

Data archiving for animal cognition research: Report of an NIMH workshop

HOWARD S. KURTZMAN

National Institute of Mental Health, Bethesda, Maryland

RUSSELL M. CHURCH

Brown University, Providence, Rhode Island

and

JONATHON D. CRYSTAL

University of Georgia, Athens, Georgia

In July 2001, the National Institute of Mental Health sponsored a workshop titled "Data Archiving for Animal Cognition Research." Participants included scientists as well as experts in archiving, publishing, policy, and law. As is described in this report, the workshop resulted in a set of conclusions and recommendations concerning (A) the impact of data archiving on research, (B) how to incorporate data archiving into research practice, (C) contents of data archives, (D) technical and archival standards, and (E) organizational, financing, and policy issues. The animal cognition research community is encouraged to begin now to establish archives, deposit data and related materials, and make use of archived materials in new scientific projects.

The emergence of the World-Wide Web and advances in computer technologies for conducting and reporting scientific research have spurred interest across many scientific fields in developing networked electronic archives for primary data and other research materials.¹ Recently, a number of animal cognition researchers have begun to consider the potential impact of electronic data archives on their field and to contemplate how such archives could best be organized and sponsored. With the goal of advancing discussion in this area, the National Institute of Mental Health (NIMH), one of the major providers of funding for research and training in the field, convened a workshop titled "Data Archiving for Animal Cognition Research" on July 19–20, 2001, in Washington, D. C.

Russell M. Church and Howard S. Kurtzman organized the workshop and served as co-chairs, and Jonathon D. Crystal served as both a primary participant and the rap-

porteur for the workshop proceedings. The 23 primary participants (listed in Appendix A) included leading animal cognition researchers whose experience spans a wide range of topics, approaches, and types of data, as well as professionals with expertise in related scientific areas and in archiving, publishing, policy, and law. Additional attendees included officials from various components of the National Institutes of Health (NIH) and the National Science Foundation (NSF). (Several other invitees could not attend due to schedule conflicts.)

Most previous discussions of data archiving in the behavioral sciences have concentrated on data from human studies. As is appropriate, the protection of human subjects' privacy has been the dominant concern. However, other important issues surrounding archiving also require careful attention. This workshop's exclusive focus on animal research enabled those issues to be more fully explored. The further limitation of the workshop's scope to animal cognition research made it possible to examine archiving realistically in the context of a specific field's goals, methods, organization, and traditions.

The workshop organizers believe that, at this early stage of thinking about electronic archiving, it is worthwhile for researchers in individual fields to undertake organized efforts, such as this one, that incorporate input from a significant number of the fields' members and other experts. This approach will ensure that archiving projects are designed in ways that satisfy each community's needs and aims, will increase the likelihood that novel ideas about archiving will gain a hearing in some forum, and will enable fields to draw from one another's experiences as efforts progress. It can be expected that

This report is a product of a workshop funded and organized by the National Institute of Mental Health, a component of the National Institutes of Health. However, it is not an official statement of the views of NIMH or of any other unit of the U.S. federal government. Individual workshop participants do not necessarily endorse all points expressed in this report. The co-chairs thank Steve Foote and Mary Ellen Oliveri of NIMH for their support of and contributions to the workshop. Correspondence concerning this report may be addressed to H. S. Kurtzman, Cognitive Science Program, National Institute of Mental Health, 6001 Executive Boulevard, Rm. 7217, Bethesda, MD 20892 (e-mail: kurtzman@nih.gov); R. M. Church, Department of Psychology, Box 1853, 89 Waterman Street, Brown University, Providence, RI 02912 (e-mail: russell_church@brown.edu); or J. D. Crystal, Department of Psychology, University of Georgia, Athens, GA 30602-3013 (e-mail: jcrystal@uga.edu).

fields will ultimately converge on solutions to various questions. Indeed, many of this workshop's conclusions and recommendations appear to be readily applicable to other fields within the behavioral sciences. Thus, the organizers hope that the present report will contribute to discussions of archiving in domains beyond animal cognition.

The first day of the workshop opened with participants briefly describing their own experiences, interests, questions, and concerns related to data archiving. The bulk of the day was then devoted to formal presentations, each 20–30 min long and followed by a discussion. (The presentations are listed in Appendix B.) The day ended with a general discussion assessing and expanding on points raised earlier. Over the course of the day, the participants considered many types of data (including numbers, text, pictures, video, and audio) and the means by which they are collected and used within various kinds of research (including laboratory, fieldwork, and theoretical work). Although other forms of data sharing were noted, the discussions focused almost entirely on electronic archives that would be publicly accessible over the Web.

On the second day, the workshop participants worked through a list of questions previously distributed by the co-chairs concerning the impact of archiving and how archives should be structured, used, sponsored, and governed. Through lively discussion, the participants reached general (though not always unanimous) agreement on a large number of observations, conclusions, and recommendations. These are summarized in the following sections.

A. Impact of Data Archiving on Research

1. Data archiving is likely to have significant positive effects on the conduct and application of animal cognition research. Currently, researchers' access to data from other investigators' studies is usually limited to the summaries provided in results and discussion sections of published works. Gaining access to complete data sets in comprehensible form is often inconvenient or impossible. As a result, researchers other than those who produced the data (and their associates) have limited opportunities to engage in further analysis and interpretation of the data.

By making complete data sets available to the entire research community, archiving would encourage broader and potentially richer treatments of the data. In particular, more extensive use of model-testing, meta-analysis, and data-mining techniques would become possible. Such work would stimulate more widespread discussion of specific data sets, possibly leading to new theoretical insights. As a further consequence of the heightened accessibility and visibility of data, new research collaborations are likely to arise, including among researchers who take different approaches or are at different institutions.

Archiving would also enhance the efficiency and quality of research efforts. Inspection or reanalysis of archived data from studies conducted by other researchers would provide answers to questions that today can often be addressed only through time- and resource-consuming replications or extensions of the work. Archives could also pro-

vide a forum for reporting of studies that have yielded null results, which, even when well designed and informative, are often not published.

Moreover, awareness by investigators that their data will become publicly available would encourage them to be especially careful in the design and conduct of their studies and in the handling and analysis of the data. When errors do occur, public archiving would make detection and corrective responses, by either the producer of the data or another researcher, more likely.

Data archives could play a useful role in graduate and undergraduate education in cognition, behavior, and statistics. Faculty could provide more comprehensive coverage of studies presented in lectures and readings, including more precise discussion of how the data from a study bear on its theoretical questions and claims. Students could gain experience in exploring alternative analyses of data and in making comparisons across studies (including their own projects). Access to archived data may be especially important at institutions that have limited laboratory resources.

Supporters and consumers of research—including government and private funding agencies, academic and health institutions, private firms, and public advocacy organizations—could also benefit from data archives. The availability of complete data sets would enable them to evaluate research outcomes and their practical implications more effectively. In the United States, laws and policies are moving toward stronger requirements for sharing of data that are collected with the support of federal funds.² Well-designed data archives would provide researchers with a mechanism for complying with such requirements in a manner that reduces the risk of distortion or misinterpretation of the data by others.

2. Many of the difficulties that data archiving potentially creates for researchers can be prevented by adoption of appropriate standards for archive design and use. Remaining difficulties are outweighed by the benefits of archiving. The prospect that data archiving may become a standard practice raises various concerns among researchers. These concerns helped motivate a number of the specific recommendations generated at the workshop.

One potential problem for researchers is that preparing data and submitting them to archives would require significant time and effort that could otherwise be spent on new research. Accessing archived data might also be a complex task. In fact, however, working with archives need not be burdensome. Much of the difficulty can be prevented by incorporating archiving into research projects from the start (as will be discussed in Section B.3 below) and by the development of straightforward user interfaces for both depositing and retrieving materials at archive websites (see Section B.4).

Another concern is that researchers do not receive adequate professional credit for archiving their data. Although this is true today, it need not remain so. As will be discussed in Section B.1, scientific societies, pub-

lishers, and academic institutions can develop mechanisms for highlighting individual researchers' contributions to archives and for measuring the impact of those contributions.

Some researchers are troubled that data archives will enable other investigators to publicly expose flaws in their research or to discover and take credit for patterns in their data that they had not noticed. However, the consensus of the workshop was that the contribution to scientific progress of such subsequent analyses outweighed the professional difficulties created for the individual researcher who has produced the data. It was noted that, although enforcement is infrequent, researchers are already expected to share their data on request with others who seek to perform additional analyses.³

Nevertheless, steps can be taken to reduce the risk of problems for individual researchers and to maintain a collegial research culture. For especially complex data, researchers can be allowed to delay archiving beyond publication of initial reports while they carry out their own further analyses (see Section B.2). Also, investigators working with archived data can be encouraged to collaborate and share credit with, rather than "scoop," the producer of the data whenever possible (see Section B.6).⁴

There is also concern that secondary analyses and interpretations presented by other investigators may be faulty because the investigator did not adequately understand the data and how they were generated. To a large extent, this problem can be prevented by including in the archive complete documentation of the data files and detailed methodological information about the study (see Section C.2).

Some researchers raise, as a fundamental objection to archiving, the concern that it undermines their ownership or control of the data they produce. The workshop consensus was that, although perhaps once legitimate, the view that researchers should be able to maintain permanent and complete control of their data can no longer be accepted. Rather, researchers must recognize that science is a collective enterprise in which they have responsibilities both to fellow scientists and to the other segments of society that support and utilize their work. Therefore, the most basic product of research—data—should be made available to colleagues and to society at large. While researchers should have priority in access to and use of the data they have worked hard to produce, they should not be able to control those data for an indefinite period.

B. Incorporating Data Archiving Into Research Practice

1. Researchers should be strongly encouraged, though not required, to archive their data. All researchers, including students, should be encouraged to archive the data they produce. Scientific societies and journals can promote archiving through articles, editorials, symposia, websites, and the like, that discuss the various conceptual and practical issues surrounding archiving and provide concrete guidance on how to set up and use archives. Societies and publishers can themselves

sponsor archives for the data produced by their members and authors.

To demonstrate the utility of data archiving, books and special issues of journals can be published that feature work in which archived data play crucial roles in advancing research. Publications by high-profile researchers in which archived data are used could be especially influential during this early period. Awards can be established for important research contributions that rely on archived data.

In order to ensure that researchers get credit when their archived data are used by others, journals should develop required citation formats for data archives that include at least identification of the data producer, the archive Web address, and the primary publications (if any) in which the data were originally reported. Such citations can also be listed in researchers' curricula vitae, perhaps along with citation and/or website access counts.

Research institutions and departments can facilitate their investigators' data archiving by providing Web space, along with technical tools and assistance, for the creation and maintenance of archives (or by making arrangements for such with other organizations). Individual researchers' efforts to archive their data should also be considered in hiring and promotion decisions. Furthermore, credit should be given to researchers' productive secondary uses of archived data, in addition to original empirical work.

Both government and private science-funding organizations can encourage archiving by allowing grant budgets to include expenses related to the preparation of data for archiving and for the creation of archives by individual laboratories. These organizations can also solicit grant proposals for research that primarily involves secondary analyses of archived data.

As was noted earlier, some U. S. federal funding agencies are moving toward requiring that researchers take active steps to share the data that they have produced with agency funds. Such requirements will, of course, be a strong inducement to researchers to place those data in public electronic archives. They might also lead eventually to broad acceptance among researchers of the view that all data, whatever the funding source, should be considered for archiving.⁵

However, the consensus of the workshop was that, at this time, no organization should make it compulsory for researchers to archive the data that they produce. Members of the research community first need to attain a better understanding of the roles archiving can play in research and how archiving can be implemented. This will be achieved through substantial discussion in various forums and experimentation with multiple approaches to archive sponsorship, standards, and policy (as will be discussed in the sections below). The question of compulsory archiving can be revisited when a significant portion of the research community has reached a high level of understanding of, and comfort with, archiving.

2. For most studies, data should be archived when the corresponding report is accepted for publication. For some complex types of research, a delay in archiv-

ing is appropriate. Data that will be reported in publications are ideally entered into an archive immediately upon the report's acceptance for publication (when it is in press). A preprint of the publication can be posted on the Web at the same time. Thus, following acceptance, there would be no waiting period during which the authors would have exclusive access to the data. Publishers should not embargo reports accepted for publication or the related data.⁶

However, for some types of studies, it is reasonable to delay data archiving for a period following publication. These include certain longitudinal and field studies, studies that collect video and audio recordings, and other studies that produce rich data sets that call for multiple, time-consuming analyses that may reasonably be reported in separate publications. Delaying archiving will enable the researchers who produced the data to maintain exclusive access while they complete their planned analyses. The likelihood that other researchers will "scoop" the producers of the data is thereby reduced. How long a delay is reasonable will depend on the particular study.

Just as archiving itself should not be compulsory, there should be no strict requirements concerning when the data for any type of research are to be archived. Although prompt archiving should be the default practice, standards for what counts as a legitimate waiting period for various kinds of studies should be allowed to evolve within the research community.

Researchers can also, at their discretion, archive data that will not be reported in publications (see Section C.1 below). Data that will not be published are most useful to others if they are archived quickly. One possible guideline as to the best time to archive such data is as soon as the researcher is confident that they will not be reported in a publication.

3. Data archiving should be built into the initial planning of research. A concern raised by some researchers is that preparing their data for deposit into an archive is burdensome and time consuming. This need not be the case, however, if considerations of future data archiving are incorporated into the early planning and conduct of the research. For example, careful initial design of data formats and file organization can ensure that the final data set will be appropriate for transfer to archives and interpretable by researchers who will access the archives. When well thought out in advance, preparation of data for archiving should not be difficult or costly.

4. Data archives should be free, open, and user-friendly. Archived data should be readily accessible via any up-to-date computer system on the Web. Individuals should not be required to pay subscription fees, membership dues, per-use fees, or other charges in order to gain access to archived data or to deposit their data in archives (although the institutions or scientific societies with which they are affiliated might support costs; see Section E.2 below). Websites should have informative names, be listed on well-publicized archive registries,

and be locatable by major search engines (as will be discussed further in Section D.1 below). Interfaces and procedures for accessing and depositing data should be clear, user-friendly, and well documented.

5. Access to archived data should be anonymous. In order not to inhibit use of archived data, access to archive websites should be anonymous, as is access to the publications describing the data. Lack of anonymity is one reason why previous data-sharing standards, such as the APA's (see note 3), have been largely ineffective.

Thus, archive users should not be required to supply identifying or descriptive information about themselves. Transfer of information about the user's computer system should likewise be limited (e.g., users' Web browsers should not be required to accept cookies).

6. Secondary users of archived data should be encouraged to consult with the data producer. Investigators who plan to make significant use of archived data for new research purposes should be strongly encouraged to consult with the producers of those data (if any of the producers are still active). The possibility of their collaborating on the new research and co-authoring resulting publications should be discussed. While such close cooperation among researchers will not always be feasible or even productive, serious efforts will help to enhance communication and mutual respect among members of the research community and prevent incidents in which researchers believe that others have inappropriately used or gained unfair benefit from the data they have produced.

C. Contents of Data Archives

1. All new data reported in publications are appropriate for archiving. Data from well-designed unpublished studies, as well as some older published data, are also appropriate. All new data reported in published articles, chapters, and books are appropriate for archiving. In general, the complete, unsummarized data for each animal in each condition should be provided.

Data that have not been reported in publications may also be archived, at the discretion of the researcher. Such unpublished data may be either from work that supplements previous publications (and should be noted as such) or from novel work. It may be especially worthwhile to archive unpublished data from studies that were apparently well designed but produced null results. These data can be used in meta-analyses and in guiding the design of subsequent work by other researchers.

Although archiving of new data should have the highest priority, it is also useful to archive data from earlier studies that are of continuing relevance for contemporary investigators or are of particular historical interest. However, the value of having older data sets available in electronic archives must be weighed against the amount of effort required to put them in appropriate form (especially for those data for which only paper records currently exist).

2. In addition to data, an archive entry should contain complete information about the study's methodology and publication status and, ideally, other ma-

terials that will help others understand the research.

In order for the data from a study to be interpretable and usable by others, additional information needs to be provided. The archive entry for a study should contain a complete description of the dependent measures, along with documentation of all coding, labeling, and file-naming conventions. If not fully provided in a published report, the entry should also contain or link to the entire set of stimuli and to detailed descriptions of the study's aims, procedures, and data analysis methods. Any missing data or idiosyncratic features of the study (e.g., animals' prior testing experiences, unusual field or laboratory conditions, equipment failures) should be characterized.

It is also important that the publication and peer-review status of each individual data set be clearly indicated. For data that have been reported in publications, complete citations for the publications, as well as links to them, should be included. For unpublished data, the archive entry should contain information regarding whether and where the research has been peer reviewed, expected publications reporting the data, and relations to other data reported in publications.

Additional material can also be included at the researcher's discretion. For example, accounts of pilot studies that motivated decisions regarding study design may be provided. It may also be worthwhile to include or link to software used in generating stimuli, performing procedures, displaying and reorganizing data, filtering and analyzing data, representing and testing quantitative models, and so forth. (For proprietary software, executable files can be provided.) Any material, including audio and video files, that might help other researchers precisely understand, evaluate, and build upon the study should be considered for inclusion.

Archive entries should also include full contact information for the lead researcher(s) responsible for the data, the names of other researchers contributing to production of the data, and acknowledgement of funding sources. All archive entries and updates should be clearly labeled and dated.

3. Video material is extremely useful but raises special concerns. Video recordings of animals behaving in field and laboratory settings are an especially rich source of information. They are useful not only for evaluation of hypotheses, but also for suggesting new research ideas and for educational purposes. Inclusion of video files, either as primary data or as supplementary information, can significantly enhance the value of data archives.

Researchers should take measures to prevent misinterpretations, by scientists and others, of the conditions and events depicted in videos. Verbal descriptions of video materials should be provided to clarify their content. When only illustrative video excerpts are included in the archive, they should be carefully selected to avoid misleading viewers about characteristics of the overall study. In addition to the study itself, it may be worthwhile to provide video of the surrounding research envi-

ronment, including animal care facilities. Researchers may find it useful to consult with colleagues concerning how to use video recordings in a manner that most accurately represents their study.

D. Technical and Archival Standards

1. Data archives should comply with the technical standards of the Open Archives Initiative (OAI). Within the OAI framework, a set of metadata elements for animal cognition research should be developed. Archive websites and the servers that support them should comply with the technical standards of the Open Archives Initiative (OAI).⁷ OAI standards are rapidly emerging as the dominant framework for designing digital repositories in all scholarly fields and for making them fully accessible on the Web. Central to the OAI framework is the use of *metadata*, which serve to describe each entry in an archive in a uniform, structured fashion. The OAI standards include a basic set of metadata elements, but this set can be supplemented within individual archives by additional elements that provide more specific information about the contents of the entries. Metadata enable services operating over the Web to search for and catalogue particular types of materials across the universe of OAI-compliant archives. Such services are especially useful when information in a field is distributed over a large number of independent archives.

A standard set of metadata elements for describing animal cognition research is needed in order to facilitate users' ability to locate and collect the particular data sets in which they are interested. These elements might refer to species, research area (e.g., Pavlovian conditioning, foraging, counting, visual categorization), experimental paradigm, data type, and so forth. Development of a useful metadata set (and updates as science evolves) would best be handled by working groups representing a wide range of the research community, perhaps convened by a scientific society.

2. Format standards for information included in archive entries should be developed. Standards are needed for the format of data and other materials included in archive entries. Such standards will provide guidance to researchers preparing archive entries and will help ensure that archives are complete, reliable, and intelligible. Working groups with a broad representation of archive users and sponsors are best suited to develop format standards.

Format standards for numerical data may include the requirement, for example, that the file types enable the data to be directly manipulated and analyzed (with the result that certain types of pictorial files would not be acceptable for presentation of numerical data). They might also address spacing, tabbing, and other aspects of the layout of numbers and identifiers.

The development of format standards for video and audio data will be especially challenging, given the wide variability of such data and ongoing developments in technologies for storing and transmitting video/audio

files. At the very least, it can be suggested that archived video/audio files be accessible via the most commonly used open-source software.

Format standards should not be so rigorous or difficult to implement that researchers are discouraged from archiving their data. It should also be recognized that data from certain types of research, such as those involving field observation, may not lend themselves to standardized formats.

3. The integrity of archived materials must be protected. It is imperative that the authenticity and completeness of archived materials be protected. The managers of the Web servers that host archives will have major responsibility for protection through the implementation of state-of-the-art firewalls and other technological means. Within the OAI framework, alterations and removal of files are noted in metadata records and can thus be readily detected and corrected.

Researchers depositing materials into archives can help protect them through such means as labeling and dating all entries (including updates and corrections) and electronic watermarking of image files. Technical staff at researchers' institutions can provide guidance on such methods.

4. Steps should be taken to ensure the long-term preservation and accessibility of archived materials. Preserving and maintaining accessibility to archived materials for the long-term future is a major challenge. Both the researchers who deposit and oversee materials in archives and the managers of supporting Web servers should make plans, extending as far into the future as may be practicable, to ensure that materials will remain available in usable forms. The need to migrate archived materials to new technologies and to adapt to new architectural and format standards should be anticipated. Techniques should be implemented for automatically redirecting users to changed URLs and e-mail addresses (perhaps via a centralized registry of archive websites and addresses). To prevent irretrievable loss of materials, remotely stored backup systems and mirror websites should be established.

Of particular concern is the preservation and accessibility of archived materials after the researcher is no longer active or is no longer affiliated with an institution that supported the archiving. Both the researcher and the institution should be involved in planning for a permanent archival home for the material. The institution might make the commitment to maintain it permanently. Alternatively, the material could be transferred to another organization that has taken on the task of long-term support of archives, perhaps a scientific society or academic consortium. Commercially supported archives, which are subject to market pressures, should generally not be relied upon for maintaining long-term accessibility.

E. Organizational, Financing, and Policy Issues

1. At this stage in the development of data archiving, a broad range of individuals and organizations should be involved in sponsoring and overseeing

archives. A wide variety of entities should be encouraged to create and manage archives and to develop guidelines for them. These include individual researchers; groups of individuals with shared research interests; established organizations, such as academic institutions (at the levels of the entire institution and of particular departments and labs), scientific societies, publishers, and information technology firms; as well as various types of consortia.⁸ It is impossible to predict which organizational and management models will turn out in the long run to be most successful in implementing the recommendations outlined here and in serving the emerging needs of archive users. Multiple initiatives will enable greater innovation, more precise comparative assessments, and more rapid dissemination of best practices.

A particular entity need not take on all functions related to archiving. In particular, development of metadata and format standards would seem best handled by an established organization, such as a scientific society, that is perceived as legitimately representing the entire field and can recruit input from a broad range of relevant researchers. By contrast, storage and technical support for archives could be handled either at researchers' home institutions or at a more centralized site. Various models can be imagined for federations of distinct archives that have some degree of shared standards, review mechanisms, governance, and facilities (including, perhaps, a single common portal website). The OAI framework is designed to facilitate such federations.⁹

2. Most costs of archiving should be incorporated into organizations' regular budgets. Costs are associated with both start-up and maintenance of archives. Start-up costs include those of server and storage hardware; technical support and assistance; design of websites, user interfaces, and related software; and development of metadata and format standards. It is anticipated that organizations will incorporate these initial archiving costs into their regular budgets for research support, digital libraries and publications, and information systems. Workshop participants recommended that government funding agencies offer seed money grants for especially significant or innovative archive projects.

Maintenance costs include those of daily operations and technical staff, long-term preservation of archived materials, and periodic evaluation and updating of archive systems. Again, these costs would be incorporated into organizations' regular budgets, perhaps with groups of organizations pooling resources. In some cases, a dedicated endowment or funding stream for ongoing archiving activities might be appropriate. Individual researchers can include expenses related to archiving their own data in grant budgets.

As was stated in Section B.4, individuals should not have to pay a fee to gain access to or deposit materials in an archive. The institutions with which they are affiliated, however, might pay fees to help support archive functions provided by other organizations. In the interest of extending scientific information and training as widely

as possible, institutions with few financial resources (including those in economically less advanced countries) should have their participation in archive activities subsidized.

3. Data archives for animal cognition research should ultimately be integrated with archives in related areas. Given the close ties of animal cognition research with other areas of behavioral and biological science (including both nonhuman and human work), data archiving for all such fields should ultimately be integrated. At a minimum, this integration would take the form of shared or overlapping metadata standards. Researchers in all fields would then be able to locate and work with data from the entire range of studies relevant to the issues they are addressing.

Data archives could also be usefully integrated, at least in part, with archives currently being developed for other types of materials. These include archives for reports and publications (which may or may not be associated with print journals), archives for information about research resources (e.g., specialized laboratories, equipment, animal supplies, software, reference materials), and archives for computational models and simulations ("theory archives"). A broader notion of *research archives*, organizing all types of materials of interest to researchers, might eventually be realized.

In order to facilitate appropriate forms of integration, each segment of the research community should stay aware of how others are approaching archiving issues and take into account the goal of integration as they develop their own archiving projects.

4. Ethical standards for archive use should be developed. All organizations that sponsor and promote research should formulate and disseminate standards for ethical and appropriate scientific conduct regarding archives. Such standards would, for example, constrain researchers from depositing inaccurate or misleading data in archives or from tampering with or misrepresenting others' archived data. They would also require that authors who discuss archived data provide citations that identify the data producer and the archive website. Also, as was addressed in Section B.6 above, researchers who make significant use of others' data should be expected, where feasible, to consult with the producers of the data and discuss the possibility of collaboration early in the research process.

These and other standards should be developed with input from a wide range of researchers and research sponsors. Like other standards for research and academic publication, they should be enforced largely through social pressure and local mechanisms, rather than legal means.

5. Evaluation of archiving projects and planning of new efforts should be ongoing activities involving a wide variety of participants. The various organizations and individuals with interests in data archiving should periodically evaluate how well available archives (both their own and those of others) are meeting the needs of

the field and develop specific plans for improvement and expansion of archives. For some organizations, it would be useful to have standing committees devoted to archiving issues. Conferences and workshops should be held to share experiences and perspectives from within and outside the field.¹⁰ All such evaluation and planning efforts should consider the impact of new developments in research methods, publication practices, technologies (including the continuing evolution of the Internet and the Web¹¹), and the domestic and international legal environment (including government science agency policies and intellectual property law¹²).

Conclusion

Data archiving can produce significant benefits for the animal cognition research community. In order to maximize the utility of archives and to make working with archives (both depositing and accessing materials) attractive and convenient for a broad range of researchers, various issues surrounding the content, structure, and management of archives need to be addressed. Institutional policies and social norms concerning how archives are used also need to be formulated. The workshop participants made significant progress in analyzing and meeting these challenges, laying the groundwork for specific projects aimed at implementing archives and developing guidelines and standards. Individuals, small groups, and organizations can each play important roles in refining and evaluating particular approaches to archiving. The sooner such efforts begin, the sooner archiving can yield its full benefits for the field.

REFERENCES

- GOVERNING COUNCIL OF THE ORGANIZATION FOR HUMAN BRAIN MAPPING (2001). Neuroimaging databases. *Science*, **292**, 1673-1676.
- JOHNSON, D. H. (2001, October). Sharing data: It's time to end psychology's guild approach. *APS Observer*, **14**, 1, 38-39.
- THE FUTURE OF THE ELECTRONIC SCIENTIFIC LITERATURE. (2001). *Nature*, **413**, 1-3.

NOTES

1. See, for example, Johnson (2001) and other articles by the same author in subsequent issues of the *APS Observer* (<http://www.psychologicalscience.org/observer>); Governing Council of the Organization for Human Brain Mapping (2001); and "The Future of the Electronic Scientific Literature" (2001) and related discussions at <http://www.nature.com/nature/debates/e-access>. Organizations promoting exploration of data archiving include the Committee on Data for Science and Technology (<http://www.codata.org>), the Coalition for Networked Information (<http://www.cni.org>), and the Digital Library Federation (<http://www.diglib.org>). The Free Online Scholarship Newsletter (<http://www.earlham.edu/~peters/fof>) provides coverage of technological and policy developments related to data archiving.

2. These measures include the Freedom of Information Act and the related "Shelby amendment" (for an overview from the NIH Office of Extramural Research, see http://grants.nih.gov/grants/policy/a110/a110_guidance_dec1999.htm); new data-sharing policies currently being developed by NIH (http://grants.nih.gov/grants/policy/data_sharing); and an existing NSF policy that applies primarily to human research (<http://www.nsf.gov/sbe/bcs/common/archive.htm>).

3. See, for example, the Ethics Code of the American Psychological Association (<http://www.apa.org/ethics>).

4. A suggestion offered at the workshop, though not fully discussed, was that editors and reviewers should be given access to the primary data reported in manuscripts submitted for publication; they could then evaluate the quality of the data and analyses and alert authors to potentially embarrassing flaws before the work is made publicly available in an archive or publication. Before such direct review of data is implemented, however, consideration should be given to its broader implications for research and editorial practice.

5. The evolving NIH and NSF policies are not absolute; they allow researchers to refrain from sharing their data if they provide adequate justification.

6. Implementation of this recommendation may require changes in the agreements that publishers typically ask authors to sign at this time.

7. The OAI standards are accessible at <http://www.openarchives.org>.

8. Examples of university-level archiving projects include DSpace (<http://www.dspace.org>), a joint initiative of MIT Libraries and the Hewlett-

Packard Company, and eScholarship (<http://escholarship.cdlib.org>), a project of the California Digital Library (University of California system). For an example of an archiving project from a single laboratory, see <http://www.brown.edu/Research/Timelab>.

9. For an example of an OAI-based archive federation, see the Open Language Archives Community (<http://www.language-archives.org>).

10. Archiving can itself be a topic for formal research, as has been evidenced by the recently founded *Data Science Journal* (<http://www.datasciencejournal.org>), sponsored by the Committee on Data for Science and Technology.

11. See the activities of the Internet2 organization (<http://www.internet2.org>) and the World-Wide Web Consortium (<http://www.w3.org>).

12. The National Academies (<http://www.nationalacademies.org>) periodically hold meetings and issue reports on legal and policy aspects of data sharing and archiving.

APPENDIX A Primary Participants in the Workshop

Michael T. Bardo, University of Kentucky
 Sarah T. Boysen, Ohio State University
 Merry Bullock, American Psychological Association
 Mary M. Case, Association of Research Libraries
 Russell M. Church, Brown University, *co-chair*
 Robert Cook, Tufts University
 Jonathon D. Crystal, University of Georgia
 Samuel A. Deadwyler, Wake Forest University School of Medicine
 Fred C. Dyer, Michigan State University
 C. R. Gallistel, Rutgers University
 Stevan Harnad, University of Southampton, England
 Michael F. Huerta, National Institute of Mental Health
 David H. Johnson, National Center for Policy Research for Women and Families
 Peter R. Killeen, Arizona State University
 Kimberly Kirkpatrick, University of York, England
 Howard S. Kurtzman, National Institute of Mental Health, *co-chair*
 Richard A. Lambert, Office of the NIH Legal Advisor, National Institutes of Health
 Heather Miller, Office of Extramural Research, National Institutes of Health
 Ralph R. Miller, State University of New York, Binghamton
 Wade E. Pickren, American Psychological Association
 David C. Riccio, Kent State University
 Gary R. VandenBos, American Psychological Association
 Edward A. Wasserman, University of Iowa

Note—Affiliations are presented as of July 2001.

APPENDIX B Scheduled Presentations (Day 1)

The Future of Research Archiving: Stevan Harnad
Perspectives of the Research Library Community: Mary M. Case
Perspectives of the Publishing Community: Mark Zdrozny (Academic Press and Association of American Publishers)—withdrew due to schedule conflict
Data Sharing in Neuroscience Research: Michael F. Huerta
Data Archiving in Behavioral and Social Science Research: David H. Johnson
NIH Policy Perspectives on Data Archiving: Heather Miller
Legal Issues Related to Data Archiving: Richard A. Lambert
APA's Interests in Data Archiving: Merry Bullock, Wade E. Pickren, and Gary R. VandenBos

(Manuscript received July 28, 2002;
 accepted for publication August 14, 2002.)