Setting the Framework of E-Collaboration for 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
Italian National Research Council, Italy

Barbara Pes
Università degli Studi di Cagliari, Italy

INTRODUCTION

Collaboration for e-science, namely executing experiments in a cooperative way by sharing data, tools, and expertise towards a common scientific goal, is becoming more and more appealing in a context like the scientific community. In such a context, a critical mass is needed to address very important and complex new questions arising because of the increasing availability of experimental data made possible by continuous technological achievements. An effective collaboration can be set in place by using the increasing empowerment of the Internet to perform such distributed laboratory. By designing a lab environment able to involve accredited actors, public research centers could benefit from technologies and tools, and could become the first promoters and the key players of collaboration initiatives. Moreover, a possible interest by private actors in adhering to such labs should be stimulated in a proactive way by contracting the mutual beneficial and burdens in such a way that both kinds of actors, as well as the civil society (which eventually funds these initiatives) could all take advantage of such arrangements. In addition, privacy concerns that could arise in this type of environment are instead easily granted under appropriate rules that leave public-only scientific data, while keeping both individual and sensitive data and information protected by copyright reserved.

In order to present such initiative as attractive for potential actors, the needed technological tools and applications should be presented as a minimal and non-intrusive tool suite for organizations. To this aim, the technology of Web services (Hey & Fox, 2005) proves very effective, in that the code to be installed at local sites is minimal, easily installable, and easily invoked through a peer-to-peer paradigm. Standard tools coming with a Web service environment should be used, such as XML, UDDI, and WSDL, to exchange data in interoperability environments, to create the services and their public registries, and to communicate in the cooperative environment.

Case studies and best practice in the area of information and communication technology (ICT)-supported collaboration (Tilley et al., 2004) show that overall architectures enabling cooperation should be based on the following logical levels:

1. **Front-end:** Acting as the presentation layer related to the e-experiment
2. **Business logic:** Dealing with collaborative applications
3. **Storage-containing:** A data base of experimental information; an index of experiments, acting as a registry that identifies data and users of the experiment; an *interface* to a set of *informative databases* (for statistical purposes data about experiments, statistics, standard classifications of terms categorized by domains (e.g., medical, biological, remote monitoring), and other reper-

tories and indexes to be consulted in data mining mode (Newman, Ellisman, & Orcutt, 2003).
4. **Interoperability:** For communication towards the centrers adhering to the cooperative environment.

The *management* of experiments between organizations is currently available either in insecure systems, which are appropriate for non-commercial academic collaborations only, or by using proprietary solutions from single vendors that all organisztions participating in the experiment are required to use. This creates a very expensive entry cost for an organization, which is worthwhile only if the collaboration survives for a long period providing substantial benefits. This can be achieved when the existing ICT endowment of each cooperating organization has the potential to interoperate easily.

BACKGROUND

Collaborative environments make use of emerging proposals of networked services that have enabled new types of applications in the field of information systems. These new applications are set in place and executed by several geographically distributed interacting organizations, exchanging data through the network and the Web. Information systems using such paradigms are called *cooperative information systems* (*CoopIS*)—that is, distributed information systems that can be either employed by users of different organizations under a common goal or operate for local activities, hence maintaining their autonomy (CoopIS, 2005).

A core issue of collaborative environments consists of *e-applications* (Mecella & Pernici, 2001), namely orchestrations of e-services provided by different organizations on the Internet. The data exchange and the interleaved execution of processes in such systems bring about issues bound to inter- and intra-organizational structures, to a plurality of actors in the distributed system, and in the heterogeneity of policies existing at the various sites where the portions of the distributed process are executed.

In order to define a framework for cooperative scientific environments, it is necessary to consider their requirements in terms of resource sharing and privacy. In particular, policies have to be defined that constrain the behaviour of system components. Tools for policy definition and management are becoming available also for dynamic adjustability of applications. Benefits of policy-based approaches include reusability, efficiency, extensibility, and context-sensitivity. Policies are often applied to automate network administration tasks, such as configuration, security, recovery, or quality of service. The adoption of a policy-based approach for controlling the cooperation in an e-science environment is appropriate, since cooperation behaviour can be represented and managed both for policies known at design time, that is, data and resource sharing, or privacy, and for policies that need to be negotiated dynamically among the cooperating processes, such as the quality of services (e.g., performances or delays). It is the responsibility of the experiment manager to specify policies through adequate support tools; for example, using WS Policy with Attachments (Bajaj, Box, Chappell, Curbera, Daniels, Hallam-Baker, et al., 2006), as suggested by standard proposals (W3C, 2004).

The complexity of the *remote components* (i.e., the participative software package at the site of the collaborating actor) that needs to be installed at each adhering to the system, in order to cooperate, has to balance:

- **Performance requirements:** Suggest exporting most of the services to lower the system workload and limit the communications among the systems
- **Privacy requirements:** Prohibit replicating data present on the central system to make them available in the local system
- **Adaptability requirements:** Make the architecture compatible with all the systems used at the various organizations with no substantial modifications
- **Minimal modification requirements:** Impose a limit to the interventions on the applications in the local system
- **Minimal programming efforts:** Required for integrating the existing legacy applications and for further extending the functionalities that have been delivered as the core system portions

Considering also the indications of Web service providers (Tilley et al., 2004), an overall view of the categories of services realizing a collaborative environment can be as follows: