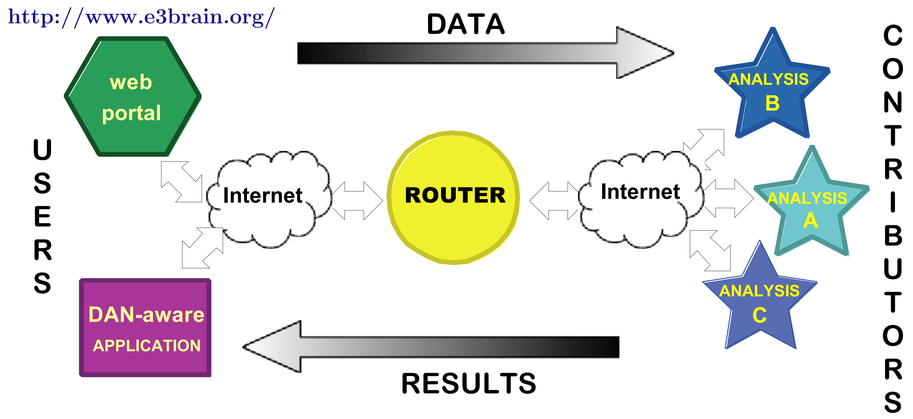
DATA ANALYSIS NETWORK: NUTS AND BOLTS



GENERAL ORGANIZATION OF DAN SYSTEM AND ITS INFORMATION FLOW.

TECHNOLOGICAL GROUND

DAN is an **open-source** project relying on libraries and third party applications that comply with the GNU (L)GPL. GNU/Linux™ is used as a developement and production platform, but no OS or vendor specific solution is used. **Portability** of the components is an essential feature. Hence the development included Microsoft Windows™ as well as Apple MacOS X™.

All descriptive parts of the project are laid down as **XML** grammars to enforce the interchange of information regardless the programming language and architecture of the computers.

Sun Microsystems **Java**™ Programming Language (http://java.sun.com/) is currently used to implement most parts of the system. The availability of many libraries for this language and the easiness of interaction with Internet represent valuable features of the language; that make it a preferred choice, at least for prototyping.

GENERAL STRUCTURE

The complete system is constructed around the idea that there are scientists submitting data to analyses that contributors have hooked to the system. The resulting architecture is then composed of five classes of components.

SUBMIT CLIENTS are the end-user applications, such as Web portals like http://dan.unil.ch/ and DAN-aware applications. All Submit Clients share a set of libraries that provide common information transport mechanisms (APIs) required by the developers to implement new software packages.

Web portals provide an access to DAN through a set of dynamic web pages for the definition and submission of the data analyses as tasks to an Internet wide queue of batch. **DAN-aware applications** embed the power of DAN to broaden their functionalities, by adding on the fly new analyses like plug-ins, offering a complete environment for complex analyses that would be difficult to describe as HTML forms and that are executed elsewhere on remote computer nodes.

EXECUTION DAEMONS manage the data streams, set up the environment for the execution of data analyses and control their completion. The security of data consistency and of system integrity is dealt by the Execution Daemons.

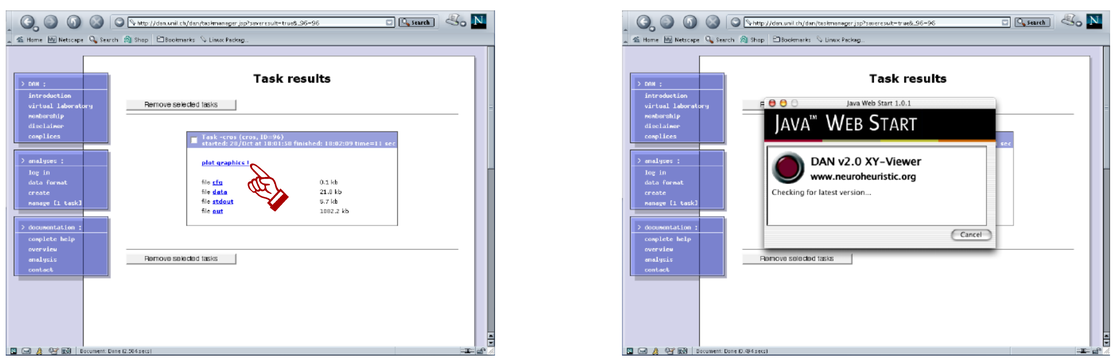
Different kinds of Execution Daemons may be developed, for example for implementation on a dedicated powerful hardware platform, or for high availability. All of them signal their status on the Internet to DAN, provide a description of their capabilities, and may host several analyses. When the execution of a task is completed, the Execution Daemons collect the results and send them back to DAN.

ROUTERS **collect the information** about all the components that connect to it, e.g. their specifications, their response time, the list of available analyses, etc. They offer also essential services like rerun crashed calculations, **temporary storage** of results between connections of the end-user, and **email notification** of task completions.

The **TRANSPORT LAYER** must rely on stable standards in order to provide the necessary ground for developing **interprocess communications** over Internet. Actually we have chosen to exchange information encoded in a **XML** grammar over the **HTTP** protocol. Among several advantages, firewalls are usually configured to let HTTP connections go through. XML and HTTP are defined and recommended by the World Wide Web Consortium (http://www.w3.org/) such that many implementations are available as open-source packages. **SOAP** (Simple Object Access Protocol) is a rising technology that presents all the advantages to suit the needs of DAN, but there is no need to stick to one specific technology.

VIEWERS may consist in standalone applications, applets, Web browser plugins or hand-held applications. Viewers are aimed to display data in a form that is easier for humans to interprete than text outputs, for example in a graphic output suited to be saved in comon exchange formats (**SVG**, Scalable Vector Graphics). A **searchable base** of scientific Viewers, suited to match **data formats** manipulated by scientists in several disciplines, may be built by DAN, just like DAN does with the list of available analyses.

We are currently developing a general purpose x-y vector graphic viewer available online from http://dan.unil.ch/. **Java WebStart**™ technology has been selected to deploy, install and run this Java™ application that displays the results of some of the analyses accessible to-date on our Web portal.

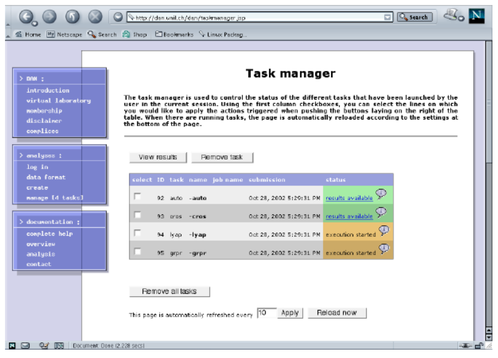


LAUNCHING A VIEWER FROM WEB PORTAL.

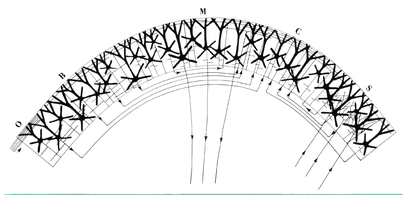
DATA ANALYSIS NETWORK: PRESENT AND FUTURE

The information circulating in the cyberspace generates an historically unprecedented richness for **sharing knowledge and methods of data analysis**. The software published at Internet in one instance becomes available for the whole society **worldwide**. However, transdisciplinary application of new methods is often limited due to the absence of professional software accessible to **users specialized in other fields of competence**. Most of new analytical methods currently developed by different laboratories and companies are based on specific software and computer platforms. Significant efforts have to be made to share the know-how due to difficulties in transmitting the software from one computer platform to another, especially if the software uses some graphical interface. This problem generates a lag until the new methods will become available for application in other fields of science. It is thus necessary to **avoid the "re-invention"** of the methods and **save time and expenses** by the whole society and even, incorrect applications. In most cases an identical conceptual frame for data representation, like multivariate time series, is applicable across several order of magnitude of scaling levels, from molecules to organisms, from human enterprises to celestial mechanics, from nanoseconds to years and centuries.

A challenge of Information Science Technologies is to provide a **transdisciplinary approach** to the development of an Internet based system for scientific data access, analysis, visualization and integration. The objective of project "DAN: Data Analysis Network" is to develop a modular computing approach, providing a flexible tool that can manage a wide set of analytical methods contributed by investigators connected to the Internet. Such flexible tool is designed in order to provide both a **software research environment** and a **tutorial basis** for users across many disciplines working with time series analyses. Our purpose is to provide a user-friendly framework of a "**Virtual Laboratory**" where computational analyses and display of results can be calculated over a distributed computer network like Internet by means of Web Browsers.



TASK MANAGER LAYOUT.

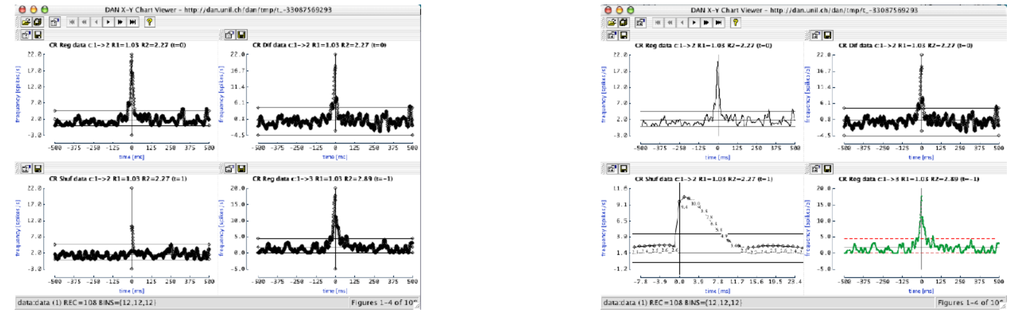


The skeleton cortex (fig. 8.3) from "On textures of Brain", by Valentino Braitenberg, Springer Verlag, 1977.

The implementation of DAN is inspired by the way **how the brain processes information**. Neural networks are characterized by **distributed information processing** where local processing power can not achieve the whole task, and **cooperation** of several units is necessary. In a similar way, we believe that it is important to allow local intelligence, in terms of specific analytical methods, to interact with a distributed amount of data in order to let emerge new and unexpected results and ideas.

There are three types of targeted users. The first type is composed of **users who develop new theories and analytical methods** and would like to make their effort available throughout the scientific community of researchers. It often happens that this information remains in the literature for a long time before practical applications, if any, become available to a wide range of users. The second type is formed by **users who collect lots of data**. They would like to test several hypotheses and related analytical methods in order to understand the message carried out by their data. Both types of users are willing to establish a correlation between theory and experiments, but they often lack the appropriate environment.

The third type is composed of academic personnel, both **students and teachers**. Our goal is to provide a support for exercise some advanced methods for data analyses by means of demo and illustrative explanations and links to other sites, as well as a support for researchers interested to submit their original datafiles for new investigation.

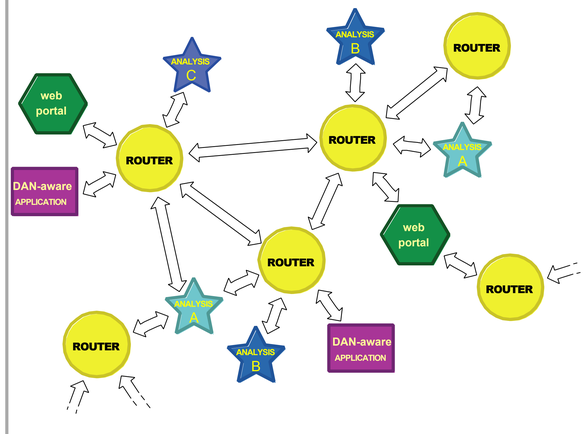


VISUALIZATION OF CORRELOGRAM ANALYSES THAT ILLUSTRATE THE POTENTIAL OF DYNAMIC EDITION OF GRAPHICAL PROPERTIES.

Our project carries a **large exploitation potential** because it is oriented towards vast areas of applications. At the moment DAN deals with **multivariate time serie**s in a suitable format for analyses originally developed in physics and successfully applied in financial forecasting and modeling of stock market, analysis of origin of glacial cycles, music, and EEG data in medicine. The software technology that we are developing may be exploited for other applications that may be networked.

Geographical origin of the visitors' domains	Percentage
Major domains (.com, .net, .gov, .mil, .edu)	31%
European Union domains	38%
Other European Countries domains	20%
Japan (.jp)	4%
Canada (.ca)	3%
Rest of the World domanins	4%

The interest raised by our approach in the **European cyberspace** is illustrated by the vast majority of contacts that we got to the prototype at our website. IST 2002 Event represents the ideal forum to establish joint ventures for further developments of this approach in areas other than the neuroscientific and medical, which is our specialized field of competence.



DAN²: UNSUPERVISED NETWORK DAN NETWORKS.

We foresee particularly attracting and innovative features at the next step of DAN, that we called DAN², to be developed within the European Scientific Space along the new **IST action of FP6**. In particular we must emphasize a totally new role played by Routers where information will be processed and dispatched among all components following a set of **"learning" rules**. These rules may be viewed as the rules that regulate the information flow in a **neural network**, for example taking into account the computing load generated by specific tasks, the number of accesses, etc. The rules themselves will evolve and optimize in a unsupervised fashion, thus allowing the emergence of dynamic links among Routers, and among Routers with Execution Daemons. The **nonlinear dynamics** that will emerge from our approach makes DAN² closer to the complexity of a **living organism**.

europa

Information Society Technologies

DAN: Data Analysis Network

Delegates of DAN

http://www.e3brain.org

http://www.neuroheuristic.org

- Prof. Alessandro E.P. VILLA, Project Leader and Coordinator
  Inserm U318, Université Joseph Fourier Grenoble (France)
- Dr. Jan L. ERIKSSON, Software Engineering & Scientific Computation
  Laboratoire de Neuroheuristique, Lausanne (Suisse)
- Mr. Javier IGLESIAS, Product Development Manager & Quality Assessment
  Institut d'Informatique-IIS, Université de Lausanne (Suisse)

Expression of interest

- Neuroheuristic Research Group (Switzerland)
- PNN Software Development Team, Kyiv (Ukraine)
- Institut d'Informatique-IIS, Université de Lausanne (Suisse)
- Inserm U318, Université Joseph Fourier Grenoble (France)
- Facoltà di Scienze Ambientali, Università di Urbino (Italia)
- Depto. Ingeniería Electrónica, Universidad Politécnica de Cataluña (España)
- Dept Biophysics, Hungarian Academy of Sciences, Budapest (Magyarország)
- Ecole Supérieure de Physique et Chimie Industrielles, Paris (France)
- School of Computing, University of Plymouth (United Kingdom)
- HONDA R&D Europe GMbH, Offenbach/Main, (Deutschland)
- European Brain Research Institute EBRI, Roma (Italia)

Partnerships for the future

- **YOU are Welcome !**

2002.istevent
4-6 November 2002 Copenhagen