This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Integrative Research: The EuroGEOSS Experience

Lorenzino Vaccari, Massimo Craglia, Cristiano Fugazza, Stefano Nativi, *Senior Member, IEEE*, and Mattia Santoro

***Abstract*—The implementation of the INSPIRE Directive in Europe and similar efforts around the globe to develop spatial data infrastructures and global systems of systems have been largely focusing on the adoption of agreed technologies, standards, and specifications to address the interoperability challenge. However, addressing the key scientific challenges of humanity in the 21st century requires a more comprehensive integrative research effort, which in turn may pose more complex requirements on the systems to be integrated, and increase the number of arrangements required to support them. This paper analyses the main challenges related to integrative interoperability, such as mutual understanding of requirements and methods, theoretical underpinning, and tacit knowledge. To illustrate our contribution to the integrative research, the paper proposes the flexible approach to interoperability, based on mediation and brokering, that has been implemented by the EuroGEOSS research project. It also demonstrates that this approach allows scientific and non-scientific stakeholders to overcome the increased complexity of the integration effort mentioned above and charts the trajectory for the evolution of current spatial data infrastructures.**

***Index Terms*—Geographic information systems, geoscience and remote sensing, integrative research, research initiatives.**

I. Introduction

ONE OF THE MOST fundamental challenges facing humanity at the beginning of the 21st century is to respond effectively to the global changes that are increasing pressure on the environment and on human society. This priority has been described by the International Council for Science (ICSU) as follows:

> "*Over the next decade the global scientific community must take on the challenge of delivering to society the knowledge and information necessary to assess the risks humanity is facing from global change and to understand how society can effectively mitigate dangerous changes and cope with the change that we cannot manage. We refer to this field as 'global sustainability research'*." [1].

ICSU identified five scientific priorities, or Grand Challenges, in global sustainability research through a broad consultation that was carried out in 2009–2010 and involved over 1000 scientists from 85 countries. These Grand Challenges include:

1) Developing, enhancing and integrating the observation systems that are required to manage global and regional environmental change.
2) Improving the usefulness of forecasts of future environmental conditions and their consequences for people.
3) Recognizing key thresholds or non-linear changes to improve our ability to anticipate, recognize, avoid, and adapt to abrupt global environmental change.
4) Determining which institutional, economic, and behavioral responses can enable effective steps toward global sustainability.
5) Encouraging innovation (coupled with sound mechanisms for evaluation) in developing technological, policy, and social responses to achieve global sustainability.

The increasing importance of linking the scientific effort necessary to underpin the sustainability agenda with innovation and sustainable economic growth is also at the heart of the European Union's Europe 2020 strategy [2], focusing on smart, sustainable, and inclusive growth.

The Global Earth Observation System of Systems (GEOSS) [3], envisioned by the group of the eight most industrialized countries in 2003 and currently halfway along its 10-year implementation plan, provides the indispensable framework to integrate the earth observation efforts of the 84 GEO-members and 58 participating organizations. A major role of GEOSS is to promote scientific connections and interactions between the observation systems that constitute the system of systems, and to address some of the scientific challenges identified by ICSU with a particular focus on nine societal benefit areas (Disasters, Health, Energy, Climate, Agriculture, Ecosystems, Biodiversity, Water, and Weather). Such interactions also stimulate the introduction of innovative observing systems techniques and technologies. In this respect, therefore, the development of GEOSS can make a strategic contribution in delivering the objectives of the Europe 2020 strategy and meet the ICSU Grand Challenges.

For these reasons, the European Commission plays a very active role in developing GEOSS. This includes participating in and co-chairing GEOSS Committees and the Data Sharing Task Force, but also implementing important initiatives to collect and share environmental information for the benefit of the global society. Examples of these are the Infrastructure for Spatial Information in Europe (INSPIRE) [4], the Global Monitoring for Environment and Security (GMES) initiative [5], and the Shared Environmental Information System (SEIS) [6]. The European Commission also contributes to the implementation of the GEOSS Work Programme through research projects like EuroGEOSS [7], which are funded from its Framework Programme for Research & Development.

Manuscript received October 31, 2011; revised January 17, 2012; accepted February 08, 2012. This work was supported by the EU-FP7 projects EuroGEOSS, GENESIS, and UncertWeb.

L. Vaccari, M. Craglia, and C. Fugazza are with the Institute for Environment and Sustainability European Commission (EC)—Joint Research Centre (JRC) Ispra, Italy (e-mail: lorenzino.vaccari@jrc.ec.europa.eu; massimo.craglia@jrc.ec.europa.eu; cristiano.fugazza@jrc.ec.europa.eu).

S. Nativi and M. Santoro are with the Consiglio Nazionale delle Ricerche, Italy (e-mail: stefano.nativi@cnr.it; santoro@imaa.cnr.it).

Digital Object Identifier 10.1109/JSTARS.2012.2190382

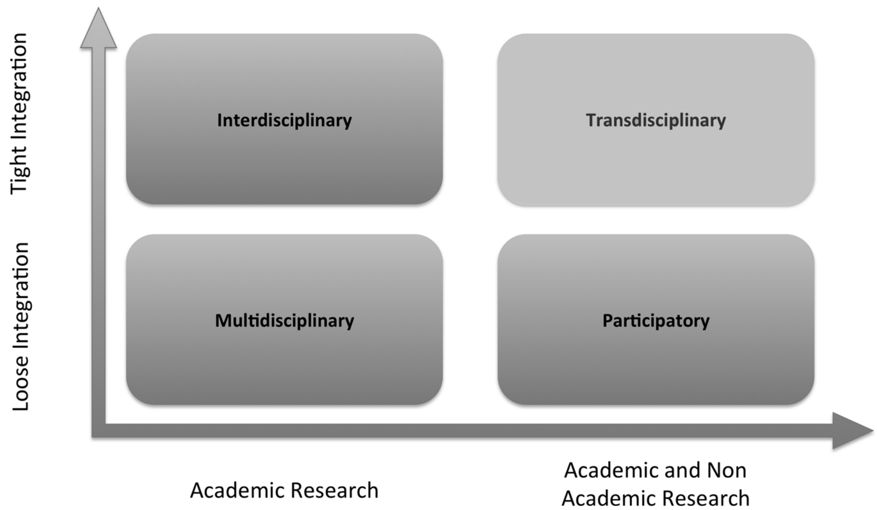

Fig. 1. Integrative research categorization: the four categories are on the one hand, *multidisciplinary* and *participatory* research (two examples of loose integrated research, the latter involving non-academic participants) and, on the other hand, *interdisciplinary* and *transdisciplinary* research. In this case, both the examples involve tightly integrated research, the latter including non academic participants.

This paper describes the contribution of the EuroGEOSS project to the integrative research effort that is necessary to tackle the challenges identified by ICSU. The project, which includes 23 partners and is coordinated by the French Geological Service (BRGM), and the European Commission Joint Research Centre (JRC), runs for the three years 2009–2012. It focuses on the three thematic areas of forestry, biodiversity, and drought, which were identified as a proof-of-concept to demonstrate how an integrative approach can support global sustainability research. In the paper, we categorize the distinct phases in the EuroGEOSS' research effort according to the terminology introduced in [8] to distinguish between the different kinds of integrative research.

This paper is organized as follows. Section II introduces the different integrative research concepts adopted in this paper. Section III reviews the main results for the three thematic areas participating in the project. Section IV focuses on the technologies that have been implemented by the project to support multidisciplinary search, discovery, and access to information resources. Section V discusses the interdisciplinary approach adopted to formalize the description of the analytical models used by domain experts and to turn these models into workflows, business processes, and ultimately chains of geoprocessing services running on the Web. Section VI moves beyond the collaboration among scientist, to the need to communicate science more effectively, and increase public participation. It therefore discusses a framework to increase two-way collaboration between data infrastructures and social networks. Section VII draws the conclusions.

II. Integrative Research

Although integrative research concepts are widely used in the literature, there is often a lack of common understanding on the meaning of these concepts. Before describing the architecture that has been followed for implementing the EuroGEOSS infrastructure and explaining how it addresses the integrative research challenge, we briefly resume the categorization of integrative approaches introduced in [8].

Essentially, the distinction between parallel (i.e., independent) and integrative research lies in whether the entities involved in the research effort are loosely or tightly integrated. Loose integration means that participants from distinct disciplines are researching on the same topic with multiple disciplinary goals. Typically, no shared knowledge is created because improvements over the state of the art in the different disciplines are independent. On the other hand, tightly integrated research efforts involve participants from unrelated disciplines that strive to create new, shared knowledge to reach a common research goal.

Orthogonal to this distinction, one can also consider whether the research is involving academic (or research-oriented) participants or not. Across the axis defined by these characteristics, we can identify the four categories of integrative research that are shown in Fig. 1: one the one hand, *multidisciplinary* and *participatory* research (two examples of parallel research, the latter involving non-academic participants) and, on the other hand, *interdisciplinary* and *transdisciplinary* (in this case, both the examples involve tightly integrated research, the latter involving non academic participants) research.

The EuroGEOSS project represents a research experience that covers both disciplinary and three of the four integrative research categories (Fig. 2). The following sections describe how the different components developed by EuroGEOSS enable both disciplinary and integrative research.

III. Disciplinary Research: Progress and Main Results to Date

Disciplinary research is the starting point to achieve integrative research. In the first part of the EuroGEOSS project, the key objectives were:

1) Achieving an initial operating capability for each thematic area, including the development of the services that are necessary to make it possible to discover, view, and access

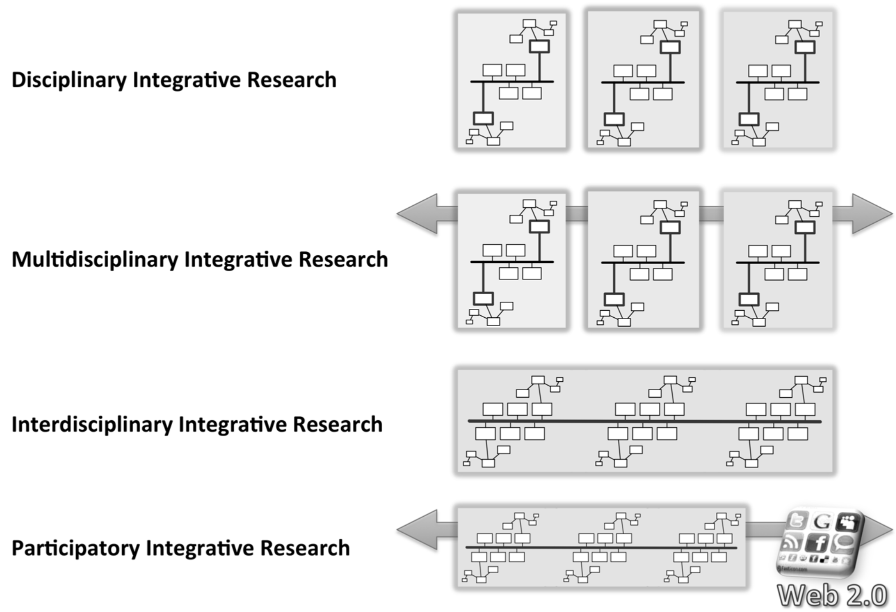

Fig. 2. The EuroGEOSS project represents a research experience that covers both disciplinary and three (Multidisciplinary, Interdisciplinary and Participatory) of the four integrative research categories.

the information resources made available by the project partners in the biodiversity, drought, and forestry thematic areas.

2) Registering these resources as GEOSS components, thus making them more widely available to the global community.
3) Developing the framework for assessing the added value of the project and of the GEOSS infrastructure to the user communities.

All of these objectives were achieved during the first 18 months. The operating capability that has been achieved in each individual thematic area constitutes the necessary component that is required to bridge between the distinct disciplinary research activities and provides the basis for the multidisciplinary approach envisaged by the EuroGEOSS project.

The **Forestry operating capability** has been achieved giving priority to the development of federated metadata (metadata is a description of an information resource, including key elements such as what it is, who is responsible for it, where can it be found, and how it can be accessed) catalogs and a map viewer, which were then integrated into the EuroGEOSS framework.

These priorities were identified through an analysis of the forestry users' requirements [9]. The forestry Metadata Catalog has been developed based on the GeoNetwork open source package (v 2.4.3) and has been populated with metadata from the European Forest Data Centre hosted by the JRC [10] of the European Commission (EC). Adjustments have been made to fit the Dublin Core [11] metadata elements and ensure compliance with INSPIRE and the relevant ISO 19115 [12], ISO 19119 [13], and ISO 19139 [14] standards. The Metadata Catalog functionalities and interface have also been tailored to meet the specific requirements of the forestry thematic domain. As a result, the Catalog provides search, discovery, and preview facilities of both spatial and non-spatial metadata. The catalog harvests metadata from national and local forestry catalogs such as those of the national Spanish spatial data infrastructure (IDEE), and is federated in the EuroGEOSS framework so that its resources are made globally accessible and viewable by the GEOSS community.

The **biodiversity operating capability** is based on an analysis of user requirements [15] and has been achieved by developing a series of metadata catalogs and services at the partners' institutions and by integrating them into the EuroGEOSS framework. A key milestone has been the development of the metadata catalog for the Global Biodiversity Information Facility (GBIF) [16] with a specialized profile using the Ecological Metadata Language. The main challenge was to support community needs better, especially for the datasets of species names and the natural history collections, and to support multiple natural languages. A metadata sharing service has been established, based on the Open Archive Initiative (OAI) [17], harvesting metadata from the participating GBIF catalogs and integrating them into the EuroGEOSS framework.

In parallel to this activity and the related developments at other partners' institutions, significant work has been done to support the creation of a Digital Observatory for Protected Areas (DOPA) [18]: a facility with initial focus on Africa but with a global reach as a component of the GEOBON observation network [19]. DOPA will be developed in an iterative way, starting with an information system capable of visualizing and interacting with the key datasets hosted by the partners through a single graphical user interface. These datasets describe the boundaries of protected areas (United Nations Environment Programme—World Conservation Monitoring Centre, UNEP-WCMC) [20], the occurrences of species (GBIF) [16], and the maps of bird distributions (Birdlife International and Royal Society for the Protection of Birds, RSPB [21]). With the support of the EuroGEOSS project, these initiatives are becoming more and more web-based, allowing the multidisciplinary integration of information made available

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

in the other thematic areas. The initial phase of the project has focused on the setting up of a prototype of DOPA that includes a specialized database, an advanced web client, and the preparation of unique datasets regarding bird distributions that become available either in the form of species occurrences via GBIF and in the form of species distribution maps directly through the DOPA.

The **Drought operating capability** has been achieved by developing the European Drought Observatory (EDO) as a distributed system to discover, view, and access drought data provided at the European level (JRC), regional level (Observatory for South East Europe), and national/regional levels (Spanish Drought Observatory, and observatory for the Ebro river basin). The goal of connecting drought data providers at the three levels (continental, national/international, regional/local) was one of the key priorities expressed by users [22] and its achievement is an important proof of concept of a nested multi-scale system of systems. All partners have in place an infrastructure for providing web map services (Open Geospatial Consortium-Web Map Server, OGC WMS) [23] and update their services regularly. Some partners (JRC and the University of Lubjana) also provide web map services of time series for accessing data sets related to a given date or period.

The integration of services from different drought partners in a common viewer, i.e., the map viewer of EDO [24], allows the linkage to services from the other thematic areas (e.g., forest) and opens new options for drought data analysis. In addition to the European perspective, an interoperable EDO contributes to a future Global Drought Early Warning System under consideration by the World Meteorological Organization (WMO) [25] and GEO/GEOSS. To this end, a prototype Global Drought Monitor has been established as a first building block of the Global Drought Early Warning System, in partnership with the North American GEO/GEOSS community, the U.S. National Integrated Drought Information System (NIDIS) [26], and the Princeton African Drought Monitor prototype. A first demonstration pilot of this Global Drought Monitoring System has been developed and has been demonstrated at the GEO Beijing Summit in November 2010 [27].

IV. Multidisciplinary Research: The Brokering Approach

In the second phase of EuroGEOSS, thematic partners are undertaking the research necessary to develop further their infrastructures into a multidisciplinary advanced operating capability. This capability needs to provide access not just to data but also to analytical models and workflows. It moves therefore beyond the quest to find data from heterogeneous distributed sources, which is often the starting an endpoint of data infrastructures, to the more important question of how to use the data effectively to address complex scientific and policy problems. To do so, the major effort needed is to elicit the knowledge of domain experts and document the theoretical underpinning and analytical processes used, so that they are more clearly understandable and usable by scientists from different disciplinary domains. This effort requires, for each thematic area, to express analytical models as workflows of geo-processing components reusable by other communities, and to advance research on the transferability of models and scientific approaches across disciplines, as well as the use of natural language to interface with the models.

A first step towards integrative research is constituted by the EuroGEOSS brokering approach, a multidisciplinary infrastructure bridging the project's three thematic areas. The approach was built upon the comparative analysis of the thematic user requirements [28], and applies several of the principles/requirements that characterize the System of Systems (SoS) approach and the Internet of Services (IoS) philosophy:

1) To keep the existing capacities as autonomous as possible by interconnecting and mediating between standard-based and non-standard-based capacities.
2) To supplement, without supplanting, the individual systems' mandates and governance arrangements.
3) To assure a low entry barrier for both the resource providers and the end users.
4) To be flexible enough so as to accommodate the existing systems as well as future ones.
5) To build in an incremental fashion upon the existing infrastructures (information systems) and incorporate heterogeneous resources by introducing distribution and mediation functionalities.
6) To specify interoperability arrangements focusing on the modularity of interdisciplinary concepts rather than just on the technical interoperability of systems.

The key features of the EuroGEOSS multidisciplinary approach are the brokering and mediation capabilities that allow for discovering and accessing autonomous and heterogeneous resources from the three thematic domains of the project. The brokering approach extends the traditional Service Oriented Architecture (SOA) archetype by introducing an "expert" component: the EuroGEOSS Discovery Broker (EDB) [29]. It provides the necessary mediation and distribution functionalities to *(i)* allow service consumers to bind to heterogeneous service providers in a transparent way and *(ii)* interact with them using a single and well-known endpoint. Such a solution addresses some of the shortcomings characterizing state-of-the-art SOA implementations, such as the lack of semantic interoperability and proliferation of standards, which jeopardize the development of complex, large, and heterogeneous infrastructures like GEOSS. Demonstrating the added value of this approach is therefore one of the main contributions by EuroGEOSS to the development of GEOSS and the IoS.

The EDB is central to multidisciplinarity because it allows the EuroGEOSS framework to read and mediate between the diverse standards and specifications that are used by the distinct scientific communities. By building bridges between the practices of these communities, the EDB makes it possible to find all the datasets and services of the partners in the three thematic areas, including multiple catalog services. By registering the EDB as a GEOSS component [30], all the project's thematic resources are also made accessible to the global research community.

The EDB provides harmonized discovery functionalities by mediating and distributing user queries against the multitude of services currently registered in the EuroGEOSS capability. In turn, many of these are catalogs or inventory services that propagate further the queries to many other resources. A key feature of the EDB is that it makes it possible to integrate widely acknowledged SOA discovery interfaces with emerging Web 2.0 resources and to easily utilize them. This list of supported formats comprises:

- Service interfaces that comply with INSPIRE and/or OGC standards.
- Service interfaces which are specific to the three thematic areas.
- Service interfaces which are widely acknowledged by other user communities, such as the Thematic Realtime Environmental Distributed Data Services (THREDDS) [31] and the Open-source Project for a Network Data Access Protocol (OPeNDAP) [32].
- Service interfaces defined by specific projects, such as the Ground European Network for Earth Science Interoperations-Digital Repositories (GENESI-DR) [33] and SeaDataNet [34].

Bridging between these different communities makes it possible to meet the multidisciplinary needs of scientific research without assuming that everyone will converge on one selected standard. Recently, the project set up a web portal providing access to the EuroGEOSS discovery broker [35].

To enhance further multidisciplinary interoperability, EuroGEOSS prototyped, in collaboration with the GENESIS FP7 project [36], the semantics-aware Discovery Augmentation Component (DAC) [37] which extends the EDB capability. The DAC implements a "third-party discovery augmentation approach" enhancing the discovery capabilities of the infrastructures that are brokered by overlaying advanced multilingual and query expansion functionalities. The DAC achieves this by issuing to existing discovery facilities (e.g., catalogs and discovery brokers) queries that have been expanded through semantic services (e.g., controlled vocabularies, ontologies, and gazetteers) by including related terms as well as translations into multiple languages. The query capabilities implemented by the DAC contribute bridging the gap constituted by semantic heterogeneity and multilingualism, which is essential to building interdisciplinary SOA infrastructures.

The EuroGEOSS DAC represents a unified entry point to the diverse resources aggregated by the EDB. Essentially, it provides functionalities for discovering data according to common geospatial constraints (i.e., what, where, when, etc.), for downloading and viewing data. It also features query expansion mechanisms that are based upon a federated set of multilingual controlled vocabularies that allow for enlarging the set of terms that are searched for in metadata descriptions.

Currently, two different augmented discovery styles are supported: *(i)* automatic query expansion and *(ii)* user-assisted query expansion. With the former, the user just selects which "axis" shall be followed for expanding the query (e.g., more general terms, more specific ones, etc.). With the latter, the user can actually browse the graph induced by the terms in the thesauri (together with the relations they define) and select the terms that s/he deems pertinent to the search. In both cases, the set of terms that are identified are further expanded with multiple translations of the terms in other languages.

The geospatial thesauri that are accessed by the DAC are provided by the GENESIS Vocabulary Service (GVS) [38] as RDF data complying with the Simple Knowledge Organization System (SKOS) format [39] and made available as a SPARQL endpoint [40]. SKOS is a widely acknowledged, ontology-based format for expressing controlled vocabularies, taxonomies, subject headings, and knowledge organization systems in general. It represents a good tradeoff between expressiveness and computational complexity that allow for leveraging the novel techniques being developed in the context of the Semantic Web and still implement efficient solutions.

Currently, the GVS is hosting the following thesauri:

- The GEMET [41]
- The INSPIRE Feature Concept Dictionary and Glossary [42]
- The GEOSS Societal Benefit Areas [43]
- The ISO 19119 geographic services taxonomy [44]
- The EuroGEOSS Drought Vocabulary [45]
- The GEOSS AIP-3 Water Ontology [46]
- The GEOSS Earth Observation parameters [47]
- The Global Change Master Directory (GCMD) scientific keywords [48]

More importantly, the repository is also hosting the relations linking many among these thesauri. It is because of these relations that the DAC can efficiently bridge between the different thematic and application domains.

The selection and harmonization of the thesauri accessed by the DAC play an important role in enabling multidisciplinary access to resources. As an example, the GVS is featuring, among the other thesauri, the thematic vocabulary on drought developed by EuroGEOSS: Thus, terms from this vocabulary can to be found in the metadata of resources related to this thematic area. However, users from other thematic areas, as well as the layman, may find it difficult to discover these resources because of terminology mismatch, that is, the terms they employ for a search may be different from those contained in the metadata. To address this common problem, and enable access by a wider audience, terms from the drought vocabulary have been related to some widely acknowledged terminology, such as GEMET and the INSPIRE Themes. As a result, query expansion takes care of crossing the thematic boundary and allows users to discover resources by using terms with which they are familiar that are more general than those provided by the domain specialist.

More interestingly, the query-expansion paradigm may enable multidisciplinary access to resources by coupling terminologies from different application domains. As an example, GEOSS resources may be annotated according to the Earth Observation (EO) Vocabulary, a selection of 142 "critical observation parameters" that are categorized in a three-level hierarchy according to 80 Global Change Master Directory topics and terms. On the other hand, discovery of GEOSS resources is likely to respond to a policy-making need in one of the Societal Benefit Areas (SBAs) defined by GEOSS. Therefore, terms from the EO Vocabulary have been related to the corresponding SBAs so that they can be retrieved by non-scientific users, such as decision makers.

The relations that were created among distinct thesauri represent also an example of knowledge that is generated by interdisciplinary and transdisciplinary research. In this case, the emergence of a common goal (supporting decision making in the SBAs) and the need for access to data across thematic and application domains.

After addressing the components needed to support multi-disciplinary search and discovery of information resources, the next step is to facilitate multi-disciplinary data access. For this purpose, a new component has been developed as part of the EuroGEOSS project by applying the same principles and technologies: The EuroGEOSS Access Broker (EAB). The EAB makes it possible for users to access and use the datasets resulting from their queries according to a common grid environment they have configured by selecting the following common features: Coordinate Reference System (CRS), spatial resolution, spatial extent (e.g., subset), and data encoding format. This feature is crucial to allow effective integration and analysis of data coming form heterogeneous sources. In normal practice, the manipulation to the data necessary ahead of the analysis has to be done by the user. The EAB takes this burden away from the user, thus providing a true added value service.

In keeping with the SoS principles, the EAB carries out this task by supplementing, but not supplanting, the access services providing the datasets requested. This is achieved by brokering the necessary transformation requests (those that the access services are not able to accomplish) to external processing services. Following the IoS and Web 2.0 principles, the EAB publishes web applications allowing users to: *(i)* select a default business logic (i.e., algorithms) implementing dedicated processing like CRS transformation and space resolution resampling; *(ii)* upload their own business logic (i.e., processing schemes) and set it as default; *(iii)* select the order of the processing steps. The EAB also publishes an interface which realizes the INSPIRE transformation service abstract specification [49].

V. Interdisciplinarity Within the EuroGEOSS Project: Enabling Service Composition

During this last phase, the EuroGEOSS project has built further components to access and use not just data but also models and analytical processes across multiple thematic areas by expressing these models in workflows and implementing them through web-based chains of services. The main impact of this development is to make EuroGEOSS resources accessible and usable not only by specialists in the individual fields, but also by scientists from multiple disciplines who will be able to better understand how to use the resources that are made available and how to adapt them to their specific needs. This implies the development of an integrated framework providing common service composition and data models. Such capabilities enable interdisciplinary research and have been implemented using the following technologies:

- The use of Business Process Modeling Notation (BPMN) 2.0 [50] to document the scientific models as a business processes.
- The use of geospatial standard web services to implement process components.
- The use of a brokering approach to implement advanced process components allowing, thus, their flexible composition.
- The use of the Composition as a Service (CaaS) approach for service chaining and process execution.

After analyzing the methodologies available to describe business processes, the partners in EuroGEOSS identified BPMN as the most suitable to describe in a formal notation the workflow described by the domain experts to analyse and solve problems specific to their area. After some experimenting, the BPMN notation is now used inside the project not only to design and manage business processes, but also to exchange business processes between scientists that belong to different disciplines. The next steps of the project will include the following activities:

- Collecting and the analyzing of the workflow models and geo-processes proposed by each EuroGEOSS thematic area (Forest, Biodiversity, and Drought).
- Investigating how geo-processing and service chaining technologies can be improved with regard to performance and reliability.
- Assessing applicability of the CaaS solution proposed by the UncertWeb [51] project.

The CaaS solution prototyped in the UncertWeb Project aims to identify an appropriate mechanism to execute an abstract business process (BP) (described by a BPMN document). To achieve this, the UncertWeb CaaS defines some conventions [52] to bind each component of the BP to proper web services that implement international standard or de-facto standard interfaces providing the needed functionality. To manage the different existing geospatial standard web service interfaces, the UncertWeb CaaS applied the brokering approach experimented in EuroGEOSS.

The CaaS has been tested and evaluated with concrete application scenarios, including the eHabitat Biodiversity scenario developed in the EuroGEOSS framework. This scenario implements a Monte Carlo simulation performed on the deterministic eHabitat ecological model. Starting from input datasets with related uncertainty description (encoded in netCDF-U), the user composes a set of available services to generate a Monte Carlo simulation on the eHabitat model in order to derive a statistics of the output datasets (again encoded in netCDF- U).

VI. Participatory Approach: Volunteer Geographic Information and Web 2.0 Crowdsourcing

The activities described in the previous sections address the important issues of facilitating the multidisciplinary effort needed to advance global sustainability research. Equally important however is to communicate science effectively to both public and decision makers, and engaging them in the understanding of the scientific process and the uncertainty inherent in predicting and understanding change. These issues were highlighted by the controversy surrounding the apparent certainty of projections made by some scientists associated with the 2007 report of Intergovernmental Panel on Climate Change (IPCC) [53]. The spiraling development of social networks with over 10 bn. accounts in 2010 (50% more than people on Earth, generating for example 65 million Tweets a day, and a store of over 4 billion photographs on Flickr), creates new opportunities to develop a two-way dialogues between science and the public. On the one hand, the information provided by the public, often geographically located, can be harnessed for a variety of applications, from emergency management and response, to risk assessment, quality of life and environmental monitoring [54], [55]. On the other, spatial and environmental infrastructures need to open up to the wider public and make the data, analyses, and scientific evidence more widely accessible and understood. Some progress in opening up government and scientific data can be noted recently with initiative like Open Data (see for example data.gov in the US, data.gov.uk in Europe, Open Science Data, the Science Commons, and the Open Data challenge [56]). More however, needs to be done to develop a constructive relationship between governmental and scientific initiatives and social networks.

EuroGEOSS has addressed this challenge by developing a Web 2.0 broker [57] that interfaces the protocols normally used by Spatial Data Infrastructures (SDI) (typically, ISO or OGC) with those used by a range of Web 2.0 services such as Twitter, Panoramio, Picasa, Flickr, OpenStreetMap, Google Search API, Wikimapia, Geonames, and Geocommons. After an analysis of the different protocols used by each of these services, it was decided to use OpenSearch [58] and its geo extension [59] as the minimum common denominator able to create a bridge between the Web 2.0 and the SDI services. Individual OpenSearch-geo interface adaptors have therefore been developed in the project for each of the Web 2.0 services identified above, so that a standard user query can be distributed by the Discovery Broker to a wide range of domain—specific catalogues (see Section III) but also via the Web 2.0 broker to a selection of social network services. The broker has been tested in a parallel research project that seeks to develop methods to extract information from social networks relevant to an emergency situation (e.g., forest fires), assess its quality, and added-value compared to information coming for the official channels. The project's early results [60] show that there are many obstacles in the process of data mining, geo-referencing, and validating (e.g., only 1% of Tweets has a geographical tag), but that the opportunities for a more integrated and participatory approach are many. In particular, integrating citizens' observations and information into global sustainability research offers two main advantages: on the one hand, it allows for the integration and analysis of qualitative information about how citizens perceive their environment and the changes that affect them, which is crucial from both policy and science perspectives; on the other, the combined flow of information from citizens and sensor networks [61], [62] move us towards a more dynamic and interactive framework for participatory science which many see as the essence of a new vision for Digital Earth [63], [64], [65].

VII. Discussion and Next Steps

There are clear challenges on using and integrating multidisciplinary resources to develop applications across different disciplines. These include:

- High Entry Barrier: users need to "learn" and develop many (often immature) information technologies.
- Limited functionalities: the international community has mainly focused on discovery functionality implementation; while, cross-domain evaluation, transfer, and use functionalities are still lacking.

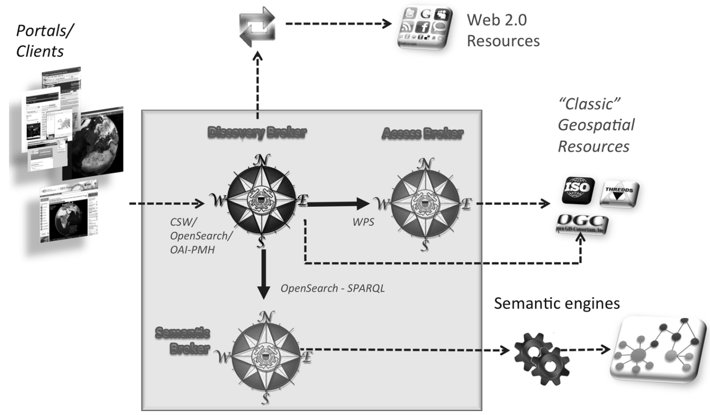

Fig. 3. The EuroGEOSS brokering framework includes different multidisciplinary EuroGEOSS Brokering components: The Discovery broker, the Access broker, the Semantic broker (Discovery Augmentation Component) and the adapters to Web 2.0 resources.

- Limited semantic interoperability: interoperability for heterogeneous disciplinary resources and different domain semantics are still main issues.
- Limited sustainability: as for scalability, a flat approach interconnecting resources is not sustainable in presence of hundreds of thousands of (heterogeneous) entries and hundreds of registered standards; as for flexibility, future systems and specifications must be easily added, as well.

EuroGEOSS experimented a brokering framework to address these challenges. This solution can provide a homogeneous discovery, evaluation, and access framework to leverage heterogeneous resources in a seamless, flexible, and scalable way, thus lowering the entry barrier for users. This is achieved by extending the SOA approach and advancing it through the use of "expert" components. Brokers proved to be effective components on which the emerging Internet of Things approach can be realized. In the EuroGEOSS project we also experimented that the brokering approach can introduce performance limitations. In this case it has to be considered that the brokering components may be replicated for load balancing and redundancy. Also, flexible computing and caching solutions can be adopted to improve performances.

In this paper, we illustrated the EuroGEOSS broker experience in the context of the integrative research, through the description of the following components:

- Disciplinary research
  - Metadata catalog
  - View services
  - Download and access services
- Multidisciplinary research
  - EuroGEOSS Discovery Broker
  - Discovery Augmentation Component
  - Access broker
- Interdisciplinary research
  - Business process description using a common notation (BPMN)
  - Service composition using a CaaS approach
- Participatory Research
  - Web 2.0 discovery

— Web 2.0 access

The diagram in Fig. 3 shows the relationships among the EuroGEOSS components illustrated in the previous sections.

Future work will include a further assessment of the brokering approach in other disciplines via a new research project [66] funded by the European Commission addressing the Water, Ocean, and Weather realms. New functionalities that will be considered for the Multidisciplinary and Interdisciplinary integration include:

- Discovery by using quality-related constraints (e.g., resolution, accuracy, lineage, etc.).
- A quality-related metrics to rank and page results matching a given discovery query.
- Advanced semantics functionalities.

As shown in this paper the brokering approach makes it already possible to connect information resources from multiple domains, building bridges among scientific communities and supporting multi-disciplinary scientific enquiry. For these reasons, the brokering approach has been recently identified at the GEO-VIII Plenary in Istanbul as a new technology to improve the architecture of the GEOSS common infrastructure [67]. Extending this approach and integrating into the Global Earth Observation System of System Common Infrastructure promises to provide a significant service to the advancement of global sustainability research.

Acknowledgment

The authors would like to thank all the EuroGEOSS partners and in particular L. Diaz for the development and support, and the discussions of the Web 2.0 wrappers, M. T. Borzacchiello for the discussion on the integrative research categorization aspects, and E. Boldrini, F. Papeschi, and L. Bigagli for their support in the development of the EuroGEOSS Brokering Framework.

References

[1] Grand Challenges in Global Sustainability Research: A Systems Approach to Research Priorities for the Decade. International Council for Science (ICSU), ICSU. Paris, France, 2010.
[2] EUROPE 2020. A Strategy for Smart, Sustainable and Inclusive Growth. European Commission, Publications Office, Luxembourg, COM(2010) 2020 Final., 2010 [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF
[3] GEOSS. [Online]. Available: http://www.earthobservations.org/geoss.shtml
[4] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). European Commission, Publications Office, Luxembourg, 2007 [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT
[5] Global Monitoring for Environment and Security (GMES). [Online]. Available: http://www.gmes.info/
[6] Shared Environmental Information System (SEIS). [Online]. Available: http://ec.europa.eu/environment/seis/
[7] EuroGEOSS. [Online]. Available: http://www.eurogeoss.eu/
[8] B. Tress, G. Tress, and G. Fry, "Defining concepts and the process of knowledge production in integrative research," in *Proc. Frontis Workshop From Landscape Research to Landscape Planning: Aspects of Integration, Education and Application*, Wageningen, The Netherlands, Jun. 1–6, 2004, vol. 12.
[9] C. Figueiredo, G. Gaigalas, F. Achard, and J. M. Iglesias, Report on User Requirements for the EuroGEOSS Forestry Operating Capacity, EuroGEOSS D3.1, 2009 [Online]. Available: http://www.eurogeoss.eu/Documents/EuroGEOSS_D3_1.pdf
[10] JRC, EC. [Online]. Available: http://ec.europa.eu/dgs/jrc/index.cfm
[11] Dublin Core. [Online]. Available: http://dublincore.org/
[12] *Geographic Information—Metadata*, ISO/IEC_19115:2003, 2003.
[13] *Geographic Information—Services*, ISO_19119:2005, 2005.
[14] *Geographic Information—Metadata—XML Schema Implementation*, ISO/TS_19139:2007, 2007.
[15] E. O' Tuama, G. Dubois, A. Cottam, I. May, and I. Fisher, Report on User Requirements for the EuroGEOSS Biodiversity Operating Capacity. EuroGEOSS D 4.1, 2009 [Online]. Available: http://www.eurogeoss.eu/Documents/EuroGEOSS_D4_1.pdf
[16] Global Biodiversity Information Facility (GBIF). [Online]. Available: http://www.gbif.org/
[17] Open Archive Initiative (OAI). [Online]. Available: http://www.openarchives.org/
[18] Digital Observatory for Protected Areas (DOPA). [Online]. Available: http://dopa.jrc.ec.europa.eu/
[19] GEOBON. [Online]. Available: http://www.earthobservations.org/documents/cop/bi_geobon/geobon_technical_summary_imp_plan.pdf
[20] United Nations Environment Programme, World Conservation Monitoring Centre, UNEP-WCMC. [Online]. Available: http://www.unep-wcmc.org/
[21] Birdlife International and Royal Society for the Protection of Birds, RSPB. [Online]. Available: http://www.rspb.org.uk/
[22] B. Hofer, S. Niemeyer, and A. Ceglar, Report on User Requirements for the EuroGEOSS Drought Operating Capacity. EuroGEOSS D5.1, 2010 [Online]. Available: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_5_1.pdf
[23] Web Map Server (WMS) Implementation Specification ver. 1.3.0, OGC_06-042, 2006.
[24] EDO. [Online]. Available: http://edo.jrc.ec.europa.eu/php/index.php?action=view&id=36
[25] World Meteorological Organization (WMO). [Online]. Available: http://www.wmo.int/
[26] U.S. National Integrated Drought Information System (NIDIS). [Online]. Available: http://www.drought.gov
[27] GEO Beijing Summit., 2010 [Online]. Available: http://www.ogcnetwork.net/pub/ogcnetwork/GEOSS/AIP3/pages/Demo.html
[28] L. Vaccari, S. Nativi, and M. Santoro, Report on Requirements for Multidisciplinary Interoperability. EuroGEOSS D2.1.1, 2010 [Online]. Available: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_2_1_1.pdf
[29] EuroGEOSS Discovery Broker (EDB). [Online]. Available: http://www.eurogeoss-broker.eu/
[30] EDB. [Online]. Available: http://geossregistries.info/
[31] Thematic Realtime Environmental Distributed Data Services (THREDDS). [Online]. Available: http://www.unidata.ucar.edu/projects/THREDDS/
[32] Open-source Project for a Network Data Access Protocol (OPeNDAP). [Online]. Available: http://www.opendap.org/
[33] Genesi. [Online]. Available: http://portal.genesi-dr.eu/
[34] SeaDataNet. [Online]. Available: http://www.seadatanet.org/
[35] EuroGEOSS portal. [Online]. Available: http://ec2-174-129-9-172.compute-1.amazonaws.com/eurogeoss/gi-portal/index.jsp)
[36] GENESIS FP7. [Online]. Available: http://www.genesis-fp7.eu/
[37] M. Santoro, P. Mazzetti, C. Fugazza, S. Nativi, and M. Craglia, "Semantics enabled queries in EuroGEOSS: a discovery augmentation approach," in *American Geophysical Union, Fall Meeting*, 2010.
[38] GENESIS Vocabulary Service (GVS). [Online]. Available: http://semanticlab.jrc.ec.europa.eu/
[39] Simple Knowledge Organization System (SKOS). [Online]. Available: http://www.w3.org/2004/02/skos/
[40] SPARQL. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/
[41] GEMET. [Online]. Available: http://www.eionet.europa.eu/gemet
[42] INSPIRE Feature Concept Dictionary and Glossary. [Online]. Available: http://inspire-registry.jrc.ec.europa.eu/registers/FCD
[43] GEOSS Societal Benefit Areas. [Online]. Available: http://en.wikipedia.org/wiki/Societal_Benefit_Areas
[44] ISO 19119 geographic services taxonomy. [Online]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890
[45] EuroGEOSS Drought Vocabulary. [Online]. Available: http://eurogeoss.unizar.es/home/thesaurus/droughtVocabulary.pdf
[46] GEOSS AIP-3 Water Ontology. [Online]. Available: http://wiki.ieee-earth.org/Best_Practices/GEOSS_Transverse_Areas/Data_and_Architecture/Ontologies/Water_Management_domain
[47] GEOSS Earth Observation parameters. [Online]. Available: http://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf

[48] Global Change Master Directory (GCMD) scientific keywords. [Online]. Available: http://gcmd.nasa.gov/Resources/valids/archives/keyword_list.html
[49] M. Howard, S. Payne, R. Sunderland, and R. , Technical Guidance for the INSPIRE Schema Transformation Network Service. 2010 [Online]. Available: http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_TG_v3-0.pdf
[50] M. Chinosi and Trombetta, "BPMN: an introduction to the standard," in *Computer Standards & Interfaces*. New York: Elsevier, in press.
[51] UncertWeb. [Online]. Available: http://www.uncertweb.org
[52] L. Bigagli, M. Santoro, V. Angelini, P. Mazzetti, and S. Nativi, UncertWeb Deliverable 2.2—Service Frameworks for Modelling Resources. 2011 [Online]. Available: http://www.uncertweb.org/documents/deliverables/service-frameworks-for-modelling-resources/download
[53] Intergovernmental Panel on Climate Change (IPCC)., 2007 [Online]. Available: http://www.newscientist.com/article/dn18363-debate-heats-up-over-ipcc-melting-glaciers-claim.html?DCMP=OTC-rss&nsref=online-news
[54] M. F. Goodchild, "Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0," *Int. J. Spatial Data Infrastructures Research*, vol. 2, pp. 24–32, 2007.
[55] S. Elwood, M. F. Goodchild, and D. Sui, "Researching volunteered geographic information: spatial data, geographic research, and new social practice," *Annals Assoc. Amer. Geographers*, Sep. 21, 2011.
[56] Open Data challenge. [Online]. Available: http://opendatachallenge.org/
[57] L. Diaz, C. Granell, O. Fonts, and J. Gil, Report on Web 2.0 and Alternative Search Mechanisms. EuroGEOSS D 2.6.1., 2010 [Online]. Available: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_2_6_1.pdf
[58] P. Gonçalves, OpenGIS® OpenSearch Geospatial Extensions Draft Implementation Standard, Version 0.0.2. Open Geospatial Consortium Inc. Ref. OGC 09-084r3, 2010.
[59] Turner *et al.*, OpenSearch Geo Extension 1.0 (Draft 2), 2010 [Online]. Available: http://www.opensearch.org/Specifications/OpenSearch/Extensions/Geo/1.0/Draft_2
[60] F. O. Ostermann and L. Spinsanti, "A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management," in *Proc. AGILE 2011*, S. Geertman, W. Reinhardt, and F. Toppen, Eds., 2011, Utrecht.
[61] M. M. Klopfer and I. I. Simonis, "SANY—An open service architecture for sensor networks," SANY Consortium, 2009.
[62] A. Broering, T. Foerster, and S. Jirka, "Interaction patterns for bridging the gap between sensor networks and the sensor web," in *Proc. 8th IEEE Int. Conf. Pervasive Computing and Communications: WoT 2010: 1st Int. Workshop on the Web of Things*, 2010, pp. 732–737.
[63] M. F. Goodchild, "The use cases of digital earth," *Int. J. Digital Earth*, vol. 1, no. 1, pp. 31–42, 2008.
[64] M. Craglia, M. F. Goodchild, A. Annoni, G. Camara, M. Gould, W. Kuhn, D. M. Mark, I. Masser, D. J. Maguire, S. Liang, and E. Parsons, "Next-generation digital earth. A position paper from the Vespucci initiative for the advancement of geographic information science," *Int. J. Spatial Data Infrastructure Research*, vol. 3, pp. 146–167, 2008.
[65] A. A. Annoni, M. Craglia, M. Ehlers, Y. Georgiadou, A. Giacomelli, M. Konecny, N. Ostlaender, G. Remetey-Fülöpp, D. Rhind, P. Smits, and S. Schade, "A european perspective on digital earth," *Int. J. Digital Earth*, vol. 4, no. 4, pp. 271–284, 2011.
[66] GeoWow. [Online]. Available: http://www.geowow.org
[67] Global Earth Observation System of System Common Infrastructure. [Online]. Available: http://www.earthobservations.org/art_017_004.shtm

Lorenzino Vaccari is a post-doc researcher at the European Commission Joint Research Centre in the Digital Earth and Reference Data Unit. He received his Ph.D. degree in computer science and telecommunications from the University of Trento, Italy, in 2009. He is now working in the field of Interoperability and Spatial Data Infrastructures. Specialties include development of global, international and local Spatial Data Infrastructures, semantic interoperability among heterogeneous geographic services and datasets, GIS management, P2P architectures for emergency response and semantic web issues. His work supports the development of the Global Earth Observation System of Systems (GEOSS) and the European Spatial Data Infrastructure established by the INSPIRE directive.

Max Craglia is a senior researcher at the European Commission Joint Research Centre in the Digital Earth and Reference Data Unit that has the responsibility for the technical coordination of the Infrastructure for Spatial Information in Europe. He is the scientific coordinator of the EuroGEOSS project discussed in this paper, and one of the founders of the Vespucci Initiative for the advancement of Geographic Information Science (www.vespucci.org). Prior to joining the JRC in 2005, Max was a Senior Lecturer at the University of Sheffield teaching GIS for urban planners, and researching areas of spatial data infrastructure deployment and use, and data policy.

Cristiano Fugazza is a post-doc researcher at the European Commission Joint Research Centre in the Digital Earth and Reference Data Unit. He obtained his Ph.D. in computer science by the Dept. of Information Technologies, University of Milan, focusing on Knowledge Management and Semantic Web technologies. He has contributed to, and published in international journals, conferences, and workshops in the domains of spatial data infrastructures, supply chain management, business modelling, digital rights management, personal information protection, and persistent identifiers. He is currently working on the creation, harmonization, and exploitation of multilingual thesauri in the SKOS format for the purpose of annotation and discovery of spatial resources. He is supporting the development of the Global Earth Observation System of Systems (GEOSS) and the European spatial data infrastructure established by the INSPIRE directive by managing a comprehensive repository of harmonized thematic thesauri.

Stefano Nativi (M'95–SM'09) coordinates the Earth and Space Science Informatics Laboratory (ESSI-lab) of the Institute of Atmospheric Pollution Research of the National Research Council of Italy (CNR-IIA). He is adjunct Professor of systems for land management at the University of Padua. He is an expert on multi-disciplinary interoperability and cyber(e)-infrastructures for Earth Sciences, gaining research and operational experience in the fields of Geospatial Information, Earth Observation, and international standardization. He is member of the GEOSS Infrastructure Implementation Board and of the INSPIRE Drafting Teams (metadata and data specification). He is president of the Earth and Space Sciences Informatics (ESSI) division of the European Geosciences Union (EGU), and co-chair of the OGC Earth Systems Science Domain Working Group (ESS DWG). He received the SIF Award medal appointed by the IEEE ICEO for the work done to advance GEO-GEOSS multi-disciplinary interoperability.

Mattia Santoro is a Ph.D. student in Methods and Technologies for Environmental Monitoring at the University of Basilicata, Italy. He obtained a degree in computer science at the University of Florence, Italy. He has been working at the Earth and Space Science Informatics Laboratory (ESSI-lab) of the National Research Council of Italy (CNR) since 2007. His research interests are semantic discovery, environmental models interoperability with a particular focus on the GEO/GEOSS initiative. He has contributed to, and published in various international journals, conferences and workshops in the field of multi-disciplinary interoperability of Geo-spatial resources. He was a member of the Global Earth Observation System of Systems Architecture Interoperability Pilots 3 and 4, in the working groups addressing applications for water, biodiversity and drought.