Who "owns" your data?

Tug of war, Kathleen Tyler Conklin, CC BY-NC 2.0

This post was originally published on the University of California Office of Scholarly Communication blog.

Which of these is true?

"The PI owns the data."

"The university owns the data."

"Nobody can own it; data isn't copyrightable."

You've probably heard somebody say at least one of these things — confidently. Maybe you've heard all of them. Maybe about the same dataset (but in that case, hopefully not from the same person). **So who really owns research data? Well, the short answer is "it depends."**

A longer answer is that determining ownership (and whether there's even anything to own) can be frustratingly complicated — and, even when obvious, ownership only determines some of what can be done with data. Other things like policies, contracts, and laws may dictate certain terms in circumstances where ownership isn't relevant — or even augment or overrule an owner where it is. To avoid an unpleasant surprise about what you can or can't do with your data, you'll want to plan ahead and think beyond the simple question of ownership.

This is a long post. Here's a quick roadmap of what's ahead:

1. Instead of starting with ownership, think about rights and responsibilities.

We tend to use the word "ownership" when we talk about data because the word invokes familiar conventions; when we talk about physical property, ownership is synonymous with ultimate control and responsibility. But rights and responsibilities in data are often more granular, and different types of rights and responsibilities might be important to different individuals or organizations.

For instance:

Data sharing. Can — or must — the data be made public? Where? When? In what format? Under what kind of reuse terms — CC0, CC BY, something else?

- This is the area where you're most likely to have to think about **copyright ownership**, discussed below. If data contains copyrightable expression, the copyright holder controls the rights of reproduction, distribution, and licensing those rights to others.
- The public sharing and licensing of data can be controlled by the copyright holder. It can also be controlled by **contracts and policies** (see below), and is increasingly being required by research funders.
- **Note**: even if your data doesn't contain copyrightable expression, your ability to share unredacted data may still be restricted by policies or contracts rather than required by them. It may also be restricted by privacy laws like HIPAA, concerns about endangered species or archaeological sites, or something else. Local offices of research and policy can be a big help in navigating privacy laws and specific policies.

Data access. Who can see the data? Who can use it, and for what purposes? Who can edit it? Who can get a copy? Who can give permission to other people?

> Individual or small group access to data is much more likely to be controlled by policies and privacy laws than by ownership.

Commercial use. Is someone planning on filing a patent claim? Producing software? Licensing the data to other users for a fee? Are there agreements with funders or partners forbidding or requiring commercial rights, or controlling how income is split?

> Ownership may very well be relevant. But so will policies and contracts.

Credit. Does a funder, partner, employer, or data provider require credit in publications arising from the data? In new datasets incorporating data obtained from them? In what format?

> Attribution tends to be controlled by scholarly norms. Sometimes attribution terms are specified in grant agreements, licenses, or other contracts. Knowing whether someone owns data doesn't tell you when or how to cite it.

Preservation. Is there an obligation to maintain the data for others to use, or to request copies of? Where? For how long? In what format? Under what kind of security controls?

> Preservation responsibilities generally come entirely from policies and contracts, and have nothing to do with ownership.

2. If you do need to figure out ownership, be prepared to argue about copyright.

What can you own? In a nutshell, you can own things that the law defines as property. Property includes real estate, objects like cars and computers, and intellectual property like patents.

Research data is often stored on computers or in notebooks — **physical items that someone can own**. Usually it's pretty easy to figure out who owns those objects, but this isn't what most people are talking about when they talk about owning data. It's more likely, *if* they really mean ownership at all, that they mean **ownership of intellectual property rights in the data**, and probably just copyright. There's a great article by Michael W. Carroll in PLOS ("Sharing Research Data and Intellectual Property Law: A Primer") that discusses different types of IP rights in research data. A few highlights:

- **Trade secrets**. "In traditional academic research, trade secrecy is unlikely to be invoked unless a member of a research team decamps to another team with confidential data."

- **Patents**. "Patents are exclusive rights in inventions. An invention is patentable if it is new, useful, and demonstrates an inventive step over what is already known within the relevant field of knowledge. Unlike [other rights discussed here], *patents only arise if they are applied for and granted by a public authority*. . . . Public disclosure of an invention prior to filing a patent application can destroy or impair one's right to obtain patent protection for the invention. However, *most research data are not eligible to be protected as inventions as such*."
- **Special database rights**. "In the EU, certain candidate countries in Eastern Europe, and South Korea, research data may also be subject to a special database right," but this isn't a concern for most researchers in the US. Read more if you're concerned about ties between your research and Europe or South Korea.
- **Copyright**. Determining whether someone owns copyright in research data is complicated. Carroll writes that it's often "more work than it is worth." And yet this is what so many people want to know, either because they want to apply a copyright license or waiver or because they assume ownership determines other control or responsibility issues.

Why is copyright such a complicated issue for research data? Because facts aren't copyrightable, but works of authorship are — and research data consist of one, or the other, or some combination, and often there's room to argue about which.

- **Is your data like a phone book?** Facts aren't protected by copyright, no matter how much work you had to do to get them. And thank goodness: imagine if someone could own the copyright in, say, temperature data. If they could claim copyright's exclusive right to reproduction and distribution of those facts, it would seriously limit other people's ability to publish work on climate change.
  - A particular arrangement of facts might be eligible for copyright protection if that arrangement demonstrates sufficient creativity, but not if the arrangement is something uncreative like chronological or alphabetical order (see this Supreme Court phone book case).
  - Even with creative arrangement, it's perfectly legal for someone else to pull out the underlying facts, rearrange them, and use them in something new. The same rules apply to other things that aren't protected by copyright, like ideas, public domain works for which copyright has expired, or federal government documents.
  - Copyrightability is a big subject, and there's plenty to read depending on what angle you're interested in. Check out the article "My unpublished research was scooped?" for a discussion of the copyrightability of equations and ideas vs. their expressions; to learn more about the relationship between facts and the way they're expressed, see Copyrightability of Charts, Tables, and Graphs; or visit this Bitlaw article on Database Legal Protection for a discussion about databases as copyrightable compilations.

- **Some data looks more like works traditionally protected by copyright**. Maybe your research data consists of photographs you took. Maybe it's research notes that resemble essays. Maybe your data includes software you've written. The less your data is trying to report objective facts about the universe and the more it embodies creative expression, the more likely it is that you hold copyright in it.
- **Maybe there *are* rights, but they're not yours**. Some scholars use a body of films or literature as the basis for their research. Sometimes research is based on data obtained from a group at another institution or from a vendor that licenses it commercially. Research relying on interview transcripts <u>may involve copyrights jointly owned</u> <u>with interview subjects</u>. (This is a separate concern from privacy and confidentiality related to interviews; that's outside the scope of this article, but you can check out <u>this guide from UC Irvine's Office of Research</u> or your local research or policy office to learn more.) The good news about these kinds of copyright and licensing issues is that people are usually aware of them, especially if a contract had to be signed; the bad news is they can be messy and restrictive.

Those are the basic outlines for assessing whether there are copyright interests to worry about. But keep in mind that, in cases where there is copyrightable expression, there may be multiple authors, transferring some of their rights or building layers on top of previous work. Throw into the mix the fact that some copyright owners will be individuals and some will be organizations, because an employer is considered the author of a work made for hire. **"It's complicated" is starting to feel like a serious understatement.**

Let's say, with all of that, you are fairly confident about a) whether the research data you're interested in is eligible for copyright protection and b) who, if anyone, owns that copyright. Good job. **Now what? You still probably don't have all the information you need in order to figure out if you can take a copy of that data with you when you switch jobs, share it in a data repository, or delete it to free up space on your hard drive.** Instead of following automatically from copyright ownership, these things are often controlled by contract or official policies.

3. See if policies or contracts provide support or create obstacles.

Realistically, you're not going to start using phrases like "right to reproduction of the full dataset for commercial and noncommercial purposes" in your average conversation and "owns" may be a reasonable shorthand. But in written documents, particularly when funding, employment, or the value of research to other potential users around the world may be at stake, it pays to be specific. Sit down and think carefully about the rights and responsibilities above and what you want to be able to do with your data; then start looking at documents that could conflict with your plans (or with each other).

Employment contracts and employer policies. The institution where you work may say specific things about whether data can be shared, copied, taken with you, or deleted. It may specify that *patent* ownership lies with the university and *copyright* ownership rests with employees. Or it may just say "data is the property of the university" without addressing what, exactly, is "owned." Check out this interim guidance document from UCLA's Office of Research Policy & Compliance for some examples of specific rights and obligations.

If you can't find a relevant policy at your institution, or the one that you find doesn't answer the questions that are important to you, ask for clarification. In writing. Better yet, ask for the policy to be updated with clearer language. If you don't know who to talk to, skim your campus directory for a name like "Office of Research" or "Policy Office." You can also try the library. They may have people who specialize in data management issues; even if they don't, or don't already know who to ask, they're good at finding out stuff.

Grant agreements and funder policies. More and more government agencies are requiring that the research data they fund be publicly shared. Sometimes they specify where and when to share the data, or what types of confidential information to redact. We have a page that describes some of the basic requirements for US federal government funders, but the best source of information is the funder itself. Some of these requirements may not be general, or public, but written into a specific grant award. Here's some example language from NOAA's Text to be included in Notices of Award and Contracts of projects anticipated to generate environmental data or peer-reviewed publications:

> *Environmental data collected or created under this Grant, Cooperative Agreement, or Contract must be made publicly visible and accessible in a timely manner, free of charge or at minimal cost that is no more than the cost of distribution to the user, except where limited by law, regulation, policy, or national security requirements.*
>
> *Data accessibility must occur no later than publication of a peer-reviewed article based on the data, or two years after the data are collected and verified, or two years after the original end date of the grant (not including any extensions or follow-on funding), whichever is soonest, unless a delay has been authorized by the NOAA funding program.*

Other contracts and policies. Employers and funders are the most common sources for granting or limiting rights or assigning responsibilities for data, but it's a big world out there. Maybe you work in a lab that has its own policies or are part of a larger collaborative that has its own partnership agreement. Maybe you're publishing with a journal that has data sharing requirements. Maybe, as discussed above, you're re-using someone else's data and you had to agree to certain terms to get it. Whatever the contract or policy is, if you've agreed to it — or you work in a group and someone representing your group agreed to it — or it's an official policy and you've agreed to something that incorporates it by reference — it can limit what you can do with your data.

Now what? If no contracts or policies are an obstacle to what you want to do with your data, you're in good shape. If there's no intellectual property in your data, or if there is but you own it, or if you don't but you have permission for what you want to do — ditto. But as you can see, ownership is a pretty small part of the picture.

4. Plan ahead to save yourself some stress.

No one wants to find themselves poring over contracts and policy documents on the verge of publication, a job change, or a funding renewal application. The best way to avoid this is to know up front you want to do with data, write it up clearly, and get it signed off on by your important stakeholders or anyone you're worried might be a roadblock later — say by using a data management plan.

A data management plan (DMP) can help you spot issues related to data before they become problems, and it's a document increasingly required by grant funding agencies. The DMPTool has all kinds of guidance to get you started, as well as funder requirements and sample plans to look at. And it's free.

So if, for example, you decide at the start that you want to post your data to DataONE with a CC0 waiver and you explain this in a DMP you submit to your funder (and you keep a copy of an email where you discuss this with your collaborators and your local office of research or tech transfer), then an issue like "how much copyrightable expression is there to own in this data?" is something you can probably avoid worrying about.

Say it with me: data management planning is my friend.

This post makes a couple of references to Creative Commons tools without going into any details. To hear more, come read the follow up post about attribution, CC licenses, and why CC BY may not work how you expect it to with data.

CDL UC3

View more posts