Scientific Geodata Infrastructures: Challenges, Approaches and Directions

Lars Bernard, Stephan Mäs, Matthias Müller, Christin Henzen, Johannes Brauner

Technische Universität Dresden, Faculty of Environmental Science, Professorship of Geoinformation Systems

01062 Dresden, Germany

Contact: lars.bernard@tu-dresden.de

This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Digital Earth on 10 Apr 2013, available online: http://www.tandfonline.com/doi/full/10.1080/17538947.2013.781244 .

Abstract

Based on various experiences in developing Geodata Infrastructures (GDI) for scientific applications, this article proposes the concept of a Scientific GDI that can be used by scientists in environmental and earth sciences to share and disseminate their research results and related analysis methods. Scientific GDI is understood as an approach to tackle the science case in Digital Earth and to further enhance e-science for environmental research. Creating Scientific GDI to support the research community in efficiently exchanging data and methods related to the various scientific disciplines forming the basis of environmental studies poses numerous challenges on today's GDI developments. The paper summarizes requirements and recommendations on the publication of scientific geospatial data and on functionalities to be provided in Scientific GDI. Best practices and open issues for governance and policies of a Scientific GDI are discussed and are concluded by deriving a research agenda for the next decade.

Keywords: scientific geodata infrastructure; sharing environmental models; geospatial cyberinfrastructure; e-science; spatial data infrastructure;

1 Introduction

It is a bit more than a decade after the former US vice president (Gore 1999) coined the term Digital Earth to envisage the future use of virtual (digital) globes based on internet and virtual reality technologies. Gore envisioned Digital Earth as an instrument to gain seamless access to various kinds of globally distributed spatio-temporal datasets each covering different parts of the world, having different scales and resolutions and describing the state of the environment and potential environmental threads. Today, a good part of that vision became reality in a number of mostly commercially driven virtual globe applications that we use on a daily basis to virtually explore places (Goodchild 2008). Gore still imagined these applications as being available in only certain institutions, which could provide the required powerful base technology. While the enormous technological progress related to geodata acquisition, computational power, internet protocols and geodata processing allows today for usage of such applications not only on our own desktops but even as 'apps' on various kinds of mobile devices. Infrastructures to share geospatial data from spatially distributed and diverse organizations form the backbone of the Digital Earth, and different types of these infrastructures can be found today.

In (re-)defining a vision for a Digital Earth 2020, Craglia *et al.* (2012) consider the *science case* to frame a prospective and requirements on future geoinformation technologies supporting earth sciences and environmental research. Taking the progress in information technologies and the on-going developments towards Spatial Data Infrastructures (SDI), Geodata Infrastructures (GDI) and the Global Earth Observation System of Systems (GEOSS) as a starting point, Craglia *et al.* (2012) identify the following key challenges for the science case in a Digital Earth 2020:

- linking of multi-disciplinary models to support forecasting and assessing global change(s),
- integration of (near) real time observations taken from the fast emerging pervasive modern sensor networks, including social networks,
- consideration of policy scenarios and their potential impacts,
- communication of scientific findings on global change effects, the related uncertainties and proposed measures to scientists, decision makers and the public, meanwhile providing participatory frameworks supporting stakeholders in sharing their concerns and formulating responses and actions.

Concurring with these aspects and based on own findings in developing GDI for scientific applications, we propose the concept of a Scientific GDI and address related issues. The paper summarizes experiences the authors gained in numerous projects related to GDI implementations for various scientific domains. Special mention deserves the GLUES project (Global Assessment of Land Use Dynamics, Greenhouse Gas Emissions and Ecosystem Services) which is the coordination project of the international research program 'Sustainable Land Management' of the German Ministry of Education and Research. Within GLUES the authors implement a scientific GDI in order to facilitate the interdisciplinary data exchange between scientists of the research program (http://modul-a.nachhaltiges-landmanagement.de/en/scientific-coordination-glues/). Additionally, as a preparation of this article, an investigation of available scientific data infrastructures has been conducted. The survey particularly addressed the provided functionalities and contents, the supported data formats and interfaces, system architectures and licensing. The results are incorporated and affected the overall conclusion of the paper. A summary of the investigated scientific data infrastructures and an approach to classify them can be found here:

http://geoportal.glues.geo.tu-dresden.de/scientificinfrastructures/.

In the following, a quick review on current GDI developments helps identifying bricks and best practices for realizing Scientific GDI. The remainder of the paper summarizes obstacles, requirements and recommendations for the implementation of Scientific GDI and discusses governance and policies to foster Scientific GDI.

2 The Status on Infrastructures for Sharing Geoinformation

The idea of establishing internet-based infrastructures for sharing digital geoinformation arose at least two decades ago: The Clinton Order from 1994 for instance marked the legal start of the US National SDI (Masser 1999). Meanwhile, a number of initiatives started on establishing GDI and developed common specifications and regulations for geoinformation sharing (Figure 1). Activities not only address the technological aspects as the Open Geospatial Consortium (OGC) but also provide organizational frames and partly even define first bricks on common semantics to establish the intended Information Infrastructures. The EU Directive 'Infrastructure for Spatial Information in the European Community' (INSPIRE; (EC 2007a)) is one of Europe's main drivers towards creating an administrative GDI and especially to improve sharing of geoinformation.

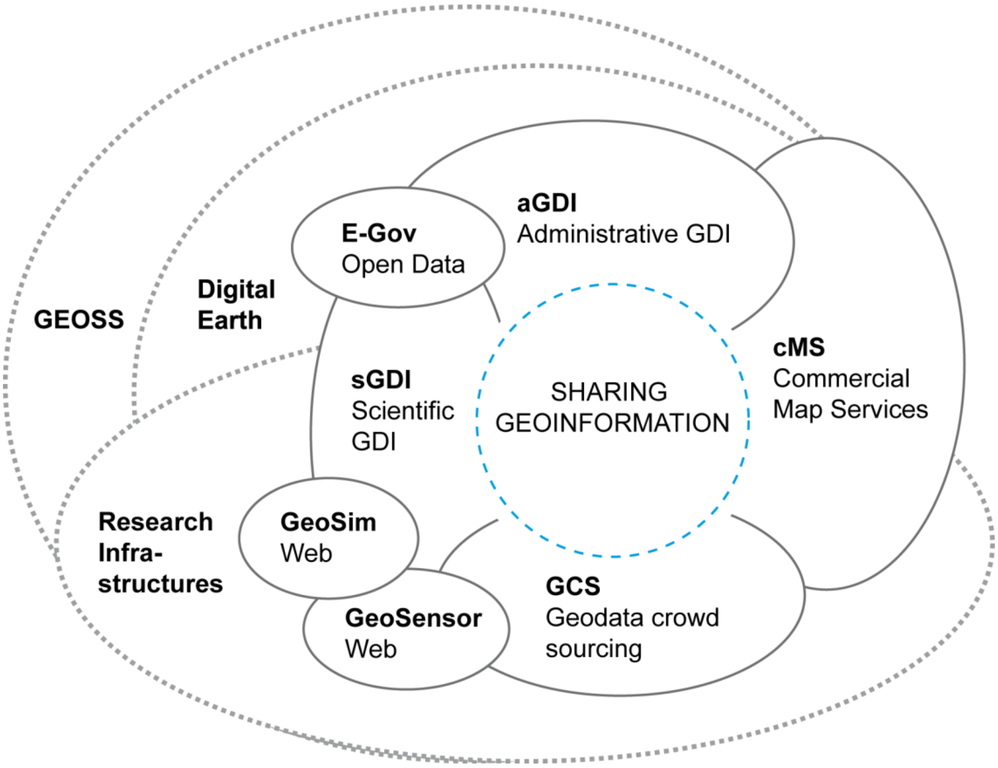


Figure 1. Overview on initiatives and infrastructures for sharing geoinformation

Consequently, a review on INSPIRE related activities provides a starting point on analysing the current state of GDI, their implementation models and their support in sharing geoinformation. A suite of INSPIRE implementing rules and technical guidance documents (see http://inspire.jrc.ec.europa.eu/) provide the reference frame on which standards should be used and on how spatial data should be made technically available to allow for interoperable applications within the EU. Additional to existing standards, the INSPIRE data specifications define harmonied data models for a number of environmental data themes thus also lay the foundation for semantic interoperability and the integration of spatial data from distributed, heterogeneous sources. INSPIRE implementations in the EU member states are on-going and full implementations shall be achieved in 2020. In mainly considering the organizational aspects, a report by De Vries *et al.* (2011) indicates that INSPIRE clearly boosts national GDI implementations in Europe but that EU member states strategies in implementing INSPIRE vary enormously such that a unified GDI implementing model seems hardly to be observed or achieved. Very recently, the INSPIRE directive together with the planned revision of the European Directive on Public Sector Information (EC 2011) encouraged a number of European Member States (e.g. the Netherlands, United Kingdom, Finland, Germany) to at least open the federal official geodata holdings and to provide even topographic data as open data. However, as the real technical INSPIRE implementations are still in their infancy, a valid and thorough report on further effects of INSPIRE cannot be given today.

As an infrastructure, any GDI realization requires to cross-cut administrative and disciplinary boundaries. Additionally, a GDI should follow a service based approach, not only in a technological but also in an organizational sense. Therefore, successfully implementing and operating GDI will most probably necessitate horizontal measures or organizational structures acting as a mediator. The main task of such a mediator - in most cases a new but small institutional entity - is to act highly flexible and to provide all tools and measures that allow an individual public authority to master the tasks being pertinent in implementing GDI.

In terms of functionality, the still emerging first generation of Web-based GDI (GDI 1.0) allows searching for distributed geodata and geoinformation services, interactively visualizing geodata in online available maps, and in best case downloading geodata in a well described format. See for instance Percivall (2010) for a recent overview on related standards. Thus, such GDI 1.0 roughly provide functionality being well comparable to services offered in public libraries: Searching on a set of well-defined attributes, browsing and getting a book (or journal, thesis, etc.) if available. Several Geoportals serve as applications to discover and explore the available geodata and geoinformation services (Bernard *et al.* 2005).

Other driving and even pushing factors for today’s GDI are the commercial Web-based Map Services (e.g. Google Maps, Bing Maps), the geodata crowd sourcing activities (e.g. OpenStreetMap, GeoNames) and the growing number of available (on-line) geosensors providing (near) real time observation of various geo phenomena (Goodchild and Glennon 2010). New combinations of these different information sources are for instance prototyped by the recently launched ‘Eye on Earth’ (http://www.eyeonearth.eu/) and by research on fusion of administrative and crowd sourced geoinformation (Wiemann and Bernard 2010). It still needs to be investigated to what extent volunteered crowd sourcing activities can offer reliable sources for environmental research, not only in (the obvious) terms of data quality but also in what are the ideal topics, scales and time frames to mobilize a sufficiently sized crowd of volunteers to monitor environmental processes or related effects.

The further improvement of distributed geoprocessing (Müller *et al.* 2010) and the usage of cloud computing (Schäffer *et al.* 2010, Yang *et al.* 2011) show how future GDI can be enhanced to offer more powerful analysis capabilities. All these activities show a number of organizational and technological components that could be clearly beneficial in implementing next generation GDI and Digital Earth to seamlessly integrate distributed environmental observations and datasets, environmental modelling systems, and processing and analysis functions. Today’s GDI that primarily focus on geodata provision could develop into a next generation of service infrastructures offering adequate and user-friendly services to derive the relevant piece of geoinformation and thus getting closer to a Digital Earth directly responding users' information requests.

3 Scientific GDI

Traditionally, the evaluation of scientific work mainly refers to published refereed articles and their impacts. Cleary, this is very often not the only outcome of scientific work. Most scientific activities produce data and methods in form of software tools which might be valuable beyond their original scope (Gray 2009). If scientific results, observations, the underlying simulation models and assumptions are sufficiently described and accessible, they can be valuable input for other scientists and other domains of users. Such exchange paired with an improved documentation of research results would make scientific work more transparent and in the optimal case even reproducible. It would allow for the evaluation of fitness for further use (Devillers *et al.* 2007), and thus would ideally increase efficiency and sustainability of research investments and potentially stimulate interdisciplinary research. Simplified and refined representations of the data through Web-based visualizations and visual analysis tools for the comparison of different datasets could support stakeholder work and provide policy makers with insights from scientific research (Bernard and Ostländer 2008). If these various forms of reuse or at least the reuse of the data by other scientists is documented in a way comparable to literature citing, the evaluation of scientific outputs would get an additional measure.

The role of scientific information infrastructures in contributing to innovation and in addressing global challenges has been recognized and addressed in research funding initiatives like NSF (2007) and ESFRI (2008). Related Initiatives in America and Asia often use the label cyberinfrastructure whereas terms as e-science and e-Research Infrastructures have been attached to equivalent activities in Europe. We introduce the term Scientific GDI, as it is felt that current GDI provide a number of useful building blocks to support scientific collaboration and it is envisioned that Scientific GDI could become one of the core components in future Digital Earth. The US National Science Foundation specifically acknowledges the need for Scientific GDI by the Earth Cube program (http://www.nsf.gov/geo/earthcube) to foster research on so-called Geospatial Cyberinfrastructures (Yang *et al.* 2010). In general, scientific information infrastructures have been approached for different science domains and with differing functions, for example the social network myExperiment for the exchange of bioinformatics workflows (Goble *et al.* 2010) or the iPlant Collaborative for plant sciences (http://www.iplantcollaborative.org/). Central functions provided are access to data catalogues and the scientific data, data analysis and visualization capabilities, high performance computing and collaboration. Although such functionality can also be aided by GDI developments, the available scientific infrastructures for the exchange of data from the environmental and earth sciences hardly refer to corresponding interoperability standards. It is only recently that GDI concepts have been taken up by scientific initiatives. A prominent example is the launch of the data service of the Intergovernmental Panel on Climate Change (IPCC; http://www.ipcc-data.org/).

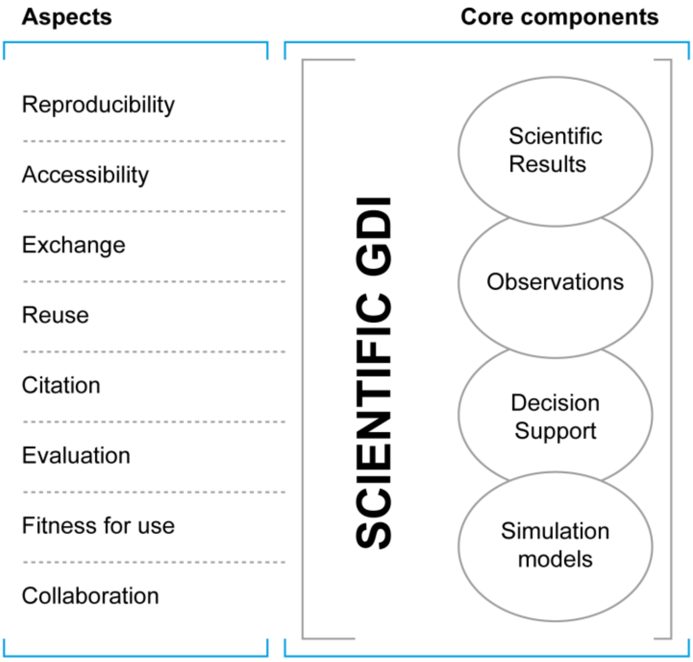


Figure 2. Aspects and core components of Scientific GDI

Creating a Scientific GDI (Figure 2) to support the environmental research community in efficiently exchanging data and methods related to the various scientific disciplines poses a number of diverse challenges on today's GDI developments. In revisiting the science case in Digital Earth (Craglia *et al.* 2012), one could consider an idealized and very simplified researchers workflow to structure these challenges:

(1) discover and access scientific geoinformation resources (data from observations, simulation and analysis as well as analysis methods and models),
(2) integrate, process and analyse scientific geoinformation resources,
(3) publish and share scientific geoinformation resources.

The remainder of the paper follows this sequence to address related issues and to hint to available approaches, best practices and potential solutions.

4 Challenges, Approaches and Recommendations for Scientific GDI

4.1 Discover and Access Scientific Geoinformation Resources

Metadata captures the basic characteristics of a geoinformation resource and ideally does not only support information discovery (like in a library) but also allows assessing the fitness for use and enables the integration with other resources. In current GDI, metadata is usually acquired according to the structure and contents as defined in ISO 19115 and ISO 19119 (ISO 2003, ISO 2005b, ISO 2005a). These standards define more than 300 metadata elements. Therefore, the automatic acquisition of metadata (Manso-Callejo *et al.* 2004, Bockmühl *et al.* 2010, Olfat *et al.* 2012) and further approaches towards user-based metadata enrichment (Olfat *et al.* 2012) are vital issues currently researched.

Typically, geoinformation metadata is managed in catalogues, which follow a quite static perspective on the generation of the underlying geoinformation resources (Fisher *et al.* 2009).

Further, standards lack in representing different levels of detail for geospatial metadata to address different (scientific) user communities.

A main requirement of publishing scientific data is the interlinking between the data and the corresponding scientific publications. This can be easily achieved by enriching metadata with references (cf. DOI in Section 4.3.1) that allows directly navigating to related publications. These publications can provide details about the scientific context, framework requirements of the research, inherent assumptions of the data processing and approaches of data interpretation. These aspects can hardly be fully covered by formalized metadata elements. Even assuming that future metadata sets do always provide these links and that these publications summarize information about the data producing methods, they are usually more focused on new scientific insights and, with regard to metadata provision, they do not sufficiently describe the data and its quality in a structured and comprehensible way. Therefore, links to publications are a part of but will not replace structured metadata.

INSPIRE, for example, requires that for a spatial dataset at least one keyword from the General Environmental Multilingual Thesaurus (GEMET) is provided to describe the relevant spatial data theme (EC 2007b). However, for the required scientific terms GEMET is insufficient. At present, there are other domain-specific vocabularies available or under development (e.g. WMO BUFR for atmospheric conditions (WMO 2010) or the planned GEOSS ontology). However, these are hardly used to unambiguously communicate the meaning across scientific disciplines. Some of these vocabularies reveal inconsistencies in their definitions when they are combined with others, such that vocabularies from different domains are partially incompatible. Beside the connection of keywords to common vocabularies, descriptions of data semantics are hardly considered in current geoinformation metadata standards, which are primarily concerned with the discovery of the data but do not yet enable data integration (Comber *et al.* 2008). Controlled vocabularies for generic scientific terms, like the science ontology developed in Brodaric (2008) and Brodaric *et al.* (2008), hardly exist. Taking the example of environmental modelling, this already starts with basic concepts like model, storyline, scenario, driver and indicator. Although frequently used, different scientific communities have a slightly biased understanding of these terms. Creating a detailed and unambiguous formal description and, particularly, communicating it to a wider audience (at least within single scientific communities) is strongly required and a pressing challenge. Public fora as Wikipedia provide a blueprint on how a common corpus could be developed in a first step. However, additional organizational and research policy measures might be required to establish or even enforce the generation and usage of such common vocabularies.

In particular for climate change and economic development data, storylines and scenarios play a major role in metadata descriptions, as they can be used to classify the data and are certainly the most prominent keywords and search terms. The IPCC defines a storyline as “a narrative description of a scenario (or a family of scenarios), highlighting the main scenario characteristics and dynamics, and the relationships between key driving forces” (Nakićenović *et al.* 2000). A storyline defines qualitative global constraints for projections and leaves a relatively big space for interpretations. Therefore, storylines are substantiated by scenarios, which are defined as “projections of a potential future” (Nakićenović *et al.* 2000): Scenarios provide a “plausible description of how the future may develop, based on a coherent and internally consistent set of assumptions (scenario logic) about key relationships and driving forces (e.g. rate of technology change, prices).” The IPCC defined a set of scenarios that are commonly used by scientists to obtain comparable model outputs when modelling effects of climate change. Therefore metadata of these output data should contain unambiguous links to the scenarios and the corresponding storylines.

4.1.1 Accuracy and Scale

Analysis and description of the quality and in particular the accuracy of the data are very often neglected by scientists. The elements for the description of accuracy and consistency of data defined by ISO 19115 (ISO 2003) are not fitting for scientific model outputs, because such descriptions are usually neither available nor very meaningful. In many cases it would be more helpful to get a description of relationships between the quality of input and output data and a rating or estimation which of the inputs has the biggest impact on the quality of the model outputs. Also, there might be conditions described in the related metadata under which a value can be considered as accurate.

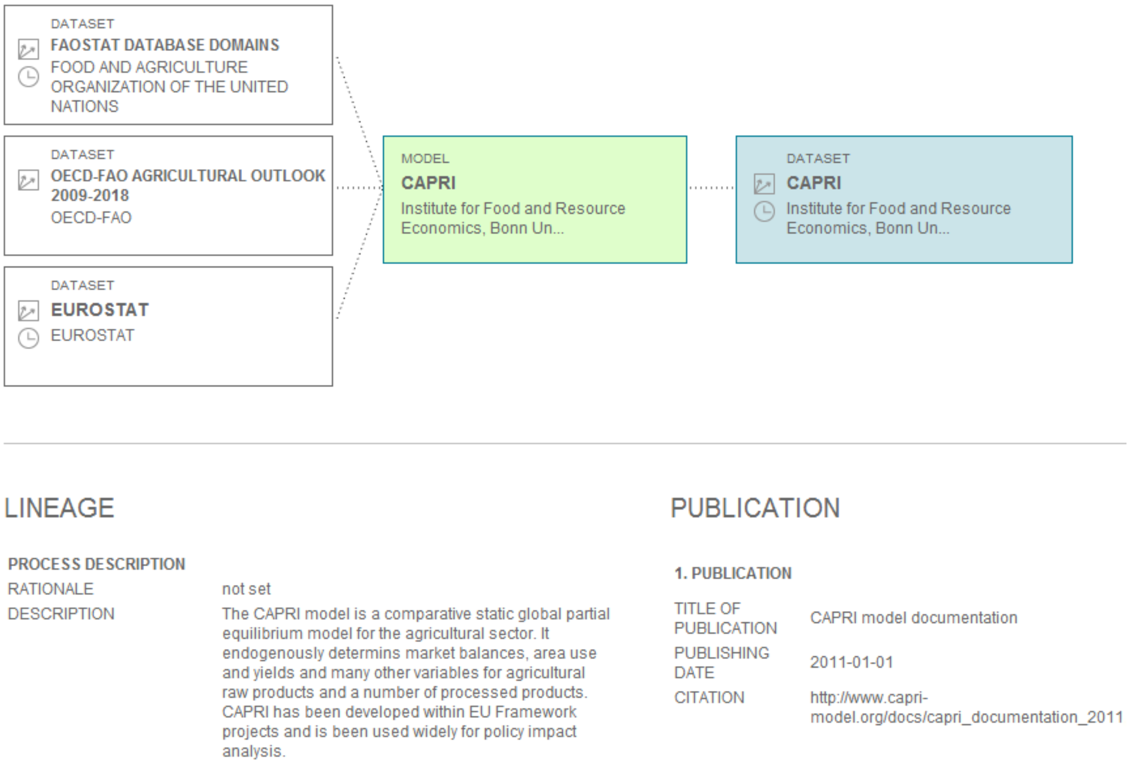
The current ISO metadata standards do not provide sufficient elements to adequately describe the spatio-temporal scale and level of detail of time series or multidimensional data. The emerging revision of ISO 19115 is already addressing this issue by allowing the statement of a temporal resolution. However, input datasets of numerical models very often include statistical data referencing discrete administrative units, like provinces and countries, as spatial resolution. We observed – for instance in economic models – that these units are not separately considered but aggregated to larger, equally sized regions to create a uniform sample size. Depending on modelling goals and the expected outputs, these aggregated spatial regions can be diverse and are task specific. Nevertheless, the aggregation procedure is hardly documented once the data is published. Beside the spatial resolution, also the scale of the geographical phenomena can be diverse. Different objectives of models lead to different thematic categories in the data, like differing terminologies for land cover or agricultural products. For example, a generic class *cereal crops* could also be represented in more detail by *wheat*, *rice*, *corn* and *barley*, which can be considered as the most important subclasses in terms of agricultural production. For a cereal crops dataset, it should be documented if it refers to these four or also includes other 'less important' crops like for example *millet*. Thus, to support related up- and downscaling tasks, the metadata must contain resolvable links to the corresponding sets of spatial aggregation units and thematic categories, as for instance provided by the future INSPIRE implementations.

4.1.2 Lineage and Usage

One of the main purposes of metadata within a GDI is to enable potential users to assess the suitability of the geodata for their specific use. In recent publications, available metadata standards have been recognized as data production oriented and it has been claimed for more user-centric metadata (Devillers *et al.* 2007, Goodchild 2007, Devillers *et al.* 2010). Due to its complexity and in particular the complexity of the data's provenance, this is especially true for scientific data. At present, even if scientific data is discoverable and accessible, the assessment of the data quality with regard to a particular use is difficult.

To evaluate the fitness for use of a dataset, information about its provenance is vital. From the data producer's perspective, lineage information can be used as an internal record of the data to ensure that the production standards are being maintained. In ISO 19115, such information is represented by the lineage element that is modelled as part of the data quality. Lineage recounts the life cycle of the dataset, from real-world abstraction, collection or acquisition, through all stages of compilations, corrections, maintenance, conversions and transformations to the generation of new interpreted products (Clarke and Clark 1995, Harding 2006, Servigne *et al.* 2006). The main components of the lineage in ISO 19115 are subclasses describing sources and process steps when generating the dataset. The sources provide information regarding the source data used to create the described dataset. Process steps are methods and processes used in the creation or maintenance of the dataset.

Since scientific environmental data is often an output of numerical models or simulations, the lineage sub-elements can be used to link to corresponding input datasets and models. If such links are systematically provided by metadata, the relationships between different datasets and models can be visually illustrated. The system MetaViz (Figure 3) has been realized to allow for using and visualizing provenance metadata in the GLUES GDI. Therewith, scientists can for example get a comprehensive view about which models provide datasets for a certain scenario or whether an input dataset also served into other models. Hitherto, gaining such information required a tedious, extensive and time consuming investigation of literature by each researcher who tried to learn about origination of a considered dataset.



DATASET
FAOSTAT DATABASE DOMAINS
FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

DATASET
OECD-FAO AGRICULTURAL OUTLOOK 2009-2018
OECD-FAO

DATASET
EUROSTAT
EUROSTAT

MODEL
CAPRI
Institute for Food and Resource Economics, Bonn Un...

DATASET
CAPRI
Institute for Food and Resource Economics, Bonn Un...

LINEAGE

PROCESS DESCRIPTION

RATIONALE	not set
DESCRIPTION	The CAPRI model is a comparative static global partial equilibrium model for the agricultural sector. It endogenously determins market balances, area use and yields and many other variables for agricultural raw products and a number of processed products. CAPRI has been developed within EU Framework projects and is been used widely for policy impact analysis.

PUBLICATION

1. PUBLICATION

TITLE OF PUBLICATION	CAPRI model documentation
PUBLISHING DATE	2011-01-01
CITATION	http://www.capri-model.org/docs/capri_documentation_2011

Figure 3. MetaViz prototype showing an example for a world economy model. Access via http://geoportal.glues.geo.tu-dresden.de/geoportal/Applications/metaviz.html.

Beside the scientific work, such comparison can also be of interest for research assessment, since it shows the 'impacts' of a dataset. This can also be useful when analysing the scientific outreach, since it represents the data exchange and collaboration between different research institutions. Therefore, lineage information can play a major role for the evaluation of scientific data, comparable to the way citations are used to rank scientific publications. Nevertheless, a detailed description of the origin of scientific data is not possible with the current ISO metadata. In particular the description lacks of details regarding the concrete model initialization, drivers and parameters and a generic description of the model and its basic assumptions.

4.2 Integrate, Process and Analyse Scientific Geoinformation Resources

4.2.1 Software Architectures

Current administrative GDI implementations are mostly realized as *Service Oriented Architectures* (SOA; (OASIS 2006, Erl 2007, OASIS 2011)). In contrast to open source programming libraries, SOA are following the black-box paradigm hiding the concrete functioning and implementation of an algorithm to the service client. Discovery, visualization, access and geoprocessing functions are made available as a set of distinct services. Each is following a well-defined interface to make the specific functions accessible and a service interoperable and easily replaceable by other services as well as combinable in different service orchestrations supporting a wide range of applications. Compared to previous generations of component based approaches or object broker architectures, SOA allow a less tight coupling of the different service components but equally support realizing applications which require certain flexibility and complexity in the usage of processing functionality. Thus, SOA are for instance well suited when dealing with real-time observations, simulations and other data-intensive decision support applications, as for instance in environmental planning or in risk assessment and prevention.

Resource Oriented Architectures (ROA) follow the idea of simply considering anything in the Web a resource which could be easily linked with other resources. They gained widespread uptake in creating Mash Ups as simple information applications (e.g. maps on air quality or a cholera pandemic).

Linked Data follows the ROA path in envisioning a network of Web accessible datasets allowing mutual linkages and ideally referring to ontologies as semantic references (Kuhn 2005). SOA, ROA and Linked Data are addressing different levels of functional complexity and user requirements and would thus require research on their adequate layering (SOA-ROA-Linked Data) to achieve and enable their complimentary usage (Figure 4).

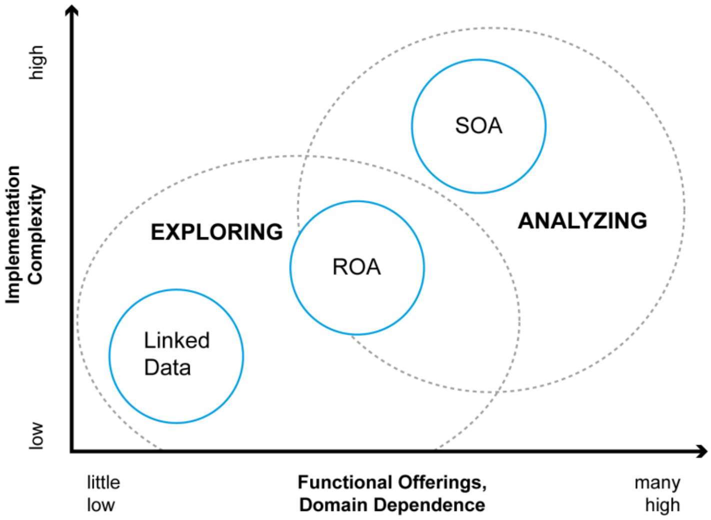


Figure 4. Classification of Service Oriented Architectures, Resource Oriented Architectures and Linked Data.

Thus, ROA and Linked Data are proposed as the future path towards integration of SOA based GDI in non-geo Web domains and fostering their multidisciplinary usage. Metadata publication using Linked Data approaches can be recommended as the first step towards this goal. Clearly, not any observation taken every second or result sets of global change model simulation for 2100 would qualify as a Linked Data item (Janowicz *et al.* 2011), so that a combination of adequate aggregation and generalization mechanisms needs to be designed.

4.2.2 Sharing and Integration of Geoprocessing Methods

Data processing logic in a GDI may be used for change and anomaly detection, to conduct general analysis as well as aggregation, fusion and transformation to convert, compare and integrate data that refers to different spatial, temporal and thematic granularities (ISO 2005b, Craglia *et al.* 2008, Brauner *et al.* 2009, Haubrock *et al.* 2009).

There are two fundamental strategies to deal with such data processing tasks in SDI (Müller *et al.* 2010). In *data-driven* approaches, geodata is sequentially shipped between a network of processing nodes, each offering a certain set of geoprocessing operators or other computational logic. Service chaining techniques as defined by ISO 19119 (ISO 2005b) are usually applied to create larger workflows by connecting data and processing services. Although a loose coupling of independent services is theoretically the most flexible approach, it has some major drawbacks in practical applications. First, purely data-driven workflows are also data-intensive and allocate a lot of network bandwidth. Second, it requires data owners to give *their* data to an unknown third party institution which prohibits to process license-, privacy- or security-constrained data in a federated GDI. Third, from a service provider perspective, offering facilities for computationally intensive jobs is a much greater commitment than the provision of a data download service.

Code-driven approaches pursue the opposite strategy – the code is shipped to the data nodes which is more challenging but offers better performance in terms of bandwidth efficiency, data privacy and provision of computational resources. Shipping code between different instances comes at the cost of the necessity to define exchange mechanisms for algorithms as the definition of common algebras and processing languages. Some of the OGC web services can be extended with basic processing capabilities by supporting standardized processing languages: The Filter Encoding specification (OGC 2010b) is supported by a variety of Web Feature Services (WFS, (ISO 2010)); for Web Coverage Services (WCS, (OGC 2010a)) the Web Coverage Processing Service (WCPS) Language Interface Standard (OGC 2009) plays a similar role. Both standards follow the principle of database query languages that allow their users to perform simple sub-setting and arithmetic tasks on the contained data. Instead of shipping the data to a processing service that performs the intended computations, the client may send a processing query directly to an extended data service where the statements are interpreted and processed. This circumvents the necessity to chain independent data and processing services to accomplish a processing task. A detailed comparison of both approaches was carried out in the OWS-8 testbed (OGC 2011).

Moving and sharing computational code as freely as data in a GDI would be the most flexible solution for code-driven scenarios and might operate equally well as data-driven strategies without suffering from the performance drawbacks. A more general framework for code exchange requires an agreement on common functionality and computing platforms, including software environments and infrastructure. Some research has been done on that matter (Müller *et al.* 2010, Müller *et al.* 2012) which demonstrates the feasibility of code-driven strategies in a GDI but has not yet led to any operational standard or widely used practice. However, the concept was found to link nicely with cloud computing since portability of the software stack is a core asset to achieve computing scalability.

Considering the efforts by individual institutions to foster software reuse, sharing software in larger groups seems to be the next logical step. Code sharing approaches may spur this development and increase the availability of reusable and portable analysis tools, which have been identified as a cornerstone in data-driven science (Gray 2009).

4.2.3 Sharing and Integration of Environmental Simulation Models

Research and developments on linking environmental models from different disciplines, as for example coupling a hydrological model with a climate model, and linking environmental models with GIS to facilitate the pre-processing of model inputs and the analysis of the spatio-temporal results has a long tradition in environmental and geoinformation science (Goodchild *et al.* 1993, Goodchild *et al.* 1996, Bernard and Krüger 2000, Haubrock *et al.* 2009, Maué *et al.* 2011). Roughly three ways of sharing and integrating of environmental models can be distinguished:

(1) *Collective model development* matured in a number of community models. Very equally to open source projects these communities from one or several research institutions normally focus on the realization of a common model for a certain discipline (e.g. the Weather Research & Forecasting Model, http://www.wrf-model.org).

(2) *Sharing model components* with the goal to allow an integrated usage of models from different disciplines is tackled by different approaches: Existing models are coupled through the usage of common interfaces (e.g. following the IEEE standards on High Level Architecture (IEEE 2010) or the Open Modelling Interface http://www.openmi.org/). Integrated models get (newly) developed to synthesize existing models from various disciplines and following different modelling styles (Beven 2007, Voinov 2010).

(3) *Sharing simulation results* allowing scientists from other disciplines to access the results from specific model simulations and scenarios. Results are either stemming from a previous simulation (as for the IPCC data) and should have been undergone the required plausibility checks and validations, or access is given to a pre-parameterized, robust model, which allows for an ad-hoc execution (e.g. an interpolation model to provide a spatial distribution on air pollution, (Stasch *et al.* 2011, Wiemann *et al.* 2012)). In considering simulation models as sensors, the suite of Sensor Web Enablement interfaces provide means to share simulation results in GDI (Maué *et al.* 2011).

These sharing paradigms can be seen as different sharing levels, which could be stacked according to the required need of expertise and the expected users and usages (Figure 5).

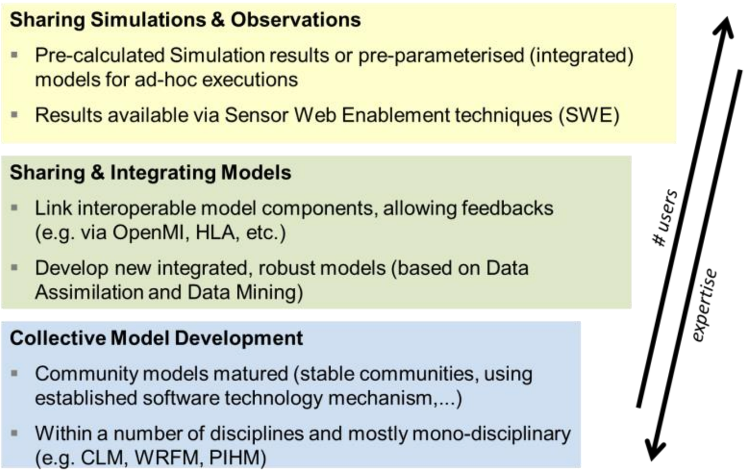


Figure 5. Layers in sharing and integrating environmental models.

GIS and GDI noticeably improved in their capabilities to manage and visualize spatio-temporal (3D) geodata, however, related processing methods (e.g. to calculate a spatio-temporal mean or sum when dealing with time-variant 3D atmospheric data) are still mostly lacking. Advancing these tools and thus better supporting data-mining on environmental simulation results and observations could provide better means to design integrated models. These integrated models are expected to act as the glue between the more complex and demanding models within the environmental disciplines and would ideally become available in a Scientific GDI as robust models to serve tools for other scientific communities and decision makers. As stated above for metadata, handling and communicating the assumptions and uncertainties underlying a model are pressing demands which should be tackled before further encouraging the sharing of environmental models and simulation results. Approaches to consider related metadata in the further processing and analysis chain - as proposed by Pebesma *et al.* (2007), Jung-Hong and Min-Lang (2012) - should be further elaborated and find their way in operational GIS and GDI implementations.

4.3 Publish and Share Scientific Geoinformation Resources

4.3.1 Publishing Scientific Data, Methods and Models

Although some scientific journals offer possibilities to link publications with data or software sources, there are not many incentives for researchers to publish their data or software. Scientific impact factors only consider textual publications and their citations. To better stimulate the publication of scientific data and scientific software they must be properly acknowledged, e.g. by having impact factors for data quotation and by assuring intellectual property rights for provided scientific data and methods. This links to current activities to make publications of scientific data and models obligatory for research work which gains public funding (EC 2012).

Published scientific data should be persistently identifiable and available in a way comparable to current textual publications. A recent open letter of the US National Science Foundation GEO Directorate exemplifies this issue and strongly recommends data citation (http://www.nsf.gov/pubs/2012/nsf12058/nsf12058.jsp). The digital object identifier (DOI) provides a system for identifying content objects in the digital environment. Data DOIs provide persistent identifiers and long term access to data. They help researchers to find and cite datasets, for example to cross reference journal articles with the underlying data. DOI is an ISO standard (ISO 2005a) and already used by numerous platforms for referencing data publications, for example PANGAEA (Data Publisher for Earth and Environmental Science; http://www.pangaea.de/) and the World Data Center for Climate in Hamburg (http://cera-www.dkrz.de/CERA/). Data DOI registration services are provided by the Data Cite organization (http://datacite.org/) which operates globally with national representations. Existing GDI hardly make use of DOI as unique data references. Conversely, data that has a unique DOI reference does not necessarily have standardized metadata or use standardized data formats or geoinformation services. Making DOI a requirement when scientific data is published in a GDI seems to be indispensable. Nevertheless, the systematic capturing of DOI citations and corresponding rewarding of the publishers is still an issue.

The unique identification of data can be implemented in different levels of granularity. INSPIRE, for example, defines persistent object IDs for individual instances. There might be cases which require that single objects can be referenced in such a way. However, the overall efforts to establish and maintain such detailed identification are considered being disproportionate to the benefits in most Scientific GDI. Thus identification on the coarser grained dataset level seems recommendable for using geodata DOIs in Scientific GDI.

4.3.2 *Licensing*

A major concern of scientists publishing their data is about licensing, intellectual property rights and warranty. Possible users must be informed about the concrete license of available data. The Creative Commons (CC) licensing framework provides a set of predefined harmonized and simplified licenses that can be combined to a license contract. CC licenses are commonly used in scientific data infrastructures. Usually, the data ownership stays with the data producing institution or scientist. The expectations of the providers of the scientific data and the infrastructure regarding citation, credits or acknowledgement must be clearly communicated and follow the “good scientific practices”. The data producers should be aware that the selection of restrictive license conditions possibly discourages potential users, for example when a license prohibits commercial use. Recently, the CC community is seeking to adjust its licenses to the specific requirements of scientific data (http://creativecommons.org/science). Scientific GDI must also guarantee the confidentiality of published data to assure the competition between scientific teams. A common strategy is for instance to make only the metadata available and to provide the actual data access later (e.g. after two years), when the data producers have finished corresponding publications (e.g. HALO project database, http://halo-db.tropos.de/).

Another issue is the access to privacy restricted data. Common privacy rules are relatively strict and in some cases they prohibit the provision of data required to explore or analyse detailed relations or interdependencies (e.g. in health or socio-economic studies). In general, GDI still lack the possibility to formalize license and privacy policies, to allow for an as much as possible automated policy negotiation and enforcement process for research applications.

4.3.3 *Governance and Policies*

Most of the current GDI initiatives and implementations are driven by public administrations and result from legal enforcement, like the European INSPIRE initiative. In contrast commercial online mapping applications successfully followed a kind of ‘supply creates demand’ approach (Craglia *et al.* 2012). A mixture of both might stimulate Scientific GDI. Legal measures will hardly work for Scientific GDI, steering needs to relate to the various research funding mechanisms (e.g. simply by making the reuse of existing data and the publishing of results an un-escapable requirement for research funding) and rewarding cultures (citation indices, best data awards, etc.). Having specific Scientific GDI implementations in place, which convince in terms of their data richness, usability and in sum get accepted by a scientific community, further demand can be generated and stimulate further Scientific GDI activities. Initiatives as the CUAHSI Hydrologic Information System (http://his.cuahsi.org), PANGAEA (http://www.pangaea.de/), the EuroGEOSS broker (http://www.eurogeoss-broker.eu/) or the GLUES GDI on sustainable land management and ecosystem research (Eppink *et al.* 2012) could for instance serve as such stimuli on Scientific GDI.

Assuring the sustainable operation and availability of future nodes in the Scientific GDI is another challenge. In general, scientific infrastructures should offer a reliable service with long term perspective and aim at a long term preservation of digital assets. Since they should be sustained beyond the lifetime of single research projects or programmes, the funding must also be sustainable and independent of projects or research programmes and most probably be linked to suited research institutions, which would act as backbone and service centres in Scientific GDI. Current INSPIRE experiences could help in identifying the best suited organizational models. As the INSPIRE directive also enforces specific service level agreements, it could provide guidance for defining comparable criteria in implementing the central nodes of a Scientific GDI. Consequently, Scientific GDI are supposed to follow a system of systems pattern, where such reliable central nodes would serve a wider scientific community in discovering the more specific scientific infrastructures, which could be bound to a specific region or discipline.

5 Conclusion

Similarly to what (De Vries *et al.* 2011) observe about INSPIRE implementations of national GDIs, best practices for the implementation of Scientific GDIs have been identified, but the choice for a certain practice and prioritizing the issues always bounds to the concrete contextual conditions and opportunities. Having broadly considered issues on the realization of Scientific GDI and based on our various experiences in realizing Scientific GDI, we conclude in prioritizing related research items for a Scientific GDI 2020 agenda. In again (as for Section 4) using the three areas (a) discovery and access, (b) integration, processing and analysis, (3) publish and share of scientific geoinformation resources, we propose the following major research activities for this areas:

(1) Discovery and access

a) (Further) establish and possibly enhance common standards and reference systems (e.g. common vocabularies) to improve interoperability especially in terms of semantics.
b) Improve methods for automated metadata creation, extraction, maintenance and provision to allow for better assessment and integration of the provided resources also addressing descriptions on provenance, accuracy, scale and models.
c) Advance usability of existing scientific data portals and services, to attract more users and usage, both in terms of data access and data provision.
d) Provide technical concepts to master the balancing act between scientific data products for scientists and mass-market users.

(2) Integration, processing and analysis

a) Allow for efficiently exchanging analysis methods (processing logic) to better support interdisciplinary and distributed data processing.
b) Further advance methods in multi-scale transformations and cross-disciplinary data integration.
c) Enhance GDI in dealing with time and spatio-temporal information.
d) Progress on having commonly accepted means to exchange and integrate environmental models and the related simulation results within and across the several thematic domains by further aligning the existing partly different developments towards community models, open modelling interfaces and sensor Web.
e) Explore ways to link traditional sensor measurements with crowd sourced observations.

(3) Publish and share

a) Further establish common policies and organizational frameworks for (cross-disciplinary) scientific data infrastructures within the frame of research infrastructures (e.g. http://ec.europa.eu/research/infrastructures), where Scientific GDI would become an integral part.

b) Better stimulate scientific data and scientific software publications by properly acknowledging these kind of publications, e.g. by having impact factors for data quotation and by assuring intellectual property rights for provided scientific data and methods.

c) Progress in formalizing license and privacy policies for research applications to allow for an as much as possible automated policy negotiation and enforcement, as well as to ease analysis of privacy restricted data (e.g. in health or socio-economic studies).

Some of the issues addressed in this agenda have not been covered within this paper but would deserve treatment and further submissions in their own rights. *Improving usability* for instance is regarded to be of utmost importance (1c and 1d in the list above). Balancing functionality and simplicity is a prerequisite to achieve convincing usability and broadly accepted information infrastructures. However current GDI and GIS developments are still felt very weak in terms of usability. Another example is *formalizing license and privacy policies* (3c) as a foundation to allow for distributed geoprocessing. Today, various databases are not accessible for scientific analyses or can only by be used in an off-line mode to ensure data and privacy protection. Thus, for instance research on health-environment effects or socio-economic patterns is not only tedious but partly even hampered as links between different data sources cannot be generated. Beside a need for harmonized policies and organizational frameworks to best enable scientific analysis on such databases, it also lacks commonly agreed formalisms and technical mechanisms. In the end, these technologies should allow for an automated policy negotiation and enforcement and an on-line usage of the restricted database for authorized scientists.

Clearly, the research agenda proposed here closely links to current GEOSS and Digital Earth activities. These are expected to provide the global and cross-disciplinary frame for the emerging more regional or discipline-bounded Scientific GDI advancements.

Acknowledgements

Several findings presented in this paper stem from work on the GLUES GDI as part of the research programme “Sustainable Land Management” funded by the German Federal Ministry of Education and Research (BMBF). The comments and suggestion of two anonymous reviewers and the journal editors helped to improve this submission and are gratefully acknowledged.

6 References

Bernard, L., Kanellopoulos, I., Annoni, A. & Smits, P., 2005. The European Geoportal - One step towards the Establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems,* 29 (1), 15-31 Available from: http://www.sciencedirect.com/science/article/pii/S0198971504000493.

Bernard, L. & Krüger, T., 2000. Integration of GIS and Spatio-Temporal Simulation Models. *Transactions in GIS,* 4 (3), 197-215 Available from: http://onlinelibrary.wiley.com/doi/10.1111/1467-9671.00049/pdf.

Bernard, L. & Ostländer, N., 2008. Assessing climate change vulnerability in the arctic using geographic information services in spatial data infrastructures. *Climatic Change,* (87), 263-281 Available from: http://www.springerlink.com/content/a786301673025334/fulltext.pdf.

Beven, K., 2007. Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. *Hydrol. Earth Syst. Sci.,* 11 (1), 460-467 Available from: http://www.hydrol-earth-syst-sci.net/11/460/2007/.

Bockmühl, T., Mäs, S. & Reinhardt, W., 2010. An Approach for Automatic Generation of Metadata based on Data Product Specifications. *INSPIRE Conference 2010,* Krakow, Poland.

Brauner, J., Foerster, T., Schaeffer, B. & Baranski, B., 2009. Towards a Research Agenda for Geoprocessing Services. *In*: Seester, M., Bernard, L. & Paehlke, V., eds. *Proceedings of the 12th AGILE Conference*, Hannover.

Brodaric, B., 2008. *Science Knowledge Infrastructure Ontology 3.0* [online]. University of Edinburgh. Available from: http://www.nesc.ac.uk/technical_papers/UKeS-2008-03.pdf [Accessed May 2012].

Brodaric, B., Reitsma, F. & Qiang, Y., 2008. SKIing with DOLCE: toward an e-Science Knowledge Infrastructure. *In*: Eschenbach, C. & Grüninger, M., eds. *Fifth International Conference on Formal Ontology in Information Systems (FOIS 2008)*, Amsterdam, The Netherlands: IOS Press, 208-219.

Clarke, D.G. & Clark, D.M., 1995. Lineage. *In* Guptill, S.C. & Morrison, J.L. eds. *Elements of Spatial Data Quality*. Oxford Elsevier Science, 13-30.

Comber, A.J., Fisher, P.F. & Wadsworth, R.A., 2008. Semantics, Metadata, Geographical Information and Users. *Transactions in GIS*, 12, 287-291.

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., Annoni, A., Ling, B., Campbell, F., Ehlers, M., Genderen, J.v., Goodchild, M., Guo, H., Lewis, A., Simpson, R., Skidmore, A. & Woodgate, P., 2012. Digital Earth 2020: towards the vision for the next decade. *International Journal of Digital Earth,* 5 (1), 4-21.

Craglia, M., Goodchild, M.F., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S. & E, P., 2008. Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. *International Journal of Spatial Data Infrastructure Research*, 3, 146-167.

De Vries, W., Crompvoets, J., Stoter, J. & Vanden Berghe, I., 2011. ATLAS of INSPIRE – conceptualizing SDI implementation through an inventory of experiences, successes and headaches of European national mapping agencies. *International Journal of Spatial Data Infrastructures Research,* 6, 97-125.

Devillers, R., Bedard, Y., Jeansoulin, R. & Moulin, B., 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science,* 21 (3), 261-282.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P. & Shi, W., 2010. 30 years of research on Spatial Data Quality: Achievements, failures, and opportunities. *Transactions in GIS,* 14 (4), 387-400.

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) 2007a.

Implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, (Commission Regulation (EC) No 1205/2008 of 3 December 2008) 2007b.

EC, 2011. *MEMO/11/891: Digital Agenda: Commission's Open Data Strategy, Questions & answers* [online]. European Commission. Available from: http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891 [Accessed July 2012].

EC, 2012. *Scientific data: open access to research results will boost Europe's innovation capacity* [online]. European Commission. Available from: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/790 [Accessed July 2012].

Eppink, F., Werntze, A., Mäs, S., Popp, A. & Seppelt, R., 2012. Land Management and Ecosystem Services: How Collaborative Research Programmes Can Support Better Policies. *GAIA - Ecological Perspectives for Science and Society,* 21 (1), 55-63.

Erl, T., 2007. *SOA Principles of Service Design*: Prentice Hall International.

ESFRI, 2008. *European roadmap for research infrastructures* [online]. European Strategy Forum on Research Infrastructures. Available from: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_update_2008.pdf [Accessed June 11 2012].

Fisher, P., Comber, A.J. & Wadsworth, R., 2009. What's in a Name? Semantics, Standards and Data Quality *In* Devillers, R. & Goodchild, H. eds. *Spatial Data Quality: Form Process to Decisions.* CRC Press, 3-16.

Goble, C.A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P. & De Roure, D., 2010. myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research,* 38, W677-W682.

Goodchild, M.F., 2007. Beyond Metadata: Towards User-Centric Description of Data Quality. *5th International Symposium on Spatial Data Quality*, ITC, Enschede, The Netherlands.

Goodchild, M.F., 2008. The use cases of digital earth. *International Journal of Digital Earth,* 1 (1), 31-42.

Goodchild, M.F. & Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3 (3), 231- 241 Available from: http://dx.doi.org/10.1080/17538941003759255.

Goodchild, M.F., Parks, B.O. & Steyaert, L.T. eds. 1993. *Environmental Modeling with GIS*, New York: Oxford University Press.

Goodchild, M.F., Steyaert, L.T. & Parks, B.O. eds. 1996. *GIS and Environmental Modeling: Progress and Research Issues,* Fort Collins: GIS World Books.

Gore, A., 1999. The Digital Earth: Understanding Our Planet in the 21st Century (Speech held on January 31, 1998 in the California Science Center). *Photogrammetric Engineering & Remote Sensing,* 65 (5), 528-530.

Gray, J., 2009. eScience: a transformed scientific method. *In* Hey, A.J.G., Tansley, S. & Tolle, K.M. eds. *The fourth paradigm: data-intensive scientific discovery.* Redmond, WA: Microsoft Research, xvii-xxxi-xvii-xxxi.

Harding, J.D., 2006. Vector Data Quality: A Data Povider's Perspective. *In* Devillers, R. & Jeansoulin, R. eds. *Fundamentals of Spatial Data Quality.* London: ISTE Ltd., 141-159.

Haubrock, S., Theisselmann, F., Rotzoll, H. & Dransch, D., 2009. Web-based management of simulation models - concepts, technologies and the users' needs. *In*: Anderssen, R.S., Braddock, R.D. & Newham, L.T.H., eds. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, Cairns, Australia, 880-886.

IEEE, 2010. *1516 IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) Framework and Rules* New York, USA, Institute of Electrical and Electronics Engineers.

ISO 19115 International Standard on Geographic information - Metadata 2003.

ISO 19115-2 International Standard on Geographic information - Part 2: metadata for imagery and gridded data 2005a.

ISO 19119 International Standard Geogaphic information Services 2005b.

ISO 19142 Geographic information - Web Feature Service 2010.

Janowicz, K., Bröring, A., Stasch, C., Schade, S., Everding, T. & Llaves, A., 2011. A RESTful Proxy and Data Model for Linked Sensor Data. *International Journal of Digital Earth*, in press.

Jung-Hong, H. & Min-Lang, H., 2012. Interoperable GIS Operations: A Quality-Aware Perspective. *In*: Gensel, J., Josselin, D. & Vandenbroucke, D., eds. *15th AGILE International Conference on Geographic Information Science*, Avignon, France.

Kuhn, W., 2005. Geospatial Semantics: Why, of What, and How? *In* Spaccapietra, S. & Zimányi, E. eds. *Journal on Data Semantics III.* Berlin/Heidelberg: Springer 1-24.

Manso-Callejo, M.A., Nogueras-Iso, J., Zarazaga, J. & Bernabe-Poveda, M.A., 2004. Automatic Metadata Extraction from Geographic Information. . *In*: Toppen, F. & Prastacos, P., eds. *7th AGILE Conference on Geographic Information Science* Heraklion, Greece: AGILE, 379-386.

Masser, I., 1999. All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science,* 13 (1), 67-84.

Maué, P., Stasch, C., Athanasopoulos, G. & Gerharz, L., 2011. Geospatial Standards for Web-enabled Environmental Models. *International Journal of Spatial Data Infrastructures Research* 6,145-167 Available from: http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/198/291.

Müller, M., Bernard, L. & Brauner, J., 2010. Moving Code in Spatial Data Infrastructures – Web Service Based Deployment of Geoprocessing Algorithms. *Transactions in GIS,* 14 (S1), 101-118 Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9671.2010.01205.x/pdf.

Müller, M., Kadner, D. & Bernard, L., 2012. Moving Code - Sharing Geospatial Computation Logic on the Web. *In* Gensel, J., Josselin, D. & Vandenbroucke, D. eds. *15th AGILE International Conference on Geographic Information Science.* Avignon, 58-62.

Nakićenović, N., Alcamo, J. & Davis, G., 2000. *IPCC Special Report on Emissions Scenarios (SRES)* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

NSF, 2007. *Cyberinfrastructure vision for 21st century discovery* [online]. US National Science Foundation Cyberinfrastructure Council. Available from: http://www.nsf.gov/od/oci/CI_Vision_March07.pdf [Accessed June 2012].

OASIS, 2006. *Reference Model for Service Oriented Architecture 1.0 OASIS Standard, 12 October 2006*.

OASIS, 2011. *Reference Architecture Foundation for Service Oriented Architecture Version 1.0. 06 July 2011. OASIS Committee Specification Draft 03.* .

OGC, 2009. *OpenGIS Web Coverage Processing Service (WCPS) Language Interface Standard* [online]. Open Geospatial Consortium. Available from: http://www.opengeospatial.org/standards/wcps

OGC, 2010a. *OGC WCS 2.0 Interface Standard - Core* [online]. Open Geospatial Consortium. Available from: http://www.opengeospatial.org/standards/wcs/

OGC, 2010b. *OpenGIS Filter Encoding 2.0 Encoding Standard* [online]. Open Geospatial Consortium. Available from: http://www.opengeospatial.org/standards/filter

OGC, 2011. *OWS-8 Geoprocessing for Earth Observations Engineering Report* [online]. Open Geospatial Consortium. Available from: http://www.opengeospatial.org/pub/www/ows8/deliverables.html

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H. & Williamson, I.P., 2012. Spatial Metadata Automation: A Key to Spatially Enabling Platform. *International Journal of Spatial Data Infrastructures Research,* 7, 173-195.

Pebesma, E.J., de Jong, K. & Briggs, D., 2007. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. *International Journal of Geographical Information Science,* 21 (5), 515-527 Available from: http://dx.doi.org/10.1080/13658810601064009.

Percivall, G., 2010. The application of open standards to enhance the interoperability of geoscience information. *International Journal of Digital Earth,* 3 (S1), 14-30.

Schäffer, B., Baranski, B. & Foerster, T., 2010. Towards Spatial Data Infrastructures in the Clouds. *In* Painho, M., Santos, M. & Pundt, H. eds. *Geospatial Thinking - 13th AGILE International Conference on Geographic Information Science.* Berlin: Springer, 399-418.

Servigne, S., Lesage, N. & Libourel, T., 2006. Quality Components, Standards, and Metadata. *In* Devillers, R. & Jeansoulin, R. eds. *Fundamentals of Spatial Data Quality*. London: ISTE Ltd., 179-210.

Stasch, C., Foerster, T., Autermann, C. & Pebesma, E., 2011. Spatio-temporal aggregation of European air quality observations in the Sensor Web. *Computers & Geosciences*, in press Available from: http://www.sciencedirect.com/science/article/pii/S0098300411003736.

Voinov, A.A., 2010. 'Integronsters' and the special role of data. *In*: Swayne, D.A., Yang, W., Voinov, A.A., Rizzoli, A. & Filatova, T., eds. *iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2010). International Environmental Modelling and Software Society*, Ottawa, Ontario, Canada S.10.19.

Wiemann, S. & Bernard, L., 2010. Conflation Services within Spatial Data Infrastructures. *In* Painho, M., Santos, M.Y. & Pundt, H. eds. *13th AGILE International Conference on Geographic Information Science.* Guimarães, Portugal, 1-8.

Wiemann, S., S., R., Karrascha, P., Brauner, J., Pech, K. & Bernard, L., 2012. Classification-driven air pollution mapping as for environment and health analysis. *In* Seppelt, R., Voinov, A.A., Lange, S. & Bankamp, D. eds. *iEMSs Sixth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2012). International Environmental Modelling and Software Society.*

WMO, 2010. *Manual on Codes (WMO-No. 306).* , World Meteorological Organization.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M. & Fay, D., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, 4 (4), 305-329.

Yang, C., Raskin, R., Goodchild, M. & Gahegan, M., 2010. Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems,* 34 (4), 264-277.