Chapter XL
Design and Managing of Distributed Virtual Organizations

Diego Liberati
Italian National Research Council, Italy

ABSTRACT

A framework is proposed that creates, uses, communicates, and distributes information whose organizational dynamics allow it to perform a distributed cooperative enterprise in public environments even over open source systems. The approach assumes Web services as the enacting paradigm, possibly over a grid, to formalize interaction as cooperative services on various computational nodes of a network. A framework is thus proposed that defines the responsibility of e-nodes in offering services and the set of rules under which each service can be accessed by e-nodes through service invocation. By discussing a case study, the chapter will detail how specific classes of interactions can be mapped into a service-oriented model whose implementation will be carried out in a prototypical public environment.

INTRODUCTION

Science is nowadays more and more a question of a critical mass of skilled people—often with complementary backgrounds—becoming a unique global organism in pursuit of a substantial common goal. This was true last century in physics, and it is going to hold true now in biology. A particular difference is worth mentioning: while traditional specialized workforces have historically been concentrated in unique sites with special instruments in physics, the availability of ICTs enables full networks to be built in remote sites, together with the absence of the need of a big site for many of the problems faced in biology, or even natural distribution, like, for instance in geology, makes it interesting to resort to a possible network of experimenters, physically not necessarily together, but logically co-present in the framework of the same cooperative big experiment involving all their complementary competences. In the paradigmatic bioinformatics application domain dealt with in this chapter within the framework of e-science (Hey & Trefethen, 2004), collaboration may be even more needed than in other areas such as financial transactions (although any ICT-assisted business interaction could be described in the same way).

A proper virtual organisation is thus required in order to manage the workflow of information over a given network in order that cooperation among the nodes does not deteriorate into an
absence of competition among potential partners with the same competence but at a different level of capability. Flexibility is essential, due to fast-developing competences and technologies, and will ensure that any redesign(s) will not have to be done “from scratch” every time. Proper orchestration would allow a better exploitation of every component. Interoperability among components is another important characteristic. A networked organisation such as this may be logically defined through workflows acting over Web services, possibly in a grid context, and made as public as the contributing actors agree—not just among them but also to a wider community—by exposing data and/or results while keeping reserved within the cooperating group the workflow. The technology addressed in this chapter would thus allow the most competent scientist to design the workflow even if he is not at a very high level of competence in IT issues (which are, in a sense, logically embedded), as is common in application contexts like the bioinformatics faced here, as well as in other application domains, where the proposed tools can also be applied. It is worth noting that such an approach is also a way to overcome a certain form of digital divide, if encompassing not just extremes being able or not to access to Information and Communication Technologies, but including also not homogeneous situation of partial access, provided that minimal requirements such the one here proposed are satisfied. In such a sense, approaches like the proposed one could provide a tremendous improvement in the capability to easily recruit a higher portion of the human capital not yet fully involved in global research.

BACKGROUND

Grid computing (Foster, Kesselman, Nick, & Tuecke, 2002) has facilitated the growth of virtual organizations: the basic idea is to have a framework, unbound to a specific technical solution (Foster & Kesselman, 2004), where people (not at the same location, but having direct access to linked ICT heterogenic resources) can achieve a common goal through collaboration with each other.

To do this, every resource shared across the network has to be associated with a sufficiently high-level interface in order to allow the corresponding service to be remotely controlled, locally granting a sufficient quality of service with respect to the usual problems of (say) faults, handling of exceptions, security, and so forth.

Web services (Booth et al., 2004; Christensen, Curbera, Meredith, & Weerawarana, 2003; Graham & Treadwell, 2004) are among the proper technologies to address such resource virtualisation problems, which form the local part of the whole networking problem.

The global side is, instead, the proper orchestration (Peltz, 2003) of the whole subset of the possibly heterogeneous and probably delocalised needed resources. Workflows (van der Aalst & Hofstede, 2003) address the orchestration of such a distributed process.

NETWORKED EXPERIMENTS

An easy way to face a complex problem is a divide-and-conquer strategy, aiming to decompose a set of simpler subproblems. In this case, the workflow is an orchestrated set of virtual procedures, each of them defined by the core computation on input data in order to produce output data related to a single activity. Such orchestration can easily be described as a way of reminding the classical general schema of petri net as being useful to define at an abstract level of sequenced interaction among actors.

Every single procedure—even if software-modeled—also usually involves human experts, adding specific value to the automatic capabilities of the integrated tools, whose geographical location is impacted, if not bounded, by the standard live(s) of the human actors, whose quality is usually of paramount importance in sophisticated applications such as e-science.

Human actors, skilled in the application domain (but not necessarily in the details of every