



gisela BOLETÍN

Grid Initiatives for e-Science virtual
communities in Europe and Latin America



2 Editorial

¡Vamos a hacerlo!

4 Especial:

- Grid para bioinformática: el reto de la región

11 Tocando el primer año

- Itacurucá: punto de encuentro
- Acuerdos
- Nuevo usuario
- Nueva versión de OurGrid
- Divulgando

13 Entrevista

- La computación distribuida como servicio en América Latina

17 Dos países con paso seguro en la e-ciencia

- Una prioridad para Argentina
- Autoridad colombiana

Boletín N°2

Octubre 2011

<http://www.gisela-grid.eu/>

[twitter](#) @gisela_grid

Coordinación General

Herbert Hoeger

Trabajo periodístico

Ysabel Briceño

Diseño y Diagramación

María Eugenia Hernández

Traducción

Alicia Bohórquez



FRANCISCO BRASILEIRO
WPS MANAGER

¡VAMOS A HACERLO!

La infraestructura de Grid de **GISELA** acaba de cumplir un año y ha madurado como una importante fuente de recursos de computación. Más de 1.700 núcleos están ahí, disponibles inmediatamente para los usuarios de **GISELA**. Sin embargo, parece que estamos frente a un problema que se vivió también en los días de EELA-2: no hay suficiente gente tomando ventaja de la infraestructura. Como representante de los proveedores de la infraestructura insto a los usuarios de todas las comunidades virtuales de investigación a hacerlo, después de todo, como dice la letra de la canción de Porter “incluso las perezosas medusas lo hacen...” Por lo tanto, “... ¡enamórenos!”.

Por supuesto, el cortejo requiere esfuerzos de ambas partes. Nosotros, los proveedores, hemos dado algunos pasos concretos en la dirección de atraer a los usuarios a la infraestructura. En particular, el objetivo principal del paquete de trabajo 6 (WP6) – Servicios Orientados a Infraestructura y Aplicaciones para las Comunidades de Usuarios - es facilitarle a los usuarios el poder ejecutar sus aplicaciones en la infraestructura. Esto se ha procurado de muchas maneras.

En primer lugar, hemos adaptado la mayor parte de los servicios que fueron heredados por el proyecto EELA-2 para que puedan ser utilizados en la infraestructura de **GISELA**, con las últimas versiones del middleware de Grid que se han desplegado. Estos son servicios que facilitan el proceso de migración de una aplicación para ejecutarse en el Grid, y también servicios que pueden facilitar la incorporación de recursos a la infraestructura.

En segundo lugar, hemos tratado de hacer frente a una de las principales quejas de los usuarios, que es la dificultad de adaptar sus aplicaciones al entorno de ejecución que se encuentra en el Grid.

El middleware DIRAC tiene como objetivo mitigar este problema. Algunos usuarios ya se benefician de las facilidades ofrecidas por DIRAC y el equipo de apoyo de DIRAC está dispuesto a ayudar a muchos más.

Por último, también hemos abordado otro punto de preocupación, que es el manejo de grandes lotes de trabajos, comúnmente llamadas aplicaciones de bolsa de tareas (bag-of-tasks application). Esta es una clase muy común de las aplicaciones. Ambos DIRAC y OurGrid cuentan con facilidades que se pueden utilizar para realizar un seguimiento de la ejecución de grandes lotes de trabajos, asegurando que los trabajos truncados en el lote se vuelvan a enviar de forma automática y transparente. Otra característica implementada en OurGrid es el seguimiento de los trabajos lentos que retardan la conclusión del lote. Este es un problema común en los grids, debido a la heterogeneidad inherente de los recursos que los componen.

También hemos desarrollado un script que puede ser utilizado para ejecutar la bolsa de tareas usando el middleware OurGrid sobre los recursos de gLite. Esto es particularmente útil para aquellas aplicaciones cuyos lotes se componen de tareas de larga ejecución (decenas de horas). Estas no son adecuadas para ser ejecutadas en el entorno oportunista sobre el que normalmente se utiliza el middleware OurGrid.

En resumen, ha habido mucha atención al desarrollo de servicios que pueden ayudar a los usuarios a tener una experiencia mucho más productiva con una infraestructura de grid. Además, hay un equipo de apoyo que está a la espera para ayudar a los usuarios a beneficiarse de la infraestructura de GISELA. Así que, ¡vengan en gran número, estaremos encantados de ayudarlos a todos ustedes!

GRID PARA BIOINFORMÁTICA: EL GRAN RETO DE LA REGIÓN

La secuenciación y otras actividades de avanzada en la biología, como por ejemplo la predicción de genes, ensamblaje del genoma y la simulación de procesos evolutivos requieren hoy día de la computación para la gestión y análisis de esos miles de millones de datos que resultan de esta parte de la ciencia.

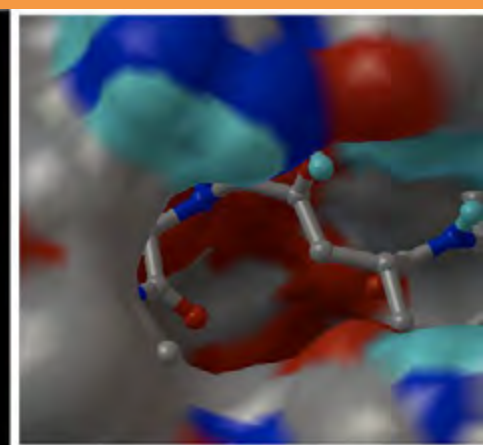
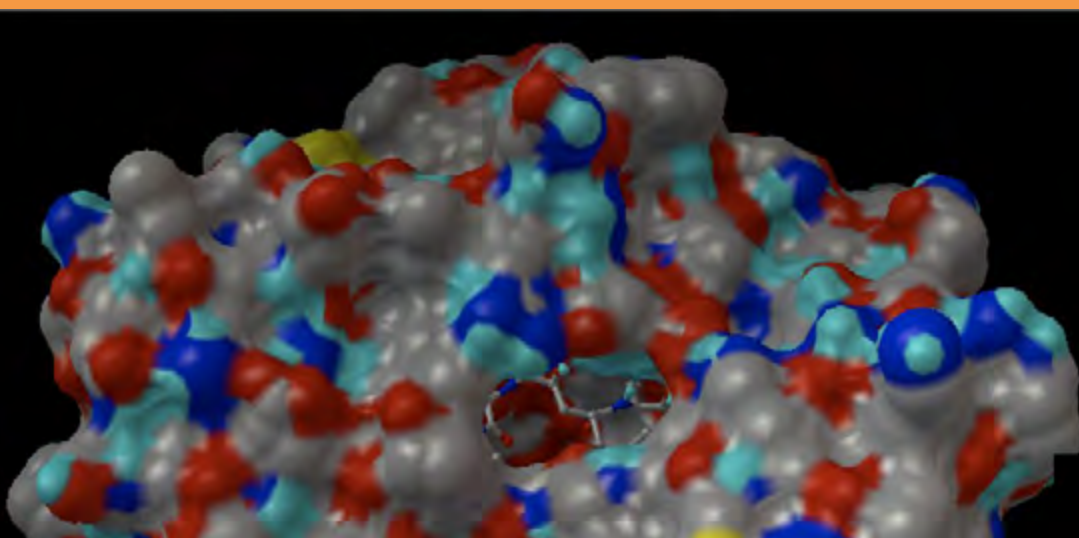
Una de las áreas de investigación que tiende a producir crecientes cantidades de datos es la relacionada con las ciencias biológicas. Luego de develarse los secretos de la doble hélice del ADN, el gran reto de dilucidar el genoma humano ha derivado en una alta exigencia del manejo de información. Miles de bases se encuentran presente hasta en el organismo más sencillo, por lo que la lectura del ADN ha requerido también esfuerzos tecnológicos para lograr la velocidad y exactitud que se requieren en este tipo de experimentos.

Ese conjunto de técnicas y métodos bioquímicos ha contado efectivamente con los beneficios de la computación para lograr soluciones que por su tiempo y costos, son llamadas de alto rendimiento. La secuenciación y otras actividades de avanzada en la biología, como por ejemplo la predicción de genes, ensamblaje del genoma y la simulación de procesos evolutivos requieren hoy día de la computación para la gestión y análisis

de esos miles de millones de datos que resultan de esta parte de la ciencia, en un área de trabajo denominada bioinformática.

En los proyectos de investigación que involucran estas actividades se exige una capacidad de cálculo y almacenamiento con la que muchas veces no se cuenta en los laboratorios de América Latina por lo que la infraestructura Grid ha resultado sensible de ser abordada desde esta área.

Durante el período de gestión de la Grid en América Latina (con los proyectos EELA, EELA2 y, más recientemente **GISELA**), investigadores del área de bioinformática conjuntamente con equipos técnicos, han ocupado experiencias que permitieron desarrollar o adaptar aplicaciones a la infraestructura disponible, con matices diversos, cuyos avances más exitosos se originaron en los laboratorios de Brasil, México, Colombia y España, como país socio europeo.





Como herencia, **GISELA** cuenta con una decena de aplicaciones en el área de bioinformática que han sido corridas en la plataforma Grid, específicas para el uso de algoritmos heurísticos, uso de álgebra lineal numérica, biología molecular y genómica, generadas durante un proceso de pruebas y aprendizaje que implicó la detección de necesidades, el desarrollo o adaptación de las aplicaciones, el entrenamiento de técnicos e investigadores, hasta lograr el servicio final de cómputo de alto rendimiento para esta área.

➔ La experiencia en espera de seguidores

* **DistBLAST (Distributed BLAST):**

Muchos de los métodos computacionales usados en genómica usan BLAST, una herramienta básica usada para comparar secuencias biológicas. BLAST es una herramienta muy eficiente pero usa grandes bases de datos y búsquedas de entradas. Dado que los laboratorios de bioinformática de la región medio oeste de Brasil necesitaban ejecutar BLAST con muchas secuencias biológicas, varias veces al día, DistBLAST fue propuesto durante el proyecto EELA para acelerar este proceso a los investigadores. Se aspiraba que los resultados con DistBlast se produjeran en un tiempo más corto, lo cual podría mejorar el rendimiento de BLAST. La Universidad de Brasilia (UnB) fue la instancia responsable de iniciar el uso de esta aplicación.

María Emilia T. Walter, investigadora del grupo de bioinformática de la UnB, comenta que “la principal ventaja de la plataforma Grid es que permite tratar grandes volúmenes de datos, mediante herramientas computacionales de bioinformática ampliamente usadas”.

T. Walter explica que existe la intención de continuar su gridificación en esta plataforma, con otro middleware. “DistBLAST fue desarrollada con OurGrid y queremos hacerlo ahora con gLite. Estas herramientas son muy útiles para los proyectos de investigación en genómica, por lo que queremos realizar otra gridificación para que en esta región de Brasil puedas usarlas”.



Maria Emilia Walter (UnB): “Queremos continuar la gridificación de esta herramienta”.

DistBlast



* **BioNMF-Grid:**

BioNMF es una herramienta Web que aplica la metodología de Factorización No Negativa de Matrices (Non-negative Matrix Factorization–NMF-) en diferentes contextos de análisis para soportar algunas de las aplicaciones más importantes en Biología.



Explica Alberto Pascual, miembro del Grupo de Bioinformática de la Universidad Complutense de Madrid (UCM), que esta aplicación web es de uso libre y anónimo (todos los trabajos se lanzan empleando el mismo usuario del grid) y ejecuta un algoritmo con grandes necesidades de potencia de cálculo. La aplicación fue gridificada para el proyecto EELA-2 y se desarrolló para aumentar la productividad de la versión “cluster local” de la UCM. “Antes del proceso de gridificación, intentamos instalar un grid privado en nuestro laboratorio para realizar pruebas, pero dada la escasísima documentación de gLite 3.1 existente en ese momento (especialmente en cuanto a instalación y configuración), finalmente se optó por emplear directamente la infraestructura de producción de EELA-2 mediante el User Interface (UI) ‘público’ del Ciemat (España)...El proceso de implementación no fue excesivamente complicado, pues habíamos asistido a cursos de manejo a nivel de usuario de gLite 3. Uno de estos cursos fue a distancia utilizando la infraestructura de GILDA”.

Esta aplicación se mantiene activa actualmente en la versión de cluster local de la UCM, pero con una media de trabajos limitados. Según Pascual, “la tecnología Grid tiene mucho que aportar a aplicaciones como bioNMF”, por lo que se aspira disponer de la infraestructura Grid adecuada, ya que varios equipos científicos han solicitado la utilización de una versión sin límites.



Alberto Pascual (UCM), “La tecnología Grid tiene mucho que aportar a aplicaciones como BioNMF”.

BioNMF-Grid



* META-Dock:

El propósito de esta aplicación es el de proporcionar un método de filtrado basado en Grid para estudios farmacéuticos. Cuando una proteína que causa una enfermedad es conocida, el paso a seguir es el de identificar una molécula que sea capaz de prevenir la acción de esta proteína. Como método alternativo para estos propósitos se ha llegado a usar la simulación *in silico* por medio de un programa de acoplamiento molecular. Inicialmente se propuso usar AutoDock, un programa con buen reconocimiento en el mundo académico, pero la idea es usar más de un programa de acoplamiento molecular. La instancia responsable del uso inicial de la aplicación fue la Universidad Nacional Autónoma de México (UNAM).

Explica Jérôme Verleyen, investigador enlace de esta herramienta, que su gridificación ocurrió durante el proyecto EELA-2. “Utilizamos muchas herramientas desarrolladas por parte del proyecto para incluirlas en el programa. El paso inicial para la gridificación de MetaDock fue acceder a recursos importantes permitiendo un estudio más amplio de docking de proteínas. La gridificación se llevó a cabo sin mayor problema, durante una de las escuelas de grid. Allí pude integrar la mayoría de las herramientas disponibles en la Grid, con tutores de muy buen nivel. Eso me permitió tener una visión general de esas herramientas, y me sigue sirviendo hasta ahora.”



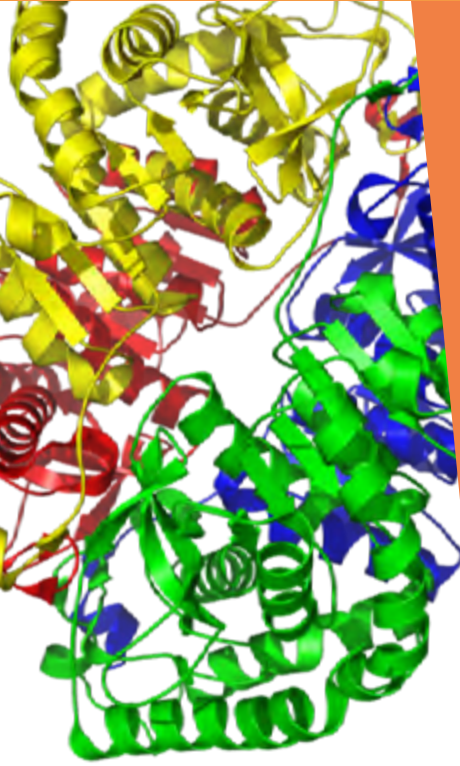
Jérôme Verleyen (UNAM) busca grupos interesados en el desarrollo de META-Dock



Verleyen es entusiasta con el posible uso de esta aplicación en la plataforma Grid, por lo que la propone a la comunidad de bioinformática, interesada en esta área. “Me gustaría encontrar grupos interesados para seguir en el desarrollo de esta aplicación”.

- ¿Crees que esta aplicación podría ser adaptada a otras áreas?

- Creo que sí, el esquema de uso de esta herramienta corresponde a otras herramientas. No me viene por ahora en la mente cuál exactamente, pero estudiar a lo grueso un conjunto de datos para elegir los mejores es algo común en la ciencia.



* GrEMBOSS

EMBOSS (“European Molecular Biology Open Software Suite”) desde EMBnet (European Molecular Biology Network) es un proyecto de paquete de software de análisis de código abierto desarrollado para bioinformáticos. Esta gridificación resultó de un proyecto conjunto de tres instituciones de la Universidad Nacional Autónoma de México (UNAM): La Dirección General de Servicios de Cómputo Académico (DGSCA), el Centro de Ciencias Genómicas (CCG), y el Instituto de Biotecnología (IBt). EMBOSS integra alrededor de 150 paquetes y herramientas para el análisis de secuencias de genoma en un conjunto perfectamente integrado. Con EMBOSS, los investigadores pueden hacer alineamiento de secuencias; búsqueda rápida de base de datos con patrones de secuencia; identificación de patrón de proteínas; análisis de patrones de secuencia de nucleótidos; identificación rápida de patrones de secuencia en larga escala; y usar herramientas de presentación para publicación.

“EMBOSS es un paquete de herramientas de análisis de secuencias genómicas muy usado y desarrollado para apoyar el trabajo de la comunidad bioinformática —explica Romualdo Zayas-Lagunas, investigador de la UNAM-. La primera vez que se portó EMBOSS sobre una infraestructura GRID fue en la Primera Escuela de Grid de EELA en Itacurucá (Brasil) en diciembre de 2005. Posteriormente, se presentó el proyecto en la 3ra Conferencia EELA en Catania 2007. Finalmente, GrEMBOSS fue el proyecto de la Universidad Nacional Autónoma de México que se incluyó en EELA-2”.

Zayas-Lagunas explica las ventajas de esta herramienta usada sobre la plataforma Grid:

GrEMBOSS fue construida pensando en que los usuarios no tienen que preocuparse por saber cómo construir un archivo JDL. Incluye un script que construye un archivo y hace un paquete con aquellos archivos necesarios para la ejecución. La recuperación de los datos es automática y el usuario no se entera del lugar donde se ejecutó su análisis.

“Los grupos que trabajan con bioinformática podrían estar interesados en GrEMBOSS para análisis de secuencias. Desde nuestro punto de vista y tomando en cuenta el buen diseño de EMBOSS es posible pensar en “gridificar” otras aplicaciones bioinformáticas sin mayor requisito que seguir los estándares de desarrollo de software de EMBOSS”.



Romualdo Zayas-Lagunas:
GrEMBOSS para el análisis
de secuencia



Los retos

Los avances con la computación de alto rendimiento no han sido nada desdeñables: hoy día estas herramientas pueden activar mecanismos que procesan millones de secuencias a la vez, sin que tenga que pararse el mundo para esperar los resultados. Sin embargo, el reto aumenta más rápido que las posibilidades: la cantidad de datos generada en la secuenciación de organismos ha sido de tal magnitud en los últimos años que aún no puede afirmarse con solvencia si alguna vez se llegarán a resolver todos los problemas de cálculo con las técnicas actuales de computación.

Para Rafael Mayo, investigador del CIEMAT y miembro del proyecto **GISELA**, “hoy en día hay problemas que el Grid no puede resolver satisfactoriamente dada la gran cantidad de resultados que se generan. El tamaño es tan grande que se pierde muchísimo más tiempo en traer esos resultados de vuelta al ordenador del investigador que el tiempo de cálculo real”.

Nuevas soluciones y avances van creando nuevos problemas en esto de la gran producción

* jModelTest

La aplicación ModelTest y su versión en Java jModelTest fueron desarrolladas por el Catedrático de la Universidad de Vigo David Posada (<http://darwin.uvigo.es>) y tienen más de 30000 usuarios registrados en todo el mundo, siendo básicas para hacer estudios filogenéticos, pues proporcionan el modelo de sustitución de nucleótidos que mejor se ajusta a un organismo. Su gridificación fue generada recientemente desde el Centro de Investigaciones Energéticas Medioambientales y Tecnológicas de Madrid (CIEMAT). Cualquier científico del mundo con interés en el ámbito filogenético o en investigaciones que requieran de modelos de sustitución de nucleótidos es un potencial usuario de esta aplicación.

Comenta Manuel Rodríguez, investigador del CIEMAT y enlace responsable de esta herramienta, que consideraron un gran avance poder portar al Grid JModelTest para que estuviera accesible a los investigadores por medio de un cálculo distribuido en lugar de secuencial. “Al haber una nueva versión distribuida que sustituye a la secuencial, es posible obtener resultados más rápidos”. El trabajo previo de paralelización de tareas se realizó en el Centro de Supercomputación de Galicia (CESGA).



Manuel Rodríguez: “Con una versión distribuida es posible obtener resultados más rápidos”

La aplicación se encuentra en producción y disponible para su uso. “Ahora mismo se van a ejecutar investigaciones en la infraestructura de GISELA. Un equipo de desarrollo del CIEMAT se encuentra trabajando en una nueva versión con una planificación más eficiente de ejecución y en un portal web para el envío de trabajos de forma transparente”.

JModelTest



de datos científicos. Como ejemplo, la Física de Altas Energías encabeza el reto mayor: el personal de la Organización Europea para la Investigación Nuclear (CERN, por sus siglas en francés) genera ingentes esfuerzos en mejorar la transferencia de grandes volúmenes de datos debido a todas las medidas y cálculos que se están produciendo con el Gran Colisionador de Hadrones (LHC).

En el caso de la Grid para América Latina, el esfuerzo por atender las dificultades en el manejo de grandes volúmenes de datos también debe ser una constante. Pasada la fase de pruebas en la gridificación de las herramientas, **GISELA** apunta a la conformación de comunidades de usuarios estables. Durante la experiencia de uso de las aplicaciones gridificadas, las propuestas han servido de

base para proyectos académicos muy puntuales. En este momento es necesario empezar a recoger los frutos de este aprendizaje.

Hacia una infraestructura fiable:

Como principales limitaciones planteadas por los usuarios de bioinformática, se asoma la necesidad de una infraestructura fiable, bien documentada y con un buen sistema de planificación de trabajos. Proveer un servicio solvente ha sido el mayor reto.

En este sentido, Francisco Brasileiro, responsable del área de Servicios Orientados a Infraestructura y Aplicaciones para las Comunidades de Usuarios del proyecto **GISELA**,



Francisco Brasileiro: "Trabajamos para optimizar las capacidades de las cientos de máquinas en los laboratorios de computación"

resume la forma como se ha ido avanzando para hacerle frente a las principales dificultades del servicio de computación avanzada, bajo la plataforma Grid en América Latina, siendo una de ellas la adaptación de las aplicaciones al entorno de ejecución. "El middleware DIRAC tiene como objetivo mitigar este problema. Además, nuestros colegas de UNIANDES, con Harold Castro a la cabeza, han seguido un enfoque diferente para simplificar el despliegue de aplicaciones en el Grid. Ellos son los responsables del desarrollo de un nuevo servicio, llamado Cluster Virtual Personalizado, o CVC por su significado en inglés Customized Virtual Cluster. El CVC utiliza herramientas de virtualización como el VMware, que permite tomar ventaja de las capacidades de decenas o cientos de máquinas en los laboratorios de computación. CVC se encuentra en pruebas beta en UNIANDES, y los primeros que quieran experimentar con CVC son bienvenidos".

En torno al manejo eficiente de grandes lotes de trabajo, Brasileiro comenta que DIRAC y OurGrid cuentan con facilidades para este objetivo. "El planificador de OurGrid utiliza una táctica de replicación simple que clona los trabajos lentos, mitigando las asignaciones desafortunadas que las lleva a ejecu-

Los grupos de trabajo de **GISELA** encargados de soporte a comunidades e infraestructura y aplicaciones, conjuntamente con la Universidad del Valle de Colombia generaron recientemente una colaboración fructífera para hacer correr en la plataforma Grid herramientas que permiten el manejo de la información química con altos requerimientos computacionales.

Gaussian ha sido exitosamente integrado a la plataforma Grid, con lo cual los investigadores de la Universidad del Valle podrán predecir parámetros de Resonancia Magnética Nuclear (RMN) con un solo clic y contar con herramientas poderosas y amigables para el uso racional de los recursos disponibles en Europa y América Latina, gracias a una interfaz denominada **myLims.org**.

El sistema **LIMS** (por sus siglas en inglés: Laboratory Information Management System), permite almacenar, manipular y compartir virtualmente cualquier tipo de datos experimentales para predecir propiedades moleculares, de una forma tal que los usuarios podrán añadir funciones e integrar sus propios programas para predecir más propiedades de estructura, espectro o cualquier información ya disponible del sistema.

Los procesos creados por LIMS se soportan en DIRAC para el envío de los trabajos al Grid. Los pasos contemplan el control del entorno de ejecución y la instalación del software necesario para ejecutar las tareas generadas por la plataforma; después de varias pruebas, los resultados (los archivos de salida) de las tareas recuperadas serían enviados de vuelta a la plataforma LIMS para ser presentados al usuario final en forma fácil y amigable.

Esta iniciativa Grid ha sido apoyada por la Comisión Europea, RENATA, CNPq, CEFET/RJ y se inició gracias a la colaboración con el proyecto **W-eNMR**.

tarse más lento que los demás”. Toda la información sobre el portafolio de servicios se encuentra disponible en el Sitio Web del proyecto **GISELA** (actividades del Grupo de Trabajo 6 –WP6-)

Grid como alternativa:

Otro de los grandes retos para el uso de la plataforma Grid, asomada por los usuarios de bioinformática, es la tendencia creciente al uso de la tecnología Cloud. Dos conceptos básicos se oponen en estas dos propuestas: Grid es un esquema colaborativo, de acceso libre a los recursos, y Cloud es una propuesta de computación distribuida por sistema de pago. Ante la idea de considerar a **GISELA** como alternativa en la región frente al cloud computing de pago, Philippe Gavillet, coordinador adjunto del proyecto acude a una categórica afirmación: “La respuesta es sí. **GISELA** es una potente e-infraestructura. Contamos con más de seis años de experiencia en el tema de la e-



Philippe Gavillet, coordinador adjunto de GISELA: “somos una potente e-infraestructura”.

ciencia en América Latina, con proyectos financiados por Europa y hemos estado presentes en casi todos los países de esta región. Somos una e-infraestructura de gran cobertura en la colaboración científica. Actualmente, el tema de Cloud está sobre la mesa y existen muchas variaciones alrededor: pública o privada, gratuita o de pago. En este sentido, necesariamente vamos a tener que enfrentar una revisión. Estamos preparando una solución que se adapte al contexto de América Latina”.

GISELA garantiza a los científicos el acceso adecuado a los recursos de computación distribuida, apoya en la adaptación de las aplicaciones científicas a la infraestructura Grid, desde su desarrollo hasta la ejecución, y soporta el uso de la e-infraestructura y servicios relacionados con las aplicaciones. Se proporciona actualmente una capacidad de almacenamiento de aproximadamente 60 TB por medio de 18 centros de recursos para apoyar a los grupos de investigación científica y aspiramos seguir creciendo.

Para mayor información, los interesados de la comunidad de bioinformática de América Latina pueden comunicarse con **wp3@gisela-grid.eu**

➔ Algunos trabajos publicados sobre Grid y Bio

bioNMF: a web-based tool for nonnegative matrix factorization in biology.

Nucleic Acids Research 2008 v.36, Oxford University Press 2008.

Estimating Conductivity Distribution of Transmural Wedges of the Ventricle Using Parallel Genetic Algorithms.

MARTINS, Daves Marcio Silva; XAVIER, Carol Ribeiro; SANTOS, Elisa Portes dos; VIEIRA, Vinícius da Fonseca; OLIVEIRA, Rafael Sachetto; SANTOS, Rodrigo Weber dos, In: Computers in Cardiology 2006, Valencia (Spain), Computers in Cardiology v.33, p.49-52, 2006.

IntegraEPI: a Grid-based Epidemic Surveillance System. In: HealthGrid,

Geneva (Switzerland), Proceedings of the HealthGrid Conference 2007.

InterProScan: protein domains identifier.

E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez,. Nucleic Acids Research v.33, Oxford Journals 2005.

Managing structural genomic workflows using Web services. Maria Claudia Cavalcanti et al., Data & Knowledge Engineering v.53 Issue 1, p.45-74, Elsevier 2005.

Performance Tests of GAMOS Software on EELA-2 Infrastructure. In: 2nd EELA-2 Conference, Choróní (Venezuela), Proceedings of The Second EELA-2 Conference ISBN 978-84-7834-627-1, p.379-385, Editorial CIEMAT 2009.



TOCANDO EL PRIMER AÑO

Durante el segundo semestre del Proyecto GISELA, las actividades se centraron en definir acuerdos de acción conjunta entre los diferentes grupos de trabajo, abordar actualizaciones tecnológicas, promover la plataforma Grid y, principalmente, revisar caminos posibles ante el gran reto de transferir a RedCLARA un modelo de computación distribuida disponible a las comunidades de investigación de la región.

Acuerdos

En abril, GISELA formó parte de los proyectos presentes en el EGI User Forum 2011 (Vilnius, Lituania) y, aprovechando la ocasión, formalizó acuerdos con EGI.eu, EGI.InSPIRE y e-Science-Talk, un proyecto europeo de divulgación en tendencias de e-Infraestructura.

El acuerdo con EGI.eu y el proyecto EGI.InSPIRE ampliará la colaboración entre GISELA y el proyecto insignia de Grid en Europa. Y el acuerdo con e-ScienceTalk aspira el trabajo conjunto con GISELA en la divulgación de las aplicaciones Grid en Europa y América Latina, haciendo uso de los mecanismos como GridCafé, GridCast, iSGTW and GridGuide, además de otras iniciativas que pudieran surgir para posibilitar la comprensión del público sobre el uso de la infraestructura Grid.



Itacuruçá: punto de encuentro

A finales de marzo los miembros de GISELA revisaron los avances de la principal iniciativa regional de infraestructura de Grid en América Latina

➔ Nuevo usuario

Durante el mes de junio el Centro de Modelado Computacional –CMC- de La Universidad del Zulia -LUZ- manifestó el interés en incorporarse como usuario Grid y compartir recursos en América Latina, para lo cual se iniciaron las primeras conversaciones en Venezuela para definir los detalles técnicos y de protocolo de la inserción de este centro en el proyecto GISELA.

El CMC es una institución con sede en Maracaibo (Venezuela) constituida por investigadores de áreas muy diversas y transdisciplinarias y que presta servicios en el modelado, simulación y visualización científicos de fenómenos físicos, meteorológicos, químicos, biológicos, médicos, económicos, entre otras áreas.

Según anunció Gilberto Díaz, gerente del grupo de trabajo encargado de servicios de redes en GISELA, este centro ya ha venido trabajando con aplicaciones de computación avanzada en el área de pronósticos climáticos. “La intención es incorporar a esa comunidad a GISELA y al Grid de Venezuela”.

En el próximo paso se aspira la firma de un acuerdo de entendimiento para incorporar las aplicaciones de este centro al proyecto GISELA y crear un centro de recursos Grid en LUZ.





y su respectiva apropiación por parte de las comunidades de investigación. Itacuruçá (Brasil) fue el punto de encuentro en el que durante tres días los grupos de trabajo del proyecto revisaron aspectos de servicios, redes, apoyo a las comunidades, disseminación y entrenamiento en Grid.


Gran parte de la jornada de la reunión se dedicó a revisar el esquema de trabajo propuesto para transferir el modelo Grid a América Latina, con una visión de sostenibilidad; esta tarea, en la que **RedCLARA** cumple un rol central, se distribuye entre los miembros de un grupo de enlace con responsabilidades técnicas y de organización para construir un modelo de servicios de computación distribuida en la región, siendo éste el principal reto del proyecto.

Todos destacaron el avance importante en esta reunión al poder definirse mejores mecanismos de trabajo conjunto entre los miembros de **GISELA**, cuyos grupos provienen de diversos países de América Latina y Europa. Tres días de dura jornada de encuentro y trabajo presencial demostraron la necesidad de discutir las distintas percepciones sobre la forma de llevar adelante los objetivos del proyecto y encontrar soluciones comunes.

➔ Nueva versión de OurGrid

En agosto, **GISELA** anunció el lanzamiento de la nueva versión del middleware OurGrid, un desarrollo regional que en el marco de EELA, EELA2 y **GISELA** ha sido utilizado por cientos de usuarios para acelerar la ejecución de aplicaciones en paralelo, cuyas tareas no se comuniquen entre sí.

El middleware OurGrid fue desarrollado en el año 2004 y desde entonces ha sido actualizado en diversas versiones que han ido mejorando su funcionamiento, generándose una comunidad activa de usuarios y desarrolladores alrededor de este software estable, libre y de código abierto. Bajo la lógica de OurGrid, los recursos informáticos y de almacenamiento son proporcionados por participantes que actúan en red para compartir y obtener el máximo de rendimiento, proporcional a los aportes de cada quien.


Hasta ahora OurGrid cuenta con cinco versiones desde su primer desarrollo. Esta última versión (4.2.4), optimizada, ya se encuentra disponible acá 



➔ Divulgando

Desde el mes de mayo, **GISELA** distribuye un resumen mensual informativo, con breves notas sobre el avance del proyecto y de las tendencias en la infraestructura Grid que inciden en América Latina.

El resumen es recibido por una red de países en la región, gracias al apoyo de la Gerencia de Comunicación de **RedCLARA** y algunas instituciones como el **CUDI** (México) y **Renata** (Colombia), cuyos aportes han favorecido la distribución de las noticias Grid en América Latina, impulsada desde **GISELA**.

También se ha contado con la iniciativa europea de divulgación de Grid: **International Science Grid this Week (iSGTW)**, desde donde han publicado permanentemente las notas derivadas del resumen informativo de **GISELA**, enlazadas a su sitio de divulgación 



LA COMPUTACIÓN DISTRIBUIDA COMO SERVICIO EN AMÉRICA LATINA

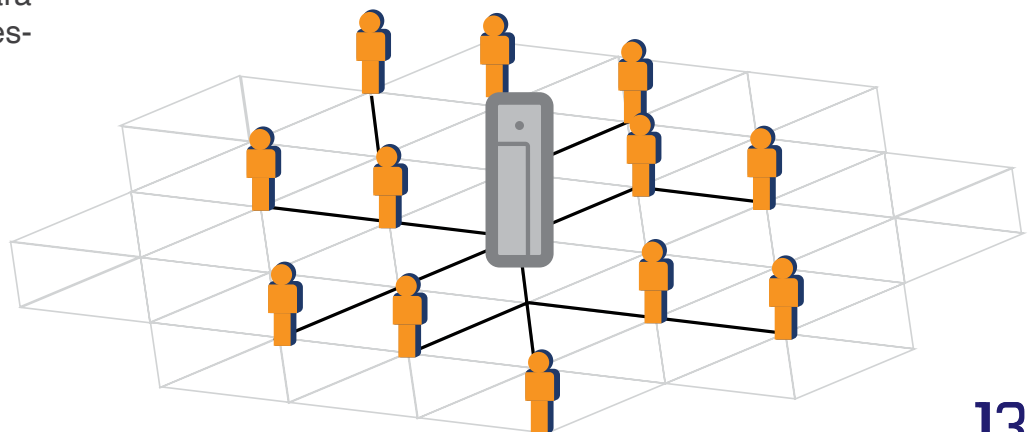
El nuevo modelo de servicios de Computación Avanzada (CA) en América Latina busca construir un puente entre la e-Infraestructura consolidada y aquellas iniciativas emergentes. Como reto, se plantea crear capacidades regionales, en igualdad de condiciones para los países. Para ello, las redes de colaboración, el entrenamiento y la sostenibilidad de las comunidades virtuales de investigación son el blanco ideal. La industria y el comercio también se incluyen dentro las posibilidades del servicio.

Salma Jalife y Luis Núñez son los encargados de proponer el Plan de Negocios para el servicio de cómputo distribuido en América Latina, luego de la experiencia de levantar la infraestructura Grid en la región. El principal propósito es definir una estrategia viable para asegurar una transición sin tropiezos del actual modelo de GISELA (en e-infraestructura y apoyo a las comunidades de investigación) a un modelo de RedCLARA. Como escenario ideal post GISELA, se aspira integrar los servicios Grid a los servicios generales de Computación Avanzada entre las redes nacionales de América Latina, con el apoyo de RedCLARA. GISELA los entrevistó especialmente para esta edición, a un año de la gestión del proyecto.

¿Cuál ha sido el principal reto en la elaboración de la propuesta del Plan de Negocios para el servicio de cómputo distribuido en América Latina?



SJ: Los diversos países miembros de RedCLARA tienen matices diferentes en su madurez técnica, por lo que plantearse cualquier nuevo servicio, de tal forma que sea sostenible, es un reto. En este caso, son 11 de los 15 países miembros de RedCLARA los que participan en GISELA y el reto es mayor porque existe en estos países niveles muy diferentes en el desarrollo de la infraestructura Grid y en las capacidades que han logrado desarrollar los diferentes representantes de los Centros de Recursos. Además, aún no se ha logrado contar con una masa crítica de investigadores, empresarios y usuarios en general que tengan necesidades de transportar sus aplicaciones a una infraestructura de Grid. Ahora el reto será incluir los servicios de Grid en un paquete de servicios de cómputo avanzado utilizando el modelo de CLARA.



¿Qué ha hecho al respecto RedCLARA, en este primer año de GISELA?



LN: Durante este primer año de GISELA, los representantes de RedCLARA se han concentrado en sensibilizar a los decisores regionales y nacionales en la importancia de apoyar la creación y el funcionamiento de la e-Infraestructura y, en general, de fomentar el desarrollo de la e-Ciencia en América Latina. Algunas iniciativas se han llevado a cabo con éxito. Por ejemplo, RedCLARA ha apoyado a comunidades virtuales de investigación, de donde se derivan los potenciales usuarios de los servicios de Computación Avanzada en la región, siendo uno de ellos la computación Grid. CLARA a través de sus redes nacionales, ha apoyado la creación de un grupo sólido conformado por instituciones nacionales que han adquirido diferentes iniciativas nacionales en Grid como puede ser la National Grid Initiative (NGI) o la Equivalent Domestic Grid Structure (EDGS). Por ejemplo, Colombia y Ecuador han conformado su NGI, México cuenta con un Grid Operation Center (GOC) regional, y Argentina ha manifestado su intención de crear un grupo de e-Ciencia que será responsable, entre otros, del desarrollo de la infraestructura de Grid nacional.



Luis Núñez: La tecnología Grid es una etapa en la evolución de la computación de alto rendimiento

La propuesta inicial era adaptar en América Latina el modelo desarrollado en Europa en el servicio Grid. ¿Ha sido viable?



LN: Según este modelo, cada país debe organizar una Iniciativa Nacional de Grid, principal encargado de impulsar los servicios de Computación Avanzada; las instituciones académicas y de investigación deberían formar parte de las NGI de cada país y ser responsables de establecer las normas para integrar la e-Infraestructura, coordinar y promover el uso de servicios entre las comunidades virtuales de investigación. Pero para poder llevar a cabo con éxito este modelo, se requiere que cada país cuente efectivamente con una iniciativa nacional bien establecida, que las redes nacionales sean sostenibles y cuenten ya con una madurez institucional y organizacional para el desarrollo de proyectos de colaboración, integrando a sus comunidades en una propuesta de e-Ciencia, soportada en políticas públicas. Desafortunadamente, este escenario que, efectivamente ha sido alcanzado en Europa tras años



de colaboración entre los gobiernos, las redes nacionales y las comunidades de investigación, tiene aún un largo camino por recorrer en América Latina. Salvo pocas excepciones, los países han construido sus redes nacionales con esquemas organizativos diferentes y todavía sigue siendo un reto importante convencer a los gobiernos que las redes avanzadas son parte del Estado y que pueden brindar aportes en el desarrollo de sus países. Tomando en cuenta la realidad latinoamericana, el estado actual de las redes nacionales y su organización no permite en estos momentos que el modelo europeo se adapte plenamente, por lo que CLARA diseñó un modelo más flexible, desagregando algunos servicios que apunten a

crear espacios de sustentabilidad. Estamos proponiendo crear servicios de cómputo puntuales, que permitan ser configurados por el usuario por un tiempo determinado. Servicios de cómputo en nube (cloud computing) que puedan ser contratados tanto por los grupos de investigación como por las redes nacionales. La computación Grid debe ser una palanca, una excusa para promover servicios de Computación Avanzada en la región.



Salma Jalife: el reto es mayor porque existen niveles diferentes en el desarrollo de la infraestructura Grid

¿Entonces, cuál es el camino que se ha dibujado en este primer año de discusión sobre los servicios de Grid en la región?



SJ: RedCLARA ha evaluado de diferentes maneras la forma de introducir los servicios Grid en la región bajo una arquitectura de varios países. Después de una serie de discusiones de grupo con el equipo de transición de CLARA (TT) y la interacción con el Gerente de Marketing CLARA, se decidió construir un Plan de Negocios que fuese viable no sólo para un solo servicio, sino también para un conjunto de servicios que tienen varios aspectos en común. El razonamiento detrás de esta decisión es que las comunidades de investigación aún no han logrado una masa crítica para consumir los servicios Grid, a excepción de la comunidad de físicos de altas energías. Otras comunidades están aprendiendo dónde y cómo hacer uso efectivo de la e-Infraestructura y correr algunas aplicaciones. Por otra parte, con el lanzamiento comercial de los servicios Cloud, algunas comunidades están explorando combinaciones diferentes de la computación de alto rendimiento y técnicas de almacenamiento para resolver sus necesidades. Hasta ahora, el Plan de Negocios define un grupo de Servicios de Computación Avanzada, en el que se incluye el servicio de Grid y el uso de la e-Infraestructura.

¿Balance hasta ahora?

LN: No es una tarea fácil manejar tales diferencias en toda América Latina. Sólo aquellos países que tienen la capacidad de interconectar recursos y poner a disposición cientos de CPUs serán capaces de formar una iniciativa nacional en Grid. Los países que no pueden alcanzar este tipo de organización deben ser tratados de manera diferente y se necesita encontrar una solución común. Ante esto, RedCLARA propone una estructura similar a la que actualmente caracteriza sus actividades con otros servicios. En este caso, una Red Regional de Centro de Operaciones arroparía a aquellos países que no tienen un grid nacional. Progresivamente, tras el crecimiento de las infraestructuras de las redes nacionales, éstas empezarían a ser capaces de instalar sus propios centros de recursos. Una vez hecho esto, se transferirían las funciones a las redes nacionales. Ahora bien, la cosa se complica con lo efímero de estas tecnologías avanzadas. La tecnología Grid es una etapa en la evolución de la computación de alto rendimiento. No podemos pensar en organizar servicios académicos centrándonos en una tecnología que puede llegar a cambiar. Tenemos que pensar en organizar servicios de cómputo para los grupos de investigación, que no estén centrados en una tecnología en particular.



¿Quiere decir que, más allá de una discusión puntual de transferencia de servicios Grid, se trata de seguir definiendo mecanismos para madurar el entramado regional en redes avanzadas para la investigación?

LN: RedCLARA tiene que encontrar la manera más eficiente para operar en la región, con un presupuesto mínimo y un máximo de beneficio para sus miembros, sin sacrificar la calidad de los servicios. La colaboración y la transferencia de tecnología son los principales activos de sus miembros. En los últimos años RedCLARA se embarcó en un ambicioso proyecto para apoyar en la construcción de las redes nacionales, a partir del fortalecimiento de comunidades de investigación específicas. Las redes de Brasil y México fueron las principales aliadas en este proyecto. Los organismos nacionales responsables de las políticas públicas en Ciencia y Tecnología, Educación Superior y de Telecomunicaciones juegan un papel específico en el desarrollo de redes avanzadas. Algunos países han respondido muy positivamente. Sin embargo, todavía hay un hueco grande que los responsables políticos deben ocupar en la idea de integrar las redes avanzadas como una herramienta para la innovación y desarrollo tecnológico de cada país.



La infraestructura de Grid y los servicios de Computación Avanzada (CA) podrían seguir el mismo proceso que otros servicios de RedCLARA. Hasta ahora, se evalúa la siguiente propuesta de operaciones:

Los países que pueden apoyar la creación de un GOC (Centro de Operaciones de GRID) tendrán uno, y los que no puedan asumirlo o no estén suficientemente entrenados como para crear uno, puede utilizar el Centro de Operaciones GRID Regional (ROC). CLARA licitaría el funcionamiento de la NOC (Network Operation Centre: Centro de Operaciones de la red) regional cada cuatro años, y la NREN (National Research & Education Network: Red Nacional de Investigación y educación) que proporcione la mayor cantidad de personal a la operación, entre otros requisitos, sería la que opere la RNOC durante ese período. Se debe crear un nuevo grupo técnico de CA, el cual puede seguir la evolución de los servicios.

Los GOCs nacionales serán responsables de sus servicios nacionales e interactuarían con el ROC cuando los servicios tengan un alcance regional. El ROC se hará cargo de las funciones domésticas de aquellos países que no tienen GOC. RedCLARA utilizaría un mínimo de personal para la prestación de los servicios del ROC, ya que la mayoría de la prestación de servicios se mantendrá bajo los RC (centros de recursos) distribuidos de los diferentes países participantes. Las instituciones o los organismos nacionales de Ciencia y Tecnología, según sea el caso, deberían cubrir los costos de operación. Al término de GISELA, el TT CLARA tomaría posesión y RedCLARA tendrá que subastar la operación del ROC entre las instituciones participantes.

DOS PAÍSES CON PASO SEGURO EN LA E-CIENCIA

Durante el segundo semestre del proyecto **GISELA**, América Latina recibió dos buenas noticias que involucran avances en la e-Ciencia, tanto en Colombia como en Argentina, dos países que reúnen esfuerzos para consolidar su infraestructura técnica y de organización en el uso académico de las redes avanzadas.

➔ Una prioridad para Argentina

Una serie de universidades y centros de investigación en Argentina firmaron un acuerdo para establecer un consorcio nacional de e-Ciencia, con el fin de promover la Grid en la investigación y educación, tanto en instituciones públicas y privadas, bajo un modelo inspirado en el proyecto **GISELA**.

La red de colaboración, coordinada por INNOVA-T, busca impulsar el desarrollo de la infraestructura Grid argentina, así como aumentar el conocimiento de la comunidad científica, industria, gobierno y otros sectores de la sociedad sobre las aplicaciones potenciales y ventajas de las tecnologías asociadas a la computación distribuida.

Esta iniciativa refuerza los objetivos del Sistema Nacional de Computación de Alto Desempeño (SNCAD), impulsado por el Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina y el Consejo Interinstitucional de Ciencia y Tecnología para consolidar en el país suramericano el trabajo conjunto que permita satisfacer la creciente demanda de la comunidad científica y tecnológica en las áreas de almacenamiento, grid, de alto desempeño y de alto rendimiento, de visualización y otras tecnologías emergentes.



**Sistema Nacional de
Computación de Alto Desempeño**



➔ Autoridad colombiana

Colombia se incorporó a la lista de países de América latina que tienen autoridades de certificación asociadas a los servicios Grid. La autoridad colombiana de certificación será la Universidad de los Andes de Bogotá –Uniandes-, instancia ya acreditada por la organización que vela por las políticas y normativas de todas las autoridades de certificación Grid –IGTF- (por sus siglas en inglés International Grid Trust Federation).

La autoridad de certificación permite avalar en forma confiable la suscripción de los usuarios y servidores en el Grid, dentro de una exigente política de seguridad que garantiza directrices comunes en el uso de esta plataforma. En adelante, los interesados colombianos en usar la plataforma Grid contarán con la gestión de certificación expedida dentro de su propio país, por medio de la Uniandes.

Para ser reconocido como autoridad de certificación, las instancias nacionales deben pasar por un largo proceso protocolar que revisa exhaustivamente requerimientos técnicos, de infraestructura, de personal y ético. En América Latina los países que cuentan con autoridad nacional de certificación Grid son: Argentina, Brasil, Chile, México Venezuela, y ahora Colombia.



Autoridad colombiana de certificación



IGTF



gisela

tailored to the needs of Latin America

A large amount of computers and storage provided by the project partners, is now available for groups of scientists working on problems that demand high quantities of computing resources, that without this e-infrastructure would be difficult to solve.

<http://www.gisela-grid.eu/> 