Transcript: Brewster Kahle builds a free digital library. (Text available at TED Talks.)

We really need to put the best we have to offer within reach of our children. If we don't do that, we're going to get the generation we deserve. They're going to learn from whatever it is they have around them.

And we, as now the elite, parents, librarians, professionals, whatever it is, a bunch of our activities are, in fact, in trying to get the best we have to offer within reach of those around us, or as broadly as we can. I'm going to start and end this talk with a couple things that are carved in stone. One is what's on the Boston Public Library. Carved above their door is, "Free to All." It's kind of an inspiring statement, and I'll go back at the end of this. I'm a librarian, and what I'm trying to do is bring all of the works of knowledge to as many people as want to read it. And the idea of using technology is perfect for us. I think we have the opportunity to one-up the Greeks. It's not easy to one-up the Greeks. But with the industriousness of the Egyptians, they were able to build the Library of Alexandria -- the idea of a copy of every book of all the peoples of the world. The problem was, you actually had to go to Alexandria to go to it. On other hand, if you did, then great things happened. I think we can one-up the Greeks and achieve something. And I'm going to try to argue only one point today: that universal access to all knowledge is within our grasp. So if I'm successful, then you'll actually come away thinking, yeah, we could actually achieve the great vision of everything ever published, everything that was ever meant for distribution, available to anybody in the world that's ever wanted to have access to it.

Yes, there's issues about how money should be distributed and that's still being refigured out. But I'd say there's plenty of money, and there's plenty of demand, so we can actually achieve that. But I'm going to go over the technological, social and sort of where are we as a whole trying to get to that particular vision. And the way I'm going to try to do this is do it like the Amazon.com website -- the books, music, video and just go step, media type by media type, just go and say, all right, how we doing on this?

So if we start with books, you know, sort of where are we? Well, first you have to, as an engineer, scope the problem. How big is it? If you wanted to put all of the published works online so that anybody could have it available, well, how big a problem is it? Well, we don't really know, but the largest print library in the world is the Library of Congress -- it's 26 million volumes, 26 million volumes. It's by far and away the largest print library in the world. And a book, if you had a book, is about a megabyte, so -- you know, if you had it in Microsoft Word. So a megabyte, 26 million megabytes is 26 terabytes, it goes mega, giga, tera, 26 terabytes. 26 terabytes fits in a computer system that's about this big, on spinning Linux drives, and it costs about 60,000 dollars. So for the cost of a house -- or around here, a garage -- you can put -- you can have spinning all of the words in the Library of Congress. That's pretty neat.

Then the question is: what do you get? You know, is it worth trying to get there? Do you actually want it online? Some of the first things that people do is they make book readers that allow you to search inside the books, and that's kind of fun. And you can download these things and look around them in new and different ways. And you can get at them remotely, if you happen to have a laptop. There's starting to be some of these sort of page turn-ee interfaces that look a whole lot like books in certain ways, and you can search them, make little tabs, and it's kind of cute -- still very book-like -- on your laptop. But I don't know, reading things on a laptop -- whenever I pull up my laptop it always feels like work. I think that's one of the reasons why the Kindle is so great. I don't have to feel like I'm at work to read a Kindle; it's starting to be a little bit more specified. But I have to say that there's older technologies that I tend to like. I like the physical book. And I think we can go and use our technology to go and digitize things -- put them on the net and then download, print them and bind them and end up with books again.

And we sort of said, well, how hard is this? And it turns out to not be very hard. We actually went off to make a bookmobile. And a bookmobile -- the size of a van with a satellite dish, a printer, binder and cutter, and kids make their own books. It costs about three dollars to download, print and bind a normal old book. And they actually come out kind of nice looking. You can actually get really good-looking books for on the order of one penny per page, sort of the parts cost for doing this.

So the idea of this technology actually may end up putting books back in people's hands again. There are some other bookmobiles running around. This is Eric Eldred making books at Walden Pond, Thoreau's works. This is just before he got kicked out by the Parks Services for competing with the bookstore there. In India, they've got another couple bookmobiles running around. And this is the opening day at the Library of Alexandria, the new Library of Alexandria, in Egypt. It was quite popularly attended. And kids starting to make their own books, and a happy kid with the first book that he's ever owned. So the idea of being able to use this technology to end up with paper where I can handle sort of sounds a little retro, but I think it still has its place. And being sort of from the Silicon Valley, sort of Utopia, and -- sort of, you know, sort of world, we thought if we can make this technology work in rural Uganda, we might have something. So we actually got some funding from the World Bank to try it out. And we found in about 30 days we could go and take a couple folks from Silicon Valley, fly them to Uganda, buy a car, set up the first internet connection at the National Library of Uganda, figure out what they wanted, and get a program going making books in rural Uganda. And it actually -- so technologically, it works.

What we found out of this is, we didn't have the right books. So the books were in the library. We could get it to people if they're digitized, but we didn't know how to quite get them digitized. Everybody thought the answer is, send things to India and China. And so we've tried that, and I'll go over that in a moment. There are some newer technologies for delivering that have happened that are actually quite exciting as well. One is a print-on-demand machine that looks like a Rube Goldberg

Machine. We have one of these things now. It's completely cool. It's all conveyor belt and it makes a book. And it's called the "Espresso Book Machine," and in about 10 minutes you can press a button and make a book.

Something else I'm quite excited about in this particular domain, beyond these sort of kiosky things where you can get books on demand, is some of these new little screens that are coming out. And one of my favorites in this is the 100 dollar laptop. And I don't mean to steal any thunder here, but we've gone and used one of these things to be an e-book reader. So here's one of the beta units and you can -- it actually turns out to be a really good-looking e-book reader. And we have a quick hack that we did to try to put one of our books on it, and it turns out that 200 dots per inch means that you can put scanned books on them that look really good. At 200 dots per inch, it's kind of the equivalent of a 300-dot-print laser printer. We're in good enough shape. You actually can go and read scanned books quite easily.

So the idea of electronic books is starting to come about. But how do you do all this scanning? So we thought, okay, well, let's try out this send books to India thing. And there was a project with -- funded by the National Science Foundation, sent a bunch of scanners, and the American libraries were supposed to send books. Well, they didn't -- they didn't want to send their books. So we bought 100,000 books and sent them to India. And then we learned why you don't want to send books to India. The lesson we learned out of this is, scan your own books. If you really care about books, you're going to scan them better, especially if they're valuable books. If they're new books and you can just, you know, butcher them because you could just buy another one, that's not such a big deal in terms of doing high-quality scanning. But do things that you love. But the Indians have been scanning a lot of their own books -- about 300,000 now -- doing very well. The Chinese did over a million and the Egyptians are about 30,000.

But we sent -- thought, OK, if we're going to need to do this, let's do it in-library. How do we go and do this, and how do we get it down so that it's a cost point that we could afford? And we sort of picked the price point of 10 cents a page. If it's basically the cost of Xeroxing to digitize, OCR, package it up, make it so that you could download, print and bind it, the whole shebang, we would have achieved something. So we started out trying to figure out, how do we get to 10 cents? And we tried these robot things, and they worked pretty well -- sort of these auto page-turning things. If we can have Mars Rovers, you'd think you could turn pages. But it actually turns out to be pretty hard to turn pages, and the volume isn't there. So anyway -- so we ended up making our own book scanner. And with two digital high grade professional digital cameras, controlled museum lighting -- so even if it's a black and white book, you can go and get the proper intonation. So you basically do a beautiful, respectful job. This is not a fax, this is -- the idea is to do a beautiful job as you're going through these libraries. And we've been able to achieve 10 cents a page if we run things in volumes. This is what it looks like at the University of Toronto. And actually it turns out to, you know, pay a living wage. People seem to love it. Yes, it's a little boring, but some people kind of get into the Zen of it.

(Laughter) And especially if it's kind of interesting books that you care about in languages that you can read, We actually have been able to do a pretty good job of this at getting 10 cents a page. So 10 cents a page, 300 pages in your average book, 30 dollars a book. The Library of Congress, if you did the whole darn thing -- 26 million books -- is about 750 million dollars, right? But a million books -- I think actually would be a pretty good start, and that would cost 30 million dollars. That's not that big a bill.

And what we've been able to do is get into libraries. We've now got eight of these scanning centers in three countries, and libraries are up for having their books scanned. The Getty here is moving their books to the UCLA, which is where we have one these scanning centers, and scanning their out-of-copyright books, which is fabulous. So we're starting to get the institutional responsibility. The thing we're missing is the 10 cents. If we can get the 10 cents, all the rest of it flows. We've scanned about 200,000 books. Now we're scanning about 15,000 books a month, and it's starting to gear up another factor of two from there.

So all in all, that's going very well. And we're starting to move out of the just out-of-copyright into the out-of-print world. So I think of -- we're kind of going from the out-of-copyright library stuff, and Amazon.com is coming from the in-print world, and I think we'll meet in the middle some place and have the classic thing that you have, which is a publishing system and a library system working in parallel. And so we're starting up a program to do out-of-print works, but loaning them. Exactly what loaning means, I'm not quite sure. But anyway, loaning out-of-print works from the Boston Public Library, the Woods Hole Oceanographic Institute and a few other libraries that are starting to participate in this program, to try out this model of where does a library stop and where does the bookstore take over. So all in all it's possible to do this in large scale. We're also going back over microfilm and getting that online. So we can do 10 cents a page, we're going 15,000 books a month and we've got about 250,000 books online, counting all the other projects that are starting to add in. So what I wanted to argue is, books are within our grasp. The idea of taking on the whole ball of wax is not that big a deal. Yes, it costs tens of millions, low hundreds of millions, but one-time shot and we've got basically the history of printed literature online. And then there's business model issues about how to try to effectively market it and get it to people. But it is within our grasp technologically and law-wise, at least for the out-of-print and out-of-copyright, we suggest, to be able to get the whole darn thing online.

Now let's go for audio, and I'm going to go through these. So how much is there? Well, as best we can tell, there are about two to three million disks having been published -- so 78s, long-playing records and CDs -- or at least that's the largest archives of published materials we've been able to sort of point at. It costs about 10 dollars a piece to go and take a disk and put it online if you're doing things in volume. But we've found that the rights issues are really quite thorny. This is a fairly heavily litigated area, so we've found that there are niches in the music world that aren't served terribly well by the classic commercial publishing system. And we've

been starting to make these available by going and offering shelf space on the net. In the United States it doesn't cost you to give something away. Right? If you give something to a charity or to the public, you get a pat on the back and a tax donation -- except on the net, where you can go broke. If you put up a video of your garage band and it starts getting heavily accessed, you can lose your guitars or your house.

This doesn't make any sense. So we've offered unlimited storage, unlimited bandwidth forever, for free, to anybody that has something to share that belongs in a library. And we've been getting a lot of takers. One is the rock 'n rollers. The rock 'n rollers had a tradition of sharing, as long as nobody made any money. You could -- Concert recordings, it's not the commercial recordings, but concert recordings, started by the Grateful Dead. And we get about two or three bands a day signing up. They give permission, and we get about 40 or 50 concerts a day. We have about 40,000 concerts, everything the Grateful Dead ever did, up on the net so that people can see it and listen to this material. So audio is possible to put up, but the rights issues are really pretty thorny. We've got a lot of collections now -- a couple hundred thousand items -- and it's growing over time.

Moving images: if you think of theatrical releases, there are not that many of them. As best we can tell, there are about 150,000 to 200,000 movies ever that are really meant for a large-scale theatrical distribution. It's just not that many. But half of those were Indian. But anyway, it's doable, but we've only found about a thousand of these things that -- to be out-of-copyright. So we've digitized those and made those available. But we've found that there's lots of other types of movies that haven't really seen the light of day -- archival films. We've found, also, a lot of political films, a lot of amateur films, all sorts of things that are basically needing of a home, a permanent home. So we've been starting to make these available and it's grown to be very popular. We're not quite a YouTube; we tended towards longer-term things and also things that people can reuse and make into new movies, which has just been great fun.

Television comes quite a bit larger. We started recording 20 channels of television 24 hours a day. It's sort of the biggest TiVo box you've ever seen. It's about a petabyte, so far, of worldwide television -- Russian, Chinese, Japanese, Iraqi, Al Jazeera, BBC, CNN, ABC, CBS, NBC -- 24 hours a day. We put -- we only put one week up, which is mostly for cost reasons, which is the 9/11, sort of from 9/11/2001: for one week, what did the world see? CNN were saying that Palestinians were dancing in the streets. Were they? Let's look at the Palestinian television and find out. How can we have critical thinking without being able to quote and being able to compare what happened in the past? And television is dreadfully unrecorded and unquotable, except by Jon Stewart, who does fabulous job. So anyway, television is, I would suggest, within our grasp. So 15 dollars per video hour and also about 100 dollars to 150 dollars per celluloid hour, we're able to go and get materials online very inexpensively and have them up on the net. And we've got, now, a lot of these materials. So we've got about 100,000 pieces up there. So books, music, video, software -- there's only 50,000 titles of it. Mostly the issues there are legal issues

and breaking copy protections. But we've worked through some of those, but we've still got real problems in Washington.

Well, we're best known as the World Wide Web. We've been archiving the World Wide Web since 1996. We take a snapshot of every website and all of the pages on it, every two months. And actually, it's really been pioneered by Alexa Internet, which donates this collection to the internet archive. And it's been growing along for the last 11 years and it's a fantastic resource. And we've made a "Way Back Machine" that you can then go and see old websites kind of the way they were. If you go and search on something, this is Google.com, the different versions of it that we have, this is what it looks like when it was an Alpha release and this is what it looked like at Stanford. So anyway, you've got basically an idea of where things came from. Mostly people want to see their old stuff out of this. If there's one thing that we want to learn from the Library of Alexandria version one, which is probably best known for burning, is: don't just have one copy. So we've started to -- We've made another copy of all of this and we actually put it back in the Library of Alexandria. So this is a picture of the internet archive at the Library of Alexandria. And we now have also another copy building up in Amsterdam. So we should put it in the San Andreas Fault Line in San Francisco, flood zone in Amsterdam and in the Middle East. Right, so anyway ... so we're hedging our bets here. If we go and put it in a couple more places, I think we'll be in good shape.

There's a political and social question out of this. Is all of this, as we go digital, is it going to be public or private? There's some large companies that have seen this vision, that are doing large-scale digitization, but they're locking up the public domain. The question is: is that the world that we really want to live in? What's the role of the public versus the private as things go forward? How do we go and have a world where we both have libraries and publishing in the future, just as we basically benefited as we were growing up? So universal access to all knowledge -- I think it can be one of the greatest achievements of humankind, like the man on the moon or the Gutenberg Bible or the Library of Alexandria. It could be something that we're remembered for for millennia for having achieved. And as I said before, I'll end with something that's carved above the door of the Carnegie Library -- Carnegie -- one of the great capitalists of this country -- carved above his legacy: "Free to the People." Thank you very much.