HathiTrust U.S. Federal Documents Program Update

Heather Christenson

I am pleased to have this opportunity to update GODORT and *DttP* readers on the progress of the HathiTrust U.S. Federal Documents Program.[1]

As of this writing in December 2019, HathiTrust includes close to 1.4 million U.S. federal documents digitized from print. Our top contributors are the University of Michigan, the University of Illinois at Urbana Champaign, and the University of Minnesota. The collaborative nature of our aggregate contributions is powerful—our collection includes digital volumes from 51 different institutions and from the Technical Report Archive & Image Library (TRAIL).

One real impact of our program has been the usage we see. For example, digital availability has resulted in 154,216 page views in 2019 for the digitized version of *History of Wages in the United States from Colonial Times to 1928*, our most popular U.S. federal document.[2] In addition to scholarly use, other users include investigative journalists such as Neela Banerjee of Inside Climate News, who relied on some of the federal documents in HathiTrust for her groundbreaking research into the record of Exxon's knowledge of climate change.[3]

As projects such as the Preservation of Electronic Government Information (PEGI) are highlighting the huge challenges of digital preservation especially in these somewhat volatile times,[4] HathiTrust's TRAC-certified preservation commitment is an assurance to the community.[5] Through this commitment almost 1.4 million digitized U.S. federal documents are preserved by an entity outside of the government or the commercial realm.

In addition to these programmatic values, HathiTrust has been working on a number of operational changes to improve access and discovery, and to set the stage for more effective collection development and management. For example, we made an adjustment to our metadata management system that resulted in 15,000 U.S. federal documents being opened to full view that were previously in limited view due to metadata issues.[6] The change also ensured that new U.S. federal documents coming into HathiTrust will be identified more effectively and thus available in full view to users. We've also been working on projects that further our understanding of needs and practices for stewardship of our very large scale HathiTrust collection, such as investigations into collection analysis use cases, and predictions of collection growth. Looking ahead, better collection analysis capabilities will enable HathiTrust to more effectively "fill in gaps" in our digital collection and make sure that corresponding print copies are retained by the HathiTrust libraries in our Shared Print Program.[7] On a practical level, predictions of collection growth enable us to calibrate infrastructure investment, such as purchases of digital storage.

At the end of 2019 new members were appointed to a refreshed HathiTrust Federal Documents Advisory Committee which will bring additional member library perspectives to the Program.[8] The Advisory Committee will focus on collective strategies to solve shared challenges such as collaborative digitization of federal documents; methods and operations to characterize, improve, and enhance the quality of federal documents metadata; and strategies to expand awareness and promote the use of the HathiTrust U.S. federal documents digital collection. In 2020 we expect to launch Program activities that will reflect these priorities.

Heather Christenson (christeh@hathitrust.org), Program Officer for Collections, HathiTrust

References

1. "HathiTrust U.S. Federal Government Documents Program," HathiTrust Digital Library, accessed March 12, 2020, https://www.hathitrust.org/usgovdocs.
2. *History of Wages in the United States from Colonial Times to 1928* (Washington, DC: GPO, 1934), https://hdl.handle.net/2027/uc1.32106007458745.
3. Neela Banerjee, Lisa Song, and David Hasemyer, "Exxon: The Road Not Taken," Inside Climate News, September 16, 2015, https://insideclimatenews.org/content/Exxon-The-Road-Not-Taken.
4. PEGI Project homepage, accessed March 12, 2020, https://www.pegiproject.org/.
5. "TRAC Metrics," Center for Research Libraries, accessed March 12, 2020, https://www.crl.edu

/archiving-preservation/digital-archives/metrics
-assessing-and-certifying/trac.

6. Heather Christenson, "Opening Up 15,000+ Federal
Documents: An Algorithm Story," HathiTrust, April 10,
2019, https://www.hathitrust.org/blogs/perspectives
-from-hathitrust/opening-15000-federal
-documents-algorithm-story.

7. "Shared Print Program," HathiTrust, accessed March
12, 2020, https://www.hathitrust.org/shared
_print_program.

8. "Federal Documents Advisory Committee," HathiTrust,
accessed March 12, 2020, https://www.hathitrust.org
/federal-documents-advisory-committee-fdac-0.