

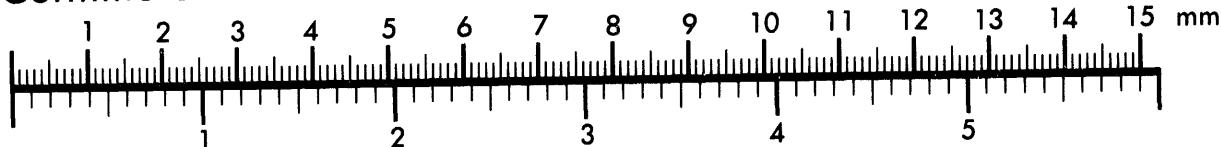


AIIM

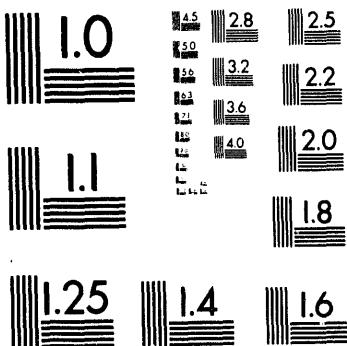
Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

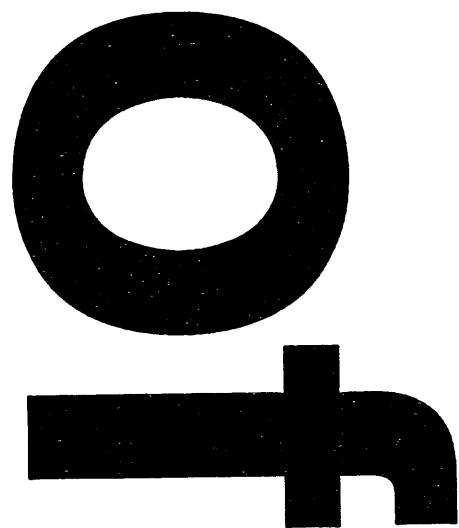
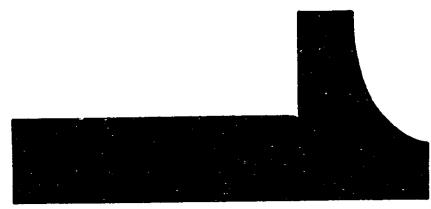
Centimeter



Inches



MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.



Conf-9403113--1

PNL-SA-23921

BUILDING A COLLABORATORY IN ENVIRONMENTAL
AND MOLECULAR SCIENCE

J. D. Myers
D. M. DeVaney
T. H. Dunning
R. T. Kouzes
J. A. Wise

March 1994

Presented at the
Environmental and Molecular Sciences
Collaboratory Conference
March 17-19, 1994
Richland, Washington

Work supported by
the U.S. Department of Energy
under Contract DE-AC06-76RL0 1830

Pacific Northwest Laboratory
Richland, Washington 99352

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
phz

Building a Collaboratory in Environmental and Molecular Science

A Computing and Information Sciences (Molecular Science Research Center) / Computer Science Department (Applied Physics Center) White Paper

Pacific Northwest Laboratory

Richard Kouzes, Principal Investigator

Contributing Authors:
James Myers, Mike Devaney, Thom Dunning, Jim Wise

I. Introduction

A Collaboratory is a meta-laboratory that spans multiple geographical areas with collaborators interacting via electronic means. The National Research Council (NRC) recently published a study of this concept entitled *National Collaboratories, Applying Information Technology for Scientific Research*.¹ Collaboratories are designed to enable close ties between scientists in a given research area, promote collaborations involving scientists in diverse areas, accelerate the development and dissemination of basic knowledge, and minimize the time-lag between discovery and application.

The Environmental and Molecular Sciences Laboratory (EMSL) is a U.S. Department of Energy (DOE) project at the Pacific Northwest Laboratory (PNL) in Richland, Washington. One of the EMSL's key missions is to bring together experts from many scientific disciplines to help solve the nation's environmental problems. The EMSL is the cornerstone of PNL's strategic objective to be the DOE's premier environmental research laboratory.² The resources assembled in the EMSL, from the staff to the computers and

instruments, will be among the best in the world. The Molecular Science Research Center (MSRC) is managing the EMSL project and will share the completed EMSL facilities with local and visiting collaborators and users.

PNL is developing the concept of an Environmental and Molecular Sciences Collaboratory (EMSC) as a natural evolution of the EMSL project. The environmental and molecular sciences are a compelling choice for collaborative development. The task of environmental remediation facing our nation is enormous, and solutions will require the integration of work in many fields. The nation cannot afford to unnecessarily duplicate work or support unguided efforts. A collaboratory is designed to address exactly these issues and promises significant advantages over the current, often uncoordinated way of conducting research.

The goal of the EMSC is to increase the efficiency of research and reduce the time required to implement new environmental remediation and preservation technologies. This new approach will decrease the costs of current projects and allow more complex tasks to be undertaken. The EMSC will leverage the resources (intellectual and physical) of the EMSL by making them more accessible to remote collaborators as well as by making the resources of remote sites available to local researchers. It will provide a common set of computer hardware and software tools to support remote collaboration, a key step in establishing a collaborative culture for scientists in the theoretical, computational, and experimental molecular sciences across the nation. In short, the EMSC will establish and support an electronic community of scientists researching and developing innovative environmental preservation and restoration technologies.

A collaboratory workstyle is well aligned with the EMSL mission and is a natural extension of PNL's laboratory-without-walls concept.² The EMSL can serve two roles in the creation of an EMSC: 1) providing technologies and expertise for development of the EMSC, and 2) acting as a test-bed for the final collaboratory. The present close interaction of EMSL researchers in computer science and the molecular sciences, and the open nature of the EMSL facility, are unique advantages for an EMSC development project at PNL.

The creation of an EMSC will require research in computer and information science and technology beyond the scope of the EMSL project, as well as the adaptation of developments from other laboratories to the specific requirements of environmental and molecular science researchers. A successful collaboratory will require developments in computer communications and integration technologies for sharing data, programs, and ideas, and the creation of a collaboratory culture that acknowledges the

benefits of collaboration and embraces the use of computer-based collaboration as a research tool. Methods of interacting with remote personnel and of accessing remote resources will have to be designed, developed, tested, and measured. Methods of fostering collaborative efforts will need to be investigated and implemented.

In both of the areas described above, there are generic concerns that will apply to any collaboratory as well as issues that are specific to the scientific domains addressed by the EMSC. Some early generic applications are already appearing:

- Collage - a cross platform shared whiteboard tool developed by the National Center for Supercomputer Applications (NCSA).
- MBones - a combined whiteboard/video conferencing application developed by Lawrence Berkeley Laboratory (LBL).
- Mosaic - a network based hypertext browsing tool developed by NCSA.
- Cello - another network hypertext browser, developed at the Cornell Law School.

There are also many programs used in the field of telemedicine that provide video conferencing, whiteboard, and image/video sharing capabilities.^{3,4}

To make the best use of resources, collaboratory development efforts should themselves be collaborative. Generic technologies can be developed with partners from other collaboratory and collaborative work development efforts. Similarly, the customization of generic tools and the creation of applications unique to the EMSC can be shared among the collaboratory participants.

Definition of the scope of an EMSC, and of the efforts needed to create it from the EMSL project, are being developed at PNL. To further this activity, PNL is hosting an *Environmental and Molecular Sciences Collaboratory Workshop* in Richland, Washington on March 17-19, 1994, to bring together motivated individuals from the molecular, environmental, computer, and social sciences to discuss the development of such a collaboratory. The workshop will address short and long term goals for an EMSC and consider molecular science (experimental, theoretical, and computational), computer technology (hardware and software), social (human interaction), and management (funding, legal) issues. The workshop will also provide an

opportunity for potential collaborators, both codevelopers and end users of an EMSC, to meet and exchange ideas.

The next section of this paper provides an overview of the EMSL mission and capabilities. This is followed by a discussion of areas of research and technology development that have been identified as important for development of an EMSC. This includes both computer technologies for information sharing and social methodologies for fostering electronic collaboration. Given the state of the art, we then discuss specific collaborative applications that could be implemented in a reasonable time frame. Finally, we propose topics for the EMSC Workshop and ask some questions that must be addressed to allow the environmental and molecular sciences community to function as a collaboratory and obtain the maximal benefit from this new research paradigm.

II. An Overview of the Environmental and Molecular Sciences Laboratory Project

The EMSL Mission

The DOE's commitment to environmental cleanup at its sites presents significant scientific and technical challenges. These challenges are exemplified by the environmental problems at the Hanford site in southeastern Washington state. The Hanford Site has approximately 1.4 cubic kilometers of hazardous and radioactive wastes, 150 square miles of contaminated aquifer, 60 millions gallons of radioactive wastes (260 MCi) stored in underground storage tanks (of which more than one-third are believed to be leaking), 270 tons of spent fuel, 9 inactive reactors, and 7 major inactive reprocessing plants. The site is the equivalent of nearly 1400 Superfund sites divided into 78 distinct groups sharing common traits and geographies.^{5,6} Pacific Northwest Laboratory, the DOE's multiprogram national laboratory in Richland, Washington, is tasked with developing innovative, cost-effective technologies to address the environmental challenges at Hanford and across the nation, and to facilitate the application and commercialization of these technologies. The EMSL is a key to accelerating and ensuring the effectiveness of this cleanup effort.

The EMSL will be a national focus for the environmental and molecular science research communities. The new laboratory will house roughly 250 resident and visiting scientists in a 200,000-square-foot building located at the Hanford site. Over a five-year period, the DOE is investing \$229.9 million to

design and develop prototypical instrumentation, construct the building, and acquire and install the advanced instrumentation that will make the EMSL a unique national resource. When complete, the EMSL's research programs and operations budget will be approximately \$60 million per year. The EMSL's special resources will enable scientists to apply advanced capabilities to research and technology development in areas such as contaminated soils and groundwater; waste analysis, characterization, processing, and storage; and human and ecological health effects. The EMSL will be part of DOE's high-performance computing network linked to the other national laboratories, and to universities and industrial laboratories, allowing data and information generated in the EMSL to be shared electronically with the national and international scientific communities. As a national collaborative research and technology laboratory, the EMSL will attract hundreds of scientists from academia, industry, and other government laboratories across the United States and around the world. These visitors will work with PNL staff and other collaborators to develop new levels of understanding and permanent solutions to the nation's environmental problems.

The EMSL project was conceived in 1986, was funded by the DOE in 1991, and is expected to be fully operational in 1997. Construction of the new facility will begin in the spring of 1994 while research, design, prototyping, and development are being done in other PNL facilities by the MSRC. Many of the EMSL resources will be in limited operation before 1997.

EMSL Capabilities

The EMSL will have six laboratory modules designed to accommodate the experimental research programs as well as three scientific computing laboratories for theoretical and computational research programs. Major EMSL research areas will include:

- chemical structure and dynamics to study the chemistry of toxic wastes and chemical processes in the environment
- structural biology to study biological processes, bioremediation, and health effects
- advanced materials for use in new chemical and process control sensors, new separations techniques, and new waste storage forms

- theory, modeling, and simulation to understand and predict the behavior of complex environmental systems
- computer hardware and software systems to enhance capabilities in the other research areas.

A wide array of equipment will be available to support these research areas. The capabilities of all the equipment will be at or beyond the present state-of-the-art and many are likely to be unique in the world:

- Ultrahigh vacuum systems for surface analyses

Vacuum chambers with standard and custom surface analytical techniques are available for studying the composition, structure, and chemistry of surfaces. Novel techniques include angle resolved x-ray photoelectron spectroscopy.

- Laser systems

A wide variety of laser systems exist to perform a multitude of linear and nonlinear spectroscopies on isolated chemical systems. Laser sources with wavelengths from the far infrared to the vacuum ultraviolet are available. It will be possible to study chemical dynamics on the femtosecond time scale using a tunable ultrafast laser system.

- Molecule, ion, and cluster beam sources

Novel sources of molecules, ions, and clusters (neutral and ionic) are being developed to provide species for study by a wide variety of experimental techniques. These sources will provide well characterized models for studying the complex environmental systems found at the Hanford site.

- Scanning probe microscopes

Scanning tunneling microscopes, atomic force microscopes, and a near-field scanning optical microscope, all of which are capable of imaging materials at atomic resolution, are under development for the EMSL project. The optical microscope can be used with an ultrafast laser system to provide spatially and temporally resolved spectra of surfaces.

- Fourier transform ion cyclotron mass spectroscopy (FTICR-MS)

FTICR-MS is a precision (better than 1 ppm) mass measurement technique based on the mass dependence of the cyclotron frequency of an ion in a high magnetic field. The technique is applicable to a wide mass range, from atoms to large biomolecules. FTICR-MS can also be used to measure chemical reactions via observation of the appearance of new product masses as a function of time. The EMSL will house three FTICR instruments with magnetic field strengths of 3, 7, and 12 Tesla. These instruments each provide a unique combination of ultrahigh resolution and sensitivity and unique sampling capabilities. A large research program, with both experimental and theoretical expertise, exists to continually improve the capabilities of these instruments and to improve the analysis and visualization of FTICR data.

- Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy can determine the structural (bond lengths and angles, spatial distributions) and dynamical (rotation and diffusion timescales) details of a wide variety of macromolecules, including proteins, enzymes, and DNA. This technique, and the related electron paramagnetic resonance (EPR) technique will be used to address a variety of structural biology and interfacial chemistry questions. These studies necessitate the development of several unique instruments, including a 23.5 Tesla (1 GHz) NMR spectrometer and a low frequency (2 GHz) and a high frequency (220 Ghz) EPR machines. The development of the spectrometers is matched by an effort to develop novel probes, for both liquids and solids, that take advantage of the new spectrometers' sensitivity and resolution.

- High-performance computing

To provide for the computing needs of the staff and collaborative users of the EMSL, a Molecular Science Computing Facility (MSCF) is being established within the EMSL as an element of the research programs. The MSCF will provide a robust integrated computing environment within the facility with links to external facilities within DOE, collaborating universities, and industry. It will consist of a High-Performance Computing Center, a Graphics and Visualization Laboratory, and an Experimental Computing Laboratory.

The High-Performance Computing Center will provide for the large-scale scientific computing needs of the research programs in the EMSL.

The Center will contain a massively parallel computer system, with a peak speed in the range of 100-200 gigaflops, for high performance modeling of molecular systems and processes. The Center will also contain a database computer system for the large-scale scientific data management needs of the computational and experimental research programs in the EMSL.

The Graphics and Visualization Laboratory will provide for the display and analysis of complex data sets from experiments and molecular simulations. This laboratory will contain state-of-the-art graphics and visualization workstations and comprehensive video recording and editing capabilities, as well as output devices for high quality color hardcopy, transparencies, slides, and video.

The Experimental Computing Laboratory will provide the MSRC's Molecular Science Software group with access to innovative computer systems that show promise for significantly extending the range or reducing the cost of molecular simulations. The ready availability of the computer systems in this laboratory will also allow the development of molecular simulation software for use on future systems to be installed in the High-Performance Computing Center.

- Modern computing infrastructure

The EMSL network infrastructure will provide the ability to link the high-performance computing system, experimental equipment, data acquisition systems, and the scientist's desktop workstation into a unified research tool.

To meet these goals, a fiber-optic cable plant has been designed in a modified star topology. This topology gives the network the ability to connect any point to any other in the laboratory while providing fault tolerance, disaster recovery pathways, and space to accommodate future growth and/or new technologies.

The logical network providing service on top of this infrastructure is currently being designed. Mature technologies, such as Ethernet, Fiber Distributed Data Interface (FDDI), and High Performance Parallel Interface (HIPPI) are currently deployed in the interim EMSL facility. New technologies being evaluated and prototyped for the final facility include Asynchronous Transfer Mode (ATM) and Fibre Channel Standard (FCS). Emerging network technologies are being considered

because the networks of today will not meet the unique bandwidth and usage demands of scientific computing in 1997.

The distributed computing system will integrate all of the EMSL computer systems with workstations on scientist's desks, experimental equipment in the laboratories, and with collaborators throughout the world. Staff and collaborators will have location-independent access to data and file management services, computational resources, local and external information services, electronic communications, and an extensive library of software applications and tools. The computing environment will be self-revealing, providing new users with quick access to commonly used services and with intuitive pathways to in-depth information on EMSL computing resources.

- Software research and development

The EMSL computers and experimental instruments are supported by a variety of software development efforts. These efforts span the range from data archiving systems to integrated chemistry environments to custom data acquisition, analysis, and visualization applications.

An Extensible Computational Chemistry Environment (ECCE) will provide integration of, and a graphical user interface for, the design, launching, monitoring, analyzing, visualizing, and recording of computational chemistry 'experiments,' from *ab initio* geometry optimizations to molecular dynamics simulations. New computational chemistry codes are being developed by the MSRC's Molecular Science Software group to take full advantage of the high performance parallel computer to be installed in the EMSL. This software will greatly decrease the time required for modeling molecular systems and will greatly increase the range of systems that can be considered. A Computer Instrumentation and Electronics group will provide sophisticated software integrating the data acquisition and analysis processes for the EMSL instruments. This software will feature an object-oriented design that will increase software reuse and allow rapid modification to enhance the instruments' capabilities.

The EMSL and Collaboration

The nature of the EMSL project is inherently collaborative – scientists with expertise in chemistry, materials science, condensed matter physics,

molecular biology, and environmental science must work together to address complex environmental problems. In the long term, the success of the EMSL, and of PNL's applied environmental research, will be judged in cross-disciplinary terms, by how well and how efficiently hazardous waste sites can be restored. Recognition of the fundamental importance of collaborative interactions to the success of the EMSL is reflected in its multidisciplinary staff and in the scheduling of EMSL resources; a significant portion of instrument time is earmarked for outside users, who will work collaboratively with MSRC scientists or independently. Research in the EMSL will be coupled to the research of other PNL scientists and that of visiting academic and industrial scientists, as well as other DOE applied research efforts in hazardous waste treatment and environmental remediation.

While the scope of the EMSL project does not reach the level of interaction envisioned in the collaborative concept, it does provide a technology and culture base for a collaborative. The EMSL community would certainly benefit from inclusion in a national EMSC and would be a critical early adopter of the collaborative resources.

III. Challenges beyond the EMSL - The Goals of an EMSC

An EMSC follows as a natural extension of the EMSL project. The technology exists to extend the PNL ideal of an environmental science "laboratory without walls" to a national scale using electronic communications. In the same way that basic research in the EMSL is tightly coupled to the environmental remediation technology development at PNL, basic research at other national laboratories and in academia can be coupled to technology applications in industry and elsewhere. Similarly, the integration of experimental and theoretical chemists, materials scientists, molecular biologists, environmental scientists, computer scientists, and human interactions experts on the EMSL project, could be extended to allow 'isolated' academic researchers to make use of sophisticated hardware and software to enhance the quality and quantity of their research.

The EMSL project succeeds in creating a collaborative environment among scientists in many disciplines because its resources are all co-located (either in the new EMSL facility or on the PNL campus) and are managed by one organization (the MSRC). An EMSC development effort would strive to find technology, social, and policy solutions to extend the collaborative model of the EMSL to a national collection of environmental and molecular science

researchers who are geographically and organizationally dispersed. The EMSL project can provide the core of resources needed to seed the development of a national collaboratory.

Generic Collaboratory Goals

Collaborative Technology

Many questions arise when beginning the development of an EMSC. The effort relies heavily on computer and communications technologies. These technologies allow or will soon allow connection by voice, video, interactive 'whiteboard', and virtual reality, thus supporting remote experimentation and analysis from a desktop workstation. A successful collaboratory development effort will need a wide range of expertise to explore these areas. Success will also require tailoring the application of these technologies to the needs and wants of environmental and molecular scientists. Having close collaborative ties in place between research chemists and computer scientists will help ensure that the best technology is applied to the most pressing problems.

The Sociology of Collaboration

Psycho-social issues, often overlooked in discussions of laboratories, are also of vital importance. Scientists' perception of being part of a group, even when geographically separated, will depend not only on communication technologies, but on the ability to develop working relationships and friendships, to have informal chats, and on all of the other ways people develop a common sense of purpose. Such dispersed groups may also have to be funded and managed in a new manner, and credit for the group's work must accrue to all of the group's members. Appendix A is a detailed discussion of the issues that must be faced in the development of a collaboratory. The resolution of these issues will rely on the knowledge about electronic communication that has been obtained over the last decade. One lesson that has been learned is that psycho-social issues can make the difference between a toy and a tool. Experts in human factors and cognitive science must be intimately involved with the computer scientists and domain researchers in a collaboratory development effort. An integrated approach, with active participation by environmental and molecular scientists, computer scientists, experts in human factors and cognitive science, and management, will be required to achieve the full promise of the collaboratory style.

Education and Support

A third requirement of a collaborative effort is a program for educating users and potential users. As with telephones, e-mail, and other ways of sharing information, a collaborative becomes most beneficial when its use is universal among the user community. An education program must inform people about the collaborative resources and opportunities available and how to make use of them. Managers need to understand the benefits of collaborative work to their organizations and how to manage employees who are engaged in collaborative research. The program should target two user communities: end users, who will use the collaborative tools and make use of the data and programs available, and users who will add data, applications, and even parts of the infrastructure as the collaborative grows.

Specific Goals of the EMSC

The NRC Collaboratory study identifies four areas of information transfer that a collaborative should support:

- data sharing
- software sharing
- remote instrument control
- communication with remote colleagues.

An EMSC should address these areas for scientists within and across several fields:

- *Small scale experiments.* Experiments done on equipment that is typically the domain of a single person or small group. These groups often have only minimal internal software development capabilities. Increased automation is a useful goal for these instruments, but remote operation may not be useful. The sharing of software and the collaborative analysis of complex data sets is of great interest in this field.
- *Large scale experiments.* Experiments on rare and expensive equipment that is shared by many users. This equipment usually pushes the state of the art and, as a result, is often "user-unfriendly" and *unstable*. However, it is often well automated or can be so instrumented. Social mechanisms for requesting time on the equipment already exist and could be extended to include distant collaborators. Remote monitoring, and remote operation, of

experiments could greatly facilitate the use of these instruments by the larger scientific community.

- *Computational molecular science.* Computational research is similar in many respects to the large scale experimental research. The limiting resource is time on a large computer, which is usually shared. The computer is tautologically 'completely automated,' but the computational software is not necessarily flexible or easy to use. The definition of standard data models and standard software interfaces, and the development of a standard computational chemistry 'environment' would greatly enhance productivity.
- *Environmental modeling and simulation.* This field is similar to computational chemistry in terms of requiring large, scarce computer power and sophisticated modeling software. An additional factor for environmental modeling is that it requires input from other fields (e.g., chemical properties and reaction rates are needed to model waste in a storage tank). Thus, there is a need to be able to 'request' information from experimental and/or theoretical chemists that is relevant to a real-world problem.

These areas and descriptions are meant to be representative of environmental and molecular science research, there is no implication that they are all-encompassing or orthogonal. They do represent logical extremes of resource and/or information needs that the collaboratory must address.

All of the researchers in an EMSC have the need to search existing literature. Databases of 'accepted,' peer-reviewed data sets for comparison with new data or as input for calculations would also have broad appeal. Methods for sharing unpublished data, with the appropriate security and privacy, will also be necessary for remote collaboration. The sharing of data will require not only the ability to exchange bytes, but standard formats for high level data items, such as molecular structures and extensive meta-data about the method used to create the data. Such data models will support the understanding and exchange of information across scientific domains. Software developed using such models will allow scientists from different domains to view the same data from different perspectives.

Software sharing is more important in the environmental and molecular sciences than it is in many other fields proposed for collaboratory development. Almost all small scale experimental systems have custom data acquisition software developed by the end users. Since there is little standardization in equipment configurations, development has usually started from commercial libraries available for the hardware components.

Similarly, data analysis software is usually written to deal with data in custom formats. For both acquisition and analysis software, sharing can be greatly enhanced if applications are written in a modular and extensible way, with at least *de facto* interface standards. Standard data models would provide a foundation for sharing software. Browsable function/object libraries would allow sharing of specialized algorithms and common graphics, *etc.*, as well as whole applications. For large scale experiments and the computational domains, several monolithic applications dominate individual subfields. New applications must again be developed *ab initio*. Advanced software environments, such as the ECCE under development for the EMSL will allow new applications to delegate tasks such as input creation, computation management, interactive visualization, and data archiving to the environment. This type of approach, based on an object oriented data model, could be extended to the experimental domains, giving sophisticated visualization and data management facilities to any new data acquisition or analysis application.

Remote experimentation varies from being the norm in fields such as computational chemistry to being unheard of in small experimental laboratories. Computational chemists already have the ability to launch remote calculations, and projects like the ECCE will provide even greater levels of control and interaction with remote processes. For large scale experimental equipment, such as NMR and FTICR spectrometers, the level of automation is sufficient to allow remote operation; software is the limiting factor. To be shareable between collaborators with different hardware and software environments, remote control software must be written to be easily extensible to use whatever resources the local environment provides. For mechanical instruments, another consideration is the information about the instrument that is normally conveyed by vibrations, noises, temperatures, *etc.*, and is not usually accessible via the control software. Remote software must transmit this information or provide alternative pathways to convey the instruments 'health.'

Communication with remote colleagues is a universal need within and across disciplines. Collaboratory communications tools must go beyond voice and electronic mail to allow researchers to jointly visualize data and to share interaction with the data. This implies interactive whiteboards, joint program control, and annotation overlay capabilities. These are the ways data would be manipulated and discussed by researchers in the same room. Collaboratory tools must also provide for spontaneous meetings and must transmit the emotional content of communications. The electronic equivalent of meeting in the hallway, or other very informal communication channel is needed to allow collaborators to exchange the seeds of ideas. A special emphasis should be placed on enhancing communications between

traditional scientific disciplines. Even within a facility, this tends to be a most difficult communication channel to promote.

Part of the informality of a hallway meeting stems from the friendship between the participants. A collaboratory must make it possible to begin, or at least maintain, working friendships electronically. Since facial expressions and body language are major pathways for conveying and qualifying information, this implies a need for high quality video connections between collaborators. The integration of voice and video into the same desktop computer environment used for remote experimentation and analysis would help to make its use natural and automatic. The less aware researchers are of the communications technology in use, the more likely they will be to collaborate freely.

Comparison with Existing Collaborative efforts

Collaboratory developments are underway in several scientific communities.¹ Molecular biologists are pooling their knowledge of gene sequences and gene maps by establishing and maintaining large databases. Space physicists and oceanographers also have large data sets that they share. A major driver toward collaboration in all of these fields is the need to enter and extract items from these databases.

The Worm Community System (WCS) is an advanced example of a data driven collaboration.^{7,8} It allows searching of the literature, including journals, newsletters, and informal notes, and the data of researchers studying the nematode *C. elegans*. Links can be created between literature and data and the distribution of new linked items can be disseminated selectively via a privacy control mechanism. These capabilities elevate the WCS from a simple tool for sharing data to an electronic forum that also allows sharing insights generated by the data.

The environmental and molecular sciences utilize a wide variety of seemingly disparate experimental techniques to understand molecular systems. While common data sets are important in the environmental and molecular sciences, there is also a great need to share data at a higher, interpreted level, after an analysis to extract chemically relevant information that is independent of the instrument and the experimental technique. Comparison of the raw data obtained from the study of one molecular system by two different techniques is not trivial. Both the raw data and the assumptions and algorithms used in their analysis affect the interpretation. Effective collaborative sharing of such data requires storing the original data, the information and algorithms used to interpret it, and the interpretation(s)

in a common format that can be readily compared with the results of other experiments. Researchers could contribute to the knowledge base at many levels: entering new data, creating methods of interpretation, and comparing experimental results.

This level of data sharing will rely on tools much more sophisticated than a shared electronic literature and database. Researchers will need to speak with each other and discuss their data using pictures, whiteboards, and shared programs. Common, extensible data models must be agreed upon and standard tools for interacting with the data models must be created. Interactive development will occur as researchers bring new experimental techniques into the common framework. Telementoring will be required as researchers strive to understand each other's data and methods.

Remote experimentation is another common driver for laboratories. Early work in remote experimentation focused on very large instruments: telescopes, synchrotrons, etc. More recently, smaller instruments such as electron microscopes and scanning tunneling microscopes have been remote enabled.⁹ These latter applications are more typical of the applications in the environmental and molecular sciences. The instruments in the EMSL will be capable of many types of experiments and are expected to undergo continual enhancement. Thus, much more consideration needs to be placed on making remote experimentation software that is flexible, modular, and well documented. Collaboratory tools must also address the issues of training potential users (via virtual instruments), scheduling the instrument, and maintaining communications between researchers for jointly analyzing data and preparing publications. Integration of the instruments with the data sharing mechanisms described above will also be important.

IV. Issues for the EMSC Workshop:

Some of the challenges in the development of an EMSC are clear; others will only surface once the effort has begun. The amount of effort required to make interactions informal, useful, and productive will vary with the individuals and the traditions of their scientific disciplines. To blindly apply collaborative technologies would be unwise and prone to failure. A more productive approach would be to ask scientists what capabilities they think are most useful and combine these ideas with knowledge of existing and possible technologies and a knowledge of human interactions. This process is iterative and requires time and effort on the part of the domain scientists and

the computer and social scientists involved. The MSRC has identified several projects, which are described below, that need to be discussed and refined in the EMSC Workshop. Ideas from those attending the workshop will be integrated into plans for the EMSC.

Other more general concerns will be addressed as well: How can collaboration be made attractive to researchers. Can the concept of cross disciplinary teams employed at PNL be extended across organizations? How should funding for collaborative work be administered? Who should review a cross-disciplinary proposal? Similarly, how will credit for collaborative work accrue to a researcher? This becomes important when one realizes that the main importance of a body of cross disciplinary work may not be in the researcher's home field. The development of an EMSC offers some opportunities to explore these issues by trying some non traditional agreements between EMSC development collaborators.

The product of the workshop will be recommendations in the following areas:

- Specific user requirements for the EMSC
 - Applications { theorists, experimentalists, applied scientists}
 - Security {industry, academia, government}
 - Information/databases
- Specific enabling technology requirements for the EMSC {areas that need research in order to meet other goals}
 - Networks
 - Distributed Computing Environments
 - Communication tools
 - Visualization Tools
 - Virtual Reality
- Specific social and managerial requirements for the EMSC
 - Incentives/support for a collaborative culture
 - Funding criteria
 - Support
- Specific requirements for maintaining an EMSC
 - Public awareness
 - Education
 - Support

- Identification of partners (academic, industrial, government)
 - Technology/infrastructure partners
 - Molecular and environmental science partners
- Individuals and organizations that will support the EMSC
 - Re: funding, management, users
- Appropriate overall scope of an EMSC
 - Physical/logical extent
 - Timeframe
- Funding requirements to achieve such an EMSC

The following specific developments that should be evaluated at the workshop have been identified by EMSL researchers:

In computational chemistry, the ECCE project will provide the basic software architecture for the creation of a local integrated environment. The ECCE could also provide the core software for the creation of a collaborative computational chemistry environment. Capabilities for collaborators to simultaneously view and interact with calculations and data, for making the ECCE environment independent of the hardware/software environment of PNL, and for enabling users to add their own computational codes, visualization routines, *etc.*, are likely requirements for a collaborative version. This environment, with its 'experiment' management and visualization capabilities, could also be extended to the areas of data acquisition and analysis.

Collaborative tools for FTICR and NMR also make sense as early projects due to the large number of high-performance instruments being developed in these areas for the EMSL. Again, the creation of an extensible, collaborative-aware architecture could allow joint development of common tools that work for a variety of equipment.

Smaller scale experiments would probably benefit most from data and software sharing. A digital library of algorithms, functions, objects, and applications would be of great benefit in many areas, including spectroscopy, time-of-flight mass spectrometry, photo-dissociation, crossed molecular beam experiments, temperature programmed desorption, laser desorption, electron stimulated desorption, and other standard molecular beam and surface science techniques. This type of effort will probably require associated

collaboratory communications tools for telementoring on the use of the software, to be successful.

A fourth area for development would be in tools for requesting information and for searching for collaborators that improve upon the idea of e-mail, bulletin boards, and newsgroups. Desktop access to peer-reviewed literature is essential for a collaboratory. Enhancing literature access with hypertext capabilities for attaching comments to papers and exchanging informal notes with the authors, as in the WCS, would speed the dissemination of knowledge. Searchable hypertext databases describing an organization's capabilities and the appropriate contact people in a given field of research would help users find potential collaborators. Hypertext might also be used to create 'idea galleries' where researchers in one field display results that suggest interesting complementary work in another field. The researchers using the gallery could be experimentalists hoping for theoretical comparisons or environmental chemists requesting reaction rate data from basic researchers.

V. Conclusion

It seems clear that Collaboratories have the potential to greatly benefit the DOE and the scientific community in general by expanding the resources available to individual researchers, increasing the efficiency of our research system, and by coupling basic and applied research efforts more tightly to national goals. The current system of communication via completed papers with occasional conferences and short visits has been with us since the 17th century. Despite the amazing advances in our ability to communicate rapidly and in great detail, there has been no qualitative change in the use science has made of it. The collaboratory concept is a qualitatively different way of using communication and information technologies. It has the potential to remove the walls around departments and organizations, and could lead to the creation of a meta-laboratory with capabilities - in both expertise and equipment - that far exceed those available in any one laboratory alone.

Pacific Northwest Laboratory has embraced the idea of an Environmental and Molecular Sciences Collaboratory as a powerful tool to connect researchers across the nation as they create solutions to our environmental problems. An EMSC project will leverage the strengths of the EMSL and other environmental and molecular science research projects around the nation to the benefit of all. The MSRC at PNL is a unique combination of physical, biological, computer, and cognitive scientists well suited for developing and participating in an EMSC, providing a coupled core of developers and end

users for such a collaboratory. This team, assembled to develop and manage the EMSL resources, is already familiar with the benefits of multidisciplinary collaboration and knows well that extension of PNL's "laboratory without walls" atmosphere to the environmental and molecular sciences community will enhance our nation's understanding of the fundamental molecular aspects of environmental problems and our ability to apply that knowledge to establish and maintain a safe and clean environment.

VI. References

1. Cerf, V.G, *et. al.*, National Collaboratories: Applying Information Technologies for Scientific Research. 1993, Washington, D.C.: National Academy Press.
2. Pacific Northwest Laboratory Institutional Plan Fiscal Year 1993-1998.
3. Martinez, R., and Chimiak, W. J. Remote Consultation and Diagnosis via the Global Medical Informatic Consortium Networks. in Medicine Meets Virtual Reality II. 1994. San Diego, CA: Aligned Management Associates.
4. Burrow, M., Toler, J., Peifer, J., Sinclair, M., Gadacz, T. A Telemedicine Testbed for Developing and Evaluating Telerobotic Tools for Rural Health Care. in Medicine Meets Virtual Reality II. 1994. San Diego, CA: Aligned Management Associates.
5. Gephart, R., Keller, J. F., "Hanford Challenges and Science and Technology Needs", Spokane Regional Workshop, 1992, PNL-SA-21392 S, Pacific Northwest Laboratory, Richland, WA.
6. Illman, D.L., Researchers Take Up Environmental Challenge at Hanford. Chemical & Engineering News, 1993. 71(25): p. 9-21.
7. Schatz, B.R., Building an Electronic Community System. Journal of Management Information Systems, 1992. 8(3): p. 87-107.
8. Schatz, B.R., *et. al.* The Worm Community System (WCS) Release 2. in 9th C. elegans Conference. 1993. Community Systems Laboratory, Univ. of Arizona
9. Mercurio, P. J., *et. al.* The distributed laboratory: An interactive visualization environment for electron microscopy and three-dimensional imaging. Comm. Assoc. Comp. Mach. 1992, 35(6), p. 54-63.

VII. Acknowledgements

Pacific Northwest Laboratory is operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under contract DE-AC06-76RL0 1830.

The authors wish to acknowledge helpful input from many of the MSRC and Applied Physics Center staff including Ray Bair, Tom Harper, Randy Heiland, Don Jones, Jan Lewis, Richard May, Doug Ray, and Jim Schroeder.

VIII. Appendix A: Psycho-social Aspects of Collaboratories

Introduction

As the power of information technologies has grown, it has brought humans to the threshold of a strange, new "collaboratory" setting. In this synthetic place, distributed across space and time yet maintained through loops of electronic information flows, individuals would convene, converse and cooperate on some of the most challenging scientific problems of the 21st century. The collaborative concept is nothing less than the village square and campfire juxtaposed to the Information Age.

The concept is undoubtedly technically feasible. The question is whether it is socially sustainable. Is it possible to electronically create a distributed organization with a suitable *genius loci*, or 'sense of place,' that permits and even enhances the successful cooperation of dispersed individuals toward common goals? Anthropologists have long maintained that it was the need of early humans for collaboration in hunting and foraging activities that drove the development of communications. Now that social equation is to be writ in reverse: How can a collaborative communications be established that in turn permits human cooperation to thrive, even though the evolutionary social mechanisms that have depended on propinquity are absent?

Although no one has yet established a collaboratory on the order of that proposed by the EMSC, there is enough diverse evidence and experience with 'human-centered automation' in different contexts to provide reasonable clues and indicants of what a collaboratory's main problems will be and what it must provide to overcome these. This section sketches some of those issues and suggests the image and character, if not the technical means of potential solutions. Its purpose is to help structure inquiry, and to convey the sense that, if addressed at the outset and approached in a 'systems' manner, a collaboratory in every sense of that inspiring term can be achieved.

Lessons from Social Communications with Information Systems

The last decade's experience with communication via electronic systems revealed the ubiquity and usefulness of human social controls that otherwise mostly go unnoticed in face to face contacts. When a technical medium of discourse removes these, their role becomes apparent.

In this respect, the first language is surely the language of gesture; and every social encounter takes place in the context of this 'silent language' of kinesics (body motions) and proxemics (spatial positions). Body orientation and movement, the interpersonal speaking distance, and making and breaking of eye contact all send silent messages that are just as meaningful as the spoken word. These are the means that maintain social controls on spoken exchanges, and yet they are often absent from the electronic medium.

In terms of the written word, e-mail is the great communications leveler. A person who would never think of calling a complete stranger to ask their assistance with a reference or a technical problem will have no compunction in contacting that stranger via e-mail. E-mail also often avoids the obligatory social salutations and closing rituals that mark personal or even telecommunications. The silent, social controls of personal discourse are uniformly absent in e-mail. There is little wonder that in every e-mail system, the phenomenon of 'flaming' (sending intensely angry messages) and 'junkmail' (adding a large number of names to a distribution list) is widely observed. 'Emoticons' or 'smileys', a set of symbol characters to be read sideways as a facial pictogram that conveys feelings in an unfeeling medium, {:-), is a relative newcomer to the scene. Yet e-mail is acknowledged as the most successful form of electronic communications, in part for the same reasons of efficiency and reciprocity that produce these unique correspondences. Reciprocity and efficiency are the historical drivers of cooperation, and should be the prime performance requirements in the basic communications protocols of a collaboratory.

Even when the image and sound of a person is restored through audiovideo communications, technical limitations can make the exchange much less than satisfying. Improper placement of a pickup camera can make it disconcertingly appear that the speaker is always looking away from the listener. Limited bandwidth in picture transmission can produce 'freeze frames' that picture the individual (seemingly forever) in the middle of a sneeze, yawn, or eyeblink. The small size and placement of monitors in an AV conference room can promote the 'talking heads' impression that socially diminishes the messenger and the message. Some of these problems will disappear with advancing AV technology, while others demand awareness and more careful attention to producing a 'sense of presence' in different communicating situations.

Groupware applications have been aimed at a variety of purposes (meeting schedulers, group decision support, joint authorship, distributed management, etc.) and have had a decidedly checkered history. Most of the failures have come from a critical lack of understanding of the intended user's workplace, and insensitivity of software developers to social and

political issues therein. Groupware has also suffered from the misconception that its implementation is akin to single user applications, when in fact it carries 'free-rider' and 'critical mass' problems that demand a different type of introduction in an organization. The sorts of groupware applications envisaged for a collaboratory will have to be selected and implemented with a clear understanding of the social and political concerns that circumscribe joint scientific work. Among these are positions of authorship, acknowledgement of contributions, esteem of peers, and recognition by professional role models.

All of the above examples show that creating a Computer Supported Cooperative Work (CSCW) environment like a collaboratory is not simply a technical undertaking, nor is it a question of applying single user applications development strategies across N members. They also illustrate that, while some psycho-social problems of a collaboratory may be new, awareness of their importance is well established, and much has been learned from the pioneering information experiments of the last decade. The need now is to extract and organize these lessons in the design of a working collaboratory.

Other Psycho-social Issues of a Collaboratory

The Issue of Autonomy

The autonomy of an organization describes how it is self governed or regulated. Formally, autonomy is conveyed through an organization chart and procedures handbook, but informally, it is practiced through the myriad of informal communications, acquaintances, and happenstance associations that occur in any organization.

The autonomy difference between a collaboratory and most organizations lies in its *locale*. Most organizations locate autonomy in a specific place (a corporate head quarters) and time (a board or stockholders meeting). But a collaboratory may be so dispersed as to evidence neither of these in quite the same ways. In this case, the autonomy of a collaboratory must be thought of more like a distributed control system whose functions are maintained through coordinated processes. One early exercise in collaboratory formation will be to decide what 'products' the collaboratory wishes to emphasize, and then to address collaboratory procedures as parts of a coordinated production system. It is almost a given that 'loose coupling' across collaboratory activities will result in the most productive arrangement, and that this might be one of the strongest arguments for a collaboratory as a 'research center without walls,' as opposed to its traditional counterpart. When the equivalent of a Collaboratory Administrative Handbook is created, it will likely owe more of

its form and contents to concurrent engineering and adaptive manufacturing than to the historical versions used to manage a research center.

The Issue of Trust

All human-to-human and human-to-machine cooperative activities proceed on the basis of trust. Among people, trust becomes reciprocal and is gained through shared experience, particularly under adversity. Trust may also be instilled and reinforced through casual observations over time, or through awareness of mutually shared histories or interests. People also learn to trust the features of a technical system through the latter's reliable performance, robustness, familiarity, understandability, and usefulness.

Yet how is trust to be established among collaborators who may never meet face to face, or work together in the same place, or even see and feel the instruments of their observations? Clearly, a collaboratory will have to engage some special means to establish the sort of trust that co-workers around each other develop over time through more informal means. For people, this might take the form of archived pictures and personal histories that collaboratory members could browse at their leisure to learn more about their colleagues. Trust might also be gained through extended, electronic Chataquas (discussions on a theme) that allow all to participate, much like the discussions around electronic bulletin boards. Occasional videoconferences, or the use of videocameras to transmit and display the user's image during working sessions, can also help instill a sense of the person and increase subsequent trust in working relationships.

Practice with 'virtual instrumentation' like that centered in the collaboratory, and the ability to monitor and review prior instrument runs will help build trust with the hardware and software components. Remote interfaces with instruments and analytical programs could be designed from a "Kan-Sei" engineering approach, which emphasizes design from the senses that is intuitively appealing and obvious in its usage.

The Issue of Sense of Place

All locations that create a strong sense of loyalty, affection, and identification in their occupants and visitors exude what has been called a 'sense of place.' Having a sense of place means several things. It means being somewhat unique from other places, being highly imaginable in that a place can be visually distinguished, recalled, and remembered, and it means entering into a mutually causal relationship with the user so that it both supports what the user wants to do, while prompting the user to actually behave in certain appropriate ways. Sense of place can occur naturally, as in the beauty of

Yosemite valley, or it can be designed in through various means that deliberately appeal to our sensory systems, habits of thought, and cultural attitudes. The absence of sense of place leads to complaints like that of Gertrude Stein's about the City of Oakland: "There is no There, there."

Many churches communicate their sense of sacred place through the deliberate inclusion of what has been called 'sacred proportion' in their facades and interior spaces. Those proportions (surprisingly) had their origins in the musical chants of the early Middle Ages. Across all denominations, most western, Christian-based churches traditionally and unconsciously adhere to the inclusion or at least the suggestion of sacred proportions in their design.

A good restaurant creates a sense of place through coordination of design elements from the small scale to the large, including the kind of tableware and dishes, table forms, scale of areal enclosures and overall decor. This attention to detail creates the kind of atmosphere the restaurateur wishes to establish for the user, which in turn becomes part of a pleasant (and profitable) dining experience.

The Disney theme parks are perhaps the most inclusively and thoroughly designed settings on the planet that create a wonderfully unique and magical sense of place that delights young and old. It is all done through enforced perspectives, carefully orchestrated movement profiles in view corridors, and color contrast hierarchies that catch and focus user attention on certain attractions. There is no detail of design or appearance in a Disney theme park that is left to chance. Here are truly 'magic places' that have been created wholesale out of an exhaustive attention to the characteristics that psychologists have learned will cogenerate in people a childlike sense of discovery and wonder.

Creation of a 'sense of place' is important in a collaboratory because the image that accompanies it has so much influence on the attitudes and likely interactions of its members. Every researcher fortunate enough remembers a favorite conference that was held in a castle center or a campus of higher learning where the setting itself seemed to demand and provoke great thoughts and peak experiences. If a collaboratory can harness some of the design strategies that have been so successful in physical group settings, it can also create a sense of place and purpose among its dispersed members that will engender an enduring sense of affiliation and cooperation toward its goals. Building a highly imaginable electronic 'sense of place' for the EMSC will not only push the frontiers of computer science, it will establish an exemplar for other future collaboratory endeavors as well.

Conclusions

The prior considerations show that building a collaboratory is, like any other construction project, a social as well as technological endeavor. The very conception of a collaboratory demands as much innovation in its human aspects as it does in its engineering and scientific ones, particularly when the social controls and communication habits that have characterized our entire social evolution can no longer operate in their accustomed ways.

While much needs to be learned about the social management and operation of a collaboratory through sheerly attempting it, this is not a blind initiative. There are reasonable analogs to follow and lessons learned from the expansion of information technologies over recent years. Also, the creation of a collaboratory setting is arguably not all that different from other special settings that have proven remarkably successful in other contexts. There is no reason why the EMSC collaboratory cannot be such a similarly pioneering effort, and a significant step forward to a fundamentally new way of doing science in the 21st century.

**DATE
FILMED**

6 / 30 / 94

END

