The Future of Portals in E-Science

Andrea Bosin
Università degli Studi di Cagliari, Italy

Nicoletta Dessì
Università degli Studi di Cagliari, Italy

Maria Grazia Fugini
Politecnico di Milano, Italy

Diego Liberati
IEIIT CNR c/o Politecnico di Milano, Italy

Barbara Pes
Università degli Studi di Cagliari, Italy

INTRODUCTION

Scientific experiments are executed through activities that create, use, communicate and distribute information whose organizational dynamics are similar to processes performed by distributed cooperative enterprise units. On this premise, the aim of this article is to discuss how a portal-based approach can support the design and management of cooperative scientific experiments executed with a strong information and communication technologies (ICT) support and in a distributed manner, hence named e-experiments. The approach assumes the Web, Web services and the grid as the enacting paradigm for formalizing e-experiments as cooperative services on various computational nodes of a network. A framework is proposed that defines the responsibility of actors of the e-experiment and of the e-nodes in offering services, as well as the portal architecture through which the e-experiment resources can be accessed. By discussing a case study in the field of bioinformatics, the article shows how an e-experiment can be planned and executed starting from a set of Web services inserted in a portal and invoked upon the possibly underlying grid structure.

BACKGROUND

According to what has been anticipated in Knuth (1993), scientists will be more and more involved in work on biological challenges, that could only be equated with computation (Adleman, 1998). These views suggest that future directions in computer science will significantly influence biological or, more generally, scientific experiments. The concept of "what an experiment is" is rapidly changing in an ICT oriented environment, moving from the idea of a local laboratory activity towards a computer and network supported application including the integration of:

- a variety of information and data sources;
- the use of existing software systems allowing the potential deviation from a predetermined sequence of actions as well as the verifiability of research work and accomplishments;
- the specific and distributed expertise of the involved scientists.

Lab experiments are still often developed in isolation, and tend to be small-scaled and specialized for ad hoc applications. On the other hand, the technology of portals can provide a strong potential for integration of data sources, applications, and tools with broad *reuse* capabilities. In order to allow researchers to be internetworked in a cooperative enterprise style, cooperative ICT environments, in particular implementing portal style interaction modes, have shown to be a feasible solution for interconnection, integration, and large information sources sharing (CooPIS, 2005). In particular, a portal for e-science could merge different competencies, could enact user interaction via multichannel access (Pernici, 2006), although privileging the Web mode, and could provide an harmonized view of differently designed experiment databases, as well as a uniform set of tools for conducting the experiments and for interpreting their results. Moreover, portals can bring together different user communities, provide a shared work area, and contain the necessary metadata enabling the discovery and management of distributed e-science facts.

Additionally, high-performance computing and communication technologies are enabling e-scientists to study and explore complex systems. These technologies allow for new

forms of collaboration over large distances together with the ability to process, disseminate, and share information (Brown, 2003). Global-scale experimental networking initiatives have been developed in the last years: the aim is to propose advanced cyber-infrastructures for e-scientists through the collaborative development of networking tools and advanced grid services (De Fanti, 2003; Newman, 2003).

As an example of what is currently proposed in bioinformatics, a systematic approach to disseminate proteomic data trough sharing an experiment data repository is presented in Taylor et al. (2003): The PEDRo system offers some distributed facilities to establish the provenance and relevance (to the researcher) of datasets, and to allow nonstandard searches in a community of users. However, such work is mainly focused to a repository for sharing experimental data, while the approach we are arguing about is a more complete environment endowed with tools also for designing and managing experiments.

E-EXPERIMENTS

On this premise, the aim of this article is to present methods and tools for deploying portals for cooperative scientific experiments (*e-experiments*). A portal is viewed as an application area with tools able to support e-experiments, for example, in the field of postgenomic, as well as conventional experiments allowing one to validate and/or to refuse hypotheses and models generated by bioinformatics in an iterative manner. These experiments will be a benchmark to test the functionality and the usage of tools and methodologies generated above, particularly by users with a biological background not very skilled in information and communication technology.

Web services (Alonso, Casati, Kuno, & Machiraju, 2004), even over a grid infrastructure, are the enabling technology considered in the portal architecture proposed by this chapter to support the simulation/execution of different classes of experiments, from visualization (browsing and search interfaces), to model identification through clustering and rules generation, in application fields such as drug discovery, micro-array data analysis, or molecular docking (Bosin, Dessì, Fugini, Liberati, & Pes, 2006a).

We now present the basic idea of a portal for e-science by illustrating the features of a distributed environment for e-experiment management, using the concepts of Web services deployed on the grid framework provided by myGrid (2005).

The purpose of the portal is to provide a public access system where (or through which) scientists, and possibly also other actors, can make their experiments, or data, public, and where other scientists can notify their availability to execute a distributed e-experiment. The e-experiment resources (data, tools, even experts) are the basic structures of the portal. Data have to be made public according to data privacy rules established in the e-science community. On these data, the portal provides functions for e-experiment planning, execution, result publication, and forums. The purpose is to provide the actors with a view (called Myportal) of the experiment status and results. Moreover, the portal is intended to support actors in fulfilling a set of tasks, such as the compilation of forms, as well as the directions on how to perform an experiment.

We observe that the portal is intended to manage e-experiments, by addressing accredited (i.e., publicly funded) and private agencies. The idea is to make a pool of data and tools public to private and public actors, thus creating a wide and therefore effective e-science area. The interest by private actors in adhering to the portal must be regulated by internal policies consistent with the visibility that the cooperative environment decides to assign to the e-science environment data. Figure 1 shows the overall architecture of the e-science portal.

Let us illustrate the portal subsystems.

- The *content management subsystem* acts via a Web front-end interface to the users and is strictly related to the portal storage subsystem containing the data of the e-experiments.
- The *profiling and security subsystem* ensures identification, authentication and tracing of users and of visits to the portal.
- The *storage/esource index subsystem* consists of:
  - an application database, containing all internal data, and referencing remote experiments and their data in the environment; and
  - an index of experiments, acting as a registry that identifies data accessible through the portal.
- The interoperability subsystem manages and integrates the communication towards the actors of the e-science environment through the interoperability back-end, which stores communication objects necessary to interface heterogeneous environments through the middleware layer. The portal interacts with such middleware (composed of Web services and possibly of a grid), in order to access a variety of network resources, as depicted in Figure 1.

In such a modular architecture, e-science providers can easily plug in the portal via the middleware and interoperability back-end components. They have just to install into their applications the Web services necessary to communicate and share resources, according to the portal standards and policies.

The described architecture is in the line of portals used for allowing distributed but federated subjects to commu-