The GEO-Cloud Experiment: Global Earth Observation System Computed in Cloud

Jonathan Becedas
Elecnor Deimos,Satellite Systems
Francia 9, Pol. Industrial La Nava, 13500 Puertollano, Ciudad Real
Email: jonathan.becedas@elecnor-deimos.es

***Abstract*—Earth Observation commercial data sales have increased a 550% in the last decade. The field is considered a key element in the European Research Road Map and an opportunity market for the next years. Nevertheless, EO still presents critical challenges that have to be overcome to cover the demand of services that are currently offered but also to advance new ones. The main challenge of this technology is the treatment of massive and large-sized data obtained from the recordings of earth observations. EO industries implement on-site conventional infrastructures to acquire, store, process and distribute the geo-information generated with the recorded data. However these solutions have the risks of over/under size the infrastructure, they are not flexible to cover sudden changes in the demand of services and the access to the information presents large latencies. These aspects limit the use of EO technology for real time use such as to manage crises. The use of cloud computing technology can overcome the previously defined limitations that present conventional infrastructures because of its elasticity, scalability and on-demand use characteristics.**

The GEO-Cloud experiment is part of the FP7 Fed4FIRE project. It is a close to reality industry driven experiment implemented in the Fed4FIRE infrastructure and its European platforms PlanetLab Europe, Virtual Wall and BonFIRE. The experiment has started on December 1st 2013. It will go beyond conventional services and infrastructures in the EO sector to implement and test in cloud a complete EO system (from the acquisition of geo-data with a constellation of satellites to its distribution to end users with remote access). The main objective is to value if the use of future internet technologies provides socio-economical viable solutions applicable to industry to offer highly demanding services such as crisis management in the EO sector. This paper presents the GEO-Cloud experiment design, architecture and foreseen research activity.

I. Introduction

At the end of the last decade, Earth Observation (EO) commercial data sales reached $1.1 billion, at the beginning only $200 million. This is a 550% increase in 10 years [1]. The forecast for this decade is$4 billion in commercial data sales at the end of 2019 [2]. This makes EO a field of new business opportunities and work. In addition, in 2002 the Group on Earth Observations (GEO) was launched, confirming that EO is a priority activity which provides social benefits. EO accumulates spatial and temporal records of the world itself and it is applied in diverse sectors; monitoring of the environment, natural disasters and civil security being of relevant interest [3]–[5]. According to this, the European Commission (EC) in partnership with the European Space Agency (ESA) and the European Environment Agency (EEA) has created the system GMES/Copernicus (Global Monitoring of Environment and Security) to provide Europe operational and autonomous capability to observe the Earth. Although the proven importance of EO in society, the access to the information obtained from satellites follows traditional and expensive paths to cover on demand services for different potential customers: conventional data centers and conventional distribution of services. This presents several drawbacks: i) the cost of acquiring recent images of the Earth is very high, which limits the access of the general public, ii) clients cannot access directly neither fast to the information they need because this has to be processed and ad-hoc distributed, iii) the service is not flexible to be adapted to sudden changes in the demand [6]. Cloud computing is presented as a possible solution to improve common services and create new market opportunities [7] because it is elastic, scalable and it works on demand through virtualization of resources [8]. However the transfer of conventional EO infrastructures and services to the cloud seems to be expensive because of the massive and large data that has to be stored during years [9], processed and distributed to different clients with different requirements. Although cloud computing reduces risks associated to over/under-size, and the total cost of ownership (TCO) [10]. Public entities and organizations promote the use of the cloud in EO. For example GMES masters presented a contest to design applications and services; ESA developed the G-POD initiative to facilitate scientists a platform to run algorithms and applications by using the ESA EO catalogue [11]; and the European Directive INSPIRE dictates guidelines and rules for the distribution of the geographic information. Nevertheless, in the private sector only specific applications have implemented low added value services based in cloud as e.g. some Geographical Information System (GIS) based applications (ESRI implemented ArcGIS on Amazon EC2). What seems to be clear is that cloud computing perfectly fits with the requirements of EO services, but it is still unknown how it can be socially, economically and technologically viable, and if the solutions that it can offer improve the services that are already offered with conventional technology. Thus, the question to solve is the following: does cloud computing and, in general, future internet technology provide socio-economical viable solutions for highly demanding services in Earth Observation? The use of cloud computing technology as IaaS is expensive at long term for EO services, but if it improves the information access, reduces the latencies, better distributes the information, shorten the distances between all the agents involved and improves the services; maybe there is a middle point where advantages and drawbacks encounter. GEO-Cloud Experiment goes beyond conventional data infrastructures used in EO industry and be-

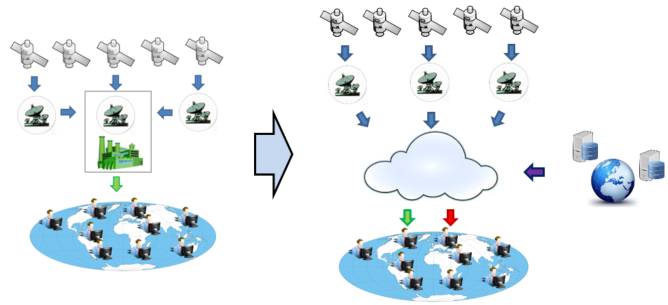

Fig. 1. GEO-Cloud Concept. The picture represents a conventional EO system in which the data is transferred to a conventional data centre where it has to be stored, processed and distributed to the clients in an ad-hoc manner (left). The concept is changed to use cloud technology to acquire data, store it, process it, integrate it with other sources and distribute it to end users and test viable solutions for its real implementation (right).

yond the implementations of applications running in cloud, to quest which parts of a complete infrastructure of EO (if not the complete infrastructure) are technologically and economically viable to be virtualized to offer basic and high added value services. Figure 1 depicts the GEO-Cloud Concept.

The GEO-Cloud experiment is under the FP7 Fed4FIRE IP project [12], [13]. GEO-Cloud will make use of the following European platforms: PlanetLab Europe [14], [15], Virtual Wall [16], [17] and BonFIRE [18], [19] of the Fed4FIRE future internet infrastructure.

The paper is structured as follows: i) Section I introduces the work. ii) Section II is devoted to present the objectives of the GEO-Cloud experiment. iii) In Section III, the description, architecture and implementation of the GEO-Cloud experiment are described. iv) Finally, Section IV is devoted to explain the main conclusions.

II. GEO-Cloud Scientific and Technical Objectives

The success of the GEO-Cloud experiment is validated by achieving the following scientific and technical objectives:

A. To validate if future internet technology provides a viable solution for Earth Observation systems

The main objective of the GEO-Cloud experiment is to validate if future internet cloud computing and networks provide viable solutions for conventional EO systems, by analysing the limitations and strengths of the technology and by comparing this with hybrid and private infrastructures. We will research the challenges in cloud implementation of conventional EO systems, which currently limit the transfer of traditional EO network infrastructures to cloud. This will establish the basis for the implementation of EO infrastructures in cloud and will advance the state of the art of cloud development by identifying the resilience, performance, security, privacy, interoperability and context awareness of user centered services.

B. To implement and trial in Fed4FIRE a complete close to real Earth Observation system

Another objective of the GEO-Cloud experiment is the implementation of an emulated but realistic Earth Observation system in the Fed4FIRE infrastructure. The Fed4FIRE infrastructure is a federated heterogeneous open infrastructure constituted by different platforms. GEO-Cloud will interconnect the Planet Lab Europe, Virtual Wall and BonFIRE platforms to emulate a complete EO system: i) The flight system constituted by a Low Earth Orbit constellation of optical satellites and the Ground Stations distributed around the world to communicate with them and download the captured images. ii) The data center in which the images are processed on demand. This will be done by implementing a cloud model with two layers architecture: one for basic services and other one for added value services. iii) A model of different end users (governments, agencies, private companies, etc) that are on line accessing the final products.

The implementation of the GEO-Cloud experiment in Fed4FIRE will improve the resilience for communication, information access, distribution, storage and processing of massive and big sized data online and on demand. GEO-Cloud will manage the interoperability of its models to cover the demands of clients with different time, storage and processing requirements.

C. To validate the GEO-Cloud Earth Observation models

To test and validate the models implemented in GEO-Cloud is also a priority. This is done by having observability, control and monitoring of the experiment at all layers to manage the interoperability and adjust the system to a viable, realistic solution. The consecution of this objective will improve the models to be more realistic.

D. To compare the performance of the cloud solution for different types of services

To evaluate the scalability of the system we will compare different types of services to cover different types of demand. This will be done by testing of different loads in storage, processing and communications and of different levels of variability in the demand to establish the comparison. In Figure 2 the different levels of demand are depicted as a benchmark to be tested in the experiment.

E. To verify if the Fed4FIRE infrastructure is appropriate for running close to reality and industry driven experiments

The final objective is to verify if the Fed4FIRE infrastructure and tools are appropriate for running large scale, complex, close to reality and industry driven experiment. The consecution of this objective will provide feedback to the Fed4FIRE infrastructure to improve its limitations and detect its strengths in order to facilitate the implementation of future experiments.

III. Experiment Description

The experiment consists on the virtualization of a conventional EO system to offer on demand services to clients with the objective of validating its viability, find the strengths and weaknesses of using cloud computing technology and establish possible solutions for a future implementation in the market. There are three components:

COMPANY'S SERVICES	Service Type		Loads in Cloud Technology		Demand Variability	USERS' DEMAND
	Basic	Basic	Processing	Low	Constant	
				Medium		
		Advanced		High		
			Storage	Low	Variable	
	High added value	Pull		Medium		
				High		
		Push	Communications	Not urgent/Not RT	Highly variable	
				Urgent/Not RT		
	Hosting			Urgent/Real Time		

Fig. 2. Benchmark to relate users demand with services and technology to be used

a) In-orbit mission: this component generates the raw data. This consists of un-processed images of the Earth captured by a constellation of satellites. The images in raw data with their associated metadata are downloaded to different ground stations distributed around the world.

b) Treatment of data: the data downloaded to the ground stations is transferred to the data center where it is stored, processed at different levels based on the services offered and distributed to the end users. The data acquired by the constellation of satellites is integrated with other sources in the data center to provide higher quality services.

c) End-users: users of the provided services with different levels of remote access rights based in subscriptions.

The system will be emulated at all levels for its monitoring and control. The in orbit mission and end-users traffic, accesses, network topology, communications and data transfer is emulated in Virtual Wall and PlanetLab Europe. The tools, models and architecture to treat the data are implemented in the BonFIRE Cloud. Different parameters are controlled (bandwidth, latencies and other impairments) to emulate a realistic system with different levels of demand. Different levels of subscriptions at end-user level are emulated.

The system follows the following requirements to cover the demand of the end-users:

- The geo-information has to be generated and transferred fast.
- The geo-information has to be easily and close to real time accessible.
- The geo-information has to be on demand processed and distributed.
- The raw data acquisition from the satellites follows an almost constant pattern.
- The data acquired has to be stored and be accessible to provide historical records.

A. *Scenario*

In our model, we have roughly estimated a constellation of 17 satellites to cover the surface of the Earth in images in a daily basis. The images are downloaded in real time to several ground stations (we have roughly estimated three: one in Spain, one in Canada and one in the in Norway; although we will design in detail and test if additional ground stations in other locations would be necessary for daily coverage). The raw data is transferred from the ground stations to a cloud computing infrastructure where the data is stored, processed on demand at different levels and integrated with information obtained from external sources to the system: sensors and other databases. All that data is processed to generate geo-information which is distributed to end-users to cover different levels of demand. There will be different types of demand of services for both basic satellite imagery services and high added value services: constant demand, variable demand and highly variable demand. The demand can also have different types of loads associated, attending to three components: processing load, storage load and communications load. Possible combinations of them can also be presented. Types of load in the demand:

- Processing: Different load levels for processing such as minimum load, medium load and high load.
- Storage: Different load levels for storage such as minimum load, medium load and high load.
- Communications: Urgent responses with real time requirement, urgent responses without real time requirement and not urgent responses without real time requirement (this also includes transfer of big data volumes).

GEO-Cloud provides three types of services to the end-users. In those types of services, representative use cases will be implemented for testing:

1) Basic satellite imagery services: It includes accessing to low and automated processed satellite images from catalog. We can differ between services offered with two types of images:

- Basic images: orthorectified images, images with minimum geolocation information as per GIS exploitation.
- Advanced images: more complex images with more geolocation information such as mosaics.

To carry out the experiment we will use images from our catalog, and public images from sources such as GMES or NASA (as e.g. Landsat images, which are freely available), but we will not integrate them in this service (just for testing, see that in Figure 2 there is no input from external sources in the Layer 1, which will implement this service). We will test the transfer of data and accessing to the information in cloud with different loads of processing. This will be used as a control for comparing with the rest of services. We will test

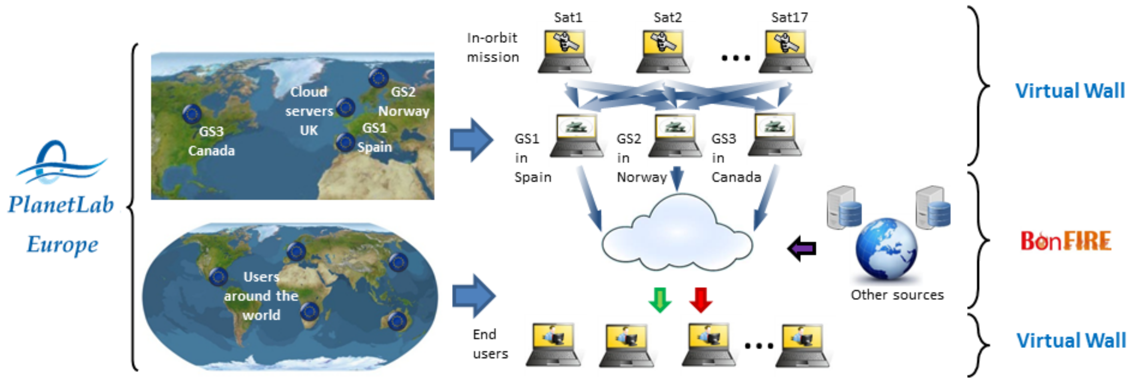

Fig. 3. GEO-Cloud in Fed4FIRE

processing loads and timing to measure the response of the system.

2) High added value services: It includes instant accessing to automatically processed satellite images and data 24/7. In this case, data from other sources will be used and integrated with satellite imagery to combine the information and provide the service: Meteorological data, environmental sensor networks, etc. Online processing will be tested with different computing loads and demands. Among the services that we will test we will focus in the following:

- Pull type service: Service under request. For example the emulation of crisis management of natural catastrophes such as earthquakes and fires (during the detailed design stage of the experiment we will evaluate the possibility of emulating other crisis services that can fit in the experiment). This service will cover all the phases of a crisis:i)Pre-crisis: To cover the demand of preliminary studies before the crisis occurs, such as the analysis of areas which are susceptible to suffer a crisis. ii)During-crisis: To cover the demand of fast response during the occurrence of a crisis. Actions such as crisis planning, definition of secure areas, prediction of the crisis expansion or identification of evacuation routes will be emulated. c)Post-crisis: Damage assessment services after a crisis has occurred.
- Push type service: The system periodically and automatically provides information to the end user without request. Tracking of specific areas or mobiles, for example the tracking of a building in construction or boats in the ocean. Different levels of subscription will be tested. This is characteristic of constant demands in daily, weekly or monthly basis. More information about pull and push strategies can be found in [20].

3) Hosting services: GEO-Cloud hosting services will reserve space and access for the end-users to offer the possibility of storing their own geo-data, which can have been generated in the previous services. This solution is designed for those clients that do not foresee to acquire specific storage hardware or do not want to invest on very large storage facilities.

Figure 2 depicts the benchmark to establish the relations between services, technology and demand. Multiple combinations can appear between services and users demand. One hypothetical path example between services and demand can be the following: Service Type Hosting with Low Storage Load for new customers with Constant Demand Variability because of periodical subscription to an information service. The classification will allow us to analyse the technological and economic viability of the services for their implementation in real life. It will also allow us to find the limitations of the cloud technology and establish requirements for its implementation. The implementation of the services in cloud will involve the emulation of three types of cloud infrastructures: public, hybrid and private. We will establish a comparison with the three infrastructure concepts.

In summary, GEO-Cloud will test and validate our models for a Global EO system, will classify the types of the end users demand of services, will link them with the type of service and the cloud technology to be used, and will find which parts of an EO system are viable to be virtualized in a public cloud. This will be done by implementing the system in the Fed4FIRE infrastructure.

B. Implementation in Fed4FIRE

The complete implementation of GEO-Cloud in Fed4FIRE is depicted in Figure 3.

1) Cloud based service: To facilitate offering the previous services we implement a multi-layered cloud model in the BonFIRE cloud infrastructure to generate on demand geo-information (Figures 4 ad 5).

The GEO-Cloud implementation in the BonFIRE cloud has two layers:

Layer 1: This layer involves the basic satellite imagery services. It acquires the raw data, stores it, has the first level of processing, distributes the processed data and offers the hosting service. The storage servers will have three functions: i) hosting service for end users, ii) storage and processing of the raw data and required metadata (this data will be stored in a long term basis) and iii) short term storage of processed data (the duration long and short term storage will be tested during the execution of the experiment) for its later distribution, processing or both. Thus the processed information can be used for other end users that require it,

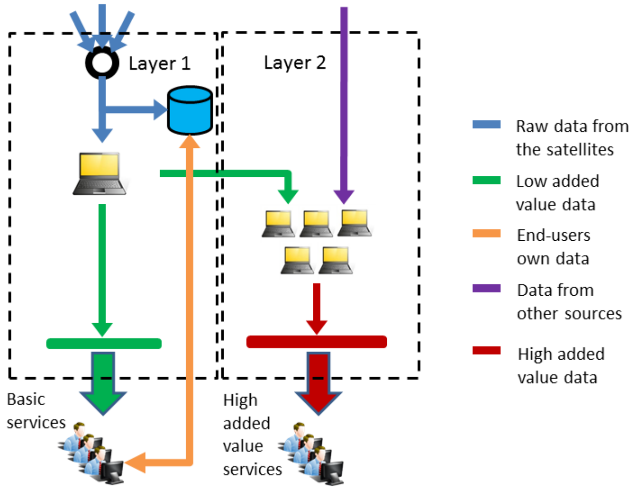

Fig. 4. Two layered EO model in BonFIRE cloud: Cloud model for storage, processing and distribution.

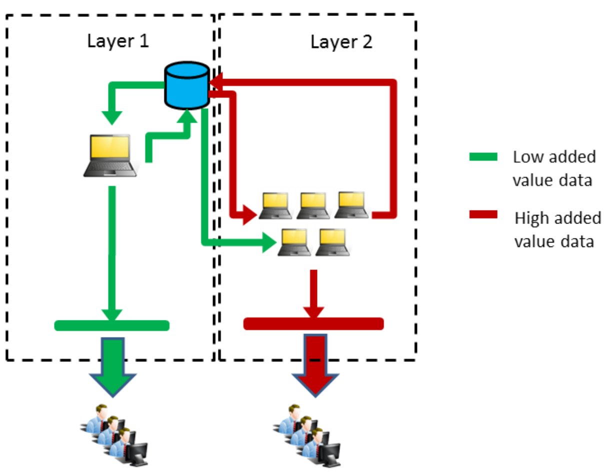

Fig. 5. Short term storage of processed information for its later distribution, postprocessing or both.

without the necessity of processing again the raw data to produce the same information. When this information is old, it can be deleted.

Layer 2: This layer involves the high added value services. It can use historical processed, real time captured and pre-processed data from layer 1. This layer processes the information for real time generation of geo-information and offers real time access and distribution to the end-users. It also includes the acquisition and integration of information coming from other sources, such as NASA and GMES.

We will also evaluate the implementation of the layer 1 as a private infrastructure to create a hybrid infrastructure. We will compare the results with those of having the complete system as a private infrastructure. This will allow us to establish a comparison in costs and performance between all possible different implementations.

2) Access to the service: The access to the services will be done with different levels of subscription. Thus, we can emulate that end-users can access the type of information they really need, and we can analyse the demand of the services we are offering. The access will be remote and will emulate a web service. A Virtual Wall network will emulate end-users accessing to the services by simulating computational loads, accesses, etc. Through the implementation of a network in PlanetLab Europe we will measure the characteristics of the real network when end users around the world remotely access the data. This information will be used to upload the implemented model in Virtual Wall. With the Virtual Wall emulation we can have complete control of the resources, compute the demand and all the factors involved in the situation to test the system and record its behaviour under the different conditions.

3) Acquisition of geo-data: The acquisition of geo-data will be obtained from the in-orbit mission and from external sources:
a) **In-orbit mission**: The constellation of satellites and the ground stations will be emulated in Virtual Wall. We will implement a network topology to communicate the different satellites with the ground stations. Every satellite in its orbit and every ground station models will be simulated in a node.

The satellite models will simulate the orbits and the pass of the satellites over the ground stations. The ground stations models will simulate the coverture of the antennas and the download of the data. When a satellite is inside this radius, the satellite downloads the data to the ground station that is visible. The downloaded data in the ground stations is transferred to the BonFIRE cloud. With the Virtual Wall network we will control, latencies, other impairments and the complete network topology. We will be able of changing the parameters of the network to simulate its limitations and its performance under realistic working conditions.

In order to determine the correct link characteristics for the connections between the ground stations in Spain, Norway and Canada (probably more), and the cloud infrastructure we will measure appropriate values for the link impairment by measuring the actual characteristics between these different geographical locations using the PlanetLab Europe testbed. Figure 3 shows the topology networks to be implemented and tested in Virtual Wall and the network to be tested in PlanetLab.

b) **External sources:** We will acquire information from other geo-data sources. That information will be directly transferred to the BonFIRE cloud for its integration. We have identified two relevant data sources to be used: GMES, providing remote sensing information related to several fields (Land, Atmosphere, Water and Sea, etc.), and the European Environmental Agency. Other Open Data portals and data repositories will be evaluated during the design phase of the end-user services to be tested.

IV. CONCLUSION

In this paper, the design and architecture of the EU FP7 GEO-Cloud experiment has been presented. The GEO-Cloud experiment is implemented in the EU Fed4FIRE infraestructure. GEO-Cloud connects the Virtual Wall and BonFIRE EU platforms to emulate a complete and realistic Earth Observation System through the implementation of three models: i) A constellation of satellites taking pictures of the Earth and dowloading them in Ground Stations distributed around the World. ii) The transfer of data to a data center implemented in cloud. iii) Distribution of added value products to end users around the World.

The topology networks that communicate the remote sensing and the end users models with the BonFIRE cloud are implemented in PlanetLab Europe. There, real parameters and impairments are measured to update the models implemented in Virtual Wall.

In the BonFIRE cloud, a complete data center is implemented in order to receive, process, host and distribute imagery data and produce added value products that will be delivered on demand to the end users.

The implementation and consecution of this experiment allow us to evaluate if future internet technologies provide viable solutions for Earth Observation systems, or if on the contrary traditional data centers are more appropriate.

In addition, GEO-Cloud will contribute to validate the effectiveness of the Fed4FIRE framework and to support its sustainability. The lessons learnt in GEO-Cloud will also contribute to improve the experience of future experimenters in the whole experiment cycle and will provide feedback to improve Fed4FIRE's functionalities.

From the state of the art we can suspect that a hybrid solution between traditional data centers and a complete cloud solution it is maybe the most appropriate. GEO-Cloud will solve this indetermination and will establish an objective view of the current future internet technology used for Earth Observation. The consecution of the experiment will provide information about how cloud computing can contribute to Earth Observation systems. The reduction of the processing, storage, communications and distribution costs of EO services will facilitate the access to the remote sensing technology of common end users, but also of a more general public. GEO-Cloud will contribute defining the basis to advance in the use of geospatial information of the nine Societal Benefit Areas defined in GEO: disasters, health, energy, climate, water, weather, ecosystems, agriculture and biodiversity; by demonstrating whether or not cloud computing offers technologically and economically viable solutions to offer highly demanding services.

Acknowledgment

The author would like to thank Flix Pedrera, Jess Quirce and Miguel Lizondo for their support in the conception of the GEO-Cloud project.

This work was carried out with the support of the Fed4FIRE-project ("Federation for FIRE"), an Integrated project funded by the European Commission through the 7th ICT-Framework Programme. (318389). It does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

References

[1] Euroconsult, *Satellite-based Earth Observation, Market Prospects to 2019*. Euroconsult, 2010.

[2] A. Keith, "Earth observation: Emerging markets, partnerships set to fuel global growth," *Earth Imaging Journal*, pp. 18–20, February 2011.

[3] D. Li, J. Shan, and J. Gong, Eds., *Geospatial Technology for Earth Observation*. Springer, 2009.

[4] D. Li, *An Overview of Earth Observation and Geospatial Information Service*. Springer, 2009, ch. 1, pp. 1–25.

[5] X. Ge and H. Wang, "Cloud-based service for big spatial data technology in emergency management," in *Proceedings of the ISPRS (ISPRS'09)*, 2009, pp. 126–129.

[6] C. Yang, M. Goodchild, Q. Huang, D. Nebert, R. Raskin, Y. Xu, M. Bambacus, and D. Fay, "Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?" *International Journal of Digital Earth*, vol. 4, pp. 305–329, 4 2011.

[7] K. Navulur, D. Lester, A. Marchetti, and G. Hammann, "Demistifying cloud computing for remote sensing applications," *Earth Imaging Journal*, vol. 16, pp. 14–19, 4 2013.

[8] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, pp. 50–58, 4 2010.

[9] J. Farres. (2012) Cloud computing and content delivery network use within earth observation ground segments: experiences and lessons learnt. European Space Agency. ESA UNCLASSIFIED - For Official Use. [Online]. Available: $http : //earth.esa.int/gscb/papers/2012/11 - Cloud\_computing\_content\_delivery.pdf$

[10] R. Perry and S. Hendrick, "The business value of amazon web services accelerates over time," White Paper, IDC, July 2012.

[11] ESA G-POD Team, "Earth observation grid processing-on-demand portal user manual. technical report," European Space Agency, Tech. Rep. GRID-GODS-EOPG-MA-06-0004, 2006.

[12] FED4FIRE Project. European Comission. [Online]. Available: $http : //www.fed4fire.eu/$

[13] W. Vanderberghe et al, "Architecture for the heterogeneous federation of future internet experimentation facilities," in *Future Network and Mobile Summit (FutureNetworkSummit), 2013*, July 2013, pp. 1–11.

[14] PlanetLab Europe Project. European Comission. [Online]. Available: $http : //www.planet - lab.eu/$

[15] S. Fdida, T. Friedman, and T. Parmentelat, *OneLab: An open federated facility for experimentally driven future internet research. In New Network Architectures*. Springer Berlin Heidelberg, 2010, pp. 141–152.

[16] Virtual Wall Facility. iMinds. [Online]. Available: $http : //www.iminds.be/en/develop - test/ilab - t/virtual - wall$

[17] S. Bouckaert, S. Becue, B. Vermeulen, B. Jooris, I. Moerman, and P. Demeester, *Federating wired and wireless test facilities through Emulab and OMF: the iLab.t use case. In Testbeds and Research Infrastructure. Development of Networks and Communities*. Springer Berlin Heidelberg, 2012, pp. 305–320.

[18] BonFIRE Project. European Comission. [Online]. Available: $http : //www.bonfire - project.eu/$

[19] A. C. Hume, Y. Al-Hazmi, B. Belter, K. Campowsky, L. M. Carril, G. Carrozzo, and G. Van Seghbroeck, *BonFIRE: A Multi-cloud Test Facility for Internet of Services Experimentation. In Testbeds and Research Infrastructure. Development of Networks and Communities*. Springer Berlin Heidelberg, 2012, pp. 81–96.

[20] D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi, *Designing and managing the supply chain: concepts, strategies and cases*. McGraw-Hill, 1999.