

Archive-It

State of the WARC Report



Web archive management and preservation in 2019-20

The *State of the WARC Report* captures [Archive-It](#) partners' practices and plans for preserving the Web ARChive (WARC) files that constitute their web archive collections. In total, 62 partners responded to a survey between November and December 2019, representing myriad collecting missions, scopes, and institution types and sizes. Now the third installment in its [series](#), the 2019 report also demonstrates the trends among these practices since they were first studied in 2015.

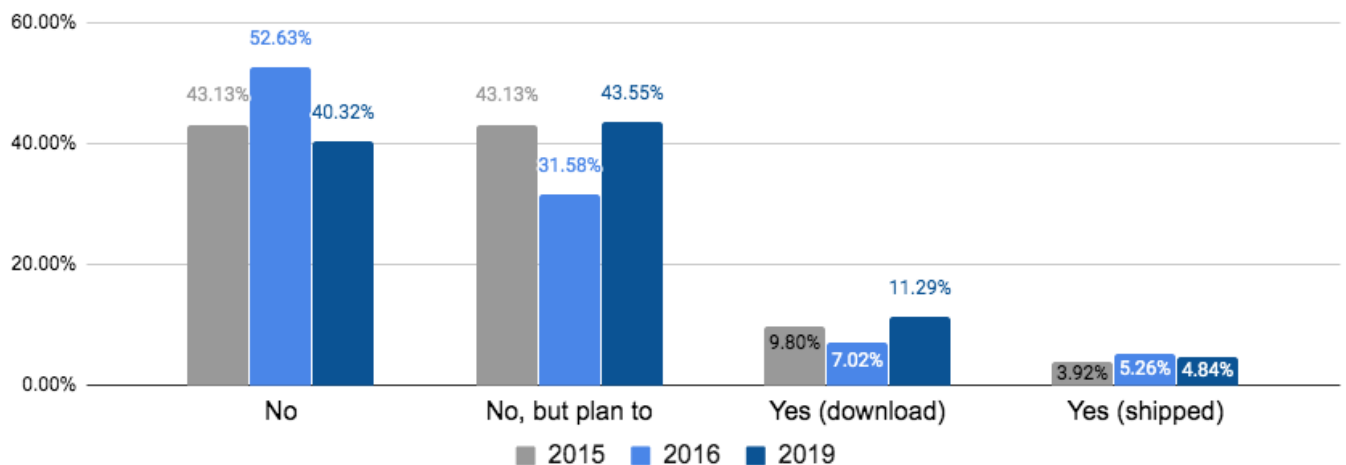
The [Internet Archive's](#) Web Archiving & Data Services Group uses these results to improve and enhance existing services for Archive-It and contract web archiving partners, peers, and prospective new web archivists. The web archiving community may also find them useful as a benchmark and complement to [reports](#) published by the National Digital Stewardship Alliance about the state of web archiving in the United States generally.

Among the survey's results we find:

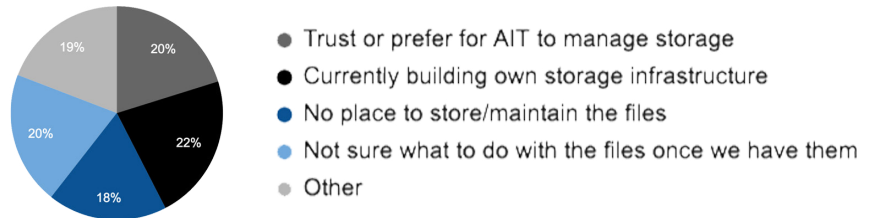


- a stable minority of Archive-It partners preserve their WARC files outside of or in addition to Internet Archive storage;
- others divide almost evenly among rationales for relying instead on the Internet Archive's storage and preservation services;
- partners' expectations for preservation services are consolidating around the themes of distributed redundancy, fixity, and interoperability.

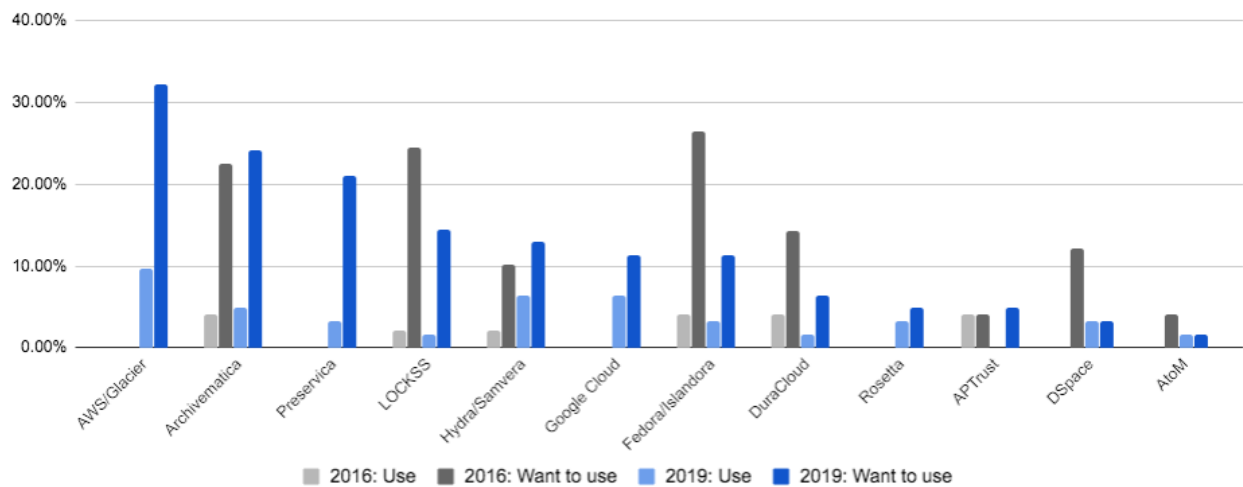
The proportion of Archive-It partners who report that they store and preserve WARC files locally (acquiring them either by download or via drive shipment) remains small and increased incrementally.



Those partners who say they “plan to” do this in the future outnumber those who do not for the first time. The most cited reason for not preserving WARC files locally was that respondents were still in the process of building necessary infrastructure, however other rationales were chosen at similar rates.



Partners were asked in 2016 and 2019 to identify the services outside of Archive-It’s that they use or hope to use for WARC file processing, storage, and preservation. Results indicate a few key preferences, such as Amazon Web Services for storage, Archivematica for processing and preservation actions, or Preservica for some or both of these functions. The Hydra/Samvera platforms also show incremental increases in both usage and desired usage.



The specific features that respondents suggested were most important to existing or imagined storage and preservation services were, in order of popularity in their comments:

- more automated and/or geographically dispersed file redundancy (13)
- auditing tools for fixity and format risk (13)
- integrations with existing access and management technical stacks (11)
- competitive pricing (8)
- user friendliness (4)
- data export options (3)
- access control capabilities (1)
- fast uptime (1)
- technical support (1)
- [Trusted Digital Repository](#) compliance (1)

In the words of one Archive-It partner who summarized priorities for further storage and preservation services succinctly: ***Greater redundancy and organizational/technological heterogeneity; operational transparency, including regular and ad hoc reporting; and affordable price point.***

We thank all participants for taking the time to respond to the survey. Community input will help Archive-It and the Internet Archive to develop or expand services and products in ways that we hope can serve different storage and preservation needs. [Please sign up here](#) for updates and opportunities to provide more feedback.



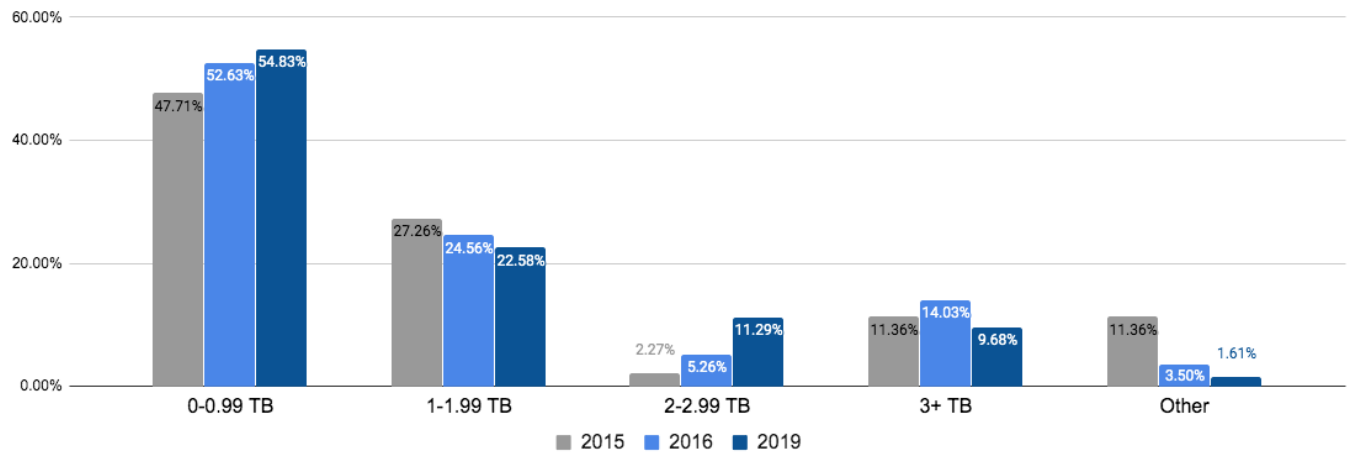
State of the WARC Report

Appendix

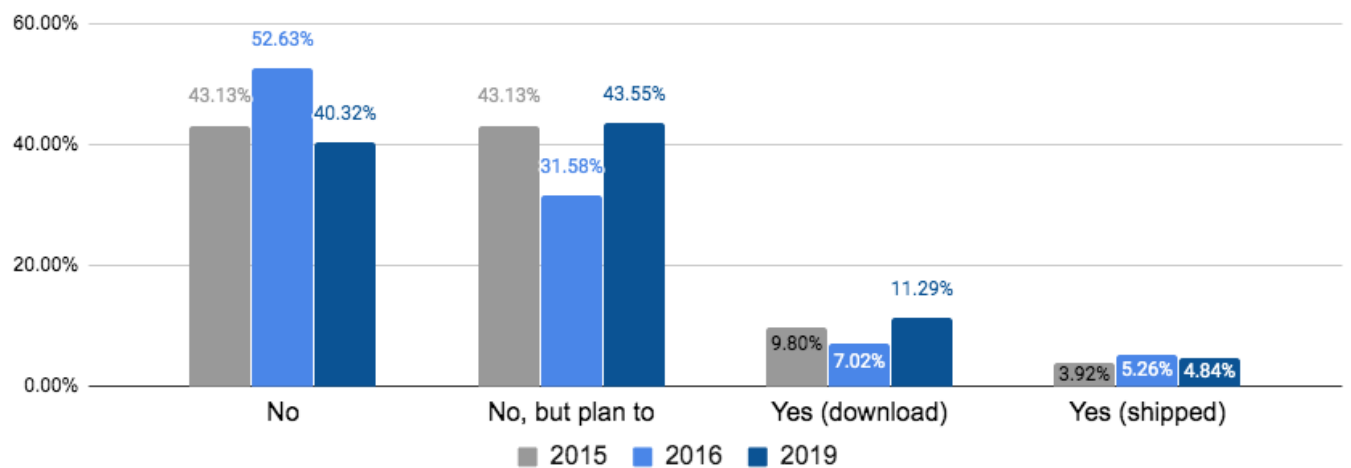
Survey questions and summary responses



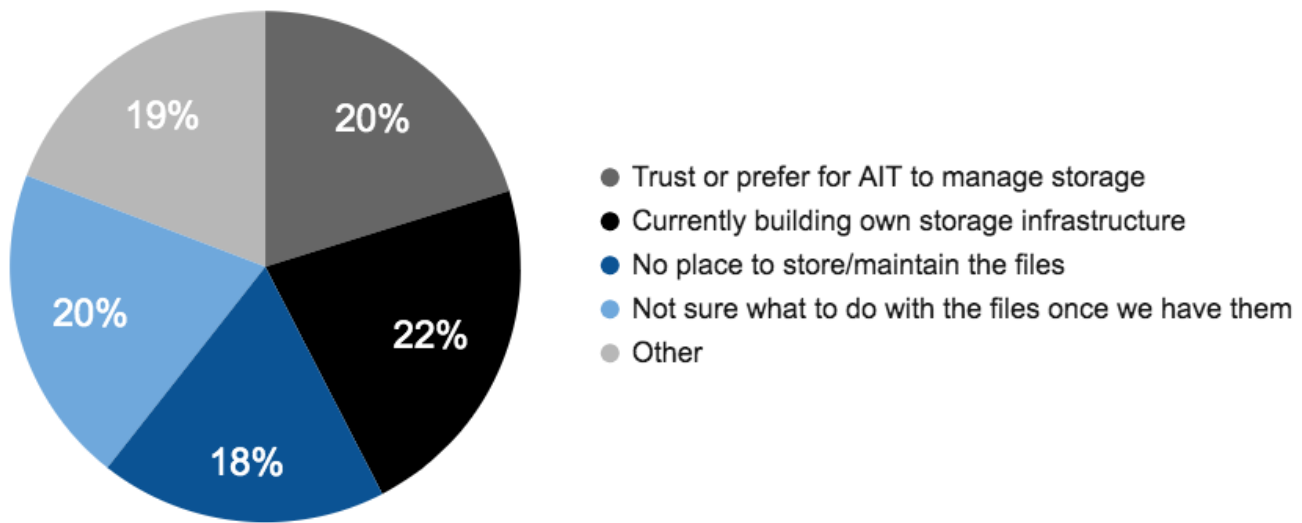
1. What is your current Archive-It data budget?



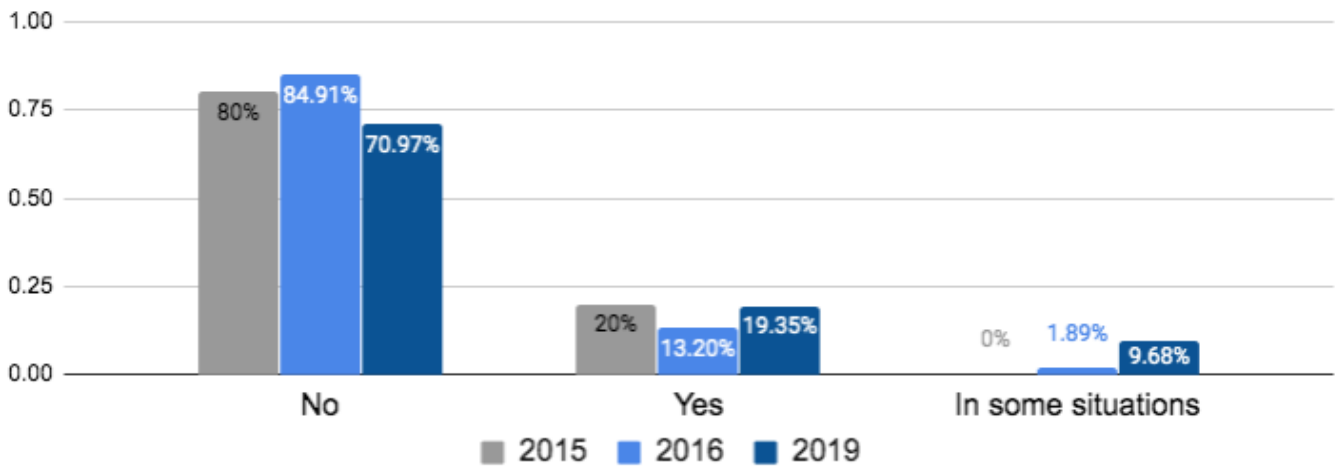
2. Do you download WARC files or have them shipped to you from Archive-It?



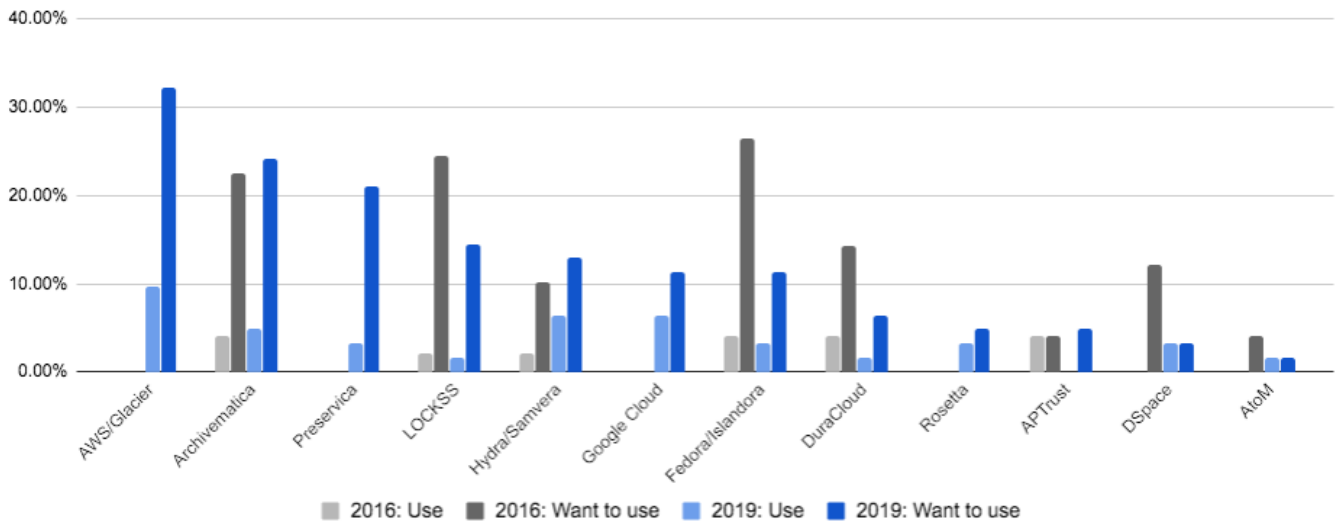
3. If no, why not?



4. Do you ingest your Archive-It WARC files into a digital preservation system and/or long-term storage repository?



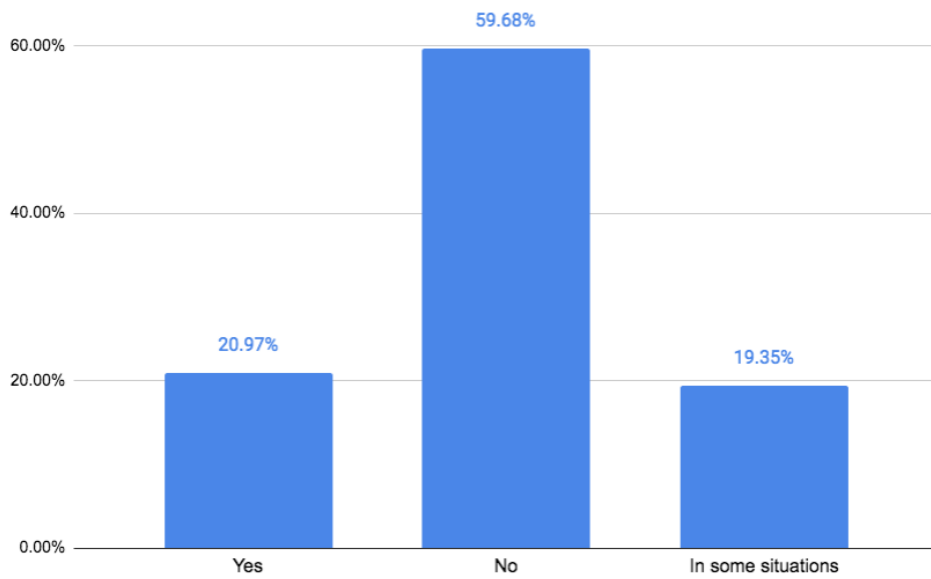
5. What external services do you use currently or want to use to preserve and/or store WARC files?



Service	2016: Use	2016: Want to use	2019: Use	2019: Want to use
AWS/Glacier	-	-	9.68%	32.26%
Archivematica	4.08%	22.45%	4.84%	24.19%
Preservica	-	-	3.23%	20.97%
LOCKSS	2.04%	24.49%	1.61%	14.52%
Hydra/Samvera	2.04%	10.20%	6.45%	12.90%
Google Cloud	-	-	6.45%	11.29%
Fedora/Islandora	4.08%	26.53%	3.22%	11.29%
DuraCloud	4.08%	14.29%	1.61%	6.45%
Rosetta	-	-	3.23%	4.84%
APTrust	4.08%	4.08%	0	4.84%
DSpace	0	12.14%	3.23%	3.23%
AtOM	0	4.08%	1.61%	1.61%



6. Do you create metadata for Archive-It WARC files?



7. Would a cost-effective, more full-featured digital preservation service be of interest to your organization?

