MERTONIANISM UNBOUND?[1]
IMAGINING FREE, DECENTRALIZED ACCESS TO MOST CULTURAL AND SCIENTIFIC MATERIAL
JAMES BOYLE[2]

[This is an edited version of a chapter of the same name that appeared in **UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM THEORY TO PRACTICE 123 (Elinor Ostrom and Charlotte Hess eds. MIT Press 2007)**]

I have written far too many pages on intellectual property, the public domain and the commons.[3] I care deeply about the future of scholarly communications, particularly in the sciences. Designing an architecture for freer and more usefully accessible scholarly work is a fascinating task, and I agree with many of the scholars in this volume that the literature on the commons has a number of insights to offer.[4] So I was pleased to be given the task of writing about the commons and the public domain in scholarly communications. This enthusiastic prologue notwithstanding, I am going to stray from that task – one that is performed ably by others in this collection – and instead suggest that we need to think still more broadly about our subject matter. My topic is Mertonianism *beyond* the world of scholarly communications.

Mertonianism of course, is a term borrowed from the sociology of science, generally used to describe a process of free, open, inquiry, without crippling secrecy norms or strong property claims, strongly reliant on the process of peer reviewed publication and citation to drive hypotheses closer to

---

[2] William Neal Reynolds Professor of Law Duke Law School, Faculty Co-Director, Center for the Study of the Public Domain. Papers: http://www.james-boyle.com E-mail: boyle@law.duke.edu

[3] See eg James Boyle, Shamans, SOFTWARE AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY (1996); *The Second Enclosure Movement & the Construction of the Public Domain* 66 Law & Contemporary Problems 33 (2003) *A Manifesto on WIPO and the Future of Intellectual Property* Duke Law. & Technology Review 0009 (2004)

[4] *See* Elinor Ostrom and Charlotte Hess, *IDEAS, ARTIFACTS, AND FACILITIES: INFORMATION AS A COMMON-POOL RESOURCE* 66 Law & Contemp. Probs. 111(2003). The entire Public Domain Symposium can be found online at http://www.law.duke.edu/journals/lcp/indexpd.htm

an underlying objective reality.[5] Access to, and citation of the peer reviewed literature is crucial to the scientific project as Merton describes it, indeed it is one of its principal method of error correction.   It is for that reason that I chose the term for my title. I am using it loosely and provocatively to suggest an inquiry that at first might seem to run partly at odds to Merton's project. My goal is to ask what impact more open access to cultural and scientific materials, both scholarly and non scholarly, by individuals and groups *outside* the academy might have on scholarship, culture and even – though this is more speculative and unlikely – on science. Merton described science as a relatively autonomous process in which specialists used the sociological disciplinary mechanisms of peer review and citation reputation to winnow results. He would have cared deeply about restrictions on access to the scholarly literature or the underlying data if those restrictions were applied to *scientists*. The issue of access by the public was simply not one that presented itself. But it is that question that I wish to raise, for culture, the humanities and the sciences as well.

One implication of the commons literature is that in attempting to construct a "comedic" commons,[6] one must think very carefully about its boundaries – the limits on who may use it and for what types of use. The tendency of my argument here is that, in the scholarly communications commons, the boundaries ought to be very wide indeed. In fact, the design principle I argue for here is that wherever possible neither use, nor the ability to participate in the fine-tuning of the system should be restricted to professional scholars.

I

"YOU CAN HAVE MY LIBRARY OF CONGRESS WHEN..."

I was searching the Library of Congress catalogue one night, tracking down a seventy year old book about politics and markets, when my son came in to watch me. He was about `eight years old at the time, but already a child of the Internet age. He asked what I was doing, and I explained that I was printing out the details of the book so that I could try to find it in my own university's

---

[5] Merton's own views are, in fact, much more subtle than this abbreviated account suggests. Robert K. Merton, ON SOCIAL STRUCTURE AND SCIENCE (1996)

[6] The phrase is Carol Rose's – used to describe cases where, contrary to the suggestion of Hardin's tragedy of the commons, resources are actually *more* efficiently used and managed collectively rather than under individual ownership. Carol Rose, The Comedy of the Commons: Custom, Commerce, and Inherently Public Property," 53 University of Chicago Law Review 720 (1986),

library. "Why don't you read it online?" he said, reaching over my shoulder and double-clicking on the title, frowning when that merely led to another information page. "How do you get to read the actual book?"

I smiled at the assumption that all the works of literature were not merely in the Library of Congress, but actually on the net: available to anyone with an internet connection anywhere in the world – so that you could not merely search for, but read or print some large slice of the Library's holdings. Imagine what that would be like. Imagine the little underlined blue hyperlink from each title – to my son it made perfect sense. The book's title was in the catalogue and when you clicked the link, surely you would get to read it. That is what happened in his experience when one clicked a link. Why not here? It was an old book, after all, no longer in print. Imagine being able to read the books, hear the music, watch the films – or at least the ones that the Library thought it worthwhile to digitize. Of course, that is ridiculous. It took Google's recent attempts to do so to fire the popular imagination, but also to reveal the massive legal pitfalls involved.

I tried to explain this. I showed him that there were some works that could be seen online. I took him to the photograph library, meaning to show him the wealth of amazing historical photographs, but instead finding myself brooding over the lengthy listing of legal restrictions on the images, the explanation that reproduction of protected items may require the written permission of the copyright owners and that in many cases, only indistinct and tiny thumbnail images display to those searching outside the Library of Congress "because of potential rights considerations." The same was true of the scratchy folk songs from the twenties, or the early film holdings. The material was in the library, of course, remarkable collections in some cases, carefully preserved and digitized at public expense – and some tiny fraction of it available online. (There is a fascinating set of Edison's early films, for example.) Most of the material available online came from so long ago that the copyright could not possibly still be in force. But since copyright lasts for seventy years after the death of the author, (or ninety five years if it was a corporate "work for hire,") that could be a very, very long time indeed. Long enough, in fact, to keep off limits almost the whole history of moving pictures, the entire history of recorded music. Long enough to lock up almost all of twentieth century culture.

But isn't that what copyright is supposed to be doing? To be granting the right to restrict access, so as to allow authors to charge for the privilege of granting it? Yes indeed. And this is a very good idea. Yet the goal was to give the minimum monopoly necessary to provide an incentive, and after that to let the work fall into the public domain, where all of us can use it, transform it, adapt it, build on it, republish it as we wish. For most works, the answer is that the owners expect to make all they money they are going to recoup from the work with five or ten years of exclusive rights. The rest of the term is of very little use to them except as a kind of lottery ticket in case the work proves to be a one in a million perennial favourite. The one in a million lottery winner will benefit, of course, if his ticket comes up. And if the ticket is "free" who would not take it? But the ticket is not free to the public, who pay higher prices for the works still being commercially exploited and, frequently, the price of complete unavailability for the works that are not.

Think of the one-in-a-million perennial favourite -- Harry Potter, say. Long after J.K. Rowling is dust we will all be forbidden from making derivative works or publishing cheap editions, or large type versions, or simply from reproducing it for pleasure. I am a great admirer of Ms. Rowling's work, but my guess is that little extra incentive was provided by the thought that her copyright will endure seventy, rather than merely fifty years after her death. Some large costs are being imposed here, for a small benefit. And the costs fall even more heavily on the other 999,999 works, works which are available nowhere but in some moldering library stacks. To put it another way, if copyright owners had to purchase each additional five years of term, the same way we buy warranties on our appliances, or insurance policies, the economically rational ones would mainly settle for a fairly short period.

Of course, there are some works which are still being exploited commercially long after their publication date. Obviously the owners of these works would not want them freely available online. This seems reasonable enough, though even with those works the copyright should expire eventually. But remember, in Library of Congress's vast wonderful pudding of songs and pictures and films and books and magazines and newspapers, there is maybe half a raisin's worth of works that anyone is making any money from, and the vast majority of those come from the last ten years. If one goes back twenty

years, maybe a raisin fleck's worth. Fifty years? A slight raisinous aroma. We restrict access to the whole pudding, in order to give the owners of the raisin sliver their due. But this pudding is almost all of twentieth century culture – and we are restricting access to it, when almost of all of it could be available.

If you do not know much about copyright, you might think that I am exaggerating. After all, if no one has any financial interest in the works or we do not even know who owns the copyright, surely the library would be free to put those works online? Doesn't "no harm, no foul" apply in the world of copyright? In a word, no. Copyright is what lawyers call a "strict liability system." This means that it is generally not a legal excuse to say that you did not believe you were violating copyright, or that you did so by accident, or in the belief that no one would care and that your actions benefitted the public. Innocence and mistake do not absolve you, though they might reduce the penalties imposed. Since it is so difficult to know exactly who owns the copyright (or copyrights) on the work, many libraries simply will not reproduce the material or make it available online, until they can be sure the copyright has expired – which may mean waiting for over a century. They cannot afford to take the risk. As for the cases where the copyright owners are identifiable, they would treat any digitizing of their work as a great new financial opportunity, though they themselves are doing nothing to distribute it, or sell it, or make it available, and have not for years.

What is wrong with this picture? Copyright has done its job and encouraged the creation, and the initial distribution, of the work. But now it acts a fence, keeping us out, and restricting access to the work to those who have the time and resources to trudge through the stacks of the nation's archives. In some cases, as with film, it may simply make the work completely unavailable.

So far I have been talking as though copyright was the only reason the material is not freely available online. But of course, this is not true. Digitizing costs money, (though less money every year) and there is a lot of rubbish out there, stuff no one would ever want to make available digitally. (Though it must be noted that one man's rubbish is another man's delight.) But that still leaves vast amounts of material that we would want, and be willing to pay to have

digitized. Remember also that if the material were legally free then anyone could get in the act of digitizing it and putting it up.

If you are shaking your head as you read this, saying it would never work, look at the Internet and think about where the information came from the last time you did a search. Was it an official and prestigious institution? A university or a museum or a government? Sometimes those are our sources of information, of course. But don't you find the majority of the information you need by wandering off into a strange click-trail of sites, amateur and professional, commercial and non, hobbyist and entrepreneur, all self-organised by internal referrals and search engine algorithms?

The most satisfying kinds of proofs are existence-proofs. Could a mammal lay eggs? The platypus provides an existence-proof. The Internet is an existence-proof of the remarkable information processing power of a decentralised network of hobbyists and amateurs and universities and businesses and volunteer groups and professionals and retired experts and who knows what else. It is a network that produces useful information and services. It frequently does so for at no cost to the user beyond the telecommunications access charge and it does so without anyone guiding it. Imagine that energy, that decentralised and idiosyncratically dispersed pattern of interest, turned loose not only on the cultural artefacts of the twentieth century, but on the universe of scholarly literature. Think of the people who would work on Buster Keaton, or the literary classics of the nineteen thirties, or the films of the Second World War, or footage on the daily lives of African-Americans during segregation, or the music of the Great Depression, or theremin recordings, or the best of vaudeville. But think also of those who are fascinated by Civil War History, or the analysis of the works of Dickens, or the latest paper on global warming, or Tay-Sachs disease. Where are the boundaries of the academy now? This is a more radical vision than making journals free available online to scholars. Imagine your internet search in such a world. Imagine that Library of Congress. One science fiction writer has taken a stab. His character utters the immortal line, "You can have my Library of Congress when you drag it from my cold, dead hand."

Now, anyone who can't sell to scholars the desirability of freer access to scholarly and cultural materials couldn't sell fire extinguishers to the burning.

But in your willingness to agree with me that this would be a fine thing, you may miss my point. Two further stories may suffice to make it. The first I owe particularly to the work of Jessica Litman[7] and Yochai Benkler. The second comes from my experiences working on digital archive projects.

<div align="center">II</div>

A GLOBAL NETWORK FOR OPEN SOURCE FACT-CHECKING...

If I had come to you in 1994 and told you that in the space of ten years, a decentralized global network consisting of a *lot* of volunteers and hobbyists and ideologues and a *few* scholars and government or commercially supported information sources could equal and sometimes outperform standard reference works or reference librarians in the provision of accurate factual information, you would have laughed. Your incredulity would surely have deepened if I had added that this global network would have no external filters, and that almost anyone with an internet connection would be able "publish" whatever they wanted, be it accounts of Area 51, the Yeti, and the true authorship of William Shakespeare, or painstaking analyses of Scottish history, how to raise Saluki dogs, and the internal struggles in the American Communist party. There is no "editor," no formal "peer review" system, and the very identity of the writers and publishers is frequently in doubt. Worse still, many inhabitants of this strange new space will wilfully and joyfully spread the wildest of rumours and speculations as facts, without going through the careful source-checking or argument-weighing, that scholars are supposed to engage in. Your first reaction to this flight of fancy, (and the correct first impression of the World Wide Web as of its inception) was that this would thus be a *uniquely* and *entirely* unreliable source of information. This seems to be the very opposite of Mertonian science – it lacks the boundaries, requirement of professional credentials and disciplinary constraints like those of peer review. And yet... when your child last had a research question from school did you go to Google, or the Encyclopedia Britannica?[8]

---

[7] In particular, I owe a considerable debt to Jessica Litman's wonderful *Sharing and Stealing* http://www.law.wayne.edu/litman/papers/sharing&stealing.pdf *See also* Yochai Benkler, Coase's Penguin or Linux and the Nature of the Firm http://www.benkler.org/CoasesPenguin.PDF   James Boyle, The Second Enclosure Movement & the Construction of the Public Domain http://www.law.duke.edu/pd/papers/boyle.pdf at 46.

[8] Jeffrey Selingo, When a Search Engine Isn't Enough, Call a Librarian, N.Y. Times, Feb. 5, 2004, at G1. (Noting that professional librarians rely heavily on search engines.)   I admit that the data on comparative accuracy

Think of the standard property account of the property regime necessary to generate a public good such as an encyclopedia or other comprehensive reference work. Strong property rights would be necessary for at least three reasons – each of them related to the tragedy of the commons. First, without the guarantee of a future legally protected monopoly called copyright, one could not attract the investment necessary to engage scores or hundreds researchers to produce a work that could easily be copied by the first free rider to come along. Second, without the ability to control the resource provided by a legal right to exclude content, quality could not be maintained: The Encyclopedia can reject the articles on Area 51 and the Yeti.   Single-entity control, backed by property rights, allows for semiotic as well as agricultural stewardship. Third, without control over the name of the resource, such as that provided by trademark, there would be inadequate incentives to generate a quality product, and inefficient signalling to consumers. Why would a publisher invest in the production of a high quality product if its name could be used by anyone? Why would consumers trust the name as a signal of quality if they could not be sure this was the *real* Encyclopedia Britannica? Names as well as pastures can be over-grazed. In other words, without single entity control and strong property rights, we will not get the generation of useful and reliable reference information. And yet, as I said before, when is the last time you turned to the encyclopedia or the World Book rather than to the Web? How many of the things you have found on the Web *could* have been found in a standard reference work? When it comes to the generation and retrieval of useful factual information the Web is an existence proof of the viability of commons-based production, validation and distribution.   In fact, as Jessica Litman points out, one reason for the success of the system is the *absence* rather than the presence of property rights in factual data – facts cannot be copyrighted.

> This information system is vital and dynamic because information
> sharing is almost frictionless. Material is passed along at low cost

---

is hard to pin down – particularly because scholars apply a set of unconscious filters to the information provided by the search engine, filters that themselves are not so easily replicated. For some older discussions, *see* Joseph Zumwalt and Robert Pasicznyuk, *"The Internet and Reference Services: A Real-World Test of Internet Utility,"* REFERENCE & USER SERVICES QUARTERLY 38:2 (1998): Joseph Janes and Charles McClure, *"The Web as a Reference Tool: Comparisons with Traditional Sources,"* PUBLIC LIBRARIES 38:1 (JAN/FEB 1999): 30-39; 165-72*;* Tschera Harkness Connell and Jennifer. Tipple, *"Testing the Accuracy of Information on the World Wide Web Using the Alta Vista Search Engine,"* REFERENCE & USER SERVICES QUARTERLY *38:4 (Summer 1999): 360-368.*

with few practical or legal barriers. Jeff Dalehite, webmaster of <scratchdj.com>, is free to post the details of the early history of the phonograph without seeking the consent of his sources. Dalehite's site tells us that Thomas Edison invented the cylinder phonograph in the 1870s and patented it in 1878. Dalehite recounts the details of the commercial standards competition between Edison's phonograph and the disk gramophone introduced to the U.S. market in 1901 by the Victor Talking Machine Company. He attributes none of his sources; he need not even know whether the information he has abstracted was original to the references he used or derived by them from some other source. Technical writer Samuel Berliner III has posted a site honoring famous people throughout history named Berliner. His site reports that the disk gramophone was invented by Emile Berliner in 1887. Berliner needs no permission from Frederick W. Nile, the author of a 1926 biography of Emile Berliner, nor the National Inventors Hall of Fame, who have posted a short profile of Berliner, from whom he initially learned that information. Neither Dalehite nor Berliner has secured a license from Tommy Cichanowski for any facts they might have learned by studying Tommy's History of Western Technology, nor have they sought the blessing of the periodical Electronic Design, whose February 1976 issue commemorating the U.S, bicentennial furnished many of the dates that Cichanowski reports. If one were unable to post facts without determining who controlled them and obtaining a license to pass those facts on, this online information space would not exist.[9]

Take a step back for a moment. The original work on the tragedy of the commons *overestimated* the applicability of the tragic commons paradigm, and *underestimated* the extent to which we could have a well-managed commons governed by a variety of formal and informal norms. Elinor Ostrom and her colleagues taught us this and a variety of intellectual property theorists have shown the applicability of their work to the world of the intellectual commons. Certainly, the world of scholarly communications is a promising place for the

---

[9] Litman, *Id.*

application. But if we confine our analysis to the world of scholarly communications *as currently constituted*, are we guilt of a similar error to the original tragedians? Are we underestimating the power of a lay audience, given free access to cultural materials and factual data *as well as* scholarly work, to add richness and depth to the world of scholarship in the same way that they have in the world of provision of factual information? Are we underestimating the power of an enlarged audience to enrich our scholarship as well as merely reading it?

Obviously, one would want to be very careful not to overstate the potential here. In the context of factual data, search engine algorithms have managed to provide a strange kind of layperson's peer review so that we can get usable quality out of contributions of distinctly varied worth. So called "water-hole ranking" relies on the assessments of other users about the relevance of a particular page; how many people link to this page on this topic? And what do other users think about the pages that provide the links? Just as markets have provide relatively good signals about the likelihood of factual events, some of them requiring considerable scientific knowledge to predict, so Google-type algorithms generally provide an aggregated sense of the collective judgement. Even if the page rank accurately reflects the collective judgement, of course, that does not mean the collective judgement is correct. Yet search engines will give us a snapshot of a debate if issues are controversial and, with surprising frequency, gives prominence to dissenting views, particularly if those seem to be well-backed by expertise and recourse to data. The result is a rough winnowing process which often allows us to free-ride on the judgement of those who expertise on the issue. Like markets, (or peer review?) the system can be distorted by intentional gaming, fads, cascades of enthusiasm and undeserved reputational advantages. Nevertheless, the results are clearly useful.

An important qualification is in order here. Most educated readers apply their own additional filters to the material retrieved by search engines. They look at several results to see if answers converge rather than merely relying on just one. (And search further if they do not.) They give different levels of credibility to work based on its origin, its author and her credentials. They assess its presentation (everything from grammar and syntax to the look and feel of the page, and the pages it links to.) They may cross-check with a recognized

authority which itself was produced through more conventional means, such as a dictionary or a book of quotations. These "filters" are often applied unconsciously, but they dramatically increase the accuracy of the results. The decentralized search engine of the web requires an entirely different level of skepticism and acquired sophistication about indicators of credibility, than does a static encyclopedia. Thus one cannot simply assume that the web, plus distributed creation and reference, plus search engine algorithms are, by themselves, enough to produce a reliable information retrieval system. Social capital, in the form of educated skepticism, is also vital. Yet the process does not stop here. Collectively created reference tools such as Wikipedia formalize the process of decentralized research. Those with a particular interest in one subject put up their own entries on it, only to have them commented upon, edited and subject to a strange form of lay peer-review. The process is often anarchic and contentious, but the results are remarkably impressive. To paraphrase a credo of open source software "With enough eyeballs, and an interested community, many errors will be caught."

Thus let me return to my central questions. Are we underestimating the power of a lay audience, given free access to cultural materials and factual data *as well as* scholarly work, to add to the world of scholarship and knowledge generation in the same way that they have in the world of provision of factual information?

My analogy might seem inapposite. Yes, decentralized systems are surprisingly good at generating factual reference material that can be winnowed through the processes I describe. But here our subject is *scholarly* communication, and surely there are differences between scholarship and simple factual reference? I completely agree. Let me stress the point*: the need for specialized expertise, sensitivity to source material, historical knowledge, and professional analytical tools means that the overwhelming majority of scholarly work will not be affected, or usefully supplemented by some imagined distributed process of lay volunteerism*. Indeed, just on the level of *reading* most scholarship would not even be of interest to a lay audience. And yet with huge numbers of potential global readers, very low costs, and the possibility of decentralized methods of assessment that mimic peer review, the possibilities of productive exchange are surely above zero. Are they sufficiently far above zero

to be worthwhile? After all, any enlargement of literacy, any broadening of the franchise, any new influx of opinion will bring with it a lot of noise as well as signal. Can current and future filtering methods, ranging from credibility assessment to peer review and search engine algorithms, manage to separate signal from noise? The answer is, I think, that we do not know. But our failure to predict the Net's role as a useful information source coupled with our experience with the tendency of individuals to underestimate the potential of 'the well-run commons' should impose on us a double dose of humility before we write off the potential of such contributions altogether.

In one sense, the question I describe here is fundamental to the division between the progressive and the populist impulses in American politics. The progressive notes the dangers of collective irrationality, of lack of understanding, of availability cascades the violently skew perceptions of risk and benefit. He puts his faith in the expertise of technocratic specialists working for the public interest, but isolated from public pressure and hubbub. The populist, by contrast, is skeptical of claims that restrict knowledge, decision-making or power to an elite group. He sees the experts as being subject to their own versions of narrowness and prejudice, their own cascades. Most sensible people acknowledge that each of these perspectives on the world has important truths to offer. The question is where the balance is to be drawn. Despite the tendency of some of my arguments so far, my goal is not to wave the banner for a populist movement in scholarship. Instead, it is to argue that we do not *know* the benefits and costs that wider access to cultural and scholarly material could bring. What's more, we have at least one reason not to reject the notion out of hand. At every stage of the development of mass literacy, it has seemed reasonable to doubt that anything productive could come out of widening the circle of participants – whether in Biblical exegesis, reading the law in English rather than in Law French, exercising popular sovereignty in the move towards mass democracy, and in the changes to politics wrought by easy internet access to public documents. Is there a lesson there?

Having thus chastened both our expectations and our tendency to discount the possibility altogether, in the remains of this paper I shall consider how a larger universe of readers might be interested in scholarly literature and how scholarship might even occasionally benefit from the process.

III
BEYOND THE SPECIALIST ARCHIVE? USERS AS DESIGNERS

I was recently at a meeting of academics, digital librarians, and technologists, talking about the construction of usable specialized digital archives. The librarians and technologists told of constructing beautiful systems, with 24 different metadata fields and incredibly powerful search capabilities. They also explained the "dirty secret" of many of these archives: no one uses them. The response from the group was a thoughtful one – academics from within the discipline should be included in the design process, so that the system fits their patterns of work, and conceptual categories, rather than being imposed based on some alien categorical scheme. Who could disagree?   Nevertheless, I was struck by the similarity of the scene to a whole series of moments in the history of technology: moments where the experts dramatically misunderstood the likely patterns of use of a technology. The telephone was, famously, initially imagined as a one-to-many communications device, useful for weather reports distributed from a central source and the like. It found such use only in Albania.[10] ATT predicted that cellphones would be used by a maximum of 900,000 people in the United States by the year 2000.[11] The FCC's prediction was lower. (Would that they had been correct!)   Who predicted that IM would be a killer app, or imagined that e-mail would replace the phone call in much of corporate culture? Indeed, to go back to my earlier example, who predicted the explosion of the Web, or the extent to which people would rush to share knowledge, impressions, opinions – generally at some inconvenience to themselves and without monetary incentives to do so? Who predicted that free and open source software written and assembled largely by volunteers would outperform proprietary software in mission-critical applications and would be endorsed by parts of the national security apparatus? (Apart from Richard Stallman, that is.)

---

[10]   ITHIEL DE SOLA POOL, FORECASTING THE TELEPHONE: A RETROSPECTIVE TECHNOLOGY ASSESSMENT (1976);

[11]   The true number was closer to 10,000,000.   http://knowledge.wharton.upenn.edu/121901_ss7.html. According to some estimates there are now more close to 1 billion cell phones world-wide.

The point is, if the history of technology teaches us anything, it teaches us that we are extremely bad at predicting *ex ante* the uses of technology. This fact has an overlooked, but absolutely vital design corollary; **whereever possible, design the system to run with open content, on open protocols, to be potentially available to the largest possible number of users, and to accept the widest possible range of experimental modifications from users who can themselves determine the development of the technology.[12] Then sit back and wait to see what emerges.** It may be that your predictions of how the technology will be used, *and even your predictions about the potential user group*, will be completely wrong. All other things being equal, the more open the system is to change from multiple sources, the more open the content is to users beyond your initial target group, and the more the system can actually accept experimental changes from multiple external sources, the quicker you are likely to find the best use of the technology. Precisely because of the limits of foresight, making the entire archive available on the web, so that anyone can develop a search engine, or simply use Google, may well be better than building a wildly sophisticated specialist system designed by experts and used by no one. It is not an accident that some of the greatest recent successes in new technologies – the Web itself and the technologies it enables – present exactly this model of development. In other words, having end-users in the design stage is definitely a step forward from having technologists or librarians dream up an archival scheme from scratch. But even end-users may misunderstand their own patterns of use, fail to anticipate important functions, or generally be unable to replicate the successes of a more open process of cybernetic adjustment.

---

[12] Eric Von Hippel is the undisputed master of the literature on user based innovation. Eric Von Hippel, DEMOCRATIZING INNOVATION (MIT Press 2005). [Available in full online at http://web.mit.edu/evhippel/www/democ.htm)] Allowing users to participate, formally or informally, in the shaping of repositories obviously runs into a non-trivial design problem because of the "negative network externalities" that can result. What is a negative network externality? Think of the Tower of Babel: the original model of a system in which all users design their own protocols. However the open source software community, Wikipedia, and our own experience in scholarly disciplines, show us that a variety of formal and informal norms can help to manage a process of commons-based production, without letting it collapse into a Babel of incompatible efforts. And to return to the linguistic example, languages themselves manage just fine without a single property owner, or authoritative Academie Francaise vetting all possible linguistic innovations; a fact that both commons theorists and Hayekian market enthusiasts noticed long ago.

Can one succeed with a closed model? Of course, We all use highly specialised databases that, for copyright or other reasons, are closed to the outside world. For lawyers, Westlaw is an example. When I want to know what the 2nd Circuit thinks about the copyright doctrine of "merger" I do not *want* an open archive, or a loose search engine. I want a very particular search restricted to a very particular set of materials, using a fairly precise and fiddly Boolean search engine that capitalises on esoteric knowledge and employs technical jargon. The system, driven by the competitive urge to be more attractive than LEXIS and relying on feedback from countless users, offers a well-designed and extremely useful service. Market pressure can make proprietary systems highly responsive to emerging user needs and desires. Open source platforms searching open content offer an attractive model, but hardly the only model. In any world I can imagine there will be a vibrant, and profitable, specialised set of "closed" information ecologies that rely on technology and proprietary rights to exclude all but high-valuation users, and offer sophisticated tools of little interest to the majority. Nevertheless, I would stick with my default design principle: whereever possible, design the system to run on open protocols, make the content available to the largest possible number of users, and accept the widest possible range of experimental modifications and additions from users who can themselves determine the development of the technology. There are two simple reasons for adopting this as the default rule. First, the traditions of the academy, of scholarship and of Mertonianism itself, dictate that openness in both content and structure should be our baseline, deviations from which require justification. Second, where one is uncertain whether a closed or open architecture is better, start with the one from which it is easier to develop alternatives if you have chosen wrongly. And shifts from open to closed are made with fuller information held by more parties (by definition) than the reverse.

IV
WITH ENOUGH BRAINS.., IS ALL CONTENT INTERESTING?

My argument depends in part on the virtues of a larger than expected audience, and on the serendipitous uses that unrestricted access and open, malleable protocols for searching can allow to develop. Is this assumption realistic? Open source software developers tell us that with enough eyeballs all bugs are shallow. With enough brains is all content interesting? Is there a lay

audience for scholarly work, and the cultural and scientific materials on which it is based? Not always, of course. But this bolsters my argument rather than undermining it. The point is that we cannot predict confidently where and when there is a broad audience for scholarly work, or archival material, still less where and when non-scholars can actually contribute to usefully to the field. And this again argues that openness to the public – rather than merely to a scholarly audience – ought to be a general design principle. Take the world of scientific medical research. This seems like the paradigmatic example of esoteric material in which laypeople have little interest and less knowledge. Yet the internet has meant a dramatic surge in lay-people using the scholarly literature to research their, or a family member's illness, to help frame questions to doctors, to look at the results of new studies and the like. NIH has actually redesigned Medline to make it more accessible to lay people.[13]

Sometimes, of course, this means that medically untrained people misdiagnose their illnesses, pester their doctors with fanciful interpretations of irrelevant studies, or refuse vaccines based on unproven charges of their effects. These are real costs, yet the consensus seems to be that the benefits are even greater – improving health knowledge, helping to catch misdiagnoses, encouraging people to seek medical care more quickly when it is appropriate, assisting in the formation of patient groups, and sometimes even catalyzing patient-led attempts to encourage development of new therapies.[14] The pre

---

[13]   *See* ELANA VARON, MEDLINE PLUS: ONLINE MEDICAL INFO FOR ORDINARY PEOPLE "Although the National Library of Medicine has always provided information to the public, its resources, including online databases, were designed for medical professionals. Now the agency has developed a World Wide Web site, Medline Plus, that aims to deliver the latest medical research and health information to lay people. The Medline Plus site collects information on common diseases and conditions and offers dozens of reference tools used by medical librarians. It is a work in progress, NLM director Dr. Donald Lindberg said. "We have known for many years that it's very, very desirable to provide biomedical information to the public, but we've not done it directly," he said. Among the reasons the agency is trying to provide such information now are that it is easier to disseminate information through the Internet, and there is growing public demand for health information." http://edition.cnn.com/TECH/computing/9901/18/medline.idg/   Medlineplus   can   be   found   at http://www.nlm.nih.gov/medlineplus/   *See also* http://www.ncbi.nlm.nih.gov/PubMed/   "PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources."
[14]   Sharon Terry, a mother of children with PXE, was named as one of the co-inventors on the patent over the PXE gene. Eliot Marshall, Patient Advocate Named Co-Inventor On Patent for the PXE Disease Gene *Science*, Vol 305, Issue 5688, 1226 , 27 August 2004. Terry has spoken frequently on the ways in which access to medical literature is vital for patient groups and advocates. See *In the Public Interest: Open Access and Public Policy*

Digital Reformation model in which a priestly intermediary always stood between the scholarly text and the laity no longer seems so inevitable. In fact, this tendency is frequently cited as a reason to encourage open access to scholarly journals. In the words of the Budapest Open Access Initiative.

> An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.[15]

Recently, this desire has even prompted a worthily-intentioned but misguided attempt to require that all articles based on government funded research be published without copyright restrictions, *precisely so that citizens can have unrestricted access to the scholarly literature.*

> Scientific research paid for by the U.S. government would be required to be given free to the public, under a bill introduced in Congress last week. Representative Martin Olav Sabo, a Minnesota Democrat, said he introduced the Public Access to Science Act (PASA) of 2003 because U.S. residents shouldn't have to pay twice — once with tax dollars and a second time with subscription fees to scientific journals — for research that improve their health or save their lives. "It is wrong when a breast cancer patient cannot access federally funded research data paid for by her hard-earned taxes,"

2005 ACRL/SPARC Forum (2005)
http://www.arl.org/sparc/meetings/ala05mw/2005MW%20Forum%20_report.pdf

[15]   (From the Budapest Open Access Initiative Web site at http://www.soros.org/openaccess/).

Sabo said in a statement. "It is wrong when the family whose child has a rare disease must pay again for research data their tax dollars already paid for. Common sense dictates we provide the most cutting-edge research to all who may benefit from it — especially when they've already paid for it with their tax dollars, and my legislation will do just that."[16]

Most, but not all, of the use by lay people of this literature is "consumptive" in the non-Jane Austen sense of the word. Citizens seek information to solve practical problems, to instruct themselves and family members. Instructional aid has always been an important and worthy goal of scholarly literature. It is also worth noting, though, that whether it is Sharon Terry, the PXE patient group advocate, or the dedicated environmental activist researching ground water contamination near his home – there are a growing number of cases in which motivated groups lay people actually help shift policy and even occasionally redirect research. Non scholars can make productive as well as consumptive uses of our work.

So much for medical scholarship. That is an area where people have a real functional need, and where smart search engines can take us an admittedly small, but important step along the road that separates the citizen from the specialist. Does this kind of interest – and the associated importance of making sure that both primary sources and secondary literature are available to the widest audience possible – exist beyond the medical realm? I would say that the answer is clearly yes, both in terms of access to scholarly literature and in terms of access to archival materials. We have examples in genealogical research, astronomy, civil war history, environmental science, with more examples popping up every day. More saliently perhaps, in those (sadly few) places where copyrights have actually expired on texts, movies, music, pictures – we have an explosion of efforts by laypeople to comment, annotate, digitize and in short make usefully available the works of the past. Project Gutenberg is only the most salient example.

---

[16] GRANT GROSS, BILL SEEKS FREE ACCESS TO FEDERALLY FUNDED RESEARCH BIO-IT World 07/01/03 http://www.bio-itworld.com/news/070103_report2813.html I support the overall goal of wider and freer access, but the tool chosen is a blunt and unfortunate one.

What does the Web teach us? It is not merely that "with enough brains all content is interesting." To paraphrase some earlier work on distributed creativity,[17]

i.)   If one has a global network, with very low barriers to entry and participation and

ii.)   If the type of creation involved is in some sense "modular" or built by accretion, and

iii.)   If there is a random distribution of interests in particular topics (ornithology, literary history open source software etc.) and

iv.)   If there is a random distribution of incentive structures (greed, pride, altruism, desire to display virtuosity, hope of attracting interest etc. etc.) then

v.)   On any given topic, one will find a lot of motivated people with useful skills.

The web has already taught us these lessons in the context of factual research. It may be that they have some application to the design principles for the "commons" of scholarly communications.

V
CONCLUSION

The literature on the commons has much to teach us about intellectual production. It teaches us that the "tragedy of the commons" is only part of the picture; that there are comedic, well-run commons. It teaches us that the commons is not the same as the public domain;[18] successful commons' are frequently characterised by a variety of restraints – even if these are informal or collective, rather than coming from the regime of private ownership. It even gives us generalisable tools that can help us to match types of resources with types of commons regimes. The Web confirms those lessons. As I pointed out earlier, standard intellectual property theory would posit that to get high quality factual reference works we need strong property rights and single entity control for at least three independent reasons related to the tragedy of the commons –

---

[17]   *See* Boyle, *Second Enclosure* and Benkler, *Coase's Penguin supra.*

[18]   James Boyle, *The Opposite of Property* 66 Law & Contemporary Problems 1 (2003)

the need for exclusive control over reproduction in order to produce the incentives necessary for large scale investment in writers and fact checkers, the need for control over content and editing in order to ensure quality, the need to control over the name or symbol of the resource itself as a signal to readers and an inducement to invest in quality in the first place. In this case though, the standard story was wrong, or at least incomplete. The fact that the Net has actually become a high quality factual reference through a distributed process run largely by volunteers, with no central organising body, is nothing short of fascinating. Indeed, it is precisely the comparative *absence* of intellectual property rights to exclude from facts and references that has been the key to the cooperative enterprise. There are provocative similarities between the possible future of digital scholarship and the remarkable successes of systems that harness lay volunteers in order to produce high quality out of individual contributions whose quality varies widely.

When coupled to our inability to predict accurately the best uses of new technologies, and the remarkable successes of free and open-source style development in which users are also designers, the Net's success as a reference work offers a persuasive analogical argument for a particular design principle in the construction of the scholarly commons: wherever possible, design the system to run with open content, on open protocols, to be potentially available to the largest possible number of users, and to accept the widest possible range of experimental modifications from users who can themselves determine the development of the technology. Then sit back and wait to see what emerges. We might be as surprised as we were when the Net stopped symbolising inaccuracy and became a default reference source.

The second implication of my argument here is even simpler. In practice, the scholarly readers of this volume have access to at least some version of the online Library of Congress that my son imagined. The wonders of interlibrary loan and subscription services can provide us with access to the resources of the world's libraries, though we cannot "click to get the out-of-print book" in the way a more rational copyright system would permit.[19]   When many of us – I

---

[19] This applies only to reading, however. Other uses of texts – republishing, annotating, using substantial excerpts in a course book – may still be suffocatingly difficult because of long copyright terms, and the prevalence of orphan works. And even gaining access to old films and music is often hard even for professional scholars. See

exempt librarians from this statement – think of a world in which one could "click to get the book" we do so with regret but little passion. Partly, that is because we think of the issue as simply one of consumptive access – it would be nice for non scholars to have a greater ability to read, see or hear the works of the past. The literature on the commons, and the past history of the Net as a factual resource, give us another reason to cherish this idea – a productive, even a scholarly one. Working in an arena where facts are largely free from intellectual property rights, the Net has assembled a wonderful cybernetically organised reference work. What might it do to the 97% of the culture of the 20th century that is not being commercially exploited if that culture was available for everyone to annotate, remix, compare, compile, revise, create new editions, link together in archives, make multi-media reference works?

The second part of my argument went beyond popular access to the cultural material of the 20th century. I suggested that the scholarly communications commons should to be designed under the default assumption that, where possible, one would seek to ensure that both the repositories and participation in the design of repositories were available to the broadest number of people. What if dramatically more scholarly material on everything from medicine to literature were freely available and easily searchable? What if specialised scholarly archives lived side-by-side with archives whose design reflected participation by both scholarly and lay users, "Democratizing Innovation" in von Hippel's terms?[20] What in, other words, if we imagined a world of potential colleagues rather than a universe of passive consumers? But that is a very large scholarly commons indeed.

Duke Center for the Study of the Public Domain: Comments to the Copyright Office on Orphan Works http://www.law.duke.edu/cspd/pdf/cspdproposal.pdf

[20] *See* Eric von Hippel, Democratizing Innovation (2005).