

SOURCES:

<https://blog.wikimedia.org/2015/06/19/meet-print-wikipedia/>

<http://bedfordandbowery.com/2015/06/an-artist-explains-why-hell-turn-wikipedia-into-a-7473-volume-print-edition/>

<http://www.bbc.co.uk/programmes/p02v2vrn>

Why are you doing this?

I'm really interested by the ways in which we define and are defined by information. The Wikipedia database is big data that we can just barely understand. It is big enough that that we have to use specialized software to handle the whole corpus, but small enough that we can imagine its size. We have no real understanding of how big the NSA's database is: we can talk about it in terms of square footage, or petabytes, but that doesn't mean much to most people. We understand a book as a marker of knowledge. We know how long it takes to read a novel? This offers the opportunity to understand where that threshold is.

How many books are there?

Print Wikipedia is 7,473 volumes, which spans over 5 million pages. In addition I have created a 91 volume Wikipedia Table of Contents and a 36 volume Wikipedia Contributor Appendix which lists the usernames of the 7.5 million contributors to English Wikipedia. Print Wikipedia contains 11,594,743 entries, which range from long featured articles to one line redirects to see other articles.

How long did it take to upload?

The upload process lasted 24 days, 3 hours, and 18 minutes and was on continuous view, with the gallery open 24 hours to the public in during the first weekend. Each time a volume was completely uploaded, the script posted to Twitter @PrintWikipedia.

Is everything printed already?

It is important to note that I have not printed out all of the books for this exhibition, nor do I personally have any intention of doing so. If someone paid the \$500,000 to fabricate a full set, then they could do so. There are 106 volumes in the exhibition,

which are really helpful for visualizing the scope of the work. It isn't necessary to print them all out: our imaginations can complete what's missing.

What is in the exhibition?

The two-week exhibition at Denny Gallery is structured around the upload process of *Print Wikipedia* to [Lulu.com](https://lulu.com) and the display of a selection of volumes from the project. The upload process will take between eleven and fourteen days, starting at 1 and ending at 5. During this time, the upload process will be open for all to see around the clock—at least during the first weekend, as the gallery will remain open through the night in recognition that the computer itself works continuously. There will be two channels for watching this process: a projection of Lulu.com in a web browser that is automated by the software, and a computer monitor with the command line updates showing the dialogue between the code and the site. If you aren't able to visit the gallery in person, you can follow the process on Twitter; we will post to the [@PrintWikipedia](https://twitter.com/PrintWikipedia) Twitter account after it finishes each volume.

How does it work?

On April 7th I downloaded a database backup from dumps.wikimedia.org, which began a month long process of preparing the data for upload. The download takes about 4 days, and it takes nearly a week to unpack the file into a MySQL database. I run some SQL queries against it to clean it up and make it run faster, then I run the Java code which I have written which turns all of those entries into PDFs. This takes a good two+ weeks. Once that is complete, the files are ready for upload.

[Lulu.com](https://lulu.com) doesn't have an API; they developed one, then deprecated it several years ago. Because there is no API, I upload them to Lulu.com with a piece of software written in Python, that uses Selenium to automate the browser; Selenium is a package that is primarily used for QA testing on websites, by mimicking a human clicking on the webpage. The software sets the titles, books sizes, choose files for upload, etc. Once each volume is uploaded, it tweets out an announcement at <https://twitter.com/PrintWikipedia>

Using browser automation software to simulate an API is a highly unusual approach, and one that certainly isn't *Enterprise grade Software*. When I told the team at Lulu.com that we had programmed the script to do this, they did the conference call equivalent of a spit-take.

Will you keep it up to date?

I will not be outputting a new version every month, or every year. I don't think it is necessary. The work is always already out of date, as it takes a month at minimum to prepare the PDFs for upload, and another two+ weeks to upload them.

I do, however, want to replicate this in other languages. I would like to upload a Chinese version from Beijing, or a German version from Berlin.

Authorship and Copyright

Wikipedia is licensed under a Creative Commons Attribution-ShareAlike license, which means that anyone can reuse the content so long as they give attribution, and preserve the same open license. So I am free to use the content, including selling the content, so long as I give attribution, and don't attempt to copyright the material.

But giving attribution to all the authors is no small task. I asked the lawyers at Creative Commons, but they demurred. The Wikimedia Foundation suggested I look at some general community guidance that says you should cite the 5 most prominent contributors to each article, but that was only guidance for when citing or using one article. Given that I was using *all the articles* I decided to create a 36 volume Wikipedia Contributor Appendix that lists *all of the contributors*. It turns out there are 7.5 million contributors (a number I established).

The Wikimedia Foundation does enforce their trademark on the logos and wordmarks, but I was able to use one of the clauses that allows for trademark use in artworks. The legal team also asked me to include a disclaimer on the copyright page, and on the back cover of the books, stating that this work wasn't endorsed by the Foundation.

Did you do this by yourself?

This was not a solitary endeavor. I was grateful to work with several programmers and designers, including Denis Lunev, Jonathan Kiritharan, Kenny Lozowski, Patrick

Davison, and Colin Elliot. I was also supported by a great group of people at Lulu.com who went above and beyond to support this wild and quite unwieldy project.

What was Lulu.com's role?

I worked closely with Lulu.com during the upload process; they were really supportive. I approached them in Fall 2014, realizing that I probably needed to give them a heads up that I was going to be uploading over 20 GB of files to their servers, and to make sure they understood that it was signal, not noise! I was impressed with the degree to which they understood the work both as a work of art and as books.

What have you learned in this process?

"Print is Dead" but people still really care about books. So many people came into the gallery and wanted to talk about the books. They wanted to talk about their memories of childhood encyclopedias, and the materiality of knowledge.

Books are a really useful unit of measurement. We don't have a great sense of how much knowledge fits inside a gigabyte of data, but we do have an embodied understanding of how long it takes to read a book... or 7473 of them!

Wikipedia is full of lists. So many lists. Almost 600 volumes of them. Knowledge coalesces around other structures too. I included 28 volumes that start with "BAT." These are not about flying bats or baseball bats, they are a compendium of *battles*. From Battle of Aachen to Battle of Żyrzyn. Likewise, the 26 volumes that start with "New" represent a structural history of colonialism and the industrial revolution.

Why are the tables so funky?

As a programmatically designed book, the layout was built off of intentional design decisions, but entirely contingent on a range of foreseeable and unforeseeable scenarios. The way that the tables are presented is one of those highly contingent situations. Initially I approached it as the the best solution for an ungainly set of information that was never meant for a three column book layout, but I've come to enjoy the absurdity of their factual accuracy, borderline usability, and abstract aesthetic beauty. In that regard, those three attributes are a pretty good distillation of the project write large.

What were some of the problems you experienced during upload?

In this work I've done a whole lot of things you are just "not supposed to do," starting with turning all of Wikipedia into books, and moving outward from there! I knew we were bound to get errors and hiccups, so we wrote the code to handle the errors. While our repeated stress

tests helped us streamline the process, the entire upload process was going to take several weeks, so it was impossible to truly simulate the whole upload without doing the whole upload.

We worked closely with Lulu on resolving a number of technical challenges prior to and during the upload performance. They helped us work through issues with storage and account size: they increased our file storage permissions and even then had to help us create special linked accounts for each ~1200 volumes, as that was the cap per account. They helped us debug a series of PDF errors, as well as help us speed up our upload script by ignoring certain http requests that were taking forever.

The biggest issue was that the upload ended up taking longer than expected. This is in part due to network issues, as our stress tests in the gallery were done at night, when network traffic is much lower and speeds are much higher. It was also due to the instability of the upload script which tended to crash every two to four hours, for a variety of reasons, all of which were due to the general principle that the code was doing "something you are not supposed to do." We wrote a script that listened to the twitter feed, and alerted us when the code had crashed because it hadn't tweeted out.

Lulu.com also had scheduled and unscheduled downtime, which ate up some of our time for fast late night uploads. Lulu was super awesome with letting us know well in advance about the scheduled downtime for system upgrades. They had to roll out new versions of their site's software, which in some cases, required us to rewrite our upload script to work with the new interface.

The bottom line is that it was a piece of endurance performance!

Tidbits + Notes

Dissonance between the print version and the digital version
Henry fox Talbot - understand the new through the old medium
Also McLuhan
Lynne Tillman effect - poetry
Inter generational gesture to reach out to the young kids who text
David Bun
Yesterday's with is today's press release
First song title and last song title. A-Z