Chapter 4.16
Collaborative Information Management System for Science Domains

Ozgul Unal
University of Amsterdam, The Netherlands

Ersin C. Kaletas
University of Amsterdam, The Netherlands

Hamideh Afsarmanesh
University of Amsterdam, The Netherlands

H. Hakan Yakali
University of Amsterdam, The Netherlands

Louis O. Hertzberger
University of Amsterdam, The Netherlands

NEED FOR COLLABORATION

With the increasing need for collaboration in different science domains, a lot of research activities are now focused on the mechanisms and infrastructures supporting advanced collaborations among pre-existing, distributed, heterogeneous, and autonomous organizations. Collaborating organizations typically share some common objectives and in order to achieve them they need to share their information and resources. One prominent requirement is to access each other's data or databases through a secure infrastructure. Biodiversity is one such science domain. Challenges in biodiversity information management are being addressed in the project ENBI (European Network for Biodiversity Information) (ENBI, n.d.). A summary of the information management

challenges in different science domains is given in the second section, Information Management Challenges in Science Domains.

In this context, the CIMS introduces promising solutions to cope with these challenges. In general terms, CIMS refers to the set of components and mechanisms that together constitute a generic information manipulation framework to support the interoperation and data sharing among collaborating members (Guevara-Masis, Unal, Kaletas, Afsarmanesh, & Hertzberger, 2004).

The proposed CIMS involves three main paradigms and technologies, consisting of: (1) federated database architecture, (2) virtual organizations paradigm, and (3) grid technology. Introductory information about these technologies and paradigms and their major benefits are covered in the third section, paradigms and technologies.

INFORMATION MANAGEMENT CHALLENGES IN SCIENCE DOMAINS

Below, a summary of challenges and requirements of information management in science domains is provided. The points addressed here correspond to the main motivations behind the need for a CIMS.

- **Distributed and Heterogeneous Databases:** One of the challenges in many science domains is to have transparent access to distributed and heterogeneous data sources. Existing organizations typically employ different data structures depending on their specific needs. A network of organizations must consider such differences for providing effective mechanisms to integrate or inter-link and homogeneously access such databases.
- **Autonomous Organizations:** Organizations from a variety of science domains represent autonomous nodes running independently of each other. Each organization must be able to autonomously decide to share a part of their local resources or services with certain other specific organizations, based on some agreements.
- **Need for Collaboration:** It has become more clear that the collaboration between different organizations, activities, and users in science domains is important for an improved understanding and thus for achieving better results in the domain. However, most organizations do not want to actively cooperate, because of for example the sensitiveness of some data that they have. Therefore, mechanisms supporting collaboration among organizations and at the same time taking these kinds of data into account are needed. With the existence of such a coordinating mechanism, organizations can more easily decide to collaborate.
- **Security and Access Rights:** The issues of trust, inter-organizational agreements, etc. play an important role in resource sharing among organizations from different science domains. Access rights and visibility levels must be taken carefully into account especially considering sensitive data. A hierarchy of visibility levels and access rights needs to be defined for different users or user roles.
- **Performance Requirements:** The amount of online data from different science domains, such as biology, physics, and astronomy, is increasing at a high pace. Considering the need for sharing these data with others and making them accessible through a network, the demands for performance and robustness are high. Furthermore, some modeling and analysis activities require computationally intensive algorithms. Thus, an infrastructure providing high performance distributed resource and data management facilities is needed.