Research Data as Intellectual Property: How the Law Influences Data Sharing
A National Symposium and Workshop

Board on Research Data and Information
National Academy of Sciences

SUMMARY

Two of the most significant impacts of digital technologies and networks on scientific research have been the great increase in the amount of data generated or collected by researchers and a tremendously improved infrastructure for sharing, aggregating or recombining data sets. With these changes have also come a series of challenges concerning preservation, authentication, annotation, and provenance. As funders and researchers contemplate methods for responding to these opportunities and challenges, they often find themselves in doubt about the legal framework that governs research data. Are data "owned"? If so, by whom? What rights do the "owners" have against those who would copy, redistribute or reuse the data without permission? Conversely, how can a researcher share data over the Internet and assure other researchers that they have any permission they might need to reuse the data as they wish? In some cases, the law has been used to impede productive data sharing by researchers. In other cases, uncertainty about the law has impeded researcher collaboration even though the law itself would pose no barrier to such collaboration.

A symposium and workshop will be convened at the National Academy of Sciences in Washington, DC to bring together key stakeholders in this area for intensive and structured discussions in order to obtain a better understanding of the ways in which the law of intellectual property and contract affect data sharing. The discussion also will cover the policies that require data management plans for the data that support published articles. Specifically, the project will be performed pursuant to the following statement of task:

1. Opportunities and Benefits of Data Sharing: What are the opportunities over the next 5-10 years for improving data sharing among researchers? What are the potential benefits to science and society of data sharing?

2. Legal Framework for Data Sharing: What rights does the law grant to data generators or their employers or funders? What rights might these parties create by contract? How have these rights been used to impede or promote data sharing among researchers? To what extent does the law provide a right to attribution when another researcher uses a data set? To the extent that it does not, how might a researcher's interest in proper attribution or citation be recognized while also encouraging data sharing?

3. Barriers to Data Sharing: What are the major legal, technological, institutional, sociological, and policy barriers to data sharing in the open online environment within the scientific community? To what extent is uncertainty or misinformation about the law a barrier to data sharing independent of the law itself? What needs to be known and studied about each of these barriers to help achieve the opportunities for interdisciplinary science and complex problem solving?

4. Range of Options: Based on the results obtained in response to items 1-3 above, define a range of options that can be used by the sponsors of the project, as well as other similar organizations, to obtain and promote a better understanding of the legal framework for data sharing. Discussion also will cover whether standardized legal tools, such as those provided by Creative Commons, can improve data sharing by removing legal barriers and clarifying the rights and expectations of researchers who generate and who reuse data. The objective of defining these options is to improve the activities of the sponsors (and other similar organizations) and the activities of researchers that they fund externally in this emerging research area.

A one-day symposium with invited expert speakers will be held to address tasks 1-3 above. This will be followed immediately by a one-day workshop to address task 4, based on the discussions of tasks 1-3 at the symposium and on the expertise of the invitees. A steering committee for the project will help to organize the symposium and workshop, and a rapporteur will prepare a summary report from the workshop, which will be published openly on the Academies' website at the conclusion of the project. The workshop report will synthesize the contributions of the experts and is expected to result in an authoritative and high-level review of the most promising and effective research opportunities in this area. Both the slides from the symposium proceedings and the symposium webcast also will be posted on the National Academies' website.

Intellectual Merit of the Proposed Activity: The proposed activity would highlight and address demand from researchers and research sponsors for relatively simple and standardized rules about data access, sharing and attribution against a legal backdrop that is complex because whether intellectual property rights attach to data or data sets depends upon a series of context-specific criteria concerning whether human creativity is found in the expression of the data or in the selection or arrangement of the data. The proposed activity would deepen knowledge about the underlying complexities while also highlighting opportunities to creatively simplify and standardize the rules through contracts, licenses or informal practices that balance the interests of data-generating researchers and follow-on researchers who seek to use or adapt pre-existing datasets for new, innovative research. In addition, attention will be given to the ways in which policies that require public access to the fruits of publicly funded research are and are not mutually reinforcing with respect to articles and data arising from federally-funded research. Finally, the activity will bring to light the independent value of clarifying and standardizing access rules, particularly when those rules are designed to foster research collaboration and the contribution of data to online commons from which a range of researchers from varying disciplines might draw.

Broader Impacts of the Proposed Activity: A better understanding of the ways in which legal rights or claims associated with research data do or might promote or impede data sharing can yield broad benefits not only to our nation's research base, but to our economy and society. Within scientific research, such improved understanding can be used to make better decisions about data management and investments, organizational models, and research management, collaboration and policy, particularly for interdisciplinary studies. A deeper understanding of the opportunities and barriers to simplifying and standardizing access and usage rules for data has the potential to accelerate the progress of scientific research, to support U.S. national competitiveness and increased productivity in information-intensive areas of research and its applications. An improved understanding of these issues also can enable research managers and policy makers to make much more informed decisions about the research enterprise, and to

explain more clearly to policymakers and to the general public how the public investment in research and digital technologies advances broader socioeconomic interests.