

**Obstacles to Data Sharing in
Ecology, Evolution, and Organismal Biology
May 30-31, 2007
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 16 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 14 professional societies and 11 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.

The third workshop, “Obstacles to Data Sharing in Ecology, Evolution, and Organismal Biology,” was held May 30-31, 2007, in Durham, NC. Thirty-nine participants representing nine professional societies and 22 other organizations assembled to work toward two goals:

- Clearly delineate what barriers exist to data sharing, for example, intellectual property concerns, proprietary and confidential business information, handling of sensitive data such as locations of

endangered species, lack of training in data sharing software, national or economic security concerns, etc.

- Develop recommendations to reduce or eliminate those barriers, for example, publication policies that encourage or require data sharing, means of providing at least limited access to business or sensitive data, and development of easily accessible training programs.

This report summarizes the discussions and recommendations from the third workshop. Reports on the first two workshops are available at http://www.esa.org/science_resources/datasharing.php.

INTRODUCTION AND WELCOME

Cliff Duke (ESA) welcomed workshop participants and provided background information on the Data Sharing Initiative.

Breakout Topics:

- 1) Commercial/proprietary obstacles – including intellectual property, confidential business data
- 2) Sensitive data – including privacy issues, endangered species data, security issues
- 3) Socio-cultural barriers – including attitudes, possessiveness, fear of being “scooped”, training students

Kathleen Smith (NESCent) welcomed workshop participants to the National Evolutionary Synthesis Center (NESCent), which is a collaborative effort of Duke University, The University of North Carolina at Chapel Hill and North Carolina State University and is sponsored by the National Science Foundation. NESCent’s mission is to facilitate broadly synthetic research to address fundamental questions in evolutionary biology. Activities include involvement in a major data sharing and registry initiative, a Visiting Scientist Program, an Informatics program, and education and outreach.

Sam Scheiner (NSF) thanked participants for attending and noted that the complexity of the issues to be discussed was shown by the diversity of the attendees, which is different from previous workshops. The partnership between NSF and the societies will help push the science community into a different mindset on data sharing. New and interesting things will result by having data shared and available.

INFORMATIONAL PRESENTATIONS

Paul Uhlir (NAS) and **Jerome Reichman** (Duke) – Intellectual Property Rights as Barriers and Incentives to Data Sharing

Paul Uhlir provided a background on intellectual property rights and a comparison of the various policies. Open availability of publicly-funded scientific data and information online promotes interdisciplinary, inter-institutional, and international research. Open access enables automated knowledge discovery and generally helps to maximize the research potential of new digital technologies and information, providing greater returns from public investments in research. Open access also helps avoid duplication of research, and it promotes new research, new types of research, and testing of new or alternative hypotheses and methods of analysis.

Excessive restrictions on access to and reuse of data and information from public research result in a disadvantage and marginalization of developing country researchers and users least able to pay; significant lost opportunity costs; monopolization problems exacerbated in scientific database markets, both public and private; higher transaction costs (not just cost of access); and less effective international, inter-institutional, and interdisciplinary cooperation.

Barriers to data sharing resulting from exertion of intellectual property rights include: privatization of government research and data collection activities; commercialization of academic research and of government data and information products and services; strategically competitive behavior by

governments, research institutions, and individual researchers; legal tools such as copyright, restrictive licenses, and statutory database protection; and other restrictive laws (e.g., protection of national security, privacy, endangered species).

Jerome Reichman provided a background on the restrictions of intellectual property laws, which are being ratcheted up at the international and national levels. The European Union (EU) now has a database protection law that applies in 50 countries, and the EU is trying to foist the law on 75 more smaller countries. The European Court of Justice, however, has cut back on the law in contorted interpretation. There is uncertainty on where scientific data falls on the continuum. Concerns over and calls for review of the EU database directive have been made, but the European Commission (executive branch of the EU) and special interest lobbies continue to push for it. Similar legislation has been blocked here in the U.S., though U.S. appellate courts are starting to have a protectionist trend towards copyright law, which protects creative selection and arrangement.

Intellectual property rights can provide incentives for work on value-added data products, private-sector provision of data products and services, and public-private partnerships.

Science has to undertake management of its own data resources to ensure desired access. Alternative mechanisms for socially beneficial mediation of rights include common-use licensing and “liability rules.” It is also desirable for universities to have treaties among themselves and for individual scientists to be able to publish data under standard licenses so they do not have to keep seeking licenses every time they publish. Peer to peer software will allow researchers to share information. There needs to be a 2-tier regime where everyone shares (horizontally) and commercial applications must pay some royalty to the data depositor (vertically).

Hal Balbach (US Army Corps of Engineers Engineer Research and Development Center) discussed issues surrounding the sharing of sensitive data using a threatened species, the gopher tortoise, as a specific example. The Department of Defense needed to know the status of the tortoise in parts of its range that are used for military training exercises. Thirteen different agencies formally agreed to protect the species and improve its management, but data sharing was required to understand what was already known about the species across its range. There was some existing geospatial data, but data sets varied in quality and extent. The gopher tortoise is a success story for data sharing, and it also forced the agencies involved to consider what kinds of guidelines or rules should govern the sharing of sensitive data.

Whether collected on private or public land, those agencies involved in sharing the data share some common concerns. There is concern about who is allowed to access certain data, especially geospatial data that could be used to harm people or property. Certain kinds of data (water sources, hazardous material storage) are more easily identified as public safety concerns. Others (rare plant and animal locations, locations of economically valuable biological resources such as timber) are ecologically sensitive. Shared data may provide an opening for legal action in some cases. For these reasons, it seems that there must be limits on who accesses some kinds of data, but it is uncertain who should establish the limits and whether they should be based in law or in regulations. Criteria must be developed by the scientific community and it must be decided how they will be implemented in practice, and by whom.

Juan Núñez Farfán (Ecological Society of Mexico) provided an international perspective on data sharing. Farfán noted that barriers are similar to those in other countries, but there are different cultural characteristics for Latin American countries to consider. These characteristics can be extrinsic to research or data producers, such as: 1) a low capability to carry out scientific ecological studies (small scientific communities and underdeveloped scientific societies); 2) the economic resources for developing a center for data sharing are not feasible for a single institution; and 3) there are few scientific journals in these disciplines that can recommend that authors openly provide data.

The intrinsic barriers include: 1) a lack of incentives—many researchers do not appreciate why data sharing is important and what the reward is for them, and they are also concerned about their competence and are afraid new analyses would contradict their conclusions; 2) competition and possessiveness; and 3) cultural barriers, including fear of exposing endangered natural resources to exploitation.

Standardization of information is an important problem. There are differences in nomenclature and language among countries and disciplines, and within disciplines that prevent comparisons. Translated versions are needed. Ornithologists in Mexico have made some good advances in standardization.

Use of databases in Mexico generally is limited to those involved in the collection of data or its production; however, during the last few years there has been an increased attention to the sharing of raw data among research groups.

One of the most representative examples of data sharing in Mexico is through the National Commission for the Knowledge and Use of Biodiversity (CONABIO), which developed and maintains the National System of Biodiversity Information (SNIB). SNIB is in charge of compiling and synthesizing information on the biodiversity and biological resources of Mexico. Projects funded by CONABIO are required to provide open access to data. Farfán hopes to work with CONABIO to get more scientists to publish data.

Kristin Brugger (DuPont Crop Protection Products) provided an overview of corporate obstacles to sharing data. The main obstacles were concerns about data protection, uncertainty about business costs versus benefit of sharing, and legal requirements. Companies that collect data on safety and efficacy of products in agriculture and medicine are concerned with protecting data from unfair commercial use and undesired disclosure of complete study results (when summary results are usually what is released). Legal protection of data varies among countries, and in the United States, while a patent excludes others from certain uses of an invention, there is not a similar protection for data.

Because there can be considerable cost associated with acquiring, entering, and recoding data (if sensitive), companies are unlikely to perceive data sharing to be in their best financial interests. Confidential business data and trade secrets are two categories of sensitive data with existing regulatory definitions, but companies may be reluctant to share other types of less sensitive data if they think it could be improperly used by others. For data sharing to become advantageous for corporations, there will need to be clear rules about ownership, intellectual property, and what uses of data are considered lawful. Other legal concerns for corporations include vulnerability to lawsuits and preserving the safety of people and organizations conducting some kinds of research (i.e., withholding author contact information and laboratory location).

Practical obstacles to data sharing include finding a person within a corporation who has the knowledge, resources, and ability to work toward a long-term goal of making data available for others to use. Because future liabilities are uncertain, clarifying expected future uses of data might be helpful. It will be especially important to address the concerns of corporations about the possibility that the data they have invested in could be passed off as a product of others' work. It will also be important to consider that data sharing by global corporations will have global implications and various obstacles that vary among different legal jurisdictions.

Ahrash Bissell (Open Context/Alexandria Institute) discussed incentives for researchers to share data, particularly for "field" sciences.

The Alexandria Archive Institute is dedicated to the open dissemination of cultural-heritage and scientific datasets (focusing on archaeology and other "field" sciences). These are highly diverse datasets that contain bio/cultural heritage information, which is complex and highly variable, and raw data that are often large in quantity. The tricky problem is that scholars demand open access but resist providing it.

Incentives to hoard data include little reward to share, fear of rivals using the data, and pressures to publish synthetic articles.

Why share? Bumpus (1898) House Sparrow Data is a good example. Cary Bumpus published all raw data along with his syntheses. This information has been reused extensively and resulted in approximately 10 subsequent groundbreaking papers by others.

Sharing data reduces duplicative effort and prevents loss of data due to neglect. It is estimated that less than 10% of all data collected by academic researchers ever gets exposed in publications. In addition, the loss of data collected by students and researchers outside of formal academic sphere is staggering. Data sharing also levels the academic playing field (young investigators with limited contacts (and clout) can still answer bigger, synthetic questions).

One problem is the issue of copyright. Copyright law typically differentiates between “facts” and “expressions.” Some solutions include:

- “Do not Compete” term: to create incentives to share (specify boundaries; require listing data author as co-author with any peer review use; helps build CVs; encourage collaboration, theoretical insights)
- “Viral” contractual agreements for some rights reserved data

Mandates to share data can come from publication agreements that include data deposition and funding agreements. Public and philanthropic funding enables research – researchers are acting in the public trust with those funds. Most scientists would like to see their data achieve its full potential.

Katja Seltmann and Greg Riccardi (Morphbank) discussed specific obstacles that they have identified for sharing of morphological images. Morphbank users are not representative of what was expected *a priori*. There is a strong community of curators using the system, and they can strongly influence other users. Primary taxonomists, however, are very important for Morphbank’s success, and there is a bottleneck of contributions from this group. In addition there are cultural differences between people studying different taxonomic groups, and this results in different attitudes toward data sharing. Morphology for the Masses is a group that has funding and a strong desire to disseminate morphological information. The goal of Morphbank is to create a flexible system that meets user needs. Common issues include crediting users, meeting needs of different users who need different levels of user control, and form psychology (avoiding so much input data that people feel overwhelmed). In general, attractors (big results with little effort) are expected to overcome user hesitations and social barriers. Users experience a learning curve, and some standards in combination with a general environment encouraging user input and flexibility help users learn how the system works.

Mike Whitlock (The American Naturalist) discussed his “editor’s perspective” on data sharing and that a main focus at this time should be on preserving/archiving the data before they are lost. The Evolution Editors Consortium developed a common archiving policy. The draft policy states that data should be archived in an appropriate public archive with sufficient details for published results to be re-created. The American Naturalist Editorial Board was surveyed regarding the policy and a majority was in favor of it. Also asked about the need for an embargo period, 11 of the board members said none was needed, 12 felt that a one-year embargo would be fine. Other responses included two years, five years, and until the authors can publish everything.

Obstacles to archiving include:

- Cultural inertia (Evolutionary journals are working to require sharing to overcome this obstacle. Once critical mass is achieved, others will follow);
- Ownership and intellectual property rights (possible solutions: in many cases, authors do not own their data; publication denotes credit; citation of original paper; archiving only required for data used in paper; possible one-year embargo with extensions at discretion of editors);

- Time and effort for archiving (Archiving should put minimal demands on authors, perhaps using a 15 minute rule for the time it takes to input data; preparation prior to data collection will improve archiving ability);
- Overambitious archiving schemes (Don't ask for too much!).

Tom Moritz (Getty Institute) provided an overview of the Science Commons' Scholar's Copyright Project. The two key issues that will affect how data sharing is addressed by the scientific community are data access and data use. The optimal use of digital objects such as data depends on heritability. A clear definition of the rights a user has to access, render, and transpose data will be necessary to lower some of the barriers to data sharing. Because Creative Commons licenses are based in copyright, there are classes of digital objects for which Creative Commons licensing is undesirable. A similar system of author addendums for digital data could be used to allow authors to retain certain key rights and to specify how their data can and cannot be used by others. Data sharing will require some sort of similar standard so that it is clear which uses of shared data (such as educational use) are allowed and which are not (such as taking credit for someone else's data).

The Scholar's Copyright Project lowers barriers to open access by reducing transaction costs and eliminating contract proliferation. More than 290 journals, some key publishers, and some self-archivers already use it. The Scholar's Copyright Addendum engine generates a PDF file that describes what rights an author retains and attaches it to a publication. It is important to note that the addendum describes rights that most authors currently assume they have but that they often do not, such as rights to reproduce, distribute, display, and make derivatives of a publication. The engine is a useful way to both simplify and standardize author rights to digital objects. It will also be important to standardize some sort of version control, since if data are published in multiple databases, there may be some variation between copies (and between derivatives if those are an acceptable use).

BREAKOUT DISCUSSIONS

Commercial/Proprietary Obstacles

Participants: Tom Moritz (moderator), Paul Uhlir (moderator), Mindy Destro (notetaker), Hilmar Lapp, Trish Morse, Kathie Hodge, Katja Selmann, Kris Brugger, Shan Duncan, Sam Scheiner, and Bob Peet.

This group first defined the scope of the discussion. "*Commercial*" was taken in the most general sense to mean both "for-profit" and "revenue generating." "*Proprietary*" was taken to mean "legally protectable" with the sense of *exclusive* ownership / control (and as distinct from "private" e.g. confidential / secret).

Summary of discussion:

- Rigorous and well-informed financial and legal analysis is fundamental to future development
- "Commercial" / "proprietary" issues, such as loss of competitive advantage/economic benefits and feelings of personal "ownership," are not a major impediment
- Solutions exist and continue to emerge
- Solutions for financial / legal problems will likely be similar to solutions for socio-cultural barriers
- Creation of supporting infrastructure, who pays and who has access? (Public? Private? Commercial?)

Categories of data owners and the normal state of their data:

- *Governmental* – data are for public benefit (most U.S. Federal data in public domain)
- *Academics* (institutions / individuals) – data are available upon publication
- *Corporate* – data are proprietary / confidential (open-access to Eli-Lilly's "Adverse Effects" Website is a counter-example of the normally unavailable corporate data.)
- *Private Sector Non-profits / NGOs* – data are available consistent with mission (social contract)
- *Private Citizens* – data are contributed as public good ("citizen scientists" etc.)

The Bayh-Dole Act gives universities, small businesses, and non-profits intellectual property control of their “inventions” resulting from federally-funded research. This might result in a disincentive to share data that may be patentable and commercially exploitable (revenue generating).

Data can be protected through copyright, embargoes on use, or made inaccessible due to privacy and confidentiality concerns. The use of Creative Commons (<http://creativecommons.org>) licenses on data can provide some protection while allowing access. Though, does application of Creative Commons licenses set a negative precedent? Some people feel that all / most ecological data may be copyrightable (due to the “unique” origin / context in which it is gathered and because the products are from individual creative scientific insight). Further research is needed in this area.

Business requires a “return on investment.” Even “goodwill” has a dollar value. Companies can also benefit from the reputation of its productive scientists. Sharing may also produce a richer commercial environment (e.g., IBM’s profitable use of open source – LINUX and APACHE). New business models built on open source / open access need to be explored.

Commercial interests may constrain data sharing. For example, a conventional “subscription model” of access to data may be more of an obstacle, particularly when it is applied exclusively without other provisions for access. Not-for-profit societies might be constrained by the need to successfully compete with commercial publishers.

Concerns about 2nd party use of data might spur proprietary reactions. Concerns include loss of priority in publication and financial exploitation. Is citation per se adequate recompense for sharing? Further investigation and analysis is needed, along with establishment of standards and protocols for citation (Resource: Micah Altman and Gary King (2007) A proposed standard for the scholarly citation of quantitative data. D-LIB Magazine. 13: 3/4. <http://www.dlib.org/dlib/march07/altman/03altman.html>)

Mandates to share can come from the funding source and publishers. Concerns regarding mandates include enforcement, archiving infrastructure, and intellectual property barriers.

Sensitive Data

Participants: Rachel Muir (moderator), Hal Balbach (moderator), Jennifer Riem (notetaker), Bruce Dancik, Charles Fox, David Leslie, Wayne Litaker, Richard O’Grady, Elizabeth Sellers, Kathleen Smith, and Michael Whitlock.

This breakout group was tasked with identifying obstacles to sharing sensitive data and considering potential solutions. The main issues discussed were what characteristics of datasets should be considered “sensitive” and how different types of sensitivity could be addressed so that some form of the valuable data in sensitive datasets can be shared by the larger scientific community.

It was generally agreed that sharing is desirable if the dataset can be made non-sensitive by delaying its release (embargo) or by concealing its most sensitive information. To conceal the locations of sensitive species, for example, some location data will need to be altered. It was also agreed that there should be clear guidelines for evaluating what data are sensitive and standard practices for handling data that is judged to be sensitive. It will be important to determine who (data center or data collector) will be responsible for storing (and allowing access to) the original versions of datasets that contain sensitive information.

The group identified several types of potentially sensitive data:

- Rare and exploitable species and other biological resources
- Cultural resources (species of traditional and/or ethnobotanical use)
- Environmental health issues

- Security issues (data on structural locations, pathogens, pests)
- Privacy (identification information, private property issues)
- At-risk data (high priority, legacy data that could be lost)

Barriers to sharing sensitive data

- Some information in a dataset should not be shared (Ex: rare species locations)
- Some data are labeled “sensitive” when really the location of the data (server) is sensitive (concern over opening up a server for access to non-sensitive data could leave the sensitive data on the same server vulnerable to a security breach)
- Distinction between political sensitivity and ecological sensitivity

General obstacles to data sharing

- Difficulty to submit
- Lack of a place to submit
- Language barriers
- Multiple formats and multiple parties hold rights/restrictions– which has priority?

Solutions to barriers

- Journals should require depositing data as a condition of publication
- Use existing or create new depositories
- Provide assistance to submit data; keep process simple
- Modify or remove certain data fields; also can embargo data
- Publicize success stories

Recommended actions

- Journal editors and society representatives should meet soon to develop a policy statement; key journals should pilot implementation
- Journals should provide guidelines/policy statement for what is considered sensitive and how it will be altered
- Government agencies could develop policies for moving non-sensitive data to outside locations and work around broad policies against public server access
- Workshop at 2008 ESA meeting to publicize workshop results and implementation (journals)
- Outreach to other science communities (AAAS, AIBS, AGU)
- Encourage review paper on data sharing successes
- Training students in data management, mining, sharing

Suggested topics for next workshop

- Challenges of managing ecological data registry
- Data standards for publication, publishers, professional societies (carrot vs. stick)
- Expansion of reasons for sensitivity

Other issues to consider

- How to handle commercial value issues?
- What about costs to access the data bank?
- Standards for making spatial data “fuzzy”
- “Security” concerns may be due to lack of skills in managing hardware, not intrinsic
- How to cite data sources? Should data use be a “publication?”
- How to handle future data mining as a publishing practice?
- Means for managing authorship via repository protocols

Socio-cultural barriers

Participants: William Michener (moderator), Todd Vision (moderator), Sarah Carrier (notetaker), David Baldwin, Ahrash Bissell, Nora Bynum, Jonathan Duncan, Peter McCartney, Juan Núñez Farfán, Greg Riccardi, Jenny Talbot, Jim Reichman, Robert Cook, William Piel

The first question addressed by the group was: Why address this issue at all? The answer is to enable interdisciplinary research, improve quality of science, and generate new ideas with old data (reuse data).

The group identified the following socio-cultural barriers: fear of data misuse or lack of use; no reward system; time-consuming data input and cumbersome archival technology; lack of training; sensitive data issues; and lack of standard data management practices.

The group discussed a number of solutions, including providing a spectrum of data services (including workflow and analytical tools) that provide immediate benefits to inputting data and accommodating a range of data types (heterogeneous, homogeneous). They also suggested a 15-minute rule for submitting material with easy to use, user friendly technology. Additional solutions include creating a flexible and sustainable data archive through improved user interfaces, integrating with controlled vocabularies, and developing workflow and analytical toolkits.

The sharing of published data can be improved through journal leverage and the existing reward system for publication; however, unpublished data, particularly negative results, also have value. Thus, journals should also encourage related (unpublished) data to be archived with the published data.

Funders should tie sharing of data to subsequent funding opportunities. Data sharing history should be evaluated by funders reviewing funding proposals.

Additional incentives for sharing data include the possibility of increased collaboration and coauthorship and improved and enforced citation credit for data used by others (See Micah Altman, Gary King, 2007. "A Proposed Standard for the Scholarly Citation of Quantitative Data", D-Lib 13(3/4).)

The group acknowledged that there are complications with sharing data from derived / meta-analyses. It was suggested that derived datasets should be published and should include citation of critical components (such as raw data) from other publications/datasets.

Standard data practices need to be developed to ensure accessibility and usability by all potential users prior to enforcement of data sharing mandates. Common access mechanisms are needed for archives and data centers. Language barriers will also need to be addressed.

Training in archival techniques and standard data practices will need to be widespread, including education of future scientists.

Solutions to technological barriers include:

- Creating a flexible and sustainable data archive
 - Minimal standards (low utility) to rigorous standards (high utility)
 - Ongoing backend management to improve utility
 - Improved user interfaces tailored to individual communities
 - Integration with controlled vocabularies and ontologies
 - Community identification of metadata and interoperability standards
 - Automated mechanisms to assess/calculate data use metrics
 - Data stewardship, security, backups and mirroring
 - Create toolkits for all of these sets, deployed across numerous communities.

It was noted that funding is needed to research and develop the archival to analysis process, quality assurance and control tools, and ontologies.

WORKSHOP CONCLUSIONS AND RECOMMENDATIONS

Obstacles to data sharing include: fear of data misuse or lack of use; loss of the competitive advantage and subsequent economic benefits; loss of personal “ownership”; lack of a reward system; time-consuming data input and cumbersome archival technology; lack of training; sensitive data issues; and lack of standard data management practices.

Journals should require deposition of data as a condition of publication. Workshop attendees recommend that journal editors and society representatives meet soon to develop a policy statement, with key journals piloting implementation. Journals should also encourage sharing of unpublished data when published data is shared.

Funders should tie sharing of data to subsequent funding opportunities.

Data archiving should be a simple process (the 15-minute rule for submitting material) with easy to use, user friendly technology. Standard data practices need to be developed to ensure accessibility and usability by all potential users prior to enforcement of data sharing mandates. Common access mechanisms are needed for archives and data centers. Language barriers will also need to be addressed. Training in archival techniques and standard data practices will need to be widespread, including education of future scientists.

Workshop attendees concluded that “commercial” / “proprietary” issues are not a major impediment and that solutions exist and continue to emerge. Rigorous and well-informed financial and legal analysis is fundamental to the future development of data sharing practices. Data can be protected through copyright (including Creative Commons licenses), embargoes on use, or made inaccessible (modification or removal of certain data fields) due to privacy and confidentiality / sensitive data concerns. Improved and enforced citation credit will help address “proprietary” concerns.

Journals should provide guidelines/policy statement for what is considered sensitive and how the data will be altered. Government agencies could develop policies for moving non-sensitive data to outside locations and work around broad policies against public server access.

One incentive for archiving data would be the provision of a spectrum of data services (including workflow and analytical tools) that provide immediate benefits to the data “owner.”

An additional incentive for sharing data is the possibility of increased collaboration and coauthorship. These successful relationships, along with other data sharing successes, should be identified (perhaps a peer reviewed paper) and publicized to promote data sharing.

Workshop attendees recommend that a workshop be held at the 2008 ESA annual meeting to publicize the results of ESA’s Data Sharing Initiative and the subsequent implementation of data sharing efforts, such as the implementation of journal data sharing policies. Outreach to other science communities (AAAS, AIBS, AGU) is another important step.

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

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 Sharing Obstacles Workshop, May 30-31, 2007
National Evolutionary Synthesis Center, Durham, NC

| Attendee | Affiliation |
|-------------------|--|
| Harold Balbach | US Army Engineer Research and Development Center |
| David Baldwin | Ecological Society of America |
| Ahrash Bissell | Alexandria Archive Institute / Open Context |
| Kristin Brugger | DuPont Crop Protection Products |
| Nora Bynum | Society for Conservation Biology |
| Sarah Carrier | University of North Carolina Metadata Research Center |
| 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 |
| Charles Fox | British Ecological Society and University of Kentucky |
| Jane Greenberg | University of North Carolina School of Information and Library Science |
| Kathie Hodge | Mycological Society of America |
| Hilmar Lapp | National Evolutionary Synthesis Center |
| Chip Leslie | American Society of Mammalogists |
| Wayne Litaker | National Oceanic and Atmospheric Administration |
| Peter McCartney | National Science Foundation |
| Bill Michener | Long Term Ecological Research Network |
| Tom Moritz | Getty Research Institute |
| Patricia Morse | American Society of Naturalists |
| Rachel Muir | U.S. Geological Survey |
| Juan Núñez Farfán | Ecological Society of Mexico / Federation of the Americas |
| Richard O'Grady | American Institute of Biological Sciences |
| Bob Peet | University of North Carolina |
| Bill Piel | Yale University |
| Jerome Reichman | Duke University School of Law |
| Jim Reichman | National Center for Ecological Analysis and Synthesis |
| Jennifer Riem | Ecological Society of America |
| Greg Riccardi | MorphBank |
| Elizabeth Sellers | USGS National Biological Information Infrastructure |
| Katja Seltsmann | MorphBank |
| Sam Scheiner | National Science Foundation |
| Kathleen Smith | National Evolutionary Synthesis Center |
| Jenny Talbot | Ecological Society of America Student Section |
| Paul Uhler | The National Academies |
| Todd Vision | National Evolutionary Synthesis Center |
| Michael Whitlock | American Society of Naturalists |