Archive-It and Web & Data Services
News and Notes!

Jefferson Bailey, Director

January 20, 2021
PARTNER STATS: 2020

Partner Stats-a-thon:

- 90 new partners
  - AIT Subscriptions: 55
  - AIT Sponsored: 14
  - Web & Data Services: 21
    - Congressional Archiving: 13
- Profile:
  - 36 EDU, 19 GOV, 21 NGO/NFP, 12 LAM, 2 COM
  - +34% YOY service revenue
USAGE STATS: 2020

Stats-a-thon:

AIT
- 491 TB (+14)
- 8.85B URLs (-16%)
- 336K seeds (+105%)
- 198K crawls (+80%); 66K test (+92%)
- 2261 collections (+43%)
- 1947 new tickets (+44%); 1968 resolved

WDS
- 400+ TB non WDS data
- 3 new National Library partners
- 2+ Petabytes processed for datasets
COVID-19 COLLECTING

Stats:
- 266 collections
- >72,000 seeds
- >65 TB data stored
- 185 institutions took part in “Covid Special Campaign”
- Varying from small data additions to ccTLD domain crawls

Support:
- Subsidies still available for new data and accounts
- Federated search, datasets, and more... in the works!
- Questions on partner opt-in, metadata, etc.

Collections cited in Nature, National Post, American Libraries, and more
OTHER UPDATES: Q4 2020

Training & Tech:

- Recording available: [Best practices for social media archiving](#)
- Blog series on web archiving tech kicks off with [browser-based web capture](#) and [automated preservation workflow](#)

**Archive-It Partner Meeting:** Archived

**Congressional Web Archiving:** New end-of-term service and partners

**New Teammates:** Welcome Alex, Tabish, and Pravin!
Community Webs updates
  ● Actively recruiting 150+ public libraries to join the cohort
  ● Deadline approaching: January 31
  ● Informational webinar at 12PM Pacific today!

Ensuring the Persistent Access of Long Tail Open Access Journal Literature updates

Preserving Open Science Data updates
STRATEGIC PROGRAMS UPDATE: Q4 2020

Scaling Research Use of Web Archives

- New partnership with the Archives Unleashed project
- First dataset release: Early Web Collections
- Prototype up and running
- User testers wanted!
ARCHIVE-IT ENGINEERING UPDATES: Q4 2020

Released:
- Improved data tracking for resumed web crawls
- 14- and 30-day crawl durations available upon request

**Engineering roadmap** updates:
- Apply the great coronavirus asterisk
- Ongoing fixes for capture/replay of social media
- More automation of account management
- Crawl controls
- More add-on products and services
- Priority: moving seeds!
- Priority: new Wayback!
PRODUCT DEVELOPMENT UPDATES

DIGITAL PRESERVATION SERVICE DEVELOPMENTS

- Additional Features
- Content Agnostic
- Intro webinar video
- Presentation slides

RESEARCH DATA SERVICE DEVELOPMENTS

- Data & text mining
- Custom analysis & datasets
- IA-wide collections
PRODUCT DEVELOPMENT UPDATES

ACCESS & OTHER PRODUCT DEVELOPMENTS

- Redirection Service
- Historic Citation Service
- Custom Web Dev

SCHOLARLY ARCHIVING SERVICE DEVELOPMENTS

- New Partnerships
- Expanded Collecting, Access, and Funding
- https://scholar-qa.archive.org/
2021 PLANNING

STRATEGIC AREAS

● **Product Development**
  ○ Bring more mission-critical products to our partner communities

● **Service Delivery**
  ○ Have existing partners find our products more valuable and effective

● **User Expansion**
  ○ Have more non-academic and global communities know and use our tools and services

● **Universal Access**
  ○ Ensure the efforts and collections of our users are widely promoted and available
2021 PLANNING

- **Product Development**
  - Digital Preservation Services
  - Data & Additional Services
  - Feature Expansions

- **Service Delivery**
  - Python Wayback & QA tooling
  - Moving Seeds
  - Account & Managed Services

- **Community Expansion**
  - Community Programs
  - Web Services
  - International Partnerships

- **Universal Access**
  - Access Websites (COVID, etc)
  - Service Integrations
Q&A

Your questions and our answers!

How can we help?

More updates and news:
- Help Center and Community Forum
- Archive-It Blog
- Twitter feed

Email us anytime:
- jefferson@archive.org
- ait@archive.org / webservices@archive.org