Review

Geospatial Cyberinfrastructure and Geoprocessing Web—A Review of Commonalities and Differences of E-Science Approaches

Barbara Hofer

Interfaculty Department of Geoinformatics—Z_GIS, University of Salzburg, Hellbrunnerstraße 34, A-5020 Salzburg, Austria; E-Mail: barbara.hofer@sbg.ac.at; Tel.: +43-662-8044-7507; Fax: +43-662-8044-525

Abstract: Online geoprocessing gains momentum through increased online data repositories, web service infrastructures, online modeling capabilities and the required online computational resources. Advantages of online geoprocessing include reuse of data and services, extended collaboration possibilities among scientists, and efficiency thanks to distributed computing facilities. In the field of Geographic Information Science (GIScience), two recent approaches exist that have the goal of supporting science in online environments: the geospatial cyberinfrastructure and the geoprocessing web. Due to its historical development, the geospatial cyberinfrastructure has strengths related to the technologies required for data storage and processing. The geoprocessing web focuses on providing components for model development and sharing. These components shall allow expert users to develop, execute and document geoprocessing workflows in online environments. Despite this difference in the emphasis of the two approaches, the objectives, concepts and technologies they use overlap. This paper provides a review of the definitions and representative implementations of the two approaches. The provided overview clarifies which aspects of e-Science are highlighted in approaches differentiated in the geographic information domain. The discussion of the two approaches leads to the conclusion that synergies in research on e-Science environments shall be extended. Full-fledged e-Science environments will require the integration of approaches with different strengths.

Keywords: geospatial cyberinfrastructure; geoprocessing web; e-Science

1. Introduction

The amount of spatial data grows substantially day-by-day through the activities of sensing technology like satellites. For example, the Earth Observing System Data and Information System of the North American Space Agency (NASA) has more than six petabytes of data stored by today [1]. The collected data are increasingly offered in online data repositories. This trend is supported by initiatives like INSPIRE—the Infrastructure for Spatial Information in the European Community [2] and the Global Earth Observation System of Systems (GEOSS, [3]). The increasing number of facilities for discovery and access to data opens new research frontiers regarding the analysis of these data.

The focus of spatial data infrastructures (SDIs) is still on providing data rather than analysis functionalities. Yue *et al*. ([4], p. 274) talk about a "data-rich yet analysis-poor period". Nevertheless, the access to data is complemented more and more by analysis functionality and tools for online geoprocessing. Geoprocessing means the application of analysis functionality to input data in order to generate transformed output data respectively information. Interoperable web services gained importance for pursuing geoprocessing online. For example, Dadi and Di [5] give an overview over the provision of GRASS GIS commands as services for constructing and executing analysis workflows online. The usage of online geoprocessing for constructing scientific analyses through experts is only a small fraction of online geoprocessing applications. Other use cases of online geoprocessing include the provision of predefined functionality to a larger user group, the development of specific services for automating repetitive tasks, and the usage of powerful distributed processing infrastructures.

Besides the access to data and analysis functionality, an additional challenge of today's researchers is to work across disciplines [6,7]. New information can be generated when thinking across boundaries of a domain and, for example, extending existing models with data from related disciplines. Craglia *et al*. [6] state that multi-disciplinary interoperability is what comes after addressing syntactic interoperability in spatial data infrastructures. To date, GEOSS and SDIs lack a framework for supporting the collaboration and exchange among natural and social scientists, policy makers, decision makers and the public [8]. This exchange is foreseen in the goals of a Digital Earth [8], which is a vision coined by Al Gore [9]. An infrastructure for sharing data, services and models is the basis for increasing the collaboration among researchers within or across disciplines. Fook *et al*. ([10], p. 379) say with a focus on the discipline of biodiversity:

> "*To improve biodiversity science, scientists need to share models, data and results, and should be able to reproduce experiments from others.*"

Conducting science in online environments, *i.e.*, conducting e-Science, has the potential to make research real-time, efficient, and cost-effective as well as supporting exchange and collaboration among scientists. E-Science environments are supposed to cover data capture, pre-processing, analysis and visualization [11]. The realization of an e-Science environment supporting data analysis and collaborative research requires the integration of diverse technologies and non-technological components.

The usage of the term e-Science in this article refers to the general principle of pursuing science in online, distributed, and collaborative environments. In literature, the term e-Science is also used for denoting the European equivalent to cyberinfrastructure [12]. The term cyberinfrastructure refers to

online processing infrastructures and has an American imprint. Stewart *et al*. [12] do not comply with this equalization of (American) cyberinfrastructure and (European) e-Science. They say that

> "*… e-Science has a sense of being more about cyber-enabled science and somewhat less about the underlying infrastructure*" ([12], p. 42).

The corresponding term to cyberinfrastructure would be e-Infrastructure [12,13]. In this article, we follow the understanding of Stewart *et al*. [12] and use the term e-Science to refer to advancing science in online environments.

The article looks at e-Science environments from the perspective of researchers, who want to use online geoprocessing for their research tasks. Researchers who are interested in online geoprocessing tool-sets need to invest considerable resources for gaining an overview over existing developments. Different names given to online geoprocessing environments introduce a separation of approaches. The main contribution of this review article is an overview over the common elements, similarities and differences of e-Science environments for GIScience applications. The overview provided in the article clarifies which aspects of e-Science are highlighted in approaches differentiated in the geographic information domain.

The focus of the review is on approaches that provide a full suite of tools for performing e-Science: the geospatial cyberinfrastructure [14] and the geoprocessing web [15]. The geospatial cyberinfrastructure has its roots in high performance computing technology. Its role is frequently reduced to the provision of an infrastructure for data storage and computation. The geoprocessing web was designed for addressing user requirements such as collaboration, exchange and communication in online geospatial analysis tasks [15]. The definitions of the two approaches show, however, that the argument of improving collaboration among scientists is used in both approaches. Additionally, spatial data infrastructures, the grid and the cloud, service-oriented architectures, and other technologies are named as elements of the approaches. What are then the differences of the two approaches? Which approach is suitable for which tasks scientists have? There are differences in the technologies used: the geoprocessing web is solely based on web technologies and avails of open web standards, whereas this restriction does not hold for geospatial cyberinfrastructures. Additionally, the geoprocessing web highlights modeling requirements of users and aims at providing workflow tools and model warehouses. Nevertheless, the geospatial cyberinfrastructure and the geoprocessing web are both realizations of e-Science environments. The author suggests finding synergies of the two approaches towards e-Science environments. Eventually, the integration of approaches with different strengths will lead to fully-fledged e-Science environments.
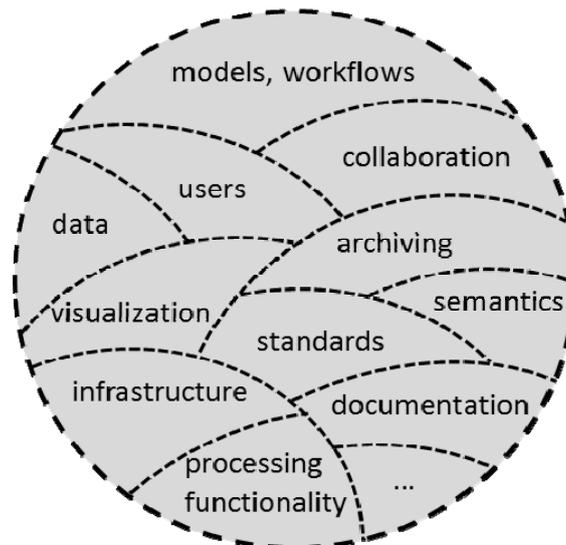
The review starts with an overview over enabling technologies for e-Science in the geographic information domain together with their development (Section 2). The review of the approaches is contained in Section 3 (geospatial cyberinfrastructure) and Section 4 (geoprocessing web). Section 5 discusses the user interaction, functions and computing infrastructure of the two approaches in general and with reference to representative examples. Section 6 concludes the article.

2. E-Science and the Geographic Information Domain

E-Science environments provide resources for pursuing science and for supporting collaboration among researchers. To make an e-Science environment possible, the combination of different

technologies as well as non-technological components is required. The combination of these components causes the complexity of these initiatives [14]. Figure 1 gives an overview of the conglomerate of components forming e-Science environments. The components fulfill different functions, that all contribute to the overarching goal of a cross-disciplinary collaborative environment for pursuing science. Approaches to e-Science may focus on different components and therefore realize different sets of components in specific e-Science environments.

Figure 1. E-Science environments: a conglomerate of components.

E-Science concepts and developments are employed in a series of disciplines [16]. E-Science environments related to disciplines like geography, Earth observation, biodiversity, oceanography, climatology, *etc.* include geospatial concepts and functionality. A series of terms refer to the usage of spatial relations for ordering information, and for the provision and documentation of online geoprocessing functionality: geospatial web [17], semantic geospatial web [18], geospatial cyberinfrastructure [14], geoprocessing web [15,19] and scientific geodata infrastructure [20]. The (semantic) geospatial web aims at developing structures for linking spatial information to data available over the World Wide Web. Semantics play an important role in this context as they are needed for improving search, retrieval and integration of data. The geospatial cyberinfrastructure offers data storage, computing power and analysis functionality to scientists for supporting knowledge generation [14]. The geoprocessing web focuses on collaboration, exchange and communication in online geospatial analysis tasks [15]. The recently introduced scientific geodata infrastructure emphasizes the documentation and exchange of data and methods used in scientific work.

The focus of this article is on the approaches geospatial cyberinfrastructure (GCI) and geoprocessing web that are providing a suite of geoprocessing functionality. Aspects of semantics, data provision and documentation of analyses and research as highlighted in the semantic geospatial web, spatial data infrastructures respectively scientific geodata infrastructures are not further discussed.

Online geoprocessing environments fulfill different functions. The main objective is to support their users at deriving information or knowledge from data. The services they provide include the provision of resources for processing large datasets, the automatizing of processing tasks, the reuse of algorithms, the modeling and documentation of workflows. Among others, the following advantages of online

geoprocessing environments have been documented: the integration of heterogeneous data and services [21]; faster update cycles [22]; the use of open-source software to make systems cost-effective [22]; grid technology and cloud technology supporting the development of low-cost, scalable and efficient systems [23,24]; and the reuse of existing services for reducing development time [25].

E-Science environments have to address a series of challenges: search and discovery of data and services, heterogeneity of resources and terminology, lineage or provenance, processing speed, scaling of infrastructures, visualization of results, sharing of models, as well as security and privacy issues. The processing of spatial data in e-Science environments poses additional or intensified challenges, e.g., workflows involving spatial data are data and computing intensive [26]; input data need to have appropriate resolution and quality for producing meaningful results [27]; spatial data are syntactically and semantically heterogeneous. These additional challenges in online *geoprocessing* environments are caused by the multidimensionality of spatial data [14].

The challenges of online geoprocessing and in particular of geoprocessing services led to the identification of three research topics [28]: service orchestration, semantic descriptions and performance improvement. Service orchestration refers to the combination of services for producing value-added processing chains of spatial data [29]. The automatizing of service orchestration requires semantic descriptions of services, which may be supported by ontologies or approaches using the semantic Business Process Execution Language (BPEL). Once a service chain has been established, the performance of its execution is essential. Grid computing and cloud computing are technologies that can support the execution of geoprocessing services chains.
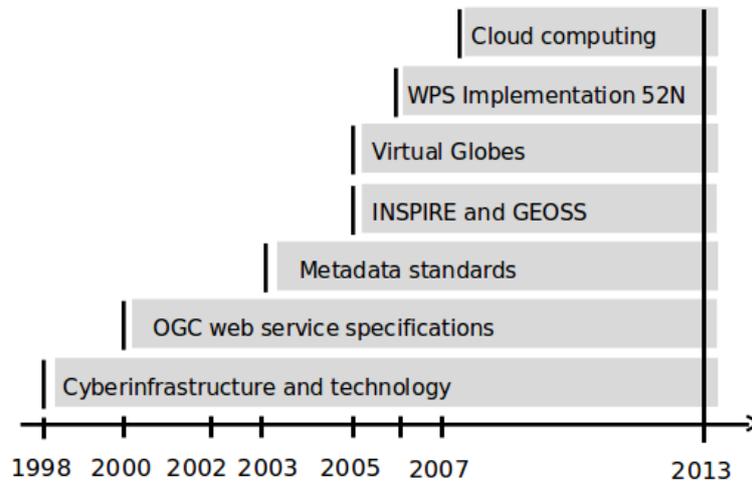
Figure 2 shows a time-line of the development of concepts and technologies that enable e-Science in the geospatial domain. These concepts and technologies include:

- standards for metadata, e.g., International Organization for Standardization (ISO) 19115 and 19119,
- standards and interface specification of data and processing web services as well as sensor observation services, e.g., Open Geospatial Consortium (OGC), World Wide Web Consortium (W3C) and ISO standards for web services,
- online visualization tools and virtual globes: Google Earth, Bing maps, *etc.*,
- spatial data infrastructures, e.g., INSPIRE and GEOSS,
- clients for geoprocessing, e.g., 52° North Web Processing Service implementation and GeOnAS (discussed in Section 5.1),
- technological developments, e.g., grid computing and cloud computing.

As Figure 2 indicates, the key components for e-Science environments are all established by 2013. Nevertheless, developments in the e-Science arena are not as widely used as expected [30]. Geographic information systems (GIS) are continuing to have a wide usage for (desktop) geospatial analyses. The transfer from a desktop-based working style to a service-based working style is not automatic. Poore ([31], p. 2) says:

> "*...it cannot be assumed that just because tools are provided they will be used, or that the affordances of new technology for collaboration will inevitably lead to the dismantling of the isolated single investigator hypothesis-driven model of science*".

Figure 2. Time-line of developments related to e-Science in the geographic information domain.

Researchers entering the arena of web-based geoprocessing and e-Science in general are faced with a series of terms, technologies and concepts. They need to classify approaches and get an overview over which tools to use when. Regarding e-Science for geospatial analyses, new terms keep being introduced, without necessarily stating their relation to other developments in e-Science. The review of the concepts of the geospatial cyberinfrastructure and the geoprocessing web provides an overview over terms in that field. The intention of the review is to provide a clarification of developments and their functions for researchers entering the world of e-Science environments.

3. The Geospatial Cyberinfrastructure

The geospatial cyberinfrastructure builds on the definition of cyberinfrastructure:

"Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible" ([12], p. 37).

In short, this definition focuses

"on the general function of a system of technology, direct involvement of people and innovation as an outcome ..." ([12], p. 42).

The general term cyberinfrastructure developed from the ideas of computational grids [12]. Grids aim at providing scientists with computing power for pursuing intensive computations. A definition given by Foster and Kesselman [32] as cited in [12] (p. 37):

"A grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities".

Nowadays cloud computing gains importance for providing scalable computing infrastructures. Cloud computing refers to distributed computing in a network that can be requested on demand. Infrastructure, platform, software and data are provided as services in cloud computing [24,33].

Stewart *et al*. [12] describe different instances of cyberinfrastructures. One usage of the term cyberinfrastructure refers to providing a scientist with access to computing power and remote control over the execution of computing jobs. Another usage of the term refers to using a cyberinfrastructure for generating information that is readily usable by researchers for their purposes. The specific example given is researchers, who require information from a discipline that is not their home discipline. The cyberinfrastructure handles the processing of data and returns the information to the users, without them having to analyze data they are not familiar with.

> "*The use of the term cyberinfrastructure here means the middleware and layers of tools that sit on top of the computing systems, data storage systems, and computer networks ...*" ([12], p. 4).

Cyberinfrastructure can be understood as science gateway in that context [12].

In the GIScience domain, two terms are based on the principles of cyberinfrastructure: distributed geographic information processing (DGIP) and geospatial cyberinfrastructure. DGIP was introduced by Yang and Raskin in 2009 [34].

> "*DGIP focuses on the technical research on how to allocate and process geographic information resources in a distributed environment to achieve a specific application objective (such as the implementation of virtual globes)*" ([34], p. 553).

Yang and Raskin [34] specify a research agenda for DGIP that covers the full range of topics from an infrastructure for spatial computing, via service oriented architectures, to interoperability, models, semantics and application sciences. The last point on application sciences includes topics like collaboration, human computer interaction, decision making, as well as the development of applications for different domains.

Zhang and Tsou ([23], p. 605) define

> "*geospatial cyberinfrastructure as a combination of distributed geographic information processing technology ..., high-performance computing (HPC) resources, interoperable Web services, and sharable geographic knowledge to facilitate the advancement of geographic information science (GIScience), geospatial technology, and geographic education*".

Yang *et al*. ([14], p. 265) see GCI as

> "*infrastructure that supports the collection, management, and utilization of geospatial data, information, and knowledge for multiple science domains*".
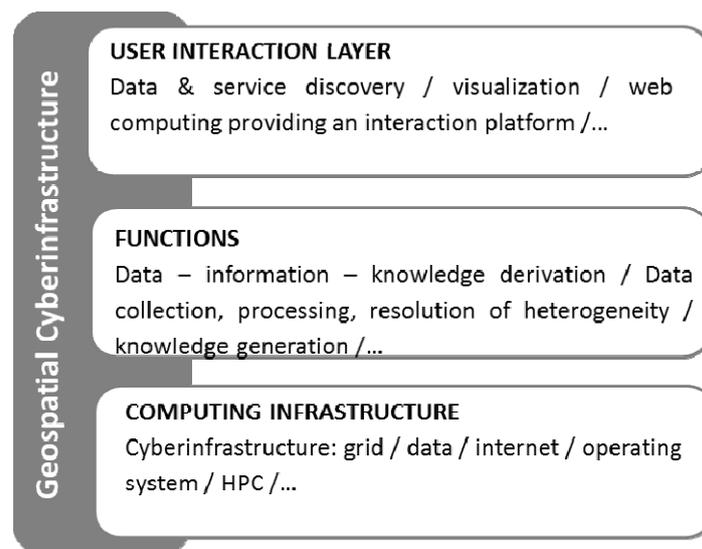
Both definitions refer to the infrastructure that is used in computations for generating information from spatial data.

The GCI framework consists of a series of building blocks that are represented in generalized form in Figure 3. The fundamental block is computing and network functions including the grid, a data center, scheduling, and security. A geospatial middleware links the cyberinfrastructure and the geospatial cyberinfrastructure. The functions of the GCI focus on the transformation of data into knowledge. This transformation process requires components like data management, information processing, cross-scale and domain management, and ontologies. The technologies supporting the GCI are among others spatial data infrastructures, DGIP, sensor networks, visualization and interoperability.

All these functions and tools together are then available to users and user domains like the environment, climate, geography and education. Yang *et al*. [14] state that cross-domain sharing and collaborations are an integral part of GCI.

Examples for GCI given in [14] are coming from environmental sciences [35], coastal and ocean studies [36], the use of Virtual Globes in education contexts, and the geosciences network (GEON) as one of several Earth-system science GCIs. "Grid processing on demand" (G-POD) is a "classical" geospatial cyberinfrastructure provided by the European Space Agency (ESA). It has its focus on the processing of satellite images and provides researchers with access to computing resources for processing large amounts of data. A GCI project that explores the integration of scientists with different backgrounds is LEAD—the Linked Environments for Atmospheric Discovery project [37]. LEAD is an example of a cyberinfrastructure being a science gateway offering capabilities for performing data discovery, running analyses and models requiring high computing power as well as designing computational analyses [12].

Figure 3. The main components of geospatial cyberinfrastructure after Yang *et al*. [14].

4. The Geoprocessing Web

The geospatial processing web is a functional framework for facilitating analyses required for progress in the geoscientific domain. Yue *et al*. [19] introduced the term geospatial processing web in 2010. They provide the following definition:

> "*The Geospatial Processing Web named here refers to a distributed, integrated, and collaborational service-oriented geoscientific research environment ...*" ([19], p. 758).
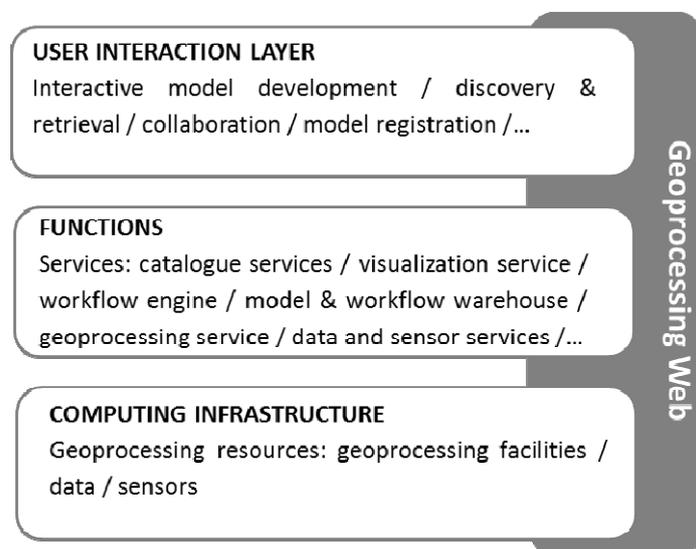
The geospatial processing web is therefore a platform designed for conducting spatial analysis online. It implements a service-oriented architecture using recent web standards and web-based services. The users of this platform shall profit from state-of-the-art technology in online data repositories, service technology, distributed computing, interoperability standards, and like technologies [19]. Yue *et al*. [19] mention cyberinfrastructure including grid technologies and cloud computing as backbone of the geospatial processing web. To differentiate GCI from the geospatial processing web, Yue *et al*. state that:

> "*For the geoscientific, education, and research users, current technologies do not directly address the issue of how distributed data and geoprocessing functions can be used to actually meet their geospatial analysis demands*" ([19], p. 757).

Zhao *et al*. [15] build on the work of Yue *et al*. [18] and alter the term geospatial processing web into *geoprocessing web*. They give the following definition: "The Geoprocessing Web allows geospatial data to be processed in real time for creating value-added information" ([15], p. 5). The geoprocessing web supports: interoperability, light-weight protocols, collaboration, distribution of resources, real-time processing, and services for infrastructure, software and platform. The geoprocessing web is a construct that brings together developments in areas like spatial data infrastructures, sensor web, grid and cloud infrastructure, web mapping, workflow development, and model sharing.

Zhao *et al*. [15] provide an overview over components of the geoprocessing web, which is provided in a simplified version in Figure 4. The figure shows three main layers: the geoprocessing resources, the geospatial data, service and model management platform and the geoprocessing modeling and application platform. The geoprocessing modeling and application platform are components that differentiate this approach from the geospatial cyberinfrastructure [15]. These components specialize on the development of models in a collaborative environment and the publishing, discovery, retrieval and execution of models for analyzing geospatial data. The other elements of the geoprocessing web have equivalences in GCI.

Figure 4. Key components of the geoprocessing after Zhao *et al*. [15].

Implementations of the geoprocessing web include GeoBrain [38] and GeoPW [19]. GeoBrain is a framework for geospatial knowledge building that offers a collaborative platform for online analysis. This platform, called GeOnAS, provides access to data repositories and offers analysis functions built on top of an interoperable web service architecture. GeoPW—the implementation of the geospatial processing web by its originators [19]—offers GIS functionality from GRASS and GeoStart GIS as web processing services (WPS). The virtual globe "GeoGlobe" is used as an interface to large data and for visualizing analyses. The GeoPW provides also a geoprocessing modeling tool called

GeoPWDesigner. This modeling tool allows the user to compose processing steps to a workflow. The workflow is then instantiated and executed through chaining of web services.

5. Differences and Commonalities of the Approaches to E-Science

E-Science environments offer a full spectrum of technology that aims at enabling scientific discovery. The components of environments that allow scientists to pursue e-Science, include workflow tools for chaining services for specific analyses, computing infrastructures, and visualization services besides others [13]. In addition, e-Science environments support collaboration, which is important for scientific progress. In the GIScience domain, the approaches of the geospatial cyberinfrastructure and the geoprocessing web are differentiated as two separate contributions to pursue e-Science. The general definitions of GCI and the geoprocessing web show that both approaches aim at providing users with functionality and computing resources to generate knowledge from data in a collaborative way. The realizations of this vision in specific implementations or applications differ. A comparison of the elements contained in the definitions of GCI and the geoprocessing web is presented in Table 1.

Table 1. A comparison of the definitions of geospatial cyberinfrastructure (GCI) and the geoprocessing web.

	Elements of the Definitions	Geospatial Cyberinfrastructure	Geoprocessing Web
Objectives	Advancement of GIScience domain	X	X
	Data collection, management and utilization	X	X
	Flexibility in geospatial analyses	-	X
	Support of collaboration	X	X
Functions (selection)	Data/service discovery	X	X
	Data integration	X	X
	Data visualization	X	X
	Data analysis and knowledge generation	X	X
	User interaction component	X	X
	Workflow development component	-	X
	Model registration and model discovery	-	X
Resources	High performance computing resources	X	-
	Distributed GI processing technology and web services	X	X
	Data	X	X

The comparison shows that the objectives of GCI and the geoprocessing web are basically identical. Only the flexibility regarding geospatial analyses based on web services is highlighted in the definition of the geoprocessing web. In terms of functions and resources the differences are that GCI puts emphasis on the computing infrastructure, whereas the geoprocessing web highlights the importance of workflow and model development components.

5.1. Characteristics of Two Representative Applications

For a more detailed discussion of similarities and differences between GCI and the geoprocessing web, two representative implementations are compared: ESA grid processing on demand (G-POD) as

example for a geospatial cyberinfrastructure and the GeoBrain Online Analysis System (GeOnAS) geoprocessing model development tool as example for the geoprocessing web. The comparison is done following the layered architecture of these applications. The layers consist of: a user interaction layer, functions or services and a computing infrastructure. In the user interaction layer, the user requests and manages the processing of data. This request is translated to a set of functions that provide intelligence to transform data into information. The computing infrastructure processes the service requests.

ESA G-POD provides processing on demand for Earth observation data available at the European Space Agency [39,40]. The technology behind G-POD includes a computational grid and access to cloud computing resources. About 180 terabyte of data are available online. G-POD provides predefined services like real-time extraction of thermal anomalies for volcanoes, flood crisis/damage mapping service, image rendering, geocoding, and visualization. Users submit requests through the G-POD portal, are informed about the processing stage and can then download the results. For adding additional services, scientists submit an application description to the G-POD team that implements the services and provides it to G-POD users.

GeoBrain is an online portal for the access to and analysis of Earth observation data. It was developed for scientists, educators and students by the North American Space Agency (NASA) [38]. One of its components is the GeoBrain GeOnAS [41,42]. GeOnAS is implemented based on interoperable, standard-compliant web services. The system allows the analysis of large data resources in an online environment; its look and feel is comparable to a GIS. Additional data sources and processing resources can be included through the GeOnAS interface. The goal of GeOnAS is the generation of knowledge from data [41].

Table 2 summarizes the characteristics of the implementations. The comparison of these two implementations thereby follows the components of user interaction, functions or services and computing infrastructure.

Table 2. A comparison of a geospatial cyberinfrastructure and a geoprocessing web implementation.

Layer Component	Specific Feature	European Space Agency Grid Processing on Demand (G-POD)	GeoBrain Online Analysis System (GeOnAS)
User interaction	Data discovery	✓	✓
	Flexibility in data analysis	-(Predefined operations)	✓ (Large set of services that can be combined)
	Collaboration (in the form of inclusion of resources)	-	✓
	Access to resources	Registration required	Public access
Functions/Services	Availability of model components	-	-
	Visualization service	✓	✓
	Catalogue and data services	✓	✓
	Analysis functionality	✓	✓
Computing infrastructure	Technology used	Hardware grid, cloud	OGC web services; HPC resources

5.1.1. User Interaction Layer

Zhao *et al*. ([15], p. 6) say

> *"[t]he unique emphasis of the Geoprocessing Web is the share and access of geoprocessing utilities from the perspectives of communication, collaboration, and participation"*.

The ability of the geoprocessing web to prepare and use models is what makes the difference to the geospatial cyberinfrastructure. In fact, the user interaction layer of GCI does not include user-centered components like a geoprocessing workbench and interactive model development components that appear in the top-most layer of the geoprocessing web. The expert knowledge on how to combine functions for a specific analysis task is not part of a computing infrastructure. Issues concerning the specification of input and output to models, reliability and quality are, however, linked to the infrastructure used for handling the models [34].

The GCI community is aware of shortcomings in respect to user-centered developments. Yang *et al*. [14] state that the focus in GCI needs to be put on the users of the systems rather than on technology. The emphasis of technology over user-related features in GCI goes back to its roots in cyberinfrastructure developments. In addition, it is in the nature of infrastructures to act in the background, without being much noticed [31].

The two examples presented in Section 5.1 support this general difference between GCI and the geoprocessing web. In G-POD, the user can access predefined functionality for processing Earth observation data. The extension of existing functionality requires the interaction with the team behind G-POD. In the GeOnAS tool, the user is provided with a large set of functions that she can combine at will. The focus here is on providing flexibility for the analysis of resources.

Nevertheless, the reduction of GCI to its role as provider of infrastructure does not hold for all implementations. Issues of collaboration, human computer interaction, and support for decision making can be considered in implementations of GCI [34]. Also the role of GCI to act as science gateway, as discussed in relation to the LEAD project in section 3, supports the conceptualization of GCI as integral e-Science environment. The LEAD platform, which is referred to as cyberinfrastructure, does have a workflow development component and offers its users flexibility in the composition of analyses.

5.1.2. Functions

Both approaches provide users with functions for deriving knowledge from data. The geoprocessing web offers all functionality as services. The services include data discovery and retrieval, visualization and geoprocessing and the specific services related to modeling and workflow generation. The provision of functions as services supports the composition of services into adapted workflows. However, the availability of services does not imply that the user has control over the execution of the services. Services can be chained in three manners: transparent, translucent or opaque (*cf.* [29]). In case of transparent chaining the user combines services. Translucent service chains are prepared chains of services. The user invokes the translucent service chain and knows which services are connected. Opaque service chains are like a black box for the user.

Functionality in GCI does not necessarily follow a service-oriented approach as in the geoprocessing web. In the GCI example G-POD, the functionality is inbuilt in processing tools. How these processing tools are implemented is not known by the users. G-POD focuses on providing ready-made processing tools and does not offer a component for developing analysis workflows.

5.1.3. Computing Infrastructures

The geoprocessing web is conceptualized as web-based service oriented architecture. It uses web service standards for its functioning as seen in the GeOnAS example. In GeOnAS the web services are executed using high performance computing resources.

The historical development of the geospatial cyberinfrastructure is driven from a technological perspective. Its initial focus is on providing a data storage and computing infrastructure that builds on grid computing and cloud computing technology. A large degree of work on cyberinfrastructures refers to these infrastructure related technologies [12]. G-POD is an example for a GCI providing computing power to its users through executing processes on hardware grids and in the cloud.

5.2. Integration of Geospatial Cyberinfrastructure and Geoprocessing Web

Computing resources offered in cyberinfrastructures can complement the components of the geoprocessing web. For example, the linkage of a service-oriented architecture with grid computing has been shown in Hobona *et al.* [43]. High performance computing resources are also used in GeOnAS. In fact, GCI frequently has its focus on the computing infrastructure underlying e-Science environments. However, this approach should not be reduced to the role of acting as a backbone for geoprocessing (*cf.* [15]).

The geoprocessing web highlights the importance of workflow and model development and flexibility on the user side regarding geoprocessing. The GCI community recognizes the importance of user-centered developments [14]. Therefore, GCI can profit from developments regarding model exchange and workflow development of the geoprocessing web.

The components of GCI and the geoprocessing web are complementary due to the respective foci of the approaches. The integration of separately developed components will lead to e-Science environments that offer flexibility in analysis to expert users and efficiency in the execution of services. An existing example for such an environment in the domain of atmospheric sciences is the LEAD platform [37]. These developments show, that differentiating e-Science environments into GCI or geoprocessing web approaches may become obsolete in the long-run.

6. Conclusions

The underlying goal of e-Science initiatives is to support scientific progress. What is required to make scientific progress? Resources need to be available, analysis functionality provided and information communicated. Collaboration of scientists within and across disciplines seems to be a key to scientific progress today and tomorrow. Collaboration includes the formalization of domain knowledge and wrapping of analysis steps in models. For executing the models to generate information, data and analysis capabilities shall be provided online and in (near) real-time.

The two approaches towards e-Science in the geographic information domain that were discussed in this article are the geospatial cyberinfrastructure and the geoprocessing web. The aims of the geospatial cyberinfrastructure and the geoprocessing web overlap, since both approaches are related to e-Science. The geoprocessing web is a new term introduced to highlight the requirements on the user management level [15]. Its specific contribution is the introduction of model development and sharing components. The introduction of the geoprocessing web seems to be borne out of a shortcoming of today's geospatial cyberinfrastructures in relation to human-centered developments. Due to historical reasons, the focus of GCI is still more on technology rather than the user [12,14]. Nevertheless, the boundaries of where GCI ends and the geoprocessing web starts are not sharp. The objectives and technological frameworks of the two approaches overlap. In the long-run, the integration of approaches with different strengths will lead to full-fledged e-Science environments that provide flexibility on the user side as well as efficiency regarding processing. Independently of the name given to an approach, what counts for its users is the provision of functionalities relevant for their use cases.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Brennan, J.; Lee, H.J.; Yang, M.; Folk, M.; Pourmal, E. Working with NASA's HDF and HDF-EOS earth science data formats. *Earth Obs.* **2013**, *25*, 16–19.
2. European Parliament and Council. *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)*; Official Journal of the European Union: Luxembourg, 2007.
3. Group on Earth Observation (GEO). *The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan*; Group on Earth Observation: Geneva, Switzerland, 2005.
4. Yue, P.; Gong, J.; Di, L.; He, L.; Wei, Y. Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *Geoinformatica* **2011**, *15*, 273–303.
5. Dadi, U.; Di, L. Creating Web Service Interfaces and Scientific Workflows Using Command Line Tools: A GRASS Example; In Proceedings of the 17th International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–6.
6. Craglia, M.; Nativi, S.; Díaz, L.; Vaccari, L. Towards Multi-Disciplinary Interoperability: The EuroGEOSS Contribution. In Proceedings of EuroGEOSS—Advancing the Vision of GEOSS Conference (EuroGEOSS 2012), Madrid, Spain, 25–27 January 2012.
7. Janowicz, K.; Hitzler, P. The digital earth as knowledge engine. *Semant. Web* **2012**, *3*, 213–221.

8.   Craglia, M.; de Bie, K.; Jackson, D.; Pesaresi, M.; Remetey-Fülöpp, G.; Wang, C.; Annoni, A.; Bian, L.; Campbell, F.; Ehlers, M.; *et al*. Digital earth 2020: Towards the vision for the next decade. *Int. J. Digit. Earth* **2011**, *5*, 4–21.

9.   Gore, A. The Digital Earth: Understanding Our Planet in the 21st Century. Presented at the California Science Center, Los Angeles, CA, USA, 31 January 1998.

10. Fook, K.D.; Vieira Monteiro, A.M.; CÃmara, G.; Casanova, M.A.; Amaral, S. Geoweb services for sharing modelling results in biodiversity networks. *Trans. GIS* **2009**, *13*, 379–399.

11. Gray, J. E-Science: A Transformed Scientific Method. In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 2009 ed.; Hey, T., Tansley, S., Tolle, K., Eds.; Microsoft: Redmond, WA, USA, 2009; pp. 16–31.

12. Stewart, C.A.; Link, M.; Simms, S.; Hancock, D.Y.; Plale, B.; Fox, G.C. What is Cyberinfrastructure? In Proceedings of ACM SIGUCCS Fall Conference on User Services 2010 (ACM 2010), Norfolk, VA, USA, 24–27 October 2010; pp. 37–44.

13. Hey, T.; Trefethen, A.E. Cyberinfrastructure for e-Science. *Science* **2005**, *308*, 817–821.

14. Yang, C.; Raskin, R.; Goodchild, M.; Gahegan, M. Geospatial cyberinfrastructure: Past, present and future. *Comput. Environ. Urban Syst.* **2010**, *34*, 264–277.

15. Zhao, P.; Foerster, T.; Yue, P. The geoprocessing web. *Comput. Geosci.* **2012**, *47*, 3–12.

16. National Science Foundation (NSF). *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*; No. cise051203; NSF: Arlington, VA, USA, 2003; p. 84.

17. Scharl, A.; Tochtermann, K. *The Geospatial Web: How Geobrowsers, Social Software and the Web 2. 0 Are Shaping the Network Society*; Springer: London, UK, 2007.

18. Egenhofer, M.J. Toward the Semantic Geospatial Web. In Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems, Mclean, VA, USA, 8–9 November 2002; pp. 1–4.

19. Yue, P.; Gong, J.; Di, L.; Yuan, J.; Sun, L.; Sun, Z.; Wang, Q. GeoPW: Laying blocks for the geospatial processing web. *Trans. GIS* **2010**, *14*, 755–772.

20. Bernard, L.; Mäs, S.; Müller, M.; Henzen, C.; Brauner, J. Scientific geodata infrastructures: Challenges, approaches and directions. *Int. J. Digit. Earth* **2013**, doi: 10.1080/17538947.2013.781244.

21. Zhang, T.; Tsou, M.-H.; Qiao, Q.; Xu, L. Building an intelligent geospatial cyberinfrastructure: an analytical problem solving approach. *Proc. SPIE* **2006**, *6420*, 64200A:1-64200A:14.

22. Kiehle, C. Business logic for geoprocessing of distributed geodata. *Comput. Geosci.* **2006**, *32*, 1746–1757.

23. Zhang, T.; Tsou, M.-H. Developing a grid-enabled spatial web portal for Internet GIServices and geospatial cyberinfrastructure. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 605–630.

24. Schäffer, B.; Baranski, B.; Foerster, T. Towards Spatial Data Infrastructures in the Clouds. In *Geospatial Thinking*; Painho, M., Santos, M.Y., Pundt, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 399–418.

25. Yu, J.J.; Qin, X.S.; Larsen, L.C.; Larsen, O.; Jayasooriya, A.; Shen, X.L. A GIS-based management and publication framework for data handling of numerical model results. *Adv. Eng. Softw.* **2012**, *45*, 360–369.

26. Jaeger, E.; Altintas, I.; Zhang, J.; Ludäscher, B.; Pennington, D.; Michener, W. A Scientific Workflow Approach to Distributed Geospatial Data Processing Using Web Services. In Proceedings of The 17th International Conference on Scientific and Statistical Database Management, Santa Barbara, CA, USA, 27–29 June 2005; pp. 87–90.

27. Dubois, G.; Skøien, J.; de Jesus, J.; Peedell, S.; Hartley, A.; Nativi, S.; Santoro, M.; Geller, G. eHabitat: A Contribution to the Model Web for Habitat Assessments and Ecological Forecasting. In Proceedings of The 34th International Symposium on Remote Sensing of Environment, Sydney, Australia, 10–15 April 2011.

28. Brauner, J.; Foerster, T.; Schaeffer, B.; Baranski, B. Towards a Research Agenda for Geoprocessing Services. In Proceedings of 12th AGILE International Conference on Geographic Information Science, Hannover, Germany, 2–5 June 2009.

29. Friis-Christensen, A.; Ostländer, N.; Lutz, M.; Bernard, L. Designing service architectures for distributed geoprocessing: Challenges and future directions. *Trans. GIS* **2007**, *11*, 799–818.

30. Schade, S.; Ostländer, N.; Canut, C.G.; Schulz, M.; McInerney, D.; Dubois, G.; Vaccari, L.; Chinosi, M.; Sánchez, L.D.; Bastin, L.; *et al*. Which Service Interfaces fit the Model Web? In Proceedings of GeoProcessing 2012: The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services, Valencia, Spain, 30 January–4 February 2012.

31. Poore, B.S. Users as Essential Contributors to Spatial Cyberinfrastructures. In Proceedings of the National Academy of Sciences of the United States of America, Washington, DC, USA, 5 April 2011; Volume 108, pp. 5510–5515.

32. Foster, I.; Kesselman, C. *The Grid: Blueprint for a New Computing Infrastructure*; Morgen Kaufmann: San Francisco, CA, USA, 1998.

33. Yang, C.A.; Goodchild, M.B.; Huang, Q.A.; Nebert, D.C.; Raskin, R.D.; Xu, Y.E.; Bambacus, M.F.; Fay, D.E. Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? *Int. J. Digit. Earth* **2011**, *4*, 305–329.

34. Yang, C.; Raskin, R. Introduction to distributed geographic information processing research. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 553–560.

35. Minsker, B.; Myers, J.; Marikos, M.; Wentling, T.; Downey, S.; Liu, Y.; Bajcsy, P.; Kooper, R.; Marini, L.; Contractor, N.; *et al*. NCSA Environmental Cyberinfrastructure Demonstration Project: Creating Cyber Environments for Environmental Engineering and Hydrological Science Communities. In Proceedings of 2006 ACM/IEEE Conference on Supercomputing (SC'06-ACM 2006), Tampa, FL, USA, 11–17 November 2006.

36. Agrawal, G.; Ferhatosmanoglu, H.; Niu, X.; Bedford, K.; Li, R. A Vision for cyberinfrastructure for coastal forecasting and change analysis. *Lect. Note. Comput. Sci.* **2006**, *4540*, 151–174.

37. Droegemeier, K.K.; Gannon, D.; Reed, D.; Plale, B.; Alameda, J.; Baltzer, T.; Brewster, K.; Clark, R.; Domenico, B.; Graves, S.; *et al*. Service-oriented environments for dynamically interacting with mesoscale weather. *Comput. Sci. Eng.* **2005**, *7*, 12–29.

38. Di, L. GeoBrain-A Web Services Based Geospatial Knowledge Building System. In Proceedings of NASA's Earth Science Technology Conference (ESTC 2004), Greenbelt, MD, USA, 2004; pp. 22–24.

39. Farres, J. *G-Pod ESA Grid Processing on Demand for Working Scientists*; ESRIN: Frascati, Italy, 2010.

40. ESA Grid Processing on Demand (G-POD). Available online: http://gpod.eo.esa.int/ (accessed on 12 July 2013).

41. Han, W.; Di, L.; Zhao, P.; Wei, Y.; Li, X. Design and Implementation of GeoBrain Online Analysis System (GeOnAS). In Proceedings of Web and Wireless Geographical Information System 2008 (W2GIS 2008), Shanghai, China, 11–12 December 2008; Volume 5373, pp. 27–36.

42. GeOnAs. Available online: http://geobrain.laits.gmu.edu/OnAS/ (accessed on 12 July 2013).

43. Hobona, G.; Fairbairn, D.; Hiden, H.; James, P. Orchestration of grid-enabled geospatial web services in geoscientific workflows. *IEEE Trans. Autom. Sci. Eng.* **2010**, *7*, 407–411.