

EELA: una infraestructura para e-ciencia en Latinoamérica

EELA: an Infrastructure for e-Science in Latin America

Recibido 26 de septiembre de 2008, modificado 31 de marzo de 2009, aprobado 3 de abril de 2009.

Harold Castro

Ph.D. Profesor Asociado, Departamento de Ingeniería de Sistemas, Universidad de los Andes, Bogotá D.C., Colombia.

hcastro@uniandes.edu.co ✉

Bernard Marechal

Ph.D. Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas CIEMAT. Universidad Complutense de Madrid. Madrid, España.

marechal@if.ufrj.br ✉

Diego Carvalho

M.Sc. Assistant professor. Department of Production Engineering, Federal Centre of Technology Education of Rio de Janeiro. Rio de Janeiro, Brazil.

dcarvalho@cefet-rj.br ✉

Leandro Ciuffo

Istituto Nazionale di Fisica Nucleare INFN. Catania, Italy.

leandro.ciufoo@ct.infn.it ✉

Inês Dutra

PhD. Departamento de Ciência de Computadores, Faculdade de Ciências, Universidade do Porto. Porto, Portugal.

ines@dcc.up.pt ✉

Philippe Gavillet

Ph.D. Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas CIEMAT. Universidad Complutense de Madrid. Madrid, España. CERN, Genève, Switzerland.

philippe.gavillet@cern.ch ✉

Rafael Mayo

Ph.D. Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas CIEMAT. Universidad Complutense de Madrid. Madrid, España.

rafael.mayo@ciemat.es ✉

Luis A. Núñez

Ph.D. Coordinador General, Consejo de Computación Académica de la Universidad de Los Andes. Director, Centro Nacional de Cálculo Científico CeCalCULA, Universidad de Los Andes. Mérida, Venezuela.

nunez@ula.ve ✉

Herbert Hoeger

Ph.D. Director, Centro de Simulación y Modelos CESIMO. Profesor, Universidad de los Andes. Mérida, Venezuela.

hhoeger@ula.ve ✉

Vanessa Hamar

M.Sc. Directora de Operaciones, Centro Nacional de Cálculo Científico CeCalCULA, Universidad de Los Andes. Mérida, Venezuela.

vanessa@ula.ve ✉

María José López

Periodista. PR Manager, CLARA, the Latin American Cooperation of Research Networks. Profesora, Universidad Andrés Bello. REUNA, Red Universitaria Nacional. Santiago de Chile, Chile.

mjlopez@reuna.cl ✉

PALABRAS CLAVES

e-Ciencia, EELA, e-infraestructura, *Grid*.

KEY WORDS

EELA, e-Infrastructure, e-Science, *Grid*.

RESUMEN

El proyecto EELA-2 (*E-science grid facility for Europe and Latin America*), cofinanciado por la Comisión Europea, busca consolidar la infraestructura iniciada por EELA (*E-infraestructure shared between Europe and Latin America*) para el desarrollo de la e-ciencia entre Europa y Latinoamérica a través del afianzamiento de iniciativas de computación en malla (*grid computing*) en ambos continentes. La nueva etapa de EELA (EELA2) pretende aprovechar ese camino recorrido para fortalecer tanto la infraestructura como las cooperaciones en curso, ampliando así el espectro de aplicaciones de impacto regional. EELA representa el mayor esfuerzo a nivel regional por contar con una infraestructura computacional de primer nivel, sostenible a largo plazo, para el desarrollo de la e-ciencia en nuestro continente.

ABSTRACT

The EELA-2 (*E-science grid facility for Europe and Latin America*) project, cofounded by the European Commission, aims at consolidating the infrastructure started by EELA (*E-infraestructure shared between Europe and Latin America*) for the development of e-science between Europe and Latin America by building a bridge between *grid* computing initiatives in both continents. The new phase of EELA (EELA2) takes advantage of the paved way to strength both infrastructure and ongoing cooperation to enhance a wider spectrum of applications with regional impact. EELA represents thus, the greatest effort in Latin America to develop a world class computational infrastructure, with long term sustainability, for the development of e-science in our continent.

INTRODUCCIÓN

Trabajar como investigador en un país de los llamados emergentes o en desarrollo no es fácil. En estos países es frecuente que las limitaciones presupuestales no permitan a las instituciones académicas y de investigación competir en igualdad de condiciones con las instituciones de los países desarrollados. Esta situación se hace evidente al comparar el número de patentes y de publicaciones científicas registradas en estos países, o al comparar los recursos computacionales disponibles en cada caso; si tomamos el ranking de los 500 supercomputadores más grandes del mundo, a junio del 2008 [1], nos encontramos una sola instalación latinoamericana en esta clasificación, que ocupa el puesto 138 (la situación africana es similar con un solo representante en el puesto 278). Aún cuando para el ranking de noviembre de 2008 [1] encontramos una mejor presencia de estos continentes (tres instalaciones en Latinoamérica en los puestos 226, 306 y 363, y una instalación en África en el puesto 128), se trata de esfuerzos aislados y sin ninguna garantía de consolidación. Una consecuencia bien conocida de esta situación es el fenómeno de la fuga de cerebros, ya que los investigadores se sentirán naturalmente más atraídos por aquellos sitios donde pueden desarrollar mejor su trabajo.

Hoy existe, sin embargo, una tecnología que comienza a jugar un rol importante en la provisión de recursos para miembros de una colaboración científica, sin importar su localización geográfica: la computación en malla o *grid computing*. Con la computación en malla, instituciones sin suficiente capacidad económica para poseer grandes equipamientos computacionales o científicos pueden hoy beneficiarse de infraestructuras distribuidas, organizadas alrededor de una malla computacional de escala mundial, para desarrollar sus actividades de investigación en las mismas condiciones que sus pares en países desarrollados.

Las aplicaciones de una infraestructura de este estilo son múltiples: desde el descubrimiento de nuevos fármacos hasta la previsión del tiempo, pasando por la biomedicina o los análisis vulcanológicos; éstos son

solo algunos ejemplos del potencial para una región como Latinoamérica.

La computación en malla ofrece a los investigadores de las instituciones participantes en la malla un marco de colaboración sin precedentes. Gracias a esta tecnología, los investigadores comparten de manera efectiva los distintos recursos a los que tienen acceso, en particular los recursos computacionales. De esta manera, los recursos particulares de un grupo de investigación dejan de marcar la diferencia a la hora de participar en un proyecto de investigación, lo que resulta de particular interés para aquellos investigadores que hasta ahora se veían en inferioridad de condiciones para alcanzar sus resultados. El desarrollo de la ciencia hoy en día está tan basado en este tipo de colaboraciones que se ha acuñado el término e-ciencia para describirla. *Teragrid* (www.teragrid.org), *OSG: Open Science Grid* (www.osg.org) y *EGEE: Enabling Grids for E-sciencE* (www.eu-egce.org) son los ejemplos más representativos de estas colaboraciones.

Es en este marco que presentamos el proyecto EELA-2, una infraestructura distribuida y compartida entre instituciones en Europa y Latinoamérica que ofrece el soporte para desarrollar una colaboración enriquecida de investigación entre los dos continentes.

ANTECEDENTES: EELA [2]

Desde finales de los 80, bajo el auspicio de la Comisión Europea, la federación de Redes Nacionales de Investigación y Educación (NREN por sus siglas en inglés) se constituyó en una red Paneuropea de alta velocidad con total soporte administrativo y operacional. La existencia de esta red ha facilitado el desarrollo de una infraestructura de *grid*, la cual provee acceso compartido a recursos distribuidos de cómputo y de almacenamiento para la comunidad europea de investigación, a través de distintos dominios geográficos. Este entorno integrado de red, almacenamiento y procesamiento se conoce como e-infraestructura

y provee una plataforma para el desarrollo de nuevos métodos de colaboración científica global: la e-ciencia; o más amplio: e-investigación. En Europa, las iniciativas líderes en el campo de las redes de alta velocidad con fines académicos y con e-infraestructura fueron los proyectos GEANT [3] y EGEE [4].

El objetivo del proyecto EELA [5] era construir, aprovechando el apoyo del programa SSA (*Specific Support Action*) de la Comisión Europea, una e-infraestructura latinoamericana con el mismo nivel de operación y calidad de servicios de las existentes en Europa para ese entonces. Con la estabilidad alcanzada a nivel de infraestructura de red gracias a los proyectos ALICE [6] y RedClara [7], el foco del proyecto EELA era generar un adelantamiento e infraestructura de *grid* y una cultura de e-ciencia basada en aplicaciones que requirieran los grupos de investigación de la región.

El corazón de la estrategia de EELA giraba alrededor del establecimiento de una red científica en el área de e-ciencia, para desarrollar y entrenar comunidades de investigación, así como la implantación de una infraestructura piloto de *grid* para la ejecución de aplicaciones de prueba de concepto. Una cantidad sustancial de actividades del proyecto se concentró en la disseminación y difusión de esta tecnología, la cual se concretó en la formación a través del desarrollo de talleres prácticos en colaboración con otros proyectos financiados por la Comisión Europea, tales como ALICE/GEANT, EGEE-II, EUCHINAGRID [8], EUMEDGRID [9], BALTICGRID [10] y SEE-GRID [11]. En paralelo, se desarrolló una infraestructura piloto de *grid* a lo largo de Latinoamérica, incluyendo el despliegue de clusters integrados en *grid* en aquellos lugares donde era posible, así como también la definición y aplicación de políticas y esquemas organizacionales y operacionales para el funcionamiento de esa plataforma. Por último, se promovió el desarrollo de aplicaciones regionales de e-ciencia y se desplegaron algunas de ellas en la e-infraestructura disponible para los miembros de la comunidad académica de América Latina.

Esta estrategia busca reducir la brecha digital que separa los países de los dos continentes, permitiendo así la participación de nuestros países latinoamericanos en actividades de nivel y calidad mundial en e-infraestructura y en e-ciencia.

Para Europa el resultado de EELA era ofrecer a los investigadores europeos la posibilidad de colaboraciones activas con institutos latinoamericanos en temas de interés común; un acceso rápido a una nueva, pero bien soportada, e-infraestructura con tasas de crecimiento garantizados. De esta manera, EELA contribuía a mejorar la investigación colaborativa entre los continentes al impulsar, de manera sustancial, el procesamiento de datos científicos desde su captura en los experimentos, pasando por los estudios de simulación, con lo cual se acelera la producción de resultados científicos. Adicionalmente, EELA fortaleció los lazos de cooperación entre comunidades científicas de los dos continentes al establecer y/o consolidar nuevas colaboraciones en nuevas áreas de aplicación.

En Latinoamérica, EELA construyó de manera exitosa una e-infraestructura al nivel de los *grids* europeos y se benefició de la experiencia y soporte de las instituciones europeas. Las comunidades científicas latinoamericanas adquirieron diferentes habilidades sobre este tipo de tecnología, lo cual les permitió autonomía para el soporte a largo plazo del *grid* latinoamericano. Este proyecto colaborativo a escala continental les permitió a nuestras comunidades mantener, cuando no incrementar, su participación activa en colaboraciones mundiales y cumplir con las condiciones exigidas en materia de recursos computacionales. Por último pero no menos importante, este proyecto contribuyó a amplificar la relevancia de las e-infraestructuras y allanar el camino para el desarrollo de nuevas iniciativas latinoamericanas de e-ciencia y/o iniciativas *grid* nacionales (NGI por sus siglas en inglés) en cooperación con las redes nacionales de investigación y educación, y con redClara.



EELA-2

EELA —calificado de “bueno a excelente”, la más alta valoración posible, por la Comisión Europea— se planteó construir un puente entre las nuevas iniciativas de e-Infraestructura en América Latina y las consolidadas en Europa, para crear una red de colaboración con el objeto de desplegar una amplia cartera de aplicaciones científicas sobre una infraestructura de *grid* piloto. En paralelo a la construcción de una infraestructura operativa, este proyecto se propuso trabajar en la formación, la difusión de conceptos y destrezas en tecnologías *Grid*, en 10 países y 21 instituciones académicas.

EELA-2 basa sus objetivos en el trabajo efectuado y los logros de EELA, busca proveer una facilidad de *grid* actualizada y una gama de servicios versátiles a través del soporte a aplicaciones, así como garantizar la sostenibilidad a largo plazo de la e-Infraestructura más allá de la duración del proyecto. EELA-2 continúa con las actividades de adiestramiento y divulgación y hace esfuerzos por captar nuevos grupos de investigación e instituciones que puedan ser potenciales usuarios y beneficiarios de la tecnología. De igual modo, labora en el desarrollo de herramientas que faciliten el uso del *grid* y trabaja con las autoridades



regionales con el fin de que la e-Infraestructura, su operación, su uso y su propagación perduren una vez culminado el proyecto.

El proyecto EELA-2 tiene por objeto el establecimiento de una facilidad *grid* de gran capacidad, de calidad y escalable, que provea acceso desde cualquier parte del mundo a computación distribuida, a almacenamiento y a la red de recursos necesarios por varios países europeos y de América Latina para sus colaboraciones científicas.

La e-infraestructura proporcionada permitirá a muchos investigadores hacer frente a importantes problemas científicos y regionales que no podían ser atacados antes, debido a la falta de recursos. EELA-2 también está aumentando significativamente la cooperación entre sus asociados, así como con otros países e instituciones.

El Consorcio EELA-2 está integrado por 14 países, 5 de Europa y 9 de América Latina, e involucra más de 50 instituciones. Unas 47 aplicaciones en unos 10 campos científicos diferentes están siendo apoyadas por el proyecto (<http://applications.eu-eela.eu>).

EELA-2 está estrechamente vinculada a EGEE, *Enabling Grids for E-sciencE*, pues es su extensión en América Latina. Se beneficia de los servicios y productos de EGEE, como por ejemplo del *middleware gLite*, y ha adoptado el mismo proceso de certificación y calidad de servicios para operación de su infraestructura. Se comparte el desarrollo de la infraestructura y aplicaciones *middleware* de interés y uso común, se crea una red de repositorios de conocimientos federados con los de EGEE y se organiza un conjunto de eventos de difusión y capacitación. EELA-2 está fortaleciendo su colaboración con EGEE y con otros proyectos con objetivos similares, tales como BalticGrid2 y SEE-GRID2. Entre los ámbitos de colaboración se dan los siguientes:

- Portar la Interfaz de Usuario y el Elemento de Computo de *gLite* a la plataforma *Microsoft Windows*.

- Simplificar el acceso a la infraestructura para los nuevos usuarios y aplicaciones.
- Facilitar la creación de archivos digitales de datos y data *grid* frameworks.
- Almacenamiento de datos de manera segura, y solución al problema de abuso de información privilegiada.
- Optimización de la operación y monitoreo entre proyectos por medio de más herramientas comunes, dando a conocer las mejores prácticas y reforzando la calidad y la seguridad.
- Apoyo común a Organizaciones Virtuales (VO) para reducir al mínimo los esfuerzos.

El proyecto EELA-2 está abierto a cualquier organización dispuesta a colaborar. Al convertirse una institución en asociada del proyecto, sus grupos de investigación pueden usar el Grid de EELA-2 para ejecutar aplicaciones y compartir los resultados científicos. Por otra parte, la organización tendrá derecho a recibir apoyo técnico a fin de integrar cualquiera de sus recursos al grid o a tener sus aplicaciones desplegadas y en operación dentro de éste. Se han organizado distintos eventos durante la vida del proyecto, tales como Tutoriales, Escuelas de Grid, Talleres y Conferencias, de los cuales se puede obtener información en [12].

LA JRU COLOMBIA

EELA-2 recoge todas las experiencias de EELA, en particular, en la manera de organizarse para la ejecución de los servicios. Ya que es un proyecto con 53 instituciones participantes que quiere extender su alcance durante sus dos años de vida, necesita una administración particular. Para ello, retoma el modelo europeo de Unidades Nacionales de Investigación Grupal (JRU por sus siglas en inglés) que son implementaciones de iniciativas de *grid* nacionales (NGI). De esta manera, las instituciones de un mismo país se agrupan bajo esta figura y seleccionan un coordinador nacional de

la JRU para que los represente ante el Consorcio del proyecto. EELA-2 tiene entonces ya conformadas y avaladas por alguna institución gubernamental del país nueve (9) JRUs. En Colombia, la JRU la conforman la Universidad de los Andes, la Universidad Industrial de Santander y la Universidad de Antioquia, siendo la primera la coordinadora para el proyecto.

La JRU Colombia nace del esfuerzo previo de una iniciativa *grid* nacional llamada *Grid* Colombia. La iniciativa *Grid* Colombia ha logrado generar la articulación y compromiso de varias universidades con un fin común: la conformación del *grid* nacional. *Grid* Colombia busca constituirse en la primera alternativa en computación distribuida en el país; así mismo, que a mediano plazo todas las universidades conectadas a la red nacional de alta tecnología RENATA[13], que estén desarrollando proyectos de computación distribuida, hagan parte de *Grid* Colombia y así ser partícipes como nación de la comunidad *Grid* a nivel internacional.

EELA-2 es, entonces, un impulso importantísimo a esta iniciativa nacional y se deben articular esfuerzos para lograr el objetivo común de mantener una e-infraestructura en la cual los científicos encuentren los recursos y los socios para el desarrollo de la e-ciencia.

CONCLUSIONES

Las actividades de EELA en sus dos fases están cambiando la perspectiva sobre la e-ciencia en Latinoamérica, pues “construye” una red humana gracias al impulso a las colaboraciones entre investigadores de los dos continentes.

Gracias a EELA, se dispone hoy de un piloto maduro de infraestructura *grid* utilizable por grupos de colaboración científicas de uno y otro continente, que resuelve problemas de interés para ambas comunidades. La expansión, tanto en recursos computacionales como humanos (objetos de la segunda fase de EELA), generará una comunidad científica inmersa en igualdad de condiciones en el contexto científico mundial.

Los aprendizajes y avances de la primera fase dejan un baremo muy alto para la nueva etapa de EELA. La consolidación del esfuerzo para lograr su autosostenibilidad en el futuro es el punto crucial de este proyecto. Por eso, en la nueva estrategia se plantean reuniones de trabajo con los tomadores de decisiones en el mayor número posible de países del consorcio. Gobierno, academia e industria deben participar en el mantenimiento de esta e-infraestructura; pero, para ello, deben hallar valor real en ella.

Entre las diversas actividades de difusión de EELA-2 se destacan sus dos grandes conferencias, una a mitad de proyecto y otra al final. Colombia fue elegida para la primera de estas conferencias: la Universidad de los Andes fue la anfitriona de este importante evento los días 25, 26 y 27 de Febrero de 2009. Se presentaron los mejores resultados surgidos de este tipo de colaboraciones y se aprovechó el espacio para discutir políticas de sostenibilidad a largo plazo en nuestra región. Reuniones del más alto nivel son necesarias para asegurar que no quedemos rezagados del contexto científico mundial. “

REFERENCIAS BIBLIOGRÁFICAS

- [1] “Top 500 Supercomputer Sites”. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.top500.org/>
- [2] **B. Marechal, P. Rausch, D. Carvalho.**
“Building a Grid in Latin America: The EELA Project e-Infrastructure”. *Proceedings of the Seventh IEEE International Symposium on Cluster Computing and the Grid LAGrid07*. September, 2007, pp. 835-839.
- [3] “Geant”. Dante. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.geant.net/>
- [4] “EGGE Enabling Grids for E-sciencE”. HealthGrid. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.eu-egee.org>
- [5] “EELA E-science grid facility for Europe and Latin America”. Philippe Gavillet. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.eu-eela.org>
- [6] “Alice”. Dante. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://alice.dante.net/>
- [7] “CLARA. Cooperación Latinoamericana de Redes Avanzadas”. Gerencia de Relaciones Públicas y Comunicaciones CLARA. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.redclara.net/en/03.htm>
- [8] “Euchinagrid”. Interconnection & Interoperability of Grids between Europe & China. Consortium GARR. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.euchinagrid.org/>
- [9] “Eumed Grid, empowering eScience across the Mediterranean”. Consortium GARR. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.eumedgrid.org/>
- [10] “BalticGrid-II Project”. BG-II Consortium. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.balticgrid.org>
- [11] “SEE-GRID”, South Eastern European GRID – enabled eInfrastructure Development. SEE-Grid2. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.see-grid.eu/>
- [12] “EELA Home”. CERN. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://indico.eu-eela.eu>
- [13] “RENATA, Red Nacional Académica de Tecnología Avanzada Colombia”. Astrolabio Imagen Web Ltda. RENATA. Fecha de consulta: 26 de septiembre de 2008. Disponible en: <http://www.renata.edu.co/>