**Module 5: Legal and Ethical Considerations for Research Data**

**Authored by Donna Kafel, Lisa Palmer, and Mary Piorun of the University of Massachusetts Medical School and by Lynne Riley of Worcester Polytechnic Institute**

Learner Objectives:

1. Explain ownership considerations related to data sharing

2. Explain and evaluate potential legal issues connected to one’s data

3. Explain ethical considerations related to data sharing

4. Recognize the importance of privacy with some forms of data

(HIPPA)

5. Understand the importance of removing key identifiers.

**Who owns the data?**

Data ownership can be very complex. There are institutional policies that can help ascertain who owns data produced at an institution and under what circumstances someone may take data, share data, or publish data. Data is seen as assets and therefore research stakeholders (institutions, funders, scientists, etc.) want to protect their intellectual property rights and there are issues to be aware of such as licensing. Before starting a project you should have an idea of the ownership issues related to any resulting research products, including data.

This question “who owns data?” can pertain to the individual or entity that has the legal rights to the data and can retain the data after the completion of the project, etc.

Ownership of the data can also depend on who funds the research. Funders sponsor research for a variety of reasons:

* Government agencies fund research to improve the general health and welfare of society
* Philanthropic organizations are interested in advancing particular causes
* Private funders are interested in profits, along with benefits to society

These different reasons often determine who claims ownership of research data.

**Federally funded grants**

In most cases for federally funded research, the government gives the research institution the right to use data collected with public funds as an incentive to put research to use for the common good (the Bayh-Dole Act). Thus the research institution owns the data but allows the principal investigator on the grant to be the steward of the data. The PI may control the course, publication, and copyright of any research, subject to institutional review. Graduate students, postdocs, or faculty involved in performing research on a particular grant would therefore be wrong to assume that they own the data that they are collecting. The PI takes responsibility for the collection, recording, storage, retention and disposal of data.

Data and lab notebooks collected by undergraduate and graduate students and research fellows for a research project belong to the grantee institution. Students should not take the data with them when they leave the institution unless they have made appropriate arrangements with the project PI.

Retaining copies of data might also be allowed, with permission.

When the PI faculty member leaves the grantee institution, they must negotiate with the institution to keep their grants and data. Many universities have offices and policies in place to ensure that such a transfer of data respects both the rights of the researcher and those of the institution(s).

**Is it a grant or a contract?**

With government funding, researchers should also distinguish between grants and contracts. Under grants, researchers must carry out the research and submit reports, but control of the data remains with the institution that received the funds.

With contracts, the researcher is required to deliver a product or service, which is then usually controlled by the government. If your research is supported with government funds, make sure you know whether it is a grant or a contract. This is a significant difference that could determine who can publish and use your data.

**Private funding companies**

Private funders seek to retain the rights for commercial use of the data.

**Philanthropic organizations**

Their policies can vary. Depending on their interests, they may retain or give away ownership rights.

As you see, ownership claims do vary from one funder to another. Therefore it is crucial that researchers be aware of their obligations to their funders before they begin collecting data.

**Reading:**

See Guidelines for Responsible Data Management in Scientific Research

<http://ori.hhs.gov/education/products/clinicaltools/data.pdf> pgs. 6-8

**Data and intellectual property**

When preparing for a research project involving data, be sure to evaluate all the legal issues: intellectual property, copyright claims, licenses needed for use, monetary charges for data and other intellectual property issues.

Also consider the different types of outputs within a research project:

· Protocols

· Description of methodologies

· Data sets

· Publications

· Software code

· Description of instrumentation/inventions

**Intellectual property overview**

According to the World Intellectual Property Organization (WIPO), intellectual property is defined as, “creations of the mind: inventions, literary and artistic works, and symbols, names, images, and designs used in commerce.” So in other words, intellectual property is basically an intangible asset. Intellectual property generated in an academic setting usually involves copyrights, trade secrets, and patents.

The creation of intellectual property is one of the expected outcomes of research conducted at universities. It benefits both the university and society to facilitate the development of these discoveries and ideas as well as to assure their availability to the public. With these goals in mind, universities develop policies and procedures relating to the ownership, use, management, and compensation for intellectual properties created with their resources. Because Intellectual Property Policies vary by institution, be sure to familiarize yourself with your institution’s policies. A few sample policies are listed here:

<http://www.umassmed.edu/uploadedFiles/otm2/Policies_and_Procedures/Intellectual_Property_UMW.pdf>

<http://www.bc.edu/research/osp/policies/intproppoly/>

<http://www.wpi.edu/offices/policies/intell.html>

Data can be licensed so you need to think about data licensure from two sides; i.e., as a creator of data and as a user of others’ data. For your research project, you will need to articulate how you will be providing permissions or licensing to your data or copyrighted works from your research project. For others’ data, you would need to obtain the appropriate permissions and cite the data appropriately.

Factors you may want to consider are:

* attribution
* notification regarding its use
* redistribution
* quality control
* risk

In cases where government funded research data is protected by intellectual property rights, rights holders should facilitate data access for the benefit of public research. As the National Science Foundation (NSF) states:

“Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.”

The NSF also now requires a data management plan and addresses intellectual policies on their web site: <http://www.nsf.gov/od/ogc/intelprop.jsp>

For further information, see the [NSF’s](http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp) [Frequently](http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp) [Asked](http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp) [Questions(FAQ)](http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp)site on data sharing [http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp#6](#http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp)

**Copyright**

A copyrightable work is an original creative work set in a tangible format that is covered by the copyright laws of the United States or other countries. Copyright protection is available for most literary, musical, dramatic, photographic and other types of creative works--including research articles, research monographs, textbooks, student theses, and dissertations, still images, computer software, teaching materials, multimedia works, proposals, and research reports. Copyright is “format blind” – that is, print and digital works are eligible for copyright protection; content on the Internet may be protected by copyright.

Copyright ownership is secured automatically when an original creative work is fixed in a tangible format; the © is no longer required. Ownership may belong to the authors/creators or their employer.

The copyright owner has the exclusive rights over the work to reproduce it, distribute copies, create a derivative work, or perform or display the work publicly. Ownership rights may be transferred in whole or in part. In the past, authors often transferred all of their rights to their works to journal and book publishers.

Facts and data are not copyrightable, but compiled databases and derived data can be. While creators of these types of work may seek exclusive rights, there is a less stringent copyright license that strikes a balance between "all rights reserved" language and terms that enable open access and reuse of data, e.g. a Creative Commons license (<http://creativecommons.org>).

Not all works are eligible for copyright protection: original works that are not fixed in a tangible format; titles, names, slogans; ideas, facts, data; lists of contents or ingredients; works in the public domain. However, trademark or patent laws may apply in some cases.

**Creative Commons (CC)**

Creative Commons licenses (<http://creativecommons.org>) are legal tools that data authors can assign to their data if they would like to allow the public to share and use their data while adhering to certain provisions. As noted in the Creative Commons site, "Creative Commons licenses let you easily change your copyright terms from the default of “all rights reserved” to “some rights reserved.” By easing some of the restrictive language of traditional copyright, Creative Commons licenses have facilitated open sharing and reuse of data. There are several types of Creative Commons licenses. The CC0 (Creative Commons Zero license) is strongly recommended for scientific data and has been widely adapted for releasing data into the public domain. In contrast to CC licenses that allow copyright holders to choose from a range of permissions while retaining their copyright, CC0 empowers yet another choice altogether– the choice to opt out of copyright and database protection, and the exclusive rights automatically granted to creators– the “no rights reserved” alternative.

Scientists who apply CC0 licenses to their databases agree to waive any copyright interests in the work and place them as completely as possible in the public domain, enabling others to freely build upon, enhance or reuse the works for any purposes without restriction.

For more details about Creative Commons licenses and data, see [CC and data[bases] http://creativecommons.org/weblog/entry/26283.](https://http://creativecommons.org/weblog/entry/26283)

**Patents**

The [United](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) [States](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) [Patent](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) [and](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) [Trademark](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) [Office](#http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp) (USPTO) defines the patent for an invention as, “the grant of a property right to the inventor” to ”exclude others from making, using, offering for sale, or selling the invention in the United States or importing the invention into the United States.” Patents are usually granted for twenty years but the term may be extended. U.S. patents are effective only within the United States, U.S. territories, and U.S. possessions. Because a patent may be challenged at any time during its twenty year term, it is important to preserve the related data for at least the term of the patent. Good data management practices could also provide more efficient problem resolution if a patent official discovers any data irregularities while evaluating a patent application.

Whether the source of the funding is federal or private, there are likely to be certain obligations regarding intellectual property, especially with relation to inventions and patents. These may be mandated by law, by contract, or both. Prior to 1980, inventions that resulted from federally funded research grants and contracts were under the control of the federal government. However, since the passage of the Bayh-Dole Act in 1980, universities, small businesses, and non-profits may choose to retain title to inventions developed with federal funding.

**Trade secrets**

Trade secrets are generally confidential commercial information such as formulas, manufacturing processes, or compilations of information, which are automatically, protected without any formal registration procedures (e.g., formula for CocaCola©). Trade secrets are generally protected under state law.

**Open source software**

Open source software is computer software often developed in a collaborative manner. The source code is made widely available through a type of license that allows users to freely modify, improve, and redistribute the software as long as they agree to the conditions specified in the license provided. Before agreeing to an open source software license, make sure that:

* your funder/sponsor agrees to the conditions of use
* the conditions do not adversely impact your intellectual property rights

**Research data sets and databases**

The U.S. Federal Government's Office of Management and Budget Circular A-110 (36.d.2.iProperty Standards; Intangible property; definition) states:

*Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include: (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.*

An important point to consider is that in the United States, while data and facts cannot be copyrighted, creative expressions of data, such as a chart or a table in a publication ARE copyrightable. In addition, be aware that in certain foreign jurisdictions such as the European Union, database compilations, including factual data, ARE protected by law.

Databases are generally protected by copyright law and are referred to as “compilations”. The U.S. Copyright Act defines a compilation as a “collection and assembling of preexisting materials or of data that are selected in such a way that the resulting work as a whole constitutes an original work of authorship.”

The individual facts or data contained within the database may or may not be protected by copyright; however, the selection and/or arrangement of the facts or data as a whole will be protected by copyright if it contains enough creative, original expression.

With only limited protection through copyright law, database developers generally protect their databases by using a legal contract, such as a license, so that users must comply with wishes of the copyright owner as to how that data may be accessed and used.

For legal responsibilities related to sharing data please refer to module 6.

|  |
| --- |
| ACTIVITY  Read scenario “Who Owns Research Data?” (Case scenario of graduate student wanting to take data, case study from Columbia University Responsible Conduct of Research Data Acquisition and Management), discuss five follow-up questions.  [http://ori.dhhs.gov/education/products/columbia\_wbt/rcr\_data/case/index.html#2](#http://ori.dhhs.gov/education/products/columbia_wbt/rcr_data/case/index.html) |

**Reading:**

“Constructing Access Permissions”, University of Oregon Libraries: [http://libweb.uoregon.edu/datamanagement/sharingdata.html#three](#http://libweb.uoregon.edu/datamanagement/sharingdata.html)

**How long do I retain data and related documentation?**

Data retention refers to the length of time one needs to keep (or securely destroy) a project’s data according to the institutional (IRB), funder, state and/or federal guidelines. Data retention, like data ownership, can be very complex so researchers should consult their IRB and institutional policies regarding the retention and secure destruction of data. Data retention also depends on a number of overlapping agencies and on factors like whether or not the data contains patient health information (HIPAA) or other identifiers. From a legal and ethics perspective, researchers would want to know what the minimum years for retention would be and how long they would need to retain data documentation for any audits or misconduct investigations. For example, while the OHRP federal guideline (45 CFR 46) may mandate data be retained for a minimum of three years, many institutional policies require 5 years of retention, and many institutions require that research data records in support of shared patents with researchers be retained for the life of the patent. Other legal issues related to this include guidelines for storage, back up and security—for more information please visit module 4.

**Ethics and data**

Any research institution (e.g. university, hospital, private research company, and soon) that accepts federal funding is required by law to have policies in place to oversee its research programs. These policies include monitoring conflicts of interest, reporting misconduct, and ensuring safety regulations are followed. They also establish standing committees to review human (Institutional Review Board) and animal (Institutional Animal Care and Use Committee) research protocols. (i.e.IRB and IACUC)

The purpose of an IRB is to protect the rights and welfare of those individuals who contribute to the research process by participating as subjects. The IRB also protects the institution and the researcher by ensuring that those individuals considering being part of a research study are adequately informed before consenting to participate, and that participants are not exposed to excessive risk.

In the context of data management the IRB has three roles. First, since funders often now ask to see data management plans, members of the IRB look more closely at these plans to see if adequate thought has been given to the plan and if what is written is feasible (e.g. cost, infrastructure, staffing). Second, the IRB reviews data collection forms to limit the amount of personal identifiable information that is being collected. Third, the IRB reviews the research protocol to see how the data will be safeguarded. This includes documenting who will have access to the data collected, and under what conditions, sometimes called the privacy or confidentiality rules. These rules need to consider who will have access to the data technically, physically and for administrative purposes.

There are federal and state rules and regulations regarding data security for specific types of data. For instance, personal identifiable data, such as names and social security numbers, are protected by many state and federal laws. At the federal level, health data are protected by the federal Health Insurance Portability and Accountability Act (HIPAA), student data are protected by the federal Family Education Rights and Privacy Act (FERPA), and financial data are protected by the federal Financial Services Modernization Act (FSMA).

As researchers work to collect and analyze data they must ask themselves if each piece of data is necessary to address the original research question or hypothesis and if the data element in combination with other data could identify an individual. For example, age alone may not identify a person, but age in conjunction with zip code and medical condition may lead to identification. To protect confidentiality in these instances, researchers should not collect the data at all, or if it is crucial, should substitute the actual data with codes known only to the primary researcher. The HIPAA Privacy Rules outline 18 data elements that need to be coded or removed (<http://healthcare.partners.org/phsirb/deidinfo.htm>).

**Privacy levels required by funding agencies and publishers**

Each funding agency and publisher has guidelines for maintaining privacy regarding human and animal subjects, as exemplified in this guideline from the National Institutes of Health (NIH):

*Data should be redacted to strip all individual identifiers, and effective strategies should be adopted to minimize risk of disclosing a participant's identity. Options to protect privacy include: withholding part of the data, statistically altering the data in ways that will not compromise secondary analyses, requiring researchers who seek data to commit to protect privacy and confidentiality, and providing data access in a controlled site, sometimes referred to as a data enclave. Some investigators use hybrid methods, releasing a redacted dataset for general use but providing access to more sensitive data through a user contract or data enclave. In most instances, sharing data is possible without compromising participant confidentiality and privacy.*

Source; NIH’s Office of Extramural Research: (See <http://www.grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm>)

Readings

1. National Institutes of Health. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule. <http://privacyruleandresearch.nih.gov/pr_02.asp>
2. Clinical Tools, Inc. Office of Research Integrity. (2006). Guidelines for Responsible Data Management in Scientific Research.  
   <http://ori.hhs.gov/education/products/clinicaltools/data.pdf>  
   pgs. 6-8
3. Columbia University. Responsible Conduct of Research. Who Owns Research Data? [http://ori.dhhs.gov/education/products/columbia\_wbt/rcr\_data/case/index.html#2](http://ori.dhhs.gov/education/products/columbia_wbt/rcr_data/case/index.html%232)
4. University of Oregon Libraries. Constructing Access Permissions. [http://libweb.uoregon.edu/datamanagement/sharingdata.html#three](http://libweb.uoregon.edu/datamanagement/sharingdata.html%23three)
5. Ball, A. & Duke, M. (2012). How to Cite Datasets and Link to Publications. DCC How-to Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
6. Altman, M. & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. D-Lib Magazine, 13(3/4). <http://www.dlib.org/dlib/march07/altman/03altman.html>
7. Green, T. (2009). We need publishing standards for datasets and data tables. OECD Publishing White Papers, OECD Publishing. <http://www.oecd.org/document/25/0,3746,en_21571361_33915056_42600857_1_1_1_1,00.html>
8. Mayernik, M.S. (2012). Data citation initiatives and issues. Bul. Am. Soc. Info. Sci.Tech., 38: 23–28. doi: 10.1002/bult.2012.1720380508. <http://www.asis.org/Bulletin/Jun-12/JunJul12_MayernikDataCitation.html>