

OpenAdap.net: a Community-Based Sharing System

Alessandro E.P. Villa^{1,2,3}, Javier Iglesias^{1,2,3}, Solange Ghernaouti-Hélie³

¹Inserm U318, Laboratory of Neurobiophysics,
University Joseph Fourier, Grenoble, France
{Alessandro.Villa, Javier.Iglesias}@ujf-grenoble.fr

²Laboratory of Neuroheuristics, University of Lausanne, Switzerland
<http://www.nhrp.org/>

³Information Systems Department, University of Lausanne, Switzerland
sgh@unil.ch
<http://www.hec.unil.ch/isi/home/>

Abstract: OpenAdap.net is an Open Source project aimed at breaking the barriers existing in the flow of data access and data processing. The infrastructure will make possible the dissemination of resources like knowledge, tools or data, their exposure to evaluation in ways that might be unanticipated and hence support the emergence of communities of users around a specific domain. The architecture is designed by analogy with a virtual distributed operating system in which the dynamic resources are presented as files in a structured virtual file system featuring ownership and access permissions. OpenAdap.net will be open to exploitation during Q3 2006 by networked organisations and alliances taking into account the vital issue of Internet security and privacy.

1 Introduction

The information circulating in the Cyberspace generates a historically unprecedented richness for sharing knowledge and methods of data processing. The software published at Internet in one instance becomes available for the whole society worldwide. However, most of the latest processing and analysis tools developed and provided by research laboratories and businesses are based on specific software and computer platforms.

The current situation generates a lag until these contributions become known within the same field of competence and restrict severely their availability, in particular for cross-fertilization application in other fields. This delay may provoke the re-invention of methods for data processing and, more generally, the re-discovery of the same knowledge. At the societal scale this delay inhibits the development of added-value activities originating from sharing the knowledge and generates repeated unnecessary expenses and even, erroneous applications. The societal benefits in terms of business developments, market diversification and creation of employment generated by the fast circulation and ease of access to the results of the Human Genome Project illustrate the interest in developing an

open and dynamic adaptive network for resource sharing with emergent properties within the Information Society.

Our project is aimed at breaking the barriers existing in the flow of data access and data processing increasing the overall cost of information processing and restricting its availability to developing countries, by avoiding re-invention of existing software and save time and expenses by the whole society and even prevent incorrect applications. The ability to evaluate and to find the best available solution to a given problem may have significant impact on areas such as economics, physics, environmental sciences, meteorology, and health. By adopting the best available method of analysing a set of data, it is possible to test alternative solutions and choose the best from its overall performance.

In the construction industry, which is playing an important role measured by investments in all economical systems, even small enhancements will make resources available to other purposes. Reduction of the energy consumption during the whole lifecycle of a building will decrease pollution and save money. By setting up an environment where people are used to working with state of the art technical solutions, companies throughout the World will be forced to increase their competitiveness work in a completeness way.

We envision the emergence of communities of users sharing resources like knowledge, tools, data, etc. for their specific domain. The key to this vision is the development of a sharing architecture that is independent of a specific type of information: **OpenAdap.net**. Through programs like web portals and standalone applications interacting with the system, the community will share the set of tools that each of these domains typically use through members contributions.

Individual users of **OpenAdap.net** can be classified as either contributors of shared resources, or end-users of such. People who develop and provide new methods of data analyses are able to share their contribution and people interested in processing their own data or access data stored elsewhere (e.g. in a community database) can extract the results of their analysis. In addition to individual use, **OpenAdap.net** is open to exploitation by networked organisations and alliances, by providing an infrastructure with which they can share and integrate resources and develop new business solutions and opportunities.

The philosophy behind **OpenAdap.net** is that users privacy is as important as contributors traceability. We believe that information sent by an end-user for manipulation by a contributors resource should be anonymized, despite the fact that all the transactions are identified and that the activity is tracked like in any computer system (not to mention the web). In contrast, meta-information concerning the contributed resource like program authorship and version number should be made available to the end-user, as a mark of diligence to the contributor, but also to point the responsibilities in the information processing chain and enhance quality assessment and reproducibility. Each task submitted to the system will be returned to the submitter with a record attesting where the information travelled, which processes were applied and who was responsible for the computers and the programs.

By sharing a program and the computer running it, for example, contributors will keep complete control over their authorship as well as the source and binary codes for the software. At the same time, they will be responsible for maintaining and checking the quality of the results that the community will use and validate. From the viewpoint of contribu-

tors, OpenAdap.net makes possible the dissemination of resources, and their exposure to application and evaluation by a broader users community. The support for broader evaluation of programs and data sets is particularly important in the research arena, as is the ability of other researchers to reproduce computational results.

In a second step, OpenAdap.net will become an environment where new techniques and methodologies will gain access to a wide range of users, possibly beyond the direct community boundaries to the adjacent domains. The system allows the dissemination of resources across domains in ways that might be unanticipated by the original contributor. For example a solution developed by experts in the field of medical statistics could permeate to the field of botany, and become a key tool in the analysis of tree distributions. In this way, OpenAdap.net supports transdisciplinarity by breaking current boundaries to resource sharing.

The OpenAdap.net project is not directly aimed at the production of new methods of analysis, but the platform helps the community to aggregate their already existing tools by dynamically composing new information processing chains using the output of existing programs as the input to others, thus opening the way to cross-fertilization and serendipity.

2 Architecture

OpenAdap.net stands in the area of complexity and aims at providing a distributed environment where tools of all kinds (applications, data, knowledge, etc.) can be accessed transparently via the Internet. At this time, three architectures are used in this field: Grid, Web-services (WS), and Peer-to-Peer (P2P). The peculiar strongholds of these architectures are briefly described here and synthesized in Table 1.

Grid: Each user has a large dataset to manipulate with one application distributed on a set of computers. The problem addressed by the Grid is to distribute the processing of large data sets. The required application is copied to distinct computers over a network with each treating a portion of the data. The results are recomposed from all the partial results at the end of the treatment. End-users have control on both data and applications, but little information on remote execution hosts.

Web Services: Many users exploit the same services permanently provided through a centralized authority. Web Services provide a secured and controlled access to applications, usually to collaborators in the same company or institution. The goal is to provide distributed access to the same references, databases or applications. The system is articulated around a repository where the service interfaces are published in a self-contained form. The architecture is rather static, as services seldom change and are expected to be permanently available. End-users have no control over the applications and little information on remote execution hosts.

P2P: Many users exchanging pieces of data in an unsupervised way. P2P (peer-to-peer) systems address the data-sharing problem. Copies of the applications installed on end-users computers keep open connections from one computer to peers, forwarding queries and results back and forth until a match is found somewhere on the network. The ar-

	Data treatment istribution	Hardware resource allocation	Hidden execution hosts	Application sharing	Published application interface	Data sharing	Highly dynamic system	Transparent user/resource connection
Grid	×	×	×					
WS		×	×	×	×			
P2P					×	×	×	
OAN	×	×	×	×	×	×	×	×

Table 1: Comparison of OpenAdap.net with the other approaches over Internet.

chitecture is open and competing implementations coexist on a self-organized network. End-users have control over their data and information on the peer hosts. It is interesting to note that end-users tolerate incredibly poor service quality and that this architecture raises social as well as technical issues.

OpenAdap.net (OAN) falls somewhere between these three architectures exploiting several of their interesting aspects, but with the intention to address a two-way problem: To provide to a community of users in the same domain a means to interchange their resources in an open, dynamic and secured way and to provide to a community of users the access to the exploitation of information processing solutions contributed by users belonging to other communities. End-users have control over their data, but do not need to manage the resources, nor do they have complete information on remote execution hosts. Collaboration within the OpenAdap.net network allows the dynamic integration of these resources, possibly yielding new or previously unexpected composite resources. This can be summarized as follows: *Many users interchanging resources (data, applications, knowledge,) dynamically provided by interconnected domain-oriented brokers.*

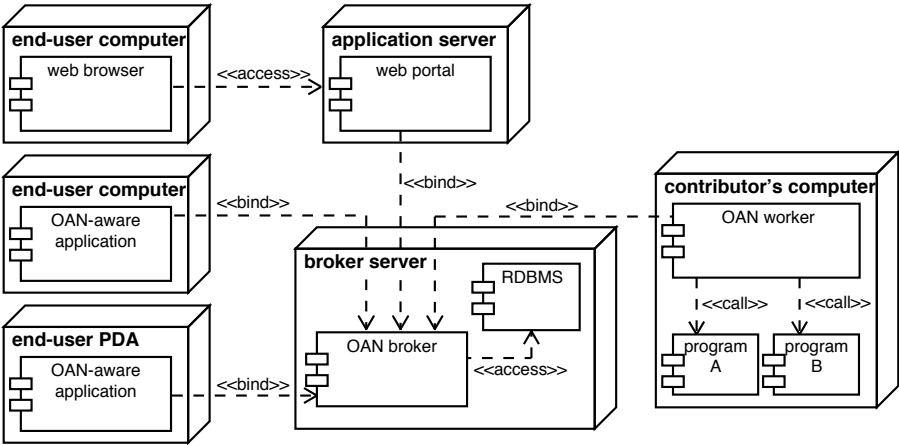


Figure 1: Deployment diagram for the OpenAdap.net components. Boxes represent different computers, rectangles represent processes, and arrows represent inter-process communications over Internet.

OpenAdap.net is a distributed system composed by three types of interconnected compo-

nents: brokers, workers and OAN-aware applications (see Figure 1). A broker is a process permanently running on a server in charge of managing a community of users and dispatching tasks and results on their behalf. Workers are processes shared by community members in charge of giving secured distant access to contributed resources like programs or data. OAN-aware applications are pieces of software (standalone applications, web portals, command line tools, etc.) providing access for an end-user to the community shared resources through identified connections to a broker.

The components are running on a network of computers, each of them defined by their specific CPU architecture, operating system (OS), amount of memory and available programs. Their resources are partially shared in a dynamic way. **OpenAdap.net** is designed by analogy with a virtual distributed OS in which all the resources are presented in a structured virtual file system. Using this high-level paradigm, resources are assigned to files in the file system tree. Security is enforced through ownership and access permissions defined on a per-file basis. **OpenAdap.net** goes beyond a traditional OS as the configuration is highly volatile. The file system structure and contents are the result of the constant runtime aggregation of several sources of information: the user database, the inter-process message queues status, the worker status, etc. A dedicated URL name scheme (`oan://`) will be proposed to identify each file in a transparent and interoperable way.

The Java 2 Platform was chosen for the implementation of the project, based on portability and platform neutrality requirements. Brokers, workers and OAN-aware applications are loosely coupled, distributed components that asynchronously communicate through a message-oriented middleware (MOM) as defined by the Java Message Service (JMS) API. The use of established industrial standards such as JMS allowed reusing existing Open Source software as a starting point for the project. It can also be expected that JMS implementations available for other platforms will allow applications written by third parties to connect to **OpenAdap.net** brokers and interoperate seamlessly.

Internally, brokers are responsible for decomposing and routing end-user tasks to appropriate workers for execution. A key element in the next stage of development consists in making brokers adaptive and dynamically interconnected into an **OpenAdap.net** network (like a neuronal network). The requests for resources will be processed and dispatched among the components of the system following a set of learning rules dynamically modifying the routing according to, for example, the computing load generated by specific tasks, availability of the resources, or the number of accesses. The rules themselves will evolve and optimize in an unsupervised manner, thus allowing the emergence of unexpected dynamics. In that sense, the required negotiation between brokers (and workers) may be compared to agent interaction. The **OpenAdap.net** network will also be able to self-adapt via learning processes derived from neural and artificial life learning. Such learning might result in new broker-broker connections, reassessments of the value of such connections, specialisation or generalisation or broker behaviour, etc.

The adaptive and behavioural models for the broker implementation represent major innovations of the **OpenAdap.net** project. This is definitely a novelty and a plus to the existing architectures for distributed environments like grids and web services that points out the project expected income to the networked computing field. We believe that SMEs and research institutions will be interested in developing novel interdisciplinary solutions

associated to the psychological and technological aspects of evolvable simulation tools, the psychological environment of remote user support and the formal aspects of artificial processing in resource-sharing.

We also expect to interface **OpenAdap.net** with established distributed systems like grids and clusters. For that purpose, specific workers will be developed to provide **OpenAdap.net** to Condor or Portable Batch System (PBS) interfaces.

Pushing existing paradigms like neuronal network inspired learning rules for the adaptable information processing or the operating system paradigm for the overall communication layout, and the lessons learned for 10 years on the self, dynamically, openly organized content on the web are key aspects of the **OpenAdap.net** philosophy and architecture for resource sharing. Besides, **OpenAdap.net** is an Open Source project designed as an open architecture. Anyone is invited to contribute their own enhancements to the system, via a set of libraries and tools provided by the consortium. Such an initiative is aimed at increasing the impact of the project with all the contributions that competent contributors will imagine and realize within their specific domains.

3 Applications

In the last few years, we have been developing, running and testing a prototype for **OpenAdap.net**. The concept proof and feasibility have been checked. In the last few months, we have been re-implementing the project from scratch based on the experience we have acquired with the prototype. **OpenAdap.net** being an Open Source project released under the GPL and LGPL licences, contributions are welcomed from developers and interested professionals. This is encouraged by the availability of publicly accessible development tools like a version control system, mailing lists and a bug tracker. There is enough experience on the impact of such coordinated and distributed development scheme on the quality of the resulting software to embrace it.

During the prototyping phase, the need appeared to have a portable and user-friendly tool that could provide a fast glimpse on the numerical output of unrelated programs. We searched for a generic front-end for plotting a wide range of graphics obtained through the web, and none could be found that was able to run on multiple platforms without requiring complicated installation procedures, and capable of producing journal-quality graphics output for free. XY-Viewer is a Java application that is being developed for that purpose as a by-product of the main project, featuring a dedicated XML data format. This standalone application can be launched directly from the **OpenAdap.net** portal to visualize files in the appropriate format that can be produced with the help of libraries made available to interested contributors in C and Java under the LGPL licence.

A particular exploitation of the project is the development of web portals tailored to the needs of end-user communities centred on specific domains, possibly by SMEs. Communities will appear first in the domains of competence of the consortium partners. These early adopters will constitute real-life case studies that will assess the validity of the concepts and the usability of the implementation before a broader diffusion. The prototype

has already attracted a few individuals from different research groups to cooperate and share resources in the domain of multivariate time series analysis.

In the future, we expect to give rise to synergies within and between both new and existing communities. The identification of such communities will necessitate further developments such as community specific ontologies and benchmark file repositories. Users will be invited to contribute resources, articles, benchmark data, and to build the domain ontology with the initial help of OpenAdap.net consortium.

Interactions with Publishers and Editors could have an incredible impact on the way scientific dissemination, including visualization, is performed, mainly by facilitating peer review in refereed publications and for results comparison and validation for peer scientific readers. Such interactions could be envisioned once the OpenAdap.net system will be fully available to the scientific community and should be encouraged. One major impact on the scientific community could be obtained by attracting well known scientific editors to encourage the scientific authors to provide source data to the community repositories, and to share their methods through an OpenAdap.net broker, hence promoting reproducibility of results and direct peer review of the published article methodologies and results by the readers.

Citizens of less favoured countries will have access to all shared OpenAdap.net resources with a basic Internet connection, thus benefiting from the knowledge transfer and available assets, and contributing back to the community with their own approaches and resources. The outcome of the expected cross-fertilization is unpredictable. Side effects are expected on the quality and harmonization of the resource documentation, as an important effort is dedicated to the elaboration of tools to enhance them.

4 Conclusion

This paper has presented the main features of OpenAdap.net, which is an intelligent network infrastructure supporting the use of shared resources, such as data, knowledge, tools, expertise, etc. aimed at providing the most advanced tools for data analysis and manipulation to a broad audience over Internet. Individual users can be classified as either contributors of shared resources, or end-users of such. OpenAdap.net will be open to exploitation by networked organisations and alliances taking into account the vital issue of Internet security and privacy.

The ability to tackle a scientific problem from a new perspective relies on both the past experience and new skills adopted by an individual. The OpenAdap.net project is based on the collaboration between information scientists, electronic engineers, computer scientists and neuroscientists having diverse scientific interests and very specialized backgrounds. We feel that such a transdisciplinary approach is a necessary way for the achievement of real advances in producing impacts in the Information Society Technologies.

The OpenAdap.net infrastructure makes possible the dissemination of resources, and their exposure to application and evaluation across domains in ways that might be unanticipated. Simulation processing tools issued from physical sciences could permeate to study

problems as different as the dynamics of societal interactions, linguistic analyses, crops forecast, traffic congestions, and life sciences. **OpenAdap.net** is aimed at breaking current boundaries to resource sharing and hence supports transdisciplinarity. End-users are provided with the ability to browse and apply shared resources, and dynamically compose and integrate existing resources to leverage new research insights.

OpenAdap.net brokers are responsible for dynamically decomposing and routing end-user tasks to appropriate resource sharers for execution. The negotiation between brokers (and workers) is inspired by the way how the brain processes information. When completed, the **OpenAdap.net** project network will be able to self-adapt via learning processes that could give rise to modifiable broker-broker connections, specialisation or generalisation of broker behaviour, etc. The nonlinear dynamics that will emerge from our approach makes **OpenAdap.net** closer to the complexity of a living organism.

Acknowledgments

This work is partially funded by the European research project SECOQC.

References

- [EBS06] G. Eisenhauser, F.E. Bustamante, and K. Schwan. Publish-subscribe for high-performance computing. *IEEE Internet Computing*, 10(1):407, 2006.
- [GH06] S. Ghernaouti-Hélie. Guide de cybersécurité pour les pays en développement. ITU publication, 2006.
- [GHSRG05] S. Ghernaouti-Hélie, M.A. Sfaxi, G. Ribordy, and O. Gay. Using Quantum Key Distribution within IPSEC to secure MAN communications. In *MAN conference*, 2005.
- [Lae05] K. Laeuffer. A hike through post-EJB J2EE web application architecture. *Computing in Science and Engineering*, 7(5):808, 2005.
- [VTI01] A.E.P. Villa, I.V. Tetko, and J. Iglesias. Computer Assisted Neurophysiological Analysis of Cell Assemblies. *Neurocomputing*, 38-40:1025–1030, 2001.
- [ZCS05] X. Zhang, S. Chen, and R. Sandhu. Enhancing Data Authenticity and Integrity in P2P systems. *IEEE Internet computing*, 9(6):429, 2005.