specifications, the GBIF discovery and access specifications, and the Common Data Index (CDI) discovery specification. CDI is a European standard [47] adopted by the operational oceanography community for discovering the availability and geographical spread of marine data across the different data centers and institutes across Europe. Finally, being a distributed catalog, GI-cat is able to distribute the query to other catalog components implementing CS-W standard profiles.

E. Typical Use Case

Based on this extended architectural framework, a typical interaction sequence (Fig. 5) might consist of the following.

Step 1) User accesses a GEO-portal and submits queries for data required by a cross-disciplinary activity defined by a use case.

Step 2) GEO-portal passes the query to the catalog component which, in turn, distributes the query to components identified as relevant based on information harvested from GEOSS and other registries. Presently, GI-cat uses an internal list of queryable components. In the final implementation, this list of components will be generated using the GEOSS registries. The query will be distributed to varied resource types being mapped into the different data models implemented by them. These include international standards like the OGC specifications for accessing mapping resources (e.g., the Web map service) and imagery resources (e.g., the Web coverage service). Resources implementing community standards are also supported, mapping the query into the data models characterizing the THREDDS/OpenDAP specifications, the GBIF discovery and access specifications, and the Common Data Index (CDI) discovery specification. CDI is a European standard [47] adopted by the operational oceanography community for discovering the availability and geographical spread of marine data across the different data centers and institutes across Europe. Finally, being a distributed catalog, GI-cat is able to distribute the query to other catalog components implementing CS-W standard profiles.
Fig. 6. Multidisciplinary interoperability process for the extended framework: The IP3 enterprise view.

Step 3) The discovered and downloaded data sets are transferred to a model server running, for example, an ENM. The process is managed by a workflow component such as the AJAX client used in the demonstrated solution for controlling the openModeller component. One of the most valuable outcomes of IP3 experimentation will be the full specification of such workflow components.

Step 4) The model server generates information products such as forecasts. These are subsequently made available to the user via the infrastructure accessing a registered resource server.

a) In this way, the new product becomes fully searchable and accessible.

b) It is possible that a model outcome will become an input parameter for a different model component (e.g., regional climate change forecasts for ENM servers), forming part of the Model Web described in Section IV-B.

V. INTEROPERABILITY PROCESS

The benefits of using an SOA-based design accrue over time. As a result, when building SOAs, it is important to take an enterprise view of all development and operations (i.e., the process view, that IP3 has the mandate to investigate and demonstrate).

Therefore, in addition to developing the described critical infrastructure components, the IP3 must refine and test the interoperability implementation process shown in Fig. 6.

The main steps for this process are as follows.

Step 1) Resource provider registers component which instantiates services, and either specifies implemented international standards or registers special interoperability arrangements.

Step 2) Every time a new well-known resource component is registered, the infrastructure catalog service (i.e., the IP3 catalog component) discovers it and automatically associates specific mediation and access capabilities with the new registered component. A component is a well-known resource when it implements a service type that is already registered in the GEOSS service registry. That is, it implements a contract that is fully specified by an international standard or by a registered special arrangement. The IP3 demonstrated this capability for several well-known component types (e.g., OGC data access components WMS and WCS; OGC catalog components CSW-ebRIM.CIM, CSW-ISO, and CSW-ebRIM.FGDC).

To register a new component that does not implement a well-known service type, users must register a special arrangement. In this case, coding must be done for implementing the opportune mediation functionalities and interface the special arrangements by the infrastructure catalog component. In the framework of the initial IP3, such coding was done to support GBIF services. We plan to do this for all the other interoperability arrangements that will be registered by the new multidisciplinary resource types used for the extended IP3 use cases.

Step 3) The infrastructure catalog registry is updated, adding the new registered resource and the new mediation functionalities where needed (i.e., the user registered component and, optionally, its implemented service type).

VI. CONCLUSION

The significance of the IP3 extends beyond simply creating a demonstration system. It is laying a foundation that future work in many disciplines or SBAs can utilize. In addition, this paper is evidence that the basic GEOSS architecture and approach to interoperability is viable and appropriate.

The IP3 exemplifies a holistic approach to Earth system science architecture. While of particular significance to the
biodiversity and ecosystem SBAs, it is pertinent to the agriculture, health, and water SBAs as well.

IP3 also demonstrates the great opportunity to consider modeling components as valuable GEOSS resources in addition to data and services. Finally, the IP3 experimentation clearly outlined the need for infrastructure mediation services, distributed discovery, and support for the heterogeneous contracts that characterize the multidisciplinary implementations of SOA for Earth and space sciences.

ACKNOWLEDGMENT

The authors would like to thank members of the GEO Architecture and Data Committee for their guidance and support. The authors would also like to thank E. O’Tuama and H. Saarenmaa of the Global Biodiversity Information Facility, and P. Mazzetti and L. Bigagli of the Italian National Research Council for their contributions to this paper.

REFERENCES


[16] OGC, OpenGIS Web Coverage Service (WCS) Implementation Specification, Ver. 1.0 (Corrigendum) (1.0.0), OGC 05-076.


Siri Jodha Singh Khalsa (M’77–SM’04) received the B.A. degree in physics from the University of California, Irvine, and the Ph.D. degree in atmospheric sciences from the University of Washington, Seattle.

His early work on boundary layer turbulence and tropical air–sea interaction evolved into an interest in large-scale interdecadal trends in atmospheric structure revealed in satellite sounding records. Since 1993, he has been supporting NASA’s Distributed Active Archive Center at the National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, where he performs science evaluation and algorithm support for cryospheric data products from NASA’s Earth observing satellites. He is also engaged in research on methods of remotely sensing glacier change as part of the GLIMS project.

Dr. Khalsa is the IEEE/GRS-S Liaison to ISO/TC211. He is the Chair of the IEEE Committee on Earth Observations Standards Working Group and leads several IEEE-initiated activities supporting standards and interoperability for the Global Earth Observing System of Systems (GEOSS). He leads the intergovernmental Group on Earth Observation Standards and Interoperability Forum, created by the GEO Architecture and Data Committee to oversee interoperability issues for all of GEOSS.

Stefano Nativi (M’95) received the first and second (Laurea) degree in electronic engineering (telecommunications field) and the Ph.D. degree in “methods and technologies for environmental monitoring” from the University of Florence, Florence, Italy.

He was appointed a PDRA Grant from the University of Bristol, Bristol, U.K. He is currently the Coordinator of the Earth and Space Science Informatics Laboratory, Institute of Methodologies for Environmental Analysis, Italian National Research Council, Florence, where he is also a Coordinator of the National Inter-university Consortium for Telecommunications Operational Unit. He is a Professor, on a contract basis, with the University of Padua, Padua Italy, teaching “systems for land management” for the informatics specialist degree (Faculty of Mathematics). He coleads the GEOSS IP3 initiative and is a member of the GEOSS SIF. He is a member of the “Metadata Core Drafting Team” for the Implementing Rules of the INSPIRE (INfrastructure for SPatial InfoRmation in Europe) European directive.

Dr. Nativi is the Chair of the Earth and Space Science Informatics group of the European Geosciences Union. He is a member of the IEEE Standards Working Group of ICEO and of IEEE GRSS Data Archiving and Distribution Technical Committee. He is co-PI of the OGC GALEON Interoperability Experiment (Geo-interface to Atmosphere, Land, Earth, Ocean netCDF) and a member of the Web Coverage Service standard working group.

Gary N. Geller (M’98) received the M.S. degree in botany from the University of Wyoming, Laramie, in 1980, and the Ph.D. degree in biology from the University of California, Los Angeles, in 1986.

In 1986, he joined the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, working on several projects, where he became the System Engineer for the U.S. ASTER Science Project in 1992. In 2001, he started a 16-month leave-of-absence, during which time he began, largely by accident, working with biodiversity conservation managers in Southeast Asia and elsewhere to understand why they did not use more remote sensing data, and the TerraLook project gradually emerged. TerraLook (http://terralook.cr.usgs.gov) makes recent and historical satellite images available as simple georeferenced jpeg images at no cost to the user and is packaged with simple and intuitive software to view and work with the images, particularly for change analysis. TerraLook was a finalist for the 2007 St. Andrews Prize for the Environment. In 2003, he returned to JPL as the ASTER Conservation Liaison, where he is currently the Deputy Manager of the NASA Applied Sciences Ecological Forecasting Program. His current focus is primarily on two areas. The first is the GEO Biodiversity Observation Network, for which he is a member of the Initial Steering Committee. The second is the Model Web, a long-term concept for a dynamic service-oriented modeling infrastructure, analogous to the World Wide Web, to serve researchers, managers, policy makers, and the general public.