Sci Eng Ethics (2010) 16:783–799
DOI 10.1007/s11948-010-9239-x

The Essential Nature of Sharing in Science

Beth A. Fischer · Michael J. Zigmond

Received: 17 July 2007 / Accepted: 16 April 2010 / Published online: 25 November 2010

Abstract Advances in science are the combined result of the efforts of a great many scientists, and in many cases, their willingness to share the products of their research. These products include data sets, both small and large, and unique research resources not commercially available, such as cell lines and software programs. The sharing of these resources enhances both the scope and the depth of research, while making more efficient use of time and money. However, sharing is not without costs, many of which are borne by the individual who develops the research resource. Sharing, for example, reduces the uniqueness of the resources available to a scientist, potentially influencing the originator's perceived productivity and ultimately his or her competitiveness for jobs, promotions, and grants. Nevertheless, for most researchers—particularly those using public funds—sharing is no longer optional but must be considered an obligation to science, the funding agency, and ultimately society at large. Most funding agencies, journals, and professional societies now require a researcher who has published work involving a unique resource to make that resource available to other investigators. Changes could be implemented to mitigate some of the costs. The creator of the resource

This paper is based on a presentation made at the conference "*New Capabilities, Emerging Issues, and Responsible Conduct in Data Management*" jointly sponsored by the U.S. Office of Research Integrity and the University of Maryland Baltimore and held in Baltimore, Maryland on September 28–29, 2006. This paper and associated references reflects the state of the field at that time.

B. A. Fischer (✉) · M. J. Zigmond
Survival Skills and Ethics Program, University of Pittsburgh School of Medicine,
3500 Fifth Ave., Suite 202, Pittsburgh, PA 15213, USA
e-mail: bfischer@pitt.edu

B. A. Fischer
Department of Family Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA

M. J. Zigmond
Department of Neurology, University of Pittsburgh, Pittsburgh, PA 15213, USA

could explore the possibility of collaborating with those who request it. In addition, institutions that employ and fund researchers could change their policies and practices to make sharing a more attractive and viable option. For example, when evaluating an individual's productivity, institutions could provide credit for the impact a researcher has had on their field through the provision of their unique resources to other investigators, regardless of whether that impact is reflected in the researcher's list of publications. In addition, increased funding for the development and maintenance of user-friendly public repositories for data and research resources would also help to reduce barriers to sharing by minimizing the time, effort, and funding needed by individual investigators to comply with requests for their unique resource. Indeed, sharing is an imperative, but it is also essential to find ways to protect for both the original owner of the resource and those wishing to share it.

Keywords Ethics · Responsible conduct · Publication · Research article · Sharing

> *"Sharing isn't easy…and learning to share is a long process…. It can be hard…to understand what's mine, yours, and ours. Sharing grows little by little, as [we] develop the ability to see things from another person's point of view and to trust that what they share will be given back."* – Fred Rogers (2004)

The great advances now being made in science are the combined result of the efforts of numerous individual researchers *and* their willingness to share the products of their work. However, sharing was not always the norm. Before the appearance of scientific journals in the 17th century, researchers are said to have been hesitant to publicize their ideas, methods, results, and conclusions with others. In the absence of a mechanism for establishing priority, they were apparently worried that other scientists might take their findings and claim them as their own (see discussion of these points by Daniel Boorstin (1983)).

Before the advent of scientific journals, many attempts were made to deal with those concerns. For example, some scientists prepared and distributed multiple copies of a letter in which they described their results to a colleague. Some results were distributed using an anagram that could be decoded if priority ever came into question. And manuscripts in sealed envelopes were deposited with the Royal Academy of Sciences of England (Boorstein 1983; Merton 1957, 1961). Although these efforts may have helped to protect the interests of the authors, they do not seem to have led to an increase in the sharing of results or in either the pace or the efficiency of science.

The pace of scientific advancement, however, changed radically with the development of scientific journals: In 1665, Henry Oldenburg of the Royal Academy of Sciences launched the publication *Philosophical Transactions of the Royal Academy of Sciences,* which remains in print today (Royal Academy of Sciences 2010). As a consequence of *Philosophical Transactions* and the general system of scientific publication that followed, research results could be made available to the broad scientific community, while ensuring that priority for discovery would be provided to the first person to publish in a peer-reviewed

journal. Today, more than three centuries later, sharing is *de rigor*—at least in principle. It has rapidly become an integral component of research and an imperative imposed by the guardians of the enterprise—funding agencies, research institutions, and professional societies.

Nevertheless, sharing is anything but simple. There are costs and benefits, disincentives and incentives, and practical considerations.

Why Share?

Sharing Permits Research to Progress Faster and Farther

Provides a Foundation of Knowledge

Isaac Newton is reported to have said *"If I have seen farther, it is by standing on the shoulders of giants."* (Koyre 1952) Surely this phenomenon is true of any successful scientist. Through the sharing of results and ideas, researchers are able to build on the efforts of those who came before them. This concept is integrated into every training program for scientists. Beginning students typically gain their first exposure to research findings through textbooks, which provide an overview of previous studies. As they progress into junior researchers, they begin a life-long inquiry into the primary literature. One cannot imagine a forward-moving research enterprise without this connection to the work of others.

Broadens Scope of Research

Being able to learn from the work of others allows individuals to specialize, while at the same time ensuring that the collective effort is broadened. In many areas of science, there is so much knowledge and so many complex methodologies that a single individual can no longer acquire expertise in all facets of a research question. When scientists share their expertise and/or the fruits of their research, the overall effort moves more quickly and to greater depths.

Diversifies Perspectives

Broader access to ideas, data sets, and unique resources increases the contributions to science that can be made by individuals from institutions having limited resources (Flournoy and Hearne 1990; Governing Council of the Organization for Human Brain Mapping 2001). Not only does this enlarge the pool of researchers who can work on a topic, it also is likely to increase the diversity of that pool. Given that an individual's personal and professional background shapes the way he or she views and interprets the world, greater diversity in the scientific workforce should, in turn, influence the questions that are being investigated, the approaches being used and, ultimately, the application of the results (National Science and Technology Council 2000). This may facilitate finding solutions to problems that had previously seemed intractable.

Sharing Allows Resources to Be Used More Efficiently

Reduces Costs—Both Money and Effort

Sharing helps to decrease the chance that several individuals will develop the same method in parallel or unknowingly perform the same experiment. That, in turn, reduces the cost of the research in terms of the funding and effort required. Of course, some duplication is essential to the progress of science, as a finding is not considered to be a fact until other investigators have replicated it. Sharing is essential to that activity, too.

Maximizes Use of Data

Some data sets are so large that they are rarely fully explored (Ball et al. 2004). Examples include the results of genomic studies, data from high-throughput screens or microarray analyses, DNA or protein sequences, X-ray crystallography, epidemiological and prospective clinical studies, and anthropological studies. A large survey might involve studying thousands of subjects over many years. The results may initially be used to examine the impact of only one variable (e.g., the impact of smoking on heart rate) yet contain data that would be useful in exploring many other questions (e.g., the impact of nutrition on cardiovascular health or of exercise on neurodegenerative disease). Placing such data in a publicly accessible repository makes it possible for other researchers to help 'mine' the data set. In these cases, shared data might lead to additional publications—perhaps even including the originator as an author—with little or no additional data gathering and thus at a fraction of the cost that would otherwise be required for the study. Moreover, in cases of historical and retrospective studies, the only source of data may be the original data set (Ceci and Walker 1983).

Corrects Error of Analysis

In some cases, broader access to data sets may result not only in new findings, but also in the correction of old ones. Honest errors can be made in the analysis and interpretation of data. Authors may misjudge the specificity of their assays, use inappropriate statistics, or fail to recognize a bias in their subject population. An example of the latter seems to have occurred in a study by the University Group Diabetes Program (UGDP). This large clinical trial was designed to explore the relative value of different treatment options for Type 2 diabetes. Data were collected between 1961 and 1978, and publications began to appear by UGDP investigators as early as 1970. A critical review of the data by an outside group concluded that several errors had been made in the design and analysis of portions of the study and that, as a result, certain conclusions needed modification (Kilo et al. 1980). These observations were only possible because of the access provided to data sets by the UGDP researchers.

Increases Impact of Findings

When resources used in a published report are subsequently made available to others who then report their own findings, more attention is likely to be given to the initial paper, which will amplify its impact. For example, in a study of sharing of cancer microarray clinical trials, it was observed that the sharing of detailed research data was associated with a 69% increase in citation rate of the originating authors' work (Piwowar et al. 2007). This effect was found to be independent of impact factor, date of publication, and the country of origin of the paper. Such increased citation is also likely to further enhance the impact of the resources used in the original research.

Reduces Subject Burden

By sharing data and thereby the need to repeat experiments, one is able to reduce number of research subjects needed, whether humans or laboratory animals (Sieber 1989; Hedrick 1998). This is an ethical imperative. It also may be an economic necessity and even a practical one—certain types of human and animal subjects are rare.

Facilitates Resource Development

Combining data sets may lead to the development of a database that is more comprehensive than any single laboratory could develop. In such cases, the resulting entity may provide a more complete representation of the phenomena under investigation and may allow for the testing of more sophisticated model systems (Stanley and Stanley; Ball et al. 2004). In addition, existing data sets are valuable resources not only to those working on a related project, but to those in complementary disciplines. For example, such data sets can serve as samples for use in developing new analytical methods and techniques, such as statistical models (Hedrick 1998).

Sharing Enhances the Climate of the Scientific Community

Discourages Fraud and Enhances Confidence

Where sharing is the norm, some types of misconduct are less likely to occur as researchers are aware that someone may try to replicate their work (Hedrick 1998). Moreover, an unwillingness to share may suggest to others that one is trying to hide something (Stanley and Stanley 1988; Sieber 1998).

Promotes creativity

One of the central characteristics of science is a climate of openness, and the sharing of ideas and resources contributes to this atmosphere of transparency (Merton 1973; Kilo et al. 1980; Stanley and Stanley 1988). The open exchange of information can

also lead to the birth of new ideas formed through the cross-pollination of thoughts from two or more individuals (Shekerjian 1991).

Sharing is a Responsibility—and a Requirement

The scientific endeavor is expensive. In 2004, U.S. federal funding for research and development was $132.3 billion. While seemingly small in comparison to the nation's overall federal budget, which was slightly more than $2.27 trillion in that fiscal year, and to the gross domestic product, which was $10.34 trillion, it represented 13.6% of the federal discretionary budget (National Science Board 2006) and is considerably larger than the budgets of the great majority of the world's nations. Moreover, this significantly underestimates what is spent in the United States for research, as additional funding is continually provided by private donors and foundations, and a great deal of research takes place within private companies.

With this enormous investment in research comes an increased need for the funders, whether public or private, not only to require accountability, but also to discourage needless replication of work. Without sharing there will inevitably be unnecessary duplication of effort. In addition, *necessary* duplication—to ensure validity and uncover errors—may be impossible (Sieber 1989; Hedrick 1998; Sieber 1998). Thus, a number of organizations, including funding agencies, professional societies, and publishers, have developed policies on sharing. The outcome of these policies is that increasingly, sharing is no longer simply encouraged. It is required.

For example, the U.S. National Institutes of Health mandates that *"unique research resources arising from NIH-funded research are to be made available to the scientific research community."* (Federal Register 1999) And the U.S. National Science Foundation has a similar policy: *"…[NSF] expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages awardees to share software and inventions or otherwise act to make the innovations they embody widely useful and useable."* (National Science Foundation 2001)

More recently, the U.S. National Research Council issued recommendations for sharing data and resources in the life sciences, entitled *Uniform Principles for Sharing Integral Data Expeditiously* (UPSIDE) (National Research Council 2003). Those recommendations reinforce and extend the policies of U.S. federal funding agencies. In particular, it proposes that (1) researchers should comply with requests to share within 60 days of an inquiry; (2) in cases of non-compliance and after adequate attempts to contact the author, the requestor may ask the author's institution or funding agency to provide assistance in securing the materials requested; (3) institutions and agencies should have procedures in place for dealing with such situations; and (4) sharing policies should apply equally to researchers across academic and corporate organizations alike (National Research Council 2003).

Many journals also indicate in their publication instructions that authors are expected to share data sets and unique resources with other qualified investigators

(e.g., see Society for Neuroscience 2000; American Physiological Society 2007; PNAS 2006). In addition, when a data set—such as protein or DNA sequences, microarray data, or atomic coordinates—is too large to be included in the paper and is necessary for replication of the work, some journals, including Science (2007) and the Journal of Biological Chemistry (2007), require that researchers cite an accession number in their manuscript to document that the data have been deposited into a publicly accessible repository.

What to Share

Data and Results

Indeed, it has often been stated that research is not complete until the results are published (Day 1979), and research results—which are the resource most frequently shared—are typically shared primarily through publication of peer-reviewed papers in widely available journals. The impetus for sharing research data in this way comes not simply from the researcher's general obligation to the scientific community and to funders. Sharing through publication also has a more direct impact on the welfare of the author. It is through such publication of research results that an individual is able to build a reputation within the scientific community, and it is this reputation that is a primary determinant of a researcher's ability to obtain employment and promotions, acquire funding for research, and attract trainees and collaborators.

Other mechanisms through which research findings are communicated include seminars, poster presentations at meetings, and informal discussions between colleagues. These enable a researcher to share their results at the early stages of a project, which provides the investigator with an opportunity to integrate suggestions and address criticisms prior to publications. Thus, it and can be extremely important. Yet, these activities do not replace the provision of an archival record of the results of a project that renders them accessible over space and time by the scientific community.

Regardless of the methods for sharing, one confounding factor is determining which results and data, to share. Not all data may be publishable or worth sharing (National Research Council 2003). Vivian Weil and Rachelle Hollander have commented on the evolving concepts of what constitutes data and the problematic *"tendency to conflate data with good data."* In response to such concerns, they suggest that scientists *"…define data as information which a person who has appreciation of the norms of her discipline will appropriately stand behind"* (Weil and Hollander 2007).

There is another term in this equation: not all useful data are published. There are a number of reasons for this. In some cases, a data set is so large that full publication in a journal is not practical. In addition, authors may be required by journals to leave some of their data out of their manuscript even when all of it was considered in drawing conclusions. Some journals offer space on their web sites so that access to such supplementary data may be provided. Another example of an incompletely

published data set might include the subset of data selected by an author to provide a simpler or more focused report. Also, many failures to replicate a published finding never appear in print. In each case, data that required considerable resources to collect may be forever lost to the scientific community.

There are a number of data repositories that provide broad access to large data sets, such as protein sequences (e.g., Genbank 2005; Brookhaven Laboratories 2007), archaeological data (e.g., Archaeology Data Services 2007), and astronomical images (e.g., the NASA/IPAC Infrared Images Science Archive (NASA/IPAC) 2007). The NIH has recently compiled a document that provides a great deal of valuable information on data sharing, as well (National Institutes of Health, Data Sharing Workbook 2004).

Methods

The Methods section of a research article serves a number of functions: It allows readers to evaluate for themselves the techniques used in the work. For example, with sufficient information about what was done, readers should be able to assess for themselves whether the specificity and sensitivity of a method justified the authors' conclusions regarding the published results. The Methods section is also designed to enable others to confirm and extend the work published.

Given the functions of the Methods section, it is essential that authors share with their readers all the details necessary to allow others to replicate their findings. If the method has been previously described in the peer-reviewed literature, it is not necessary to republish that information in whole, but rather to cite the previously-published method and provide enough information so that the reader can understand the type of method used and be alerted to any methodological modifications that were made. In this regard, it is important for authors to describe what they actually did to collect *these* data, rather than what they have done in the past. Senior authors may not always be aware of seemingly minor modifications made by their co-workers and are well advised to work closely with those most intimately involved in the data collection.

Assuming that the method has not already been published, what sort of information should be included in the description? These details vary greatly depending on the type of work performed. A general guideline is that anything critical to performing work should be reported. This might include details on the subjects, key pieces of equipment (company, model number, and company location) and the settings used, chemicals used (reagent, catalogue number, company, and location), the sources of any non-commercial reagents or samples, the timing of the procedures, and the methods of analysis.

In reality, it is a rare paper that provides all of the details of how the experiments were carried out. This may be the result of the author's forgetfulness or sloppiness in writing. In some cases, the omission of critical details may even be an intentional act on the part of the author for the purpose of hampering the progress of competitors. However, this would be a short-sighted tactic because replicability is critical to the acceptance of results by the scientific community. Indeed, findings that cannot be replicated can damage a researcher's reputation On the other hand, a

method that is broadly disseminated can bring recognition if it becomes widely used.

In other cases, a lack of detail in the methodology may result from limitations that the publisher has placed on the length of the Methods section. Some journals have recently developed a remedy for this by allowing authors to post supplementary information (methods and/or additional data) to the journal's website. This is a useful approach for dealing with the space limitations of print journals, however readers should realize that, whereas some journals subject all supplementary information to peer-review (e.g., Science 2007; Journal of Biological Chemistry 2007), others do not (e.g., Addiction 2007; Genes, Brain and Behavior 2007).

Some investigators opt to post their protocols and/or data sets to their personal websites, which they reference in their research articles. In general, this mechanism for sharing is less preferable than posting the information to a journal's website, as personal websites are less likely to be permanent, are not always reliably maintained, and may not provide archival versions of the files. Indeed, some have argued that even journal websites should not be used for this purpose and, instead, information should be stored in public repositories that have tools for identifying and querying relevant data sets (Santos et al. 2005).

Unique Research Resources

Given the importance of being able to replicate and extend work published in the scientific literature, it is essential that researchers also share any unique resources developed in the course of the work if those resources are not commercially available. Researchers may choose to replicate and distribute these materials themselves. Alternately, there are a number of national and international repositories to which researchers may submit their resources. Such organizations take responsibility for maintaining, replicating, and supplying the item to other investigators. These include The Jackson Laboratories, a repository for transgenic mice, (Jackson Laboratories 2007) and American Type Culture Collection, which maintains and distributes cell lines (American Type Culture Collection 2007).

Barriers to Sharing

Despite all the benefits that can be derived from sharing, individual scientists may not be immediately rewarded for sharing their resources (Hedrick 1998). Indeed, given that they incur the costs of sharing, such individuals may feel that it is the recipient researchers who are the beneficiaries and that the originators lose out in the transaction (Ceci and Walker 1983), In fact, this may not be an accurate perception. As noted above, sharing can increase citation rate which, in turn, is likely to improve the originating author's reputation and may lead to increased funding, access to more trainees, and more rapid advancement. Nonetheless, it should not be surprising that, whereas most researchers will acknowledge publicly that sharing is the correct thing to do, their actual behavior may be inconsistent with that position. As a case in point, in a recent survey of geneticists at 100 U.S. universities, 47% of

respondents indicated that at least one of their requests for additional information or resources related to a published article had gone unfulfilled (Campbell et al. 2002).

There are many ways to avoid sharing, such as ignoring requests, agreeing to share but never following through, or providing some explanation of why it is impossible. Rather than simply crediting all failures to share as malevolent, however, it would be more useful to recognize legitimate barriers to sharing and then determine if they can be overcome.

Negative Career Impact

Need for Publications

Some resources may take a great deal of time to develop. For example, it may take a year or more to develop a transgenic animal or a new methodology. Given the importance of publication to a scientist's career, individuals are likely to invest such effort in developing a resource only when they feel that they can expect to realize multiple publications from it. If investigators are required to share a resource immediately upon the first publication in which they cite that resource, they may be less likely to pursue such long-term endeavors, because they may not feel that they will have the opportunity to adequately profit from their effort.

Junior scientists face a special problem: Imagine a graduate student or postdoctoral fellow investing 2 years in the development of a transgenic mouse only to have Dr. Megalab request that mouse immediately after the first publication. Megalab may have interests in the mouse that are similar to those of the junior investigator as well as a large research group ready to carry out the relevant experiments. Filling the request may dramatically reduce the opportunity for the junior investigator to generate additional publications from his or her efforts that, in turn, may affect his or her chances of landing a job, a grant, or both.

There are several approaches that might protect the sharers' interests by limiting the disincentives to sharing. For example, authors might be allowed a window of time between publication and dissemination of data or reagents. Another approach would be to use a Material Transfer Agreement (MTA) designed to limit the use to which another individual can put a shared resource. A third approach would be to delay publication until the investigator has developed a series of papers (though this opens up the possibility that the author may in fact be "scooped" during this period of delay).

Unfortunately, whereas each of these options may appear to protect the originator's interests, they also run counter to the most basic objectives of policies on sharing—to maximize the efficiency of resource utilization and to promote collegiality within science. Moreover, in the case of MTAs, just the complexity introduced by the varying requirements of these documents and of the institutions involved and the time required to negotiate them can be cited as a barrier to sharing (National Science Board 2006). Indeed, we know of instances in which institutional constraints blocked the sharing of materials despite the willingness of the authors to do so.

Potential for Non-Replication

Some individuals are reluctant to share because they are concerned that their work will not be replicable. This is not necessarily a reflection of fraud or lack of rigor in experimentation, but because of worries that the recipient might not have the requisite skills and qualifications necessary to use the materials appropriately (Hedrick 1998). This is a particularly relevant concern when sharing complex sets of raw data (Chicurel 2000) or a reagent whose use requires special methods. Indeed, in such cases, it can often be to the advantage of both the requestor and supplier of the data set to work in collaboration.

Lack of Recognition and Increased Competition

Currently, in many venues, scientists are not recognized for sharing their resources and data sets. Instead, it is peer-reviewed research articles that are considered in terms of grants, promotions, and recognition. If anything, sharing can be seen as disadvantageous as it increases the competition (Hedrick 1998; Theologis and Davis 2004). Indeed, sharing can provide others with a head start on publication by eliminating the need to collect the data or construct the reagent.

Limited Resources

Financial Cost

Another reason that an investigator might hesitate to share is that it reduces the time and money they have to spend on their own research activities (Hedrick 1998). Sharing resources requires resources. For example, extra effort or money may be necessary for converting the data set into a form that is intelligible to other researchers, e.g., developing a codebook that is much more detailed than one might otherwise develop, (Hedrick 1998) scaling up the production of the reagent, or learning the software necessary for depositing data to a central repository (Ventura 2005). Indeed, some authors have estimated that it may take several weeks for a novice to successfully deposit experimental data from approximately 30 microarrays into a data repository (National Science and Technology Council 2000).

One might think that the cost of sharing a unique product could be covered by charging the recipient of the item. However, in many cases, funding to support sharing is needed prior to anyone accepting requests for the material. For example, researchers may need to hold the data set in a readable format despite changes in computer technology, or maintain a transgenic mouse colony after its usefulness to the investigator has been completed. Researchers may not have built into their budgets funding to support such sharing efforts. Moreover, even when money is requested in a grant application to support the activity of sharing, granting agencies tend to under-fund such activities (National Science and Technology Council 2000), and seldom if ever fund such activities beyond the project period of the grant.

Personnel Required

Experienced personnel may be needed to prepare the material for distribution. In principle it may be only the data set, details of the method, or a specific resource that are shared. However, given the potential impact that a failure to replicate could have on the reputation of the original investigators, they may feel obliged to train the recipient in the use of the data or other resource, or to assist with troubleshooting. Such training would also likely require the time of skilled personnel involved in ongoing research in the investigator's laboratory.

Availability

Once investigators complete their work with a given resource there may be little reason to maintain the resource. For example, maintaining colonies of genetically engineered animals after they are no longer in use by an investigator generates costs that are difficult to justify or to cover. If a third party (e.g., Jackson Labs) cannot be engaged in the process of storage and distribution, such types of animals may not be available for very long after the original researcher has completed his or her work with them.

Intellectual Property Rights

There are specific situations in which authors may have commercial and/or legal reasons not to share. For example, many reagents, processes, and inventions have commercial potential. In cases in which patents are sought, it may be essential to limit disclosure of the resource prior to filing a patent application. In such cases, however, we feel it incumbent on the author of a manuscript to clearly indicate to the editor their intention *not* to share certain information or material. The editor can then decide whether to proceed with the review process or return the manuscript to the author without review.

A second such example relates to MTAs. As noted above, in obtaining unique resources from a colleague, an individual may be required to sign an MTA restricting them from sharing it with others, and thus requiring any future requestors to approach the inventor of the resource directly if they also wish to obtain the material. This can pose a difficulty for the initial recipient of the resource if third parties subsequently request the item from them following publication as it adds a level of complexity to the process of sharing and may restrict the use of the materials, if and when they are received. We believe that the use of MTAs should be minimized and, when they are used, be limited to an understanding of the rights to commercial gain rather than to impose limits on research.

Ethical and Legal Issues

Subject Confidentiality

It is essential that subject confidentiality be maintained in any sharing transaction. However, in some cases, it may not be possible to maintain the value of the data set

while stripping the data of all personal identifiers. Moreover, although a single data set may not identify a given individual, it may be possible to identify that individual through the use of multiple types of data sets (Nature Neuroscience 2000).

Informed Consent

Material from human subjects (such as survey data and blood samples) collected for one purpose cannot necessarily be used for another purpose without obtaining additional consent from the subjects themselves. Investigators who wish to make use of such material must carefully examine the original consent form and, if necessary, obtain additional permissions from the original subjects (Sieber 1989).

National Security

Government regulations may prohibit researchers from sharing some types of information or materials (e.g., pathogens, missile technology, and encryption software) because of security concerns (Kilo et al. 1980; Bok 1982). Even genome data have been cited as having the potential for dual (i.e., destructive) use, as such data could be used to facilitate the development of more virulent pathogens (National Research Council 2004).

International Transport

It is not possible to share all types of research resources across international borders. This may be due to national security concerns, or in some cases, because of a government-imposed embargo against a specific nation. Moreover, even if export/import regulations of the countries involved permit the sharing of a given resource, an exceptional effort may be required to complete the transaction. Not only may there be many forms to complete, but given the perishable nature of some resources, long delays in customs offices can render the material useless.

Ways to Facilitate Sharing

Given the great potential for the scientific community—and eventually the general public—to benefit from the sharing of research resources, it seems essential to develop ways to overcome barriers to sharing. This will require effort on the part of the individual researchers, their institutions, and funding agencies, as well as a change in attitudes within the scientific community.

Plan Ahead

One of the most valuable ways for the scientist to facilitate sharing is to design research projects assuming a need to share the results as well as data sets. That is, sharing becomes part of research design, because it is at this stage that specifications/protocols for sharing can be established and integrated most easily

into the research plan. Key activities are to (1) anticipate the need for sharing data when developing informed consent procedures and requesting IRB approval for studies (Governing Council of the Organization for Human Brain Mapping 2001), (2) structure the data collection process so as to facilitate sharing and, if necessary, to format the data to conform to the requirements of a given data repository, and (3) ensure co-development of documentation necessary for use of the data. Furthermore, the U.S. National Research Council recommends that (4) scientists anticipate the data or resource their readers are likely to request, and (5) authors include information on how to obtain that material in the text of their research articles (National Research Council 2003).

Explore Opportunities for Collaboration

Rather than avoiding sharing data or reagents, or trying to constrain their use, the originator of the data set could try to develop requests for research resources into opportunities for collaboration, seeking informal understandings regarding the experiments to be done by the originating and requesting laboratories. Such an approach—when successful—can convert a potential short term hazard to the originator's research productivity into an advantage for the author, the persons requesting the materials, and the field.

Provide Funding and Expertise

Sharing may require substantial funding and perhaps even experienced personnel such as bioinformaticians to design the databases and to develop the software to deposit, store, search, and analyze data sets (Nature Neuroscience 2000). In addition, dedicated site curators may be needed to assist users with data deposition and queries, ensure quality control, and develop software to accommodate emerging forms of data (Ball et al. 2004). Finally, as others have noted *"No standards for organizing supplementary data collections have been adopted either across journals or even for supplemental data collections associated with articles in the same journal"* (Santos et al. 2005). Without standards, the number of formats and ways of organizing the data can make the use of shared data unwieldy. Thus, it seems incumbent on those who wish to ensure that sharing occurs to make the necessary financial support and technical expertise available to the individuals with the resources to be shared.

Improve the Culture

There is also much that can be done at the institutional level and by the research community at large to promote sharing. Primary among these is for universities, funding agencies, and reviewers to provide researchers with substantial 'credit' for developing resources and sharing them. Evaluation of candidates for hiring, tenure, promotion, and funding could include documentation of the extent to which data sets and reagents had been shared as well as the results of that sharing, such as published papers, whether or not the researcher is a coauthor. Sharing might also be

facilitated if institutions modified their training programs and courses on research methods to include in the curriculum information on sharing research data and resources as well as results. The stimulus for such changes could come from those agencies and professional organizations wishing to ensure that resources be shared. Additional impetus could come in the form of editorials in leading journals together with policy recommendations that are widely distributed.

Enforce Guidelines, While Avoiding Legal Remedies

In addition to providing incentives and instruction for sharing, organizations should enforce policies on sharing. Without such efforts, compliance may suffer and the directive may devolve into more of an aspirational goal than an essential part of research practice. As noted above, the U.S. National Research Council has recommended that institutions and funding agencies develop and enforce procedures for addressing non-compliance with sharing policies (National Research Council 2003). Time will tell whether its proposal will be adopted by the community.

There are many reasons why sharing is good for science as well as reasons why it can be difficult and problematic. Nevertheless, sharing is essential; it is both a responsibility and a requirement. We hope that the community of science will continue to overcome the barriers and move forward—together—with even greater efficacy.

Acknowledgments Funding for the Survival Skills and Ethics Program, which Drs. Fischer and Zigmond oversee, has been provided by the National Institutes of Health, the National Science Foundation, the Office of Research Integrity, and the University of Pittsburgh.

References

Addiction (2007). Guidelines for authors on acceptable materials for submission. www.blackwellpublishing.com/bauthor/suppmat.asp. Accession 9 July 2007.

American Physiological Society (2007). Instructions for preparing your manuscript. http://www.the-aps.org/publications/i4a/prep_manuscript.htm. Accession 9 July 2007.

American Type Culture Collection (2007). About ATCC. http://www.atcc.org/About/AboutATCC.cfm#mission. Accession 10 July 2007.

Archaeology Data Services (2007). About the ADS. http://ads.ahds.ac.uk/project/about.html. Accession 10 July 2007.

Ball, C. A., Sherlock, G., & Brazma, A. (2004). Funding high-throughput data sharing. *Nature Biotechnology, 22*, 1179–1183.

Bok, S. (1982). Secrecy and openness in science: Ethical considerations. *Science, Technology, and Human Values, 7*, 32–41.

Boorstein, D. J. (1983). *The discoverers* (p. 1983). NY: Random House.

Brookhaven Laboratories (2007). RCSB protein data bank. http://www.rcsb.org/pdb/home/home.do. Accession 10 July 2007.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., et al. (2002). Data withholding in academic genetics: Evidence from a national survey. *Journal of the American Medical Association, 287*, 473–480.

Ceci, S. J. (1998). Scientists' attitudes toward sharing data. *Science, Technology, and Human Values, 13*, 45–52.

Ceci, S. J., & Walker, E. (1983). Private archives and public needs. *American Psychologist, 38*, 414–423.

Chicurel, M. (2000). Databasing the brain. *Nature, 406*, 822–825.
Day, R. A. (1979). *How to write and publish a scientific paper*. Phoenix, AZ: Oryx Press.
Federal Register. (1999). Principles and guidelines for recipients of NIH research grants and contracts on obtaining and disseminating biomedical research resources: Final notice. *Federal Register, 64*, 72090–72096.
Flournoy, N., & Hearne, L. B. (1990). Sharing scientific data III: Planning and the research proposal. *IRB: A Review of Human Subjects Research, 12*, 6–9.
Genbank (2005). What is GenBank? http://www.ncbi.nlm.nih.gov/Genbank/. Accession 10 July 2007.
Genes, Brain and Behavior (2007). Guidelines for authors on acceptable materials for submission. http://www.blackwellpublishing.com/bauthor/suppmat.asp Accession 9 July 2007.
Governing Council of the Organization for Human Brain Mapping. (2001). Neuroimaging databases. *Science, 292*, 1673–1676.
Hedrick, T. E. (1998). Justifications for the sharing of social science data. *Law and Human Behavior, 12*, 163–171.
Jackson Laboratories (2007). JAX services information. http://jaxmice.jax.org/. Accession 10 July 2007.
Journal of Biological Chemistry (2007). Instructions for authors. http://www.jbc.org/misc/ifora.shtml. Accession 2 April 2007.
Kilo, C., Miller, J. P., & Williamson, J. R. (1980). The achilles heel of the University Group Diabetes Program. *Journal of the American Medical Association, 243*, 450–457.
Koyre, A. (1952). An unpublished letter of Robert Hooke to Isaac Newton. *Isis, 43*, 312–337.
Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Journal, 22*, 635–659.
Merton, R. K. (1961). Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society, 105*, 470–486.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago.
NASA/IPAC (2007). Infrared images science archive: What is IRSA? http://irsa.ipac.caltech.edu/about.html. Accession 10 July 2007.
National Institutes of Health, Data Sharing Workbook (2004). http//grants1.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf.
National Research Council. (2003). *Sharing publication-related data and materials: Responsibilities of authorship in the life sciences*. Washington, DC: National Academies Press.
National Research Council. (2004). *Seeking security: Pathogens, open access, and genome databases*. Washington, DC: National Academies.
National Science Board (2006). Science and Engineering Indicators, 2006 vol. 1. (NSB 06-01), National Science Foundation: Arlington, VA.
National Science Foundation (2001). General Grant Conditions (GC-1). http://www.nsf.gov/pubs/2001/gc101/gc101rev1.pdf.
National Science and Technology Council (2000). Ensuring a strong U.S. scientific, technical, and engineering workforce in the 21st century. http://www.ostp.gov/html/workforcerpt.pdf.
Nature Neuroscience. (2000). A debate over fMRI sharing. *Nature Neuroscience, 3*, 845–846.
Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One, 2*, e308.
PNAS (2006). PNAS information for authors. PNAS, 103, i–iv.
Rogers, F. (2004). Sharing. http://pbskids.org/rogers/parentsteachers/theme/1711.html. Accession 9 July 2007.
Royal Academy of Sciences (2010). Philosophical Transactions of the Royal Academy of Sciences. http://www.pubs.royalsoc.ac.uk/index.cfm?page=1000. Accession 9 July 2007.
Santos, C., Blake, J., & States, D. J. (2005). Supplementary data need to be kept in public repositories. *Nature, 438*, 738.
Science (2007). Information for authors: Preparing your online supporting materials. http://www.sciencemag.org/about/authors/prep/prep_online.dtl. Accession 2 April 2007.
Shekerjian, D. (1991). *Uncommon genius: How great ideas are born*. New York: Penguin Books.
Sieber, J. E. (1989). Sharing scientific data I: New problems for IRBs. *IRB: A review of Human Subjects Research, 11*, 4–7.
Sieber, J. E. (1998). Data sharing: Defining problems and seeking solutions. *Law and Human Behavior, 12*, 199–206.

Society for Neuroscience (2000). Responsible conduct regarding scientific communication. J Neurosci. 20: i–xiv. (Also available as http://www.sfn.org/index.cfm?pagename=responsibleConduct).

Stanley, B., & Stanley, M. (1988). Data sharing: The primary researcher's perspective. *Law and Human Behavior, 12*, 173–180.

Theologis, A., & Davis, R. W. (2004). To give or not to give? That is the question. *Plant Physiology, 135*, 4–9.

Ventura, B. (2005). Mandatory submission of microarray data to public repositories: How is it working? *Physiological Genomics, 20*, 153–156.

Weil, V. & Hollander, R. (2007). Sharing scientific data II: Normative issues. IRB: A review of human subjects research, 1990. 12, 7–8.