Published Version DOI: 10.1016/j.cageo.2018.03.008

Interoperability Challenges in River Discharge Modelling: a Cross Domain Application Scenario

Mattia Santoro, Volker Andres, Simon Jirka, Toshio Koike, Ulrich Looser, Stefano Nativi, Florian Pappenberger, Manuela Schlummer, Adrian Strauch, Michael Utech, and Ervin Zsoter

Abstract— **River discharge is a critical water cycle variable, as it integrates all the processes (e.g. runoff and evapotranspiration) occurring within a river basin and provides a hydrological output variable that can be readily measured. Its prediction is of invaluable help for many water-related tasks including water resources assessment and management, flood protection, and disaster mitigation. Observations of river discharge are important to calibrate and validate hydrological or coupled land, atmosphere and ocean models. This requires using datasets from different scientific domains (Water, Weather, etc.). Typically, such datasets are provided using different technological solutions. This complicates the integration of new hydrological data sources into application systems. Therefore, a considerable effort is often spent on data access issues instead of the actual scientific question.**

This paper describes the work performed to address multidisciplinary interoperability challenges related to river discharge modeling and validation. This includes definition and standardization of domain specific interoperability standards for hydrological data sharing and their support in global frameworks such as the Global Earth Observation System of Systems (GEOSS).

The research was developed in the context of the EU FP7-funded project GEOWOW (GEOSS Interoperability for Weather, Ocean and Water), which implemeted a “River Discharge” application scenario. This scenario demonstrates the combination of river discharge observations data from the Global Runoff Data Centre (GRDC) database and model outputs produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) predicting river discharge based on weather forecast information in the context of the GEOSS.

Keywords – **Interoperability; Brokering; River Discharge; Standardization; GEOSS.**

I. INTRODUCTION

River discharge is a key variable of the global water cycle as it provides an integrated signal of all the hydrological processes occurring within a river basin. Its observation and prediction is an information source for many application areas such as water resources assessment and management, design and operation of technical facilities (dams, reservoirs), aquatic ecosystem management or flood protection and disaster mitigation (Group on Earth Observation, 2014). Additionally, the Global Climate Observing System (GCOS) programme (World Meteorological Organization, 2010) (Bojinski, et al., 2014) defined river discharge as an Essential Climate Variable (ECV) because its observation is critical for the characterization, understanding and prediction of the Earth’s climate and its changes. In these various application areas, the exchange of river discharge data across disciplinary and institutional boundaries is of high importance (e.g. for the validation and calibration of hydrological prediction models or regional and global analysis of river discharge). Thus, scientists and policy makers would highly benefit from a common interoperability framework that addresses the needs of the Earth science community for improved information sharing among its different domains through an eased access to diverse hydrological data sources and their integration into application systems.

From a technological perspective, the main challenge to achieve this stems from the diversity of the involved components (i.e. geospatial data sharing infrastructures, services, data models, etc.). In fact, when creating a geospatial data-sharing infrastructure for a given scientific domain, a set of discipline-specific needs must be taken into account and addressed. This has an impact on the adopted technological solutions as far as data, metadata, processing models, services protocols and interfaces, semantics, and embedded knowledge are concerned (Craglia, Nativi, Santoro, Vaccari, & Fugazza, 2011). As result, the landscape of existing

This work was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 282915 (GEOWOW).

M. Santoro and S. Nativi are with the Consiglio Nazionale delle Ricerche, Italy (e-mail: mattia.santoro@cnr.it; stefano.nativi@cnr.it).
V. Andres and U. Looser are with Global Runoff Data Centre at the Federal Institute of Hydrology (BfG) (e-mail: volker.andres@mac.com; looser@bafg.de).
S. Jirka is with 52°North GmbH, Germany (e-mail: s.jirka@52north.org).
T. Koike is with University of Tokyo, Japan (e-mail: tkoike@hydra.t.u-tokyo.ac.jp).
F. Pappengerger and E. Zsoter are with European Centre for Medium-Range Weather Forecasts, UK (e-mail: florian.pappenberger@ecmwf.int; ervin.zsoter@ecmwf.int).
M. Schlummer and A. Strauch are with University of Bonn, Germany (e-mail: manuela.schlummer@uni-bonn.de ; astrauch@uni-bonn.de).
M. Utech is with KISTERs AG, Germany (e-mail: michael.utech@kisters.de).

Published Version DOI: 10.1016/j.cageo.2018.03.008

discipline-specific infrastructures for geospatial data sharing is highly heterogeneous (Laney, 2001). Multidisciplinary interoperability is essential to achieve an effective integration of such information systems and to provide integrated access to a range of advanced information and processing resources for the environment and the policy-makers support (Mazzetti & Nativi, 2012).

This research focuses on the work performed to address multidisciplinary interoperability challenges for river discharge modeling and validation. The research was developed in the context of the EU FP7 project GEOWOW (GEOSS Interoperability for Weather, Ocean and Water) and of the Global Earth Observation System of Systems (GEOSS). A cross-domain application scenario was implemented to combine river discharge observations data from the Global Runoff Data Centre (GRDC) database and model outputs produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) predicting river discharge based on weather forecast information.

This paper is structured as follows. Section II introduces the scientific needs for this research. Section III describes the main interoperability challenges stemming from the scientific needs. Section IV describes the adopted multidisciplinary interoperability approach, the resulting system architecture and the developed application scenario and the results. Secion V provides an example of possible further applications. Finally, Section V draws the conclusions and describes possible applicability in a different application scenario.

II. SCIENTIFIC FRAMEWORK: RIVER DISCHARGE AND PREDICTION

According to the Emergency Event Data Base (EM-DAT) of the Centre for Research on Epidemiology of Disasters, floods (including general river floods, flash floods, storm surges/coastal floods) accounted for 30% of all recorded natural disasters between 1900 und 2006, while 6,899,095 people were killed and more than 3 billion people affected (Adikari & Yoshitani, 2009). Since, furthermore, flooding is an increasingly frequent thread in many regions of the world (Adikari & Yoshitani, 2009) (Intergovernmental Panel on Climate Change, 2014) flood protection and risk management are a urgent issues for policymakers and managers to reduce the vulnerability of societies against flood damages. Crucial elements of flood disaster prevention are river discharge forecasting systems based on numerical weather prediction (NWP) models that provide medium range forecasts and early warnings to civil protection authorities, flood forecasting services and the public (Cloke & Pappenberger, 2009). A recent trend in the field of medium term flood forecasting goes to Ensemble Prediction Systems that are used as input for hydrological models to forecast river discharge (Cloke & Pappenberger, 2009). For the calibration and validation of river discharge prediction systems data centers such as the Global Runoff Data Centre (GRDC), that collects river discharge data from more than 9000 stations in 160 countries in various data formats, provide valuable databases. A subset of these river discharge data sets is made freely available as GEOSS Data-CORE[1] and is discoverable in standardized formats (see section IV.B) via the GEOSS Portal[2].

General issues regarding the collection and combined use of data from different sources (in the case of the GRDC various National Hydrological Services) and different domains (e.g. river discharge observations from national hydrological services and ensemble predictions from meteorologists) are the large variety of interfaces and data formats. This heterogeneity regarding data formats and data exchange mechanisms within the hydrology domain combined with data quality assurance and data policy issues lead to the fact that many datasets are uploaded to global databases such as the Global Runoff Data Centre with a delay of a few years. This is an obstacle to all applications that require near-real-time data access (access to data within one year of the measurement). Scientists and decision makers often suffer from the low interoperability within their specific domain and between different domains and therefore have to spend considerable effort on data access issues besides the actual scientific question and application.

Application areas such as the validation and calibration of river discharge prediction models or regional and global analysis of river discharge would highly benefit from an increased multidisciplinary interoperability that allows a harmonized and timesaving access to different data sources and an eased integration of new hydrological data sources into application systems.

The river discharge forecasting performed within the framework of the GEOWOW project was performed for NWP models available in the THORPEX Interactive Grand Global Ensemble (TIGGE) archive (Bougeault, et al., 2010). TIGGE is a major component of the WWRP-THORPEX research program, whose aim is to accelerate improvements in forecasting high-impact weather. TIGGE provides a database of ensemble predictions, collected from leading operational NWP centers for scientific research on various topics since October 2006, and has been instrumental in supporting cooperation between the academic and operational meteorological communities. In turn, it has provided a basis for research on objective evaluation, predictability and dynamical processes. The research undertaken includes studies on multi-model combination, correction for systematic errors, tropical cyclones and the dynamics of extra-tropical storm tracks. Although TIGGE is a research project, it has proved invaluable

---

[1] https://www.earthobservations.org/geoss_dsp.shtml
[2] http://www.geoportal.org/

Published Version DOI: 10.1016/j.cageo.2018.03.008

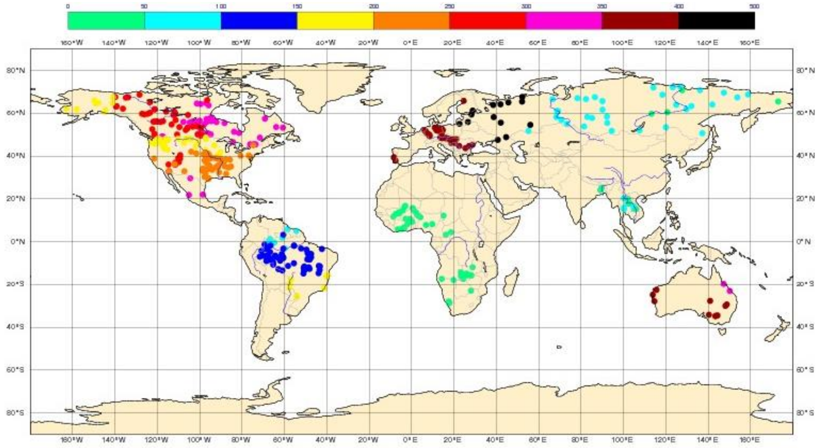


Fig. 1. GRDC stations used in the TIGGE discharge modelling use case scenario

for the development of several applications in relation to operational forecasting. Examples include the development of multi-model tropical cyclone tracks, severe weather early warning products for heavy rainfall or strong winds, and various applications in flood forecasting using coupled hydrological models.

The modelling work was based on the HTESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land) land-surface model used operationally at ECMWF (Balsamo, et al., 2009). The offline version of HTESSEL was extended to accommodate ensemble forecast runs from TIGGE with also using ECMWF climate and initial conditions. The HTESSEL runs require atmospheric forcing of temperature, humidity, pressure, wind, radiation and precipitation. The radiation parameters are not archived in TIGGE therefore ERA Interim (ECMWF global atmospheric reanalysis from 1979) was used to provide common radiation forcing replacement. The rest of required forcing parameters were available from four TIGGE models, ECMWF, UKMO (UK Met Office), NCEP (National Centers for Environmental Prediction in the US) and CMA (China Meteorological Administration). The other models had some missing parameters. The land-surface model output is runoff, which is provided in the grid structure of the HTESSEL model. To aggregate the runoff into river discharge we used the CaMa-Flood river routing scheme and produced discharge for about 400 global river catchments (Fig. 1).

The production of the TIGGE based hydrological discharge forecasts covered the period 2008-2013. Within the framework of the GEOWOW project, the analysis of this discharge data set was performed using GRDC observations and the skill properties of the different TIGGE model's hydrological predictions were highlighted. Based on the developments described in the following sections to enhance multidisciplinary interoperability, a dedicated data server with PostgreSQL database support and an OGC Sensor Observation Service instance (from 52N) to serve the data was installed at ECMWF to support the data provision required by the analysis.

III. INTEROPERABILITY CHALLENGES

To meet the identified scientific objectives a set of interoperability requirements needs to be addressed, both at a discipline-specific level and at a multidisciplinary level. Discipline-specific requirements include:

- Accommodate discipline-specific needs (e.g. supporting a specific data format, applying conventions to narrow down generic concepts defined in a standard, etc.); Section IV.B describes in detail the domain-specific needs stemming from the application scenario and how they were addressed;
- Develop ad-hoc applications to provide users with easy exploitation tools for the discipline-specific functions.

Both these requirements can be addressed in several ways: (i) implement use case specific solutions, (ii) extend existing interoperability solutions with ad-hoc additions, and (iii) define and implement new interoperability arrangements for the domain needs. Each of these approaches presents advantages and drawbacks. Implementing use case specific solutions is certainly the most rapid approach, at least in the short-term. Besides, this approach provides a high degree of freedom in the development phase. The main drawback is that these solutions will remain isolated from the rest of existing systems, which represents a major issue when an effective interoperability is aimed at.

Extending existing interoperability solutions by implementing ad-hoc additions – e.g. implementing existing standard specifications with new non-standard functionalities – represents a step forward for more effective interoperability. In fact, in this case, the interoperability with external systems is ensured by the existing solution – i.e. with the existing standard specification the new discipline-specific needs are addressed by the implemented extensions. The main drawback here is that the new functionalities will remain isolated, i.e. not interoperable and in turn not useable by external systems.

The last approach has the clear advantage of a high interoperability level, since it defines new interoperability arrangements that external systems can use (implement) to exploit the new domain-specific functionalities. The drawback here is that a

Published Version DOI: 10.1016/j.cageo.2018.03.008

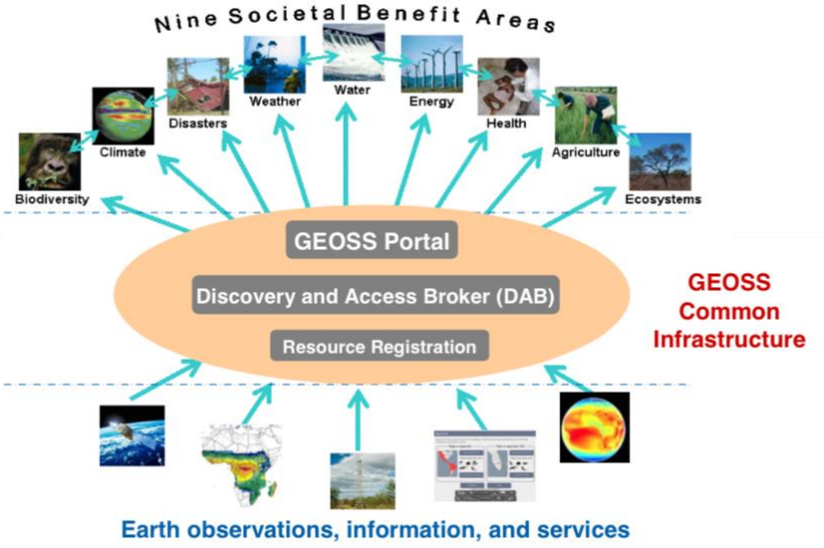


Fig. 2. High-level GCI Architecture (Group on Earth Observation, 2014a)

standardization process, which might be time consuming, is required. Section IV.B describes the application of the third approach for this particular application scenario.

Multidisciplinary interoperability requirements include:

- Homogeneous discovery of resources from different discipline-specific infrastructures;
- Homogeneous access/download of resources from different discipline-specific infrastructures.

The two main approaches for achieving a cross-domain interoperability infrastructure are (Nativi, Craglia, & Pearlman, 2013): (i) interconnecting the existing (or under-development) discipline/domain infrastructures (*Brokering* approach) or (ii) developing a new overall infrastructure whose domain covers all the Earth Sciences disciplines (*Federated* approach). Both these approaches can be used to build a multidisciplinary *System of Systems* (SoS). The notion of *System of Systems*, and the related *System of Systems Engineering* (SoSE) process, emerged in many fields of applications with the aim of integrating different, autonomous systems to satisfy a global goal (Karcanias & Hessami, 2010). SoS can be usefully described as large-scale integrated system consisting of heterogeneous sub-systems that operate within their own mandates, but are networked together for a common goal (Jamshidi, 2011).

With a federated approach, the federation model ensures interoperability – i.e. a set of specifications and open standards defining the data model and communication interfaces/protocols (Pyster & Olwell , 2013). The advantage of this approach is a very high interoperability level, as all federated systems (i.e. discipline-specific infrastructures) conform to the same federation model and client applications only need to implement the federal model to be able to interact with all federated systems. On the other hand, conforming to the federation model results in an additional overhead for discipline-specific infrastructures and has several drawbacks, as described in detail by (Nativi, Craglia, & Pearlman, 2013), including: (i) complexity, (ii) lack of flexibility, (iii) sustainability, (iv) cost, and (v) managing different maturity levels. Examples of federation-based infrastructures are: the Infrastructure for spatial information in Europe (INSPIRE[3]), the WMO Information System (WIS[4]) , the CUAHSI Hydrological Information System (Maidment, Hooper, Tarboton, & Zaslaksky, 2009), the GlobalSoilMap (Arrouays, et al., 2014), etc.

With the brokering approach, the requirements stemming from the federal model approach are relaxed by separating client applications and discipline-specific infrastructures with a set of new intermediary components: the Brokers (Nativi, Craglia, & Pearlman, 2013). These components are dedicated to provide the necessary mediation and transformation functionalities in a

---

[3] http://inspire.ec.europa.eu
[4] http://www.wmo.int/pages/prog/www/WIS/

Published Version DOI: 10.1016/j.cageo.2018.03.008

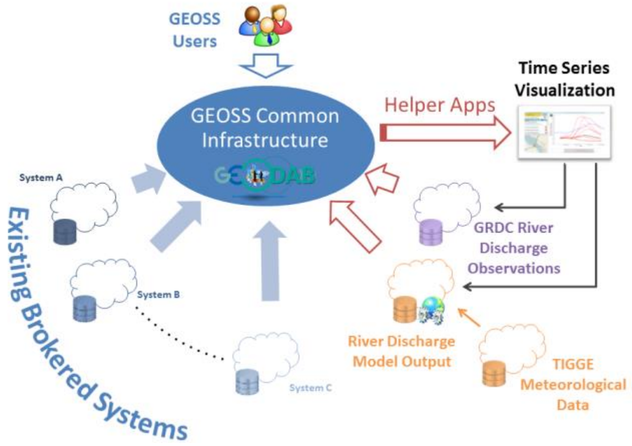


Fig. 3. High-level Multidisciplinary Architecture

transparent way to the SoS components (i.e. the discipline-specific infrastructures) and the client applications. This design allows a much higher degree of freedom to discipline-specific infrastructures and client applications. In fact, the former are allowed to develop and improve their infrastructures while contributing, at the same time, to a cross-domain capability; the latter can choose to implement the most effective interoperability protocol(s) for the functionalities they want to provide to their users. Examples of brokering-based infrastructures include: GEOSS, the ICSU World Data System (Sarkisyan, Goncharova, Gurov, & Markov, 2015), the eReefs Data Brokering layer (Car, et al., 2014), the NSF EarthCube Brokering Building Block (BCube) BCube (Khalsa, 2017) (Santoro M. , Nativi, Pearlman, Khalsa, & Fulweiler, 2015), etc.

It is not in the scope of this manuscript to provide an in-depth discussion about advantages and disadvantages of the brokered and federated approaches for multi-disciplinary interoperability. The data science community is discussing this topic in several arenas, including scientific publications (see citations in above paragraphs) and fora (e.g. as the Research Data Alliance, RDA[5]).

Section IV.A describes how the Brokering approach was applied to this cross-domain application scenario, based on the existing GEOSS brokering framework. Section IV.C provides a description of the extensions that were developed to support this application scenario. Section IV.D describes a use-case of the application scenario with searching, selection and analysis of observed and predicted discharge time series using the newly developed extensions. A more advanced use-case of evaluating ensemble flood prediction is presented in Section V.

IV. TOWARDS IMPROVED MULTIDISCIPLINARY INTEROPERABILITY

GEOSS is composed of contributed Earth Observation systems, ranging from systems collecting primary data, to systems concerned with the creation and distribution of information products. The infrastructure that coordinates access to these systems is the GEOSS Common Infrastructure (GCI); Fig. 2 depicts a simplified diagram of GCI architecture.

One of the key components to achieve multidisciplinary interoperability of the GCI is the GEO Discovery and Access Broker (DAB). The GEO DAB stems from results achieved in the EuroGEOSS project[6] (Vaccari, Craglia, Fugazza, Nativi, & Santoro, 2012) (Nativi, Craglia, Vaccari, & Santoro, 2011) and applies the Brokering approach to implement broker components for discovery (Nativi & Bigagli, 2009), access (Boldrini, Santoro, Papeschi, & Nativi, 2013) and semantic interoperability (Santoro, et al., 2012).

The multidisciplinary interoperability architecture for this application scenario is based on the GEOSS Brokering framework (Nativi, et al., 2015), and extends it where needed in order to accommodate the specific needs for the two domains of Hydrology and Weather. In particular, to address the interoperability requirements identified in the previous section, the following improvements were developed:

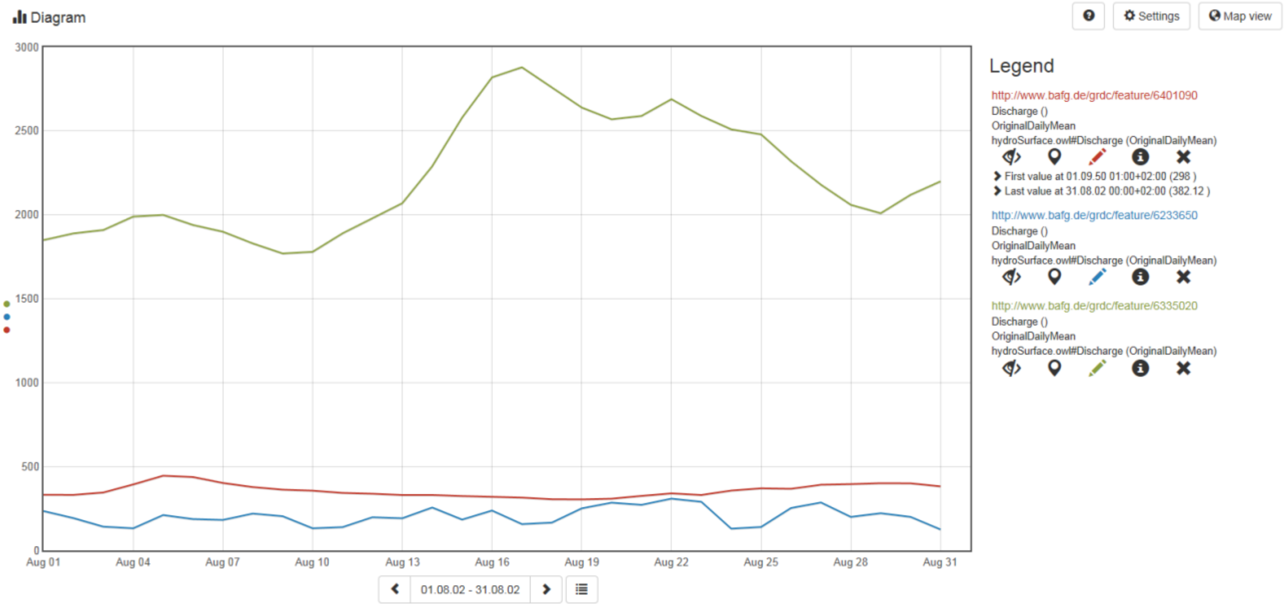
- Development of a Hydrology Profile for the OGC Sensor Observation Service (SOS) 2.0 standard;
- Development of a web application for the visualization and comparison of time series served through this new standard;
- Enhancement of the GEO DAB to support the application scenario:
    - Develop new interoperability arrangements for GRDC and ECMWF capacities;

---

[5] https://www.rd-alliance.org
[6] http://www.eurogeoss.eu

Published Version DOI: 10.1016/j.cageo.2018.03.008

- ○ Select multiple time series for comparison.



First value at 01.09.50 01:00+02:00 (298)
Last value at 31.08.02 00:00+02:00 (382.12)

Fig. 4. Screenshot of the JavaScript SOS-Client

A. Architecture

Fig. 3 depicts the high-level multidisciplinary architecture for this application scenario. For the sake of simplicity, client applications interacting with the GEO DAB are not included in the diagram; in the following, it is assumed that all user actions are executed through a client application.

Users interact with the GEOSSS Common Infrastructure to search, discover, visualize and access data provided by the GEOSS contributing systems. When a user needs a resource, she/he sends a query to the GEO DAB. This, in turn, executes in parallel the two following actions – details of the GEO DAB technology and its query execution, including distribution and harvesting, are available in online[7] and in (Nativi, et al., 2015):

- forwarding the query to the different systems contributing to GEOSS (hereafter, brokered systems);
- executing the query on its internal metadata repository that contains metadata from previously harvested systems.

After aggregating the results from both the distributed and the local queries, the GEO DAB returns the result set back to the user.

In Fig. 3 existing brokered systems are depicted at the bottom left of the diagram, while the new discipline-specific systems developed by this pilot are depicted at the bottom right. These are composed of service components providing access to hydrological/meteorological data and an ad-hoc client application that was developed for visualizing and comparing multiple time-series. Red arrows represent the needed enhancements to interconnect the GEO DAB with the new systems.

It is worth noting that this allows enriching the overall multidisciplinary infrastructure without the need to make any change on the rest of the system – e.g. existing interconnections between the brokered systems and the GEOSS Common Infrastructure (GCI) or between the GCI and the users (i.e. client applications).

The next section B describes in details the developed solutions for the discipline-specific needs, while section C provides details about the new interoperability arrangements, developed to connect the new infrastructures to the GEO DAB.

B. Discipline-specific Infrastructures

The OGC Sensor Web Enablement (SWE) framework is a set of standards to allow the interoperable integration of sensors and their measurements into spatial data infrastructures (Bröring, et al., 2011). Within this standards framework, the OGC Sensor Observation Service (O&M) 2.0 specification (Bröring, Stasch, & Echterhoff, 2012) is important for serving data such as discharge predictions and observations. This standard defines a set of operations to query measurement data as well as the associated metadata. Complementary to the SOS standard, the OGC WaterML 2.0 standard (Taylor, 2012) is highly relevant. WaterML 2.0 is a specialization of the ISO/OGC Observations and Measurements standard (International Organization for Standardization, 2011) specifically adjusted to the needs of hydrology. While the SOS standard defines the operations and query parameters, WaterML 2.0 specifies a data model and format to specify how an SOS server shall encode the hydrological data it returns in a standardized form.

Based on several studies performed in the context of the GEOWOW project, it was possible to observe differing ways on how to apply and implement the SOS standard in practice. Especially in different thematic domains, there was a need to apply the SOS

---

[7] http://www.geodab.net

Published Version DOI: 10.1016/j.cageo.2018.03.008

standard in specific ways. For example, central concepts such as "procedure" (= "what is a sensor") were interpreted differently. This flexibility is important for ensuring the best possible support of different thematic domains but can cause at the same time interoperability issues. To increase interoperability within the hydrology domain and to offer a consistent approach on how to provide hydrological data to the GCI, the GEOWOW project has worked on a hydrology profile of the SOS 2.0 specification. This profile offers a clearly defined approach on how to use the SOS standard for hydrological applications. In 2014, the OGC adopted this profile as an OGC Best Practice Document (Andres, Jirka, & Utech, 2014).

The SOS 2.0 Hydrology Profile defines several conventions on how to apply the SOS concepts to hydrology but also requirements regarding the functionality of SOS servers. For example, the profile requires that SOS servers shall support WaterML 2.0 as a response format, so that client implementers have clear information on which types of SOS responses are able to be supported . Other conventions concern the types of sampling features (shall always be a WaterML 2.0 monitoring point) or the use of SOS concepts/terms such as procedure or observed property. Another important topic concerns a clear distinction between fields that can be used as identifiers in SOS requests and fields that may only be used by clients as labels. Finally, additional types of exceptions and errors are specified (e.g. mechanisms how to deal with empty responses and how to enforce limitations on response sizes).

Furthermore, the SOS 2.0 Hydrology Profile specifies an enhancement of the SOS 2.0 standard to facilitate the retrieval of information about the time series offered by an SOS server. For this purpose, the profile specifies the additional "GetDataAvailability" operation, which is not yet included in the official OGC standard. This operation allows clients to determine for which locations, parameters, time spans, etc. observation data is available. Compared to regular SOS servers this is an important enhancement, which significantly facilitates the development of client applications.

To validate and evaluate the developed SOS 2.0 Hydrology Profile, the GEOWOW partners 52°North and KISTERS have developed corresponding client and server implementations. On the server side, the 52°North Sensor Observation Service 4.x[8] was enhanced to support the SOS 2.0 Hydrology Profile including WaterML 2.0. This implementation offers the option to enable a broad range of existing hydrological databases with an SOS server interface. Complementary, the KISTERS Web Interoperability Solution (KiWIS) time series server application[9] was enhanced, too, to support the SOS 2.0 Hydrology Profile as well as WaterML 2.0. This allows users who rely on KISTERS software for managing hydrological data, to serve easily their data as specified in the SOS 2.0 Hydrology Profile. As a result, two independent implementations of the SOS 2.0 Hydrology Profile are available, which were subsequently used for compatibility testing and evaluating the profile.

For demonstrating the interoperability gained through the SOS 2.0 Hydrology Profile both, 52°North and KISTERS, developed client applications, as well. KISTERS enhanced its DataServiceConsumer component with a SOS 2.0 plugin able to read time series data from the KiWIS and 52°North server implementations of the SOS 2.0 Hydrology Profile. At the same time, 52°North has worked on a similar enhancement of its Sensor Web client such as the JavaScript SOS Client[10]. Fig. 4 shows an exemplary screenshot of this application. Both clients fully support the SOS 2.0 Hydrology Profile and utilize all operations defined in it to provide users with the possibility to discover, visualize and download WaterML data from a SOS 2.0 Hydrology Profile compliant server.

The resulting implementations were deployed to serve openly available data of the GRDC. In addition, the ECMWF has installed an instance of the 52°North SOS to serve discharge prediction data.

C. *Enhancing the GEO DAB*

In order to add the previously described infrastructures to the set of GEO DAB brokered systems, it was necessary to enhance the GEO DAB framework with new interoperability arrangements – i.e. new modules in charge of communicating with the new systems. The GEO DAB interacts with external systems using a set of software components, called *Accessors*. When a new external system must be accessed, the GEO DAB configuration is updated the by GEO DAB administrator, providing the link to the external system service, the service type and version (e.g. OGC SOS 2.0 Hydrology Profile); this way the GEO DAB has all required information to instantiate the proper Accessor to communicate with the external system. Accessors are in charge of implementing the needed adaptation/mediation functionalities to interact with external systems (Nativi, Craglia, & Pearlman, 2013) (Nativi & Bigagli, 2009). Briefly, these components translate requests expressed according to the GEO DAB internal query language into requests expressed according to the language of the external system. Besides, these components execute data and metadata model mapping from the external systems to the GEO DAB internal system.

A new *Accessor* for the SOS 2.0 Hydrology Application Profile was developed to interact with the new systems. The new Accessor implements a mapping between the SOS 2.0 Hydrology Application Profile metadata model and the GEO DAB internal one (which is based on ISO 19115 (International Organization for Standardization, 2003) and 19115-2 (International Organization for Standardization, 2008)). This way it was possible to harvest the content from the SOS Hydrology Application Profile servers in order to make it available through the GCI. The harvesting task is a periodic task which is executed by the GEO DAB system according to the schedule defined by the GEO DAB administrator in agreement with the data provider.

---

[8] http://www.52north.org/sos

[9] http://kiwis.kisters.de/KiWIS/

[10] https://52north.org/software/software-projects/helgoland/

TABLE I
Mapping sos:ObservationOffering to ISO 19115/19115-2

SOS 2.0 Hydrology Profile	GEO DAB
swes:identifier	gmd:fileIdentifier
sos:observedArea	gmd:geographicElement
sos:phenomenonTime	gmd:temporalElement
swes:name	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:title
swes:description	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:abstract
sos:responseFormat	gmd:distributionInfo/gmd:MD_Distribution/gmd:distributionFormat/gmd:MD_Format
swes:procedure	gmi:LE_ProcessStep/gmd:description
swes:relatedFeature	gmi:acquisitionInformation/gmi:MI_AcquisitionInformation/gmi:objective*

* To avoid listing all sampling features available in a SOS server, the *swes:FeatureRelationship* is used to list sampled features (e.g. river, lake) only where sampling features provide measurements for.

TABLE III
Mapping sos:featureMember and sos:dataAvailabilityMember to ISO 19115/19115-2

SOS 2.0 Hydrology Profile	GEO DAB
parentIdentifier+gml:identifier	gmd:fileIdentifier
sams:shape	gmd:geographicElement
sos:dataAvailabilityMember/om:phenomenonTime*	gmd:temporalElement
sos:dataAvailabilityMember/om:observedProperty*	gmd:MD_keywords/gmd:keyword
gml:name	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:title
swes:description (Inherited from Parent) + gml:name	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:abstract
sos:responseFormat (Inherited from Parent)	gmd:distributionInfo/gmd:MD_Distribution/gmd:distributionFormat/gmd:MD_Format
*Server base URL ***	gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:online/gmdce/gmd:linkage
*urn:ogc:serviceType:SensorObservationService:2.0.0:HTTP ****	gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:online/gmdce/gmd:protocol
gml:identifier	gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:online/gmdce/gmd:name
sf:sampledFeature	gmi:acquisitionInformation/gmi:MI_AcquisitionInformation/gmi:objective

* The *GetDataAvailability* operation is used to retrieve this information.
** This element is created from the base URL of the SOS server.
*** This URN is constant for SOS 2.0.0 servers.

For each *sos:ObservationOffering* in the Capabilities documents provided by the SOS servers, the Accessor generates an ISO metadata record of type collection (i.e. setting the ISO element *gmd:hierarchyLevel* to *series*) applying the mapping defined in Table I.

The above-described mapping generates metadata records describing the provided observation collection. This is further enriched by issuing specific *DescribeSensor* requests to the Hydrology Application Profile server and mapping the *swes:SensorDescription* to ISO 19115-2 as defined in Table II.

After generating the collection-level metadata records from *sos:ObservationOffering*, these records are scanned and for each of them a *GetFeatureOfInterest* request is used to retrieve actual sampled feature associated with the observation procedure. For each *sos:featureMember* element in the *GetFeatureOfInterest* response, an ISO metadata record of type dataset (i.e. setting the ISO element *gmd:hierarchyLevel* to *dataset*) is created as child of the collection record. The mapping defined in Table III is applied to create such records.

Plugging the new *Accessor* into the GEO DAB allows discovery and access of data from the new discipline-specific infrastructures described in section B. These infrastructures provide also a new visualization component that is able to display multiple time-series appropriately for the application scenario. This component is linked to the overall multidisciplinary architecture by means of the *GEOSS Helper Application* (Santoro, Nativi, & Menard, 2013). This is a prototypal functionality of the GEO DAB that enriches metadata records with information about specific client applications – i.e. Helper Applications – that are able to access and/or visualize the referenced dataset. The *Accessor* developed for brokering SOS 2.0 Hydrology Application Profile servers generates appropriate linkages that can be clicked by users to open the SOS-client (see section B) and visualize the discovered observation time-series.

TABLE II
MAPPING SWES:SENSORDESCRIPTION TO ISO 19115-2

SOS 2.0 Hydrology Profile	GEO DAB
sml:input/swe:Quantity/@definition	gmi:LE_ProcessStep/gmd:source/gmi:LE_Source/gmd:description
sml:input/@name	gmi:LE_ProcessStep/gmd:source/gmi:LE_Source/gmd:sourceCitation/gmd:CI_Citation/gmd:title
sml:identification/sml:IdentifierList/sml:identifier/sml:Term/sml:value	gmi:LE_ProcessStep/gmi:processingInformation/gmi:LE_Processing/gmi:identifier
sml:method/sml:ProcessMethod/sml:rules/sml:RulesDefinition/gml:description	gmi:LE_ProcessStep/gmi:processingInformation/gmi:LE_Processing/gmi:procedureDescription
sml:ProcessModel/gml:name	gmi:LE_ProcessStep/gmi:processingInformation/gmi:LE_Processing/gmi:algorithm/gmi:LE_Algorithm/gmi:citatic gmd:title
sml:ProcessModel/gml:description	gmi:LE_ProcessStep/gmi:processingInformation/gmi:LE_Processing/gmi:algorithm/gmi:LE_Algorithm/gmi:descri
sml:output/swe:Quantity/@definition	gmi:LE_ProcessStep/gmi:output/gmi:LE_Source/gmd:description
sml:output/@name	gmi:LE_ProcessStep/gmi:output/gmi:LE_Source/gmd:sourceCitation/gmd:CI_Citation/gmd:title

D. The Pilot

The application scenario is composed of three main steps (depicted in Fig. 5):

i. **Search**: the user searches for available observations in the area of interest and during a specific time extent (Fig. 5 depicts this search performed on the GEO DAB test web client);

ii. **Select**: among the different results, the user selects which time-series (i.e. observations) she/he is interested in comparing (Fig. 5 depicts this selection performed on the GEO DAB test web client);

iii. **Compare**: the user visualizes the previously selected time-series, compares them and makes the needed analysis (Fig. 5 depicts this comparison performed on the SOS client developed for the discipline-specific infrastructures).

(i) Search (ii) Select (iii) Compare

Fig. 5. The Application Scenario Steps Executed by Users

The first step is executed by the GEO DAB; which retrieves the results from the SOS servers content that was previously harvested. The second step takes place on the GEO DAB test web client that displays the possibility to use a Helper Application (i.e. the SOS-client) to visualize the discovered data – users can select one or more results to be visualized. Finally, after selecting all the results of interest users are automatically redirected to the SOS-client. The SOS-client displays the previously selected results showing them in an appropriate way to let users compare and analyze the discovered data.

E. Benefits

The use case pilot demonstrates the following benefits of the enhanced multidisciplinary interoperability:

(1) An eased data provision through the GRDC;
(2) Free and harmonized access to river discharge observations for everyone;
(3) An eased data exchange across institutional boundaries;
(4) An interactive platform for a quick investigation of observed and/or predicted river discharge data for potential end-users.

(1) + (2) The adaption of the GRDC architecture through the deployment of the developed software suites strongly eases the data provision as a handling of each individual data request is no longer necessary. The potential end-user can search and download the data directly via the GEO DAB connected to the GRDC SOS server. Through the registration as GEOSS Data-CORE every user can freely access a subset of the GRDC data sets. The remaining data sets – which are only free for scientific use – are only made available by individual identified requests.

(3) The new architecture of the GRDC and the standardized data formats therefore also eases the data exchange between organizations. The ECMWF predicts river discharge and accesses river discharge observations from the GRDC for calibration and validation purposes through the GEO DAB. When the final model validation is achieved the river discharge predictions are in turn provided through a SOS server.

(4) The web clients enable the user to compare different time series and to load new time series from various data providers. In this example, the end-user is interested in river discharge predictions and its accuracy. The user searches for river discharge data for a specific station through a client application connected to the GEO DAB (e.g. the GEOSS Web Portal[11]). The search results deliver observed and predicted river discharge data. The user firstly selects the predicted river discharge time series. A visualization client is offered that allows a quick overview about the time series. In order to assess the accuracy of the model predictions the user selects the observed time series for the station and plots it into the same graph so that she/he gets a first and quick overview about the predictive power of the model.

V. APPLICABILITY: ENSEMBLE FLOOD PREDICTION IN THE VOLTA RIVER BASIN

To demonstrate the applicability of the developed application scenario, the University of Tokyo applied the TIGGE and GRDC data together with the water and energy budget-based distributed hydrological model WEB-DHM (Wang, et al., 2009) to the Volta River Basin. The WEB-DHM is a distributed-biosphere hydrological model that can simultaneously estimate river discharge, groundwater level and soil moisture at basin scale. It has been applied to various river basins and has shown reliable accuracy in simulations of fluxes, river discharge, and soil moisture. There is a strong need for basin-wide and multi-sectoral coordination of water resources management among flood control, hydro-power and agricultural water use in the Volta River Basin. Especially, the flooding in 2007, historically the fourth largest flood in Ghana, was considered to intensify when water was released from a hydro-electric dam in neighboring Burkina Faso. The 9-day ensemble rainfall forecasts of the five operational models (as described

---

[11] http://www.geoportal.org

in Section II) were used as inputs to the calibrated WEB-DHM. Since the observed river discharge data was not available in the year, the simulated river discharge by inputting the GLDAS data (Global Land Data Assimilation System) was used and compared with the ensemble flood prediction results. Fig. 6 shows reasonable applicability of the ensemble flood prediction for effective flood management, while quantitative analyses and compilation of case studies should be followed.

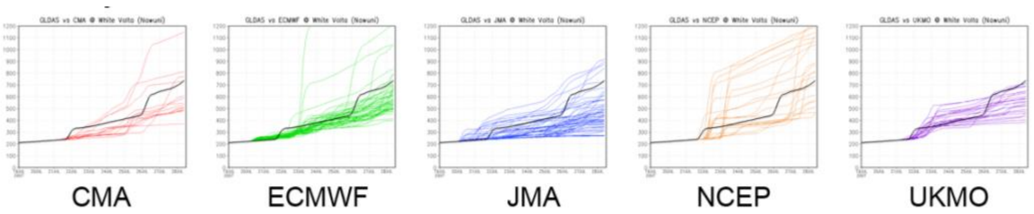


Fig. 6. The river discharge by using the GLDAS forcing (black line) and the 9-day ensemble flood predictions (colored lines) by using the five operational ensemble rainfall forecasts.

The use of the standards and services which were developed within GEOWOW and applied in this scenario can help addressing an important obstacle for real predictive applications: the time lag in the availability of measured data. In fact, these standards and services can be applied for real time sensor data sharing and improve the real time processing of such data into prediction models which have been calibrated with quality assured historical data. Another challenge which the described scenario addresses is the possibility to seamlessly discover and access required data, e.g. GLDAS data, from GEOSS facilitating the development of applications which require the use of data from different sources and domains.

VI. CONCLUSION

Validation and calibration of hydrological prediction models or regional and global analysis of river discharge requires the possibility to exchange river discharge data from different data sources, from different disciplines and institutions. The main technological challenges stem from the diversity of the involved components (i.e. geospatial data sharing infrastructures, services, data models, etc.). In fact, when creating a geospatial data-sharing infrastructure for a given scientific domain, a set of discipline-specific needs must be taken into account and addressed. This has an impact on the adopted technological solutions as far as data, metadata, processing models, services protocols and interfaces, semantics, and embedded knowledge are concerned.

To address such challenges, in the context of the FP7 project GEOWOW (GEOSS Interoperability for Weather, Ocean and Water) a cross-domain application scenario was defined. Different interoperability approaches were taken into account and have been analyzed at the domain-specific as well as the multi-disciplinary level (see Section III). This drove the design of the high-level architecture described in Section IV.A. Domain-specific interoperability challenges were addressed by applying a standardization process, in particular contributing to extend the OGC Sensor Web Enablement (SWE) framework with the definition of a Hydrology Profile of the OGC Sensor Observation Service (SOS) 2.0 standard (see Section IV.B). Multi-disciplinary interoperability was achieved by adopting the broker pattern (Nativi, Craglia, & Pearlman, 2013) and by extending the GEOSS brokering framework to accommodate the new discipline-specific systems that were developed.

The developed pilot was demonstrated at the GEO X Plenary held in Geneva in January 2014.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 282915 (GEOWOW).

REFERENCES

Adikari, Y., & Yoshitani, J. (2009). *Global trends in water-related disasters - An insight for policymakers.* Paris: ICHARM, UNESCO.

Andres, V., Jirka, S., & Utech, M. (2014). *OGC Best Practice: OGC Sensor Observation Service 2.0 Hydrology Profile (OGC 14-004r1).* Wayland, MA, USA: Open Geospatial Consortium Inc.

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A., & Scipal, K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology, 10*, 623-643.

Published Version DOI: 10.1016/j.cageo.2018.03.008

Bojinski, S., Verstraete, M., Peterson, T., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of Essential Climate Variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*. doi:10.1175/BAMS-D-13-00047.1

Boldrini, E., Santoro, M., Papeschi, F., & Nativi, S. (2013). Data Access Services Interoperability in the Geosciences by means of the GI-axe Brokering Framework. *EGU General Assembly 2013.*

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., De Chen, H., . . . Worley, S. (2010). The thorpex interactive grand global ensemble. *Bulletin of the American Meteorological Society, 91*(8), 1059-1072. doi:10.1175/2010BAMS2853.1

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., . . . Lemmens, R. (2011). New Generation Sensor Web Enablement. *Sensors, 11*(3), 2652-2699.

Bröring, A., Stasch, C., & Echterhoff, J. (2012). *OGC Implementation Specification: Sensor Observation Service (SOS) 2.0 (12-006).* Wayland, MA, USA: Open Geospatial Consortium Inc.

Cloke, H., & Pappenberger, F. (2009). Ensemble Flood Forecasting: A review. *Journal of Hydrology*, 613-626.

Craglia, M., Nativi, S., Santoro, M., Vaccari, L., & Fugazza, C. (2011). Inter-disciplinary interoperability for global sustainability research. *Fourth International Conference on Geospatial Semantics*, (pp. 1-15).

Dubois, G., Skoien, J., de Jesus, J., Peedell, S., Hartley, A., Nativi, S., . . . Geller, G. (2011). eHabitat: a Contribution to the Model Web for Habitat Assessments and Ecological Forecasting. *34th International Symposium on Remote Sensing of Environment.*

Group on Earth Observation. (2014). *The GEOSS Water Strategy: From Observations to Decisions.*

Group on Earth Observation. (2014a). *GEOSS Implementation - 2013 Highlights.* Geneve. Retrieved January 6, 2015, from http://www.earthobservations.org/documents/geo_x/geo_x_p_2013_geoss_highlights.pdf

Intergovernmental Panel on Climate Change. (2014). Summary for Policymakers. In *Climate Change 2014: Impacts, Adaptation and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1-32). Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White.

International Organization for Standardization. (2003). *Geographic information — Metadata.*

International Organization for Standardization. (2008). *Geographic information - Metadata - Part 2: Extensions for imagery and gridded data, as sent to the ISO Central Secretariat for publication .*

International Organization for Standardization. (2011). *ISO 19156:2011 - Geographic information -- Observations and measurements.* Geneva, Switzerland: ISO TC 211.

Jamshidi, M. (2011). *System of systems engineering: innovations for the twenty-first century* (Vol. 58). New York: John Wiley & Sons.

Karcanias, N., & Hessami, A. (2010). Complexity and the notion of system of systems: Part (II): Defining the notion of system of systems. *2010 World Automation Congress, WAC 2010.*

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety.* Retrieved 03 23, 2015, from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Mazzetti, P., & Nativi, S. (2012, June). Multidisciplinary Interoperability for Earth Observations: Some Architectural Issues. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5*(3), 1054-1059.

Nativi, S., & Bigagli, L. (2009). Discovery, mediation, and accessservices for earth observation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 233-240.

Nativi, S., Craglia, M., & Pearlman, J. (2013). Earth Science Infrastructures Interoperability: The Brokering Approach. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE*, 1118 - 1129.

Nativi, S., Craglia, M., Vaccari, L., & Santoro, M. (2011). Searching the New Grail: Inter-Disciplinary Interoperability. *Proceedings of the 14th AGILE international conference on Geographic information science.*

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big Data challenges in building the Global Earth Observation System of Systems. *Environmental Modelling and Software*.

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big Data challenges in building the Global Earth Observation System of Systems. *Environmental Modelling & Software*.

Pyster, A., & Olwell , D. (2013). *The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 1.2.* Hoboken, NJ: The Trustees of the Stevens Institute of Technology. Retrieved 09 12, 2014, from http://www.sebokwiki.org/w/downloads/SEBoKv1.2_full.pdf

Santoro, M., Mazzetti, P., Nativi, S., Fugazza, C., Granell, C., & Diaz, L. (2012). Methodologies for augmented discovery of geospatial resources. In L. Diaz, C. Granell, & J. Huerta, *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 172-203). Information Science Reference.

Santoro, M., Nativi, S., & Menard, L. (2013). *GEOSS Common Infrastructure (GCI) Research Engineering Report GEO Architecture Implementation Pilot, Phase 5.*

Taylor, P. (2012). *OGC Implementation Specification: WaterML 2.0: Part 1 - Timeseries (0-126r3).* Wayland, MA, USA: Open Geospatial Consortium Inc.

Published Version DOI: 10.1016/j.cageo.2018.03.008

Vörösmarty, C., Askew, A., Barry, R., Birkett, C., Döll, P., Grabs, W., . . . Webster, F. (2001). Global water data: A newly endangered species. *AGU Eos Transactions*.

Vaccari, L., Craglia, M., Fugazza, C., Nativi, S., & Santoro, M. (2012). Integrative research: the EuroGEOSS experience. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(6), 1603-1611.

Wang, L., Koike, T., Yang, K., Jackson, T., Bindlish, R., & Yang, D. (2009). Development of a distributed biosphere hydrological model and its evaluation with the Southern Great Plains Experiments (SGP97 and SGP99). *Journal of Geophysical Research: Atmospheres, 114*.

World Meteorological Organization. (2010). *IMPLEMENTATION PLAN FOR THE GLOBAL OBSERVING SYSTEM FOR CLIMATE IN SUPPORT OF THE UNFCCC.* Geneva. Retrieved from http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf