VIRTUAL LABORATORY STRATEGIES FOR DATA SHARING, COMMUNICATIONS AND DEVELOPMENT

Enrique Canessa[1], Carlo Fonda[2] and Sandro M. Radicella[2] [*]

[1]*International Centre for Science and High Technology (ICS-UNIDO), Trieste, Italy*
[*2]*The Abdus Salam International Centre for Theoretical Physics -ICTP, Trieste, Italy*
E-mail: radionet@ictp.trieste.it

ABSTRACT

We present an overview of the VL approach to promote research and education in developing countries and to help reduce the technology gap of the digital divide. We discuss software tools for instrument control, data sharing and e-collaboration and communications with special attention to low-bandwidth networks. We analyse the tentative costs involved in VL and the skills needed for VL administration. We conclude by identifying some VL strategies for development.

Keywords: Virtual Labs, Collaboratories, Information Technology, Digital Divide, Development.

1 INTRODUCTION

Information and Communication Technologies (ICT) are being heralded as the foundation for a new world order with the Internet as the means to participate in this global village. In particular, information retrieval systems on the Internet, for the transfer of knowledge and technology and the sharing of ideas and content in synchronous or asynchronous ways, have to be considered essential for sustainable development. Dissemination and management of knowledge as well as activities and transactions conducted by '*moving bits rather than molecules*' occupy less '*road*' space, and consume fewer resources.

In this context, we shall focus on Virtual Laboratories (VL), which are projects that involve collaborative research carried out over a distance, the performance of remote (or distributed) computing, and the sharing of data between groups of scientists, each of whom remains in their home institutions. Our motivation is to identify some VL strategies for sustaining development and scientific progress. We discuss some software tools for instrument control, data sharing and e-collaboration with special attention on low-bandwidth networks.

On a technical level, Open Source technologies provide a cost-effective alternative for promoting distant electronic collaboration. For example virtual laboratories, based on the Linux O.S., and the extensive use of digital communication tools can help reduce scientific isolation, while filling the need to transfer knowledge to-, and from-, countries in less developed geographical areas in an unprecedented way. Open source software (OSS) is usually freely distributed under the GNU General Public License (GPL). While free, this does not mean free in terms of support, training or maintainance. OSS is also particularly interesting in relation to VL environments since it is itself created in a VL. Distributed authors coordinate their work via the net with tools like CVS and Bugzilla.

The limited bandwidth of the few available telecommunication lines in countries that are trying to join the Internet cause line congestion and make access exceedingly slow, often beyond the limit of usability (see *e.g.*, Postogna, Fonda, Ajayi, Canessa & Radicella 1998; Canessa & Onime, 2000). However, to meet such problems 'pre-prepared kits' can be offered with programs for enabling e-collaboration and documentation in conjunction with some reviews tailored by experts. This can help to increase interest in establishing virtual communities and enhance efforts for their integration.

2 THE VL APPROACH

The VL approach embraces a more general definition than Collaboratories (or, `*center without walls*') coined by Wulf (1989), namely, to interact with colleagues, access instrumentation, share data and computational resources, and access information in digital libraries. Similarly to collaborative work, VL aims to create something more than the simple sum of individual efforts.

VL is '*an electronic workspace for distance collaboration and experimentation in research or other creative activity, to generate and deliver results using distributed information and communication technologies'* (Vary, 2000).

VLs around the world could share scientific data in return for access to collectively generated data sets, which are likely to become true `*gold-mines*' for data mining over the coming decades. Dissemination and management of such knowledge will be essential for scientific progress.

Another potential of VLs is that they may generate an enhanced ability to attract R&D contracts from the private sector due to the flexible access to professionals, laboratory facilities, expertise, geographic presence, *etc.* The private sector can provide prototype computer applications and also perform research on the generated VL data sets for specific purposes such as the development of new drugs.

In a practical sense, virtual working teams or communities, regulated by commonly agreed agendas are responsible for shaping the steps needed to enforce their virtual relations, and thus open new doors for interaction.

Scientific thoughts live in the minds of human beings, not inside equipment and machines. Science feeds on science, scientists wish to work where other scientists are present, and where they have access to facilities (Salam, 1989). The Virtual Lab approach could become a worthwhile service that helps to reduce scientific isolation while filling the need to transfer knowledge and technology to the South.

Good planning, implementation and management are each necessary steps for creating virtual laboratories that will serve as *'anti brain-drain*' devices. Motivated by new VL technological and human resources, a real time e-collaboration between developing countries themselves could also begin to flourish to safeguard sustainable development (Canessa, Postogna & Radicella, 1999).

3 SOFTWARE TOOLS FOR VL

Some software tools for instrument control, data sharing and e-collaboration which have been tested on low-bandwidth networks are listed next.

3.1 Person-*to*-person (P2P) communication in a network of collaborators

The conventional techniques of human interaction are ususally carried out via conversations, telephone calls, books, TV, or for example letters. The computer-supported equivalents are, respectively, videoconferences, Internet telephony, the WWW, and e-mails. Of the many commercial, shareware or free software available for P2P communication, we can mention only a few

- Collaborations via E-mail:

  NewsGroups (*e.g.*, NetNews - *Google Groups*, n.d.)
  Electronic Mailing Lists (*e.g.*, Majordomo, n.d.; GNU MailMan, n.d.).
  Electronic Bulletin Board via Web (*e.g.*, HyperMail, n.d.).

- Collaborative Document Authoring:

IBM Lotus Domino - provide a multi-platform foundation for collaboration and e-business, driving solutions from corporate messaging to Web-based transactions and everything in between (Lotus Notes, n.d.)

PHPcms Content Management System (PhpCMS, n.d.).
IPlanet - designed to support a collaborative enterprise work environment (iplanet, n.d.)
NTE - Network Text Editor designed for Mbone usage (Network Text Editor, n.d.)
WebCT - e-Learning hub on the Web (Webct, n.d.).

- Group and Community Calendars:
  Web-based programs written in PHP that use a MySQL database at the backend. Some of its key features include: Public Calendar, Private/Group Calendar, Full month view, Single day view, and more (Hotscripts Web Development Portal, n.d.).

- On-line Discussions:

  Shared White Boards (Collaborative Drawing and Ideas) and Real-time *Chat* (Text). Many types of software available.

  Instant Messaging - *e.g.,* Jabber (Jabber, n.d.)), ICQ (ICQ, n.d.)), Internet Relay Chat (IRC, n.d.).
  ScientificTalk is a profession-specific prototype tool for scientists, students and teachers to exchange information via a web browser, using a synchronous display of math equations. It focus on users' interests in such things as mathematics and scientific notation. And follows an early goal of the Web to be a readable and writable collaborative medium (S & T Collaborium, n.d.)

- Voice over TCP/IP and Video:

  VIC is a video conferencing tool. It can be used for either point-to-point (unicast) conferencing, involving a direct link between two computers or for multiparty (multicast) conferencing when several participants in different locations are linked via a multicast-capable network (VIC, n.d.).

  RAT (Robust-Audio Tool) - is an audio conferencing tool that can be used for either unicast or multicast conferencing (RAT, n.d.)).

  Microsoft NetMeeting – is a Video and Audio Conferencing tool, which also includes a Whiteboard, real-time text Chat, Internet Directory to locate people to call on the Internet, FTP, Program Sharing, Remote Desktop Sharing, Advanced Calling to send a mail message to a NetMeeting user (Microsoft NetMeeting, n.d.).

  Gnomemeeting - is the first Linux H.323 compatible client and it is designed for the Gnome desktop. It is compatible with Netmeeting and other H.323 products. It is the first free H.323 videoconferencing tool with a powerful GUI to be available under GNU/Linux (GnomeMeeting, n.d.).

  RealAudio/Video - is a product of Real Networks to stream video on demand and live video performances across the Web or a corporate Intranet (RealNetworks, n.d.).

  Apple QuickTime – is a multi-platform, industry-standard, multimedia software (Quicktime, n.d.).

- Electronic Green Spaces:
  Shared 3D virtual environment to teach science and distributed design (Furness, T. & Kawahata, M, n.d.).

3.2 Shared scientific data - syncronization

The exchange of scientific data among different computing environments presents major problems that can hinder the full exploitation of the Internet and new Information technology. What is needed today is the sharing of scientific data on the net via intelligent databases that can be synchronized with each other. In this way, new data inputs from experiments that are added to a database would automatically be updated and/or labelled via the net. This is what directory access services such as LDAP do but what at present MySQL and PostrgreSQL (the two big Open Source database packages) cannot do.

The Lightweight Directory Access Protocol (LDAP) is a software protocol enabling the location of organizations, individuals, and other resources such as files and devices on the Internet or on an Intranet. LDAP is a "lightweight" (smaller amount of code) version of Directory Access Protocol (DAP), which is part of X.500, a standard for directory services in a network. An LDAP hierarchical directory can be distributed among many servers. The OpenLDAP Project (www.openldap.org) is a collaborative effort to develop such a robust, commercial-grade, fully featured, and open source LDAP suite of applications and development tools. The project is managed by a worldwide community of volunteers.

Information about standards used specifically for the exchange of scientific data are being collected within the Diffuse Project (www.diffuse.org). The objective of the Diffuse project is to provide a single, value-added, entry point to up-to-date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information, on for example XML – Extended Markup Language, DOM -Document Object Model (DOM) and CSS – Cascading Style Sheets and Resource Description Framework.

Related work is the Platform for Privacy Preferences Project (P3P) of the Web Consortium (www.w3c.org) that will enable Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate.

To share data from databases that may be synchronized, we recognise that the concept of the Semantic Web introduced by Berners-Lee, Hendler & Lassila (May, 2001), which offers a natural way to describe the vast majority of the data processed by machines, can play an important role. The Semantic Web will enable machines to comprehend semantic documents and data. This will open up the information to meaningful analysis by software agents.

At the file system level, the free advanced networked file system Coda (Coda File System , n.d.), the Andrew File System (AFS) (AFS, n.d.) - which is composed of cells with each cell representing an independently administered portion of file space (www.angelfire.com/hi/plutonic/afs-faq.html), or even the ideas of the serverless Gnutella community (Gnutella, n.d.) may be used to achieve data syncronization.

Whereas use of the ‘rsync’ remote-update protocol to only transfer the differences between two sets of files across the network using an efficient checksum-search algorithm (*e.g.,* via the secure encrypted shell ‘ssh’) is feasible when sharing and synchronizing data, but it may not be overly efficient in some cases, *e.g.,* for accessing real-time upper atmospheric and space sciences data (SPARC, n.d.).

3.3 Shared workspaces

Since the geographical distribution of participants is essential to the scientific and technical goals of a VL project shared virtual workspaces are needed.

With the evolution of cost-effective tools and processes to share both instruments and thought, scientific and technological programmes are increasingly being distributed among remote geographic facilities and organizations.

The Basic System for Collaborative Work (BSCW, n.d.)

> BSCW supports asynchronous (not simultaneous) and synchronous cooperation over the Internet (or Intranets). For asynchronous cooperation, BSCW offers shared workspaces to store, manage, jointly edit and share documents. It is entirely Web based and the advantages are:
> Workgroups can share documents – therefore platform independent. There is no need to install the extra software needed –only a Web browser is needed. BSCW workspaces can be accessed, folders browsed and documents downloaded to a local system just like `*normal*’ web pages. BSCW informs

users of new relevant events in a shared workspace. Documents are uploaded to shared workspace via the Web browser.

Virtual Network Computing (VNC, n.d..)

VNC allows a remote display system to view a computing '*desktop*' environment not only on the machine where it is running, but from anywhere on the Internet and from a wide variety of machine architectures. For example: a Unix desktop can be accessed from a native Macintosh viewer, or an X desktop viewed from Microsoft Internet Explorer on a PC.

Vmware (Vmware, n.d.)

This is a commercial product that provides a virtual machine capable of emulating a full x86 class computer at the hardware level. Able to run as a task under Linux it allows, for example, a virtual windows machine to be run from within the Linux O.S.

SourceForge (Sourceforge, n.d.)

This is another example of an asyncronous content management package, which focuses in on software development.

The Collaborative Virtual Workspace (CVW, n.d.)

CVW is a software environment that provides a `*virtual building*' where teams can communicate, collaborate, and share information, regardless of their geographic location.

VIPOL (VIPOL, n.d.)

Virtual Portfolios for Learning. This project broadly explores the potential of virtual portfolios by functioning as a pedagogical/educational tool for supporting learning, in both on-campus and off-campus studies.

3.4 Instrument control/data sharing

Graphical development environment for data acquisition and control, data analysis, and data presentation: e.g., Interchangeable Virtual Instrument (IVI) controls for digital multimeters (DMMs) and oscilloscopes are needed Also needed is a Data Socket that simplifies live data exchange between different applications on one computer or between computers connected through a network.

For data sharing, there are many WWW and FTP mirroring programs (*e.g*., getwww, rsync, Wget) that can be used to download files or full directories, whereas to upload files via Web the CPAN script NET::FTP via secure Web server HTTPS can be used over a network.

3.5 Peer-*to*-peer computing (metamachine)

This is a style of networking in which a group of computers communicate directly with each other. System prototypes include:

JXTA Project (Project JXTA, n.d.)

Its goal is to explore a vision of distributed network computing using peer-to-peer topology, and to develop basic building blocks and services that would enable innovative applications for peer groups.

Jini Network Technology (Jini Network Technology, n.d.)

It provides simple mechanisms that enable devices to plug together to form a community without any planning, installation, or human intervention. Each device provides services that other devices in the community may use.

Grid-based Computing (GGF, n.d.)

The computational grid is analogous to the electric power grid and it couples geographically distributed resources and offers consistent and inexpensive access to resources irrespective of their physical

location or access point. The grid will connect, for example, a geographically distributed national computing infrastructure, databases, instruments, and people into a seamless web.

4 VL ADMINISTRATION

It is essential to familiarize the scientific communities with the many freeware tools that are currently available under the Open Source Initiative. Enlarging the demand for virtual laboratories means that people will adapt to working at a distance.

In these virtual sessions, human diversity, generational differences, and human relations based on the geographical location (and hence time difference) of the partners need to be considered. The electronic realm should meet the needs of the whole virtual community, avoiding an overload of information and unnecessarily complicated usage. Participation in interactive sessions should make it easier to do science, not become another barrier to surpass.

The administration of VL should be simplified so that they can be run by scientists for the scientists in their own countries.

VL require tools featuring the most advanced techniques of instrument control, data sharing, electronic collaboration to mention but a few. Hence, in order for VL to be widely adopted in less developed regions adequate computer literacy must be supported in the academic community of countries that are new to the Internet.

The VL tentative costs indicators are generally given by

- Face-to-face meetings: to discuss the projects and build trust among researchers.
- On-site and distance training on ICT.
- Necessary hardware and software: for monitoring, e-collaboration, data sharing, instrument control, *etc.*
- Salary to technicians: ensuring outputs are delivered via the VL, following standards and Quality of Service.

4.1 Examples of VL

The Human Genome Project, with funding from the US National Institute of Health since 1990, can be considered a VL example. Laboratories around the world shared their gene-sequencing data via central repositories in return for access to the collectively generated data sets. The private sector provided contract services for the repositories and performs research on the data sets for specific purposes, such as the development of new pharmaceuticals. Press coverage of this VL project is extensive, often focusing on the intellectual property issues and their variances from one country or region to another.

The Toolkit for Collaboratory Development of the Environmental Molecular Sciences Laboratory (MSL) provides a comprehensive suite of capabilities for remote scientific collaboration that can be easily integrated with instruments and analysis tools, creating a customized Collaboratory environment tailored to the needs of specific communities.

4.2 The S&T Collaborium Initiative and VL

The Science & Technology (S&T) Collaborium initiative (S & T Collaborium, n.d.) focuses on the transfer of knowledge and Open Source technology using state-of-the-art Information and Communications Technologies (ICT) to Universities, Research Centers, private Corporations and Governments from less developed countries to help sustain economic competitiveness and scientific and social progress.

The development and implementation of electronic Collaboration Tools for Low & Medium Bandwidth Campus Services in conjunction with Research & Development via Demo Projects, Articles and Reports show the

greatest promise for benefiting communities in remote areas via the Collaborium. For example The Collaborium Web server offers free services like '*www4mail - Web Navigation & Database Search by E-Mail*'. This is free software which has evolved since 1998 to enable researchers in developing countries, where full access to the Internet is often unavailable or extremely expensive, to access on-line databases and any other Web-based information resources they need to carry out advanced research. Typically, a www4mail server can register about 293 Hits per Hour with a maximum of about 700. The number of Hits per Day is 7038 (with peaks of 11148). The maximum number of Files sent per Day is usually about 10000 and the KBytes per Day transferred are typically 380000.

A `*Web Portal for Developme*nt' and a VL Toolkit for Person to Person Communication are being implemented to join independent sources for the work being done in relation to collaborative research and education efforts world-wide. The hope behind this S&T Collaborium effort is that it will grow over the years to become an important partner in the pursuit of shared North-South and South-South opportunities.

5 CONCLUDING REMARKS

While the technological revolution of the World-Wide Web has continued to expand rapidly since 1990, there is still a large community of Internet users who only have access to e-mail, because their Internet providers do not offer full Internet connections (due to the lack of adequate infrastructure and low-bandwidth lines), or because they simply cannot afford to pay for full Internet capabilities.

Many of these users live in the developing world, *e.g*., countries in Africa, and rely on e-mail to access essential information (*e.g.*, medical, business and news) and for interpersonal communication. Without broad and unlimited access, scientists, artists, (young/old) people are limited in the amount of creative and diverse content they can produce and communicate.

We all should support Virtual Lab solutions for narrowing this technology gap: To do this it is necessary to optimize bandwidth usage and minimize costs using open source low-cost technologies (mainframe, application). It is also necessary to increase ICT awareness in remote regions by organizing e-learning, and on-site capacity building.

When considering collaborative projects to be generated within VL it is not the number of partners, nor the method of organization that will matter, rather it is the principle of (multi-) cultural awareness, and mutual respect between participants upon which the project is to be based.

The framework of collaboration uses ICT for communicating with others, assumes active roles in the learning process rather than the passive role of being a recipient of information (as transmitted, *e.g*., by a teacher, textbook or TV broadcast).

To conclude, the use of VL can help to increase the size and scale of efforts from different fields. In particular, scientists could share their data in return for access to collectively generated data sets, and also share both instruments and thought. VL facilitates the increase of multi-regional participation, helps avoid scientific isolation, and promises to help sustain further development in the South.

6 ACKNOWLEDGEMENTS

The stimulating discussions with Douglas Fils (from Iowa State University) on VL are greatly appreciated.

7 REFERENCES

AFS (n.d.) Homepage of the IBM Pittsburg Laboratory. Available from: http://www.transarc.ibm.com/

Berners-Lee, T., Hendler, J. & Lassila, O., (May, 2001) The Semantic Web, *Scienctific American. Retrieved August 22, 2002:* http//www.scientificamerican.com/2001/0501issue/0501berners-lee.html

BSCW (n.d.) Homepage of the Basic Support for Cooperative Work. Available from: http://bscw.gmd.de

Canessa E. & Onime C. (2000) Low-Bandwidth Communication Tools for Science, *Linux Journal 75*(July) Retrieved from the World Wide Web: http:// www.linuxjournal.com/article.php?sid=3825

Canessa E., Postogna F. & Radicella S. (1999) Enhancing Electronic Collaboration in the South, *Nature* 398, 744.

Coda File System (n.d.) Retrieved from Computer Science Dept, Carnegie Mellon University: http://www.coda.cs.cmu.edyu/index.html

CVW (n.d.) Homepage of the Collaborative Virtual Workspace. Available from: http://cvw.sourceforge.net

Furness, T. & Kawahata, M. (n.d.) Homepage of The GreenSpace Project. Retrieved from the Human Interface Technology Laboratory, Univ of Washington Website:http:// www.hitl.washington.edu/projects/greenspace).

GGF (n.d.) Homepage of Global Grid Forum. Available at: http://www.gridforum.org) Grid-based Computing

GnomeMeeting (n.d.) Homepage of GnomeMeeting. Available from: http://www.gnomemeeting.org

GNU MailMan (n.d.) Homepage of GNU MailMan. Available from: http://www.list.org

Gnutella (n.d.) Homepage of Gnutella. Available from: http:// www.gnutella.org

Google Groups (n.d.) Homepage of Google Groups. Available from: http:// groups.google.com

Hotscripts Web Development Portal (n.d.) Calendars page. Retrieved from Hotscripts Web Development Portal Web site:http://www.hotscripts.com/PHP/Scripts_and_Programs/Calendars

HyperMail (n.d.) Homepage of the HyperMail Development Center. Available from: http://www.hypermail.org.

ICQ(n.d.) Homepage of ICQ. Available from http://www.icq.com,

iplanet (n.d.) Homepage of iplanet. Available from: http:www.iplanet.com.

IRC (n.d.) Homepage of IRC. Availablr form; http://www.irchelp.org

Jabber (n.d.) Retrieved from the *Jabber Software Foundation* Website: Http:// www .jabber.org.

Jini Network Technology (n.d.) Home page of Jini Network Technology. Available from: http://www.jini.org

Lotus Notes (n.d.) Homepage of IBM Lotus Notes. Available from: http: www.lotus.com

Majordomo (n.d.) Retrieved from the *Great Circle Associates* Website: http://www.greatcircle.com/majordomo

Microsoft Netmeeting (n.d.) Retrieved from the *Microscoft* website: http://www.microsoft.com/windows/NetMeeting/Features/default.ASP.

Network Text Editor (n.d.) Retrieved from the UCL Network and Multimedia Research Group, UCL Website: http://www-mice.cs.ucl.ac.uk/multimedia/software/nte

PhpCMS (n.d.) Homepage of PhpCMS Available from: http://mcyra.homeip.net/homepage/phpcms/index.htm

Postogna F., Fonda C., Ajayi G.O., Canessa E. & Radicella S. (1998) Wireless Networking in Africa, *Linux Journal*, 56 (Dec), 42.

Project JXTA (n.d.) The Homepage of Project JXTA. Available from (http://www.jxta.org)

Quicktime (n.d.) Homepage of Quicktime. Retrieved from the *Apple* Web site: http://www.apple.com/quicktime/

RAT (n,d) Robust Audio Tool. Retrieved from the UCL Network and Multimedia Research Group, UCL Website:www-mice.cs.ucl.ac.uk/multimedia/software/rat.

RealNetworks (n.d.) Homepage of Real Networks. Available from: http://www.realnetworks.com

S & T Collaborium (n.d.) Homepage of S & T Collaborium. Available from: www.collaborium.org/Scientific Talk

Salam A. (1989) *Science in the Third World.* Alna Press

SourceForge (n.d.) Homepage of SourceForge. Available at: http://sourceforge.net.

SPARC (n.d.) Homepage of the Space, Physics and Aeronomy Research Collaboratory. Available from: sparc-1.si.umich.edu/sparc/central).

Vary J.P. (Ed.) (2000) *Report of the Expert Meeting on Virtual Laboratories.* Retrieved from International Institute of Theoretical and Applied Physics (IITAP), Iowa State University Website: http://www.iitap.iastate.edu/reports/vl/vlfinal.pdf

VIC (n.d.) Homepage of VIC. Available from: http://www-nrg.ee.lbl.gov/vic.

VIPOL (n.d.) The Virtual Portfolio for Learning Project. Retrieved from the Insitute for Communication, Aalborg University Website:(www.hum.auc.dk/~vipol)

Vmware (n.d.) Homepage of VMWare. Available at: (www.vmware.com)

VNC (n.d.) Homepage of Virtual Network Computing. Available at: (www.uk.research.att.com/vnc)

W3C (n.d.) Homepage of the Platform for Privacy Preferences Initiative. Retrieved from the *W3C* Website: http://www.w3c.org/P3P/

WebCT (n.d.) Homepage of WebCT. Available from: http://www.webct.com

Wulf W. (1989), The national collaboratory - A white paper. In J. Lederberg, and Uncapher, K. (Eds) *Towards a National Collaboratory: Report of an Invitational Workshop*. USA: Rockefeller University.