

Análisis de planificadores e interoperación para ambientes HPGC

Diego Alberto Rincón
Universidad Católica de Colombia
Diagonal 47, Cl. 47 #15-50, Bogotá
+57 3005707966
darincon@ucatolica.edu.co

Cesar Orlando Díaz
Universidad Católica de Colombia
Diagonal 47, Cl. 47 #15-50, Bogotá
+57 3114996464
codiaz@ucatolica.edu.co

Holman Diego Bolívar
Universidad Católica de Colombia
Diagonal 47, Cl. 47 #15-50, Bogotá
+57 3003117177
hdbolivar@ucatolica.edu.co

ABSTRACT

In this article an analysis for a High Performance Grid Computing (HPCG) scheduler environment will be presented. We will exhibit the relevant standards by Open Grid Forum to establish a clear understanding of all its components, and in this way approach a grid problematic, known as interoperation. It'll also evaluate ways to perform this action in a standard way, focusing in the interaction of the basic grid components.

CCS Concepts

- Software and its engineering → Real-time schedulability
- Computer systems organization → Grid computing

Keywords

Grid Computing, Scheduling, Globus, Interoperation, interoperability, Grid.

1. INTRODUCCIÓN

El termino Grid Computing, se originó temprano en la década de los 90, como una metáfora para hacer computación poderosa y fácil de usar, como la red eléctrica tradicional en el trabajo seminal de Ian Foster y Carl Kesselman[2]. Como podemos observar en la Figura. 1 existen 3 componentes fundamentales, los cuales son necesarios para respetar la arquitectura propuesta [10].

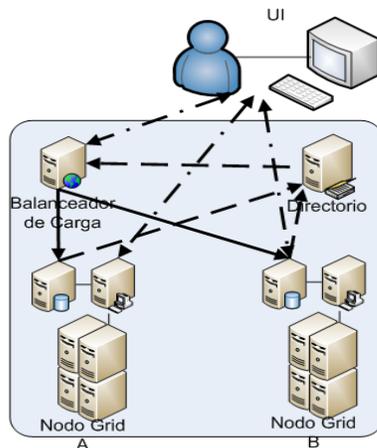


Figura. 1. Esquema General del Grid: La interfaz de Usuario (UI) somete un trabajo al Grid(puntos y líneas), esta petición es recibida por el Balanceador de Carga (WMS). El WMS contiene un listado de los recursos computacionales activos y consulta la disponibilidad (línea continua) al Compute Element (CE), una vez el CE devuelve la respuesta el WMS le envía la localización a la UI del recurso que puede usar.

Debido al repentino crecimiento e implementación de esta tecnología en la comunidad científica, desde al año 2001 el Open Grid Forum (OGF) ha creado una estructura jerárquica enfocada en el descubrimiento, solución y estandarización de problemas [3]. Dada la necesidad de implementar la tecnología de forma ágil, se han generado implementaciones variadas para problemas específicos, dando nacimiento a las grandes comunidades Grid alrededor del mundo entre estas se destacan *Open Science Grid (OSG)* [4], *Enabling Grids for E-sciencE (EGEE)* [5], *German Grid Initiative (D-Grid)* [6], *NorduGrid* [7] y *NAtional REsearch Grid Initiative (NAREGI)* [8]. Cada comunidad teniendo en cuenta el acercamiento a la tecnología, sus trasfondos en la coordinación de recursos computacionales, dieron origen a lo que actualmente conocemos como las diferentes implementaciones del muy conocido Globus Toolkit [9].

Cada una de estas implementaciones ha generado una rama de Globus como lo podemos observar en la Figura. 2. Éstas a través del tiempo han añadido servicios diferentes, generando así múltiples versiones de Globus, totalmente diferentes

Una de las filosofías del grid, el cual dice que todo debe estar implementado por estándares abiertos [1], ha creado la necesidad de hacer que estas grandes “islas” se comuniquen unas a otras y generen intercambio de datos y recursos computacionales para ofrecer una mayor cantidad de servicios [13]. A esta forma de intercambiar servicios y usarlos en cada “isla” se le conoce como interoperación [14].

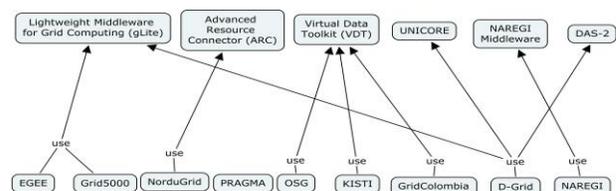


Figura. 2. Proyectos Grid Internacionales o Iniciativas de Grid Nacionales (NGI) (abajo) vs Implementaciones

2. PLANIFICADORES HPGC

Open Grid Forum (OGF), es una comunidad de usuarios involucrados en el área de la investigación, el desarrollo y el soporte de actividades relacionadas a los sistemas distribuidos de altas capacidades [3]. El objetivo fundamental del OGF es usar los estándares creados por Internet Engineering Task Force (IETF, www.ietf.org) [11] dándole soporte y complementando dichos estándares para fortalecer los procesos HPC propuestos por esta institución [12].

Debido a que parte del trabajo realizado por el OGF es dar las métricas claras para el trabajo colaborativo en el desarrollo de sistemas distribuidos de alto desempeño, se desarrolló una arquitectura de servicios (Open Grid Service Architecture – OGSA [15]) y de infraestructura (Open Grid Service Infrastructure – OGSi [16]). Usando los estándares y respetando los esquemas previamente definidos por el OGF se pueden realizar cambios para lograr cierto nivel de personalización a cada una de las soluciones grid aplicadas a los casos previamente expuestos se puede permitir un despliegue de los servicios expuestos en OGSA de manera mucho más desacoplada de cualquier plataforma específica.

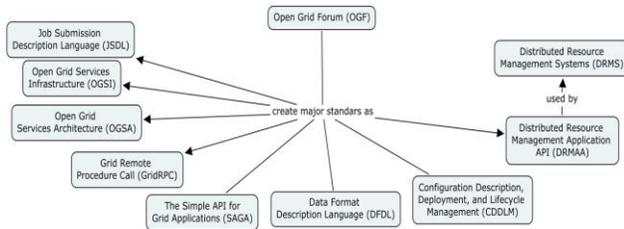


Figura. 3. Estándares del Open Grid Forum.

A partir de estas dos especificaciones, OGF describió una serie de estándares como se puede observar en la Figura 3, los cuales están enlazados a cada uno de los servicios descritos en OGSA y utilizando la infraestructura de puntos de control y de enlace definidos en la OGSi. Algunos de los estándares son:

- Job Submission Description Language (JSDL): Una extensión de XML para la descripción de tareas no interactivas usando los Manejadores de Recursos Locales (LRM)
- Grid Remote Procedure Call (GridRPC): Descripción de llamado a procedimientos remotes para ambientes Grid
- Data Format Description Language (DFDL): Descripción del modelado de datos generales, texto o binarios.
- The Simple API for Grid Applications (SAGA): Describe una interface para la programación de aplicaciones grid a alto nivel.
- Distributed Resource Management Application API (DRMAA): Conjunto de librerías para el envío de trabajos de control a sistemas de recursos distribuidos.
- Configuration Description, Deployment, and Lifecycle Management (CDDL): Describe el manejo, despliegue y configuración de ciclos de vida de los servicios grid entre dominios organizacionales.
- Grid Laboratory Uniform Environment (GLUE): Modelo de representación para la información de recursos grid.

Paralelo al desarrollo de los estándares y a la consolidación del OGF, se fue creando la implementación que se conoce como Globus Toolkit. A la fecha es considerado una implementación de facto en cuanto al uso de del grid.

3. INTEROPERACIÓN

La interoperabilidad, es la capacidad de dos o más sistemas para inter operar entre si [14]. Lo que significa que para que haya interoperación, los sistemas involucrados deben saber previamente como intercambiar información entre ellos, para que

una vez la información sea intercambiada, pueda usarse para prestar servicios en cada uno de los sistemas participantes de dicha interoperación.

Para lograr la interoperación de sistemas Grid hay que tener en cuenta dos aspectos claves, la arquitectura y el mecanismo de operación. La arquitectura define la estructura lógica de un sistema “Multi-Grid” y el mecanismo de operación define como los Jobs son manejados y ejecutados [17]. Teniendo en cuenta los aspectos se desprenden cinco retos los cuales son necesarios para lograr la interoperación:

- Selección de Recursos: Se refiere a la capacidad del sistema de seleccionar el recurso más apropiado para el trabajo bajo altas cargas de procesamiento.
- Propiedad de Entradas: Tiene relación con el uso de los sistemas de información y las entradas (prioridades) que tiene un sistema con respecto a otro al momento de usar un recurso.
- Escalabilidad: Traduce la habilidad del sistema a operar con normal desempeño una vez los recursos, número de usuarios y trabajos son incrementados.
- Confianza: Confianza permite a los Grids intercambiar información crítica entre ellos y participar de la toma de decisiones, aceptando que la decisión tomada es la correcta.
- Contabilidad: Habilidad del Grid de llevar el estado del uso de los recursos por usuario, trabajos y sitios.
- Disponibilidad: Es la habilidad que tiene un sistema a superar fallas en alguno de sus componentes.

Para que los retos de interoperación se puedan cumplir hay que tener en cuenta las habilidades de los Grids para el intercambio exitoso de información. Esta habilidad de “hablar” un mismo idioma entre los Grids es precisamente la Interoperabilidad [18].

4. INTEROPERABILIDAD

El primer paso para la interoperación es conseguir la interoperabilidad de los sistemas [18], teniendo en cuenta lo anterior, se puede decir que la interoperabilidad es la habilidad de un sistema para intercambiar información y usar aquella información que ha sido previamente intercambiada [14].

Dentro del esquema general la Computación en Grid se puede dividir en dos fases: la localización de recursos y el uso de los recursos localizados. La interoperabilidad juega un papel fundamental en cada una de estas fases que dentro de este contexto puede ser llamado de las siguientes maneras:

- Interoperabilidad Sintáctica: Es el intercambio de información relevante para el uso del sistema adyacente. Lo que quiere decir que se construye la información de un Grid a partir de parámetros específicos o características del mismo interoperable. Dando como resultado un mapa general de la localización de los recursos computacionales de cada Grid interoperable.
- Interoperabilidad Semántica: Toma como entrada la Interoperabilidad Sintáctica, teniendo la información de localización de los recursos del sistema foráneo. y decidiendo que recurso es el más indicado para el trabajo, se procede a el acceso a dicho recurso para usarlo.

El uso del Grid, en general da dos acciones generales que son: la localización y el uso de recursos. Para que las acciones puedan

realizarse satisfactoriamente en los sistemas interoperables deben tener en cuenta los siguientes problemas:

- **Acceso y Seguridad:** Hace referencia al uso de deferentes técnicas para lograr confiabilidad en el uso y el paso de información de los sitios Grid. Está enfocada más al uso de las organizaciones virtuales para lograr la interoperabilidad en los sitios que al uso de certificados digitales ya que desde el 2003 está constituida la International Grid Trust Federation (IGTF) y las Police Managment Authorities (PMAs).
- **Descubrimiento:** Una vez la seguridad y la confiabilidad está garantizada, el segundo servicio para garantizar la interoperabilidad y tomando como referencia los grandes Grids de producción, es el uso de servicios de directorios y servicios de información en el Grid. Lo anterior debido a la gran cantidad de recursos puestos a disposición de la comunidad. Los servicios de localización deben proveer el Fully Qualified Domain Name (FQDN) del gatekeeper y el nombre del LRM respectivo del host.
- **Despliegue de Aplicaciones:** Apunta a la implementación de estándares abiertos para el uso del los gatekeepers, implementando mecanismos de invocación remota. El primer paso es la construcción de un ambiente de envío de trabajos estándar.
- **Compatibilidad de Aplicaciones:** Uno de los problemas más significativos del uso del Grid probablemente sea el traspaso de aplicaciones de sistemas viejos a entornos de computación de alto desempeño. A este proceso se le conoce como Gridificación de aplicaciones. El objetivo es asegurar la portabilidad de la aplicación para garantizar la correcta interacción con el procesador y el sistema operativo.

Como vimos en la Figura 1, el esquema general de un Grid lo podemos dividir en dos fases. Podemos usar esas dos fases para crear la habilidad de interoperación, dichas fases las podemos denominar: Descubrimiento de Recursos y Uso de Recursos descubiertos

5. INTEROPERABILIDAD PARA EL DESCUBRIMIENTO DE RECURSOS

Teniendo en cuenta la alta heterogeneidad de los recursos pertenecientes al grid y el dinamismo de los mismos, los métodos de descubrimiento son herramientas vitales para poder brindar la transparencia en el uso del grid al usuario final. De la misma manera para responder a las necesidades de interoperabilidad en los múltiples sistemas grid, es necesario realizar una fase de descubrimiento de aquellos recursos que harán parte de esa interoperabilidad. Se han identificado dos tipos de métodos que solucionan el problema de intercambio de información entre sistemas Grid heterogéneos, son descubrimiento Orientado al Manejador de Recursos (CE2WMS) y descubrimiento Orientada al Balanceador de Carga (WMS2WMS)

5.1 Orientado al Manejador de Recursos (CE2WMS)

Cuando nos referimos al manejador de recursos se está haciendo referencia a la interfaz grid del LRM del sitio grid que quiere ofrecer sus servicios y hacer parte de la interoperabilidad. De esta manera el Manejador de recursos le envía la información pertinente al balanceador de carga y este se encarga de registrarla y tomar la decisión cuando llega la petición del cliente para el uso de los recursos tal como se muestra en la Figura 4. Este tipo de interoperabilidad tiene una desventaja importante, se debe tener

en cuenta el método de registro del CE en el balanceador de carga perteneciente a la Organización Virtual (VO) con la que se va a colaborar.

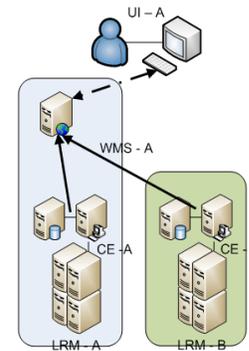


Figura. 4. En este esquema el CE reporta directamente al WMS, teniendo en cuenta que el WMS puede ser el de su misma implementación o de la que se quiere lograr interoperabilidad.

5.2 Orientado al Balanceador de Carga (WMS2WMS)

El Balanceador de carga es equivalente a lo que se conoce como el Workload Manager System (WMS). Lo que se intenta con esta aproximación es conseguir modificar la menor cantidad de componentes y simplemente crear un protocolo de intercambio de información entre los WMS's cooperantes como lo indica la Figura 5, garantizando de esta manera que la información del Grid Tipo A estará "replicada" en el Grid Tipo B para ser tenida en cuenta al momento de resolver la disponibilidad de recursos computacionales.

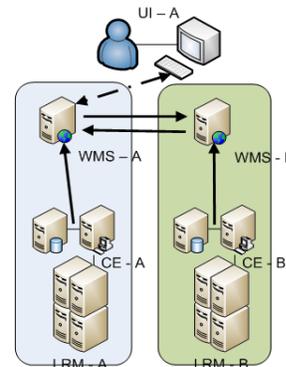


Figura. 5. En este esquema cada CE reporta a su respectivo balanceador de carga como normalmente lo harían, pero con la diferencia de que existe una comunicación entre los dos WMS para el intercambio de información

6. INTEROPERABILIDAD PARA EL USO DE LOS RECURSOS DESCUBIERTOS

Una vez se conoce la localización de los recursos y se sabe a qué tipo de sistema Grid pertenece, se continúa con la fase de uso de los recursos previamente localizados, para lograr lo anterior se existen las siguientes aproximaciones: Orientada al Usuario, Despliegue en Paralelo, Gateway, Adaptadores.

6.1 Orientado al Usuario

Se podría decir que el WMS genérico de grid logra la interoperabilidad entre sitios así sean del mismo tipo de middleware. Pero comparten información de los recursos y ubicación, usando más de uno y logrando interoperabilidad a nivel de nodos computacionales y de homogeneidad de middleware. Hay que tener en cuenta que para lograr este tipo de interoperabilidad es necesario un gran esfuerzo por parte de las organizaciones virtuales para solventar el problema de coordinación de recursos generales, por lo tanto puede resultar en trabajo duplicado por parte de los administradores de las VOs.

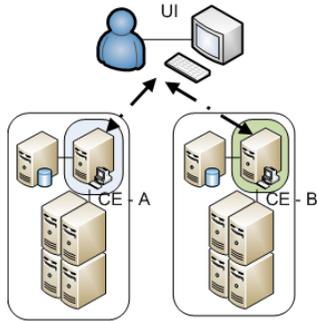


Figura. 6. Interoperabilidad para el uso de recursos descubiertos – Esquema Orientado al Usuario

6.2 Despliegue paralelo

La idea de este acercamiento es que cada sitio tenga desplegado en su interior ambos servicios de localización de Jobs para que el acceso a los recursos se pueda realizar a través de los dos gatekeepers de cada implementación. Hay que tener en cuenta que significa una sobrecarga a las instituciones debido a que el administrador se tiene que convertir en un experto en ambas implementaciones, haciendo que el despliegue de infraestructuras de este tipo se vuelva inmanejable si se quiere abarcar más servicios usando interoperabilidad. Recomendado solo para grandes centros de datos, no valdría la pena para centros pequeños debido a la anterior problemática de complejidad.

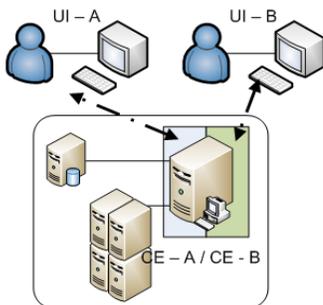


Figura. 7. Interoperabilidad para el uso de recursos descubiertos – Esquema de despliegue en paralelo

6.3 Gateway

Es un servicio “centralizado” que traduce los mensajes con el fin de crear una capa de transparencia entre un sistema multi Grid y otro, permitiendo que no haya que realizar cambios en el CE heterogéneo. Aquí hay un problema muy grave que es el cuello de botella que la infraestructura Grid de Tipo B es un solo recurso si hablamos gatekeepers. Sin embargo, este esquema se puede convertir en el expuesto previamente en la sección de

descubrimiento (WMS2WMS). De esta manera se solucionaría el cuello de botella para cada tipo de infraestructura.

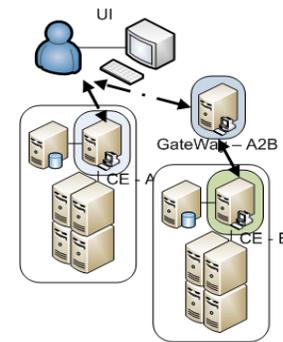


Figura. 8. Interoperabilidad para el uso de recursos descubiertos – Esquema de Gateway

6.4 Adaptadores

Los adaptadores van enfocados en parte en el balanceador de carga, pero también requiere esfuerzos en el nodo. Lo anterior en razón a que este “adaptador” debe conocer cuáles son las interfaces, protocolos y puntos de conexión de cada uno de servicios que se prestan. Los adaptadores tienen una ventaja significativa y son realmente útiles cuando se usan estándares, de lo contrario se vuelve una gran cantidad de trabajo teniendo en cuenta que cada uno se comunica e implementa diferentes protocolos de interacción y habría que desarrollar cada solución para cada middleware por aparte.

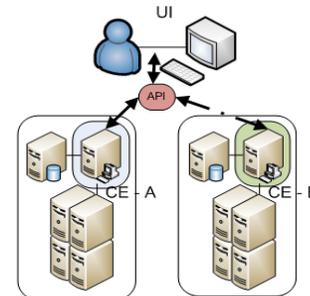


Figura. 9. Interoperabilidad para el uso de recursos descubiertos – Esquema de Adaptadores

7. CASOS DE ÉXITO

Teniendo en cuenta los anteriores modelos de interoperabilidad, tanto de descubrimiento como de uso de los recursos existentes, presentamos varios casos de éxito que consiguieron interoperabilidad parcial, estos proyectos han tenido éxito gracias a la creación grupos de trabajo colaborativo y la subdivisión de tareas imitando el esquema usado por OGF [3].

7.1 GLUE Schema

El proyecto GLUE nació como una iniciativa de OGF para crear un árbol de información estándar similar a la labor que cumpliría un LDAP. Uno de los logros más importantes de GLUE aparte de la unión de dos de los proyectos Grid más grandes del mundo es la creación de un esquema de descripción de información estándar de los recursos computacionales pertenecientes a cada nodo Grid[19]. El esquema de GLUE está basado en descripción de Entidades para la descripción de información y UML para la descripción grafica de dichas entidades, dentro de los atributos

más relevantes están: Localización, Recurso, Políticas, Contacto, Punto de Enlace y Manejador de Recursos.

7.2 Asia 2007

Esta iniciativa de interoperabilidad fue adelantada por los proyectos EGEE y NAREGI. El objetivo era la integración de gLite con lo que los Japoneses llaman SuperScheduler el cual es el planificador de que utiliza NAREGI, se podría asimilar con lo que normalmente se conoce con un CE [20].

Si bien este proyecto produjo un experimento exitoso, existen algunos problemas de diseño en cuanto a las fases tanto de descubrimiento como de uso de los recursos de SuperScheduler a gLite como de gLite a SuperScheduler. El inconveniente se da debido a que no hay igualdad en cuanto consumo de recursos computacionales como de la resolución de los mismos de un lado a otro. En este punto existe una contradicción de conceptos [2], el primero dice que los recursos deben ser compartidos homogéneamente y el segundo dice que las VO son independientes y autónomas de los usos de sus recursos.

7.3 Kebnekaise

Por último está el caso más reciente de éxito de interoperabilidad, el cual une 4 implementaciones de Globus ya existentes, las cuales son las más importantes de Europa que son: Advance Resource Connector (ARC), Lighthouse for Grid Computing (gLite), UNICORE, dCache[21]

Este proyecto liderado por el CERN usa GLUE y la aproximación WMS2WMS realizando un intercambio de información entre los balanceadores de carga, replicando los datos de los recursos de cada sistema Grid en los WMS respectivos. Creando un intercambio de información exitoso en los WMS respectivos. Creando un intercambio de recursos de manera exitosa

8. CONCLUSIONES Y TRABAJO FUTURO

La interoperabilidad es un campo de acción en los Grids computacionales, apuntando a uno de los principios el cual es la agrupación de recursos de manera transparente para el usuario.

Hacer que dos sistemas multi grid logren la interoperabilidad no es tarea trivial, sobre todo si se tiene en cuenta la gran carga que cae sobre los administradores de los centros de datos y los desarrolladores de cada una de las implementaciones asociadas a dichos centros. Pero si se logra crear un mecanismo de interoperabilidad sintáctica y semántica que tenga logre acoplarse a un middleware específico sin la necesidad de modificar los componentes del mismo la carga de los administradores y desarrolladores se reducirá.

El esquema de WMS2WMS es una buena manera de completar la fase de descubrimiento tal como lo vemos en el caso de éxito Kebnekaise. Sin embargo esta tarea también se podría delegar a un segundo componente que se encargue de mezclar la información de los sistemas multi grid tomando la información de los respectivos WMS que deseen interoperar.

En el caso del uso de los recursos se visualiza el problema típico de escalabilidad ya que las aproximaciones existentes no tienen en cuenta el gran dinamismo de los recursos, por lo tanto se ve la necesidad de crear una aproximación que abarque este problema y tome como entrada la fase de descubrimiento para de esta manera completar el ciclo de vida de un trabajo en la Grid.

Este estudio sirvió como base para crear un modelo que permita lograr la interoperación entre middlewares Grid teniendo en

cuenta los principios de trabajo conjunto y colaborativo y a la posibilidad de compartir más recursos computacionales con algunas Organizaciones Virtuales pequeñas que posiblemente no tengan acceso a gran cantidad de recursos como los que posiblemente pueda tener un gran laboratorio de investigación.

Como trabajo futuro se busca crear un modelo para ambos aspectos de la interoperabilidad (descubrimiento y uso) que cumpla con los principios del grid teniendo en cuenta los estándares definidos por la OGF.

9. REFERENCIAS

- [1] I. Foster, "What is the grid? A three point checklist", *Grid Today* 1 (9) www.mcs.anl.gov/~itf/Articles/WhatIsTheGrid.pdf, 2002.
- [2] I. Foster, C. Kesselman and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations", *International Journal of High Performance Computing Applications* 15 200_222, 2001.
- [3] I. Foster, W Johnston and C.Catlett, "Grid Global Forum Structure", Abril 2002.
- [4] Open Science Grid (OSG)- <https://opensciencegrid.org/>. revisión Marzo-2012.
- [5] Enabling Grids for E-science (EGEE)<http://public.eu-egee.org/>. revisión Mar-2012.
- [6] Deutsch Grid- D-Grid <http://www.d-grid.de/> revisión 3-2012
- [7] Grid Solution for Wide Area Computing and Data Handling (Nordugrid) <http://www.nordugrid.org>. revisión Mar-2012
- [8] Center for Grid Research and Development (NAREGI) <http://www.naregi.org>. Última revisión Mar 2012.
- [9] I. Foster, "A Globus Primer Or Everything You Wanted to Know about Globus, but Were Afraid To Ask, Describing Globus Toolkit Version 4", Draft, 2005.
- [10] I. Foster and Kesselman, C.: "The Grid: Blueprint for a New Computing Infrastructure" *Morgan Kaufmann*, San Francisco, 1999.
- [11] S. Bradner, "IETF Working Group Guidelines and Procedures," *RFC 2418*, Sept 1998.
- [12] S. Bradner, "The Internet Standards Process – Revision 3," *RFC 2026*, Oct 1996.
- [13] B. Bethwaite;D. Abramson;A. Buckle; , "Grid Interoperability: An Experiment in Bridging Grid Islands," eScience, 2008. eScience '08. *IEEE Fourth International Conference on*, vol., no., pp.590-596, 7-12 Dic. 2008.
- [14] ATIS Telecom Glossary 2012, <http://www.atis.org/glossary>. revisión Feb – 2012.
- [15] I. Foster, A. Savva, H. Kishimoto and T. Maguire, "Grid Service Architecture", *Second Revision, Open Grid Forum*, Jul 2006.
- [16] S. Tuecke, K. Czajkowski, I. Foster and J. Frey, "Open Grid Service Infrastructure", First Revision, Jun 2003.
- [17] A. Iosup, "Grid Inter-Operation Mechanisms, The CoreGRID European Virtual Institute on Grid Scheduling", 2008, pp :121 - 152.
- [18] L. Field and M. Schulz, "Grid Interoperability: The Interoperation Cookbook", *International Conference on Computing in High Energy and Nuclear Physics*, 2009.
- [19] S. Andreati, S. Burke, F. Ehm, L. Field and G. Galang, "Grid Laboratory for a Uniform Environment", *International Conference on Computing in High Energy and Nuclear Physics*, Mar. 2009.
- [20] H. Nakada, H Sato, M. Hatanaka, S. Matsuoka, "Job invocation interoperability between NAREGI Middleware Beta and gLite", 2008.
- [21] The EMI Grid Middleware, <http://www.eu-emi.eu/middleware>, revisión Mar 20.