

---

## SCOPE

As we saw in chapter 1, any kind of content can in principle be OA. Any kind of content can be digitized, and any kind of digital content can be put online without price or permission barriers. In that sense, the potential scope of OA is universal. Hence, instead of saying that OA applies to some categories or genres and not to others, it's better to say that some categories are easier and some harder.

OA is not limited to the sciences, where it is known best and moving fastest, but extends to the arts and humanities. It's not limited to research created in developed countries, where it is most voluminous, but includes research from developing countries. (Nor, conversely, is it limited to research from developing countries, where the need is most pressing.) It's not limited to publicly funded research, where the argument is almost universally accepted, but includes privately funded and unfunded research. It's not limited to present and future publications,

where most policies focus, but includes past publications. It's not limited to born-digital work, where the technical barriers are lowest, but includes work digitized from print, microfiche, film, and other media. It's not limited to text, but includes data, audio, video, multimedia, and executable code.

There are serious, practical, successful campaigns to provide OA to the many kinds of content useful to scholars, including:

- peer-reviewed research articles
- unrefereed preprints destined to be peer-reviewed research articles
- theses and dissertations
- research data
- government data
- source code
- conference presentations (texts, slides, audio, video)
- scholarly monographs
- textbooks
- novels, stories, plays, and poetry
- newspapers

- archival records and manuscripts
- images (artworks, photographs, diagrams, maps)
- teaching and learning materials (“open education resources” and “open courseware”)
- digitized print works (some in the public domain, some still under copyright)

For some of these categories, such as data and source code, we need OA to facilitate the testing and replication of scientific experiments. For others, such as data, images, and digitized work from other media, we need OA in order to give readers the same chance to analyze the primary materials that the authors had. For others, such as articles, monographs, dissertations, and conference presentations, we need OA simply to share results and analysis with everyone who might benefit from them.

A larger book could devote sections to each category. Here I focus on just a few.

## 5.1 Preprints, Postprints, and Peer Review<sup>1</sup>

Throughout most of its history, newcomers to OA assumed that the whole idea was to bypass peer review. That assumption was false and harmful, and we’ve made good

progress in correcting it. The purpose of OA is to remove access barriers, not quality filters. Today many peer-reviewed OA journals are recognized for their excellence, many excellent peer-reviewed toll-access journal publishers are experimenting with OA, and green OA for peer-reviewed articles is growing rapidly. Unfortunately many newcomers unaware of these developments still assume that the purpose of OA is to bypass peer review. Some of them deplore the prospect, some rejoice in it, and their passion spreads the misinformation even farther.

All the public statements in support of OA stress the importance of peer review. Most of the enthusiasm for OA is enthusiasm for OA to peer-reviewed literature. At the same time, we can acknowledge that many of the people working hard for this goal are simultaneously exploring new forms of scholarly communication that exist outside the peer-review system, such as preprint exchanges, blogs, wikis, databases, discussion forums, and social media.

In OA lingo, a “preprint” is any version of an article prior to peer review, such as a draft circulating among colleagues or the version submitted to a journal. A “postprint” is any version approved by peer review. The scope of green OA deliberately extends to both preprints and postprints, just as the function of gold OA deliberately includes peer review.<sup>2</sup>

We could say that OA preprint initiatives focus on bypassing peer review. But it would be more accurate to say

In OA lingo, a “preprint” is any version of an article prior to peer review, such as a draft circulating among colleagues or the version submitted to a journal. A “postprint” is any version approved by peer review.

that they focus on OA for works destined for peer review but not yet peer reviewed. Preprint exchanges didn't arise because they bypass peer review but because they bypass delay. They make new work known more quickly to people in the field, creating new and earlier opportunities for citation, discussion, verification, and collaboration. How quickly? They make new work public the minute that authors are ready to make it public.

OA preprints offer obvious reader-side benefits to those tracking new developments. But this may be a case where the author-side benefits swamp the reader-side benefits. Preprint exchanges give authors the earliest possible time stamp to mark their priority over others working on the same problem. (Historical aside: It's likely that in the seventeenth century, journals superseded books as the primary literature of science precisely because they were faster than books in giving authors an authoritative public time stamp.)

Preprint exchanges existed before the internet, but OA makes them faster, larger, more useful, and more widely read. Despite these advantages, however, preprint exchanges don't represent the whole OA movement or even the whole green OA movement. On the contrary, most green OA and most OA overall focuses on peer-reviewed articles.

As soon as scholars had digital networks to connect peers together, they began using them to tinker with

peer review. Can we use networks to find good referees, or to gather, share, and weigh their comments? Can we use networks to implement traditional models of peer review more quickly or effectively? Can we use networks to do better than the traditional models? Many scholars answer “yes” to some or all of these questions, and many of those saying “yes” also support OA. One effect is a creative and long-overdue efflorescence of experiments with new forms of peer review. Another effect, however, is the false perception that OA entails peer-review reform. For example, many people believe that OA requires a certain kind of peer review, favors some kinds of peer review and disfavors others, can’t proceed until we agree on the best form of peer review, or benefits only those who support certain kinds of peer-review reforms. All untrue.

OA is compatible with every kind of peer review, from the most traditional and conservative to the most networked and innovative. Some OA journals deliberately adopt traditional models of peer review, in order to tweak just the access variable of scholarly journals. Some deliberately use very new models, in order to push the evolution of peer review. OA is a kind of access, not a kind of editorial policy. It’s not intrinsically tied to any particular model of peer review any more than it’s intrinsically tied to any particular business model or method of digital preservation.

With one exception, achieving OA and reforming peer review are independent projects. That is, we can achieve

OA without reforming peer review, and we can reform peer review without achieving OA. The exception is that some new forms of peer review presuppose OA.

For example, *open review* makes submissions OA, before or after some prepublication review, and invites community comments. Some open-review journals will use those comments to decide whether to accept the article for formal publication, and others will already have accepted the article and use the community comments to complement or carry forward the quality evaluation started by the journal. Open review requires OA, but OA does not require open review.

Peer review does not depend on the price or medium of a journal. Nor does the value, rigor, or integrity of peer review. We know that peer review at OA journals can be as rigorous and honest as peer review at the best toll-access journals because it can use the same procedures, the same standards, and even the same people (editors and referees) as the best toll-access journals. We see this whenever toll-access journals convert to OA without changing their methods or personnel.

## 5.2 Theses and Dissertations<sup>3</sup>

Theses and dissertations are the most useful kinds of invisible scholarship and the most invisible kinds of useful

scholarship. Because of their high quality and low visibility, the access problem is worth solving.

Fortunately OA for electronic theses and dissertations (ETDs) is easier than for any other kind of research literature. Authors have not yet transferred rights to a publisher, no publisher permissions are needed, no publisher fears need be answered, and no publisher negotiations slow things down or make the outcome uncertain. Virtually all theses and dissertations are now born digital, and institutions expecting electronic submission generally provide OA, the reverse of the default for journal publishers.

The chief obstacle seems to be author fear that making a thesis or dissertation OA will reduce the odds that a journal will publish an article-length version. While these fears are sometimes justified, the evidence suggests that in most cases they are not.<sup>4</sup>

Universities expecting OA for ETDs teach the next generation of scholars how easy OA is to provide, how beneficial it is, and how routine it can be. They help cultivate lifelong habits of self-archiving. And they elicit better work. By giving authors a foreseeable, real audience beyond the dissertation committee, an OA policy strengthens existing incentives to do rigorous, original work.

If a university requires theses and dissertations to be new and significant works of scholarship, then it ought to expect them to be made public, just as it expects new and significant scholarship by faculty to be made public.

Sharing theses and dissertations that meet the school's high standard reflects well on the institution and benefits other researchers in the field. The university mission to advance research by young scholars has two steps, not one. First, help students produce good work, and then help others find, use, and build on that good work.

### 5.3 Books<sup>5</sup>

The OA movement focuses on journal articles because journals don't pay authors for their articles. This frees article authors to consent to OA without losing money. By contrast, book authors either earn royalties or hope to earn royalties.

Because the line between royalty-free and royalty-producing literature is bright (and life is short), many OA activists focus exclusively on journal articles and leave books aside. I recommend a different tactic: treat journal articles as low-hanging fruit, but treat books as higher-hanging fruit rather than forbidden fruit. There are even reasons to think that OA for some kinds of books is easier to attain than OA for journal articles.

The scope of OA should be determined by author consent, not genre. Imagine an author of a journal article who withholds consent to OA. The economic door is open but

the author is not walking through it. This helps us see that relinquishing revenue is only relevant when it leads to consent, and consent suffices whether or not it's based on relinquishing revenue. It follows that if authors of royalty-producing genres, like books, consent to OA, then we'll have the same basis for OA to books that we have for OA to articles.

Even if books are higher-hanging fruit, they're not out of reach. Two arguments are increasingly successful in persuading book authors to consent to OA.

1. Royalties on most scholarly monographs range between zero and meager. If your royalties are better than that, congratulations. (I've earned book royalties; I'm grateful for them, and I wish all royalty-earning authors success.) The case for OA doesn't ask authors to make a new sacrifice or leave money on the table. It merely asks them to weigh the risk to their royalties against the benefit of OA, primarily the benefit of a larger audience and greater impact. For many book authors, the benefit will outweigh the risk. The benefit is large and the realistic prospect of royalties is low.

2. There is growing evidence that for some kinds of books, full-text OA editions boost the net sales of the priced, printed editions. OA may increase royalties rather than decrease them.

The first argument says that even if OA puts royalties at risk, the benefits might outweigh the risks. The second argument says that OA might not reduce royalties at all, and that conventional publication without an OA edition might be the greater risk. Both say, in effect, that authors should be empirical and realistic about this. Don't presume that your royalties will be high when there's evidence they will be low, and don't presume that OA will kill sales when there's evidence it could boost them.

Both arguments apply to authors, but the second applies to publishers as well. When authors have already transferred rights—and the OA decision—to a publisher, then the case rests on the second argument. A growing number of academic book publishers are either persuaded or so intrigued that they're experimenting.<sup>6</sup>

Many book authors want a print edition, badly. But the second argument is not only compatible with print but depends on print. The model is to give away the OA edition and sell a print edition, usually via print-on-demand (POD).<sup>7</sup>

Why would anyone buy a print book when the full text is OA? The answer is that many people don't want to read a whole book on a screen or gadget, and don't want to print out a whole book on their printer. They use OA editions for searching and sampling. When they discover a book that piques their curiosity or meets their personal standards of relevance and quality, they'll buy a copy. Or, many of them will buy a copy.

Evidence has been growing for about a decade that this phenomenon works for some books, or some kinds of books, even if it doesn't work for others. For example, it seems to work for books like novels and monographs, which readers want to read from beginning to end, or which they want to have on their shelves. It doesn't seem to work for books like encyclopedias, from which readers usually want just an occasional snippet.

One problem is running a controlled experiment, since we can't publish the same book with and without an OA edition to compare the sales. (If we publish a book initially without an OA edition and later add an OA edition, the time lag itself could affect sales.) Another variable is that ebook readers are becoming more and more consumer friendly. If the "net boost to sales" phenomenon is real, and if it depends on the ergonomic discomforts of reading digital books, then better gadgets may make the phenomenon disappear. If the net-boost phenomenon didn't depend on ergonomic hurdles to digital reading, or didn't depend entirely on them, then it might survive any sort of technological advances. There's a lot of experimenting still to do, and fortunately or unfortunately it must be done in a fast-changing environment.<sup>8</sup>

The U.S. National Academies Press began publishing full-text OA editions of its monographs alongside priced, printed editions in March 1994, which is ancient history in internet time. Over the years Michael Jensen, its

director of web communications and director of publishing technologies, has published a series of articles showing that the OA editions increased the sales of the toll-access editions.<sup>9</sup>

In February 2007, the American Association of University Presses issued a Statement on Open Access in which it called for experiments with OA monographs and mixed OA/toll-access business models. By May 2011, the AAUP reported that 17 member presses, or 24 percent of its survey respondents, were already publishing full-text OA books.<sup>10</sup>

The question isn't whether some people will read the OA edition without buying the toll-access edition. Some will. The question isn't even whether more readers of the OA edition will buy the toll-access edition than not buy it. The question is whether more readers of the OA edition will buy the toll-access edition *than would have bought* the toll-access edition without the OA edition to alert them to its existence and help them evaluate its relevance and quality. If there are enough OA-inspired buyers, then it doesn't matter that there are also plenty of OA-satisfied nonbuyers.

Book authors and publishers who are still nervous could consent to delayed OA and release the OA edition only after six months or a year. During the time when the monograph is toll-access only, they could still provide OA excerpts and metadata to help readers and potential buyers find the book and start to assess it.

Even the youngest scholars today grew up in a world in which there were more print books in the average university library than gratis OA books online. But that ratio reversed around 2006, give or take. Today there are many more gratis OA books online than print books in the average academic library, and we're steaming toward the next crossover point when there will be many more gratis OA books online than print books in the world's largest libraries, academic or not.

A few years ago, those of us who focus on OA to journal literature were sure that journal articles were lower-hanging fruit than any kind of print books, including public-domain books. But we were wrong. There are still good reasons to make journal literature the strategic focus of the OA movement, and we're still making good progress on that front. But the lesson of the fast-moving book-scanning projects is that misunderstanding, inertia, and permission are more serious problems than digitization. The permission problem is solved for public-domain books. Digitizing them by the millions is a titanic technical undertaking, but it turns out to be a smaller problem than getting millions of copyrighted articles into OA journals or OA repositories, even when they're written by authors who can consent to OA without losing revenue. OA for new journal articles faces publisher resistance, print-era incentives, and misunderstandings in every category of stake-

holders, including authors and publishers. As the late Jim Gray used to say, “May all your problems be technical.”

#### 5.4 Access to What?<sup>11</sup>

Not all the literature that researchers want to find, retrieve, and read should be called knowledge. We want access to serious proposals for knowledge even if they turn out to be false or incomplete. We want access to serious hypotheses even if we’re still testing them and debating their merits. We want access to the data and analysis offered in support of the claims we’re evaluating. We want access to all the arguments, evidence, and discussion. We want access to everything that could help us decide what to call knowledge, not just to the results that we agree to call knowledge. If access depended on the outcome of debate and inquiry, then access could not contribute to debate and inquiry.

We don’t have a good name for this category larger than knowledge, but here I’ll just call it research. Among other things, research includes knowledge and knowledge claims or proposals, hypotheses and conjectures, arguments and analysis, evidence and data, algorithms and methods, evaluation and interpretation, debate and discussion, criticism and dissent, summary and review. OA to research should be OA to the whole shebang. Inquiry and research suffer when we have access to anything less.

We want access to all the arguments, evidence, and discussion. We want access to everything that could help us decide what to call knowledge, not just to the results that we agree to call knowledge. If access depended on the outcome of debate and inquiry, then access could not contribute to debate and inquiry.

Some people call the journal literature the “minutes” of science, as if it were just a summary. But it’s more than that. If the minutes of a meeting summarize a discussion, the journal literature is a large part of the discussion itself. Moreover, in an age of conferences, preprint servers, blogs, wikis, databases, listservs, and email, the journal literature is not the whole discussion. Wikipedia aspires to provide OA to a summary of knowledge, and (wisely) refuses to accept original research. But the larger OA movement wants OA to knowledge and original research themselves, as well as the full discussion about what we know and what we don’t. It wants OA to the primary and secondary sources where knowledge is taking shape through a messy process that is neither consistent (as it works through the clash of conflicting hypotheses) nor stable (as it discards weak claims and considers new ones that appear stronger). The messiness and instability are properties of a discussion, not properties of the minutes of a discussion. The journal literature isn’t just a report on the process but a major channel of the process itself. And not incidentally, OA is valuable not just for making the process public but for facilitating the process and making it more effective, expeditious, transparent, and global.<sup>12</sup>

To benefit from someone’s research, we need access to it, and for this purpose it doesn’t matter whether the research is in the sciences or humanities. We need access to medical or physical research before we can use it to tackle

a cure for malaria or devise a more efficient solar panel. We need access to an earthquake prediction before we can use it to plan emergency responses.<sup>13</sup> And we need access to literary and philosophical research in order to understand a difficult passage in Homer or the strength of a response to epistemological skepticism.

For this kind of utility, the relevant comparison is not between pure and applied research or between the sciences and humanities. The relevant comparison is between any kind of research when OA and the same kind of research when locked behind price and permission barriers. Whether a given line of research serves wellness or wisdom, energy or enlightenment, protein synthesis or public safety, OA helps it serve those purposes faster, better, and more universally.

## **5.5 Access for Whom?**

Answer: human beings and machines.

### **5.5.1 OA for Lay Readers<sup>15</sup>**

Some have opposed OA on the ground that not everyone needs it, which is a little like opposing the development of a safe and effective new medicine on the ground that not every one needs it. It's easy to agree that not everyone

needs it. But in the case of OA, there's no easy way to identify those who do and those who don't. In addition, there's no easy way, and no reason, to deliver it only to those who need it and deny it to everyone else.

OA allows us to provide access to everyone who cares to have access, without patronizing guesswork about who really wants it, who really deserves it, and who would really benefit from it. Access for everyone with an internet connection helps authors, by enlarging their audience and impact, and helps readers who want access and who might have been excluded by central planners trying to decide in advance whom to enfranchise. The idea is to stop thinking of knowledge as a commodity to meter out to deserving customers, and to start thinking of it as a public good, especially when it is given away by its authors, funded with public money, or both.<sup>14</sup>

Some lobbyists for toll-access publishers argue, in good faith or bad, that the goal of OA is to bring access to lay readers. This sets up their counter-argument that lay readers don't care to read cutting-edge research and wouldn't understand it if they tried. Some publishers go a step further and argue that access to research would harm lay readers.<sup>16</sup>

This is a two-step argument, that OA is primarily for lay readers and that lay readers don't need it. Each step is false. The first step overlooks the unmet demand for access by professional researchers, as if all professionals who

wanted access already had it, and the second overlooks the unmet demand for access by lay readers, as if lay readers had no use for access.

One reason to think the first step is put forward in bad faith is that it overlooks the very conspicuous fact that the OA movement is driven by researchers who are emphatic about wanting the benefits of OA for themselves. It also overlooks the evidence of wide and widespread access gaps even for professional researchers. (See section 2.1 on problems.)

The problem with the second step is presumption. How does anyone know in advance the level of demand for peer-reviewed research among lay readers? When peer-reviewed literature is toll-access and expensive, then lack of access by lay readers and consumers doesn't show lack of demand, any more than lack of access to Fort Knox shows lack of demand for gold. We have to remove access barriers before we can distinguish lack of access from lack of interest. The experiment has been done, more than once. When the U.S. National Library of Medicine converted to OA in 2004, for example, visitors to its web site increased more than a hundredfold.<sup>17</sup>

A common related argument is that lay readers surfing the internet are easily misled by unsupported claims, refuted theories, anecdotal evidence, and quack remedies. Even if true, however, it's an argument for rather than against expanding online access to peer-reviewed research.

If we're really worried about online dreck, we should dilute it with high-quality research rather than leave the dreck unchallenged and uncorrected.

Many of us medical nonprofessionals—who may be professionals in another field—want access to medical research in order to read about our own conditions or the conditions of family members. But even if few fall into that category, most of us still want access for our doctors, nurses, and hospitals. We still want access for the nonprofit advocacy organizations working on our behalf, such as the AIDS Vaccine Advocacy Coalition, the Cystinosis Research Network, or the Spina Bifida Association of America. And in turn, doctors, nurses, hospitals, and advocacy organizations want access for laboratory researchers. As I argued earlier (section 1.2), OA benefits researchers directly and benefits everyone else indirectly by benefiting researchers.<sup>18</sup>

A May 2006 Harris poll showed that an overwhelming majority of Americans wanted OA for publicly funded research. 83 percent wanted it for their doctors and 82 percent wanted it for everyone. 81 percent said it would help medical patients and their families cope with chronic illness and disability. 62 percent said it would speed up the discovery of new cures. For each poll question, a fairly large percentage of respondents checked “neither agree nor disagree” (between 13 and 30 percent), which meant that only tiny minorities disagreed with the OA propositions. Only 3 percent didn't want OA for their doctors, 4

percent didn't want it for themselves, and 5 percent didn't think it would help patients or their families.<sup>19</sup>

The ratio of professional to lay readers of peer-reviewed research undoubtedly varies from field to field. But for the purpose of OA policy, it doesn't matter what the ratio is in any field. What matters is that neither group has sufficient access today, when most research journals are toll-access. Professional researchers don't have sufficient access through their institutional libraries because subscription prices are rising faster than library budgets, even at the wealthiest libraries in the world. Motivated lay readers don't have sufficient access because few public libraries subscribe to any peer-reviewed research journals, and none to the full range.<sup>20</sup>

The argument against access for lay readers suffers from more than false assumptions about unmet demand. Either it concedes or doesn't concede that OA is desirable for professional researchers. If it doesn't, then it should argue first against the strongest opponent and try to make the case against OA for professionals. But if it does concede that OA for professionals is a good idea, then it wants to build a selection system for deciding who deserves access, and an authentication system for sorting the sheep from the goats. Part of the beauty of OA is that providing access to everyone is cheaper and easier than providing access to some and blocking access to others. We should only raise costs and pay for the apparatus of exclusion when there's a very good reason to do so.<sup>21</sup>

## 5.5.2 OA for Machines<sup>22</sup>

We also want access for machines. I don't mean the futuristic altruism in which kindly humans want to help curious machines answer their own questions. I mean something more selfish. We're well into the era in which serious research is mediated by sophisticated software. If our machines don't have access, then we don't have access. Moreover, if we can't get access for our machines, then we lose a momentous opportunity to enhance access with processing.

Think about the size of the body of literature to which you have access, online and off. Now think realistically about the subset to which you'd have practical access if you couldn't use search engines, or if search engines couldn't index the literature you needed.

Information overload didn't start with the internet. The internet does vastly increase the volume of work to which we have access, but at the same time it vastly increases our ability to find what we need. We zero in on the pieces that deserve our limited time with the aid of powerful software, or more precisely, powerful software with access. Software helps us learn what exists, what's new, what's relevant, what others find relevant, and what others are saying about it. Without these tools, we couldn't cope with information overload. Or we'd have to redefine "coping" as artificially reducing the range of work we are allowed to consider, investigate, read, or retrieve.<sup>23</sup>

Some publishers have seriously argued that high toll-access journal prices and limited library budgets help us cope with information overload, as if the literature we can't afford always coincides with the literature we don't need. But of course much that is relevant to our projects is unaffordable to our libraries. If any problems are intrinsic to a very large and fast-growing, accessible corpus of literature, they don't arise from size itself, or size alone, but from limitations on our discovery tools. With OA and sufficiently powerful tools, we could always find and retrieve what we needed. Without sufficiently powerful tools, we could not. Replacing OA with high-priced toll access would only add new obstacles to research, even if it simultaneously made the accessible corpus small enough for weaker discovery tools to master. In Clay Shirky's concise formulation, the real problem is not information overload but filter failure.<sup>24</sup>

OA is itself a spectacular inducement for software developers to create useful tools to filter what we can find. As soon as the tools are finished, they apply to a free, useful, and fast-growing body of online literature. Conversely, useful tools optimized for OA literature create powerful incentives for authors and publishers to open up their work. As soon as their work is OA, a vast array of powerful tools make it more visible and useful. In the early days of OA, shortages on each side created a vicious circle: the small quantity of OA literature provided little incentive

to develop new tools optimized for making it more visible and useful, and the dearth of powerful tools provided little extra incentive to make new work OA. But today a critical mass of OA literature invites the development of useful tools, and a critical mass of useful tools gives authors and publishers another set of reasons to make their work OA.

All digital literature, OA or toll access, is machine-readable and supports new and useful kinds of processing. But toll-access literature minimizes that opportunity by shrinking the set of inputs with access fees, password barriers, copyright restrictions, and software locks. By removing price and permission barriers, OA maximizes this opportunity and spawns an ecosystem of tools for searching, indexing, mining, summarizing, translating, querying, linking, recommending, alerting, mashing-up, and other kinds of processing, not to mention myriad forms of crunching and connecting that we can't even imagine today. One bedrock purpose of OA is to give these research-enhancing, utility-amplifying tools the widest possible scope of operation.

In this sense, the ultimate promise of OA is not to provide free online texts for human reading, even if that is the highest-value end use. The ultimate promise of OA is to provide free online data for software acting as the antennae, prosthetic eyeballs, research assistants, and personal librarians of all serious researchers.

Opening research literature for human users also opens it for software to crunch the literature for the benefit of human users. We can even hope that OA itself will soon be old hat, taken for granted by a new generation of tools and services that depend on it. As those tools and services come along, they will be the hot story and they will deserve to be. Technologists will note that they all depend on OA, and historians will note that OA itself was not easily won.<sup>25</sup>

