Exploration of the InfoMall Concept
"Building on the Electronic InfoMall"

Geoffrey C. Fox, Wojtek Furmanski, Kenneth A. Hawick, Donald M. Leskiw *
Northeast Parallel Architectures Center,
111 College Place, Syracuse, NY 13244-4100

August 1994
Revised and Extended May 1995

*The Ultra Corporation, 110 Harwell Avenue, Lexington MA 02173

Contents

1 Introduction

This document describes: the **InfoMall concept**; how it is employed by the Northeast Parallel Architectures center (NPAC) as a technology transfer program, how it could be used by **Rome Laboratory** and by the United States **Air Force Materiel Command** (US AFMC). A description of the "Electronic InfoMall" system built on the World Wide Web as a pilot project for Rome Laboratory is also given as well as some experiences building WWW systems for academia, commerce and industry. It should be emphasized that this document is primarily a description of the **potential uses** of the **InfoMall** concept and the human interactive processes involved in InfoMall and is not primarily about the HPCC technologies that make the InfoMall process work. These technologies are well described elsewhere although we summarize the main their main features in this document.

2 What is the InfoMall Process?

Technology is developing rapidly and it is no longer possible for a single, static, hierarchically structured organization to possess all the right knowledge to make the best use of technology. The full knowledge base to do this can only come from **virtual organizations.** The InfoMall process [2] is a mechanism for enabling technology transfer by building virtual organizations whose members can interact and exchange multimedia information rapidly yet in an easily controlled manner using high performance computing and communications (HPCC) technology which is already widely available.

The InfoMall process works by networking together the following participants:

Scientists and engineers working together on a project but at geographically separate locations can exchange technical information in the form of documents; diagrams; software; photographs; videos; audio sounds and speech. Teams working in overlapping but distinct areas can choose to make available certain information to each other and prevent duplication of effort.

Line managers can review technical work-in-progress rapidly by accessing the technical information placed on line by the technical staff and can adjust cost estimates and schedules accordingly.

Program managers can review material in-progress across several projects and programs and identify synergistic opportunities.

Favored customers or end-users can be given access to early software products, or on-line catalogs of product information. Mechanisms exist for customers to order products directly via the network and also to provide feedback in the form of carefully controlled questionnaires and product evaluation sheets.

Ordinary customers can also make requirements specifications directly available to the product development teams and both parties can exchange information rapidly to come to a correct, achievable, agreed specification.

Production teams can receive technical information accurately and rapidly from the **product development teams** thus avoiding costly delays and errors.

Strategic planners can review material across all programs and sites that are part of the InfoMall and can decide what areas of work to promote in the future on the basis of up-to-date information on what is available and what feedback is being returned.

Marketing Managers can make product information available not only to existing participants of their InfoMall network but through the main InfoMall USA (see section 3) to a potential market covering the nation and indeed the world. Mechanisms exist to advertise products and carry out market research very cost effectively compared with traditional means.

The participants are thus **empowered** by the InfoMall process.

The InfoMall people-networking process can be enabled by building an Electronic Mall based on technology integration servers (TIS). These are networked repositories of integrated information which can be in the form of:

- multimedia "encyclopedia" review articles;
- on-line glossaries and acronym explanation lists;
- technology data-sheets;
- on-line "normal" documents;
- on-line videos;
- on-line software including modeling and simulations-on-demand;
- on-line software product demonstrations.

There are many other forms of information that can also be included. This process of enabling people using technology is shown in figure 1.

3 How is NPAC using the InfoMall process?

The InfoMall process [2] was invented by Geoffrey Fox at the Northeast Parallel Architectures Center (NPAC) of Syracuse University as a mechanism to manage the technology transfer process associated with High Performance Computing and Communications (HPCC) technology.

The metaphor of a **shopping mall** is used to illustrate the various people and organizational networking transaction activities that are necessary for successful technology transfer.

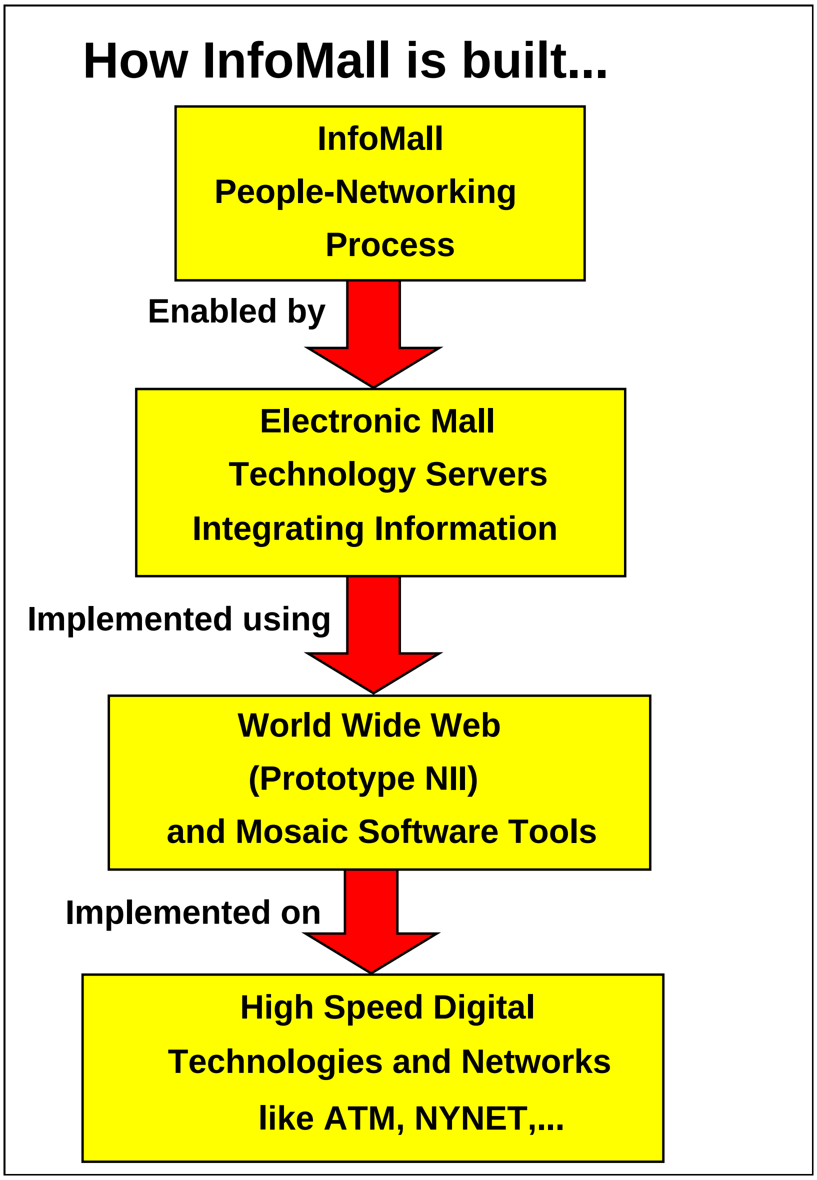

Figure 1: The InfoMall process enables people using layers of software and hardware technology.

NPAC manages the **InfoMall USA** program with over 50 member organisations (including Rome Laboratory, IBM, Kodak, Martin Marietta, Niagara Mohawk, NYNEX and the Center for Research on Parallel Computation). The resulting stream of technological and business related information pouring through NPAC has proved an excellent means of promoting technology transfer in New York State and indeed nationally. NPAC has considerable experience of making the InfoMall process work effectively for a diverse collection of organizational entities.

The InfoMall concept is being adapted by a number of other organizations for technology transfer, including a consortium based around Southampton University in the United Kingdom, which has openly acknowledged their "borrowing" of the InfoMall concept for use in their HPCC technology transfer program funded by the European Union. NPAC joined forces with Southampton in demonstrating the InfoMall concept at the recent European conference on High Performance Computing and networking (HPCN) in Milan, May 1995 [3].

A full description of InfoMall USA as managed by NPAC is available on-line and a paper information pack is also available on request by email to: **infomall@npac.syr.edu**.

NPAC has its own Electronic Mall server (see section C providing integrated information services on NPAC related activities in HPCC. Links are also provided to activity at other organizations nationally and internationally. This system is constructed with various hierarchical levels of access and security, ranging from "advertising" material publicly available to anyone on the Internet; a set of material available under password lock to collaborators, and a private internal set of documents relating to NPAC project management.

4 How could Rome Laboratory and AFMC use the InfoMall process?

The InfoMall process is well-suited for a leading-technology organisation like AFMC which is spread over many geographically distinct sites, has many separate operating teams both internally and in the form of contractors.

The Defense sector relies heavily on the work of contractors - either individuals on-site or contracting organisations off-site. Using the networking InfoMall process, it is possible for engineers and scientists within AFMC to work much more closely with contractors and ensure technological information is not lost but is available for re-use. It is also possible for managers within AFMC to stay in closer touch with work-in-progress off-site if technical information is made available on-line by contractors.

More generally it is likely to be much easier for AFMC to retain control of technologies for which it has paid the development costs if it can use rapid multimedia information exchange with its contractors.

4.1 An InfoMall Scenario

Consider the following scenario:

A project is initiated by Rome Laboratory to satisfy some customer need. A project team at Rome is formed and start to work together using a restricted access version of the WWW information server. These project members can pull information from anywhere on the Internet and start to produce their project design documents and so forth on this restricted access server. They are enabled to work rapidly even although they may be in different parts of the laboratory because they can access each others information as soon as it is available, and each can maintain an accessible up-to-date copy of the documents each is working on.

The project grows, and staff at other USAF laboratories become involved. They can be seamlessly integrated into the restricted access server for the project and with the high bandwidth presently available can work in the project just as effectively as the "local" participants.

The participants within the USAF can access restricted information like phone lists, organisation charts, facility catalogs and so forth and are therefore enabled to work rapidly within their own organisation by having rapid access to up-to-date information.

USAF contractors or favored customers can be involved with the project at an early stage by either including them on the access list for the restricted server or by setting up a different level of server access for them. They can be geographically spread across the USA but can still work closely with the most up-to-date information. The standard software is available on the Internet for them to be brought into the project collaboration very rapidly. They do not need to purchase any special hardware to work in this way.

The project participants may need to access software and information products from suppliers either nationally or internationally. They are enabled to do this because of the de-facto standardization of the WWW/Mosaic interface software (see section 5). For example, an ADA compiler might be required for a new system. Rome Lab project members or their contractors might be able to access demonstration ADA compilation software made available across the network by compiler vendors. The products could be reviewed in advance of a purchase decision.

The project participants need to review patent information for example. Rather than access their own out-of-date paper copies they connect directly to the US patent office to obtain the most up-to-date information. The distributed maintainance of such documents frees up the USAF from having to have dedicated personnel just to maintain and distribute copies of this information internally.

The project proceeds and delivers products which can be made available for testing and evaluation to selected customers directly using the WWW serving technology. Software licensing token technology means that customers can obtain Rome Lab products directly by accessing the software directly from a public Rome Lab server and the software is activated by the purchase of a license code by telephone or fax.

The project has delivered a final product which can be advertised on a public Rome Lab server. Specific feedback can be sought from customers, who will be using the same WWW software technology.

A list of USAF technology transfer personnel contacts posted on the public server directs potential customers where to seek further information.

Other USAF and DoD organisations may also be using this technology, and all can make available the information they want to be available, and can obtain the most up-to-date copies.

AFMC planners can monitor program and project progress by reviewing the information available at the AFMC sites. Opportunities for synergy across programs and projects may be identified. Market opportunities may also be identified from customer feedback and ultimately end-users within the USAF and more broadly within the Defense community may themselves be on-line within the InfoMall process network to feed customer requirements specifications and needs directly to the providers.

4.2 Integrating InfoMall within USAFMC

It appears that the Integrated Weapons Systems Management (IWSM) process already in use within USAF is analogous to the people-networking aspect of InfoMall and therefore IWSM would be more fully enabled for success by the Electronic Mall Information Technology Servers mechanism upon which the InfoMall process relies.

The recent 3 month pilot project between Rome Laboratory and NPAC has uncovered many of the key issues that could make an AFMC InfoMall really successful.

The project illustrated the concept and value of InfoMall using Electronic Mall Technology Integration Servers which integrated information about Rome Laboratory and its contractors' technologies.

A natural conclusion from the project was that it would be very valuable to Rome Lab and to AFMC in general if contractors were encouraged to present results and perhaps even work-in-progress in a form that could go on-line in the Electronic InfoMall.

Many of the information exchange mechanisms have been explored and developed in the course of this project. These mechanisms are now ready to be shrink-wrapped into "information integration products". These include multimedia encyclopedia articles; glossaries of terms and technology product sheets as well as more technologically innovative ones such as static high resolution image sequences and video clips.

The project also demonstrated the mechanism of obtaining expert analysis in the form of invited multimedia encyclopedia review articles. This in-depth study was in the three target fields of software engineering; broadband networking and distributed computing. These three areas were chosen as key HPCC technology areas, although the mechanism is applicable to technology areas outwith HPCC. An across-breadth study looked at 50 Rome Lab technology data-sheets, to many of which NPAC was able to add value by setting them in the context of national and world wide activities in their fields or in related fields.

4.3 Recent Technical Innovations

Recent work, carried out by NPAC outwith the original contract with Rome Laboratory indicates a number of other technologically innovative information delivery and access possibilities. A number of organizations, including USAFMC, already have proprietary database systems which offer the facility for searching, database-joining and other manipulations.

Recent work at NPAC has demonstrated that proprietary database systems using the industry standard Structured Query Language (SQL) can be interfaced with the Mosaic/WWW system demonstrated by NPAC for Rome Laboratory. In particular we are able to interface to Oracle (including a parallel version), Sybase and other proprietary database systems. The implications for networked access to existing data using the existing infrastructure of UNIX/Macintosh and IBM compatible PC's within USAFMC are potentially attractive. The integrity of existing database systems can be maintained by employing the secure (password protection) features of the system.

SQL proprietary database systems today generally store and manipulate textual data, but another possibility is to build database systems consisting of multimedia information items. This is currently an active research area.

NPAC is working with its partners in the virtual research organization "Center for Research on Parallel Computation" to investigate possibilities of "data mining" of publicly posted information systems on the Internet. This has the attractive possibility of being able to ask for the most up to date information on a subject from organizations and experts around the whole world. NPAC is investigating how the potential flood of information that could result from such a query can be organized to provide a manageable response.

5 What HPCC technology makes InfoMall work?

The InfoMall process of networking people is enabled by "Electronic Mall Technology Servers" [1, 3] which are on-line repositories for integrated information. These technology servers are implemented using HPCC software tools such as Mosaic and the World Wide Web, as described below. These software tools are implemented on HPCC hardware in the form of high speed digital networks such as ATM-based networks NYNET[1] and JADE[2]. The **specific** software and hardware technology that is used to do this is not important, but it **must** have the following properties:

1. The necessary hardware must be relatively cheap and widely available and preferably should adhere to widely adopted standards so that an InfoMall participant does not need to buy special purpose hardware but just uses the system already on his desk.

2. The software must use standard widely used and well understood protocols, so a large base of developers will be attracted to it, ensuring many high quality, well tested, cheap tools become and remain available.

3. The software should be in a reasonably advanced state of maturity already for reliability so that early adopters are not turned off by poor performance.

4. The software should be easily extensible and should have a large base of compatible tools so InfoMall participants are not restricted from interacting the way they want to by the limits of the software capabilities.

[1]NYNEX's high speed network, part of which already connects Rome Lab., and NPAC at Syracuse University.

[2]Joint Advanced Development Network

5. Ideally the software should interface easily with existing user environments to minimise learning time.

6. The software should allow sufficient mechanisms to set up various levels of access restriction and afford a degree of security protection so InfoMall participants can feel comfortable about the integrity of their on-line material.

7. The system as a whole should support tele-communication - i.e. person-to-person interaction by video-conferencing on-line.

8. The system should support the distributed maintainance of databases so the owner of information can keep the definitive copy up to date.

9. The system should support automatic information caching, so that although a definitive copy is maintained by its owner, the information system itself will maintain copies of critical databases at more than one cache location to avoid bottlenecks of access for "popular" information and to avoid the problems of critical points of failure in the network itself.

10. The system should allow information servers to be constructed from a wide range of hardware platforms, and should allow interaction with Massively Parallel Processing (MPP) systems to fulfill demand for very high performance database operations, datamining activities and also complex simulations.

11. The underlying high speed digital networking technology for the information system must be able to meet the high bandwidth requirements of tele-conferencing, exchanging multimedia documents, and transmitting very large data sets for datamining and simulation-on-demand applications.

Fortunately there is a candidate system that meets many of these criteria. The most widely used and standardized multimedia information exchange software at present is the Mosaic package from NCSA, which implements protocols defined under the internationally adopted World Wide Web (WWW) system which originated from the High Energy Physics community in CERN, Switzerland. The Mosaic software has been implemented on Macintosh microcomputers, IBM compatible PCs and a wide range of workstations. Mosaic itself is a "viewer" for the hypertext information which is exchanged between participants across the Internet. The WWW protocols determine how a user can embed hypertextual links connecting information components in hypertext documents, with these documents residing on disk space on servers all around the world.

The WWW system allows for exchange of:

- text in many standardized formats such as PostScript; plain ASCII; LaTeX; TeX as well as the widely adopted hypertext markup language (HTML) itself, with which existing information components can be linked.

- photographs stored digitally in a range of standard formats.

- video sequences or movies.
- audio sounds.
- almost any item of software can be set up to be run on the target server system. This allows demonstration software to be made available, under a controlled environment.

In addition a number of tools are available to link hypertext to existing databases, to allow interaction with other users and to integrate the information components with other software packages on the users' host computer systems.

The Mosaic system allows a number of access restriction levels to be set up ranging from completely open readable access to anyone on the Internet, through named Internet address access to individually locked password access.

There are many documents available to explain and discuss the Mosaic, World Wide Web and the other software components mentioned above. At present virtually all this software is available free on the Internet.

It is interesting to note that although the WWW mechanism originated in Europe, it has taken the superb networking infrastructure of the Internet enabled in the USA by ARPA and DARPA investment to make this system viable for exchanging the large amounts of information needed for true multimedia interaction.

Networking systems like the ATM-based networks NYNET and JADE will be able to meet the demands of InfoMall participants interacting together as described above. It will be important that the right sites are linked together using these networks in order to demonstrate the InfoMall process in action.

6 Contents of the Prototype Server for Rome Laboratory

It is not possible to convey the full power on the on-line information delivery system built for this project in a static paper document. Nevertheless, this section gives a list of some of the information products that were constructed or adapted from other sources.

The server was constructed as a guide to Rome Laboratory Technological Assets with hypertext links provided to complementary resources available elsewhere on the Internet. A number of hypertext "clickable" image maps were designed to provide a quick way of navigating around the various information items provided. Figure 2 shows the guide map used as the "front page" provided for the Technology Integration Server.

This "clickable map" technology can be used to illustrate the inter-relationships between the various components of the project work, Rome Laboratory itself, InfoMall, industry, academia, the federal Government and other assets on the World Wide Web. Central to the project were a number of consultant experts in various HPCC technologies that NPAC sought advice from. The distinguished group of consultants (see section 7 provided expert advice on the three technological focus areas of: *Distributed Computing*, *Broadband Networking* and *Software Engineering*. Consultant reports were integrated into the system by NPAC and were hypertext cross-referenced to the Rome Laboratory documents and technology data sheets provided.

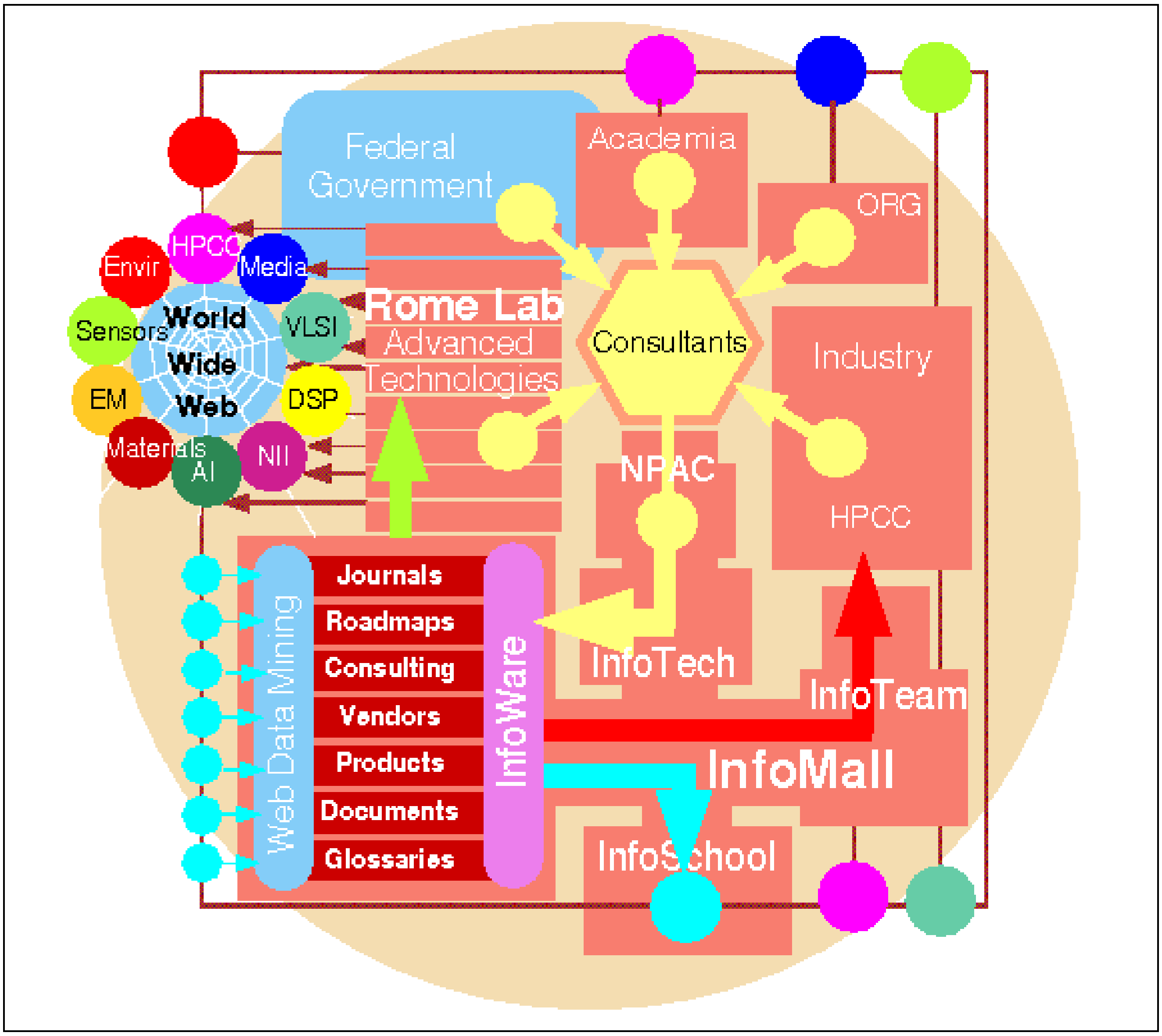

Figure 2: The "Front page" of the Technology Integration System server built for Rome Laboratory to demonstrate the InfoMall concept. This image is "click-sensitive" and each box provides a hypertext link to the appropriate pages on the Technology Integration System server.

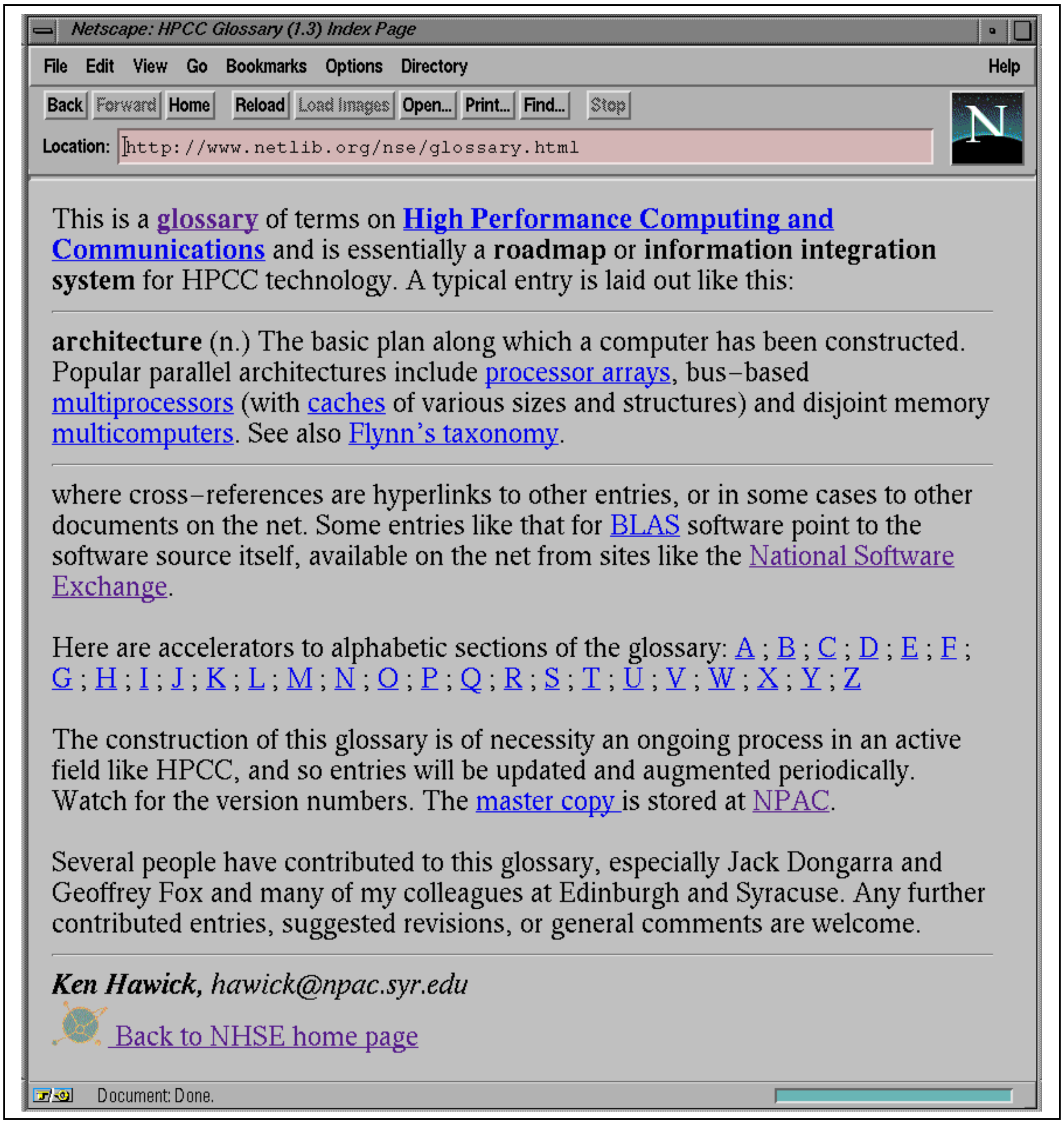

Figure 3: A Hypertext Glossary of HPCC, as part of the National HPCC Software Exchange, illustrating hypertext cross referencing capabilities..

The following information products were set up, not as complete items in themselves but as representative of *how* information could be packaged and delivered.

Consultant Reports and "home pages" See Section 7.

Rome Laboratory Independent Server is a World Wide Web information server constructed by Rome Laboratory personnel outwith the present project, and was linked in as part of the Electronic InfoMall demonstration system.

Collected "scanned" documents - a collection of non-classified material was obtained through Rome Laboratory public relations office for conversion to on-line information packages to demonstrate the concept of distributed information systems.

Rome Lab, Advanced Technology Pages - publicly available data-sheets were also scanned electronically. In addition some of these were "integrated" by adding relevant hypertext links to synergistic activities posted elsewhere nationally.

AFMC Navigation Aid - a number of "clickable" maps and review documents were constructed to demonstrate how the "Electronic InfoMall" could be navigated by a decision maker.

Collected Photographic Imagery was placed on-line to demonstrate the ease with which this could be integrated with hypertextual material and to illustrate the additional power that on-line images, diagrams and maps could have over plain text.

Links to Other On-Line Server Sites were provided as information integration aids. Many organisations are posting material on the Internet advertising their work and as an aid to long distance collaboration. These networked information assets are available to InfoMall participants.

A Conference Proceedings Summary - was placed on-line to illustrate the wider audience that could be reached by a networked information source over a paper one. The example chosen was relevant to the technology areas considered by this project and was for the EFDPMA Conference "Engineering the NII", June 8-9, 1994, Washington, DC. This summary provides a "database" of technology experts from several relevant domains.

Glossary of High Performance Computing and Communications - this glossary of HPCC terminology was written as a contribution to the National HPCC Software Exchange [5] (an on-line collection of HPCC software and documents maintained by the Center for Research on Parallel Computation) and a copy of this material was made available to this present project as an example of a networked information asset. The current release of this glossary is illustrated in figure 3 which shows how hypertext links (blue in colour version or underlined in monochrome copy) can be use to make the glossary cross references clickable. Letters of the alphabet can be made clickable to allow rapid access to sections of the glossary. The Glossary is also available as a downloadable PostScript Document [4].

Survey of High Performance Computing Systems - is an on-line catalog of HPCC hardware systems and is also available as part of the National HPCC Software Exchange.

Acronym expansion lists such as military terminology examples were specially written for this project to demonstrate the general utility of relevant on-line "dictionaries".

InfoWare in Virtual Reality - a number of Technology Demonstrations and Assessment packages were linked to this project to show how simulations could be demonstrated on-demand across the network. This is potentially valuable for advertising purposes as well as for collaborative work.

InfoWare Software Demonstrations PROTO Software (Space Debris Hazard Example) was one particular software package which could in principle be run remotely as a demonstration. A number of other simulations such including applications of computational electromagnetics and financial option pricing are available as simulations on demand within the Syracuse University Internet subdomain. These demonstrations run using integrated NPAC HPCC systems.

InfoWare Video Demonstrations such as a sequence showing First Lady Hillary Rodham Clinton and Senator Daniel Patrick Moynihan visiting Geoffrey Fox at the Northeast Parallel Architectures Center for a demonstration of a prototype Telemedicine system. This illustrates the potential for live interaction between InfoMall participants for decision support for example. A battlefield operations video sequence was also included.

Server Installation Notes - were also included to allow Rome Laboratory personnel to construct their own server systems based on the concepts demonstrated by this project. Additional information on some of the server technical details are given in appendix F which is based on the consultant report by Dr Paul Coddington on Mosaic and the World Wide Web.

Links to other on-line information services and products such as HPCwire (an online news agency for HPCC), NCSA software and CERN software for constructing an electronic server system.

The prototype server was constructed for Rome Laboratory with a strong emphasis on technology of direct interest to Rome Laboratory. To illustrate the potential value of the Electronic InfoMall concept to the USAFMC more widely, a system of "dummy entries" for other USAFMC bases and laboratories was constructed. This idea can be conveyed in a "clickable" map of the USAFMC. The map provides hypertextual links to on-line pages which are dummy entries for each site. If these ideas were adopted by USAFMC then the links could point across the Internet or some more powerful networking system to data physically stored and maintained at each USAFMC site.

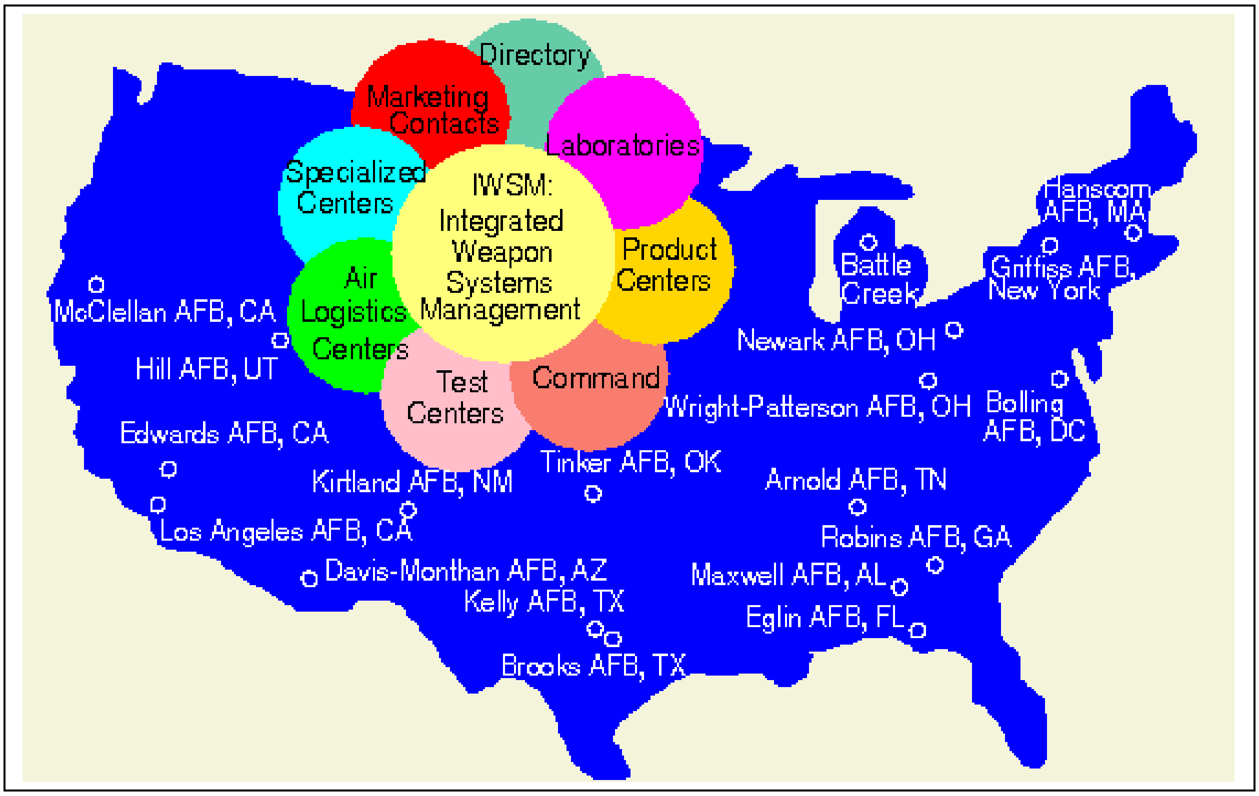

Figure 4: A "clickable" guide map to USAFMC sites. Regions of the map are sensitive and when clicked, provide hypertext links which could potentially be across the Internet or some other network to data stored physically at each site. The IWSM Logo is also "clickable" and provides a navigation aid to IWSM sites.

To summarize, the information applications NPAC demonstrated involve the exchange of hypertextual data consisting of text (in various formats, static images in various formats, video sequences in MPEG format, and "clickable" image maps.

Other technical capabilities known to NPAC at the time of writing include:

Interfacing to (existing) SQL Databases NPAC is engaged in work interfacing Oracle database systems to a World Wide Web server using the WOW system.

On-Line search facilities for on-line data A number of search engines have been integrated with the National HPCC Software Exchange system to facilitate information access. Some of these like harvest and freeWAIS are available at very low cost. Others such as WAIS are available commercially and can be used to index all the contents of a WWW server.

Data entry interfaces We have built a number of data entry systems, amongst which is a system for developing online glossaries of cross referenced information.

Interfacing to other software packages NPAC has also developed a simulation on demand system which allows certain HPCC application programs to be run, on-demand on HPCC [platforms at NPAC. Due to the need to restrict access to the number of machine cycles consumed by web users, this system is only demonstrable within the Syracuse Internet subdomain at present.

7 InfoMall Consultants and On-Line Contributed Reports

To provide the necessary breadth of expertise on the technology areas reviewed by this project (see section 4.2, a number of consultants were asked to provide review reports.

Consultants were carefully chosen from a shortlist of around 50 potential contributors based on the authors' contacts through the existing InfoMall as operated by the Northeast Parallel Architectures Center. The following consultants and their contributions are available as part of the demonstrated server system:

Marina Chen and James Cowie, (Cooperating Systems Corporation) on "CRAFT Compiler and Runtime"

Peter Christy, (Apple Computer, Inc.) on "CORBA"

Paul Coddington, (NPAC) on "Mosaic: A Network Hypermedia Information Browser" (This report is included as an a example in appendix F.

Jack Dongarra, (University of Tennessee and ORNL) on "The National Software Exchange"

Geoffrey Fox, (NPAC) on "Virtual Machine Framework and InfoVision"

Wojtek Furmanski, (NPAC) on "Middleware and Televirtuality"

Ira Goldstein, (Open Software Foundation) on "Portability, Scalability, Interoperability and Usability Research for the National Information Infrastructure (NII)"

Ken Hawick, (NPAC) on "Software Engineering for High Performance Computing"

Robert Jacobson, (Worldesign, Inc.) on "The Crafting of Experience: Designing Virtual Worlds as Environments for Shared Decision-making"

Walter Johnston, (NYNET S&T) on "Future Telecommunication"

Carl Kesselman, (Caltech) on "CC++, HPC++ and CORBA"

Adam Kolawa, (Parasoft Corporation) on "Commercialization of Technology Products"

John Latta, (4th Wave, Inc.) on "Interactive Media"

Donald Leskiw, Ultra Corporation on "HPCC Technology for the Air Force"

Paul Messina, (Caltech) on "CASA Gigabit Testbed"

Roman Markowski, (NPAC) on "ATM/SONET Technologies"

Paul Nielsen, (Rome Laboratory) on "Rome Laboratory Overview"

Janusz Niemiec, (NPAC) on "Operating Systems and High performance Distributed Computing"

Marek Podgorny, (NPAC) on "High Performance Network Integration"

Peter Mills and John Reif, (Duke Univ.), and Jan Prins, (Univ. NCCH) on "A Refinement-Based Framework for the Development of Parallel Applications"

Richard Schantz, (BBN Systems and Technologies) on "Current Status and Planned Extensions for CRONUS, an Enterprise Wide Object Based Distributed Computing Environment"

William Scherlis, (Carnegie-Mellon University) on "Software Engineering, Technology Transfer and Security on the NII"

Shar Zand, (International Software Systems Inc.) on "PROTO: A CASE Tool for Rapid Prototyping of Concurrent Software Systems"

The consultants participated in what was essentially an electronic brainstorming session - a key component of the InfoMall process. Consultants contributed material and were able to review each other's material as well as that assembled by our selves during the course of the project. Mutual feedback was thus enabled. In addition we were able to monitor which participants and project personnel were accessing which entities of the server material which information in itself gave us valuable insight into important and popular parts of the project. One consultant report is included as an example in appendix F. This was generated using the **html2latex** conversion tool, which takes hypertextual material from the World Wide Web as input and produces LaTeXtypeset document information as output. The reverse process can be used to place written documents such as this one on the Web as hypertext.

The consultant forum is thus a demonstration of part of the InfoMall concept and we believe this project conveys the utility of such a system.

NPAC is currently working with other Infomall participants to develop a more fully integrable interaction "whiteboard" system. This system integrates text exchange, email to hypertext conversion, and essentially integrates the multimedia information exchange mechanisms and allows InfoMall participants to use their existing client computers systems on the World Wide Web Internet, whether they be PC, Macintosh or Unix workstations. All the important computer services are provided by the NPAC Web server. This has significant potential for organisations like the National Research Council, as well as a number of businesses, who wish to use the InfoMall human interaction model.

8 Other Information Integration Systems built by NPAC

In addition to the information integration system we prototyped for Rome Laboratory, we describe here some experiences in constructing information integration packages for academic and commercial use. The Northeast Parallel Architectures Center (NPAC) maintains its own academic server as part of a number of servers at Syracuse University. The focus of this system is much more towards the activities, staff and students involved and hypertext links emphasise integration within the academic and research community of HPCC.

We have also recently established a server oriented towards InfoMall business activities and on which InfoMall partners and customers can "rent space" for advertising and promotions purposes.

8.1 Design Styles for Technical and Business Servers

We have chosen rather different design concepts for the front pages of these servers. The Rome Lab prototype server was built based around a series of concept maps as the main access points. These involve a click sensitive graphical image that conveys the main mission, or organizational structure of the organisation involved.

The NPAC academic server is highly technical in its content and is designed in what has become a more traditional style involving the organisations logo and a detailed list of contents in the form of textual bullets. We have generally forum this form of Web page successful for around a dozen bullet points. We draw attention to an special new features to our NPAC page with a "Hot link of the month" and by periodically revising the ordering list of items on the page. This approach is illustrated in figure 5.

For the purposes of our Infomall Business server we adopted another approach. This server is generally accessed by non-academics and it must be very clear and must summarize

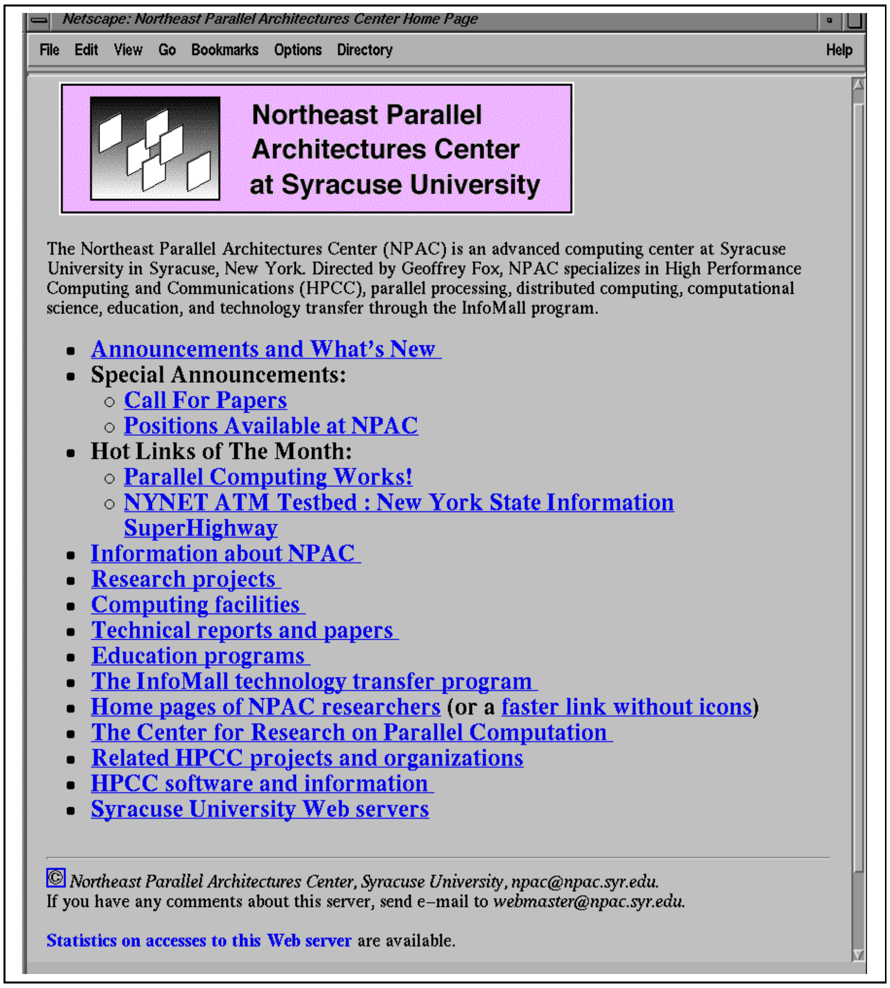

Figure 5: The Front Cover page of the NPAC Web Server - in the logo plus detailed list of contents style.

only a few ideas in a short space. We therefore adopted the approach of a very few iconized bullets for this front page. This approach is illustrated in figure 6.

This design of iconized bullets was originally conceived by Coddington for an educational collection of material known as Kid's Web. It transpires that this approach is also well suited to the busy business executive who may have a longer attention span than school children but has limited time and needs a very compact executive summary of what information is available to allow quick judgement on the value of pursuing a hypertext link.

A third approach appears to be useful as a supplement to any of the three described above, and involves a very simple directory list of contents, giving hypertext links that describe contents in a way that is recognizable to someone who has already read the pages and wishes to find a particular item again. Generally when designing Web pages we try to modularize the information and in fact to tell a story with a train of suggested links given

Figure 6: The Front Cover page of the InfoMall Business Web Server - in the style of iconized bullets.

to the reader. This has proved valuable for first time readers but has proved frustrating for the reader returning to examine a specific item.

8.2 Logging Server Usage Statistics

A recent technical capability we have added to our servers involves the logging of usage statistics. An automated program can be configured to analyse access log data and produce graphical representations of this data available on the server itself. Figure 7 illustrates some of this data for our NPAC server.

A number of interesting trends emerge. For example, in the case of the NPAC server:

- The Kid's Web Educational Package is the most popular item on NPAC's server. (This is partly biased by the fact that accesses of all the graphic icons it contains generate additional "hits").
- The countries that originate the most accesses to this server are in order: USA, Canada and Japan.
- Daily access times (mainly biased by USA originating hits) peak between 11am and 4pm Eastern Standard Time. This spread towards afternoon is probably caused by lunchtime access spread across the US continental time zones. This is illustrated in figure 8.
- Accesses drop off substantially during the weekend (Saturday and Sunday), suggesting that the majority of accesses originate from workers using their university or company client platforms.

We envisage the powerful marketing possibilities from this sort of data mining as being very valuable for business ventures on the World Wide Web, using an InfoMall-like environment.

A number of InfoMall business participants including small start up businesses in the regional InfoMall at Mid-Hudson, are using our InfoMall business server sytem as an advertising medium, and are very interested in collecting the access addresses of browsers as a potential list of customers. It is also possible to mine the data collected in this way for regional correlations, as well as indicators of the sort of organization - academic, commercial or othrwise that these potential customers belong.

9 Concluding Remarks

We have described the InfoMall process for the successful implementation of large scale complex technological projects.

This project has established an effective *mechanism* for rapidly building a a world-wide "Multimedia Technology Encyclopedia" using on-line review articles provided by consultants drawn from the large pool of expert contacts available to InfoMall.

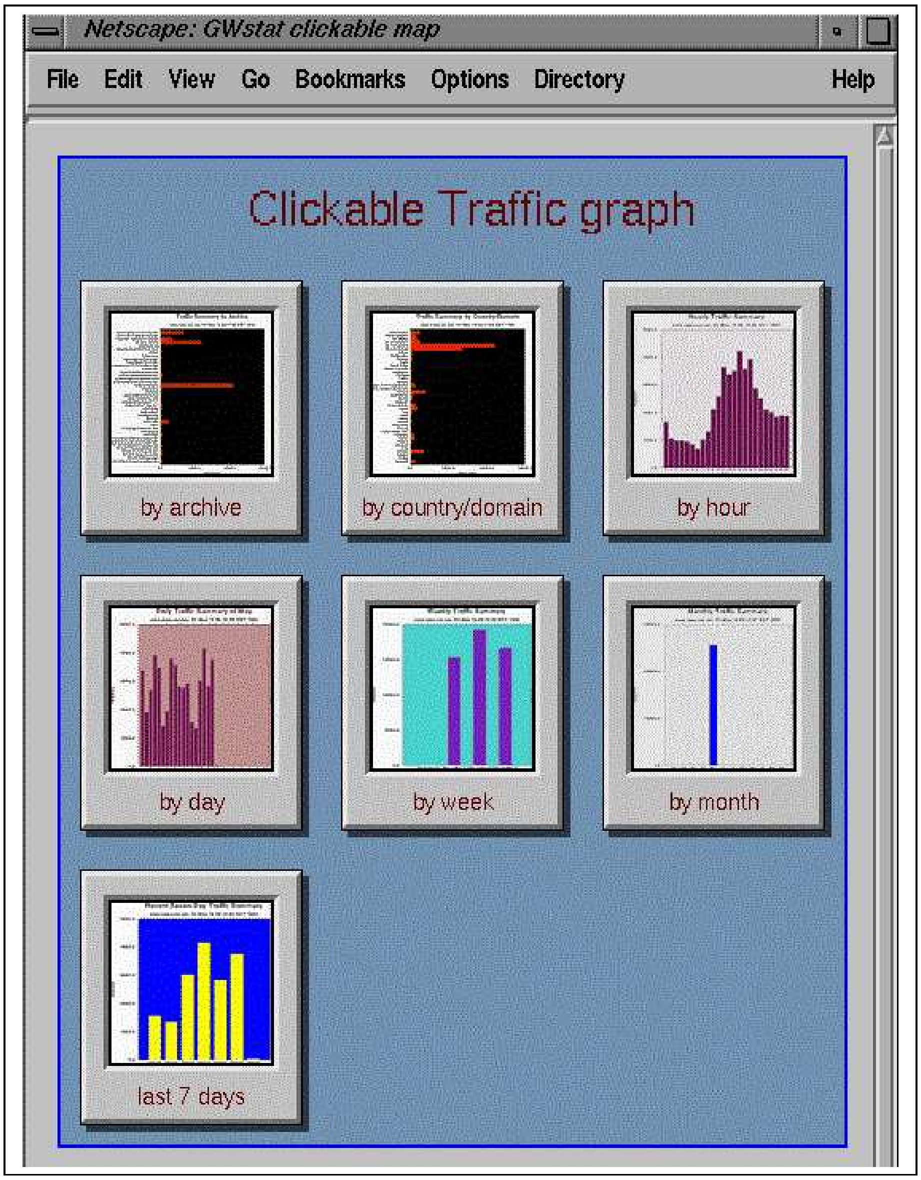

Figure 7: Usage Statistics data for the NPAC Server, showing a number of important trends.

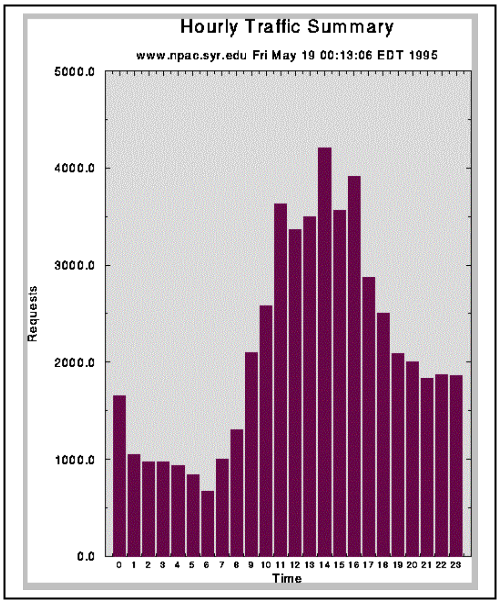

Figure 8: Hourly Usage Statistics data for the NPAC Server, showing peak use during lunchtime, with a spread due to USA timezones.

The technology of the World Wide Web is an excellent vehicle for the InfoMall concept. All the participants involved in a complex project can participate in the InfoMall process using standard software run from their existing client PC, Macintosh or UNIX workstations.

We believe that organisations carrying out complex technological projects **must** use a human interaction model such as the InfoMall process to achieve their missions. The electronic InfoMall system we have described here is particularly well suited to project participants who are distributed geographically or across different sub-organisations.

The InfoMall process appears to be ideally suited to the very complex challenging projects undertaken by the US Air Force involving personnel across many bases and contractor sites.

A An InfoMall Glossary

The following is a glossary of terms used to describe various parts of the InfoMall and which are referred to elsewhere in this document. A much more substantial glossary of terms in High Performance Computing and Communications was developed by NPAC as part of the National Software Exchange project and was also made available to this project in preliminary form. This is available on-line from sites (NPAC and NHSE) listed in section C, and is also available as a NPAC Technical Report [4].

HPCC (n.) High Performance Computing and Communications is a field which covers the state-of-the art computing and networking technology in its broadest sense. A full glossary of HPCC terminology is available.

HTML (n.) Hyper Text Markup Language, a simple mechanism for embedding hypertext commands in a plain ASCII text file.

InfoLink (n.) is a relatively new concept and refers to the networking infrastructure that enables the InfoMall process. The Shopping Mall analogy of InfoLink is the Mall services and utilities that are normally hidden from the shopper, but which are crucial to the smooth running of the Mall as a whole.

InfoMall (n.) is a process that enables technology transfer by building effective virtual corporations. The process is enabled by HPCC technology so that groups of individuals, businesses, or their organisations can work effectively together, exchanging information rapidly and in many forms. The InfoMall process not only allows product development teams to carry out technical work together, but assists planners, managers and also marketers to do their jobs effectively by giving them control of up-to date information. Two key features of the InfoMall process are that a vast information structure can be navigated rapidly in a controlled manner and that the information elements are maintained by their originators in a distributed way. The InfoMall concept was conceived by Geoffrey Fox at the Northeast Parallel Architecture Center (NPAC) as a method for enabling NPAC's technology transfer program. The original InfoMall USA is managed by NPAC and is concerned with HPCC technology. The InfoMall idea is however broadening to encompass many other HPCC-enabled technologies across the USA.

InfoMall USA (n.) is the specific instance of the generic InfoMall process that is run by NPAC for HPCC technology.

InfoMarket (n.) is concerned with linking product developers with customers for partnership marketing, group solutions, systems and software demonstrations, world-class facilities access, market requirements, and product sales. InfoMarket provides the primary link between the product development process (InfoTeam), (InfoTech) and the HPCC consumers. This link initially establishes technical and market criteria for commercialization potential. It also performs the commercialization functions ultimately leading to product sales. NPAC's own InfoMarket is concerned with HPCC products and services and includes in the marketing arms of HPCC vendors and also other market organizations with unique consumer relationships, e.g., computer services (outsourcing), consulting and system integration companies.

InfoSchool (n.) is a program concerned with training, education and consulting. NPAC's own InfoSchool is concerned with HPCC related technologies and offers a carefully designed set of workshops, courses and seminars.

InfoTeams (n.) are development teams drawn from either small businesses or from groups within large corporations. **Product development InfoTeams** take reusable technologies and integrate and develop them into domain specific reusable application components. **Business development InfoTeams** exist within NPAC's own InfoMall and can assist small companies and entrepreneurs wishing to start new businesses. To reduce the risk of failure in new software development ventures, assistance with access to financial resources, business development services, management and technical services can be provided. NPAC's InfoMall helps new businesses apply the latest HPCC computer and communication facilities in their day to day activities.

InfoTech (n.) is a program concerned with gathering, evaluating and integrating the world's best HPCC technologies. InfoTech manages the InfoMall technology warehouse (InfoWare). In this role, the InfoTech program analyzes and maintains current "stock" elements of world class HPCC technology, while providing direct technical expertise for transitioning next generation components into truly demonstrable and re-usable application (RAP) and technology pieces (RTP). Other InfoTech activities include gathering and classifying technologies; developing a methodology for the decomposition of applications into component technologies (a generalized library function); development and maintenance of metrics and benchmarks for systematic evaluation; and promotion and development of open interfaces to allow integration of technologies (i.e. vendor neutral open system "seals of approval"). InfoTech maximizes the re-use and dual-use of technologies and minimize unnecessary duplication in technology development. NPAC's own InfoTech program is managed within the context of the Center for Research on Parallel Computation (CRPC)

InfoWare (n.) is the InfoMall warehouse of Reusable Technology Pieces (RTPs)) and Reusable Application Pieces (RAPs). InfoTech provides these components to builders (InfoTeams) of sophisticated HPCC applications products which can then be sold in the growing HPCC marketplace

InfoWeb (n.) is relatively new concept and describes a subset of the World Wide Web where the links included serve InfoMall purposes - linking together InfoMall participants. Nodes in the InfoWeb are Technology Integration Systems for the various domains; InfoMall tenants; roadmaps and navigation aids. InfoWeb is the "Electronic InfoMall".

Mosaic (n.) is a hypertext viewing tool which understands hypertext markup language (HTML). Mosaic is available from NCSA (see section C).

NII (n.) abbreviation for the National Information Infrastructure, which is being built from existing and planned digital networks.

NPAC (n.) abbreviation for the Northeast Parallel Architectures Center at Syracuse University.

RAP (n.) Reusable Application Piece - a generic term for an item of technology such as a software package that is essentially a shrink-wrapped end-user product.

reusable application piece (n.) see RAP.

reusable technology piece (n.) see RTP.

RTP (n.) Reusable Technology Piece - a generic term for an item of technology such as a software package or tool or a hardware element that is not necessarily a end-user product in its own right but is a reusable building block for making application products.

Technology Integration System (n.) see TIS.

TIS (n.) an anchor point for integrated information on one (perhaps broad) topic. A prototype TIS was constructed for Rome Laboratory technologies to demonstrate how information about Rome and what it does and has achieved could be integrated with complementary world wide activities.

World Wide Web (n.) an international system for the distributed access of multimedia information across multiple sites connected by the Internet.

WWW (n.) see World Wide Web.

B Bibliography

References

[1] Fox, G.C, Furmanski, W., Hawick, K.A, Leskiw, D., "Exploration of the InfoMall Concept", NPAC Technical Report SCCS 634, August 1994.

[2] Fox, G.C., Bogucz, E.A., Jones, D.A., Mills, K., Podgorny, M. and Hawick, K.A., "InfoMall: A Scalable Organisation for the Development of High Performance Computing and Communications Software and Systems" Proc. HPCN 1994, Munich, Germany April 1994. Volume 2, PP137

[3] "The Electronic InfoMall - HPCN enabling industry and commerce" Fox, G.C., Hawick, K.A., Podgorny, M., Mills, K., Proc. HPCN Europe 1995, The International Conference on High Performance Computing and Networking, Milan, 3-5 May 1995.

[4] Hawick, K.A., "High Performance Computing and Communications Glossary", Release 1.3, NPAC Technical report SCCS-629, National HPCC Software Exchange **http://www.netlib.org/nse/glossary.html**.

[5] Browne, S., Dongarra, J., Green, S., Moore, K., Rowan, T., Wade, R., Fox, G.C., Hawick. K.A., "Prototype of the National High Performance Software Exchange", Accepted for publication in IEEE Computational Science & Engineering.

C Internet Access to On-Line Information

The following Uniform Resource Locators (URL) provide Hypertext transfer protocol (HTTP) links to information assets relevant to this project:

http://www.npac.syr.edu/ The Northeast Parallel Architectures Center (NPAC) Main Server.

http://www.npac.syr.edu/techreports/ The NPAC Technical Reports Repository - including access to this document as report SCCS 711.

http://www.infomall.org/ The InfoMall main business server.

http://king.syr.edu:2001/ The Pre-prototype Server built for Rome Laboratory by NPAC. (Note: this is an access-restricted server, email the authors if you are interested in viewing the material on it.)

http://www.rl.af.mil:8001/ The Rome Laboratory Server, built by Rome, following the success of this demonstration project.

http://www.netlib.org/nse/home.html The National HPCC Software Exchange.

http://info.cern.ch/ The World Wide Web Main Server, at CERN in Switzerland.

http://www.ncsa.uiuc.edu/ The National Center for Super-computing Applications (NCSA) Server.

D About the Authors

Prof. Geoffrey Fox is Professor of Physics and Computer Science at Syracuse University and Director of the Northeast Parallel Architectures Center (NPAC) (Email: gcf@npac.syr.edu).

Prof. Wojtek Furmanski is Research Professor in the Physics Department of Syracuse University and Associate Director of NPAC (Email: furm@npac.syr.edu).

Dr. Ken Hawick is Senior Research Scientist and Project leader at NPAC and was formerly Technical Manager at the Edinburgh Parallel Computing Centre, Scotland. (Email: hawick@npac.syr.edu).

Don Leskiw is President of The Ultra Corporation and a Project Manager at NPAC (Email: leskiwd@npac.syr.edu)

E Acknowledgments

This report was written for Rome Laboratory, one of the US Air Force Super-Laboratories as part of a contract to "Explore the InfoMall Concept" in the context of the US Air Force Materiel Command.

In carrying out this work we received invaluable help and assistance from Roman Markowski and Janusz Niemiec at NPAC, and from Donald Spector and Scott Gregory at Rome Laboratory.

It is also a pleasure to thank Marina Chen, Peter Christy, Paul Coddington, James Cowie, Sam DiNitto, Jack Dongarra, Ira Goldstein, John Graniero, Robert Jacobson, Walter Johnston, Carl Kesselman, Adam Kolawa, John Latta, Paul Messina, Richard Metzger, Peter Mills, Paul Nielsen, Marek Podgorny, Jan Prins, John Reif, Richard Schantz, William Scherlis and Shar Zand who provided either technical advice and expert opinion on the technological areas used as exemplars or other material used in the project.

F Consultant Report on Mosaic and the World Wide Web

A Network Hypermedia Information Browser

by Paul Coddington

Research Scientist and Project Leader, Northeast Parallel Architectures Center
111 College Place, Syracuse University, Syracuse, NY 13244-4100 — paulc@npac.syr.edu

Contents

What is Mosaic?

Mosaic is a network information access tool recently developed at the National Center for Supercomputing Applications (NCSA). It is available for Unix workstation, Macintosh and PC environments. It is the first easy-to-use multimedia interface to the vast repositories of information on the global Internet, and has therefore generated a huge amount of interest. It has been dubbed "the killer application of the 90s".

Mosaic provides a hypertext interface to the Internet. Hypertext is text which contains highlighted links, known as hyperlinks, to other hypertext documents. Each highlighted phrase (in color and/or underlined) is a hyperlink to another document or information resource somewhere on the Internet. The click of a mouse button on any highlighted phrase will follow the hyperlink, which means that Mosaic will retrieve the document associated with the selected hyperlink and display it.

Hypertext can be plain text or multimedia (i.e. images, audio, or video), also known as hypermedia. Mosaic allows the display of formatted text that can include inlined images. Multimedia sources that Mosaic cannot handle internally, such as MPEG movies, sound files, Postscript documents, and JPEG images, are automatically sent to external viewers (or players) for display. NCSA has a Mosaic Demo Document giving examples of hypermedia on the Internet.

The World Wide Web

Mosaic is a basically a graphical user interface to the World Wide Web (WWW), a networked system of information servers using a common protocol (HyperText Transfer Protocol or HTTP) that allows access to hypertext information across the Internet. The World Wide Web was developed by CERN, the European high-energy physics research center, as a way for research groups in many different countries to share information. The Mosaic client communicates with HTTP servers, and can also interface with other Internet protocols such as FTP, Gopher, WAIS, NNTP, etc.

Since Mosaic was introduced about a year ago, the number of WWW servers and the WWW traffic over the Internet has increased dramatically. The number of people accessing the NCSA WWW server increased at 11% *per week* in the second half of 1993, and the server is now handling over 1 million accesses per week. Sun Microsystems set up a Mosaic/WWW server for information on the Winter Olympics that was handling up to 100,000 users per day and 32,000 page requests per hour.

Uses of the WWW and Mosaic

Many universities, research institutions, companies and government departments are setting up WWW servers to provide information about themselves and their activities.

Most government departments have set up WWW servers (or are converting from Gopher to WWW) to make government information readily accessible to researchers, educators, lawmakers, and the general public. Examples include the Department of Education, the Department of the Interior, the Census Bureau, and the National Coordination Office for High Performance Computing and Communications (HPCC). Most government research labs also have WWW information servers, for example NASA, The National Center for Atmospheric Research, and Los Alamos National Laboratory.

Many of the Department of Defense research labs have WWW servers containing a lot of information, for example the Army Research Laboratory, the Army Corps of Engineers, the Army High Performance Computing Research Center, and the Naval Research Laboratory. The Navy has a gateway to all Navy on-line resources, known as NavyOnLine.

Here is a more comprehensive list of WWW servers for government departments and national research labs.

A number of universities are setting up WWW servers in the role of Campus-Wide Information Systems, providing information for current or prospective students. University departments are also setting up servers to provide information on courses, faculty, and research activities. Here is a list of American universities on the World Wide Web.

The commercial use of the WWW has been increasing dramatically. A number of companies are either setting up their own WWW servers to advertise themselves and their products, or going through other organizations who offer access to the WWW. Some servers allow products to be ordered over the network. Examples of commercial servers include CommerceNet, the Internet Business Directory, and the Internet Shopping Network.

The NPAC/InfoMall WWW Server

I have been heavily involved in setting up a WWW information server for the Northeast Parallel Architectures Center (NPAC) at Syracuse University. We have found that using the WWW is an invaluable way of disseminating information about NPAC, including information on our research projects, as well as finding out about related work going on elsewhere in the world.

We are also using Mosaic to provide information about InfoMall, NPAC's technology transfer program. This provides a way for the InfoMall partners to keep in touch with what is going on with InfoMall and NPAC, and to advertise their own projects and products. InfoMall also aims to provide access to information on state-of-the-art HPCC technologies via InfoTech, and a "warehouse" of HPCC software via InfoWare. These two projects are being organized in conjunction with the HPCC National Software Exchange that is being developed by the Center for Research in Parallel Computation.

A lot of research at NPAC is in the area dubbed "InfoVision" - Information, Video, Imagery and Simulation on Demand. All NPAC InfoVision applications and demonstrations, including video on-demand, simulation on-demand, image and text on-demand, and access to educational information, have been or can be made available across the Internet under Mosaic.

The Future of the WWW and Mosaic

Mosaic seems to be poised to make a big impact in the commercial world. Most of the developers of Mosaic have been lured away from NCSA to a new company headed by the founder of Silicon Graphics. Their goal is to provide a variety of commercial services using Mosaic. There is also a large consortium of Silicon Valley companies and research institutions developing CommerceNet, a WWW-based business network. One of their main goals is to provide a secure service using encryption techniques, so that private information such as credit card numbers can be securely transferred over the Internet.

The use of WWW and Mosaic for the dissemination of information by US Government departments has been growing rapidly. ARPA has been strongly pushing for research information to be made available via Mosaic. Mosaic provides a simple way for information to be shared within a collaborative project for which the partners may be at different institutions. It also allows workers in independent groups or departments in the same large institution or company to be aware of what is going on in other sections of the institution, and hopefully avoid duplication of effort and create synergy between different projects.

One of the main problems to be faced now and in the future, given the explosive growth of the amount of information available on the Internet, is how to find specific information. *Data mining* is an area of great interest in the academic, military and commercial worlds. A substantial amount of research is currently being directed towards *search engines* that try to make it easier to find information on the Internet. There is particular interest in recent ideas such as *intelligent agents*, or *knowbots*, which search the Internet looking for new information on particular topics and update the user with relevant information. There are a currently number of information-gathering programs for the WWW, known variously as wanderers, robots, worms, or spiders.

Another approach is for experts in an area to act as "Internet librarians" and collate information for a particular area. The is being done by the World Wide Web Virtual Library, EINet Galaxy, and the Global Network Navigator Whole Internet Catalog. Currently these types of services are just pointers to available information resources. What is really needed is for this information to be placed in context. For example, resources could be grouped as to whether they were more suitable for research groups, undergraduates, high school students or younger children. Some explanatory structure is also helpful, rather than just having a list of hyperlinks pointing to various information resources. This kind of material is currently quite rare and often hard to find, and generally is only available in a few specialized subject areas.