Purdue University

Purdue e-Pubs

Building Trust When Truth Fractures

Brewster Kahle
Internet Archive

Building Trust When Truth Fractures

Brewster Kahle, Internet Archive

The following is a lightly edited transcript of a live keynote presentation at the 2019 Charleston Conference. A video of this talk is available at https://youtu.be/bA67X9y-ozc.

Courtney McAllister: Good morning. It is an honor to introduce our opening keynote speaker, Brewster Kahle. We are delighted to welcome Brewster back to Charleston and I'm sure you all are eager to hear his thoughts on how digitization efforts can help counteract the widespread dissemination of fake news. Brewster founded the Internet Archive in 1996 but even before that he was an influential figure transforming the ways we generate, archive, and access information. Shortly after graduating from MIT, he was the lead engineer on Thinking Machine's parallel supercomputer Connection Machine. In 1989 he created WAIS, Wide Area Information Server, which was the Internet's first publishing system. In 1996 he cofounded Alexa Internet, which was subsequently sold to Amazon three years later, and that's just a tantalizing snapshot of his noteworthy contributions. Brewster's vision and passion for expanding access to information has impacted all of us working in libraries and publishing. Please join me in welcoming Brewster Kahle.

Brewster Kahle: So, thank you to Beth and Anne for inviting me back to the Charleston Conference. I've been really, really looking forward to this. Thank you. This can be our day. This should be our day. I would say actually this has to be our day. People need us. We now have a lot of information out there on the open Internet that isn't true, that is specifically not true, and people are looking for better, solid information out there. They need the libraries, they need publishers, they need authors, they need access to materials in new and different ways to be able to go and understand their world. It's important to do and it's important to do now.

Let me start with a story. Three years ago, there was an election, a presidential election in the United States and no matter who you wanted to be president of the United States, how it all came down was a train wreck. We had all sorts of things that just went really wrong, how the journalism system didn't function very well, for an interference, as injected into the Internet, into the United States election

system, people starting to not trust the actual voting machines, the primary system sort of not reflecting what the majority of people wanted, all sorts of things kind of showed all sorts of creaks and problems and the like, and part of this was these Internet companies that were pursuing maybe their own interests to the expense of their users. I think there's a growing idea and a growing understanding that the Web is betraying us, that it's not working very well, it's not on our side anymore, it's not from us, by us, it seems to be coming at us filtered by people and reasons that aren't reflecting what it is we want to have happen.

This is horrifying, right? I'm really in the center of this whole Internet information generation and I didn't know what maybe the Internet Archives, or some of the other organizations I work with, their responsibility was for what happened three years ago but more importantly, probably I didn't know what to do next. So, I looked for help from my friends. I'm part of an executive director club. It's a set of the executive directors of the high-tech open nonprofits and I said, "Time to get together for dinner." This is the head of Wikipedia, the head of Electronic Frontier Foundation, the head of Mozilla Foundation, Public Library of Science, so I said, "Please come to the Internet Archive and let's have dinner." So, they all came, and I said, "What happened here? And what should we do next?" And I offered to put the Internet Archive on the table. I said, "What is the Internet Archive for now? What should we be? How should we operate? What can we do that would do the maximum public good with the skills that we have?" There was a lot of conversation through this and Katherine Maher, she's the executive director of Wikipedia, she said something that chilled me to my spine. She said that she was worried that "truth might fracture." Truth might fracture. I said, "That sounds horrible. What does this mean?" She said Wikipedia is built on the idea that on any particular subject a truth will emerge, true that consensus, at least, that it will be pushes and shoves but it will stay coherent and that there's something that will evolve to be a consensus of you on that particular subject. She was worried that it might become two or three or four or this idea that you can have a consensus idea of what's going on might not work anymore. So, that would be bad and when I think about where do

people turn for some concept of truth out there on the open Internet and it's probably Wikipedia, We refer to it all the time, millions of people every day go and use this service and it's a crowd source system and it's got these sort of interesting characters but it works very well.

So, how can we go and reinforce Wikipedia? I asked her, "How does this work? And how would truth fracture?" She said, "Citation wars." I'd never heard that term before: citation wars. Citation wars. She said that behind the scenes on every Wikipedia page there's a debate going on and people are trying to figure out what are the assertions that should be there and what shouldn't be there and the way those debates happen is based on the weight of the citations and the citations, they have a rule in Wikipedia that they bias toward things that you can click on and see, so that makes sense from a Wikipedia reader's perspective, you'd want to be able to click on a footnote and see why they think that that is in fact true so you could fact check it or you could go deeper. Makes sense. But that means that we really leave out a lot of material from being cited in Wikipedia because it may not have ever been digitized. If it's a book it may not have ever been digitized. If it is it might just be snippets in Google world. It might not be available to a Web user at all. It might be in a Kindle version but it's not in the Web world. So, the books may not be there, the journal literature may not be there, and the Web links might not be there. So, I committed myself at that point and the Internet Archive to help reinforce Wikipedia, that we were going to go and try to help Wikipedia become as good and solid an information resource going forward as it has been going backward. I said, "Let's weave millions of books and webpages into Wikipedia and the Web. Let's go and take the best information we can and make it one click away from people through Wikipedia."

Okay, that's three years ago. Now, I'd like to report a little bit on how we're doing toward that and, more importantly, how we all can work together going forward toward this and sort of similar goals of getting to digital learners. We had been collecting all of the out links from Wikipedia for about five years before. We basically subscribed, if you will, via robots to go and find every time a new author wrote a link, we would instantly go and archive it and put it in the Wayback Machine. We knew we wanted to secure Wikipedia in the future and with this three years ago we said, "Okay, now, let's go and make robots go through and find broken links and go and

replace them with links into the Wayback Machine." It turns out that a webpage only lasts about 100 days before it's changed or it's deleted; 100 days, so we are building our culture on sand out there. It is constantly shifting, so it may not reflect even what the author originally wanted to say or it might be completely gone, and I'd like to announce today that we've achieved a milestone. We have now replaced and fixed over 10 million broken links in Wikipedia. That's good, and we went in Wikipedia and did an analysis of where do people go from Wikipedia and it turns out 6% of those people that visit a webpage in Wikipedia click on one of the footnotes, so 6% of page views people want to know more, which is quite encouraging to me, and the most sites, the place that people go to most at this point is the Wayback Machine by a factor of three and that's because we've gone and fixed so many broken links. The next is books and then journal literature, so that is our agenda. Let's go and reinforce books and journal literature and this is how people using books in a very different way than going necessarily and buying a book and reading it from one place to another. It would be just going in and reading and then coming back out.

I'm going to do something here that I've constantly been advised against doing, which is I'm going to do a live demo. Okay. So, here is the Martin Luther King page. Is it working? Yeah, score, okay. So, here's the Martin Luther King page, a very popular page in Wikipedia land, and it turns out it has about 100 books at the end, if you scroll way down to the end it's got 100 books that are there. We only had about 30, so we went and bought some more and we used this as sort of a proof of concept and if you go to the first footnote, hover over it, then it's a book. It's all deliberate speed, reflection, and has a page number and if you click on the page number, then it opened right to the right page. Yay! It worked! You can only go one page forward and back and beyond that then you have to go and log in and borrow the book from the library or buy the book or go to your physical library to get it but at least you get enough information to be able to understand, is this the right thing? Is it correct? And what you can do.

So, I was next door at my next-door neighbor's house when my next-door neighbor, Carmen Steele, she's 15 years old, and I was describing what I wanted to do with Wikipedia. You click on it, you open the page to just the right page, and she lit up. She said, "I want that!" Okay, I don't usually get a rise out of my 15-year-old neighbor, Carmen Steele, but she

said, "I want that," and I said, "Okay, why?" And she said, "In my school I'm not allowed to quote from Wikipedia." Right? "I have to go and get—I have to quote from the real book and if I had that I could do it in the middle of the night." So, I said, "Okay, now we're on to something here." That we actually have something that we really should move forward on and make this whole thing happen, so we've started working much more closely with Wikipedia and so we had written this bot to go and fix the webpages and then we wanted to go and weave in the books themselves, and we started working along with the English Wikipedia group and not only did they say yes, please do go and make better hypertext links. Moriel Schottlender said Wikipedia loves Internet Archive in this presentation that they gave and they gave us the best tool of the year award for impact in the wiki mania community, so there's a real hunger out there within the open online world to improving the information that is available and to be able to do that in a respectful way is absolutely critical for everybody to work together on this.

This news went out about two weeks ago. It's been covered in *Wired* and a bunch of the online press and there's been generally "Hurrah!" Right? It's sort of, you know, what's to argue with out there but this is only the first step and soon I'm going to be getting to where everybody here can really help participate because we don't have the right books online yet. We really just don't. So, we started to go and analyze where are we? At the Internet Archive collections, for instance, we've been digitizing books at about a thousand books a day for the last bunch of years, 15 years now, and working with libraries, we have a room inside the Library of Congress, it's been cranking away and we went and did a study of the books by decade that we have on the Internet Archive and if you look at that graph, it's going great up to 1923 and like in all these information graphs it's supposed to continue up exponentially but it doesn't. It craters. It craters all the way through the 20th century and it doesn't even get up to 1923 levels until the end of the 20th century so we're missing a century of books. So, those that are turning to the Internet Archive for information that is deeper than you can get through Google or Wikipedia are not getting access to a large percentage of the materials that were published during the 20th century. Not good enough. It was a pretty impactful century. I don't think we want to bring up a generation not knowing about the 20th century. So, where do people go? If you try to figure out we know, okay, if it's not in the Internet Archive maybe it's available as a physical

book and where many people go to go and find a physical book, you can go to the library but a lot of people turn to Amazon, so there was a study of Amazon of books by decade, one more time so it's decade, decade, decade, goes up to 1923, crash. Craters. Even books available do from Amazon—they're not in print and they're not available from Amazon, so we've got a real problem out there that we don't have access, very easy access, if it's anyplace it's within our libraries and our libraries in general have been quite nervous about how to go and make post-1923 books available respectfully, and here there's been an evolution that has worked very well. It's called controlled digital lending. It's very library-like and frankly a little lame that you can basically go and find and check out a book and if you have a physical book, if it's been digitized then one reader at a time can have access to that digital book and if somebody else wants that book they have to wait until that book is returned and it's protected with the same protection mechanisms that the author, that the publishers use for their in-print stuff so our "dusty mustys," our century of old books we're going and protecting with the same sort of clunky interfaces to go and make these digital versions available. So, this has been up and running since the year 2011 and Boston Public Library has been doing it for cookbooks and genealogies, lots of other libraries, and it's doing fairly well. There's a position paper and some legal white papers now and in terms of how to argue all of this and this sort of basic balance of having some level of access to these materials has been adopted by a very large number of organizations. So, now we have not only a distribution system, Wikipedia, we now have a mechanism of going and making things available respectfully in terms of copyright to go and take the modern materials and make those available sort of on our own terms as libraries in a library-like way.

The next key was to ramp up digitization. So how to do this? Well, we've been digitizing books up a storm but we needed more books and we wanted to be able to go much faster, so we started to work with libraries on not just in-library scanning but where they would go and donate enormous quantities of books to the Internet Archive that we could cost-effectively go and digitize. One of the first on this was Trent University and Trent University was remodeling one of their libraries and they were going from 500,000 books down to 250,000 books because they wanted to free up space for more student activities and meeting rooms and the like and the students weren't using these so they basically took all of the

books that hadn't been checked out in 10 years and they wanted to deaccession them respectfully, so instead of throwing them out or giving them to Better World Books they decided to donate them to the Internet Archive. So, we went and we packed up all the books and they're nicely kept in boxes so they're not very accessible but they are preserved without destroying the books so they are respectfully kept, and we've been digitizing them and there's now 171,000 books that have been digitized and made available on the Internet and something that's kind of interesting to me is these books—Trenton University hadn't circulated them in 10 years or more but of these 171,000, 25,000 of them are checked out at this point so somebody wants these things. We just have to find a new and different way to give a second life to a lot of these books.

Georgetown Law Library, they've been digitizing their modern books and then they've been weaving those books into their card catalogs themselves, so they're using the same discovery mechanisms but they're now just giving their patrons an alternative. So, you can go and see, all right, I want this book, *Liquid Assets*, and I want to be able to see this book. I can either go to the shelves and get the physical book or I can get a digital book but only one or the other is available at any one time, so there's no more copies that are being circulated and this book has now been checked out, I guess, 24 times or at least viewed 24 times, which is more than probably that book has seen in 10 years within the Georgetown Law Library.

We've been working on our digitization technologies to be able to take better and better pictures to really update and upgrade the quality, so it really hopefully only needs to be done once because it is painstaking. There's people actually turning the pages one by one. We tried the robots but the robots didn't work very well. They were kind of expensive and they broke the books, so we went back to going and turning the pages and it's been working out very well, and we're now digitizing over a thousand books a day of modern books that have been donated or we've been working with. We just finished a project with Phillips Academy Andover where they were remodeling their library so they had all of their books off their shelves, and before they went and put them back on the shelves they said, "Let's digitize all of them." We worked with Phillips Academy Andover and actually I think it's completely great. This is one of the top prep schools in the United States, so a really elite collection of books for this and they said, "It's really important to us at Phillips Academy

Andover to make sure that not just those that come to Phillips Academy Andover can leverage the books that are in our library." So, we've digitized all of those books and they are now available on the Internet Archive site to the public so that somebody in Oakland, California, or anywhere can go and borrow these books in the same way that somebody can at Phillips Academy Andover. So, that's really great where a library I think is the first complete library that has come forward and said, "Yes, let's go completely online under our own name." So, thank you very much to Phillips Academy Andover.

But all this was to try to figure out what books we needed, so we started to do this study of the books that are actively referenced on Wikipedia or they're in a lot of course catalogs or course syllabi or they were commonly held in libraries. We came up with a top list at 1.5 million but a more expanded list of 4 million books, and I'd say if we did the right 4 million books we would have a Yale, a Princeton, a Boston Public Library that would be available to all. That will be tremendous or at least one copy of them and then we would be able to work with other libraries so they'd start to lend more copies on their own behalf.

Where are we going to be able to get these books? The libraries in general were fairly nervous about having modern books available though they're getting more bold on this front as I've been showing. People have been donating books and we went back to our friends at Better World Books. Better World Books is a special bookstore. It is one of the three biggest used bookstores. They sell new and used books in the United States. They mostly work with, well, us, we libraries, where they take the deaccession books and they find new life for these books, but they are also a B Corp, so when B Corps switch to these beneficial corporations when they were being started, there were 29 inaugural classes, there was the first 29, they were one of those, they have a mission to help and be not just a profit-seeking organization and they have always been mission aligned with us. We basically went to them and over the years they would be giving us thousands and thousands of books that we've been digitizing but as part of this program we wanted to do millions and we were ready for it. We knew what books we wanted, we had the digitization equipment ready, and we knew how to do the controlled digital lending and all of this, so we went to Better World Books and we said, "Can we work together to get millions of books?" and they said, "Brewster, we'd love to. We would absolutely love to, it's within our mission, it's

how we like to work, but we've got a problem out there. Why don't you go and make us more mission aligned? Why don't you go and buy us?" and I said that was not the conversation I thought I was going to be having. I said, "I don't want to buy you. I want you to be Better World Books. I want you, Better World Books, to be more Better World Books," but they said, "You can help us," and so what we did is we set up a separate nonprofit called Better World Libraries and we funded it out of my foundation based on my Internet winnings, and we basically bought Better World Libraries in such a way that now it's mission aligned with the Internet Archive and the whole library world in a much more closed way, so that's the big announcement today. So, with Better World Books now part of I'd say closer into the whole library ecosystem, we are coming up with mechanisms of going and having books see all of their positive attributes leveraged in the different parts of the lifecycle of books, and toward this I'd like to invite out Dustin Holland, the CEO of Better World Books.

Dustin Holland: Thank you, Brewster. I can tell you we owe you a debt of gratitude. Thank you for believing in us and thank you for investing in our mission. So, Brewster told you a little bit about the story of how it happened, but let me tell you how it really happened. Last fall Brewster came to us. He shared his vision with us. After walking him through the 330,000-square-foot distribution center that we operate in Indiana.

Brewster Kahle: This is completely great! You have to go visit. It is acres and acres of books!

Dustin Holland: You are welcome to come; send us an e-mail or drop by anytime. Anyway, we told Brewster about the 30 million books a year that we process at one of our four distribution centers and he told us about his mission and we wanted to help, so we sat down in our conference room and he thanked us for the hundreds of thousands of books that we had provided over the past decade or so and he told us, he kept throwing out the word "millions." Millions of books, millions of books. So, we said, "Brewster, I've got two options for you. You can buy those books or you can buy the company," so here we are today. Thank you. And I am so pleased to be here. Today is truly a glorious day. For 15 years I've been coming to the Charleston Conference and many of you librarians in the room have come up to me over the years and told me our program is too good to be true. How can a company offer such a great free service where you collect the books, you keep them out of landfills, you

get them to where they are needed most, and you even give back money to literacy in libraries and education causes around the world. I've also heard, and I've heard this from many people in the room, that sometimes you wrap up your books in newspapers and you put them in the dumpster or you meet at the library at midnight and you join your colleagues and you drive the books somewhere far out in the county and you dump them down an abandoned well shaft, so I'm here today to tell you it's—our program is going to get even better. You've heard Brewster's vision for weaving millions of books into Wikipedia and beyond, and Better World Books is pleased to play a role in fulfilling that mission but this is not about us. This is about you making that happen. Collectively, over the past 15 years, you all have sent 330 million books to Better World Books so that we can find a better use for those books and that's why I'm here today to talk about the better option. So you're going to continue to ship books to Better World Books as you always have, nothing changes there, but what we're going to do is we're going to be taking the millions of books that we process at one of our distribution centers and we're going to be checking those books against Internet Archive's wish list, and we'll be keeping those books and the books that we were unable to resell out of the recycling stream, and we're going to be setting them aside for the Internet Archive so they may preserve and digitize those books and make those books universally accessible to all people and all generations to come. So, I hope you agree that today is truly a historic day for readers around the world and for future readers that follow. Together we are preserving knowledge so that knowledge does not become extinct, and you can rest assured that you'll truly be maximizing the value of every book that you send us. That's why today is a glorious day. Thank you so much.

Brewster Kahle: Really, it's been a delight to work with Better World Books over the years and it is really getting better, so please change your buying habits from another website where you might be buying your books to buying from Better World Books.com, please, and continue to donate books to them as that will keep it a vibrant ecosystem. It is not easy to compete against Amazon. Amazon is a juggernaut out there and I think that we in the library world can and should work together more concretely and better to have a full lifecycle of books that reflects our values that can go and support the publishers and authors on the front end as libraries buy, I understand, about 20% of the whole book publishing industry goes to libraries, so that when it's

time to have fewer books on the shelves they can go to Better World Books, those books can be digitized, physically preserved, and they can be injected, bits and pieces of them, into the World Wide Web and read in a very different way than how you might've read a history book by buying it and reading it all the way through. It's more the sort of in and out. I think of it like newspapers are really designed for today's readers like today and yesterday's newspaper is fish wrap, but we know that there's history in the newspapers from other days but you read it in a very different way. I believe that's what we're seeing on the Internet Archive as we've been weaving these books into the Web, they're being used in a fairly different way but there's also driving demand back up into buying the books or getting them out of libraries.

So, we've heard now about millions of books and how this whole thing could work but really the whole thing comes down to people and supporting communities. I work with Wendy Hanamura. She's a Japanese American and her mother and her grandparents were in a Japanese internment camp during World War II and she's gotten—she is fabulous and working with the Internet Archive, you'll hear later on our panel of Dustin and a bunch of us at 1:45, please come, but she will be there but she's distraught about how the Japanese internment experience is being come to be understood out there. There was an article a few weeks ago in the *New York Times* that showed that one of the physical Japanese internment camp facilities was being used for the migrant children that are coming over the borders to intern them there and that this conflict and the sort of idea that these camps are being used again and in this way is sort of historically not a great thing from the perspective of the Japanese American community and beyond that, so there have been some protests about this and the like but they're starting to find that the Japanese internment program is starting to be apparently deliberately misremembered for what it was, that it even happened or that it might have been good for the Japanese. That they wanted to be in these internment camps. This is just not true but the people that were in those Japanese internment camps are starting to pass away and when the last person dies who actually was involved in those, then history is needing to be gotten from our books and other sources.

How are we going to keep this all alive? The Internet Archive has been working with the Japanese American community through Densho, which is a nonprofit in Seattle, and they came up with the 500 key books that represent the Japanese American experience and with funding from the National Park Service, went and digitized these books and made them available on the Internet Archive to be able to be borrowed. If you were to go and search for them you'd be able to see *No-No Boy* and be able to check that out and it looks like, oh, yeah, it's actually already checked out, so you can check it out and be able to see this generation, this set of books for a generation that is if it's not online it is as if it doesn't exist. Right? Really. If it's not online it's as if it doesn't exist and we know that the best we have to offer is not online. A lot of good stuff is online, but the best we have to offer is not online, so we're bringing up a generation without access to the best materials and that is going to give us the generation we deserve and I think we can upgrade this, but going and just putting it on the Internet Archive wasn't enough. We said, "Let's go back to Wikipedia and go and take"— the key article on this was the Japanese internment of Japanese Americans and what I was amazed by out of this was that Wikipedia, to their credit, they go in and chart over the course of a year how many people have read this particular page and on this particular page over 700,000 people went and read this page but for a lot of them they stopped there because they couldn't go further. What we can do together is to go and make it so that deeper information, the controversies, not just the consensus view but the controversies around these are available from these websites so that one click away from that you can go and open a book to the right page, but that is the type of thing that we can deliver to communities of all sorts based on the resources that are in our libraries, that our publishers go and make sure that these stories come and are made available in high quality for the long-term, that authors are spending years of their lives making available and are often not reflected in the digital sphere. We can bridge that gap now by working together. Digital learners need us now. It is our chance. It is our day. We have a mechanism of being able to pull this off. We can pull this off at scale. Together we can complete the lifecycle of books from authors, to publishers, to libraries, to used book vendors, to preservation, digitization, reinsertion of the potentially just clips into the live Web and Wikipedia and what it becomes. Thank you very much. I really appreciate your time.