

Open Access and the Last-Mile Problem for Knowledge

From “Open access and the last-mile problem for knowledge,” SPARC Open Access Newsletter, July 2, 2008.

<http://dash.harvard.edu/handle/1/4322587>

After Hurricane Katrina hit the US gulf coast in August 2005, the Federal Emergency Management Agency (FEMA) bought 11,000 mobile homes for \$431 million and shipped them to Arkansas for the evacuees. Six months later the homes were still sitting unused in an Arkansas cow pasture because federal rules—ironically, FEMA rules—prohibited the use of mobile homes in a flood plain.

<http://tinyurl.com/4rks8t>

In September 2005, Britain donated \$5.3 worth of military rations to Americans displaced by Katrina. A month later the food was still sitting on a tarmac at Little Rock Air Force Base while officials tried to figure out whether US rules banning British beef allowed them to distribute the food to the needy. Meantime, the US was paying \$16,000/month to store the food, and its expiration date was approaching fast.

<http://www.washingtonpost.com/wp-dyn/content/article/2005/10/13/AR2005101302084.html>

When Cyclone Nargis hit the coast of Burma in May 2008, dozens of governments around the world shipped relief supplies to the country. At first the Burmese junta refused to accept the aid; then it accepted the aid but not aid workers; then it accepted aid and air workers but not ships or helicopters. Even after allowing aid into the country, much of it was stolen by the Burmese military and much was delayed on airport tarmacs or offshore until it was unusable. On June 5, more than a month after the disaster, a carrier

group of four US ships returned to the US after being denied permission to unload its relief supplies.

http://en.wikipedia.org/wiki/Cyclone_Nargis#Military_junta.27s_blockade_of_aid

You could call this the “tarmac problem” for disaster relief. Or you could use a telecom analogy and call it a disastrous version of the last-mile problem.

In telecommunications the “last-mile problem” is the problem of connecting individual homes and businesses to the fat pipes connecting cities. Because individual homes and businesses are in different locations, hooking up each one individually is expensive and difficult. The term is now used in just about every industry in which reaching actual customers is more difficult than reaching some location, like a store or warehouse, close to customers.

We're facing a last-mile problem for knowledge. We're pretty good at doing research, writing it up, vetting it, publishing it, and getting it to locations (physical libraries and web sites) close to users. We could be better at all those things, but any problems we encounter along the way are early- or mid-course problems. The last-mile problem is the one at the end of the process: making individualized connections to all the individual users who need to read that research.

The last-mile problem for knowledge is not new. Indeed, for all of human history until recently it has been inseparable from knowledge itself and all our technologies for sharing it. It's only of interest today because the internet and OA give us unprecedented means for solving it, or at least for closing the gap significantly.

The problem is not that librarians “warehouse” knowledge in the pejorative sense of that term. On the contrary, they go out of their way to help users find and retrieve what the library has to offer, and often do the same for much beyond the library as well. The problem is to make individualized connections between knowledge, wherever it lies, and users, wherever they are. Even a well-stocked and well-organized library staffed by well-trained librarians can only solve a subset of that problem and connect a subset of users with a subset of knowledge.

A journal is many things, for example a collection, a periodical, a brand, and a peer-review filter. But it's also a tarmac. It's not the final destination for new research, just a landing place close to the final destination. If the articles inside only reach the journal and not the readers beyond, or if they only reach some but not all of the readers who need to read them, then every player up and down the chain of scholarly communication is frustrated: authors, author funders, author employers, editors, referees, publishers, librarians, and readers.

It helps to distinguish two reader-side stages of the last-mile problem for knowledge. Stage One is getting access to texts or data, and Stage Two is getting answers to questions. The first treats scholarly communication as a delivery system. When there's

a problem, it's the failure to complete the delivery. The second treats scholarly communication as a knowledge system. When there's a problem, it's the failure to convey understanding.

Consider the difference between a conference lecture and the question period afterwards. The lecture is accessible to the people in the room, even unavoidable to them. There's no Stage One problem. But not even a good speaker can customize the talk for every member of the audience. For some listeners, the talk may be in the wrong language, at the wrong speed, at the wrong level of abstraction, or on the wrong subtopics. It may presuppose too much background or too little. It can still leave an unclosed gap between the speaker's knowledge and a given listener's understanding. The Q&A period can close that gap, at least when people with questions actually ask them and people with answers, perhaps the speaker, actually answer them.

Unfortunately, most existing knowledge isn't even as close to us as the conference lecture, let alone the individualized answer to our question. That's one reason, by the way, why Plato preferred speech to writing. Speakers are interactive and can close knowledge gaps in real-time Q&A, while writers are generally unavailable for interrogation about their writings, sometimes for the good reason that they are dead. Plato had a point: speech usually surpasses writing at solving the Stage Two problem. But the reverse is true for the Stage One problem. If we had to depend on live speakers for transmitting knowledge, when knowledgeable speakers were few and far between, then Stage One of the last-mile problem would be much more difficult than it already is.

First Things First: Stage One

You solve the last-mile problem for a published journal article when you put your hands on a hardcopy or display a digital copy on a screen in front of your face. This requires open access (OA) or money to pay for toll access (TA).

Acknowledging that money solves the problem, at least for some researchers, is just as important as acknowledging its limitations as a solution. It works for lucky individuals who have the money or who work at institutions that have the money. The snag, of course, is that all of us are unlucky for some priced literature, and most of us are unlucky for most of it.

The fact that the money solution doesn't work for everyone is the chief reason why the last-mile problem is a problem. Paying to make individualized connections for *some* individuals is clearly feasible and clearly affordable. But the problem is to make individualized connections for all individuals, or all those who need connections. Money just doesn't scale to the size of this problem. If the supply of published knowledge were fixed, money might have a chance to catch up with the demand. But the supply is growing exponentially and money to buy access to it is not. Inevitably,

then, as the volume of TA literature grows, the percentage of it accessible to the average researcher declines, and the faster the volume of TA literature grows, the faster the accessible percentage declines. If all literature put a price on access, or if money were the only solution to the last-mile problem, the problem would worsen every year.

OA is the only solution that scales to the full size of the problem and keeps pace with the growth of published knowledge. No matter how fast the OA literature grows, you'll only need an internet connection to have access to all of it.

<http://www.earlham.edu/~peters/fos/newsletter/03-02-04.htm#scaling>

Until the 19th century, mailed letters were like scholarly journals: the costs were paid by readers or recipients. Senders could send mail without charge, but recipients had to pay to pick up their mail from the post office. If they couldn't pay, they had to do without. They often had to do without, creating a last-mile problem begging for a solution. Rowland Hill introduced the postage stamp in 1837 precisely to shift costs from recipients to senders. The sender-pays model made the system scale for the first time, triggered an explosion in the use of mail, and solved the last-mile problem.

<http://www.earlham.edu/~peters/fos/newsletter/03-11-02.htm#analogies>

When I say that the money solution doesn't scale, I mean money to pay for access to published literature, not money to do the research or publish the literature in the first place. That is, I mean money to solve the last-mile problem, not money to solve one of the early- or mid-course problems. Hence, the conclusion that the OA solution works better than the money solution doesn't imply that publishing can be made costless, any more than mail delivery can be made costless. The question raised by OA is not whether the production costs can be reduced to zero, but whether there are better ways to pay the bills than by charging readers, creating access barriers, and aggravating the last-mile problem.

Non-OA publishers, even those who lobby hard against OA policies, want to solve the last-mile problem as much as anyone. Non-profit or for-profit, green or gray, they have nothing to gain by leaving end-users disconnected from the knowledge they publish. The difference is that they want to steer stakeholders toward the money solution, not the OA solution. That's why Elsevier's Crispin Davis used to argue "that the government needs to lay down guidelines on the proportion of university funds that should be set aside for the acquisition of books and journals, or even increase funding to ensure that universities can buy all the material they need. ..."

<http://www.theguardian.com/business/2005/feb/19/science.comment>

<http://www.earlham.edu/~peters/fos/2005/02/more-from-elsevier-ceo-on-university.html>

If the telecom version of the last-mile problem had to be solved with copper wire or optic fiber, it would be much more difficult and expensive than it need be. Wireless

connectivity solves the last-mile problem at a stroke for everyone with the right equipment. OA is the analog of wireless in the last-mile problem for knowledge. OA solves the Stage One problem at a stroke for everyone with an internet connection. Wireless and OA are revolutionary shortcuts that make connections to individual users without the need for individualized labor and expense.

If OA doesn't solve the problem for literally everyone, it's only because we haven't finished solving the telecom version of the last-mile problem. If both the money and OA solutions leave some people out, at least the OA solution will leave out fewer and fewer people as the digital divide continues to shrink, and the money solution will leave out more and more people as the volume of TA literature continues to grow.

Stage Two

Suppose you have a question. You're lucky if some careful, curious researchers have already asked the same question and done some of the needed research. You're luckier if some of them have taken the research far enough to answer the question, write up their answers, win the approval of peer reviewers, and publish them. You're even luckier if there's a scientific consensus on the right answer to your question and that among the published papers on it, at least one is up to date, written in your language, and written at your level of understanding. You're even luckier if the Stage One problem has been solved and, thanks to OA or money, you have access to at least one of the enlightening papers which meets all your conditions.

It may look like this scenario goes about as far as it can to close the gap between you and existing knowledge. But it leaves some nagging parts of what I'm calling Stage Two of the last-mile problem. How do you go beyond access to answers? We grant that you're darned lucky, and that if you could find one of the enlightening papers, then you could retrieve it, and if you could read it, then you could understand it. But not all published papers meet your conditions for an enlightening paper. In fact, nearly all of them don't. How do you know that an enlightening paper even exists? When you go looking, how can you find one that meets your conditions, and distinguish it from other papers which happen to use the same keywords or even address the same question?

Without solutions to these problems, you might as well be trapped in a maze knee-deep in conflicting maps thrown over the wall by people trying their utmost to be helpful.

For people with less luck, Stage Two problems are more numerous and more difficult. How do you do find a good answer when there's no consensus? When there is a consensus answer, how do you learn what it is when papers describing it are mixed together in your search results with papers describing discredited answers? How do you learn the consensus answer when there isn't a good paper in your language or at your level of understanding, or when the best papers use terms you'd never think to use

in your search query? How do you get answers when nobody has yet posed the question exactly as you have posed it, and when partial answers lie scattered in dozens or hundreds of different papers in different journals in different languages and even different fields?

To solve these problems, access to the papers is necessary but not sufficient. But while OA is only part of the solution to the Stage Two problem, it's a precondition to most other parts of the solution. No tools yet suffice to solve the Stage Two problem, and maybe no tools ever will. But the tools that help us inch toward a solution presuppose OA literature and data the way telescopes presuppose open access to the sky. In fact, one of the primary benefits of OA is to provide the inputs to a new generation of sophisticated tools to facilitate research, discovery, and analysis. Whatever methods we use to attack Stage Two problems, OA will streamline our solutions and lack of OA will limit their scope and slow us down.

We already have some means to help us solve the Stage Two problem. Some are fairly mature and some are very rudimentary, but in every case talented people are working hard to improve them. I'm thinking of means to learn about the existence of relevant new work (alert systems), find the texts and the passages we need (search engines), find work already found by colleagues (tagging and social networking systems), find articles similar to ones we know to be relevant (recommendation systems), find articles in our own language (machine translation), navigate to cited sources (reference linking), navigate to different versions of cited sources or other relevant destinations (multiple-resolution hyperlinks), convert a text to speech when we can't read the screen (voice readers), paraphrase articles we don't have time to read (text summarizers), digest larger volumes of literature than we could ever read (text mining), combine independent resources to create new synergies and utility (mash-ups), find information relevant to our questions even when we don't know the relevant keywords (semantic web), distill uncopyrightable facts from natural-language texts and enter them into queryable OA databases (knowledge extraction), pose our search queries in our own words and sometimes even get back direct answers rather than mere pointers to literature that may contain answers (natural language search engines).

Most Stage Two problems can only be solved with human judgment. But that doesn't rule out the possibility of technologies to lend us a hand. The reason is simply that we are building technologies that harness human judgments, at least when those judgments are digital, online, and accessible to the tools. Stage Two solutions don't require machine-generated answers to our questions or magical forms of artificial intelligence. They only require barrier-free access to human-generated answers, human evaluations of those answers, and human evaluations of the evaluations. Tools to do these jobs are multiplying, they are improving, and they are interconnecting so that the output of one is the input to another. We don't have to predict the future in order to know that this kind of incremental, recursive progress can

continue indefinitely, just as the compounding of mathematical functions can continue indefinitely.

As long as the last-mile problem remains unsolved, rapidly growing human knowledge will coexist with rapidly growing unmet demand for that knowledge. As long as the problem remains unsolved, the uses we make of recorded knowledge will fall far short of its usefulness.

It's staggering to think about what could happen if the knowledge we have painstakingly discovered, articulated, tested, refined, validated, gathered, and delivered to the tarmac were systematically distributed to all who need it. Imagine if what was known became more widely known, especially among those who could put it to use. Imagine if we became even 10% more effective at using what we know.

