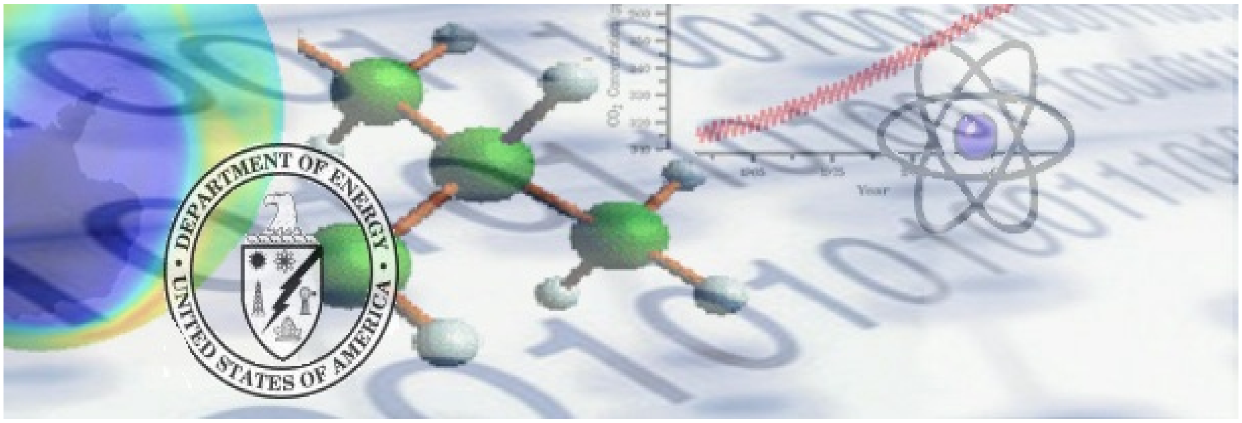

THE STATE OF DATA MANAGEMENT IN THE DOE RESEARCH AND DEVELOPMENT COMPLEX

Report of the Meeting
“DOE Data Centers: Preparing for the Future”

Held July 14-15, 2004
Oak Ridge, Tennessee

November 5, 2004

Prepared by the DOE Office of Scientific and Technical Information

Finding

DOE needs a Department-wide data policy that recognizes life-cycle data management as an integral part of research programs and projects.

Introduction

DOE invests billions of dollars in R&D projects that generate and use data as a critical resource integral to science. Examples abound. Through its large-scale experimental facilities, DOE provides its researchers with an ever-growing array of world-class instruments that produce vast amounts of data in probing the fundamental workings of science. Science communities have developed many new ways to exploit data collections. Today, state-of-the-art computer simulations and modeling have become as important to the advancement of science as theory and experimentation always have been. Recently, Ray Orbach, Director of the Office of Science, highlighted the growing emphasis on computational science as a fundamental shift in research. Simulation-based science is dependent on the availability, accuracy, and usability of the data previously generated by scientists and funded by DOE and its predecessor agencies.

Increasingly, valuable source data are being stored in digital databases. The charter for the National Science Board's (NSB) Task Force on Long-Lived Data Collections (LLDC) states, "the number of digital databases is increasing rapidly, and database sizes are in many cases growing exponentially. Digital databases are now an essential and increasingly large component of the cyber-infrastructure that underpins research and education."

Early in 2004 the NSB created the LLDC Task Force to study data management practices across the federal science agencies and determine how they should shape future NSF funding policies. Some R&D agencies already have data management policies, while other R&D agencies are now exploring the needs relating to data preservation and use.

On July 14 and 15, 2004, data management experts from seven data centers in the Department of Energy, as well as representatives of the Office of Scientific and Technical Information (OSTI) and other DOE organizations with interests in data management, met in Oak Ridge, Tennessee, to discuss data management issues and concerns from a DOE perspective. Also participating were Christopher Greer, Executive Secretary to the National Science Board's LLDC, and Jeffrey Hayes from the Space Science Program at the National Aeronautics and Space Administration (NASA) Headquarters Office. (See Appendix A for a summary of the meeting and Appendix B for a list of attendees and their affiliations.)

The participants agreed the first priority in addressing the issues and opportunities associated with data management in the Department is the need to develop a unified agency-wide data policy. Such a policy would encourage best practices in the use and preservation of valuable data resources in the advancement of science.

It is envisioned that a DOE data management policy would be developed by representatives from the DOE data community, working in collaboration to define and determine the best interests of the Department. The policy would encourage program managers and project managers, working together, to address such life-cycle issues as: when data should be made available and to what audiences; how data are to be preserved and for what period of time; what documentation needs to accompany the data to assure validation and future use; and what types of funding mechanisms and infrastructure are needed to assure appropriate retention and access. It would also suggest guidelines that data-intensive programs and facilities may adopt to assure that data generated are effectively managed and made available. It would help to collect and retain data that might otherwise be lost to future scientists, and would assure that DOE maximizes the return on its research investment by securing the output of the billions of dollars annually spent on R&D.

At the same time, a data management policy would not be prescriptive in nature and would not abrogate the roles and responsibilities of programs, data centers, or other data management facilities. It would be an "umbrella" policy for data generators, collectors, curators, and users. Such a policy would not impose operational requirements on individual data centers and would not seek to "standardize" data form or format. However, for data that are not currently managed for long-term availability and re-use, the policy would reflect the best practices of the data community and offer guidelines for establishing mechanisms for managing data.

This paper summarizes the issues associated with establishing a DOE data policy suitable for the 21^{st} century.

The Challenge

"Computation is now seen as an equal partner with theory and experimentation in the advancement of science," according to an article in *DOE Pulse* in 2002. [Taken from ORNL Review, Vol. 35, No. 1, 2002, "Visualization Tools: Interacting with Data in Many Dimensions."]

Science as a process has changed dramatically in the last century, often conducted by geographically dispersed teams using remotely located computers to run sophisticated simulation and visualization software.

Experimental and observational data are the traditional building blocks of science. New discoveries or applications are considered valid when the original, underlying data can be analyzed, verified, and reproduced with the same results time after time.

But the paradigms that define data management are changing. What constitutes data, how data are collected, who owns the data, how data are organized and stored, how data sets may be reused, and what ultimately happens to data are significant issues that are surfacing and demanding attention. Old truths have become new challenges, simply because of exponential growth of data and the capability to collect, organize, store, and re-use it for future scientific endeavors. Sharing of data in multi-disciplinary and international collaborations has blurred traditional lines of scientific communication. New issues have arisen as technology enables new kinds of analyses and as numeric data and text data are integrated. End-users of scientific data are demanding better access to more collections and expecting better quality. Information organization and retrieval issues, once considered essential for published research findings, now also apply to data. The awareness of these changes among the federal agencies that support America's scientific research is growing. Indicative of this phenomenon is the decision of the National Science Board in 2004 to establish a task force to study the issues posed by long-lived data collections and make recommendations to the National Science Foundation on how to deal with these changes as they affect NSF-funded science projects.

Within the Department of Energy, there are about 10 designated Data Centers, each with its own set of criteria for collection, standards and format, methodologies for capturing supporting documentation, and policies for archiving various levels of data. The content scope for these Centers is scientific subject or discipline based. Individually, each of the Data Centers has a tradition of effectively managing its respective collection in spite of such constraints as budget shortfalls and staffing limitations. However, there is no infrastructure or policy within the DOE community to facilitate the sharing and use of data across disciplines and areas of scientific research, to provide a forum for sharing of lessons learned and best practices, or to provide unified DOE principles for data management issues.

In addition to the data collected by these Centers, there are a number of data-intensive programs and facilities that may archive their own unique data sets, but do not have a data center. DOE needs to consider if data contained in such collections are being managed to fully capitalize on their value now and for re-use in future scientific research endeavors.

Another phenomenon occurring in the DOE data community is the development of multi-purpose user facilities, such as the Spallation Neutron Source (SNS), that provide the resources for data generation in a number of different scientific disciplines and areas of interest. The results from the SNS experiments will not fit into the traditional mold of a narrow, discipline-bounded set of data.

Researchers will come from many disciplines across the global science community, each generating data for a specific need. Issues such as data ownership and DOE rights of re-use compound the problem of how to manage resulting data in a way that will maximize its value to scientific advancement.

As terabytes at a time of scientific data are fed into sophisticated computer programs and models, the results generate even more data that must be manipulated by supercomputers, shared via high-speed, broadband networks, analyzed via visualization techniques, and archived intelligently so that they can be retrieved and re-analyzed in the future with even more powerful tools.

To ensure successful simulation and support the increase in computational sciences, serious attention should be paid to numeric data and its documentation and preservation. Scientific data should be thoughtfully managed, made retrievable, and archived for extensive, multiple uses now and for re-use and mining in the future.

Data management, data archiving, data access and sharing, and the future re-use of archived data are increasingly topics of concern across the Research and Development (R&D) agencies of the United States government. While many of the challenges facing data management are not new, the quantum growth in the amount of data generated and needed have highlighted those challenges. Other challenges are new, arising from today's collaboration models and the layer upon layer of complexity added as computational tools for simulation and remote experimentation generates new types of data.

The LLDC charter states the case simply: "It is timely to consider the policy ramifications of this rapid growth of data collections in the NSF-supported community."

Sharing Issues and Concerns of the DOE Data Centers

This section summarizes the data management issues that were identified and discussed throughout the meeting as areas of interest that all DOE Data Centers share. A common thread throughout these discussions was the difficulty in addressing issues due to misconceptions that exist about data, its value, and the importance of re-usability. Therefore, efforts to reconcile issues must be addressed concurrently with efforts to overcome misconceptions.

There is an emerging recognition that data need to be managed. With the growing capabilities offered by technological advances, the value of data as the raw material for computational simulation increases as a critical component of scientific discovery; if data are not collected, managed, retained, and optimized for re-use in an ever-changing technological environment, simulation-based science is limited. As Charles Dunford, Deputy Director of the National Nuclear Data Center, stated, "Critically analyzed data are the underpinning for discovery."

Merely storing data after project completion is no longer sufficient. Long-term preservation along with ease of accessibility needs to be planned and implemented. Some of the most critical and careful analyses of project data may not take place until many years after project completion. University and lab researchers may rely on previously gathered data in projects to develop new concepts and new technologies. If the data have not been maintained and upgraded, documented and indexed, and made available for retrieval, these analyses cannot be done. Almost every data manager can relate cases where old data was suddenly needed for some critical, high visibility project. The point was also made that it is nearly impossible to judge at the time of initial storage how important that data might be to unknown researchers in the future. In his presentation, Jeffrey Hayes talked about the huge increase in the NASA Space Science Program in re-visiting and re-analyzing older data and predicted that the trend will continue to rise.

Better processes and practices are needed to alleviate the difficulty in obtaining source data. During informal planning meetings prior to the July gathering, one data management expert commented that some of the most fundamental data sets in the world are not available to U.S. researchers. Reasons vary, from a lack of defined requirements to gaps in the life cycle process, researchers who are pushed by the need for funding to move on to the next project before they wrap up the details of previous ones, benign neglect, and even the fact that some scientists are reluctant to share data. This is not unique to the DOE environment, but as DOE becomes more computationally intensive and data driven, a proactive role needs to be taken to address the challenges to good data availability and management.

The need to manage and preserve data should be explicit considerations in project planning and management. DOE has an increasing number of large-scale user facilities where visiting researchers use the instrumentation to produce unique sets of experimental data. The question of what happens to these data when the visiting researcher leaves the facility and then wraps up his or her project is a pressing one. The research done on these facilities is usually selected based on investigator-initiated, peer-reviewed proposals. Turning over data files is not mandatory in many cases, nor is there a clear path from each facility to a corresponding data center. Because of the disconnects that seem to exist, data management solutions in the Department of Energy will probably need to involve funders, user facility managers, and data generators. There needs to be recognition that data should be preserved for re-use. Ideally this recognition should be reflected in policies for data deposit that then need to be reflected in the experiment proposals. In addition to the user facility proposals, a way to address this challenge is to specify timing and procedures for providing source data in the contract or grant language, just as technical reports or peer-reviewed articles that provide R&D project results are specified as deliverables. This is beginning to happen in various instances and, if adopted agency-wide, could be very effective. Greer commented that the LLDC Task Force has found strong support for this approach.

Better processes and practices are needed to alleviate the difficulty of obtaining documentation. The group agreed that simply turning over the source data is not adequate. In order to reproduce the results of an experiment or re-use the data in future analyses, information accompanying and documenting the data is just as important as the data sets themselves. This information could include protocols used, details of instrument settings and descriptions of other conditions, software and software manuals, specialized thesauri, images, etc. Another very important reason behind the difficulty in getting the documentation along with the data is that the cost of preparing the documentation can be significant and is often not specifically funded. Quality documentation should be prepared before projects are completed and the funds have been spent or ended. There is a need for a Department-wide understanding of what constitutes quality documentation, and how it supports data generation and long-term availability for re-use.

Technology provides new opportunities to address data management issues, but it cannot provide solutions without thoughtful planning and application. Focusing on the enabling technology rather than the principles and best practices governing collection, search and retrieval, distribution, archiving, and re-use of the content and its intrinsic value is a common problem in information and data management. Today's emerging technology offers exciting opportunities to address data management issues in ways never before possible. The capabilities, the ever-increasing capacities, and the innovative tools, however, cannot solve problems unless they are thoroughly integrated with the intellectual effort that goes into careful planning, utilizing best practices, and anticipating future needs for data that have not even been expressed yet.

Data retention (what to archive and for how long) should be addressed discipline by discipline. The group noted that there are many layers of data, leading to considerable confusion as to what should be archived, for how long, and how to maximize its usefulness. Besides the accompanying documentation mentioned above, another key challenge in data management is determining which data sets should be archived from a given project. The answer to these questions is more a function of the specific field of science, or even a specific project, rather than a single set of retention criteria for all data generators. The question of whether there is a common level of specificity required to enable interdisciplinary sharing of data also needs to be answered.

A related question is whether only "good" data should be archived, i.e., data that meet a benchmark for certainty, or whether preliminary sets of data should also be kept. The issue of "raw" source data vs. data that have been analyzed, critiqued, perhaps even corrected becomes more important as simulation-based science generates additional levels of data. Such issues need to be addressed explicitly by cognizant managers working with experts in the various disciplines.

Metadata must be optimized for future retrieval, assimilation, and re-use. Talk of cross-referencing led to the topic of metadata in general. Some disciplines have no standards for metadata that describe data for future retrieval; others have several conflicting standards, and in some cases new standards are in the process of replacing old standards. Larry Voorhees, Manager of the Distributed Active Archives Center (DAAC) for Biogeochemical Dynamics, was one of several participants voicing the opinion that metadata and the indexing of metadata is key to ever being able to find specific data again.

The issue of metadata is one that looks deceptively simple to those outside the field of data or information management. But how do you ensure that all data related to a particular scientific concept, process, or phenomenon can be retrieved in a search across multiple databases? As Hayes related, in space science, a single star may have as many as 50 different names and 300 synonyms for those names.

"Indexing is everything – and the thesaurus that goes along with it," Voorhees also noted. But, given that semantic connections change across scientific boundary lines, how many cross-references can successfully be built into a system? Another difficulty is that operational practices affecting indexing vary from data center to data center. Whether data collections in a particular discipline are usually aggregated in one location or distributed across many would be another factor that could affect indexing. Raymond McCord, Data Archive Manager of the Atmospheric Radiation Measurement (ARM) program, agreed with the critical importance of good indexing by referring to the "Wal-Mart Syndrome," that is, just because today's storage mechanisms can hold anything and everything doesn't mean you will find the one thing you want.

A related issue concerns protecting, or at least identifying, the quality of data sets. Should metadata be minimal citation-like information or should it carry information that helps a future analyst determine the pedigree and the relative integrity and certainty of the data it describes?

Data sets need to be referenced in order to be easily located by users of scientific literature. There is a need for better linkage, both conceptually and physically, between scientific literature and data stored in collections. Those who use the literature as a starting point for their own research projects often want to verify and cross-check observational data or replicate experimental data. However, unless the original source data can be found, identified with certainty, and retrieved along with information about processes, variables, and analyses made at the time of collection, this is nearly impossible. Jim Myers of the Computational and Information Sciences Directorate at Pacific Northwest National Laboratory, stated, "Journal literature is the ultimate level in compression of data." It also continues to be the first line of reference. There needs to be a better and more consistent way to guide researchers from the "lossy" record of the journal article to the source data itself.

Retention and re-use of data need to be addressed in the context of emerging needs for long-term management and curation. A key concern is that funding models for long-term archival care of data often do not exist in DOE. A research agency like DOE typically funds projects. When the project is completed, the funding typically stops. Too often, data are lost.

Among the best examples of sound data management in DOE are those that exist at DOE's data centers. Data centers do get operating budgets that may cover the storage of the data, even the increasing volume of data with improved, less expensive storage capacity. But the budgets of these centers cannot cover the cost of maintaining the data for long-term preservation and re-use in the future. Hayes noted that figures in NASA show that maintenance of large data archives requires up to ¼ as much funding as the project that generated the data. Many in the group agreed with this concept and expressed concern that the long-term perpetual care of data is not routinely funded.

Models which bundle the costs of data management into the costs of doing the research or ones that might create a national fund or endowment have been discussed in the scientific and information communities. Further examination of these and other alternative models would be useful.

Data, like information, should be widely accessible and available at no cost to the user. While better funding mechanisms are needed in many areas to support data management operations and archival care, the meeting attendees appeared to be in agreement that charging for access to data is not the way to alleviate the funding gap. In any case, OMB Circular A-130 dictates that the most that can be charged is the cost of incremental dissemination, which leaves the major base costs of operations unrecovered. In addition, charging for anything requires spending in order to support the accounting process. And when scientific data are not freely shared, the impact on entrepreneurial research, developing countries, and others with limited resources can be severe, slowing or even halting progress in some fields. Dunford pointed out that the impact could also include the occurrence of "poor quality" science, since those who cannot get the best information have to make do with less stringently analyzed data.

Raymond McCord recommended a paper written by Peter Weiss of the National Oceanic and Atmospheric Administration (NOAA) [*Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts*] that cites statistics purporting that information access limited to paying customers adversely impacts the overall economy of a country. When information is free and open, there are multiple users who will pursue different creative, developmental directions; this leads to a stronger, better economy.

Scientists should lead data management centers. Professional staffs of scientists and information specialists are needed to manage, maintain, and operate data centers effectively. They must be recognized in their scientific fields because they are the curators and custodians of highly technical knowledge. In fact, scientists who are familiar with the existing bodies of work are needed to provide expertise in established areas of science.

Attendees related several concerns about recruitment and retention of professional staff, from the need for specific expertise to the funding to hire and train recruits in the intricacies of scientific data management and curation. There was also a general perception that the field is "aging." As older staff members retire, the science knowledge they applied to the retrieval of certain data sets is lost. This loss of knowledge and experience in an established field of science can result in significant knowledge gaps as newer interdisciplinary sciences emerge. Science specialists are needed to successfully guide these new users across multi-disciplinary lines and back into the past areas of expertise. Investing in quality data over the long term requires an understanding of the value of the data.

Conclusions

A Data Policy Is Needed

The group reached consensus that DOE needs a policy for data management. It is envisioned that a DOE data management policy would be developed by representatives from the DOE data community, working in collaboration to define and determine the best interests of the Department. The policy would encourage program managers, when planning R&D programs, to consider such issues as: when data should be made available and to what audiences; how data are to be preserved and for what period of time; what documentation needs to accompany the data to assure validation and future use; and what types of funding mechanisms and infrastructure are needed to assure appropriate retention and access. It would also suggest guidelines that data-intensive programs and facilities may adopt to assure that data generated are effectively managed and made available. It would help to collect and retain data that might otherwise be lost to future scientists, and would assure that DOE maximizes the return on its research investment by securing the output of the billions of dollars annually spent on R&D. Planning for data needs to take into account a near-term, a 3- to 5-year outlook and, even more important, a long-term perspective. The plans also need to be based on input from user communities.

One attendee suggested that a data policy office might need to be established in DOE in order to handle the complexities and to power the data management models that would cover the range of scientific disciplines.

Establishing the Value of Data

The importance of establishing, in some practical and tangible way, the value of data preservation and sound data management was reiterated by the group several times. This is key to overcoming misconceptions and improving the funding position of data-intensive programs and facilities. It was suggested that a formal study, perhaps done by the National Academy of Science, would provide a tangible product that would have influence and credibility. Stories from actual experience, stories that would illustrate the pros and cons of the various issues were recommended as an integral part of this study and the resulting paper. Comparison costs must be included also, such as the cost to collect and preserve data versus the cost (or loss) to science of not having that data five years from now and therefore having to totally re-invent conditions and re-do projects.

Developing Expertise in Scientific Informatics

Finally, developing the scientific informatics community was recognized as a critical underpinning for more effective data stewardship. It was recognized that thought needs to be given to the recruitment, training, and incentivising of data scientists, specialists who are comfortable in both a scientific discipline and in data and information management. Related to this, it was emphasized that reward structures or motivational incentives need to be developed to encourage data generators to prepare and make pertinent data and supporting documentation available for long-term re-use as well as to encourage technical experts to work in the scientific informatics of these disciplines.

Creating a Unified Voice

The group recognized the value that a unified voice from data centers would provide in working on issues. It was agreed that a second meeting of DOE Data Centers would be useful for following up on ideas generated at the July meeting.

Next Steps

Create a DOE Data Management Community

key interest is to facilitate communication among the DOE's Data Centers to resolve or better handle some of the known issues. Data centers have best practices and valuable lessons learned to share. To foster the creation of a unified message from the data management community, several steps will be taken:

Attendees agreed to meet again (perhaps late Fall 2004) to continue discussions of the issues and on whether to establish a formalized group of data management experts.
Data management experts will be invited to discuss the issues and share best practices.
Current communications among DOE data centers will be improved.

The offer by OSTI to host a web site with an online discussion forum to facilitate information sharing among the group met with approval from attendees. OSTI will proceed immediately to set these up.

Work for Change

Efforts will continue to:
Actively work toward the goals of improving data management within DOE,
Improve visibility of the data centers,
Position DOE for potential changes in the federal guidelines for data management~~, and~~, and
Provide strategic direction for DOE's data management that will ensure a strong foundation for the science of tomorrow.

Document the Issues

As a result of the July meeting, a document will be prepared to present to the Scientific and Technical Information Advisory Board (STIAB) the need for developing a sound data management policy for DOE. We propose to ask STIAB to agree that OSTI should coordinate the development of a DOE data management policy. To begin this process, OSTI will coordinate the preparation of a paper summarizing the current challenges and the opportunity to resolve data management issues within DOE.

Appendix A

The Meeting

Directors and key staff members from Data Centers in the Department of Energy met in Oak Ridge, Tennessee, on July 14th and 15th to identify issues common to data center managers across the DOE complex and to determine if addressing these issues in a coordinated way to present a common voice for data centers would be useful. (See participants in Appendix B.) The Department's Office of Scientific and Technical Information (OSTI) initiated the meeting in response to the NSB/NSF activities and the needs expressed in subsequent discussions with DOE data managers across the Agency.

Walter L. Warnick, Director of OSTI, opened the meeting by highlighting the convergence of factors that had triggered the meeting. He also discussed OSTI's role in facilitating the meeting. "We understand alliance," he said, noting that fostering collaboration has been key to OSTI's success with multi-agency projects such as Science.gov, with cross-agency cooperation in CENDI, and in working with INIS and other international partners. In fact, OSTI coordinates DOE's Scientific and Technical Information Program (STIP) via a Department-wide collaboration. The purpose behind all of these working relationships is stated in OSTI's mission: "to advance science and sustain technological creativity by making R&D findings available and useful to DOE researchers and the American people."

Invited speaker Christopher Greer, Executive Secretary to the National Science Board's Task Force on Long-Lived Data Collections (LLDC), then previewed the work of the Task Force and some potential implications for DOE Data Centers. Among various topics, he discussed the need for a "federator," or a site that keeps track of data and sets standards, formats, and structures queries to find data; the importance of agency-wide data management plans; and direct and indirect models of funding support. The National Science Foundation hopes to receive recommendations from the Task Force as to which model would be better suited to the long-term financial support of data collections, a need that will, no doubt, outlive the projects and possibly even the organizations that spawned the data originally. Greer encouraged the group's comments throughout his presentation, saying that part of his job is to take back to the Task Force input from other agencies.

The next invited speaker was Jeffrey Hayes from the Space Science Program at NASA Headquarters. He shared the evolution of the Space Science program and policies for data management and archiving. This addressed a second goal of the July meeting, the need to look at what other agencies have done or are doing in the area of data management.

Hayes discussed the increased need for older data sets for use with new models and simulations. The results can create a new category of data with a whole new set of issues. Hayes noted that NASA projects based on re-analysis and computation using older data sets have risen from 10%

requests coming in to space science archives to 45%. He sees the increase continuing, especially with multi-wavelength data. This led to a discussion of whether there is such a thing as an expiration period for data.

At NASA, the distinction between conditions or experiments that can be replicated and those that can't lead to the realization that, without good data, no scientific experiment can really be replicated. If science can't be replicated, then it is not proven and may not actually be "science."

Certain questions were inherent in the meeting:

- Do the issues described by the LLDC Task Force exist in the DOE complex?
- Are there other data management issues that are specific or unique to DOE?
- Do the DOE data centers share with each the same issues or do those issues differ across scientific disciplines?

A key interest is to facilitate communication among the DOE's Data Centers to resolve or better handle some of the known issues. Data centers have best practices and valuable lessons learned to share.

More pressing was the group's recognition of the need for DOE to be proactive rather than reactive as groups such as the LLDC Task Force, Committee on Data for Science and Technology (CODATA), and others address the need for better data management in federal agencies. The goal is not simply to avert potential negative impact on the current operations of DOE data centers, but to act in partnership with other organizations and agencies to prepare for the next generation of data management, for the sake of science now and for the scientists of the future.

The discussions covered a number of subjects, most of which will require further consideration by the DOE data management community.

Appendix B

ATTENDEE LIST

DOE Data Centers: Preparing for the Future
July 14-15, 2004, Oak Ridge, Tennessee
Pollard Technology Conference Center

Mark Bannister
Controlled Fusion Atomic Data Center
Physics Division, Oak Ridge National Laboratory
Bldg 6010, MS 6372
P.O. Box 2008
Oak Ridge, TN 37831-6372
Telephone: (865) 574-4700
bannisterme@ornl.gov
http://www cfadc.phy.ornl.gov

Thomas A. Boden
Director, Carbon Dioxide Information Analysis Center (CDIAC)
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6335
Telephone: (865) 241-4842
bodenta@ornl.gov
http://cdiac.esd.ornl.gov/home.html

Bonnie Carroll
President
Information International Associates, Inc. (IIA)
1009 Commerce Park Drive, Suite 150
P.O. Box 4219
Oak Ridge, TN 37831-4219
Telephone: (865) 481-0388
bcarroll@infointl.com
http://www.iiaweb.com

Charles Dunford
Deputy Head
National Nuclear Data Center
Brookhaven National Laboratory
Building 197D
Upton, NY 11973-5000
Telephone: (631) 344-2804
dunford@bnl.gov
http://www.nndc.bnl.gov/index.jsp

Jannean Elliott
DOE Office of Scientific and Technical Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (865) 576-6784
elliottj@osti.gov
http://www.osti.gov

Christopher L. Greer
Program Director for Research Resources,
Division of Biological Infrastructure
National Science Foundation
BIO/DBI, Rm 615 N
4201 Wilson Boulevard
Arlington, VA 22230
Telephone: (703) 292-8470
cgreer@nsf.gov
http://www.nsf.gov

Jeffrey J.E. Hayes
NASA Headquarters, Code SZ
300 E St SW
Washington DC 20546-0001
Telephone: (202) 358-0353
jhayes@nasa.gov
http://www.hq.nasa.gov

Hamilton T. Hunter
Director, Radiation Safety Information Computational Center (RSICC)
P.O. Box 2008, MS6171
1 Bethel Valley Road
Oak Ridge, TN 37831-6171
Telephone: (865) 574-6297
hunterht@ornl.gov
http://www-rsicc.ornl.gov/rsicc.html

Sharon Jordan
Assistant Director
DOE Office of Scientific and Technical Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (865) 576-1194
jordans@osti.gov
http://www.osti.gov

Bernadette Kirk
Radiation Safety Information Computational Center (RSICC)
P.O. Box 2008, MS6171
1 Bethel Valley Road
Oak Ridge, TN 37831-6171
Telephone: (865) 574-6176
kirkbl@ornl.gov
http://www-rsicc.ornl.gov/rsicc.html

Predrag Krstic
Controlled Fusion Atomic Data Center
Physics Division, Bldg. 6010, MS-6372
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6372
Telephone: (865) 574-4701
krsticp@ornl.gov
http://www-cfadc.phy.ornl.gov

Raymond McCord
Data Archive Manager
Atmospheric Radiation Measurement (ARM) Program
Oak Ridge National Laboratory
PO Box 2008, MS6407
Oak Ridge TN 37831-6407
Telephone: (865) 574-7827
mccordra@ornl.gov
http://www.archive.arm.gov

James D. Myers
Pacific Northwest National Laboratory
P.O. Box 999/K1-85
Richland, WA 99352
Telephone: (610) 355-0994
jim.myers@pnl.gov
http://collaboratory.pnl.gov

RL Scott
Deputy Director
DOE Office of Scientific and Technical Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (865) 576-1193
scottrl@osti.gov
http://www.osti.gov

Larry Voorhees
DAAC Manager
Distributed Active Archive Center (DAAC) for Biogeochemical Dynamics
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6407
Telephone: (865) 574-7309
voorheesld@ornl.gov
http://www.daac.ornl.gov/

Walt Warnick
Director
DOE Office of Scientific and Technical Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (301) 903-7996
warnickw@osti.gov
http://www.osti.gov

Thurman Whitson
DOE Office of Scientific and Technical
Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (865) 576-1088
whitsont@osti.gov
http://www.osti.gov

David Wojick
Consultant
DOE Office of Scientific and Technical
Information
1 Science.gov Way
P.O. Box 62
Oak Ridge, TN 37831
Telephone: (865) 576-3390
wojickd@osti.gov
http://www.osti.gov