Smart Instruments and the National Collaboratory

Barry M. Leiner (ed.)

Research Institute for Advanced Computer Science
NASA Ames Research Center

IN-35
43104
P16

RIACS Technical Report TR-89.45
November 1989

RIACS

Research Institute for Advanced Computer Science

Smart Instruments and the National Collaboratory

Barry M. Leiner (ed.)

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report TR-89.45
November 1989

The National Collaboratory is a concept for greatly enhancing both collaboration between scientists geographically dispersed across the network and access to remote resources and facilities. At the request of Dr. William Wulf, Director, NSF Directorate for Computer and Information Science and Engineering, an invitational workshop was convened at Rockefeller University on March 17-18, 1989. The workshop developed a set of recommendations for a research agenda leading to the National Collaboratory.

The workshop was organized as a set of sessions addressing different aspects of the National Collaboratory. This report documents the results of one of the sessions focussed on Smart Instruments. This session discussed the needed research to assure that networks can provide full and effective access to remote instruments and facilities.

Work reported herein was supported in part by
Cooperative Agreement NCC 2-387 from
NASA to the Universities Space Research Association (USRA).

TABLE OF CONTENTS

1. Introduction and Summary

Modern computer and communications technologies are revolutionizing the conduct of scientific research. Through the incorporation of microprocessor and other computing capabilities into scientific instrumentation and the attachment of such instrumentation to computer communication networks, experimental facilities are accessible to teams of scientists operating from geographically dispersed locations. Data can be gathered, moved, and analyzed from the investigator's home institution. Multiple instruments may be used in coordinated experiments to investigate the relation between different scientific parameters. Scientists can collaborate in all phases of scientific research.

The process of scientific investigation ranges from the planning and design of experiments and the required facilities through the actual conduct of the experiments to the analysis of the data and publishing of results. In each of these stages, scientists need access to remote resources and to cooperate with engineers, system operators, and other scientists in order to achieve the desired scientific objectives.

With the widespread use of computing to control scientific experiments and instruments, people are learning how to use instruments more effectively, but often at the cost of more processing power and storage. The costs of such hardware and software are often not factored into planning for instrument budgets. Thus, understanding the *smart instruments* and their requirements is critical to the planning for future scientific investigations.

In this report, we explore the process of scientific experimental investigation and ask what capabilities are required of the collaboratory to support such investigations. We first look at a number of examples of scientific research being conducted using remote instruments. We then examine the process of such research, asking at each stage what are the required capabilities. We finally integrate these results into a statement of the required set of capabilities needed to support scientific research of the future.

2. Experimentation Using Remote Instruments

In assessing the impact of the national collaboratory, one must look at the results of the experiment being conducted and its resulting impact on the scientific community. Many examples currently exist of scientists attempting to utilize remote experimental facilities in the conduct of scientific research. These include:

- Operation of high energy physics experimentation (e.g. LEP)
- Access to remote observatories (e.g. Sonderstrom Observatory and space-borne telescopes.)
- Operation of onboard Space Station experiments such as microgravity material science experiments,
- Operation and coordination of the airborne and spaceborne earth observation sensors, and
- Access to remote supercomputer simulation facilities, where the experiment is conducted via simulation rather than "wet labs."

Note that it is important to distinguish between the experiments and the instruments performing those experiments. Even today, but especially in the future, we are going to find more and more that the scale of scientific investigations requires a sharing of resources. Thus, while "small scale" experiments may still involve a single instrument designed specifically for that experiment, more typical will be major experimental facilities (e.g. Superconducting Supercollider) designed for use in many different experiments. Furthermore, many experiments (e.g. earth observation) will require the simultaneous use of several instruments.

To understand more fully these requirements on the collaboratory, we now discuss in somewhat more detail one example: the Global Change research program. The FCCSET Committee on Earth Sciences has recently issued a report[1] presenting an initial strategy for a comprehensive, long-term U.S. Global Change Research Program. The goal of the Program is *to provide a sound scientific basis for developing national and international policy on global change issues. The scientific objectives of the Program are to monitor, understand, and ultimately predict global change*. In particular, the objectives are:

1. Establish an Integrated, Comprehensive Monitoring Program for Earth System Measurements on a Global Scale.
2. Conduct a Program of Focused Studies to Improve Our Understanding of the Physical, Chemical, and Biological Processes that Influence Earth System Changes and Trends on Global and Regional Scales.
3. Develop Integrated Conceptual and Predictive Earth System Models.

The Program contains seven integrated and interdisciplinary science elements: 1) Biogeochemical Dynamics, 2) Ecological Systems and Dynamics, 3) Climate and Hydrologic System, 4) Human Interactions, 5) Earth System History, 6) Solid Earth Processes, and 7) Solar Influences. Thus the program is heavily interdisciplinary and involves multiple federal agencies: NSF, DoE, DoI/USGS, NASA, DoC/NOAA, EPA, and USDA. The proposed budget totals over $130 million in FY89 and $190 million in FY90. This budget is only for the increment to conduct the research in global change, and does not include currently funded activities in related and supporting activities.

To accomplish this program will require a high performance computing and communications infrastructure. Earth sensors generating data rates of hundreds of megabits per second must be fed into combined databases for global analysis. Supercomputers must be used to model and predict behavior, to analyze observational data, and to compare models to observed data. Powerful graphics workstations must be used to assist scientists in visualizing global environmental processes. Advanced collaboration technologies will be needed to help coordinate the gathering of sensor data, the analysis of that data, and the comparison of results across disciplines. High performance networks will be required to connect scientists, supercomputers, sensors, and databases. Furthermore, these networks must connect scientists from a variety of

[1] *Our Changing Planet: A U.S. Strategy for Global Change Research, A Report by the Committee on Earth Sciences To Accompany the U.S. President's Fiscal Year 1990 Budget*

disciplines and organizations, and provide them access to resources managed by a variety of institutions around the earth.

Thus, the capabilities of the national (in fact international) collaboratory are not only desirable, but required to achieve the goals of the Program. Collaboration between instruments is required to achieve coordinated measurements of specific areas and/or events. Simultaneous data collection from such instruments must take place to support future analysis, and this analysis will require databases spanning many years. The sensors involved will generate very large data rates and therefore methods will be required for dealing with massive datasets from different types of sensors. Multiple disciplines must be involved, and this will require support for collaboration between a wide variety of scientists from such disciplines.

2.1. A Model of Interactions

In this report, we are particularly interested in the capabilities that must be provided in order to support the interactions between smart instruments and scientific investigators. To do this, we use a model of intelligent agents moderating the interactions shown in Figure 1.

In particular, we think of each entity involved in the collaboratory (e.g. investigator, instrument, database, computing resource) as having an intelligent agent between it and the network. These intelligent agents are responsible for acting on behalf of the attached entity, negotiating with other agents to carry out the assigned tasks. They are also responsible for interacting with their attached entity to provide the required interfaces. For example, a user might request the local agent to develop a plan for using several instruments in a single coordinated earth observation. The local agent would interact with the instrument agents to control the timing and scheduling of the measurement, and report status back to the scientist.

Thus, the desired set of interactions between the various users, instruments, and other resources of the collaboratory (indicated by the dotted lines of Figure 1) is created through the support of the interacting agents. This method of interaction support is already in use in a number of settings. For example, scientists using remote telescopes do so through the support of attached computing resources acting on behalf of the scientists.

This model particularly addresses the need for coordinated sets of measurements involving multiple instruments and multiple investigators. An alternative model would be to have observations made by independent researchers and instruments recorded in a common depository (archive). The coordinated experiments then result from joint analyses of these data. The difficulty with this model is that it does not provide for coordination in the taking of the data. Since it is impossible to take data from all possible parameters and directions simultaneous, some sort of coordinated decision making process is required. A prime example of the such a requirement is the earth observation process. The proposed model permits scientists to jointly decide what experiments to perform (i.e. what observations to take) using a relatively arbitrary set of instruments and then obtain automated assistance in organizing and controlling the experimental data collection.

A further simplication in the corresponding software model may be achieved by treating instruments as computer processes in the distributed system. This will enable an easy substitution of simulator modules for the instruments themselves, thus enabling a smooth transition from simulated experiments (for design and analysis) to and from the physical experiment.

2.2. The Process of Scientific Experimentation

The typical process of scientific experimentation is shown in Figure 2. The process begins with a scientist or team of scientists developing a proposed experiment. Typically, a proposal for funding this experiment will be developed and the required instruments and other facilities identified. Once the experiment and instruments are agreed upon, the process of design and implementation of such facilities is begun. We can refer to this initial stage of the experimentation as the design stage. Note that it involves multiple scientists working with government funding agencies, systems designers, instrument developers, and other such organizations and people to assure that the experimental facilities are developed in a way that satisfies the scientific requirements.

After the experimental facilities have been designed and installed, and the experiment plan developed, the actual experiment is conducted. During this phase, a (team of) scientist(s) uses (multiple) instruments to gather data for future analysis. At the completion of the experiment, the scientist(s) analyze the data, often comparing it to data gathered in prior experiments and computer simulations and to previously published results.

While this process has been portrayed in a somewhat linear fashion, it clearly is iterative. New experiments are proposed based on prior results. Operating parameters for the experiment are changed based on preliminary analysis. Design changes are made to the instruments to respond to new requirements and deficiencies identified during initial experiments and through simulations and analysis.

With the widespread availability of computer networking, much of this process of scientific experimentation can be carried out without the investigators leaving their home institutions. By incorporating *telescience* capabilities into the experimental facilities and by providing the required information infrastructure to permit access and manipulation of those facilities, vast increases in scientific productivity can be achieved. Furthermore, through these same capabilities, new types of coordinated experimentation can be conducted, enabling new, previously impossible investigations.

3. Required Infrastructure for Telescience

This new capability for *telescience* (the conduct of science using remote resources) does not come without a price. Achieving these increases in productivity requires the development, installation, and maintenance of an information system infrastructure to permit access to remote resources along with the capabilities in the instruments themselves to permit remote access and operation. We can conveniently discuss these required capabilities in terms of the three phases of the scientific process described above (design, operations, and analysis), recognizing that many of the capabilities are common to two or three phases.

3.1. Required Capabilities

Designing the experiment and required experimental facilities requires remote access to both people and resources. The following are examples of the kinds of interactions between people that need to be supported:

- The joint development of a proposed experimental plan,
- Iteration on a proposed experiment between the principal investigators, the instrument developer or facilities manager, and the government program managers,
- Debugging of the software residing in the various instruments, the data archives, and the workstations,
- Design validation for scientific payloads to assure that they meet spacecraft specifications.

Furthermore, even in the design and development phase, there is a need to test hardware and software. Thus, there is a need to support remote access to testbed and simulation facilities as well as design and specification databases.

During the operations phase, when the experiment is actually being carried out, there is a need for real-time interaction between the scientists and the experiments. As we saw above, many modern experiments, particularly those in the observation sciences (e.g. Earth Systems Science) involve making near-simultaneous observations from widely dispersed sensors developed and operated by dispersed scientists. Thus, in real-time, scientists often dispersed around the globe need to be able to assess the state of their remote instruments, determine quickly whether the data being gathered is satisfactory, and make any required adjustments. This requires not only real-time access but also the ability to rapidly understand the data. In turn, this may require computer-supported collaboration to permit the various scientists and engineers to jointly carry out their roles. For example, a principal investigator on the ground may need to advise a payload specialist on the Shuttle as to required steps to be taken in a materials science experiment being conducted in microgravity.

Another form of real-time capability to assure high-quality data may be the access to simulations. As data is gathered, it can be compared to simulation results and the parameters of both simulations and the physical experiment adjusted to permit suitable comparison of the underlying models to the physically observations. This is facilitated by having a user be able to interact with the simulation and the ''wet experiment'' in a similar manner.

Finally, once the data is available, scientists need to collaborate in the analysis of the data. This requires the ability to jointly view analysis results from analysis facilities which may require the ability to deal with massive amounts of data. For example, the SAR (Synthetic Aperture Radar) currently under design for earth observation can generate 300 Mbps of raw data. Analysis also needs the ability to compare with prior results, and therefore access to data archives.

To support the above modes of operation requires a high-performance, highly functional information system infrastructure having such capabilities as:

- Standard data representations for the exchange and archiving of experimental results,

- Instruments designed for real-time interactive monitoring, control, and "quick-look" data reporting,
- Intelligent instruments capable of doing data processing and reduction,
- User interfaces that are similar for both accessing simulations and "wet experiments".
- High-bandwidth low-latency networks, and
- Remote access capabilities that provide adequate safety and security to protect both the instruments and the people around them.

3.2. Technical Issues for Smart Instruments

In order to achieve the desired and required capabilities in the national collaboratory to support the development and use of smart instruments in the anticipated experiments of the future, a large number of technical issues need to be addressed. To facilitate the discussion, we break the issues into those mainly related to design, then operations, and finally analysis. It must be kept in mind, though, that many of these issues, while primarily of interest in one phase of the process, in fact cut across all three phases. For example, while issues related to simulation may be of primary interest in the design phase, they are also needed during operations (e.g. for fault diagnosis and correction) and analysis (to validate models against experimental data.)

Issues in Design Support

Many of the needed tools for the design phase were addressed in other sessions of the workshop and we just note here the need for such tools to support collaboration between people. These tools are also needed for collaborative operations and analysis.

There are a number of issues that are specific to the design of smart instruments. These include:

Smart Agents for the Design of Experiments

As instruments become more complex and expensive, are used in multi-sensor experiments, and are shared by multiple investigators, the design of the experiments become more complex. Standard collaboration tools facilitate the design of such experiments, but in addition, tools specifically aimed at the planning, scheduling, coordination, and operation design of the experiments must be developed. These tools should themselves incorporate sufficient intelligence to be a real "assistant" to the scientists and engineers doing the experiment design. As an example of such a facility, the DARPA-sponsored MOSIS facility for sharing and brokering VLSI manufacturing and design is a cost-effective way of meeting the needs of many people in a timely manner.

Methods for Dealing with System Complexity

As instruments become larger and incorporate more intelligence, and as experiments attempt to deal simultaneously with several instruments, the complexity of the overall system increases significantly. We have already seen that this complexity, particularly in the software design area, has resulted in significant system development issues (witness the Advanced Tactical Fighter.) Appropriate methods for dealing with such complexity must be developed.

Simulation of Instruments

Prototyping instruments, while ultimately necessary, is an expensive process. Advances in computing technology permits these prototypes to be done, at least in the initial stages, through computer simulation. By having ''totally compatible'' hardware and software instantiations of the instruments, debugging of the instrument software both in the design phases and even during operations can be facilitated.

Distribution/transmission of Simulated Instruments

Once a complicated instrument can be simulated, it can then be distributed (by electronic transmission or otherwise) to a large number of scientists. This can allow multiple scientists to work on the development and evolution of shared instruments, or to conduct a large simulated distributed experiment (e.g. in robotics research.)

Instrument Software Evolution

As instruments evolve with new capabilities, there will be a need to incorporate them into existing systems. For example, new instruments will have to be installed on Space Station over its lifetime. Methods for designing self-describing instruments will have to be developed to allow communications between the instruments and the underlying infrastructure.

Instrument Auto-Calibration

If computer controlled experiments and instruments are to become the norm (permitting reduction in human involvement in the data gathering phase), it will be necessary for automatic calibration techniques to be developed. This could involve either gathering and storing calibration data, or better yet, the instrument itself being able to calibrate its data providing the scientist pre-calibrated data.

Issues in Operations Support

The bulk of the research issues to be dealt with for smart instruments show up in the experiment operations phase. Teleoperations, while normally thought of as being remote operations, is better interpreted to mean indirect, i.e. interactions mediated by computer. Designing such instruments and the overall system to allow experiment operations to take advantage of embedded intelligenece requires attacking a broad set of questions:

Smart Data Gathering

Instruments are becoming capable of generating more and more data (e.g. the SAR and HiRIS mentioned above.) This taxes the communication and data archiving system. In addition, many of the instruments are capable of having parameters (such a direction of look and spectral band) adjusted under computer control. Many experiments only provide a single ''shot'' at gathering the data (e.g. deep space probes, experiments where the underlying conditions change). Incorporation of intelligence into the instruments allows the possibility of ''self-directed'' data gathering, with the instrument itself deciding when data is significant and should be transmitted, setting parameters based on local feedback, and doing preliminary data reduction. This can lead to both reduction in communications and archiving requirements and better scientific data.

Support of *Quick-Look* Analysis

Similarly, the ability of a remote scientist to interact in real time with the instrument, determining the quality of the data through preliminary analysis and adjusting instrument parameters has a similar set of benefits.

Multi-Sensor Experiment Data Gathering

The success of experiments involving multiple sensors/instruments depends on the coordination and control of the data collection. The system (perhaps through the agents) must provide the capabilities required for such coordinated control, such as time synchronization of data collection.

Self-Describing Instruments

To support later analysis, it is critical that the conditions be recorded and encoded under which the data was gathered. The instrument agents can play a critical role here, interacting with sensors separate from the instrument itself to integrate the overall status (e.g. date, time, temperature).

Robust, Safe, Unattended Operations

Many instruments have safety issues inherent to them. For example, telescope operations have to be conducted with regards to potential human hazards. As instruments begin to incorporate intelligence and hence more autonomous behavior, careful consideration must be given to robustness and guarantees of safe operation. This issue also relates to the robustness of the instrument itself. For example, a space telescope must be smart enough to not point itself into the sun, thereby burning out its sensors.

Real-time Control

Distance (and corresponding delay) between scientist and experiment exacerbates the issues of real-time control. To allow remote control of instruments will require an appropriate division of labor between local intelligence in the instrument and the higher level control required by the remote scientist. It is likely that, in practice, that the software interdependencies of the collaboratory will be like those of a real-time distributed operating system.

Software Models

To facilitate real-time operation and feedback, software models of the instrument can be helpful. Through such models interacting with models of the environment, sources of problems in instrument operations and unexpected data can be understood, leading to corrective hardware or software actions.

Remote Distributed Control for Widespread Access

As experiments become broader in scope, including multiple scientists and multiple instruments, there will be a need for a common model of automated instrument control and remote interactions. This model must go beyond the basic communications paradigms and provide essentially a distributed system control architecture that all relevant instruments can use. The model described above of an agent acting on behalf of the instrument, with the agent responsible for interactions with other agents, is the beginning of such a model.

Resource Allocation, Instrument Ownership and Access Control

It will be desirable to design an operations architecture that allows instruments attached to the collaboratory network to be accessible to anyone using the collaboratory, subject to authorization. This raises the issue of who owns and controls the access to the instrument, and how can access control be provided in an appropriate manner. For example, if multiple scientists desire to use a nationally sponsored instrument but with incompatible instrument settings, automated techniques should be developed that facilitate the resolution of access control and resource allocation questions.

Data Sharing

Similarly, as instruments gather the data, there is a question as to the appropriate data dissemination mechanism. Historically, shared data has been handled through centralized data archives. With the development of the national research network and the availability of low-cost satellite terminals, a new possibility arises - broadcast scientific data. Real-time sensor instrument data (perhaps preprocessed by the instrument itself) can be broadcast with any and all interested scientists using a smart agent to detect when the data is of interest to them, and then grabbing it off the airwaves.

User Interface to Smart Instruments

Advanced workstation capabilities with appropriate networked interactions to the remote instruments will be critical to facilitating remote and distributed control of instruments. Such interfaces will need to not only permit direct interaction but also visualization of preliminary data and tools for understanding the current state of the overall experiment.

Issues in Analysis Support

The purpose of building better and smarter instruments is to facilitate gathering data for future analysis. Thus, the smart instruments must be built in a way that supports the analysis. Issues that arise in this area include:

Flexible and Long-Term Data Structures

Instruments will evolve over the years, and new information will be gathered. However, important to the scientific process is the ability to compare newly gathered data with older archival data. Methods for formatting and structuring the gathered data must be developed that permits new analysis tools to be able to deal with data over many years. This most likely will require the development of techniques for self-describing data formats. These descriptions must contain not only the data itself, but methods for describing the conditions under which the data was gathered.

Data Storage

The evolution (perhaps revolution) in instruments themselves is leading to a rapid explosion in the amounts of data being gathered. Techniques for archiving and retrieving the massive amounts of data obtained over many years must be developed.

Interactive Analysis Support

These techniques must facilitate the scientific analysis of the data, accessed and obtained through the collaboratory. Because of the massive amounts of data involved, the techniques must include intelligent agents to act on behalf of the scientists in their search for relevant data and preliminary as well as ultimate analysis.

Multisensor Data Fusion

Many of the future scientific activities (e.g. Global Change research) involve the integration of data from multiple sensors. Having well-understood data formats facilitates the fusion of data from such sensors, as does an overall architecture that allows agents to gather data from different data archives. This capability must be developed and explored.

Data Ownership and Access Control

Scientists often have proprietary interests in the results of experiments under their control. Yet the premise of the collaboratory is collaborative scientific investigation. Techniques will be required for specifically handling the ownership of and access to scientific data and changes in such ownership, particularly in the difficult case where multiple instruments and scientists are involved in the experiment.

Sharing of Analysis Code

Not only is it desirable to share data, but also tools to analyze that data. Issues here range from software portability through to proprietary interests of the developers.

4. Conclusion and Recommendations

A major goal of the collaboratory with respect to smart instruments is to develop a software architecture for instrument management and data capture, even within a single laboratory. This architecture should allow the linking of instruments and researchers from very different fields.

Recognizing the purpose of this workshop was to develop a set of recommendations for NSF/CISE actions that could further the national collaboratory, we suggest the following as concrete steps in that direction.

- Research and development should be undertaken into the development of techniques for preserving data formats and analytical tools, particularly for data gathered from multiple instruments about the same physical phenomena.
- Basic research should be undertaken into trusted, real-time, distributed systems suitable for the support of a collection of intelligent instruments operating in a coordinated manner to conduct a joint experiment.
- Collaboration technologies should be developed and demonstrated that support coordination of data gathering amongst multiple instruments, real-time experiment planning, and distributed operations of experiments.

Recognizing that CISE cannot attack the development of smart instruments by itself, we recommend that CISE undertake collaborative activies with other offices and agencies in the design and development of smart instruments and their use. In particular:

- CISE Program Managers should interact with the Program Offices responsible for the development of large instruments to encourage their design in a manner that facilitates incorporation into the national collaboratory in general and teleoperations in particular. CISE should undertake some pilot projects jointly with such Program Offices to demonstrate the feasibility and advantages of such capabilities.

- CISE, in cooperation with Program Offices responsible for existing instruments, should underake some prototyping pilot projects to develop and demonstrate teleoperations "wrap-arounds" on existing instruments. These wrap-arounds should show how intelligent control and interfaces may be added to existing instruments.
- CISE should promote experiments in the broader accessibility of instruments for collaborative work, with the aim of a better understanding of what is commonly required for all instruments to made part of a computer-controlled environment.
- All three of the above recommendations can be facilitated through the funding of computer/communications experts working with application scientists in real laboratories to conduct prototype experiments.
- In light of the fact that current and future instruments are in large part software, CISE should assist in the review of such issues as software lifetime and plans for instruments under development and in planning. CISE should encourage that adequate attention be paid to software issues in the early design and planning for major instruments.

5. Session Participants

Barry Leiner - Chair

Vint Cerf - Rapporteur

David Farber

Larry Smarr

Robert Sproull

Y.T. Chien

Herbert Schorr

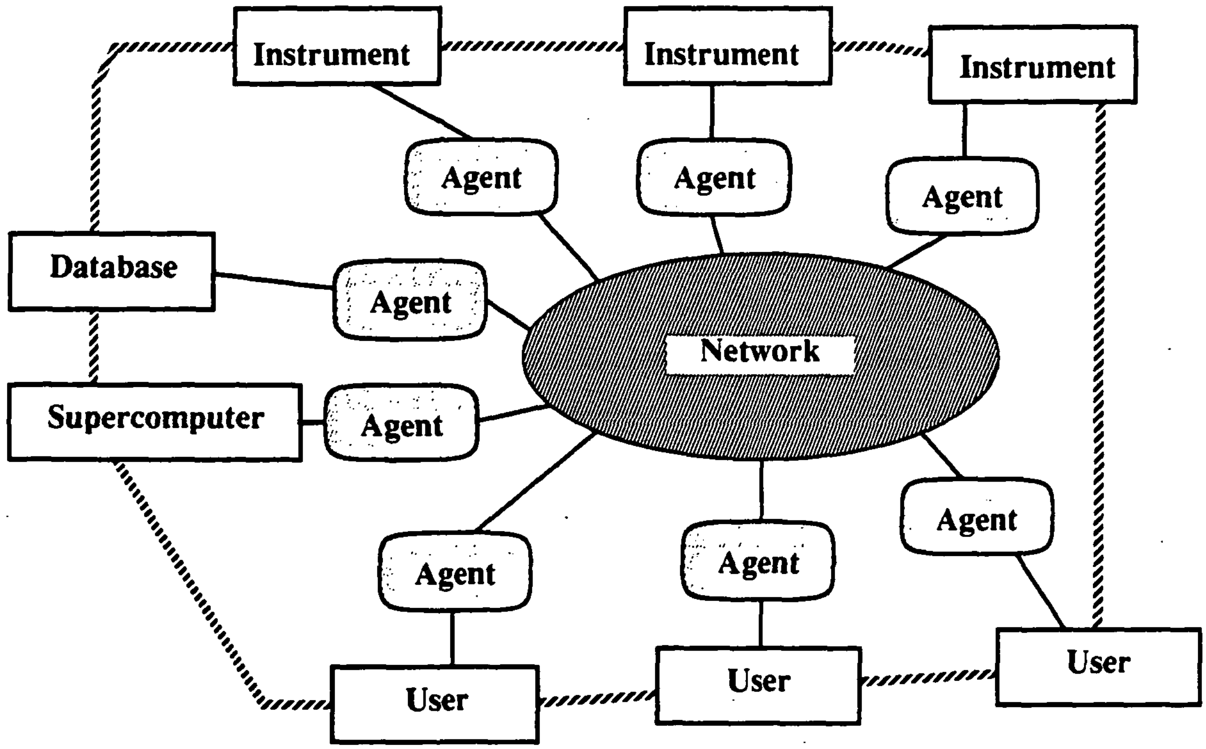

Figure 1: Agents for Collaboratory Interaction

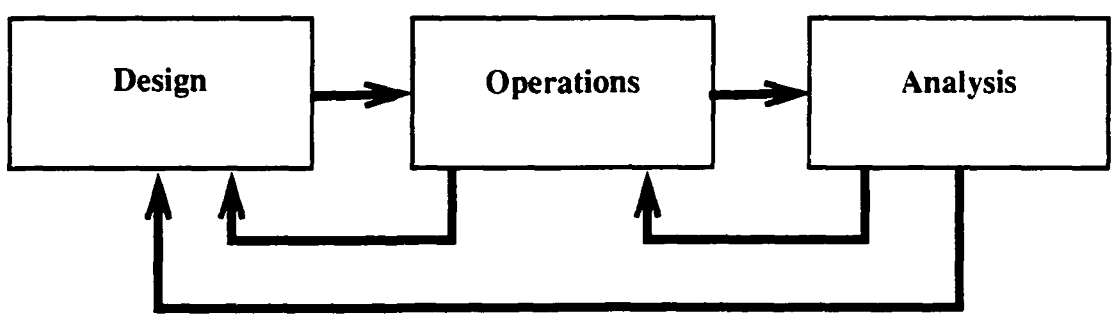

Figure 2: The Scientific Experimental Process

RIACS

Mail Stop 230-5
NASA Ames Research Center
Moffett Field, CA 94035
(415) 694-6363

The Research Institute for Advanced Computer Science
is an Institute of the
Universities Space Research Association
The American City Building
Suite 212
Columbia, MD 21044
(301) 730-2656