Data Centers for Ecology, Evolution, and Organismal Biology
December 8-9, 2006
Workshop Report

BACKGROUND

A Joint Working Group on Data Sharing and Archiving (JWG), representing major professional societies that publish ecology, evolution, and organismal biology journals, was formed at a September 2004, NSF-sponsored workshop on data sharing and archiving, hosted by the Ecological Society of America (ESA). Attendees adopted a consensus statement that "Our vision as members of the scientific community is to promote the advancement of science through the process of documenting, archiving, and making available the research information and supporting data of published studies." In support of that vision, the JWG made the following recommendations:

- Facilitate continuing communication among professional societies on data sharing and archiving issues via a dedicated web site and periodic e-mails;
- Widen participation in these activities by professional societies and international organizations; and
- Support three workshops to (1) develop a strategy for creating data registries that describe datasets and provide information on how to access them, (2) identify, and develop means to reduce or eliminate, cultural and other barriers to data sharing, and (3) develop a set of requirements and recommendations for data centers in ecology, evolution, and organismal biology.

The first of these three workshops, "Data Registries for Ecology, Evolution, and Organismal Biology," was held July 11-12, 2006, in Washington, DC. Twenty-five participants representing sixteen professional societies and nine other organizations assembled to work toward three goals:

- Identify a set of common needs for, and desirable features of, data registries for ecology, evolutionary biology, and organismal biology, based on an understanding of existing resources.
- Develop recommendations, as appropriate, for shared or independent data registries for the disciplines and societies represented.
- Develop preliminary plans for implementing those recommendations.

The second workshop, "Data Centers for Ecology, Evolution, and Organismal Biology," was held December 8-9, 2006, in Santa Barbara, CA. Thirty-two participants representing fourteen professional societies and eleven other organizations assembled to work toward three goals:

- Identify gaps between existing data centers and needs, including specific issues such as quality assurance procedures needed for contributions to centers, types of data that should be archived, etc.
- Identify roles of professional societies, funders of research, and users of research in developing – or encouraging the development of – data centers, along with where data centers should be housed and who should operate and maintain them.

- Assess likely cost to establish and maintain data centers required to meet community needs, including identification of potential funding mechanisms and models for data centers.

INTRODUCTION AND WELCOME

Cliff Duke (ESA) welcomed workshop participants and provided background information on the Data Sharing Initiative. The goals for the Data Centers meeting were to: identify needs for and desirable features of data centers; develop recommendations for funding, locating, operating, and maintaining data centers; and adopt a consensus statement on those needs.

Jim Reichman (National Center for Ecological Analysis and Synthesis - NCEAS) also welcomed the participants and provided background information on NCEAS (the workshop's local host).

INFORMATIONAL PRESENTATIONS

Matt Jones (NCEAS) provided an introduction to data centers. The purpose of a Data Center is to facilitate science by supporting novel synthesis studies, cross-disciplinary studies, and evaluation of traditional studies. Jones provided a number of project examples of cross-discipline synthesis and re-use of data. These studies represented massive investments to compile, integrate, and analyze existing data. Building custom databases for each of these diversified projects is not logistically feasible. Instead, loosely-coupled systems that accommodate heterogeneity are needed.

What is a data center? A data center is a virtual organization that provides infrastructure supporting a collection of data and documentation. It enables preservation of data, enables discovery of data by searching (e.g. space, time, taxa, keywords), enables evaluation of data through examination (title, abstract, methods, etc), and enables access to data.

Jones highlighted some critical issues that participants were to keep in mind during the workshop. These included permanence, data dispersion, management and governance, and architecture. Jones noted that many existing centers developed for particular purposes or groups are not necessarily available for contributions from other scientists. Furthermore, the cross-disciplinary nature of ecology, evolution, and organismal biology requires that data centers have the ability to cross domains. Thus, unique identifiers for each dataset are necessary to prevent multiple representations, along with the ability to update archived datasets.

Participants noted the tremendous need for educating people on the use of data centers and what they could provide. The question of whether such education should be part of the role of a data center was posed to the group for consideration.

Chris Greer (National Science Foundation) discussed the role of universities and libraries in data preservation and noted that the key issue in thinking about data as infrastructure is sustainability - both technical and economic sustainability. The technical framework should have a decadal horizon. Economic sustainability requires partnerships among universities, colleges, non-profit organizations, commercial and international researchers, and local, state, and federal agencies, all representing diverse business models and funding sources.

Universities and libraries should be involved in preserving digital data, which coincides with their current roles. Universities are charged with creating, disseminating and preserving knowledge. Academic libraries currently preserve “traditional” research products and thus have special capabilities, together with the long-term horizon necessary for true preservation of digital data.

NSF’s role is to catalyze and support development of a system and tools for data acquisition, mining, integration, analysis, and visualization. This role is highlighted in a draft NSF cyberinfrastructure vision document.

The challenges are the need for full commitment from society (technically, legally, economically, and organizationally) and a cultural change among individual scientists.

Participants noted that it would be extremely powerful if all funding agencies had a requirement similar to NSF’s that encouraged, and ultimately required, data sharing. It was also noted that the funding for data archiving is an important consideration for funders encouraging/mandating data sharing.

Concern was voiced regarding the poor communication among the various sectors and the difficulty in reaching a point at which all of these diverse sectors communicate in a way that allows the user to gain access to data.

Peter Joftis spoke about the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan, which has been archiving data for 45 years. ICPSR is a membership-based organization that maintains and provides access to a vast archive of social science data for research and instruction and offers training in quantitative methods to facilitate effective data use. ICPSR preserves data, migrating them to new storage media as changes in technology warrant. ICPSR also provides user support to assist researchers in identifying relevant data for analysis and in conducting their research projects.

The mission of ICPSR is to work with its member institutions to: (1) acquire and preserve social science data, (2) provide open and equitable access to these data, and (3) promote effective data use.

Joftis noted that researchers are not good at providing documentation and metadata, and that we need to build a culture that values archiving efforts and rewards them.

Advantages of ICPSR membership include:

- Free access to all data holdings. Nonmembers are restricted to a subset of holdings paid for by grants/contracts that specify public access (Special Topic Archives).
- Reduced fees and first access to the courses taught by ICPSR’s Statistical Training Program.
- Voting on membership of the governing council and other issues specified in the constitution and bylaws.

Subscribing members help sustain the effort financially and assist in seeking grants. Other funding comes from grants and contracts, which provide resources to pursue special projects and to make those data available. The current funding balance has tipped towards grants and contracts; approximately 65% of ICPSR’s revenue is derived from grants and contracts versus approximately 25% from members. ICPSR has a tiered fee structure based on the institution's level of research.

Joftis also mentioned the data documentation initiative, which is an effort to establish a metadata standard and to move beyond machine-readable to machine-actionable data.

Robert Cook discussed the Oak Ridge National Laboratory's (ORNL) Distributed Active Archive Center (DAAC). ORNL DAAC is one of 8 DAACs archiving Earth Observing System (EOS) data generated by NASA. The ORNL DAAC focuses on biogeochemical dynamics and terrestrial ecology data. It incorporates multiple methods for data discovery and access, including: FTP, Clearinghouse/Registry; WebGIS, Google Earth (metadata mashup), and tools for subsetting and visualization of remote sensing products. The data are available at no cost. The ORNL DAAC requires about 12 full-time equivalent personnel (FTEs) to operate.

Cook highlighted the need to archive enough metadata to find, understand, and use the data, along with needing a convenient, well defined format. ORNL DAAC requests citation of its datasets when used in publications, thereby enabling readers to find the data files for possible future use.

Participants raised the issue of preserving models and whether the goal is to be able to re-run a model and what needs to be preserved to do that. Cook responded that they do need to be sure they understand why a model result is what it is. But there is a concern over how one can run a model if the software on which it was originally run is not preserved.

Roman Olschanowsky spoke about the San Diego Supercomputer Center (SDSC), which is one of nine Teragrid sites, specializing in high-end computing as well as data, data service, and knowledge. SDSC has a lot of experience with archiving and processing large amounts of data and has recently been asked by NSF to lead the way with regard to "Storage Allocations." In 2005, SDSC announced formal deployment of Data Central, a coordinated set of data-oriented services, resources, and capabilities to support the scientific community, at http://datacentral.sdsc.edu.

William Piel (Yale University) presented on TreeBase (www.treebase.org), which stores data from papers published in peer-reviewed journals and books. TreeBase accepts data using a popular format, and data in other formats can be converted into the TreeBase format. Currently 2,771 authors and 1,489 studies are represented in TreeBase. By 2000, about 12,000 taxa were represented in TreeBase; now there are almost 70,000 taxa. Mycologists' contributions to TreeBase have increased faster than those of researchers studying other taxa, perhaps because the notion of archiving and sharing data was accepted by mycologists earlier on.

One of the challenges for TreeBase is the reconciliation of semantic heterogeneity. Phyloinformatics requires names for "things," such as taxon labels or names of genes. This presents problems in that one name may be used for two different genes in two different species.

Human nature poses another series of challenges. First, to motivate researchers to share data, one must make them an offer they cannot refuse. Second, some researchers may want to make "blob" data deposits, but these do not work well. For example, when one submits a zipped group of text files, it can be very difficult to make much sense of such data. There is no quality control. Third, even when the data repository provides a means of checking for errors, some data providers will try to circumvent the error-checking features. TreeBase is attempting to normalize data by encouraging data providers to spell taxon names correctly, format their data correctly, and check for errors.

Piel noted that a fair bit of thinking must go into design of a data center or repository. He suggested requiring more work from authors/data providers to encourage them to be more conscientious when entering their data.

Robert Peet (University of North Carolina at Chapel Hill) spoke about VegBank and ESA's Vegetation Classification Panel and its place in the greater information infrastructure. The Panel has developed guidelines for vegetation classification covering requirements for vegetation field plots, documentation and description of floristic types, submission and peer review of proposed types, and archiving of supporting vegetation data.

VegBank is a public archive for vegetation plot observations (http://vegbank.org) that allows vegetation plot data to be archived, searched, viewed, cited, annotated, and downloaded. VegBank supports the vegetation classification enterprise, but can be used for any research involving species co-occurrence. As such it is expected to play a central role in the developing field of ecoinformatics. The VegBank strategy for community participation recognizes the need for usability, incentives, and connectivity. VegBank's strategy for data sharing includes embracing established standards and establishing new standards as needed, designing for idiosyncratic data, planning for long-term use and preservation of data, respect of intellectual property (allow for embargoes, licenses, and confidentiality), and leveraging agency mandates.

Peet noted the particular challenge of scientific names of organisms. One taxon may have multiple names depending on which authority/reference one uses, and one name can refer to multiple taxon concepts, particularly as taxa are lumped and split with the original name always tracking the type specimen rather than a particular group of organisms. This makes integration of data collected by multiple observers at multiple times and places using different taxonomic authorities extremely challenging. VegBank deals with this by tracking not simply taxon names but taxon concepts (name-reference couplets) and embedding relationships between taxonomic concepts so that the user can potentially integrate data based on different but related taxon concepts.

Peet summarized more broadly the likely future distribution of roles and responsibilities in data preservation and distribution. In his view institutional repositories and libraries will provide data preservation and security at the level of the individual investigator with a significant mandate provided by granting agencies with funding provided through overhead. Granting agencies will set requirements for archiving and sharing of data, and will pay for archiving and publication either directly or indirectly through overhead. Data centers will maintain portfolios of critical, discipline-specific databases along with key infrastructure including digital identifiers, data registries, and compendia of common objects (taxa, publication lists, etc.). Publishers and professional societies will require that certain types of data be archived in standard repositories (e.g., gene sequences, vegetation plots), will embed links to data underlying published research, and will provide archives for and links to supporting documentation. Government agencies will develop and adopt federal standards, mandate their use, and assure existence of critical components of the overall information infrastructure.

William Michener (University of New Mexico) discussed the Long-Term Ecological Research (LTER) Network and the National Ecological Observatory Network (NEON). LTER site data come from heterogeneous data sources including terrestrial, aquatic, coastal, and open ocean ecosystems. LTER has a data-sharing policy requiring all data to be publicly available on each site's website. Data include archived site data, a remote-sensing archive, sensor data, sentinel data, imagery, and Network-level databases. LTER provides processing tools and access to both raw data and processed/analyzed data.

NEON data will include sensor array and satellite data. NEON will present challenges for environmental informatics. The planned sensor arrays are likely to generate 50 terabytes (TB) of data per year, including raw data as well as many products. Remote sensing and other data will probably be around 100 TB per year.

Michener discussed raw data versus various levels of derived data products. He suggested that the tendency now is to focus attention on the preservation of raw data, but predicted that in the future, raw data will be a very small part of what is archived and used; the derived products will be used more. But one of the conundrums we will face is that when funding gets tight, it will be the more highly derived products that drop out. This will then require scientists, students, and educators to expend more effort trying to deal with raw data.

Michener also raised the question of when one should throw away data. It is simply not practical to save every bit forever but this question has not been answered.

Matt Jones addressed NCEAS' approach to data centers, the main principles being: control of data by contributors (e.g. access restrictions); the use of open, nonproprietary systems; geographic replication, which is critical to preservation (multiple storage locations); clear separation of system components (well-defined interfaces, well-defined interchange standards, and support of multiple access methods); and easy, cost-effective ways to change components over time as desires and goals change. NCEAS stores raw data as well as derived data products (models and analyses). The Knowledge Network for Biocomplexity (KNB) and Science Environment for Ecological Knowledge (SEEK) projects were discussed.

Hilmar Lapp described the efforts of the National Evolutionary Synthesis Center (NESCent) to create the cyberinfrastructure for sharing data in evolution. NESCent, funded by NSF, launched the Digital Repository for Information and Data in Evolution (DRIADE), a collaboration with the Metadata Research Center (MRC) at the University of North Carolina. DRIADE is a specialized data repository that will allow for data sharing, discovery, preservation, and synthesis. Three workshops are being planned to determine requirements, priorities, and challenges, and to develop an implementation plan and funding strategy.

Most research in evolution is "small science." Single researchers are manually acquiring data, with little to no automated data collection. The datasets are typically small, but highly heterogeneous, and data and metadata often are only logged in field notebooks. The human element in data deposition is key (data stewardship is highly dependent on individual researchers). A cultural change driven by societies and journals is necessary.

BREAKOUT DISCUSSIONS

Participants in the "**Needs for Data Centers**" breakout session, facilitated by Nancy Grimm (Arizona State University) and Chip Leslie (American Society of Mammalogists), were tasked with identifying gaps between existing data centers and needs, and discussing specific issues such as quality assurance procedures needed for contributions to centers and types of data that should be archived. Participants included: Bruce Dancik, Charles Fox, Peter Joftis, Bill Michener, Tom Moritz, Jim Reichman, Mark Schildhauer, and Mindy Destro.

Issues discussed included:

- Current data centers are typically:
    a) supported by funders and restricted in subject matter, or
    b) supported by institutions and restricted to affiliated researchers, or
    c) discipline-based.
- Small, individual projects and researchers are in most need of data centers.
- Challenges include:
    a) heterogeneity of the data
    b) varied project types
    c) usability for researchers to archive data, as well as discover and use archived data
    d) lack of necessary metadata, limiting use by other researchers
    e) new tool development
    f) use (incentives need to be developed for researchers to archive data)
- Top data archiving priorities include:
    a) legacy data not yet in digital form
    b) at-risk data (data on researchers' hard drives, data of retiring researchers)
    c) published data
    d) data essential to answering really important questions (determined by the ecological community)
- Quality assurance should be the responsibility of the submitters, not the data centers.
- A common vocabulary (ontology) is necessary.
- Training is crucial to ensure quality of data submitted.
- Societal needs are the main driver behind data sharing – the need to collaborate, synthesize, and integrate for larger-scale, longer-term, more robust science.
- Interconnection among data centers is crucial for cross-discipline research.

A reoccurring issue of discussion concerned the types (levels) of data to be archived. Raw sensor data is just a bunch of numbers, which have no meaning until they are combined with additional information (descriptors). This level of data is then aggregated, outliers might be removed, averages might be used to represent subsets of data, etc. At what level of abstraction and analysis should archiving occur – raw numbers, data with descriptors added, further transformed data, or multiple levels?

A National Data Center infrastructure would include:

- free and open access (with some exceptions, e.g, remote-sensing data)
- open source and model-sharing/community model development capabilities (e.g., Kepler work flows can be used to share, build, and replicate analysis)
- a directory of connected data centers
- a how-to manual (online)
- training

One immediate need expressed by the group is a directory of data centers that currently exist.

A white paper on topics discussed is being developed by the Data Center Needs breakout group members for possible publication.

The “**Development of Data Centers**” breakout group, facilitated by Peter McCartney (NSF) and David Baldwin (ESA), was tasked with identifying the roles of professional societies, funders of research, and users of research in developing – or encouraging the development of – data centers, and to consider where data centers should be housed and who should operate and maintain them. Participants included: Shan Duncan, Erica Fleishman, Paul Kemp, Susan Mazer, Allen Moore, John Pearse, Dennis Stevenson, Callie Bowdish, Hilmar Lapp, and Cliff Duke.

Issues discussed included:

- The difficulty of separating the overall cultural issues related to data sharing (e.g., getting people to openly share data) from the technological issues (e.g., providing data storage facilities and other resources needed to facilitate data sharing).
- How to manage different scales of data archiving, considering differences in dataset size, complexity, and degree of derivation (i.e., raw data vs. variables derived from raw data).
- The need to establish the credibility of specific data centers as reliable repositories for and sources of data.
- The appropriate degrees of connectedness to establish among data centers of different kinds (e.g., genetic versus ecological) and the value of relatively centralized versus dispersed data centers.
- How to provide incentives or rewards for researchers to contribute data to repositories.
- General desirable features of data centers.

The group responded to its charge by developing a 24-month implementation plan for beginning development of data centers, with roles for various stakeholders identified as follows:

<u>Societies</u>

1) publish journal editorials supporting data center development and data sharing generally
2) issue shared statements on the topic, signed by multiple societies
3) conduct outreach to members
4) develop publication incentives to encourage data sharing and deposition in public repositories
5) engage in discussions with publishers about citation conventions for datasets
6) develop presentations/workshops for their annual meetings
7) issue press releases about activities relating to data sharing

<u>Participating organizations</u>

1) support users with systems for data deposition and sharing
2) ultimately provide mirror systems
3) participate in an identity management system to uniquely identify datasets

<u>Funding agencies</u>

1) encourage cross-directorate/cross-discipline discussions of the lowest (least complex in terms of metadata requirements) level data file deposition that should be accepted
2) emphasize the importance of archived data for nonfunded studies

<u>Developers</u>

1) try to implement data deposition with off-the shelf tools as much as possible

2) make data deposition and sharing as painless for users as possible

Users

1) try the tools that are made available and provide feedback to developers, funders, professional societies, and participating organizations

The group also identified a consensus set of desired features for data centers:

- Close interweave with users' existing file management practices
  a) shared storage folders mountable as user file systems with minimal software overhead
  b) an easy to use or scripted checkin/checkout system
- Metadata requirements as simple as possible
  a) minimal Dublin Core
  b) richer metadata is optional
- Author control – write access, read access could even be limited
  a) all data registered is public
  b) read access could be variable
  c) write access is variable
  d) global identity management
- Associated collaboration tools available through the data center
  a) user profiles, groupware features
- Replication (full/selective) to provide backup/availability
  a) distributes a commitment to sustainability
- Supports basic versioning of datasets so that authors can revise them as appropriate while letting data users know which version they are working with

The group also discussed how membership in data centers might work. It seems clear that researchers and/or their organizations would be eligible for membership. Issues that would need further discussion involve the roles of the international community, for-profit organizations, educational organizations, and other disciplines outside the primary ones involved in the creation of a given data center.

It was also noted that data centers would be more attractive to researchers if the researchers could archive the data for their own use prior to making the data public. Data input takes time, but if researchers were already storing data at a data center, then they would be more inclined to "hit a button" to allow public access rather than spend the time uploading data that is already stored elsewhere.

Participants in the "**Operation and Maintenance of Data Centers**" breakout session, facilitated by Matt Jones (NCEAS) and Robert Cook (ORNL), were tasked with assessing the likely cost to establish and maintain data centers required to meet community needs, and identifying potential funding mechanisms and models for data centers, including the question of whether user fees should be charged and how this might impact usefulness of data centers. The team decided, however, that a better handle on the kinds of functions required of a data center was necessary before a cost model could be produced. Participants included: Jonathan Duncan, Chris Greer, Roman Olschanowsky, Annette Olson, Bob Peet, Bill Piel, Hannu Saarenmaa, and Bette Stallman.

Ideas discussed included:

- The importance of longevity of data centers and the necessity of a long-term perspective. Libraries are the most likely candidates (or models) for heading up data preservation, due to their overall mission of preserving information and making it accessible. Their mission should be shifted to digital information. Libraries are risk-averse, but to be relevant and innovative, they will need to be willing to take some risk or to partner with less risk-averse entities.
- Conversion of libraries and library science is already underway. The framework will include library, computer, and domain science; cyberinfrastructure; and archiving.
- Individual repositories may increase their capacities for sustaining data collections by joining with other repositories in a network. It may be beneficial for individual repositories to specialize in a particular kind of data or research, but each repository should be capable of housing the entire spectrum of collection types, as defined below.
- The spectrum of collections includes research, resource, and reference collections. The different collection types can be difficult to define but, generally, they are as follows:
    1. A research collection is from a specific researcher or research team and may be used only or primarily by that team. Research collections are relatively heterogeneous and idiosyncratic and are broader in scope than resource collections.
    2. A resource collection is a community-level collection such as a museum collection. A resource collection is more standardized and international and typically more focused in scope than a research collection. It is used by a community of researchers.
    3. A reference collection is integrated, well formatted, standardized, has lots of value added, many tools, its own line of funding, and offers very long-term preservation. A reference collection is used by everyone, not just one type/community of researcher.
- Each type of collection is important. For example, while many research collections will never be used again, some will prove to be crucial; groups at NCEAS have found small, idiosyncratic research collections to be crucial to their work. Reference collections need to be considered in defining standards and providing mechanisms for discovery.
- Who can pay? Existing models include user fees, grants, contracts, institutional user fees, and government funding. Future models will need to include a diversity of funding sources from multiple components of society, including academic institutions, federal and state agencies, nonprofit organizations, for-profit companies, etc.
- What will the cost be? One problem in estimating costs is that costs could change quickly. Factors contributing to overall costs include: (1) storage costs; (2) hardware and software migration; (3) the level of quality assurance provided, both of the science (reliability) and of the systems (accuracy, problems such as bit flips); and (4) replication/backup.
- Existing models can provide a rough idea of potential costs. The range of costs could be less than (or much less than) $1 million per year for a research collection, $1 – 10 million per year for a resource-level collection, and $10 million or more for a reference collection. It will be useful to do a more careful analysis of the costs of existing data centers, including the handful of supercomputer centers like SDSC and other models such as the DAACs to determine the range of costs and to better estimate the costs of the data centers we wish to establish.
- What to preserve? To attempt to digitize legacy data would be an immense challenge in terms of time and cost. Legacy data can be preserved in various ways, not necessarily

by digitizing. What may be at greatest risk (and most worth the trouble to digitize) are the data from individual investigators from the 1960s on that are currently only stored digitally with software and/or media that are no longer usable.

- How long to preserve data? There is controversy over how long research collections should be preserved. Some believe we should keep everything, considering the potential for unexpected usefulness of data (e.g. climatological data). But there are practical issues, such as forward migration of formats, storage space, and funding. For some data it might cost less to re-do the experiment (if possible) than to keep the data.

DISCUSSION OF NEXT STEPS AND ELEMENTS OF WORKSHOP REPORT

Next Steps:

- ESA staff will prepare a draft Workshop Report, circulate it to participants for input, and then distribute a final version to be shared with society leaders.
- ESA will make the workshop report available to NSF as part of its responsibilities under the workshop grant and will explore ways to highlight the specific recommendations concerning the roles of funding agencies.
- Participants are invited to prepare editorials based on the workshop recommendations for publication in their journals, and to share them with other participants as templates for their own editorials or other written reports.
- ESA will circulate the ESA Governing Board statement on data centers as a model for other societies.
- The Data Centers Needs breakout group will develop a white paper and will circulate it to participants for input and will explore possible publication.

ESA would like feedback on what other big topics are amenable to this workshop format that might be missing from the process.

NESCent is considering hosting the 3rd Workshop in Durham, NC in late Spring 2007. The main goal of the 3rd Workshop is to discuss major cultural obstacles to data sharing and what could possibly be done to resolve them. ESA will be requesting input from participants regarding date availability, the agenda, and specific discussion items.

ESA staff also request that representatives of societies and other organizations that have made tangible progress on data sharing issues (e.g., establishing registries, publishing editorials, writing data sharing into editorial policies) share that information with the group. ESA will distribute such information and add it to the data sharing initiative page on ESA's website.

ESA Data Centers Workshop, December 8 - 9, 2006
National Center for Ecological Analysis and Synthesis, Santa Barbara, CA

Attendee	Affiliation
David Baldwin	Ecological Society of America
Callie Bowdish	National Center for Ecological Analysis and Synthesis
Bob Cook	Oak Ridge National Laboratory, Distributed Active Archive Center
Bruce Dancik	National Research Council Canada
Mindy Destro	Ecological Society of America
Cliff Duke	Ecological Society of America
Jonathan Duncan	Consortium of Universities for the Advancement of Hydrologic Science
Shan Duncan	Animal Behavior Society
Erica Fleishman	Society for Conservation Biology
Charles Fox	British Ecological Society and University of Kentucky
Chris Greer	National Science Foundation, Office of Cyberinfrastructure
Nancy Grimm	Ecological Society of America
Peter Joftis	University of Michigan, Inter-university Consortium for Political and Social Research
Matt Jones	National Center for Ecological Analysis and Synthesis
Paul Kemp	American Society of Limnology and Oceanography
Hilmar Lapp	National Evolutionary Synthesis Center
Chip Leslie	American Society of Mammalogists
Susan Mazer	Society for the Study of Evolution
Peter McCartney	National Science Foundation
Bill Michener	Long Term Ecological Research Network
Allen J. Moore	European Society for Evolutionary Biology
Tom Moritz	Getty Research Institute
Roman Olschanowsky	San Diego Supercomputer Center
Annette Olson	National Biological Information Infrastructure
John Pearse	Society for Integrative and Comparative Biology
Bob Peet	University of North Carolina
Bill Piel	Yale University
Jim Reichman	National Center for Ecological Analysis and Synthesis
Hannu Saarenmaa	Global Biodiversity Information Facility
Mark Schildhauer	National Center for Ecological Analysis and Synthesis
Bette Stallman	Ecological Society of America
Dennis Stevenson	Willi Hennig Society