ZERO-IN
BUILDING INSIGHTS, BREAKING BOUNDARIES
ISSUE 3 - OCTOBER 2009

SUCCESSFUL CASE STUDIES OF eSCIENCE/eRESEARCH PROJECTS IN EUROPE AND GLOBALLY

e-INFRASTRUCTURES FOR CULTURE, ARTS & HUMANITIES APPLICATIONS

STANDARDS

e-INFRASTRUCTURE TECH

e-INFRASTRUCTURES FOR SUSTAINABILITY STRATEGIES AND POLICIES

e-INFRASTRUCTURES FOR DEVELOPMENT

ZERO-IN: BUILDING INSIGHTS, BREAKING BOUNDARIES

SUCCESSFUL CASE STUDIES OF eSCIENCE/eRESEARCH PROJECTS IN EUROPE AND GLOBALLY

3

Review Panel

Dear Readers,

Welcome to this third edition of Zero-In! I hope you find the contributions as inspiring as I do. We focus this time on e-Research: a fundamental e-Infrastructures application, not only for researchers themselves but for policy-makers who increasingly see it as a platform for enhanced competitiveness. In e-Research we tend to think of the technology that facilitates our collaborative research across global communities of practice. This technology focus, although fundamental, tends to overshadow two key prerequisites: people and law. In the article about Epikh, we see training exchanges between Africa, Latin America and the EU that promote grids as a means for developing e-Science applications and skills in the future. As well as people, the law is key. In 'ERIC' we see a proposal for a new legal instrument for research infrastructures. The community needs to discuss this and implement it as soon as possible: realising ERIC will uncork the innovative talent bottled up by legal uncertainty in sustained e-Research. How far it be applied globally is the next, very serious question to tackle.

As ever there is a wide range of rich articles from many research domains in this edition. Splash, for instance, is an exciting example of e-Infrastructures. In fact, the very technology that has helped us discover our history beneath the waves can also prevent more of humanity from disappearing under rising sea levels in the future by enabling climate modelling and forecasting.

Get reading…and be inspired!

Stephen Benians
BELIEF-II Project Coordinator

METAFOR: MANAGING METADATA FOR CLIMATE MODELS

By Sarah Callaghan, Reinhard Budich, Gerry Devine, Eric Guilyardi, Bryan Lawrence and Sophie Valcke, Metafor Project, United Kingdom

Climate model outputs are increasingly important, not only to climate scientists, but also to those concerned with the impact of climate change on disease, water resources, forest fires and fish stocks, including researchers, policymakers and the enlightened public.

Identifying, accessing and using the climate data stored in digital repositories requires accurate and complete metadata, or "data which describes the data". Unfortunately, metadata is not always complete or understandable, and hence there is a need to standardise the description of climate data and the models that produce it.

EASY AS CIM

The Metafor project is funded by the EU's 7th Framework Programme and aims to develop a Common Information Model (CIM) to document and compare climate models. As modellers know, discriminating between two simulations is not easy, even if you were responsible for creating and running them both. Documentation currently revolves around runtime, but not the scientific detail or relevance of the model components. To counter this, the CIM exists to document the "simulation context and models", otherwise known as the whys, wherefores and issues associated with any particular simulation.

The core Metafor team consists of eleven international partners from Europe and the USA, with additional collaborators from the US-led Curator project. As Metafor progresses, more climate modelling groups will become involved in the development and governance of the CIM, and via interaction with the Coupled Model Intercomparison Project Phase 5 (CMIP5) metadata questionnaire.

THE CMIP5 METADATA QUESTIONNAIRE

The Working Group on Coupled Modelling(WGCM), via the Coupled Model Intercomparison Project (CMIP) panel, gave Metafor the task of developing a questionnaire to define and collect the model and experiment metadata for CMIP5, to be used for the next Intergovernmental Panel on Climate Change assessment, due in 2013. CMIP5 will have a globally distributed data archive using technology primarily developed by the U.S. Earth System Grid (ESG) team, with chief responsibility for the archive held by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). The core archives are expected to comprise around 750 Terabytes of model-run data, hence the need to capture all the metadata to allow users to differentiate between and compare the experiments, and better understand the climate models used. To develop the questionnaire, the Metafor team has been collecting controlled vocabulary from domain experts; this collection will also be used to populate the CIM with a standard structure. The questionnaire is web-based and will collect information from the CMIP5 modelling groups, including details of the models used, how the simulations were carried out, how the models conform to CMIP5 experiment requirements, and what hardware or platforms were used.

SPINNING OFF THE RESULTS

The questionnaire results will be made available to anyone interested in the metadata, whether or not they plan to use the CIM. The controlled vocabulary generated during the development of the questionnaire will also be spun-off and governed independently from the CIM, providing a valuable resource for the climate modeling community. The metadata collected by the questionnaire will be used by ESG and stored in the Metafor CIM archive.

Climate modeling is a complex process, so the CIM has been designed to be highly generic and flexible. The information provided in each CIM instance will vary according to the model and parameters described, but the CIM will store this information in a standard way and thus enable the development of tools to search for and compare CIM documents, and hence different climate models.

INTEGRATION AND COLLABORATION

Metafor does not exist in isolation. A key aim of the project is to integrate with other larger projects such as GEOSS and INSPIRE, and to collaborate with other complementary projects alongside ESG, such as Curator, IS-ENES, GENESI-DR and EuroVO-AIDA. Partnership with technology providers such as OGC GML projects and UK NERC DataGrid will ensure the development of new technology that will use the CIM to benefit users.

Metafor is halfway through its three years of operation and its project team has developed into a dedicated and tightly organised group of experts. This team have proposed a CIM development strategy, including a conceptual level and meta-model, and released CIM v1.0, which is freely available for comment and criticism at http://metaforclimate.eu/trac/browser/CIM. Excellent international collaborations and links have been established with US colleagues and there is strong buy-in by members of the climate modelling community, as demonstrated by Metafor's key contribution to the CMIP5 metadata collection.

Further information about the Metafor project and the CMIP5 questionnaire can be found at http://metaforclimate.eu

7

GRID INFRASTRUCTURES IN EUROPEAN RESEARCH: LESSONS FROM EGEE

By Danielle Venton, Enabling Grids for E-SciencE/CERN, Switzerland

Budgets limit research, and computing budgets are often considered last. Grids can help alleviate these cash restraints by pooling distributed resources, thus spreading the computing and cost load. Grids also allow contributing countries to retain their existing resources, rather than contribute new resources to a centralized location.

The grid computing solution grew from the needs of the high energy physics (HEP) community and has been adopted by scientists in scores of disciplines. User communities have often invested much of their own energy and resources to make grid computing a success, convinced of its value for their work and as a framework for collaboration. For example, in forming the Worldwide LHC Computing Grid (WLCG) – a global collaboration of more than 140 computing centres in 33 countries – the HEP community showed that grids:

- are an economical way to meet data-management needs (avoiding reapplication to funding agencies);
- promote collaboration and cooperation, since members receive recognition for contributing resources;
- allow public participation, as seen in volunteer computing initiatives such as LHC@home.

Another example is the grid computing infrastructure managed by EGEE, an e-Infrastructure that allows global and secure access to processing power, data, software and data storage. EGEE's software is scalable, dynamic, and can be extended as required. In the final year of its third phase, EGEE has set its eyes on sustainability, technology transfer and integration with cloud services.

IN THE CLOUDS

Cloud computing makes heavy use of virtualization and grid computing may be able to learn from this approach, enhancing user experience by simplifying operations and building friendlier user interfaces (see "Clouds and grids: evolution or revolution?").

Increasingly, "proof-of-concept" developments are investigating ways of integrating cloud-based services with existing grid-enabled applications across a number of disciplines. Projects such as RESERVOIR and StratusLab have shown such integration can be positive, but the transition to everyday production requires more work. Potential benefits include simpler application porting and reduced operational costs.

Cloud computing currently entails a number of issues. For example, if resources are located in the United States – the home of many cloud providers – the U.S. government will have access to users' data. This may

not be acceptable for many international scientific projects. Also, if a cloud provider goes out of business, access to hosted data is not necessarily guaranteed. Transferring data between clouds is also difficult.

TAILORING COMPUTE SOLUTIONS

Most user communities are primarily interested in accessing easy-to-use, powerful and secure data management facilities, without concern for how these facilities are provided. Since each style of e-Infrastructure has advantages and drawbacks, users should be able to select solutions that best suit their needs. Hence commercially operated clouds can be a good solution for users who need additional resources on demand, while users who operate complicated applications may require sophisticated and expensive high-performance computer resources, such as supercomputers. (To opt for supercomputing, users must be prepared for significant financial and engineering investments, since matching their software to the resource's architecture is generally no small task.)

The interoperability work EGEE has undertaken with supercomputing structures such as DEISA, volunteer grids such as EDGeS, and cloud systems such as DigitalRibbon has been driven by the needs of users who wish to access a range of resources and systems; such work will surely gain importance in the future. Standards bodies such as Open Grid Forum are also working to simplify interoperability between clouds, grids and supercomputer installations.

LESSONED LEARNED: THE PLANNING STAGE

When building an e-Infrastructure, the planning stage is most important, since it lays the foundation for the rest of the project.

Early planning of the networking layer is critical. In the WLCG, for example, institutions are arranged in "tiers" corresponding to their computing responsibilities; network capacity must match the tier hierarchy. Achieving this required much negotiation and joint work between National Research and Educational Networks (NRENs) and the GEANT project.

Common data formats are important to enable data sharing. If possible, collaborations should adopt existing data standards, which are more likely to have open source and commercial product support. Otherwise, projects must agree on a common set of data formats.

Practice makes perfect: a project should gradually deploy and test its infrastructure. For example, the WLCG organised a series of "service challenges" held over a number of years, each with defined and successively more demanding quality of service targets.

LESSONS LEARNED: TRAINING

Spreading knowledge about new techniques and technologies is key to unifying and extending the community.

Hands-on tutorials and interactive eLearning events, each tailored to specific needs, are more effective than generic presentations.

Outreach to young researchers, who are often unaware of new research tool, is particularly important.

Training courses are a great opportunity to port home-grown applications; prior to training, many scientists are not aware of the advantages grid technology can bring to their research.

LESSONS LEARNED: TECHNOLOGY'S HUMAN DIMENSION

The grid is not flat: even if all sites are technically equal, they have varying levels of importance to each user community. Importance grows if a site

- is run by colleagues,
- is reliable,
- holds important data sources,
- has important capacity (the WLCG, for example, has defined a hierarchy of tiers).

People want to help: collaborators are their own strongest allies. Initially the WLCG planned on having just a handful of Tier-1 centres; now it has more than 100 sites, and the resources provided by Tier-2s outweigh that of Tier-1s, at nearly 60%.

Grids empower contributors, turning them into stakeholders instead of spectators. Partners take pride in contributing to a project and making it work. The same phenomenon can be seen with volunteer grids such as BOINC's SETI@home.

Working through cultural differences is an interesting and fundamental process for all international collaborations. For example, the WLCG agreed to publish the state of grid sites, reaching consensus from all collaborators, even those cultures in which transparency, which can be interpreted as criticism, is frowned upon.

CONCLUSIONS

As science changes, e-Infrastructures must also adapt. Since the benefits of grid computing parallel the requirements of many research collaborations, grids will continue to be an effective way to meet data management needs. EGEE has built the world's largest grid for research science and is the foundational cornerstone of the European Grid Initiative, which will sustain this grid infrastructure for the European Research Area.

The primary value of e-Infrastructures such as EGEE is their role as a framework for collaboration, on both a physical and human level. Grids, when well planned and managed, support and encourage the best science possible.

E-INFRASTRUCTURE FOR SCIENTIFIC DATA IN THE PSYCHOLINGUISTICS: A SURVEY BY PARSE.INSIGHT.

By Beate Sturm, Goettingen State and University Library and Max Planck Digital Library, Germany

Scientific data are at risk: they are scattered among various institutions and administered by researchers themselves. As a result, they are rarely accessible or preserved for the future.

To counter this, the EU-funded project PARSE.Insight (http://www.parse-insight.eu/) will develop a roadmap and recommendations for shaping an e-Infrastructure to maintain the long-term accessibility of scientific digital information in Europe. As part of PARSE.Insight, several case studies have been conducted, including one on the use of data in psycholinguistics. Psycholinguists study the relationship between the human mind and language, and create data during experiments and observations or use data stored in language archives. The case study covered persons within and outside of the European Union.

PSYCHOLINGUISTICS SURVEY COVERAGE

A survey conducted as part of the psycholinguistics case study generated 401 responses (205 completes; 196 partials), the vast majority of which (90.1%) were researchers. Most were individual researchers (almost 50%), but also research group leaders/managers (27.7%) and members of research groups (24.1%). The researchers usually represented their department (40.5%) or working group or project (39.9%), and not their entire institute (less than 20%).

Almost 70% of respondents work in the EU, three quarters of them in Germany, the United Kingdom, the Netherlands or France. Eastern Europe was underrepresented, and more than half of responses from outside the EU came from America. The survey therefore best represents the central-European psycholinguistic community, where significant psycholinguistic institutes are located.

E-INFRASTRUCTURE FOR PSYCHOLINGUISTIC DATA

When questioned as to the relevance of specific e-Infrastructure components (tools for automated data generation, researching language resources, automated data exchange and conversion, metadata generation, persistent identifier systems, storage, recommended standards, preservation policies and training in data curation), most respondents consider all components to be fundamental or important. The most important are thought to be basic work components ("research tools for language resources" and "storage"), while components highly relevant to the long-term preservation of data (such as "tools for metadata generation" and "persistent identifier systems"), are judged less important.

"Research tools for language resources" and "storage" are the only tools that respondents believe already exist. Thus it can be assumed that these components are most frequently provided within existing e-Infrastructures for psycholinguistic data. Hence, the most important components are most frequently available, and presumably, most frequently used. The researchers are not sure whether the other components exist. Hence, it can be assumed that they do not use these components, even if they understand their importance.

"Training in data curation" is the component researchers miss most, even though most consider it an important component.

Nearly 40% of researchers do not think that existing components fit their needs, and 37.3% are not sure. Even though researchers are not sure which e-Infrastructure components already exist, more than 23% are satisfied with them. Although almost a quarter state that the components fit their needs, more than 54% see the need for improvements; more than 42% are uncertain about it, and only 3% think nothing should be improved.

Although most researchers consider all given components as fundamental or important, most of them do not know whether existing e-Infrastructures provide these components, whether these components fit their needs, and what should be improved. Despite this uncertainty, 76% express the need for a networked and transnational e-Infrastructure for psycholinguistic digital research data. Only 7.5% do not see this need, and about 16% are not sure.

SUMMARY

Within the surveyed psycholinguistic community there is interest in the long-term preservation of data and awareness that efforts are needed to shape e-Infrastructures for these data. However, respondents are uncertain about the existence of current e-Infrastructure components and how to use them – they state a need for more training. Hence, there is a need for more information on how to use the e-Infrastructure components required for the long-term preservation, (re)use and exchange of scientific data.

GIVING COMPUTATIONAL SCIENCE A FRIENDLY FACE

By Jano I. van Hemert and Jos Koetsier, University of Edinburgh, United Kingdom

Today, most researchers from any discipline will successfully use web-based e-commerce systems to book flights to attend their conferences. But when these same researchers are confronted with compute-intensive problems, they cannot expect elaborate web-based systems to enable their domain-specific tasks.

Instead, they have to deal with archaic command-line tools or generic portals that mimic the technical complexity of the underlying infrastructure. These interfaces are expensive to use, require much training, and their laborious and intricate processes often lead to errors.

Of course exceptions to this situation exist. The most fortunate researchers have access to web portals or scientific gateways that were built to enable their specific tasks. But here we find a significant difference with e-commerce systems: where flights are purchased by millions of customers, these science gateways serve a specific niche inside an already small scientific community. Most of these communities cannot meet, let alone sustain, the cost of developing and maintaining a gateway specific to their requirements. Scientific tasks are not only complex, they also change frequently to reflect progress in the scientific field. The development of scientific gateways is thus expensive since each specific task has its own specific requirements. In contrast, purchasing a flight is equivalent for almost everyone.

COST-EFFECTIVE SCIENTIFIC GATEWAYS

Researchers at the University of Edinburgh's National e-Science Centre have developed a cost-effective and scalable solution to delivering scientific gateways for computational science. They have recognised that compute-intensive tasks have common use-patterns, and encapsulated these patterns in a declarative language. Using this language, a portal designer can rapidly deploy several portlets within days, forming a scientific gateway. Each portlet guides researchers through their task and enables delayed collection of results. Their solution is appropriately named "Rapid".

SO HOW DOES RAPID WORK?

- The portal designer specifies the user interface and logic flow of a task. This specification is created in one XML file.
- The designer then uses the Rapid Translator, which reads this XML file and creates a portlet
- The portlet can be deployed into any portal container compliant with the Java Specification Request 168 Portlet Specification, such as GridSphere, WebSphere or Liferay.
- A domain specialist can use a web browser to log in to the portal and access the new portlet.
- The user can then configure and perform the task at hand, using domain-specific jargon and graphical user-interface elements such as drop-down menus, radio buttons and check boxes.
- The computational task manager embedded in each Rapid portlet runs the appropriate compute jobs on designated compute resources. Rapid can submit jobs to most high-performance computing vendors and grid computing infrastructures, and cloud computing resources will be added soon.
- The progress of all tasks can be monitored
- When the compute jobs finish, the results can be transferred and analysed by embedding web-based visualisation components.

SUCCESS STORIES AND MORE

Rapid has been successfully applied in several specialist domains:

- Medicine: to enable wider uptake of image analysis on MRI brain scans of stroke patients (Scottish Funding Council Brain Imaging Research Centre)
- Chemistry: to create portlets for teaching and research to improve the usage of computational chemistry tools (joint Chemistry Research School of the Universities of Edinburgh and St. Andrews)
- Engineering: to run non-linear finite element analyses to develop real-time emergency response systems (University of Edinburgh)

- Seismology: to allow analysis of data available in the Orfeus Data Center, the primary EU centre for seismic data; these data no longer have to be downloaded and interpreted by individual researchers and thus can be more effectively used (Observatories and Research Facilities for European Seismology)
- Biology: to process microscopy data in conjunction with an Open Microscopy Environment, thus lowering the adoption barrier to these processing techniques (The Swedlow Lab at the University of Dundee)

Stable software releases, documentation, tutorials and examples are available at http://research.nesc.ac.uk/rapid/. Rapid is developed under an Open Source model and is available freely through a GNU General Public license. Feel free to contact us at j.vanhemert@ed.ac.uk to discuss potential applications.

CLIMATE-G: AN INTERNATIONAL TESTBED ON CLIMATE CHANGE

By Giovanni Aloisio and Sandro Fiore, Euro Mediterranean Center for Climate Change (CMCC), Italy

Climate-G is an interdisciplinary and unfunded research project aiming to create a Virtual Research Environment for climate change. It is built on top of the EGEE infrastructure and exploits transparently accessible compute and data resources through the Climate-G Portal.

MANAGING HUNDREDS OF PETABYTES

The size of climate datasets continues to spiral as model complexity, resolution levels, and the number of experiments continue to increase. Thus, the first challenge for Climate-G comes from the need to address scalability, performance and local site autonomy, and from the unfeasibility of moving such large data volumes in a centralized repository. Large-scale data federation is one opportunity to effectively share data produced by several centres. Such a distribution schema can include replication strategies to increase data availability.

DISTRIBUTED GRID AND P2P METADATA

To make such a large volume of widespread data really accessible, a strong metadata framework is needed. Scalability and autonomy are currently addressed using a distributed approach, exploiting a CMCC metadata management solution that leverages P2P and grid technologies. The service, called GRelC, is included in the EGEE RESPECT program, and can manage metadata information held in heterogeneous grid data sources. It enables sharing, search, and discovery of geographical data.

CREATION OF A SCIENTIFIC GATEWAY: THE CLIMATE-G PORTAL

The third challenge addressed by the Climate-G testbed is the creation of a seamless and ubiquitous access point: the Climate-G Portal. The portal is intended for scientists and researchers who want to easily and transparently manage the climate change experiments available in the Climate-G digital library. It provides functionalities including data search and discovery, metadata annotation and validation, data access, and visualization. The Climate-G Portal aims to be an "integrated working environment", where scientists can access huge volumes of data with complete metadata support and a wide set of data access services, data visualization and analysis tools, and easy access to the underlying EGEE infrastructure to run via web processing, analysis software and so on.

A new EGEE Virtual Organization has been established to support the climate change community via the Climate-G testbed, and EGEE have allocated additional storage and compute resources to support our experiments. Already, about 60 users are accessing the system via the portal (measured from April 2009).

COLLABORATIVE SUCCESS

The EGEE NA4 Steering Committee, EGEE Activity Management Board and the European Commission called Climate-G "indicative of the excellent scientific work being done on the grid and of the advancement of grid services/tools"; Climate-G representatives also demonstrated the testbed to an EC-appointed panel during the 2009 EGEE-III review.

The Climate-G collaboration continues to grow. The initial phase involved Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC, Italy), Institut Pierre-Simon Laplace (IPSL/CNRS, France) and Fraunhofer Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI, Germany). Now other centres and institutions have joined the testbed, including the National Center for Atmospheric Research (NCAR, USA), Rensselaer Polytechnic Institute (RPI, USA), University of Reading (Reading, UK), University of Cantabria (UC, Spain), and University of Salento (UniSalento, Italy).

LOOKING TO THE FUTURE

Moving forward, Climate-G aims to (i) extend its existing functionalities, services and tools (ii) act as a virtual laboratory for partners to test and validate software developments useful in the climate change domain and (iii) establish strategic synergies with new partners and projects at an international level. Full information is available at: http://grelc.unile.it:8080/ClimateG-DDC/

14

EMOTIVE: THE BSC'S ENGINE FOR CLOUD SOLUTIONS

By Iñigo Goiri, Jordi Guitart, Mario Macias, Jordi Torres, Eduard Ayguadé, Jorge Ejarque, Raul Sirvent, Daniele Lezzi, Rosa M. Badia - Barcelona Supercomputing Center, Spain

Cloud computing, in which IT resources and services are accessed as utilities, has recently emerged as a promising paradigm for exploiting e-Infrastructures.

Cloud computing is strongly based on virtualization, allowing applications to be multiplexed onto a physical resource while isolated from other applications sharing that physical resource. This technology simplifies the management of e-Infrastructures, but also requires additional effort if users are to benefit from it. Cloud computing must hide its underlying complexity from users: the key is to provide users with a simple but functional interface for accessing IT resources "as a service", while allowing providers to build cost-effective self-managed systems for transparently managing these resources. System developers should be also supported with simple tools that allow them to exploit the facilities of cloud infrastructures.

ELASTIC MANAGEMENT OF TASKS IN VIRTUALIZED ENVIRONMENTS

The Barcelona Supercomputing Center (BSC) is contributing to the e-Infrastructures research community with the EMOTIVE (Elastic Management of Tasks in Virtualized Environments) middleware, mainly the result of work by the BREIN and SORMA projects.

EMOTIVE provides users with elastic fully customized virtual environments in which to execute services. Further, it simplifies the development of new middleware services for Cloud systems by supporting resource allocation and monitoring, data management, live migration, and checkpoints.

LAYERED ARCHITECTURE FOR IMPROVED SUPPORT

EMOTIVE comprises several layers and provides users with basic tools for supporting tasks:

- the Virtualization Resource Management and Monitoring (VRMM) layer wraps each virtualized node and monitors its state
- within the VRMM layer, the Virtualization Manager (VtM) provides an application-specific virtual machine (VM) for each application; the application is given full control of its execution environment without introducing risk to the underlying system or other applications
- the VMs are created on demand, according to application requirements, and are consolidated in the provider's physical resources to optimise their use
- a Resource Monitor (RM) monitors task and resource status, including historical information.

EMOTIVE is currently based on Xen virtualization technology, with plans to support several virtualization solutions using the libvirt toolkit. It can manage local and remote virtual machines (i.e. those running in a provider's nodes or in a third-party providers' such as Amazon EC2). In this way, users can benefit from several Cloud implementations when running applications.

EMOTIVE also provides facilities for supporting resource management in Cloud environments. These include:

- VM placement and fine-grain dynamic resource distribution based on Service Level Agreements (SLAs),
- efficient live migration of VMs across provider nodes,
- a checkpoint/recovery system to resume task execution upon hardware failure (thus achieving fault-tolerance),
- ability to create additional VMs on external clouds when the local provider is overloaded, and
- data management services for supporting VM creation and the migration and checkpoint mechanisms, and also to allow users to provide input (i.e. data stage-in) and retrieve output (i.e. data stage-out).

Middleware services for Cloud systems can also be developed on top of EMOTIVE. For example, within the BREIN IST project, BSC has developed a Semantically-Enhanced Resource Allocator (SERA) that distributes resources using semantic information. Similarly, within the SORMA IST Project, BSC has developed an Economically-Enhanced Resource Manager (EERM) that manages resources using economic algorithms.

PARTICIPATION IN NUBA

Future development of EMOTIVE will also be fostered through participation in NUBA, the "Normalized Usage of Business-oriented Architectures" strategic research program, funded by the Spanish Ministry of Industry, Tourism and Trade. NUBA aims to advance state-of-the-art in business models and technology to deploy a federated Cloud platform that will simplify the deployment of business services.

OUTCOMES OF THE BELIEF WORKSHOP ON DATA PROVENANCE

By Eleni Toli, Michael Pantazoglou and Dimitra Keramida, National Kapodistrian University of Athens, Greece

A hot topic for scientific communities is the challenge of dealing with massive volumes of digital research objects.

For such data to be of long term worth, they must be traceable in terms of provenance (data origin) and authenticity (data validity). Achieving this would also extend return on e-Infrastructure investment and increase competitiveness. However, managing provenance and authenticity data also poses significant challenges in terms of formal modelling, storage, and maintenance.

THE ATHENS CORE

The 5th BELIEF brainstorming workshop, held in Athens, April 2009, aimed to tackle these issues, furthering discussions started at the 6th eConcertation Meeting in Lyon, November 2008. The Athens workshop led to several tangible results and recommendations, including the need to formally define a minimal set of provenance information and to develop a roadmap towards data provenance. Together, these actions form what is called the "Athens Core".

The need for high-level definition of a roadmap towards data provenance is part of this "core"; the roadmap would involve Communicable Information Packages (CIPs) and their definition, implementation, testing, real-world application, and so on.

The "Athens Core", if further elaborated and accepted, could act as a cornerstone for future research and standardisation activities in data provenance and authenticity.

Other important technical outcomes included the need for:
- Abstract definition of data provenance information as a "relationship", with a graph model
- Layered viewing of provenance information to reflect their granularity and differences in stakeholder perspectives

RECOMMENDED ACTIONS

The European Commission will play an important role in directing policy in this area, and the workshop recommended they take a number of actions:

- Encourage and fund seminal research to reach a critical mass of provenance offers, thus encouraging investors to adopt it.
- Use digital resources produced by the EU (e.g. Official Journal [OJEU]) as vehicles/pilots for the systematic provision of provenance information.
- Encourage key stakeholders (e.g. EMBL, ECMWF, ESA, CERN) and providers of digital information to include provenance information in their digital assets.
- Set the provision of provenance information as a primary topic in the research agendas of future EU- or nationally funded grid/e-Science/data preservation programmes.
- Fund pilot projects/demonstrations that address provenance within the lifecycle management of domain-specific information.

Thus the workshop recommended that the initial roadmap be further elaborated and refined, aiming to create a revised roadmap that can contribute to and influence future research agendas in data provenance and e-Infrastructures.

A report on the Athens Brainstorming, as well as full presentations, are available in the BELIEF Digital Library.

THE DARIAH E-INFRASTRUCTURE

By Mark Hedges, Centre for e-Research, King's College London, United Kingdom; Andreas Aschenbrenner, State and University Library, Goettingen, Germany; Tobias Blanke, Centre for e-Research, King's College London, United Kingdom; Eric Haswell, Department of Scandinavian Research, University of Copenhagen, Denmark

DARIAH (Digital Research Infrastructure for the Arts and Humanities; http://www.dariah.eu) aims to conceptualise and build a virtual bridge between different humanities and arts resources across Europe. Funded under the ESFRI programme (http://cordis.europa.eu/esfri/), DARIAH is in its preparatory phase, which involves design of the infrastructure and construction of a sound business and governmental model. DARIAH will begin its construction phase from 2011.

Just as astronomers require a virtual observatory to study galaxies, researchers in the arts and humanities need a digital infrastructure to access, combine and collaboratively work on their scholarly resources, which include digital content, services and methodologies. DARIAH will be such an infrastructure: it aims to combine various national infrastructures – from the UK's arts and humanities e-Science initiative projects (http://www.ahessc.ac.uk/) to the German e-Humanities infrastructure TextGrid (http:///www.textgrid.de) – and also to help other EU countries establish their own arts and humanities e-Infrastructures.

The DARIAH vision

DARIAH will promote, support, and advance research in the digital humanities. Digital humanities is a long-established research field, and in the last sixty years numerous digital humanities centres and organizations have developed. This trend continues, and we do not perceive the digital humanities to be a closed field of existing centres, but rather an open, developing research environment. Everybody interested in using digital means for arts and humanities research is part of the DARIAH community. Thus, the DARIAH infrastructure will be a connected network of people, information, tools and methodologies for investigating, exploring and supporting work across the broad spectrum of digital humanities.

The DARIAH network will be as decentralised as possible, empowering individual contributors (e.g. researchers, national centres, thematic centres) to work with and within the DARIAH community and to shape its features to their needs. Each con-

tribution builds DARIAH and is linked via DARIAH's architecture of participation. At the same time, collaboration across the borders of individual centers requires use of common technologies for authentication or federation of archive contents, for example.

(Re)use by anybody, anywhere

DARIAH is an eResearch environment focused on the (re)use of digital research resources by anybody and anywhere. DARIAH will not prescribe standards, but will encourage their use to foster interoperability and compatibility. Researchers do not have to support interoperability and openness, but they may want to support and benefit from the opportunities these features facilitate, such as collaboration and re-usability. DARIAH will provide community-driven recommendations and foster interoperability and collaboration through incentives. This approach means less central control over what DARIAH contains and provides.

When DARIAH is operational after the construction phase, its technical products will be manifold:

- services and tutorials to help existing humanities data archives link their systems into the DARIAH network
- a software and consultancy/training package that supports emerging data centres in establishing their own technology environment
- an interoperability layer that will connect data centres
- a means of linking into DARIAH for countries or disciplines that do not yet have e-Humanities infrastructure and cannot afford it in the near future
- best practices and guidelines for individual researchers that foster data interoperability and preservation across the DARIAH network.

We imagine DARIAH not as one large infrastructure, but as a means for linking people, services and data for research in arts and humanities. Most likely, DARIAH will not be one technical solution, but many, according to community needs and willingness to collaborate. The definition of a repository found in the DuraSpace midterm report also fits DARIAH: a "trusted intermediary that makes content (…) usable with a 'chinese menu' of added-value services". Of course DARIAH will not be a single repository, but otherwise the idea of a trusted intermediary fits well.

VIDEO ACTIVE – EUROPEAN TELEVISION HERITAGE ONLINE

By Johan Oomen - Netherlands Institute for Sound and Vision, The Netherlands; Vassilis Tzouvaras - National Technical University of Athens, Greece and Marco Rendina - Cinecittà Luce, Italy

Only a few percent of the many millions of hours of archived audiovisual material can be found online, but before audiovisual archives can set up meaningful online services, they must overcome obstacles in intellectual property management, digitisation technologies, metadata standardisation, and source presentation.

Video Active aims to address these challenges to create multilingual access to Europe's television heritage, providing access to a balanced collection of thousands of videos from 14 archives across Europe, each selected to reflect the cultural and historical similarities and differences of television across the European Union. Complementing this archive is a set of well-defined contextual metadata, allowing the portal to support textual search modes as well as faceted, thematic and timeline-based browsing.

Video Active uses Semantic Web technologies to provide expressive representation of its metadata, mapping heterogeneous metadata schema in the common Video Active schema, sophisticated query services, and interactive presentation modes. Video Active is thus fully compliant with the interoperability specifications of Europeana, the EU's massive digital library, due to launch in 2010 with links to 10 million cultural items.

Defining user requirements

Video Active worked to understand user requirements to ensure user satisfaction and revisits, using surveys, interviews, usability tests and desk research to help define its technical specifications and architecture. The excellence of the portal was acknowledged during the 2009 Museums and the Web conference, where Video Active won the Best of the Web award.

High level architecture

The Video Active system comprises various modules that manage its entire workflow, including annotating and uploading material, transcoding material, extracting keyframes, and storing and searching for metadata. Each module exploits semantic web technologies, enabling automation, sophisticated query services (based on the SPARQL standard) and semantic interoperability with other digital archives. The metadata are represented using the Resource Description Framework (RDF) and the Simple Knowledge Organizational System (SKOS), and are stored in the Sesame semantic metadata repository. This use of semantic web technologies enables light reasoning services (the use of implicit knowledge through subsumption and equivalence relations), merging/aligning metadata from heterogeneous sources, and sophisticated query facilities using the SPARQL RDF query language. Relational databases and full-text search technologies have also been used where semantic processing is not required. The Video Active metadata are public and can be harvested using OAI-MPH technology.

Storing and querying data, the semantic way

The Video Active metadata schema is based on the Dublin Core set of metadata schema with additional elements where necessary (i.e. genre, English title). The video metadata are generated automatically and represented in an MPEG-7-based schema. The metadata are then transformed in RDF triples and stored in a semantic metadata repository.

The annotation process is either manual or semi-automatic. When semi-automatic, the archives export their metadata using a common XML schema. Elements that cannot be mapped to the Video Active schema are inserted manually using the Web Annotation Tool, which contains a module that transcodes the original format to Flash and Windows Media streaming formats, creates low and medium bit rates for the streaming service, and performs keyframe extraction for thumbnail creation. The Web Annotation Tool produces an XML file that contains metadata, based on Dublin Core, as well as content encoding and key frame extraction information. The XML is then transformed into RDF triples and stored in the Sesame semantic repository. Sesame is an open source Java

19

framework for storing, querying and reasoning with RDF. It allows storage of RDF triples in several storage systems (e.g. Sesame local repository, MySQL database). The use of an ontology language such as RDF that has formal semantics enables rich representation and reasoning services that facilitate sophisticated query, process automation and semantic interoperability. Search and retrieval in Video Active is performed using a combination of structured RDF queries in SeRQL (an optimization of SPARQL query language for Sesame) and full text search queries using the high-performance, full-text search engine library Lucene.

All metadata stored in Sesame are exposed to external systems/archives with the help of an OAI-PMH compliant repository. Distributed OWL/RDF query mechanisms will be employed in a future release.

Multilingual access: using SKOS

Video Active supports eleven languages in four distinct ways.

- Includes localized interfaces for each of the languages covered
- Translates key metadata elements into English, thus providing the platform with a monolingual baseline
- Uses a timeline view to provide a visual overview of milestones in the development of television in Europe
- Includes multilingual controlled vocabularies for Keywords, Genre and Location metadata: the keyword vocabulary stems from the International Press and Telecom Council's 1500-term thesaurus (translated by the Video Active project into eleven languages); genre vocabulary uses the ESCORT 2007 EBU System of Classification of Radio and Television Programmes; and geographical names use the ISO 3166 English Country Names and Code Elements. A specialized application called ThesauriX handles the translation of these terms and their export to machine-readable XML. To achieve semantic interoperability, the thesaurus taxonomy has been transformed into a semantic web language using the recommended Simple Knowledge Organisation System (SKOS) standard, which is built on top of the RDF language.

Conclusion

Video Active exploits recent advances in Semantic Web technologies to provide sophisticated web services using machine-understandable metadata. Semantic Web technologies such as RDF, SKOS, OWL and SPARQL have be used for the representa-

tion, query, presentation and exchange of the Video Active Metadata.

Video Active is a content enrichment project under the eContentPlus programme, and an invited member of EDLnet, the network initiated in 2006 to build consensus to create the European Digital Library. Europeana, a library bringing together hundreds of collections across Europe, has already indexed the data from the Video Active repository. www.videoactive.eu

AUDIOVISUAL AND FILM ARCHIVES CREATE FUTURES: THE EUROPEAN FILM GATEWAY

By Georg Eckes, Deutsches Filminstitut, Germay; Monika Segbert, Eremo srl, Italy and Paolo Manghi, CNR-ISTI, Italy

The challenges of accessing and preserving film – an essentially trans-national medium – cannot be solved at the national level alone. Thus in meeting modern digital challenges, European film archives are benefiting from valuable technical, legal and operational expertise within the framework of the Association des Cinématheques Européennes (ACE), an affiliation of 38 national and regional film archives from all over Europe.

Already, ACE has worked on a number of successful cooperative projects: classical analogue restoration; expansion of a joint European filmography; defining recommendations for the digitisation of films and other audiovisual content (FIRST); providing a gateway to information about the existence, location and copyright of documented material (MIDAS); and leading standardisation processes (CEN) on metadata for cataloguing and indexing of cinematographic materials.

These projects are the building blocks for the European Film Gateway, a new EU-funded project uniting 20 ACE partner institutions from 14 European countries. EGF will be the key to unlocking Europe's film archives for users of Europeana, an EC-funded online library that will feature millions of digital items from Europe's cultural heritage. ACE is a founding member of the EDL Foundation, the governing body of Europeana.

Technical and semantic interoperability

Common interoperability standards are not yet widespread in the film archival community. While many European libraries and non-film archives already enforce common cataloguing and metadata standards as well as interoperability protocols, most film archives lack these basic preconditions

for technical interoperability. By involving a large number of European film institutions, EFG aims to develop a comprehensive strategy to close this interoperability gap.

The EFG system will be based on the D-Net software toolkit produced by the DRIVER project (www.driver-community.eu). Running D-Net enables a service-oriented distributed infrastructure that offers:

flexible and configurable tools for constructing a uniform Information Space by collecting, transforming, and indexing metadata records from a set of heterogeneous data sources, and

customisable functionality including user interfaces, recommendation services, and user profiling. This openness will allow EFG to be easily adapted and extended to deal with specific metadata formats and functional requirements.

Intellectual property rights

In many European countries, works of film are especially affected by copyright regulations, rendering their digital preservation and access a difficult and legally complex task. In addition to European copyright directives, a variety of legal regulations exist for non-commercial and educational use of material in film archives and cinémathèques. Further, access to works considered "orphans" can be complex and time-consuming, as film heritage institutions face the task of trying to find and contact creators or their successors. As a result, most films and film-related objects in Europe's film archives are effectively inaccessible. To help tackle this, EFG will develop guidelines and best practices that support film archives in avoiding legal pitfalls.

EFG and Europeana

As an aggregator of aggregators, Europeana depends on the spadework being done within the four domains: libraries, museums, archives and audio-visual archives. As a leading aggregator project, EFG will make a huge contribution to Europeana, making Europe's rich and valuable film archive content more accessible and thus enriching user experience with a popular form of cultural expression.

EFG launched in September 2008 and will run for three years. EFG's public operational service will launch in 2010. EFG is co-funded by the EC's eContentplus programme and coordinated by the Deutsches Filminstitut.

EARLY HUMAN HISTORY: WAITING UNDER THE WAVES

By Danielle Venton, Enabling Grids for E-sciencE/ CERN, Switzerland

For our ancient ancestors, coastlines were attractive places to settle and experiment with what became the foundations of civilization. However, between twenty and six thousand years ago the major glaciers melted, and these sites — where humans first began to make fishing equipment, build boats and create permanent settlements — became engulfed by the rising seas.

In many cases, instead of destroying these ancient landscapes, the rising sea level actually preserved them, along with many details from the story of our past.

"There are large gaps in our general knowledge of early history," says Geoff Bailey of the University of York's Department of Archaeology. "We have a lot to learn by looking underwater. There are many sites to discover and examine, and preservation is in fact often better than on land."

Working in places as distant as the Southern Red Sea, the shores near Gibraltar, and off the coast of England, Bailey and his colleagues look for sites containing the well-preserved ancient remains that are rare in inland sites: wood, woven fibers, beetles and insects, plant material, pottery (sometimes with the remains of food inside) and even bones and other organic material containing DNA traces.

To aid in the collection and sharing of this information, Bailey and colleagues have started a new European research network, called SPLASH, or "Submerged Prehistoric Landscapes and Archaeology of the continental Shelf." The project was recently awarded funding by the European Union

under the Cooperation in Science and Technology (COST) Action framework. Its first phase of work will last four years.

"Over the past twenty to thirty years, enormous amounts of submarine data have been gathered on the seas," says Nic Flemming, SPLASH member and research fellow at the National Oceanographic Center in Southampton, England. "No one has previously thought of systematically using this data for archaeology, but in fact we can use it to reconstruct past, once-dry landscapes. What we need is a system to tap into these archives all over Europe."

Beginning with a splash

SPLASH with begin by providing the research and cultural management communities with the means to access, browse, recover, manipulate and integrate all relevant earth science and archaeological data currently archived in European digital repositories. This will weave together large stores of information about Europe's coastlines, wetlands and soil profiles, including mineral and organic content.

Using e-Infrastructures, researchers will be able to respond to EU Marine Policy/Strategy initiatives (DG MARE), and parallel initiatives on pan-European electronic data access (DG INFSO) and support for operational oceanography (DG RTD and DG ENTR) to provide rapid access to maritime data in a standardised way, allowing researchers to dig into large data sets from distributed archives.

"We will be able to mobilise knowledge and skills that are currently dispersed across national and disciplinary boundaries," says Bailey, "and begin to combine the efforts of individuals in marine and environmental sciences, the archaeological community, heritage organisations, and industry in tackling archaeology's last frontier: the deeply submerged landscapes and prehistory of the continental shelf "

Researchers have found some unexpected things beneath the waves, including this wooly mammoth head. Image courtesy Dick Mol and Nic Flemming/SPLASH

GLITE: LEADING THE DEVELOPMENT AND ADOPTION OF STANDARDS

By Steven Newhouse, Enabling Grds for E-sciencE, Switzerland

The Enabling Grids for E-sciencE (EGEE - http://www.eu-egee.org) project supports the high-throughput data-processing needs of the European research community and their international collaborators, supporting over 17,000 users across 160 projects and covering domains such as astrophysics, biomedicine, computational chemistry, earth sciences, high energy physics, finance, fusion, geophysics and multimedia. EGEE also supports business applications in the geophysics, finance and plastics industries.

The middleware that makes this possible is gLite, a software distribution produced by EGEE and composed primarily of open-source software from many sources – most developed within the project, some from external providers. gLite allows researchers to access physical resources attached to the EGEE infrastructure – including disks, computers, and instruments – using their own applications. gLite services now have stable interfaces, often in the form of web services and defined by technical standards, which encourage application and tool developers to build their own software on top of the gLite platform.

The advantages of standards

Adoption of distributed computing infrastructures is expanding across all disciplines, making it increasingly important to have standard interfaces. Standards allow developers to protect their investment in the software they develop by enabling access to any infrastructure that supports these standards. Standards prevent developers from being locked-in to one particular platform. Standards encourage competition between middleware providers, allowing resource providers to replace under-performing software components with better components that use the same standard interface, with no disruption to the user community.

As EGEE's infrastructure has matured, the project has increased its investment in standards and its liaison with standard organisations. EGEE uses many protocols common to the internet (e.g. TCP and IP), the web (e.g. http and https) and web services (e.g. XML, SOAP and WSDL); these protocols will not be covered in this article.

Instead, we will concentrate on the protocols that allow secure access to shared resources. These protocols come primarily from the Open Grid Forum (OGF), of which EGEE is an active technical participant, and the Organization for the Advancement of Structured Information Standards (OASIS) and Internet Engineering Task Force (IETF), where EGEE and the grid community are primarily external consumers.

Standards within EGEE: Security

The resources that make up the EGEE infrastructure are extremely valuable and access to these resources needs to be strictly controlled. EGEE, in common with other infrastructures, use the X.509 certificate and X.509 proxy certificates from the IETF to represent the permanent and delegated identity of an individual. Through the EGEE Virtual Organisation Membership Service (VOMS), the X.509 proxy certificate can include a SAML (Security Assertion Markup Language) assertion from OASIS, which provides information that oth-

23

er services can use to make authorization decisions. Authorization requests use XACML (eXtensible Access Control Markup Language) from OASIS to define the content of messages in the SAML message format passed to the authorization service.

Standards within EGEE: Information

With many thousands of services potentially available, discovering which one to use presents many challenges. The GLUE 2.0 specification from OGF provides a defined abstract information model that uses a set of known attributes with defined semantics to facilitate interoperation between grid infrastructures. These attributes can be rendered through a number of defined protocols. EGEE is currently adopting GLUE 2.0 with deployment planned for production use in early 2010.

Standards within EGEE: Data

Many researchers use EGEE's infrastructure to analyse data stored in files. EGEE, in common with many other infrastructures, uses the GridFTP v2 protocol (an OGF specification that extends the FTP protocol) to provide high performance transfer of large data files over wide area networks. The storage resources at the source and destination of the file transfer are manipulated through OGF's Storage Resource Management (SRM) protocol, which governs access methods when moving files to and from the permanent storage space (disk or tape) to the temporary space accessible to the GridFTP server.

Nearly all communities supported by EGEE analyse file-based data by starting an application that processes that data. Two specifications developed within OGF and relevant to EGEE are the Basic Execution Service (OGSA-BES), which describes an interface for running and controlling applications, and the Job Submission Description Language (JSDL), which specifies how applications should be run. EGEE is examining how these specifications could be adapted to serve its users.

Standards within EGEE: Accounting

Since EGEE's resources are contributed by different organisations for shared use by third parties, it is important that this use is accounted for. Within EGEE, and other infrastructures, OGF's Usage Record specification is used as part of the protocol for transferring records between resources and the central accounting server. Partners within EGEE are examining the use of the Resource Usage Service as an interface for retrieving stored records, but this needs further development.

Summary

The adoption, development and implementation of standards is crucial to the health of the grid computing community. EGEE has contributed significantly to this work, thanks to its experience in running a large e-infrastructure. We are cooperating with other infrastructure providers within OGF's Production Grid Infrastructure Working Group (PGI-WG) to 'fine tune' these specifications to our particular use cases and thus build interoperable e-Infrastructures.

The encapsulation of this 'know-how' into published use cases and the specifications of organisations such as OGF may represent a significant part of EGEE's eventual legacy. This work also provides a route for new software providers to enter the grid world, and allows resource providers to deploy any software solution that meets the required standards. Standards provide an essential foundation for allowing independent National Grid Initiatives to federate their resources into a European Grid Infrastructure, ensuring Europe's researchers have access to the computing resources they need for many years to come.

LEGAL ISSUES IN E-INFRASTRUCTURES

By Damien Lecarpentier and Panos Louridas, e-IRGSP2 Project, Finland

Legal issues may directly influence the development and sustainability of e-Infrastructures and are as such increasingly important. In 2008, the e-IRG-SP2 (e-Infrastructure Reflection Group (e-IRG) Support Programme 2) project conducted a study of key legal issues concerning research infrastructure technologies and collaborations. The report examined three areas that need to be tackled to ensure sustainable development of e-Infrastructures:

- the new legal form of the European Research Infrastructure Consortium (ERIC);
- the pre-commercial procurement initiative put forward by the European Commission;
- other legal issues pertaining to the use of electronic infrastructures, such as privacy and intellectual property rights (IPR).

The report gives an overview and seeks to foster discussion on these crucial issues.

A new legal framework for research infrastructures: ERIC

The legal form of a research infrastructure has long-term implications for its creation, management, finances, human resources, development and sus-

tainability. As such, choosing the right legal form can be complicated. The e-IRGSP2 report outlines the different legal forms available under national, international, and Community law, and examines ERIC: the European Research Infrastructure Consortium proposed by the European Commission as a new legal instrument for research infrastructures of pan-European interest.

ERIC has several benefits: it provides a legal personality recognised by all Member States; it offers some of the advantages enjoyed by international organisations, for example, VAT exemption; it meets non-profit criterion while allowing limited economic activity and cooperation with industry (essential for promoting innovation and progress); and it is easy to establish and flexible regarding the possible types of partnership.

However, the lack of precedent concerning this legal form raises certain issues. In particular, the regulation of several issues will rely on agreement between participating members and the national law of countries hosting an infrastructure. Therefore, the statutory provisions that define the organisation (including its decision-making processes, voting rules, member obligations regarding performance, liability, quality of service and security, etc.) will be of primary importance for the future implementation – and examination – of this new legal form.

Pre-commercial procurement issues crucial for innovation

Another important issue is public procurement, which, if adequately implemented, can have a significant impact on innovation. The Pre-Commercial Procurement (PCP) approach supported by the European Commission will undoubtedly stimulate R&D spending by public Member State entities. In this approach to PCP, a public purchaser can acquire, via competitive tender, research services leading to the development of product prototypes, prior to these products becoming commercial. Involving a public purchaser in this way can result in the production and supply of better and cheaper products. Thus PCP is likely to stimulate economic growth and new jobs if used as a strategic policy instrument. PCP can also complement other research and innovation policy instruments, such as structural funds, R&D tax incentives, joint technology initiatives, or state aid for R&D. Limiting PCP to European suppliers can also contribute to the development of new sectors of European leadership in the global market, thus strengthening European innovation and competitiveness.

25

Data management and intellectual property rights: a growing concern for e-Infrastructures

Data is at the heart of the knowledge society. However, the use of large volumes of data in e-Infrastructures raises problems with security, reliability and privacy protection. Tackling data protection issues in e-Infrastructures requires examination of related issues, such as data processing, responsibilities of data producers and managers, national and/ or territorial application of data protection directives, and international data transfer.

Management of Intellectual Property Rights (IPRs) is another important issue, and although the research community has identified and analysed several aspects of this topic, there is still a long way to go before consensus can be reached and global directives put forward. Both researchers and e-Infrastructure users and providers are interested in exploring and settling this issue, which is felt to be a barrier to the further development of e-Science and collaboration with business. Currently, e-Infrastructure users, providers, researchers and developers must be aware of the risks and liabilities they may face in violating third parties' rights; at the same time, e-Infrastructures must continue to promote innovation and development. It is thus important to find a good set of IPR directives that that protect users and developers without jeopardizing the main objectives of e-Infrastructures.

The full e-IRGSP2 report can be downloaded at: http://www.e-irg.eu/images/stories/publ/e-ir-gps2_public_deliverables/e-irgsp2_d4.1.pdf

GOING GREEN WITH E-INFRASTRUCTURES

By Manisha Lalloo, GridTalk Project, United Kingdom

Information Communication Technologies (ICTs) are found in all aspects of our lives, from health and science to transport and manufacturing. But the widespread use of ICTs has come at a price: the ICT sector is estimated to produce two percent of Europe's carbon emissions, a figure equal to that produced by the aviation industry. And, as the ICT industry continues to grow, emissions for the production and use of ICTs are set to rise by six percent a year.

To continue to benefit from ICTs we must reduce the sector's carbon footprint and energy consumption. Cutting emissions is key to sustainable growth in the ICT sector: without it, providers may be forced to pay large fines or offset emissions, potentially leading to unfeasible operation costs.

The transition to an energy-efficient, low carbon economy

But the news is not all bad. In a recent Communication, the European Commission identified ICTs as key to cutting emissions. Entitled "Mobilising Information and Communication Technologies to facilitate the transition to an energy-efficient, low carbon economy", the Communication stated that ICTs can be used to enable energy efficiency improvements and quantify energy consumption.

The technological solutions ICTs offer can reduce energy usage, not just in the technology sector, but across the whole economy. For example, by improving logistics, ICTs could help increase the efficiency of travel, therefore reducing emissions by 27%.

However, engaging the help of e-Infrastructures offers one of the best hopes for greening the ICT sector.

Virtually green

Innovative technologies such as grid computing, cloud computing and virtualisation are being touted as some of the best ways to reduce redundancy, and therefore, energy consumption.

Virtualisation, for example, allows users to run multiple applications using multiple operating systems, all on the same physical computer. This

can maximise server capacity, cutting the number of machines required to process information, and so cutting power consumption too. Many see virtualisation as the future of IT, with organisations outsourcing their computational needs, rather than dealing with them in-house. However, companies need to ensure they are not simply offloading their emissions to another source, but are instead enabling dedicated server hosts to aggregate their processing needs onto a smaller number of machines.

In many cases, moving to the cloud has already reduced the number of servers required. An example is Mimecast, a company that provides cloud-based email management. By moving electronic communications to the cloud, Mimecast users have eliminated an estimated 8,300 servers worldwide.

Towards more eco-friendly data centres

While e-Infrastructures can offer a way to cut power consumption, a move to cloud computing must be carefully managed. Aggregating processing needs could actually increase energy consumption, due to the cooling requirements of large data centres owned by cloud providers. This is a major problem: data centres are estimated to account for three percent of the world's energy consumption.

Many providers and organisations are already taking steps to reduce this, finding more eco-friendly ways to run data centres. Using renewable energy – as well as simple measures such as using thicker cabling, "free cooling" using outside air, and leaving more space around processors – can help reduce emissions and power consumption.

Increasingly, plans for new centres are taking green issues into consideration, not least because going green also reduces operating costs. For example, plans for a new computer centre at CERN will likely use green design, water cooling, and energy reuse that could redirect excess energy to the heating systems of nearby buildings.

Researchers are already looking for ways that e-Infrastructures can adapt to meet the growing demand for reduced energy consumption. For example, the GREEN-NET project, supported by INRIA, is scheduling jobs on the grid in a way that reduces energy consumption. Their research has found that scheduling jobs at quiet times and following green policies can dramatically reduce energy usage. For a platform like Grid'5000 in France, GREEN-NET methods can yield energy savings of 30-35% a year. The OpenNebula initiative is also working to reduce power usage: their newly developed consolidation scheduler allows grid services to run on fewer physical machines, cutting both power use and cooling requirements.

Although a reduction in energy consumption should lead to a cut in emissions, providers must ensure that in response, they don't simply increase the load on machines. Nevertheless e-Infrastructures can play a key role in helping the ICT industry reduce rather than add to carbon emissions. By going green we can realise many benefits for our environment, our budgets and a sustainable future.

E-INFRASTRUCTURE FOR ENVIRONMENT SUSTAINABLE DEVELOPMENT

By Paraskevi Mentzelou, Technological Educational Institute of Thessaloniki, Greece

Environmental problems require immediate action and sustainable long-term strategies. To this end, in 1998 fifty scientists from Balkan countries formed the Balkan Environmental Association (B.EN.A, http://www.gen.teithe.gr/~bena/bena-home.html) as an international nonprofit and nongovernmental scientific organization tasked with preserving the environment and culture of Balkan countries.

B.EN.A.'s members work to develop international collaborations aimed at reducing transboundary air and water pollution. B.EN.A. thus serves as a Balkan "think tank", working to:

- analyse interactions between environmental, economic and community trends;
- establish scientific ways for environmental protection;
- promote strategies and institutional changes to improve environmental and natural resources support environmental decision making processes through objective and independent scientific research, studies and professional training, and ultimately;
- improve people's lives.

Behind B.EN.A.: our e-Infrastructure

Environmental protection and sustainable development has been a "non-stop-watching" process, especially for transboundary pollution. Thanks to Information and Communication Technologies, more effective ways for collaboration and cooperation are available in this research area.

B.EN.A.'s e-Infrastructure was developed in a decentralised way, with management based on authority delegation and on the scientific interests of its members. The Alexander Technological Educational Institute of Thessaloniki is responsible for overall management, aiming to make best practise in common research and experimental approaches available across the Balkan countries. Thus B.EN.A. is a reference point for scientists interested in environmentally sustainable development in the Balkans.

The e-Infrastructure comprises three Virtual Research Environmental Communities (VREC), each with a distinct role:

- The National Bureaus involve scientific collaboration between universities for the "Sustainable Development and Better Life of Balkans". Partners include: the University of Tirana in Albania, University of Plovdiv in Bulgaria, University of Sarajevo in Bosnia and Herzegovina, University of Zagreb in Croatia, Institute of Public Health in FYROM, National Institute for Research and Development of Industrial Ecology in Romania, Trakya University in Turkey and Institute of Nuclear Science in Serbia.
- The Liaison Offices identify emerging environmental issues and are based in the Russian Federation (National Research Institute of Azov Sea Fishery Problems), in Ukraine (Ukrainian Scientific Centre of the Ecology of Seas), and in Georgia (Technical University of Georgia).
- The International Centres and Offices focus on certain environmental topics, involving specialists from all Balkan countries.

Current studies and future steps

B.EN.A's specialists are involved in the study of pollution (air, water, soil, agricultural, industrial); ecology; risk assessment; waste management; radioactivity; clean technologies; environmental education, legislation, management and informatics; and computer application to environmental issues. The results of B.EN.A' research can be found in the Journal of Environmental Protection and Ecology (www.jepe.gr), which is starting to serve as a digital library for environmental issues.

B.EN.A.'s next step is to activate the Balkan Local Authority Environmental Network, or BLAEN, which will be based on regional cooperation and substantial support for a clean and healthy community environment. Up today BLAENS members are the Municipalities of Albania, Bulgaria, FYROM, Romania, Serbia-Montenegro and Turkey.

VERY HIGH SPEED RESEARCH NETWORK TO BOOST SOUTH AFRICA'S NATIONAL CYBERINFRASTRUCTURE

By Christiann Kunn and Duncan Martin CSIR, South Africa

SANReN (South African National Research Network) forms part of a comprehensive South African Government plan to ensure South African researchers successfully participate in the effort for global knowledge production. Together with the Centre for High Performance Computing (CHPC) and Very Large Data Sets, the initiative for data curation, SANReN will form a key component of the core scientific research infrastructure for South Africa.

SANReN will be a high-speed network that enables and supports scientific and technological research, including research into networking and broadband infrastructures. SANReN's deployment commenced in 2007 and, when complete, will connect 108 university campuses and research institutions across the country at very high speeds. SANReN will also join the global research and education network, which connects over 3000 research and education organisations from all over the world.

Planning and implementation

The Council for Scientific and Industrial Research (CSIR)'s Meraka Institute was given responsibility for implementation planning in early 2006, when it was working to establish the CHPC. This has helped to ensure a comprehensive and holistic approach to developing South Africa's research cyberinfrastructure. In 2006 CSIR and the South African Department of Science and Technology (DST) signed a three-year contract, ending 31 March 2010, for the implementation and management of this project..

The SANReN team has since been active in designing, negotiating and building the SANReN network, including the establishment of partnerships with local companies to leverage economies of scale through co-investment initiatives. The 10 Gbps SANReN national backbone network is scheduled for commission this year.

SANReN and TENET

SANReN will be operated by the Tertiary Education and Research Network of South Africa, or TENET. TENET is a non-profit association of South African universities and research councils that has provided networking and Internet services to 40 universities, research institutions and associated support institutions since 2001. While the DST, through the CSIR, has provided the capital to build and equip SANReN, TENET will recover its operating costs from user institutions.

TENET will also provide, and user institutions will bear the costs of, all external connectivity. This includes connectivity at 10 Gbps to Telecity, London, via the new SEACOM submarine cable; transit interconnections with UbuntuNet, the African regional research and education network; and via UbuntuNet, connections with the European regional research and education network, Géant. TENET is a founding member of the UbuntuNet Alliance, and operates UbuntuNet's hub and connection to Géant in London on behalf of the Alliance.

Through their contractual and collaborative relationships, the DST, CSIR's Meraka Institute, TENET and UbuntuNet are securing affordable, high-speed national and international connectivity services for Southern Africa, connecting the region's universities, research institutions and research facilities, and enabling researchers and students to participate fully in the global research community.

ENABLING KNOWLEDGE TRANSFER WITH THE EPIKH PROJECT

By Giuseppe Andronico INFN, Division of Catania, Italy, Roberto Barbera INFN Catania and Department of Physics and Astronomy of the University of Catania, Italy and Valeria Ardizzone for the EPIKH Project

e-Science and e-Infrastructures are key enablers of progress and sustainable development, and concrete means by which to address the 'digital divide' and 'brain drain' endemic in large parts of the world.

To turn the vision of a European Research Area (ERA) into reality, Europe is heavily investing to create a grid-based continental e-Infrastructure. In the context of its 6th and 7th Framework Programmes, the European Commission has co-funded several projects that stimulate and foster e-Science and grids well beyond its borders, in parts of the world such as Asia, Latin America and the Mediterranean.

Exchange Programme to advance e-Infrastructure Know-How

However, adoption of the 'grid paradigm' and effective use of research e-Infrastructures requires knowledge dissemination and training. It is to achieve just this that the 'Exchange Programme to advance e-Infrastructure Know-How' (EPIKH) (www.epikh.eu) was created. EPIKH aims to reduce the digital divide and brain drain by:

- Reinforcing the impact of e-Infrastructures in scientific research by defining and delivering stimulating educational programmes, including Grid Schools and High Performance Computing courses;
- Broadening engagement in e-Science activities and collaborations both geographically and across disciplines.

These ambitious goals translate into the following specific actions:

- Use an extensive training programme to spread knowledge of the 'grid paradigm' to potential users, including system administrators and application developers;
- Provide trained people with easy access to relevant e-Infrastructures;
- Foster the establishment of scientific collaborations between countries and continents involved in the project.

The exchange programme will be implemented in alternating phases:

- first, a select team of brilliant young researchers will visit EPIKH's EU partners for around one month to be trained as trainers of grid technology (including site administration and application support ('gridification'));
- second, EPIKH will organise and run at least two educational events per year and per continent (Africa, Asia, and Latin America), training new users to access and apply a pilot e-Infrastructure on which applications can be deployed, developed, and then used as exemplar use cases in future events.

The EPIKH project will shift the educative mission by developing an intensive and diversified training programme in which grids are not the 'goal' but rather the 'means' by which to develop e-Science applications, gather scientific communities from four continents, and access globally distributed production quality e-Infrastructures.

The EPIKH consortium unites 23 institutions from 18 countries across four continents: Africa, Asia, Europe and Latin America. EPIKH will mobilize about 115 people for more than 650 researcher-months, not counting, of course, those who will be trained by and benefit from the project. These are huge figures, confirming the strong interest of these four continents in setting up a programme to improve dissemination of know-how about grid and e-Infrastructures.

Two EPIKH Grid Schools have been planned so far: "Latin America 1 - 2009" was held in Queretaro, Mexico City, from 28 September to 10 October

2009. This event was co-organized with the EELA-2 project (www.eu-eela.eu). For more information visit http://agenda.ct.infn.it/conferenceDisplay.py?confId=179.

"Africa 1 - 2009" will be held in Cape Town, South Africa, from 16 November to 4 December 2009. This school will be held immediately prior to the CHPC/BELIEF-2 (www.chpcconf.co.za/) conference. The technical nature of the EPIKH event will complement the political and strategic aims of the CHPC/BELIEF-2 one. For more information visit http://agenda.ct.infn.it/conferenceDisplay.py?confId=182.

GLOBAL E-INFRASTRUCTURE OF IPV6 EXPERIMENTAL LABORATORIES

By Socrates Varakliotis and Peter Kirstein, University College London Dept. of Computer Science, United Kingdom

The global network of experimental IPv6 laboratories is set to be extended thanks to a coordinated effort involving Cisco, European Commission and NATO projects and University College London (UCL). The network, which currently includes Europe, Latin America and Africa, will be extended to include the Caucasus, Central Asia and India. UCL will use this e-Infrastructure to train the staff of National Research and Education Networks (NRENs) in the use of Internet Protocol (IPv6), including remote coordination using global-scale 'voice over

IP' systems. Participating NRENs and their associated scientific communities will thus be better able to engage in consolidated networking and partnership building.

Federated training laboratories

A number of projects are working together to make this e-Infrastructure a reality: partners of the EC-funded 6DEPLOY (www.6deploy.org) project offer training to organisations in Europe and developing countries, and support real IPv6 deployments. 6DEPLOY's global workshop activity has worked with major Regional Internet Registries (AfriNIC and LACNIC) to develop a substantial laboratory infrastructure, as well as training and e-learning materials aimed at training engineers and network administrators, and transferring knowledge and best practice to other trainers.

Cisco is the exclusive equipment donor for these activities. Existing IPv6 training laboratories in Paris, Sofia and Mauritius can be accessed remotely to perform hands-on exercises on routing and configuration. RENATER in France coordinate these 6DEPLOY laboratories; UCL coordinate the voice and video collaboration (to ease communication, and as an important application of IPv6 in its own right).

Collaborating along the Virtual Silk Highway

The Virtual Silk Highway project, funded by the NATO Science for Peace Programme, has built a substantial network infrastructure and sustainable National Research and Education Network organisations in three Southern Caucasus and six Central Asian countries, including Afghanistan. As a result, the communication requirements of the regional research communities has grown multi-fold, to the point where transition is now taking place from satellite-based to fibre-based network infrastructure (through the NATO-funded Silk project, the EC-funded BSI project in the Caucasus, and the EC-funded CAREN project in Central Asia (www.silkproject.org, www.blacksea-net.eu, caren.dante.net)).

From Bangalore to Bishkek

Cisco is currently installing new laboratories in Bangalore, Tbilisi and Bishkek. Laboratories are also planned for Turkey, Kenya and beyond. These new laboratories will be equipped with advanced Unified Communications components (IPv6 phones, Call Managers and Conference Servers) to allow interaction between trainers and remote site

31

administrators, and for future course and skills development.

Many other projects are involved in this venture. UCL is working to unite a wealth of human, knowledge and infrastructure resources from a wider spectrum of EC projects, including 6CHOICE (www.6choice.eu) and GLOBAL (www.global-project.eu). We expect these efforts to have the following benefits, both for developing regions and for Europe:

- Host organisations will manage the e-Infrastructure (IPv6 laboratories) locally and will benefit from knowledge transfer on network and compute resource management, as well as increased links with European libraries, network providers and suppliers of network equipment. Hosts can also develop new services and deploy state-of-the-art technologies such as virtualisation and networked embedded systems.
- Hosts will maximise use of their National Research and Educational Networks by enabling remote access and availability.
- Regional research communities will grow closer to their European counterparts, facilitating joint research on core scientific areas such as seismic research, water conservation and energy generation.
- These interactions will ignite further development of national e-Infrastructures and align common social and infrastructure development interests (such as the use of future internet protocols and advanced network services).
- Scientific and cultural gaps between geographically isolated communities will be bridged.
- Europe will expand its research context and broaden the focus of its research framework.
- Human resources and knowledge will feed back into EC regions.
- Economic stability will be strengthened in participating regions.

More information is available from {socrates, kirstein}@cs.ucl.ac.uk

CLOUD COMPUTING: AN E-INFRASTRUCTURE FOR COLLABORATIVE RESEARCH

By Tavleen Oberoi and Surjeet Mishra, Tata Consultancy Services, India

Research institutions and organisations are looking beyond their own walls for innovation, opting to create research collaborations to cope with the challenges of a rapidly changing world. Such research collaborations are joint ventures involving two or more organisations with the common objective of producing new scientific knowledge by collating their individual strengths.

Collaborative research involves sharing ideas and concepts through personal meetings and electronic communication; fabricating and materialising those ideas; implementing the ideas in parallel; sharing the results; testing the products; and finally, releasing those products. This scenario becomes more complicated when the collaborating parties are distributed, occupying different geographies and time zones. Thus, a research partnership requires good communication, fast data sharing and high processing power for complex operations. Cloud computing is emerging as a suitable platform to meet these demands.

What is cloud computing?

Cloud computing provides compute resources – such as scalable virtualized servers, data centres, and economical bandwidth – over the Internet. In addition, clouds can provide complex and traditionally expensive database services, such as automatic data indexing data and querying. Major cloud service providers – such as Amazon, IBM, Microsoft and Google – provide these resources using the Infrastructure as a Service (IaaS) model,

one of three types of cloud service (Platform as a Service (PaaS) and Software as a Service (SaaS) being the other two). Cloud resources are available on a pay-per-use basis: the consumer pays for the amount of resources they use and the duration of their usage.

Cloud computing assuages the initial investment required to set up hardware, software and networking equipment, and also reduces maintenance costs. It allows for dynamic scaling, such that computing capacity can be increased or decreased as required. Clouds also avoid common problems such as poor network connection or hardware failure since they distribute work across multiple computers. Further, cloud computing can reduce the cost of running applications with lower access rates, since cloud servers can be automatically started on request, or stopped after certain duration of idleness, hence reducing the duration of resource usage.

Cloud usage in collaborative research

Cloud computing can support collaborations between institutes and businesses. It can provide a platform for communicating and exchanging ideas, thus helping organisations of all sizes to work together and beyond their individual boundaries. For example, according to a Reuters' article, Sogeti Groups, a leading provider of IT services, used 'IBM idea factory for cloud computing' to run the first collaborative brainstorming session of its kind. In collaborative research, clouds also permit collaborative development and/or testing, allowing each of the collaborators to access the product centrally, without keeping multiple copies at each location. For example, TCS Innovation Labs is working with a research partner to create an advanced visualization portal for social video documentaries. While TCS has been responsible for technology, its partner has been responsible for content creation. A convenient approach has been to provide a set of powerful content creation and visualization tools on a shared server, accessible to both. Amazon Web Services (AWS) was selected as the platform for this, providing cheap storage costs and dynamic scalability of processing power and storage; this allowed the incremental addition of content and sophisticated video processing tools, without requiring upfront capital investment from either partner. Another advantage of such shared environments is the speed of content ingestion and technology feedback. Clouds also reduce the costs of travelling to meet partners, and facilitate remote and mobile operations.

What does the future hold?

Cloud computing is evolving and is predicted to have a great future. We can expect partnerships between different companies to add a new dimension to cloud services. For instance, Yahoo and TATA Computational Research Laboratories (CRL) recently announced an agreement to jointly support cloud computing research in India: CRL's supercomputer 'EKA' (the world's fourth fastest supercomputer) will run Yahoo's open-source distributed computing software to support data intensive computations.

However questions still remain regarding cloud's data security. Currently, there is no provision for inter-cloud operability of an application; such a feature would allow users of a single application to choose between various services from different providers.

With the evolution of cloud we can expect the present services to grow, with many new and attractive features enhancing cloud computing as an e-infrastructure for research and business solutions.

NEWS

The world's largest digital library of 18th century printed books grows even larger for UK academic community
Released: 30 June 2009
Web link: www.jisc.ac.uk/news/
stories/2009/06/ecco.aspx

Following an agreement between JISC Collections and Gale, part of Cengage Learning, Part 2 of the ambitious digitisation project, Eighteenth Century Collections Online (ECCO), is now available free of charge to all UK academic institutions currently subscribing to ECCO. With 45,000 new titles selected from the English Short Title Catalogue, ECCO Part 2 adds close to seven million pages to the size of the collection - at no extra cost for subscribing institutions.

The new collection includes 96 new editions of Shakespeare, a unique edition of Allan Ramsay's 'Above Nine Hundred Scots Proverbs' (Glasgow, 1781, from the Bodleian Library) and a new author (Joseph Fox) with his 'Parish Clerk's vade mecum' arising from his work as parish clerk at St. Margaret's, Westminster. As with Part 1, ECCO Part 2 includes almost 13,000 titles from British Library's extensive holdings of eighteenth-century material – new additions include 'A Catalogue of Vocal and Instrumental Music' engraved, printed and sold by James Blundell in 1781 and a unique copy of 'The Fairing: or, a Golden Toy', a small illustrated book for children.

New supercomputer to reel in answers to some of earth's problems
Released: 29 July 2009
Web link: www.innovations-
report.com /html/
reports/earth_sciences/
supercomputer_reel_an-
swers_earth_039_s_prob-
lems_136822.html

EMSL's Chinook supercomputer by HP commissioned for research. The newest supercomputer in town is almost 15 times faster than its predecessor and ready to take on problems in areas such as climate science, hydrogen storage and molecular chemistry. The $21.4 million Chinook supercomputer was built by HP, tested by a variety of researchers, and has now been commissioned for use by Pacific Northwest National Laboratory and the Department of Energy.

DARIAH Publishes First of Series of Major Studies of National Digital Infrastructure Development
Released: 14 July 2009
Web link: www.dariah.eu/

DARIAH is pleased to announce the publication of an in-depth study of the digital infrastructure for arts and humanities in Slovenia (http://www.dariah.eu/index.php?option=com_content&view=article&id=76&Itemid=75). This important analysis of an emergent infrastructure focuses on the development process itself; identifying individual initiatives, strategies, collaborations, and accomplishments, as well as the obstacles encountered. Painstakingly researched with a series of in-depth interviews with researchers, infrastructure providers, and government representatives, the report maps all relevant recommendations, strategies, initiatives and key figures supporting and enabling the development of current and future policies in this area. The paper also includes a comparative component – considering developments in Slovenia within a European Union and wider international context. This study, along with several forthcoming national reports, is of interest to all those working towards the development of enhanced opportunities for collabora-

tive and independent research in the arts and humanities. In particular, the focus on the construction and deployment of digital research infrastructures will provide examples and be a helpful guide to those working toward the building of an infrastructure within similar contexts.

Argonne develops program for cyber security "neighborhood watch"
Released: 21 July 2009
Web link: www.innovations-report.
com/html/reports/informa-
tion_technology/argonne_
develops_program_cyber_
security_039_136322.html

Cyber security team wins 2009 DOE innovation, technology achievement award. U.S. Department of Energy laboratories fight off millions of cyber attacks every year, but a near real-time dialog between these labs about this hostile activity has never existed – until now. Scientists at DOE's Argonne National Laboratory have devised a program that allows for Cyber Security defense systems to communicate when attacked and transmit that information to cyber systems at other institutions in the hopes of strengthening the overall cyber security posture of the complex.

First projects granted access to the PRACE Prototype systems - 4.4 Million Core hours to 3 projects
Released: 30 July 2009
Web link: www.prace-project.eu/
news/first-projects-grant-
ed-access-to-the-prace-pro-
totype-systems-4-4-million-
core-hours-to-3-projects

Three projects, two from Ireland and one from France, have been granted access to the PRACE prototype systems. The purpose of this granting access is to enable future Tier-0 users to assess the prototypes and to prepare their applications for the Petaflop/s Infrastructure. The evaluation process has therefore focused on technical feasibility and the expected benefits both for PRACE and the users.

The aim of this feasibility study by Niall English (University College Dublin) is to test new approaches in applying Carr-

Parrinello Molecular Dynamics (CP-MD) methods to simulate interfaces of biological systems with nanomaterials. In particular, the detailed mechanisms of protein and dye binding and adsorption to metal and metal oxide surfaces raise interesting and unresolved questions which only fully dynamic, electronic simulations by CP-MD methods can answer at the present time.

More specifically, the goals of this feasibility study are to assess how these methods may be applied efficiently on PRACE-provided HPC architectures, in particular the IBM BlueGene/P at Jülich, Germany, the Cray XT5 at CSC/CSCS (Finland/Switzerland), and the Intel Nehalem cluster at CEA/Jülich (France/Germany). Code-scaling experiments up to 16.384 cores are planned.

How networking is transforming healthcare

Released: 21 Jul 2009
Web link: www.networkworld.
com/news/2009/072009-internet2-healthcare.html

The Obama Administration's push towards health IT (http://www.networkworld.com/news/2009/022609-obama-e-health-plan-health-it.html) -- including electronic health records and a nationwide health information network -- has one important prerequisite: reliable, high-bandwidth Internet connectivity.

Internet2 (http://www.internet2.edu/about/), a consortium of 200 U.S. universities, 70 corporations and 45 government agencies, has been applying advanced networking to medical research, education and clinical practice for several years. With its Health Network Initiative, Internet2 is underpinning the Federal Communications Commission's Rural Healthcare (http://www.fcc.gov/cgb/rural/rhcp.html) pilot program, which promotes the use of telehealth and telemedicine services nationwide. Mike McGill, program director for Internet2's Health Sciences Initiative, talked with Network World's Carolyn Duffy Marsan about how he believes gigabit-speed networks can transform the healthcare industry.

Medical researchers, educators and clinicians are all looking to Internet2 to show them what is possible to do with a network in the medical environment.

Newly Enhanced pS-Performance Toolkit Version Released

Released: 21 Jul 2009
Web link: mail.internet2.edu/wws/arc/i2-news/2009-07/msg00000.html

Easy To Install Suite of Tools Provides Performance Metrics, Monitoring and Diagnostics To Significantly Improve Users' Network Experience

Indianapolis, IN - July 21, 2009 – The perfSONAR-PS development team today announced at the ESCC/Internet2 Joint Techs Workshop, the release of an enhanced version of the pS-Performance Toolkit. The pS- Performance Toolkit (formerly called the pS-Network Performance Toolkit) is available as a live CD download, which can be burned to disc and used to turn any Intel-based computer into an active network performance node. Based on feedback from early adopters, the latest release bundles the suite of recommended network performance tools in an easier-to-use package with new functionality. With these tools, researchers and network engineers can not only monitor and debug their own networks, but access information from external networks.

CSIR technology supports access to government information and services in rural South Africa

Released: 09 Sept.2009
Web link: www.beliefproject.org/news/csir-technology-supports-access-to-government-information-and-services-in-rural-south-africa

The final pilot of project Lwazi (a Zulu word meaning knowledge or information) takes place on 11 September 2009 at the Thusong Service Centre in Atteridgeville, outside Pretoria. At the heart of this project, which facilitates access for all South Africans to government information and services, lies a winning combination of funding commitment by the Department of Arts and Culture (DAC) and the skills of a group of researchers at the CSIR's human language technologies (HLT) research group. The HLT research group is one of the research groups of the CSIR's Meraka Institute.

Centre for High Performance Computing puts new supercomputer into operation

Released: 15 Sept. 2009
Web link: www.beliefproject.org

The Centre for High Performance Computing (CHPC) has launched phase 2 of its operations. The latest addition to its facilities is the Sun Microsystems hybrid supercomputer. The Minister of Science and Technology, Mrs Naledi Pandor, was the guest of honour at the launch event, which took place on 8 September 2009 at the CHPC in Rosebank, Cape Town. The event was hosted by the CSIR (Council for Scientific and Industrial Research).

EGI Council elected its chair

Released: 28 Sept.2009
Web link: www.eu-egi.eu

The EGI Council elected its chair on 24 September 2009, during the EGEE'09 conference in Barcelona, Spain. Per Öster, representative of the Finnish National Grid Initiative (NGI), was chosen by the European Grid Initiative (EGI) Council members to lead the project's governing body. This election constitutes another important step towards the implementation of the EGI, which aims to establish a sustainable grid service for the European scientific communities. This will be achieved through the creation of a long-term, pan-European grid infrastructure.

A new legal entity (EGI.eu) is currently being established in Amsterdam, and will act as the coordinating body for EGI. Its main role will be to facilitate the interaction and collaboration between the national grid infrastructures of the different participating countries. Together EGI.eu and the NGIs will operate and further develop the pan-European grid infrastructure, in order to guarantee its long-term availability for performing research and innovative work. The EGI Council, which is made up of NGIs as well as other members such as European institutions represented in the EIROforum will govern the direction of the EGI collaboration on a long term basis, and is therefore one of the key bodies in the EGI.

Per Öster has been working since 2007 as Director of Application Services at CSC, the Finnish IT Center for Science.

He has been also actively involved in the EGI Design Study (EGI_DS), in which he has been responsible for the EGI promotion and links with other initiatives. In his statement to the Council members, Öster emphasised the importance of creating a working atmosphere based on trust and transparency, and setting up good working principles, in order that the Council can carry out its mission efficiently, and meet the future challenges.

"In the next few months, important tasks await the EGI Council, including ensuring the success of the EGI proposal to the European Commission and establishing the EGI.eu organization", Öster said. "EGI is a fantastic opportunity not only for scientific collaboration but also to make the latest and most advanced computing resources, applications, tools, and data available to all European researchers. It is a great honour to me to be able to contribute to it as chair of the EGI Council."
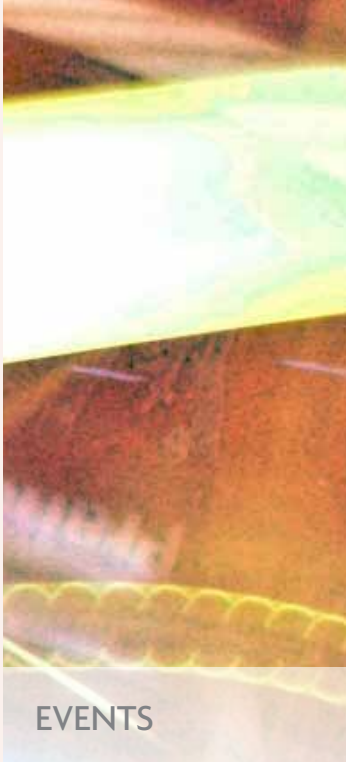
EVENTS

7th eConcertation Meeting
12-14 October 2009, Brussels, Belgium
This event is organised by the Commission with the support of the BELIEF-II project and with the active participation of projects in the e-Infrastructures area. It will bring together the e-Infrastructures' community in order to take stock of the current developments and to discuss the future orientations. The ultimate aim is to develop world-class e-Infrastructures as part of a strategy to achieve European leadership in e-Science. For further information please visit www.beliefproject.org

E-IRG workshop
14-15 October 2009, Uppsala, Sweden
The e-Infrastructure Reflection Group organises workshops twice a year in collaboration with the country holding the EU presidency. Workshops are open to all, and function as incubators for feeding new information and trends into the e-IRG plenum work. Internal meetings for the e-IRG delegates are also scheduled four times a year. For further information please visit www.e-irg.eu

WINDS-Caribe Seminar "Fostering EU-Caribbean cooperation in ICT research"
20 October 2009, Brussels, Belgium
Key ICT research leaders from the Caribbean will meet in Brussels to inform European researchers about outstanding ICT research initiatives from Cuba, Martinique, Dominican Republic and Jamaica and to discuss future strategies to foster EU-Caribbean cooperation in ICT research. During the event, participants will get an overview of the ICT research capacities and dynamics present in the Caribbean region, as compiled by the WINDS-Caribe project, and will have the possibility to find research partners from the Caribbean region. Further information available at http://www.winds-lac.eu/brussels_seminar.

eResearch Australasia
09–13 November 2009, Sydney, Australia
This year's theme of the eResearch Australasia is No boundaries. What challenges are raised by a world with no boundaries? What potential can we unlock? The conference features speakers and demonstrations from the Australian and international eResearch community and provides a catalyst for innovation and collaboration, by bringing together researchers, practitioners, and educators from diverse disciplines and a forum to support the development, enhancement, and harmonisation of national, regional and discipline-specific eResearch infrastructures and services. For further information please visit http://www.eresearch.edu.au/index.html

SC 2009
14-20 November 2009, Portland, Oregon, USA
SC09 returns to Portland, Oregon for its 21st annual conference. Recognized globally as the premier international conference on High Performance Computing (HPC), networking, storage and analysis, SC09 will feature the most interesting and innovative HPC scientific and technical applications from around the world. For further information please visit http://sc09.supercomputing.org/index.php

Second EELA-2 Conference
25-27 November 2009, Choroní, Venezuela
The Second EELA-2 Conference aims to be both an "open conference" and a "user forum" and its Scientific Programme is conceived to reach two main objectives. The first objective is to present a selection of the most im-

pressive results obtained by scientific communities using not only EELA-2 but also other e-Infrastructures in Latin America and the rest of the world and the second goal is to discuss policies and plans for the long term sustainability of Regional e-Infrastructures, with special emphasis on the Latin American one. For further information please visit http://indico.eu-eela.eu/conferenceDisplay.py?confId=201

CHPC National Meeting 2009 and 5th BELIEF Symposium
7-9 December 2009, Johannesburg, South Africa

This event aims to highlight the development of the CHPC within the national cyberinfrastructure intervention supported by the Department of Science and Technology (DST). The importance of the cyberinfrastructure will also be illustrated through a selective successful research projects from different local and international reputable institutions. Furthermore, as a member of the BELIEF-II consortium, the CHPC will make use of this event to cultivate dialogues that are aimed at enhancing its service delivery and strengthening an integrated partnership amongst local and international stakeholders. For further information please visit http://www.chpcconf.co.za

SEE-GRID-SCI User Forum
9-10 December 2009, Istanbul, Turkey

This international event dedicated to distributed and parallel computing with a main focus on research fields of Earth Science. Contributions that investigate parallel and distributed techniques, algorithms, models and applications; present innovative software tools, environments and middleware; focus on various aspects of grid computing; introduce novel methods for development, deployment, testing and evaluation are invited.

5th IEEE International Conference on e-Science
09-11 December 2009, Oxford, UK

The e-Science 2009 conference is designed to bring together leading international and interdisciplinary research communities, developers, and users of e-Science applications and enabling IT technologies. The conference serves as a forum to present the results of the latest research and product/tool developments and to highlight related activities from around the world. For further information please visit http://www.escience2009.org/

PTC'10
17-20 January 2010, Honolulu, Hawaii, USA

Through a network of members extending across more than 50 nations, PTC promotes the development and use of telecommunications and ICT to enhance communications in the Pacific hemisphere. PTC's acclaimed annual conference held in Honolulu, Hawaii every January is the pre-eminent meeting place for telecom and ICT industry executives, professionals, government representatives, educators, regulators, lawyers, researchers and many others. This senior-level event attacts over 200 Presidents, Chief Executives and other top-level industry leaders along with over 500 "C-suite" and senior executives (COO, CTO, CFO, Executive Vice President, etc.) among the participants. For further information please visit http://www.ptc.org/ptc/index.php?q=about-ptc

eLearning Africa 2010
26-28 May 2010, Lusaka, Zambia

eLearning Africa is delighted to announce that next year's conference, the fifth in the highly successful series of pan-African gatherings, will take place in Zambia. From May 26 – 28, 2010, the Continent's largest annual assembly of eLearning and education professionals from Africa and beyond will convene in the capital, Lusaka. For further information please visit http://www.elearning-africa.com/

e-INFRASTRUCTURE

The new generation of integrated ICT-based infrastructure. E-Infrastructures, which exploit and seamlessy interconnect several separate components and layers, such as networks, supercomputers and other computing resources, storage, and other remote resources.

eSCIENCE

An innovative approach to research, thanks to the use of advanced technologies of communication and regardless to geographical location of instruments, resources and last but not least, brains. Today, the paradigm is used in several application fields: Astrophysics, High Energy Physics, Computational Chemistry, Biomedicine, Earth Sciences, Metereology, Environmental Sciences but also Finance, Cognitive Sciences and Archaeology.

GRID

Grids are a set of services over the Internet, allowing geographically dispersed users to share computer power, data storage capacity and remote instrumentation. The term Grid was coined in the mid-1990's to indicate the "coordinated resource sharing and problem solving in dynamic, multi institutional virtual organisations". Although Grids are still in a prototype phase, experts believe that they will have a dramatic impact, comparable to WWW, in the next few years.

MIDDLEWARE

A software layer able to manage and allocate resources in an optimal way to all users and applications that need them, just like the Operative System does with programs running on your PC.

NREN – NATIONAL RESEARCH AND EDUCATION NETWORK

A NREN is a unique organization that plans, manages and operates the IP telecommunication network infrastructure devoted to R&E in a country. They implement the network requirements of the scientific and academic community and are the privileged partner for dealing with Telecom operators and provider of apparatuses on their behalf.

NGI – NATIONAL GRID INITIATIVE

An NGI is an organization that coordinates on a national basis the Grid projects and initiative and is entitled to deal with counterparts worldwide on themes such as standardization and interoperability of middleware and procedures. It is the NREN-equivalent at a Grid level.