Speaker 1:   As part of the University of Michigan's bicentennial celebration, the U of M library brought StoryCorps to campus, to capture personal stories of those who make up the university's rich history. In the early 2000s, a young Michigan alum had a bold vision for his up and coming company, and he approached his alma mater for help. Google cofounder, Larry Page, wanted to digitize just about every book ever printed, and he wanted to start with the university's library collections. For our latest installment of StoryCorps, former U of M president, Mary Sue Coleman, sat down with former provost, Paul Courant, to discuss how U of M became the first institution to sign on to the Google Books project, and how that decision still resonates today.

Paul Courant: So let's talk about the project. I remember how I got started in it, but I don't remember the point at which I brought you into it.

Mary Sue Coleman: You know, I was trying to recall that, too. I do remember having a conversation with Larry Page at Google, who is an alum of the computer engineering program at Michigan and obviously a brilliant young man, as one of the founders of Google. He made some comment about digitizing all the literature in the entire world. And I thought, oh, 29 year old, big ambitions. I didn't think it was even possible, and then you and I talked, and I remember I was so skeptical that somebody had to prove to me, with the number of scanning stations and the number of people that would be employed and the rate of turning the pages, to photograph the pages, that we could do a million a year, which is what you had to do.

Paul Courant: And which we actually, once the whole thing got rolling, we hit that target relatively easily.

Mary Sue Coleman: But do you recall that? Do you recall how mind boggling it was, because Michigan had been doing digitization for a long time. We'd done some very interesting projects and we were doing 7,000, 8,000 books a year.

Paul Courant: I remember making the calculation when we went public a couple of years later with the project. The calculation, which was not quite accurate but close enough, that it would take us 1,000 years to digitize our current collection. Actually, it would've only taken about 800, but that assumed that we wouldn't buy any more things, which of course, we would have. Basically, it would've taken forever to digitize the collection, and Larry waltzed in here and said, "I think I can do it in six years."

                    Actually, we pushed on it and we thought we could. We didn't quite in the end do the whole project, but we did six million volumes, and we were certainly working at the rate of million a year.

Mary Sue Coleman: Do you remember that people are working 24 hours a day, turning the pages of books, because to digitize you still had to turn the pages. I mean, nothing happened by magic.

Paul Courant: It was a very unmagical process. I had to sign an NDA in order to be able to see it, but I think I'm not breaking any rules at this point. Basically, it was two cameras and an easel, and people would turn the pages, boom, click, boom, click, boom, click, like that. That was a huge operation.

Google being Google, designed their own library cart. The library cart is, of course, a library cart; it's sort of one of the objects that all libraries have and they're all the same. No no. Google designed a special library cart that fit perfectly into the trucks that they were using to move the books into the scanning station and then move them back. They did it just like graduation. Everybody lines up. You follow the person in front of you out, and then you follow the person in front of you back in.

They'd clear a whole set of shelves of books. They'd move the carts. They'd move the carts in order, the books staying in order down to the scanning station. They'd scan them in order. They would then come back the same, and nobody had to think very much about where the books were supposed to go. It was remarkably efficient process, and they were doing, once we really got started, 30,000 a week.

Mary Sue Coleman: Don't you remember being so excited about potential power of the project and the fact that we would have the power here to open up information to anybody who had an internet connection around the world? Since we'd been engaged, even back in the 1940s, with helping restore libraries that had been destroyed through war in other countries, that for us, it was making it accessible, but it was also preserving a treasure that might be destroyed.

You remember Katrina, when all those government documents at Tulane were underwater? They were gone.

Paul Courant: Yup, and partly, actually because of this project, it was possible to restore other parts, make available to Tulane and other libraries that have had such difficulties, replacement copies of works that they already owned.

I just want to go back a little bit to the story, the very beginning as I understand it. There was a long period of quiet.

Mary Sue Coleman: It was two years.

Paul Courant: Yeah, really two years between when we agreed that this was a good idea and we signed the first cooperative agreement.

Mary Sue Coleman: You know, right away we were concerned about the issue of litigation. I remember having many conversations with the general counsel at the university about this, but also with the Board of Regents, the governing board. We didn't want to put the university in a position that it was going to be somehow liable, and at the end of the day we worked out an agreement, so that we had

indemnification from Google, so we actually didn't get sued by the Association of American Publishers, but Google did. And I think, unfortunately, I think all that litigation and all those lawsuits caused them, a number of years later, to sort of back down from the project.

What we did then was to engage with forming the HathiTrust, and you might want to talk about that a little bit, because I think that's a resource that people still have, that's really a very…we were sued there, too, but we won at every stage, because of the fair use doctrine.

Paul Courant: So the HathiTrust is a cooperative digital library. The partners are more than 100 libraries around the world. The founding partners were the University of Michigan and Indiana University. When this Google digitalization project was going on, the original idea that Google had was, they would keep copies of all the scans. If we wanted to look at them, we'd send them an email. Obviously a little more elaborate than that, but basically we'd search their records. We, from the very beginning said, "No no, that's great. We want our own copy. We're an independent public university. We have different values and interests. We want our own copy. We want to be able to us it to allow people with print disabilities to read. We want to use it for these preservation purposes, so you get a copy, obviously, and that's fine. We get a copy, too."

Then we realized that other libraries were also working with Google, and if we all got together and worked on the same infrastructure, we could have a higher quality, more reliable and secure version. If we each did it alone, they'd be okay, but not nearly as good. So we used this collection, largely scanned by Google, to start this cooperative digital library. And from the very beginning, we had put into the original memorandum of agreement, that we were going to be allowed to use the works in consortial ways, in the academic sector. I think that made all the difference, because we, we the academy, not just Michigan at all, but 100 universities and more have this great digital library that is now moving up towards being almost everything held in this set of libraries, published before the end of the 20th Century.

Mary Sue Coleman: Really, the value here is, all these works that are out of copyright, that are available to anybody with an internet connection.

Paul Courant: And even the ones that are in copyright, if you go and search, you can find out whether a word or phrase or a name is in the book. If you want to know when the word gay changed its meaning, you can do searches on all of the uses in the collection of that word and the words adjacent to them over time. I'm just picking one example there, but you have a computationally usable record of essentially all the published words from this period of.

Mary Sue Coleman: And that's the kind of research, Paul, that without this you cannot do.

Paul Courant: You cannot do it.

Mary Sue Coleman: First of all, you can't get access to the physical copies of the books. You couldn't possibly read them all, so this is just another level of human endeavor and scholarship that I think has been extraordinarily valuable, and really, that's what was transformative about it. I'm disappointed that the litigation caused Google to back off. I'm very happy with the HathiTrust, because I think that's been extraordinarily important. And I'm proud, I am so proud that we did it.

Paul Courant: Yeah, it really is…and we involved many other universities. But I have to say, in those early stages, everybody but us was cautious. We were not cautious. In fact, our lawyer put it very well. You don't expect lawyers to necessarily be so enthusiastic. He said, "We were intrepid." And I think that's a very good description.

Mary Sue Coleman: Well, but you know what? That's what the University of Michigan is, and so many times it has stepped up to the plate and done things that others are hesitant to do. You know, I had written about this story, that Michigan has believed in the power of books forever. The first gift to the of Michigan was a German encyclopedia, that was brought to the university by a fur trader from Wisconsin, shortly after the founding of the university in 1817, and we still have it. You can go read it in the special collections. That says something about Michigan.

Speaker 1: That was former U of M president, Mary Sue Coleman, speaking with former provost, Paul Courant. These interviews are recorded by StoryCorps, www.storycorps.org. For more from this series, visit arts.umich.edu/storycorps.