



# Beyond the Book®



## **Copyright, Open Access, Subscriptions and Permissions: What Editors Need to Know in the New Digital Publishing Environment**

**Presented at Council of Science Editors Annual Meeting  
Pittsburgh, May 4**

WILBANKS: Copyright is one of the most talked-about things and one of the least understood things that you run into in the content industries.

So, because I knew he was going to talk about this, I've sort of focused my presentation a little more on the interaction of the law and the technology on the network.

When Creative Commons, which is the organization I work for, thinks about this, we think less about sort of the aspects of the law and more about the aspects of the network.

So, when you look at, for example – this is the ARPANET logical map from 1977. How did this become the Internet? What was it about the ARPANET design that allowed it to scale to become the Internet?

One of the things that you see when you look around at both the Internet protocols and the Web protocols is that this idea of making and distributing copies and making and distributing derivatives is really encoded in the culture of the technical standards that underpin all the networks we've got. And it's completely at odds with copyright. Completely.

But you have these things called RFCs, requests for comments. The entire Internet comes from this paper, which outlines a standard by which bits get moved around between computers. And copyright wasn't really asserted over these things. It existed but it wasn't really enforced. It was sort of allowed that you could make and distribute copies of these protocols, because what was desired was the maximum number of implementations of the protocol. People wanted their computers to talk to each other.

And the Web sort of recapitulated this, so it wasn't even just the right to make a copy on a Xerox machine of that Internet standard. Tim Berners Lee, who I used to work for, actually technically enabled copying at the very core of the Web. So if I go to a Web page, I can do what's called a view source and I can see this HTML source code for the page and I can make a copy of that, paste it, change what's inside the tags and I'm up on the Web.



# Beyond the Book®



Me and almost everyone I know in 1994 when we all got on the Web wrote our first Web pages this way.

There were competing hypertext systems that enforced strict copyright controls. There's a very famous one called Xanadu, which was actually one of the original hypertext systems and it protected this idea of infinite copyright licensing and nesting and fees like a lot of the other hypertext systems.

But the one that actually scaled was the one that was open. And unfortunately, it's enabled a pretty significant disconnect between the reality of copyright law and the reality of how most of the Web content got built.

So what you sort of ask is – this is called Metcalf's Law and what it says is that computers get more valuable as they sit in a network with each other. And what we've found is that documents get more valuable as they sit in a network with each other as well. But it's actually in this concept of compatible communication that you get to this.

So TCTIP, that network standard, let computers compatibly communicate. HTML on the Web allowed documents to compatibly communicate. But in the end, it's really about devices and pages, not about people and not about knowledge. So when we're talking about science, when we're talking about journal articles, it's a different way to convey knowledge.

One of the problems we have, one of the reasons my piece of the organization exists, is because knowledge isn't really communicating compatibly. So we don't have any network effects for knowledge.

Knowledge can be lots of different things. I asked once at a conference, what do you think is knowledge, and this is sort of the spaghetti that got thrown back at me.

Journal articles and data make a lot of sense, but as we take these things digital and they need things like ontologies and annotations, we have blogs and wikis coming in, no one really knows how this all fits in.

There's an old saying in the free software movement – and I'm not going to even get into whether or not the copyright laws need to change because that's actually not my role here. My role is going to be to talk about organizationally, what we can do with what we have.



# Beyond the Book®



But one of the old sayings in the free software movement is that if you were to wrap the world in Internet and spin it, software moves through the network. So we don't ask why the software moves for the same reason we don't ask why electrons move along in a wire. We call that an emergent property of the system, that electrons move one direction across a wire. You don't ask why electricity works. The question is, what's the resistance of the wire?

And the argument is that the field strength of the intellectual property system is the resistance in the Internet to software moving around. And what you find is that free software emerged as a resistor, basically a way to resist the resistors in this concept.

Now, this had to be done in a way that was legal because it's not very good to have a culture where everyone violates the law all the time, right? I actually don't forward copyrighted content around. We're one of the people – we have to be the most scrupulous in my organization about not breaking copyright.

But what you see is that there's not just copyright resistors when you're talking about knowledge. What you have is a system where legal resistors come together with incentive resistors as well as sort of human capital resistors, and they interlock in a way that actively prevents the emergence of a knowledge network like we have for documents and for computers.

So copyright is only one of these things. Copyright is like any powerful system. It all depends on how it gets used, and in the scientific and journal space – and this is an Elsevier contract, but I'm not doing it just to beat up on Elsevier. This was actually made public, so I praise Elsevier for letting me analyze their contract.

This is part of the general terms and conditions. This was specifically signed by the University of California. This is the contract that launched the open access movement, just so you know.

So as we moved into the digital world, we stopped buying things and we started renting them through licenses. It's the same thing that allows you to sell a used CD but not your iPod songs. It's like being an owner versus a renter in an apartment. You have a lot less rights when you rent than when you own.

So this whole idea of first sale goes away. This whole idea that you can mark up your copy and do stuff with it goes away.

So what's banned here in this sentence is that the subscriber shall not use spider Web crawling or other software programs, routines, robots or other mechanized



# Beyond the Book®



devices to continuously and automatically search an index, any content access to online under this agreement.

So my guy then at the University of California wasn't allowed to build his own search engine on the content and he went and started the Public Library of Science as a result because he was really pissed off.

But basically, this is 50 percent of the biomedical literature that you can't index. You can't go add hyperlinks to it because it doesn't have hyperlinks. It's got citations. You can't build your own markup of it because you've lost that right as part of the licensing process.

So this is one way that we've resisted webification, if you will, of the scholarly literature, and it's a side effect of the licensing, and if I ran Elsevier, I would do this too. I've run a for-profit company. This is shareholder value.

Now, another resistor – and this is where I eat my own dog food. I do believe this is fair use. This to say it's why I've maintained the notice and the contact info of the author.

However, this is what comes up in Google Images when you look for academic incentives, which is basically anything that doesn't take you towards the tenure door takes you towards the McDonald's door.

So even in the absence of copyright, if you don't get rewarded for sharing stuff, you're not going to go to the effort of sharing it effectively. So there is no incentive to go share stuff. The incentive is to keep your data private, keep your lab notebooks private, not send your biological materials around, because that helps you get the next paper, which is how you get graded towards the tenure door.

And if you solve the tenure problem, you have an actual human capital problem. This is a blog post – and I do actually have permission for this one from the author – a blog post about how hard it is to change a hyperlink in an institutional repository where someone that sends an article to your journal may well try to make a self-archived copy of that paper available. It takes an hour to change one URL in the most commonly used institutional repository software.

So if we solved the copyright problem magically and we solved the incentive problem magically, we would overwhelm the poor librarians and still be unable to serve the users.



# Beyond the Book®



So we have these three resistors that are just not set up right and so it comes down to this concept of compatible communication. Our offices are at MIT in the computer science lab and one of the local jokes is that citations between documents, when you compare them to how efficient the Web is, are like the string between two cans.

Because what you really want are hyperlinks that go into the data. You want to be able to order the materials. You want to be able to run the software that's mentioned in an article. All of these things we take for granted on the Web, the ability to right click and order something, the ability to right click and replicate, analyze. You can't do that in the scholarly literature. And it's not any one problem. It's a nested set of problems.

So, the commons is what we envision as an infrastructure layer that makes it easier to begin addressing this. And it's not the complete answer but it's based on the idea that digital property is different from a fundamental level in terms of what you can do with it than physical property.

So, if you've heard of the commons, you've probably heard it in the context of the tragedy of the commons. There was an article in 1968 in *Science* that said, essentially, rivalrous goods will become overgrazed if there aren't property rights.

Chris and I each have a sheep we put on the Boston Common. We both have an incentive to add more and more sheep. You're from Boston, yes?

KENNEALLY: Yes. I can tell.

WILBANKS: Yeah. I've been there for 13 years. I recognize the accent.

But we both have an incentive to keep putting more sheep on the Common, and over time, we're going to destroy all the grass because each of us has a local incentive to add more sheep because we can shed more wool, harvest more milk, harvest more meat. But we don't represent that common interest, and so it gets destroyed.

But it's based on this idea of a rivalrous resource and it's an argument for fences and for property rights. And this has been the sort of traditional way that the commons was dealt with, and it really does reflect an analog view of the world.

There didn't used to be a lot of ways to violate copyright on a massive scale, as Chris very rightly pointed out. So you had the sort of all possible uses which involved sleeping with your head on the book, underlining pages in the book,



# Beyond the Book®



selling the book in a used book store, and the uses that implicated copyright law were a small part of the total possible uses, and fair use is inside there.

I love the point that Chris made, which is that it's a defense, not a right. So Steve at NY's Law School says fair use is the right to call a lawyer, and that's a pretty good point.

But what's changed on the network is that every use on the network makes a copy. And this was completely transformative. You visit a Web page, your browser makes a cache copy. And suddenly there's no rivalrous use anymore and that creates an enormous amount of both benefits and incredible complications in terms of the way that the copyright law works with the network.

And this is a point where I completely agree with Chris, which is ignoring the law doesn't scale. So the TCPIP protocol I showed you ignored the copyright law. The Web copy and edit, paste and view source sort of ignored the copyright law. But that doesn't really scale, especially as we all become creators of content, not just consumers of content. The ability to create and distribute, as was rightly pointed out, is ubiquitous. It's completely democratized.

So for us, what we think of is that the digital commons is a different layer of the network. We showed you TCPIP, I showed you HTML, and the idea is that for knowledge and for cultural works, this is just like those standards, a voluntary, private way to interoperate, if you want to interoperate. If you don't want to interoperate, you don't have to.

And the sort of thing that emerges out of this is something like Wikipedia. If I had stood before you 20 years ago and said there's going to be an online encyclopedia created by a certain percentage of the 1.59 million Internet users, none of them will be paid and it will outperform the Encyclopedia Britannica, you would have thought I was insane 20 years ago. Yet it happened and here it is.

It was stitched together through a network of common licensing – not one of ours but the GNU Free Documentation License – and standard software. And this leverages the fact that the digital property is somewhat different. You can actually assemble this sort of thing bit by bit and edit by edit, and even though the original quality might be extremely low, over time, if you have the permissions set right, it can improve and still be fully within copyright law.

And the lawyers actually talk about this as the comedy of the commons instead of the tragedy of the commons. And it's only really possible with digital stuff. It doesn't work that way with rivalrous physical goods.



# Beyond the Book<sup>®</sup>



So Creative Commons, which is the nonprofit organization I work for, what we try to do is build something that leverages that power of the network without getting outside of the copyright system. The idea is that we want sharing this kind of stuff to be easy, legal and scalable. And scale is the most important word there, in many ways.

So the generic phrase for what we do is free culture. The idea is that there are some people who really want to share their copyrighted works. I'm one of them.

If you would like to be able to do that, you need a user interface to copyright because it's pretty complicated. That's why the CCC has traditionally existed, to give that option to owners of rights.

But the idea is that on the net, we can do some things with standards that used to require organizations. If you wanted to wire together two networks, you used to have to call IBM. Now you use the Internet. If you wanted to wire together two libraries, you used to have to call IBM. Now you use the Web. If you want to knit together rights, this is the way to do it.

So if you go to [creativecommons.org](http://creativecommons.org), this is what our website looks like.

But the most important thing you do is make some choices as an author and this is where I'm actually going to make a small correction to what Chris said, which is that we enable four conditions in our licenses. Attribution is required. You're not allowed to waive attribution in our system.

What you do is you choose the answer to two questions. One is, would you like to allow commercial use or not? And the other one is, would you like to allow a derivative work or not? And if you choose to allow a derivative work, you have the option to require what we call Share Alike, which is copy left. It basically passes on your license to the derivative work that gets created.

But you do indeed have the right to say, I'm going to let you make and distribute copies of my presentation, but you can't sell them. That's one of the choices you can make. It's actually the most popular choice that's made.

And we have a technology that allows you then to attach. Let's say that I choose to make this available under a noncommercial. You can even then indicate that you'd like me to go buy the rights for a commercial use from the CCC. That's actually fully enabled in what we do because the whole point is, this is about the authors making a choice because you're making a one-to-many offer to the world that says, you're allowed to make and distribute copies, but these are the conditions that I attach to that.



# Beyond the Book®



What this creates is a universe of six possible licenses, a set of standards, from the least free, which is essentially by noncommercial, no derivatives to the most free, which is all you have to do is give me credit and if so, if you do that, you're allowed to make and even sell copies.

It generates what we call the human readable version of the license, as opposed to the lawyer readable or the machine readable versions, where the rights and the obligations are expressed in very clear ways. We summarize them in plain language and we represent them with icons and you can even say, this is how to give me attribution. You put in a URL. Here's how I'd like to be attributed.

The licenses are what we call a three layer cake. There's the human readable version and then there's the legal code, and we also have a digital code version of it, which interoperates with Google and Yahoo search so you can say, find me only songs I'm allowed to remix for commercial purposes.

It creates, essentially, a private space where you know you have the rights to whatever you want to deal with and you don't have to go clear any of the permissions.

This started out as sort of a thought experiment that came out of Harvard Law School after a failed Supreme Court challenge at the beginning of the decade in which the challenge was to the term of copyright. Suffice to say that the Supreme Court sided with Disney and not us and the reaction was, you know what, let's build something that works anyway. And it was a thought experiment, but it's exploded, and now it's exploded internationally as well.

So these are the places where we've translated the licenses legally and lexically, so that if you upload a file in Brazil under a CC license, you can download it in the United States under a CC license and it interoperates. And we've seen exponential growth in the number of the contracts that are actually deployed and used.

So by the middle of last year, we're at about 140 million. We do the count every summer. It's actually very complicated in how to do it. Early indications are that we will have doubled again this year. Some of that is a function of the way blogs work. Every blog post gets its own licensed trackback, so we think that's a piece of it. But Flickr-licensed images under Creative Commons are over 100 million already as of January.

And we're becoming part of the infrastructure, so if you leave a comment on the whitehouse.gov website, it's licensed under a Creative Commons license. This



# Beyond the Book®



actually happened during the inauguration speech, which was sort of surreal. During the inauguration, suddenly you start getting cell phone messages and twitters basically saying, did you know that the White House just adopted your licenses?

This is becoming a fairly standard way to deal with the idea that if you have something you want to share, you should do so in a standardized, open and low transaction cost way.

Now, Science Commons is another piece of this. We come out of the open access movement. The way to think about this is if you've ever read the Budapest or the Berlin or the Bethesda declarations, our attribution license is essentially a word-for-word implementation of that in a free contract. That's why you see us all over the place.

Notably, I would point out PLoS, BioMed Central and Hinari. Plosst, of course, you guys know. BioMed Central, using the most liberal license. They granted their users the right to resell their journals and they did \$15 million in revenue last year and got bought by Springer.

These licenses are compatible completely with the business world. It's a matter of different business models.

We're looking at what these things mean in a world of data. It's very unclear whether and if and how these things should even apply in a world of data. Attribution is a copyright concept, not something that applies to factual data, as Chris very well pointed out earlier.

But people are, ironically because of the success of what we've done, are trying to fit our licenses onto databases in a way that doesn't really work and in a way that we think actually doesn't really scale.

So what we promote actually is a waiver on data and databases that make it easy for commercial journals to link into databases and to recapitulate databases without any fear of a lawsuit.

Now, some good examples of this recently would be the Tropical Disease Initiative, Disease Kernel, which maps genomes to drug targets and known compounds. This was released two weeks ago under the CC zero waiver. The Personal Genome Project, which is moving terabytes of fully sequenced genomes, health records and stem cells around the world, again, under the waiver. You can see the same sort of thing. This is what it looks like in use. You can actually see the icon.



# Beyond the Book®



Immediately after we did this, a file sharing network, a legal file sharing network, picked up and started distributing the terabytes of genome data and again, provide it now with a stable URL, so anytime someone writes a paper about that and submits it to you, they can link back to it and you can actually begin to have that weblink appear.

And the idea is that the same three problems – the copyright problem, the incentive problem and the workload problem – if they get distributed right on the network, they don't have to be interlocking problems. Copyright doesn't have to be a problem. Copyright can be part of the solution so instead of locks, what you have is a set of gears that can come together and make things like Wikipedia happen.

But we have to take some active steps to make that happen in the sciences. And here's an example of what I mean.

The Allen Brain Atlas in Seattle was created as a set of maps of the mouse brain and where the genes express in them. They never thought that anyone but them would do anything valuable with it and so their default terms of use said you can't do anything with these images.

So we actually thought that it might be kind of cool if you could do the same sorts of things with mouse brain pictures that you could do with maps, like annotate collaboratively, like drag and drop and zoom, or even show two things next to each other.

So we shoved the Allen Brain Atlas into Google Maps and went and showed it to them. And it was just a matter of the default setting. They never intended to keep us from doing it. They weren't going to sue us. It had just never occurred to them that there might be value generated by the crowd and not just by them.

The most recent example of this would be the partnership we announced last month with Microsoft. Microsoft is not the one people most often think of when they think of open licensing and open systems, but when it comes to scholarly literature, Microsoft is really interested in making sure that as the point of contact to most authors, they provide good services to them.

So what we've worked with them on is a plug-in that allows you to take words inside the article as you write them, and using something very similar to spell check, start to integrate them out into the dataweb. And this happens before the copyright transfer typically does.



# Beyond the Book®



But this sort of thing could be a batch process by a librarian. It could be run by you guys as a value add that you actually sell to your users. What it lets you do is add these very clear, unambiguous markups for things like tuberculosis, mortality, linked directly out into the dataweb and create these sort of connections.

You can do it all sorts of terms. What you find is that scholarly papers actually ought to have a lot of hyperlinks in them. On average, it's 40 to 50 percent of the words in any given sentence have a link that you could put in if you actually enabled that. And you can even go ahead and start browsing it.

And the reason this is possible is it sits on top of this massive government investment in public data and ontology infrastructure. That plug-in runs on top of what's called the National Center for Bio-ontology's bio portal.

Now, all of these things are only possible either if the user, the author, starts to do these things or if you as scientific editors and journal providers ask them to do this sort of thing or provide it as a service. It's not going to happen magically no matter how many plug-ins we create toward and how open they are.

But the commons provides enough of an infrastructure that you know whether or not you have the rights to an article to make this sort of transformation, whether you're the author or the user or the publisher or the reader.

Just to sort of finish, there's really two ways this goes. I give a lot of open access talks and I hear a lot about the strategies of the other players in the space, and so one of the futures is the Web. This is why the NIH and Autism Speaks and the Wellcome Trust and Howard Hughes have mandated access to the articles that they create and they go into PubMed central in XML where you can do stuff with them like make assertions and links.

But the other future is iPhone. iPhone is great if you own an iPhone. iPhone is great if you use AT&T. But it's limited to the Apple subscribers and Apple can decide what goes into the App store and what doesn't go into the App store. It is a completely different model than the distributed creativity of the Web.

Facebook is a similar thing. These are systems that are inherently closed systems that leverage distribution to help one company or one owner.

And the Web is inherently a messy, semi-anarchic open system, but it scales. It scaled well beyond AOL, which if you go back in time, the iPhone is basically AOL with better graphics, because they tell you what happens when you click the News button and they tell you what happens when you click the Weather button. It's exactly the opposite of the way the Web functions and the Web works.



# Beyond the Book<sup>®</sup>



So the choices that we have to make when we think about how we do this stuff is do we want the Web or do we want AOL? Because those are the real choices we have as content users and citizens on the Web.

So what you can do – it's actually really hard and I feel sorry for people in content industries because the network really screws with content business models.

Now, one of the things you can do is to get a little bit ahead of the curve. Most of the people in this room have already dealt with the NIH mandate. That's getting ahead of the curve on one aspect. But you can get ahead of the curve by linking into data and tools.

This is what Elsevier is doing. This is what Nature is doing. This is what the major businesses in the science journal space are doing. They're providing value added semantic markup to what they do because in a data world, the machine-readable version of the article is much more economically valuable than the one that has no hyperlinks.

In an open access world, you can sell RSS feeds of markup without ever having any of this trouble with someone like me coming in and complaining or any trouble with the NIH, because that's a value add that you created that went through your peer review. It's going to be a lot more valuable than the Wikipedia version of markup on an article that comes through or the author's version of the markup.

And if you do that sort of stuff, you can begin to expect the unexpected. If you begin to allow hyperlinking into your articles – I know for a fact that Google really wants to sell ad word space that connects to things like plasma an cell lines and reagents.

There's a lot of different ways to begin defraying the content business model changes you're facing because of open access mandates, but they don't come from being sort of like half open, because that's going to be NIH and funders sort of forcing you farther and farther down the line and you have these embargoes.

If you allow people to begin to link in, they're going to link in, and the key is to be the ones that get out there and form the content partnerships and the B2B partnerships that let you make more money. (inaudible) culture repositories, mouse repositories. They want your business. They need revenue flow to maintain what they do, but if you have to do all the linking, if they have to do all the linking, it's not going to scale.



# Beyond the Book®



If you let the users make the linking, if you let the businesses come in and make the linking, that's the way the Web took off and that's the way that this is going to take off.

And with that, I'll stop. I'm at my 25 minutes. I give thanks we're funded by various charitable foundations as well as the Nike Corporation.

Thank you.

END OF PROGRAM