Scientific Information Management in Collaborative Experimentation Environments

Kaletas, E.C.

Publication date
2004

Appendix B

Related Projects

In addition to the VLAM-G Project, which formed the base for this thesis work and described in Subsection 1.5.1, the author has been involved in many other projects that contributed to the work described in this thesis. This appendix provides an overview of these projects.

Dutch ICES/KIS II VLAM-G – MACS Project (1999-2002)

The main goal of the MACS System [9] was the development of innovative mechanisms and functionality for storage and handling of large amounts of data generated by the Material Analysis of Complex Surfaces (MACS) experiments. Specific requirements of these experiments, e.g. managing large data sets and metadata management, contributed to the refinement of the generic data model developed for experimental information (i.e. EEDM) and the VIMCO functionality. MACS System also benefited from the generic functionality provided by VLAM-G, in specific from the information modelling and management functionality of VIMCO. A database called MACS DB is designed and developed based on the EEDM to store experiment descriptions and metadata. Several experiment templates are defined and stored in the MACS DB. Data is stored as files in the NetCDF format [74].

Dutch NWO-BMI Flexwork Project (2001-2004)

The aim of the Flexwork project is to combine data from different biomolecular techniques in a flexible manner for the Dutch scientific community. The project will allow the users to define their own data gathering and analysis by linking bioinformatics tools in a virtual laboratory environment that can be approached via Grid. Among the key objectives, a flexible data integration framework to support system biology and object-oriented data modelling of DNA-sequence and gene expression data can be named. Highly heterogeneous data types involved in the target experiments, and the requirement for the modelling and storage of annotation data allowed to define generic models for the analysis results.

European 5FP FETISH Project (2000-2002)

The goal of the FETISH–Federated European Tourism Information System Harmonization project was to integrate the fragmented tourism information systems and their IT-based services into a federation of distributed resources. The project developed a Jini-based service repository, a catalogue for service interface definitions, an ontology for tourism, and a

business process editor/manager for defining and executing value-added services. FETISH employs a service-oriented architecture [172]. The main task of the CO-IM Group in this project was to design and develop the Service Interface Definitions Catalogue and the Federated Access Rights Manager for defining and supporting the enforcement of access rights on tourism services within the virtual enterprize context [173, 172].

Dutch Electronic Toll Collection Project (1996-2001)

The Dutch Government intended to implement an Electronic Toll Collection system to reduce the traffic congestion. An Automatic Debiting System (ADS) was required for the toll collection not to interfere with normal traffic flow. The role of UvA was to evaluate the technical feasibility of ADSs. For this purpose, a modelling and simulation approach was developed, and a software environment to perform these simulations was realized [21, 20, 22]. The data collected along the road was stored in a Matisse database, together with the meta-information already produced by the ministry. Studying both data sets as well the planned access patterns to this data, a database model with proper indexes was developed. The data consisted of two very large files containing over four million entries, which were parsed and loaded into the database. The database is used for the analysis of traffic data in MATLAB through an interface which is developed to transparently access Matisse from the MATLAB environment. Currently there is access to this database with traffic measurements with a unique level of detail [22]. The same database and the parser tools were also used for evaluating the performance of the Matisse ODBMS.

Dutch JERA Project (1996-1998)

The JERA project [174] focused on the realization of a fault tolerant distributed Web server. Several aspects were taken into account; namely: object caching, databases, fault tolerance, and scalability. The main task of UvA in this project was the design of a Web server architecture to cope with the ongoing increase of the Internet requirements. The design addressed two of the parameters: the need for a powerful data management system to support the increasing complexity of user requests, and an efficient caching mechanism to reduce the amount of redundant traffic. The proposed solution was based on a distributed database framework to replace the file system and a collaborative caching system that builds up a dynamic hierarchy of cache servers [175]. In order to achieve scalability, a distributed database framework integrated with the Web server was designed [175, 176]. A prototype of this architecture was developed and used in many other projects; for instance for storing large scientific data sets [177] and in the MegaStore project for storing music data [178].

European ESPRIT 4 ARCHES II Project (1995-1999)

The prototype developed in the JERA Project was also used in the ARCHES project for evaluating a gigabit switch [179]. The purpose of this project was to bridge the gap for the utilization of existing high speed network technologies (HIC) developed in various European projects for the emerging standard of Gigabit Ethernet (GE). New, scalable switches based on the HIC technologies were designed and put to the test. The role of UvA was to validate the technology developed in the ARCHES project as an expert end user of this technology. Web technology was chosen as the evaluation application area and, the Web server software with a distributed database replacing the file system, which was developed in the JERA project, was extended and used as the demonstration application [179, 180].