Asian Journal of Innovation and Policy (2015) 4.3: 381-396
DOI: http//dx.doi.org/10.7545/ajip.2015.4.3.381

An Empirical Analysis of Worldwide Cyberinfrastructure

Manhyung Cho*

Abstract Cyberinfrastructure is a research infrastructure that provides an environment in which research communities can get access to distributed resources and collaborate at unprecedented levels of computation, storage, and network capacity. The Worldwide LHC Computing Grid (WLCG) is a global collaborative project of computing or data centers that enables access to scientific data generated by the Large Hadron Collider (LHC) experiments at CERN. This case study analyzes the WLCG as a model of cyberinfrastructure in research collaboration. WLCG provides a useful case of how cyberinfrastructure can work in providing an infrastructure for collaborative researches under data-intensive paradigm. Cyberinfrastructure plays the critical role of facilitating collaboration of diverse and widely separated communities of researchers. Data-intensive science requires new strategies for research support and significant development of cyberinfrastructure. The sustainability of WLCG depends on the resources of partner organizations and virtual organizations at international levels, essential for research collaboration.

Keywords Cyberinfrastructure, WLCG, computing grid, scientific data, data center

I. Introduction

Today, scientists are overwhelmed with data that are captured or generated by instruments, simulations, and sensor networks (Hey et al., 2009). Data are growing exponentially in all sciences, and science is increasingly driven by data. Scientists often no longer have to operate scientific instruments or equipment; they do their research work using data digitally captured, cured, synthesized, or visualized (Critchloco and van Dam, 2013). Gray (2007) described this new data-intensive science as the 'fourth paradigm of science', distinguishing it from empirical, theoretical, and computational science. Thousand years ago, science was empirical describing natural phenomena

Submitted, November 25, 2015; 1st Revised, December 9; Accepted, December 9.
* Department of Public Administration, Hannam University, Daejeon, 305-796, South Korea; mancho@hnu.kr

(1st paradigm). The last few hundred years, a theoretical branch using models and generalizations emerged (2nd paradigm). Last decades saw computer simulations become an essential third paradigm, simulating complex phenomena (3rd paradigm). Data-intensive science (4th paradigm) is based on cyberinfrastructures, which include resources such as computing, storage, as well as network needed to capture, cure, and analyze scientific data. Data-intensive science is characterized as collaborative, networked, and data driven, synthesizing theory, experiment and computer simulations with scientific data.

Cyberinfrastructure is the set of tools and technologies to support data-intensive science for analysis, data mining, visualization and exploration, and scholarly communication. Many researchers depend on experimental or observational data generated from accelerators, satellites, telescopes, sensor networks, and supercomputers. However, such datasets are too big and complex for individual scientists to manage. Accordingly, new types of computing technologies have been developed that make possible data management and analysis. Cyberinstrastructure is a kind of new infrastructure that supports data-intensive science for research collaborations by interconnecting computing powers around the world (Foster et al., 2001).

The Worldwide LHC Computing Grid (WLCG) provides an excellent model of worldwide cyberinfrastructure in research collaboration. The project is a global collaboration of more than 170 computing centers in 42 countries, linking up national and international cyberinfrastructures. WLCG can store and share experiment data generated by the Large Hadron Collider (LHC) at CERN (Organization Européenne pour la Recherche Nucléaire/European Organization for Nuclear Research) with the computing resources supported by many associated national and international grids across the world, such as the European Grid Initiative (Europe-based) and the Open Science Grid (US-based), as well as many other regional grids.

This paper purposes to analyze the WLCG as an empirical model of cyberinfrastructure in research collaboration. Cyberinfrastructure is increasingly central to the research endeavors to meet the data-intensive paradigm of research. The case of WLCG provides an indication of how important cyberinfrastructure can be in facilitating collaboration in different kinds of data-intensive sciences. Based on the analysis of the case, some policy implications will be suggested as related to collaborative science and research innovation under the fourth paradigm of science.

II. Theoretical Background

1. Cyberinfrastructure

As mentioned above, data-intensive science can be described as a new research methodology, supported by computing capabilities and scientific data (Hey et al., 2009). It is distinguished from traditional experimental and theoretical methodologies in terms of large-scale, data-driven, computing capabilities. Cyberinfrastructure is fundamental to data-intensive science. The concept of cyberinfrastructure can be traced back to a foundational work in computer science that resulted in the creation of 'grid computing' in the 1990s. Foster et al (2001) defined grid computing as 'a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.'

The National Science Foundation (NSF) has led the development of cyberinsfrastructure. The NSF report (2003) introduced a new concept of cyberinfrastructure composed of elements such as computing power, scientific data, instruments, virtual organizations, etc. It proposed to use cyberinfrastructure to build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments that operate at unprecedented levels of computational, storage, and data transfer capacity. Later NSF (2007) suggested more specific areas of investments for cyberinfrastructure: (1) high-performance computing, (2) data, data analysis, and visualization, (3) virtual organizations for distributed communities, and (4) learning and workforce development. NSF (2011) gave an advanced definition of cyberinfrastructure as 'the broad collection of computing, systems, software, data acquisition and storage systems, and visualization environments, all generally linked by high-speed networks, often supported by expert professionals.'

In sum, cyberinfrastructure can be defined as a research infrastructure that provides a digital environment in which researchers (scientific communities) can share and utilize distributed resources for collaboration at a high level of computational, storage, and network capacity. It combines all related technologies such as computer hardware, software, storage, and network to provide an environment where scientific communities (or virtual organizations) can collaborate in the cyber space of research. Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments linked together by middlewares and networks to enable research not otherwise possible. Cyberinfrastructure links technology elements through software and

high-performance networks to create a larger system of collaborative research environment.

2. Components of Cyberinfrastructure

2.1 Grid Computing

Grid computing is a service infrastructure for sharing data, computing, and storage resources over the internet (IBM, 2005). A grid is a collection of machines, sometimes referred to as nodes, resources, members, donors, clients, hosts, engines, and many other such terms. It turns the global network of computers into one vast infrastructure for solving large-scale data-intensive research applications. It enables collaboration among geographically dispersed communities in the form of virtual organizations (Bird, 2009). Grid computing is distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation (Foster et al., 2001). A fully developed grid computing makes possible to share all resources for research collaboration around the world. Grid computing may provide a solution to the problems of storing and processing the large quantities of data that are captured or generated by scientific equipment and instruments.

Most common resources for grid computing include processors, data storage, communication networks, middlewares, and platforms (IBM, 2005). Computing capabilities are provided by the processors of computers on the grid. The processors can vary in speed, architecture, software platform, and other associated factors, such as memory, storage, and connectivity. Data storage provides some volume of storage for grid use. Storage can be memory attached to the processor or it can be secondary storage, using hard disk drives or other permanent storage media. The distributed nature of data-intensive science requires data transfer capabilities in order to support research collaborations. Communication networks include communications within the grid and external to the grid. Communications within the grid are important for sending jobs and their required data to points within the grid. Machines on the grid may have connections to the external internet in addition to the connectivity among the grid machines. Middlewares let users simply submit jobs to the grid without having to know where the data is or where the jobs will run. The software can run the job where the data is, or move the data to where CPU power is available. Platforms on the grid will often have different architectures, operating systems, devices, capacities, and equipment. Each of these items represents a different kind of resource that the grid can use as criteria for assigning jobs to machines.

2.2 Scientific Data

Today, research across all fields of inquiry is increasingly data intensive (Hey et al., 2009). Advances in experimental and computational technologies have led to an exponential grow in the volume, variety, and complexity of data (Critchloco and van Dam, 2013). Huge volumes of scientific data are generated or captured by instruments, computers simulations, or sensor networks. Some areas of science are inundated with unprecedented increases in data volumes from colliders, satellites, telescopes, sensors, and supercomputers, compared to the past. For example, in astronomy and particle physics petabytes of data are generated from experiments (Bell et al., 2009).

Scientific data as an infrastructure and a tool for scientific discovery can play a crucial role in the fourth paradigm of science. Scientists can do their research works by analyzing second-hand data provided through grid computing systems even though they do not own or operate instruments. Digital data environments make possible diverse research communities to collaborate over cyberinfrastructure by experimenting data. Science can be data-intensive as a result of the development of grid technologies and high-speed networks through which data are collected, generated, shared and analyzed. Scientists can access and analyze scientific data stored in computing centers around the world via networks. New scientific opportunities are made possible with the combination of scientific and new computing technologies for capturing, curing, and analyzing these data. The enormous growth in the availability of scientific data also contributes to research productivity and accelerates collaboration by sharing data and research outcomes.

2.3 Virtual Organization

The virtual organization is a coalition of dispersed researchers that pool resources to achieve common objectives of research, collaboration and sharing of scientific data on cyberinfrastructure. Virtual organizations built upon cyberinfrastructure are becoming a reality for today's distributed, collaborative, data-intensive research (Borgman et al., 2009). It is a kind of online community where distributed researchers work together and share relevant data and tools to solve common research goals. Virtual organization is a fundamental part of data-intensive paradigm of scientific research. Researchers form virtual organizations to collaborate, share resources, and access common data through cyberinfrastructure. Grid computing technologies provide services that make collaboration among researchers. Virtual organization can be characterized by dispersion/distribution, diversity, coherence, security, coordination, and flexibility (Bird et al., 2009). Such an

organization is highly complex, particularly when enabled by large-scale, data-intensive, and distributed infrastructures such as those in data-intensive science.

In scientific communities virtual organizations usually include a group of researchers who use and share common research resources such as scientific experiment data. Virtual organizations provide them with distributed resources for a collaborative research (DeSanctis and Monge, 1998). Grid computing infrastructure such as WLCG is based on the concept of virtual organization. Researchers have formed WLCG to share experiment data and collaborate via the grid infrastructure. Virtual organizations build on cyberinfrastructures and enable science communities to pursue their research without constraints of time and distance. Usually, virtual organizations are created by a group of individuals whose membership and resources may be dispersed geographically, but can be shared through the use of cyber-infrastructure systems.

III. Analysis of the WLCG Case

1. WLCG Project

The WLCG project was launched in 2002 to provide a global computing infrastructure to store, distribute and process the data annually generated by the LHC experiments at CERN. It integrates thousands of computers and storage systems in hundreds of data centers worldwide. CERN itself provides only about 20% of the resources needed to manage the LHC data. The rest is provided by the member countries' national computing centers through grid computing networks. The LHC experiments rely on distributed computing resources. WLCG is a global solution, based on the grid technologies/ middleware, distributing the data for processing, user access, local analysis facilities etc.

1.1 Structure of WLCG

The project aims at the ‘collaborative resource sharing’ between all the scientists participating in the LHC experiments. The infrastructure is managed and operated through collaboration between the experiments and the participating computer centers to make use of the resources regardless of their location. WLCG provides the computing resources needed to process and analyze the data generated by the LHC experiments. It interconnects hundreds of computer centers that provide data storage and computing

resources needed by the experiments and operate these resources in a shared grid-like fashion.

WLCG is managed and operated by a worldwide collaboration between the experiments (scientific communities) and the participating computer centers making use of partner resources no matter where they are located. The resources are distributed for funding and technological reasons. Significant costs of maintaining and upgrading the necessary resources are imposed upon member centers (called 'tiers') where individual partners fund their resources while contributing to the whole WLCG community. The hierarchical tier organization is based on MONARC (Model of Networked Analysis At Regional Centers) network topology (Legrand, 2004). The basic principle underlying the model is that every scientist should have equal access to the data and computing resources. The model allows for no single points of failure, multiple copies of the data, automatic reassigning of tasks to resources, access to data for all scientists independent of location, round-the-clock monitoring and support.

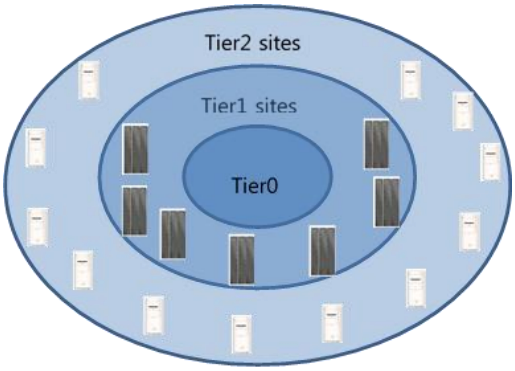

Figure1 WLCG basic architecture

1.2 Governance of WLCG

WLCG is governed with diverse layers of collaborative bodies (WLCG, 2015). The Collaboration Board is comprised of computing centers (Tiers) and experiment scientists and provides the main technical direction for WLCG. The Overview Board is a standing committee of the Collaboration Board and has the role of overseeing the functioning of the collaboration. It also acts as a clearinghouse for conflicts that may arise within the collaboration. The Management Board supervises the work of the project, maintaining the overall program of work and all other planning data necessary to ensure the smooth execution of the work of the project.

WLCG comprises almost 175 sites in 35 countries with about 350,000 computing cores. For worldwide collaboration, WLCG is composed of layers called 'Tiers' – Tier0, Tier1, Tier2, and Tier3. Each tier is made up of several computer centers and provides a specific set of services. Between them the tiers process, store and analyze all the data from LHC. Different computing centers have different roles according to their resources and their geographical location.

Tier0 is the computing center in CERN. As a hosting site of the LHC experiments, Tier0 at CERN has the central facility for data processing and archival that records all original data to permanent storage. Processed initially at CERN, the data are sent to other computing centers for storage and analysis. Tier1s are big computing centers with high quality of service used for most complex and intensive processing operations and archival. Tier1s consist of 13 computer centers around the world. They receive data directly from Tier0 and provide additional permanent storage. Tier1s are the hub of raw data at partner countries and provide researchers computing resources required for data analysis. Tier2 are computing centers across the world used primarily for data analysis and simulation. They are typically universities and other scientific institutes that can store sufficient data and provide adequate computing power for specific analytical tasks. There are around 155 Tier2 sites around the world. Tier3s consist of local clusters in a university department or even an individual laboratory. Their main role is to provide analytical capacity for local users.

2. Analysis of WLCG Components

2.1 Worldwide LHC Computing Grid (Grid computing)

CERN scientists initiated the LHC Computing Grid (LCG) project to create a computing environment that would meet the LHC experiments' unprecedented storage and computing needs (CERN, 2014). While the scale of data processing made grid computing a natural choice for the LHC experiments, locating sufficient computing resources at CERN is also limited by practical considerations (Bird et al., 2009). It is more effective to install computing and storage in universities and at national laboratories than to send them to CERN. Thus, they decided to design the LHC computing as distributed in order to use the available resources dispersed around diverse institutes or computing centers.

WLCG has used models derived from the MONARC Project (Legrand, 2004), which was based on a distributed infrastructure to deploy diverse computing resources linked through networking, using a grid middleware as connecting among computing centers. The model adopted the grid computing

paradigm of distribution and sharing of computing resources for data storage and services. Originally, it was started with LCG built on a grid computing of around 40 sites where LHC experiment data were used for simulation work. Later, the LCG evolved into the WLCG (Worldwide LCG), a collaboration that now includes some 175 sites in 35 countries. The computer centers that participate in WLCG join the community signing a memorandum of understanding that defines the levels of resources to be provided, as well as service levels. Computing centers rely on various grid-computing infrastructures to provide the tools and services required to fulfill the requirements.

The four main component layers of WLCG are physics software, middleware, hardware and networking (CERN, 2014). WLCG computer centers are made up of multi-petabyte storage systems and computing clusters with thousands of nodes connected by high-speed networks. They need software tools that satisfy the demands of researchers. Middlewares allow users to access computers distributed across the network. It is located between the operating systems of the computers and the applications software that is dedicated to a particular problem. Each data center manages a large collection of computers and storage systems. Optical links working at 10 GB connect CERN to each of the Tier1 centers around the world. This dedicated high-bandwidth network is called the LHC Optical Private Network (LHCOPN).

2.2 Data Generated from Experiments (Scientific Data)

Particles collided within the Large Hadron Collider (LHC) are recorded by detectors as a series of electronic signals and then sent to data center (Tier1) at CERN for digital reconstruction. LHC delivered billions of recorded collision events to the LHC experiments from proton-proton and proton-lead collisions. This translates to larger than 100 petabytes of data recorded at CERN. Around 30 petabytes or so of data are produced annually at experiments such as ATLAS, ALICE, CMS, LHCb, etc. These experiments are run by collaborations of scientists from institutes all over the world. Each experiment is distinct, and characterized by its detectors. For example, the biggest one of these experiments is ATLAS, which uses general-purpose detectors to investigate the largest range of physics possible. ATLAS investigates a wide range of physics, from the search for the Higgs boson and standard model studies to extra dimensions and particles that could make up dark matter.

ATLAS uses an advanced trigger system to tell the detector which events to record and which to ignore. The trigger system selects 100 interesting events per second out of 1,000 million in total. The data acquisition system channels the data from the detectors to storage. If all the data from ATLAS were to be

recorded, this would fill 100,000 CDs per second. This would create a stack of CDs 450 feet high every second, which would reach to the moon and back twice each year. The data rate is also equivalent to 50 billion simultaneous telephone calls. ATLAS actually only records a fraction of the data (those that may show signs of new physics) and that rate is equivalent to 27 CDs per minute.

The ATLAS computing model develops and operates a data storage and management infrastructure able to meet the demands of an annual data volume of utilized data processing and analysis activities spread around the world. The ATLAS Databases and Data Management Project (DB Project) leads and coordinates ATLAS activities in these areas, with a scope encompassing technical data bases, online databases, distributed database and data management services. The project is responsible for ensuring the coherent development, integration and operational capability of the distributed database and data management software and infrastructure for ATLAS across these areas. The ATLAS computing defines the distribution of raw and processed data to Tier1 and Tier2 centers, so as to be able to exploit fully the computing resources that are made available to the collaboration. Additional computing resources are available for data processing and analysis at Tier3 centers and other computing facilities to which ATLAS may have access.

2.3 Collaborative Communities (Virtual Organization)

High-energy physics has a long history of collaboration in creating and accessing unusual and expensive equipment such as particle accelerators. It has adopted a modern cyberinfrastructure and evolved coordination mechanisms that allow a distributed community of scientists to collaborate across long distances and over significant periods of time. Experiments at the Large Hadron Collider use detectors to analyze the myriad of particles produced by collisions in the accelerator. These experiments are run by collaborations of scientists from institutes all over the world.

ATLAS is one of the largest collaborative efforts involving more than 3000 investigators from nearly 38 countries working on labs, institutes, departments, and universities. International collaboration has been essential to this success. These scientists from more than 177 universities and laboratories pursue different research areas mostly in small groups working at their home institutions. All interested ATLAS researchers are invited to analyze the data by being part of analysis teams. ATLAS started before the term ‘virtual organization’ became popular, but it has been highly collaborative from the start through cyberinfrastructure of the ATLAS collaborative project.

ATLAS demonstrates the long-standing importance of cyberinfrastructure-enabled collaboration.

WLCG governance structures and decision-making processes are decentralized. In-kind contributions and collective sharing is encouraged. The actual implementation of the computing centers and is governed by the WLCG Collaboration, which encompasses more than 200 computing centers pledging resources to ATLAS. Each detector subsystem has its own management team. The ATLAS Executive Board and Spokesperson maintain general oversight of the Project. The Technical Coordination team is responsible for making sure that all the separate subsystems can fit together. In parallel, there are national representatives whose functions are to oversee the distributions of resources from each participating country to all the collaborating groups from that country, and make sure that those resources are well-used. The Collaboration Board sets out policy issues, and is not involved with their execution, which is the domain of the management

Candidates for a particular leadership position at ATLAS collaboration are proposed by the groups working on that system. Candidates for Spokesperson, and Collaboration Board (CB) Chair are proposed by the membership of the whole collaboration. The leadership is then elected, and their proposed 'management teams' have to be endorsed. The 'electing bodies' are the institutions relevant to the particular position, namely the whole collaboration for the CB Chair or Spokesperson, or the institutions participating in a given detector system in the case of a project leader for that system.

The ATLAS virtual organizations allow production and analysis users to run jobs and access data at diverse tier centers using grid computing tools. The main computing operations run centrally on Tie0, Tier1s, and Tier2s. Strong high-level links are established with other parts of the ATLAS organization, such as the T-DAQ Project and Physics Coordination, through cross-representation in the respective steering boards. The Computing Management Board, and in particular the Planning Officer, acts to make sure that software and computing developments take place coherently across sub-systems and that the project as a whole meets its milestones. The International Computing Board assures the information flow between the ATLAS Software & Computing Project and the national resources and their Funding Agencies.

IV. Implications for Collaborative Science and Innovation

As data-intensive science has emerged as a new research methodology, synthesizing theory, experiment and computation with statistics, a new way

of thinking is required. Data-intensive science requires new strategies for research support and significant development of cyberinfrastructure. It has the potential to transform not only how we do science, but also how quickly we can translate scientific progress into complete solutions (Critchloco and van Dam, 2013). Data-intensive science touches on some of the most important challenges we are facing. Meeting these challenges requires the collaborative effort of teams and also significant contributions form enabling data-intensive technologies. Data-intensive technologies have become an essential foundation in many different domains for sustainable progress in research and innovation. To address most pressing challenges, scientific communities need to move beyond the traditional research paradigms of theoretical, experimental, and computational science and move forward to make a full use of the data-intensive technologies.

Cyberinfrastructure is the backbone of a new research infrastructure that supports a whole new way of doing collaborative science. It enables different communities to come together and create virtual organizations to exploit a wide variety of distributed resources. Advances in high-performance computing, grid computing, advanced networking, data repositories, and visualization are all contributed to the development of cyberinfrastructure. Cyberinfrastructure plays the critical role of facilitating collaboration of diverse and widely separated communities of researchers. Critical scientific challenges will require unusual coordination of and collaboration between the diverse communities of researchers. Corresponding advances in cyberinfrastructure will facilitate these collaborations. Collaboration is essential to meeting the challenges, and cyberinfrastructure-based virtual organizations offer promise for improving research innovation.

Many areas of science are now becoming data-driven sciences. The imminent flood of data from the new generation of research facilities and equipment will also pose significant challenges for cyberinfrastructure to assist the process of capturing, curing, and sharing scientific data. Data-intensive science is characterized by the huge volume of scientific data required for management, analysis, visualization, and re-using. Because data used in the data-intensive approach to science are often diverse in distributed locations, scientists are not only to find ways to acquire the data, but also to develop new cyberinfrastructure tools for utilizing them. Data-intensive science requires more advances in cybernfrastructure technologies and services.

As data-driven science continues to increase in its impact, government needs to invest in more budget to develop cyberinfrastructure such as high-performance computing, software, data storage, and high-speed network, etc. Data center which provides an operational support for research groups is also a crucial part of cyberinfrastrucuture. In the case of LHC experiments, several

100 petabytes of additional storage are needed across the WLCG to provide space for archival, replication, simulation and analysis. The challenge how to process and analyze the data and produce timely results depends on the level of resources committed to the global collaborative efforts. Funding for data centers is needed to allow research communities to maintain experimental data sets and keep up with growing data sets. It is very important to maintain reliable data centers independent of project funding cycles so that scientific data are stored and reused regardless of project cycles.

Cyberinfrastructure technologies are becoming more complex and difficult to use. More advanced tools need to be developed to close the gap between the advanced cyberinfrastructure and scientists' ability to utilize it. Training and expertise for data service is essential to use advanced cyberinfrastructure to solve complex problems. Keeping sufficient staff support at a reasonable cost is a continuing concern. This calls attention to the importance of data center which is the hub of operation and technology development of cyberinfrastructure. New research paradigm requires more policy concern for the advancement of cyberinfrastructure in the areas of hardware, software, computational science activities, and human capital.

WLCG shows the possibility of worldwide cyberinfrastructure linking up national and international computing grids. A federation of more than 200 national computing or data centers join together to make a virtual organization, sharing resources connected by fast networks. This worldwide cyberinfrastructure allows scientists access to resources from national data centers distributed all around the world. They can rely on large-scale computing resources to address the challenges of analyzing large data sets, operating computations and simulations, and allowing for voluntary participation of large groups of researchers (Bloom and Gerger, 2013). These resources have become more distributed over a large geographic area, and some resources are highly specialized computing powers.

WLCG provides an empirical case of worldwide cyberinfrastructure working successfully in all aspects of large-scale data processing, data management and user analysis. Data is exported from CERN to main computing centers (Tiers), where data are stored, processed and further distributed to collaborating institutes for analysis. All members of the experiment groups have equal access possibilities to all experiment data, independent of their geographical location with the help of WLCG. The purpose of WLCG is to share the resources needed to process and analyze the data gathered by the LHC experiments. CERN provides only around 20% of the resources (computing power and storage).

WLCG has worked to link up the national and international data centers by worldwide cyberinfrastructure. The case demonstrates that distributing computing resources across many locations can be shared by international

collaboration of computing or data centers. It leads to greater engagement in the projects by individual institutions (Bloom and Gerger, 2013). While each data center contributes its own resources to WLCG, it has access to the resources of others and aggregation of the whole. Ultimately more resources which would not be possible without cyberinfrastructue can be available for any scientists regardless of nationality and location.

Finally, the WLCG case shows the important of virtual organizations as a special form of policy network. Generally, policy network refers to the structure of network and structure of actors who participate in the policy process (Rhodes, 1997). Three major factors for policy network analysis include actors, interaction, and network structure. Actors refer to all individuals and groups who participate in the policy process. Interaction is the practical process of mobilization and trade of policy resources to pursue policy goals and the results of strategies among actors. Last, network structure is the pattern of relations among actors. The network involved in WLCG grew complex and dynamic as different background of participants from laboratory (CERN), computing centers and scientists (research communities) came together. WLCG has favorable relationships and strong ties among network actors in the process of operating the project. CERN acted as a central organizer, but was very much dependent on participating computing centers for resources. Thousands of scientists around the world voluntarily contributed to address common research goals and thus made the effort really international through virtual organization for research collaboration. WLCG has been totally dependent on each other because participants are the major source of program funding. Over time, WLCG network actors interacted under conditions of reciprocal interdependence and developed into cooperative ties.

V. Conclusions

This paper provided a concise statement of the cyberinfrastructure and analyzed WLCG as an empirical case of cyberinfrastructure from a worldwide perspective. Cyberinfrastructure as a critical research infrastructure provides a digital environment in which research communities can get access to distributed resources and collaborate at unprecedented levels of computational, storage, and network capacity. The case of WLCG demonstrated how cyberinfrastructure can integrate and operate resources distributed all over the world and make all these resources accessible and usable for researchers. WLCG provides the channel for distributing, archiving

and processing the data produced by the LHC and gives a community of over 8,000 physicists near real-time access to LHC data.

WLCG has invested heavily in research collaboration both at the national and international levels, with impressive results. It has become a world leader in the field of grid technology. This technology has become a fundamental component of cyberinfrastructure. Many other research communities make reference to WLCG collaboration model when they design data centers and collaborative projects of their own. But it should be kept in mind that the sustainability of WLCG depends on how to secure resources and funds necessary to cover the costs of maintenance and operation. As a cyberinfrastructure, WLCG's success depends on the network stability of partner Tier organizations' cooperation to the point that they are willing to contribute more resources to the general pool.

References

Bell, G., Hey, T.L. and Szalay, A. (2009) Beyond the data deluge, Science, 323, 1297-1298.

Bird, I., Jones, B. and Kee, K.F. (2009) The organization and management of grid infrastructures, Computer, 36-46.

Bloom, K. and Gerber, R. (2013) Computing Frontier: Distributed Computing and Facility Infrastructures, Community Planning Study: Snowmass.

Borgman, C.L., Bowker, C.G., Finholt, T.A. and Wallis, J.C. (2009) Towards a virtual organization for data cyberinfrastructure, Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries.

CERN (2014) Annual Report 2014.

Critchloco, T. and van Dam, K.K. (2013) Data-Intensive Science, Boca Raton, FL.: CRS Press.

DeSanctis, G. and Monge, P. (1998) Communication processes for virtual organizations, J. Computer-Mediated Communication, 13(4), 25-37.

Foster, I., Kesselman, C. and Tuecke, S. (2001) The anatomy of the grid: enabling scalable virtual organizations, Int'l J. Supercomputing Applications, 15(3), 200-222.

Fox, P. and Kozyra, J. (2015) eScience and informatics for international science programs, Progress in Earth and Planetary Science, (12), 1-9.

Gray, J. and Szalay, A. (2007) eScience - A Transformed Scientific Method, Presentation to the Computer Science and Technology Board of the National Research Council. http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt.

Hey, T., Tansley, S. and Tolle, K. (eds.) (2009) The fourth paradigm: data-intensive scientific discovery, Microsoft Research Corporation, 978-0982544204.

IBM (2005) Introduction to Grid computing.

Legrand, I. (2004) MONARC Simulation Framework, ACAT'04, Tsukuba.

NSF (2003) Revolutionizing Science and Engineering through Cyberinfrastructure.

NSF (2007) Cyberinfrastructure Vision for 21st Century Discovery.

NSF (2011) A report of National Science Foundation Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges.

National e-Science Centre (2007) Developing the UK's e-Infrastructure for Science and Innovation.

Rhodes, R.A.W. (1997) Understanding governance, Buckingham: Open University Press. WLCG. http://wlcg.web.cern.ch/