Archive-It
13th Partner Meeting
2006-2019
[teen years achieved]
Landscape Thoughts

Wide
● Sustainability Questions
● Economic Forecasts
● Cloud, Hosted, E2E

Focused:
● “Secondary” Outputs
● Degrees of Distribution
● Capture/Replay/Affiliated
Staff Updates & Shout Outs

NEW AIT TEAM MEMBERS

- Pete Mitchell (pete@archive.org), Technical Support Specialist
- Madeline Carruthers (madeline@archive.org), Partner Coordinator

OTHER NOTABLES

- Here ye: James, Jillian, Karl, Kyrie, Lori, Sylvie
- Wish you were here: Adam, Barbara, Mouse, Neil, Noah, Connor, Helge
- Gone but not forgotten: Maria, Mary

#aitTX | @archiveitorg
# Community Updates New Partners (47)

<table>
<thead>
<tr>
<th>Organization</th>
<th>Organization</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cape Ann Museum</td>
<td>University of Colorado - Boulder</td>
<td>Phillips Academy - Andover</td>
</tr>
<tr>
<td>WGBH</td>
<td>Congresswoman Niki Tsongas</td>
<td>East Carolina University</td>
</tr>
<tr>
<td>Museum of Contemporary Art</td>
<td>Pepperdine University</td>
<td>University of Adelaide</td>
</tr>
<tr>
<td>Scottish Poetry Library</td>
<td>Indiana University Purdue University</td>
<td>Heinrich-Heine-Universität Düsseldorf</td>
</tr>
<tr>
<td>Corning Museum of Glass</td>
<td>Indianapolis (IUPUI)</td>
<td>Hyetert</td>
</tr>
<tr>
<td>University of Pennsylvania</td>
<td>Iowa University SILS</td>
<td>Yale University Libraries IT</td>
</tr>
<tr>
<td>Greenpeace International</td>
<td>California State Library</td>
<td>Center for Cooperative Media</td>
</tr>
<tr>
<td>Joshua Meyer-Gutbrod (Digital Campaigns Project)</td>
<td>MacEwan University</td>
<td>California Digital Library</td>
</tr>
<tr>
<td>Institut français du Proche-Orient</td>
<td>Arizona State University</td>
<td>Abilene Christian University</td>
</tr>
<tr>
<td>North American Coordinating Council on Japanese Library Resources</td>
<td>New York City Department of Records</td>
<td>University of Northern Colorado</td>
</tr>
<tr>
<td>Senator John Heinz History Center</td>
<td>Marine Biological Laboratory</td>
<td>North Park University</td>
</tr>
<tr>
<td>99 Antennas</td>
<td>Worcester Polytechnic Institute</td>
<td>Art Gallery of Ontario</td>
</tr>
<tr>
<td>Fashion Institution of Technology</td>
<td>Expatriate Archive Centre</td>
<td>Wichita State University</td>
</tr>
<tr>
<td>Norwich University</td>
<td>Los Angeles County Metropolitan Transportation Authority</td>
<td>Algoma University</td>
</tr>
<tr>
<td>News Measures Research Project</td>
<td>University of Georgia</td>
<td>Beckman Research Institute of City of Hope</td>
</tr>
<tr>
<td>(Branchhead Consulting)</td>
<td>Publications Office of the European Union</td>
<td>LIBER Europe</td>
</tr>
</tbody>
</table>

**2006-today: 664 Archive-It partners | 29 in 2019 so far**

#aitTX | @archiveitorg
AIT Community Updates

Partners Per Year

Year


Partners

27 8 36 31 48 35 38 53 54 88 66 89 62 65

#aitTX | @archiveitorg
AIT Community Updates

Total Partners by Year

Year


Partners

0 100 200 300 400 500 600 700 800

27 35 71 102 150 185 223 276 330 419 485 574 635 695

#aitTX | @archiveitorg
AIT Community Updates

Archive-It Total Size of All WARCs

Data in TBs


#aitTX | @archiveitorg
AIT stats-a-thon

- Total URLs: 40 Billion URLs preserved
  - +33% from 30B 1Y ago
- Total Data: 1660 TB
  - +34% from 1235 TB 1Y ago
- Total Collections: 11,800 public+private
  - +16% from 10,100 1Y ago
- Total Seeds: 1.76 Million
  - +15% from 1.53M 1Y ago
- Total Crawls: 488,000
  - +38% from 388,000 1Y ago

#aitTX | @archiveitorg
AIT Community Updates

Consortial & Other Partnerships

- Ongoing consortia: Consortia Canada, LIPA, CDL, KAIC, Yale, Stanford, BSZ, Harvard, Five Colleges
- Transitioned IMF partners - additional international and managed partnerships
- Discussions with ArchivesSpace, DuraSpace, and more about service integrations/partnerships

#aitTX | @archiveitorg
Grants & Strategic Activities

“Continuing Education for Web Archiving”
- Workshops and training for technical tools

“Global Events & Trends Archive” (GETAR)
- Events archiving, data mining, K-12 education

“Building Infrastructure and Integrations for Open Data Preservation and Access”
- Data sharing, access and preservation

“Broader Web-Scale Provision of Parallel Corpora for European Languages”
- Machine language data mining
Grants & Strategic Activities

Since 2017 the Cohort of 27 Public Libraries:

- Created over 200 community focused collections
- Collected over 27TB of data
- Spoke at over 15 different conferences

At the end of the grant we have:

- A strong group of leaders in the field
- A free online course on web archiving geared toward public libraries, developed with the input of the cohort

https://communitywebs.archive-it.org/

#aitTX | @archiveitorg
Grants & Strategic Activities

- Symposium and half-day workshops among 50 librarians, archivists, and technologists.

- A follow-up meeting for roadmapping with stakeholders.

- Roadmap of a collaborative collecting model for art materials on the web

https://archive-it.org/blog/learn-more/art-libraries/

#aitTX | @archiveitorg
AIT Special & Affiliate Projects

- Edu: 10 MLIS program partners
- Training & Support:
  - Special Topics Webinar Trainings
- Partner meet(ings)(ups):
  - New models for partner engagement
- Professional Participation:
  - IIPC Training WG & SC Member
  - DPC Web Archiving & Preservation Working Group
  - 30+ conference presentations/workshops/etc
AIT Special & Affiliate Projects

• Pitch a Collection Contest Winners:
  – Five Collections from 3 partner orgs
  – Topics include: Beer and Brewing, Mission History, and San Diego Refugees
  – Collected 92 seeds and 120GB data since Nov 2018

• Examples of IA-Sponsored event collections
  – Puerto Rican Protests 2019, UPR and UConn
  – Malibu Community Woolsey Fire, Pepperdine University
  – 2018 Mexican Government Web Archive - El Colegio de Mexico

#aitTX | @archiveitorg
Archive-It New Features

- Brozzler for all!
- Bulk seed scoping rules
- Per-seed WARC for all crawls
- Custom user agents
- Redirection Service
- Token-based authentication for data download
Archive-It backend changes

Completed:
- Operational Resiliency
- Postgres
- Phase 1 - Renewals Streamlining

In Progress/upcoming:
- OutbackCDX - Faster wayback!
- Crawlboss - additional crawl customization/scalability
- Rules Engine - more flexibility in access restrictions/rewrites
AIT Current & Planned Development

- Partner API - Advanced training Aug 21!
- Crawl management/video replay redesign
- Python Wayback
- Partner Wayback Controls
- Public Site redesign
- Renewals in web application
- Data budget enforcement in UI
Final Thoughts

- AIT Application Extensibility
- Beyond Libraries/Archives
- Beyond Crawling
- Interoperability/Integration
- Promotion/Use