



LOCKSS

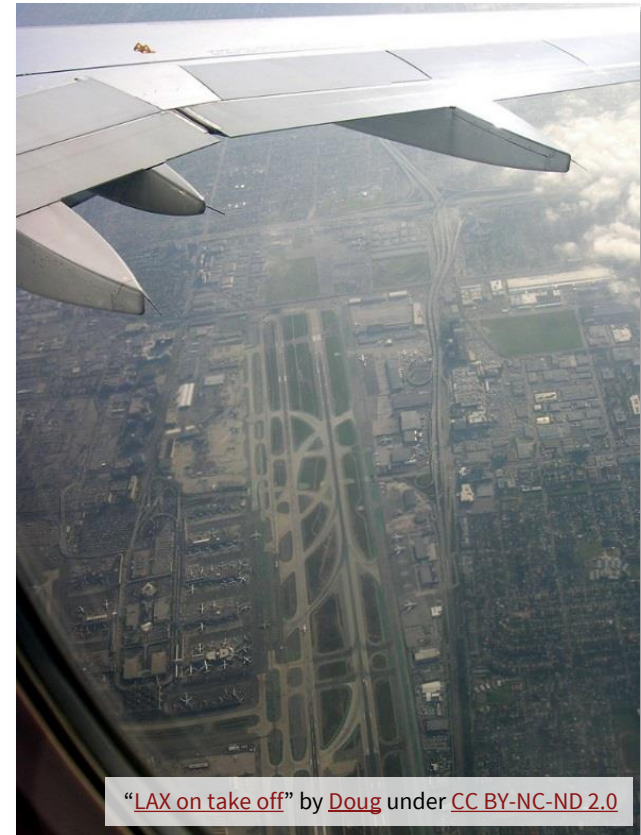
Fall 2018 Web Archiving Updates

Nicholas Taylor ([@nullhandle](#))
Program Manager, [LOCKSS](#) and [Web Archiving](#)
[Stanford Libraries](#)

LOCKSS Community Call
27 November 2018

overview

- IIPC Web Archiving Conference
- LOCKSS + web archiving
 - LAAWS
 - WASAPI
 - Ivy Plus network

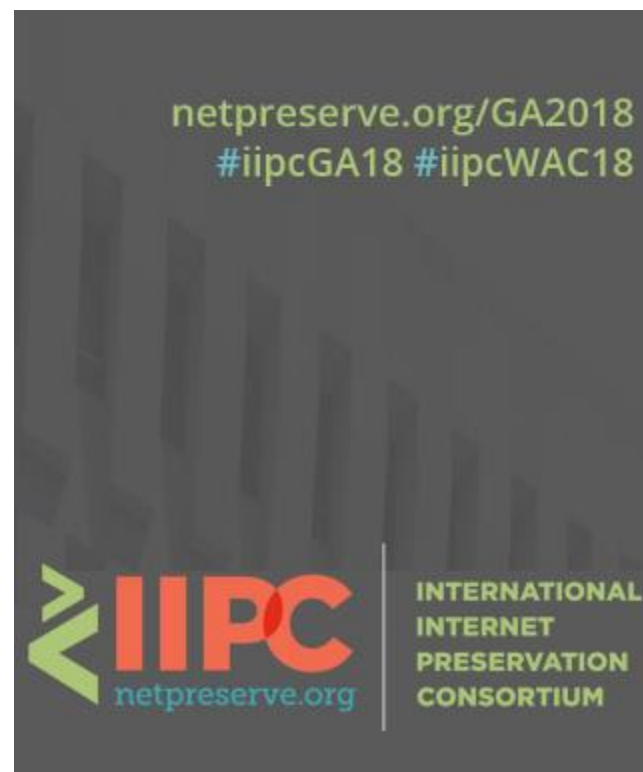


International Internet Preservation Consortium (IIPC) Web Archiving Conference (WAC)



IIPC developments

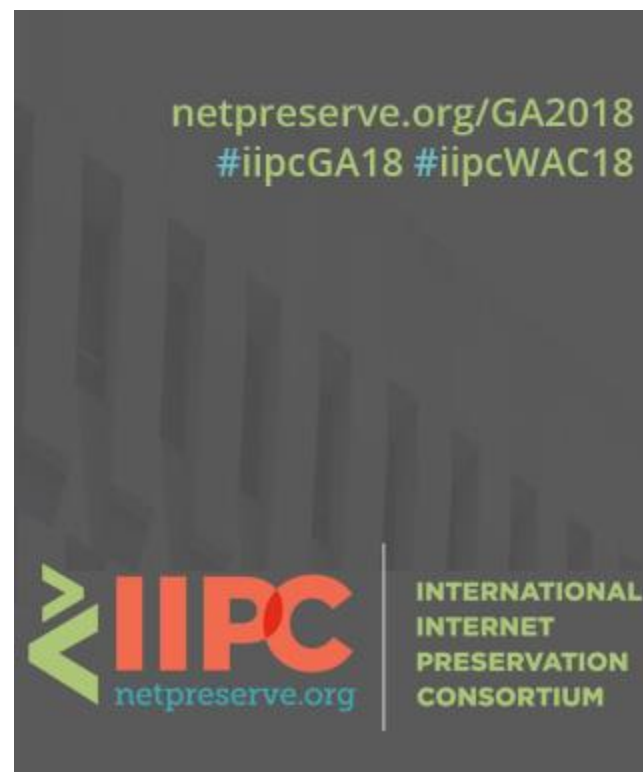
- fiscal / **organizational transition** complete
- strong **financials**
- investments in OER **training** materials
- new **proposal process**
- new **members**
- new **research working group**





select presentations

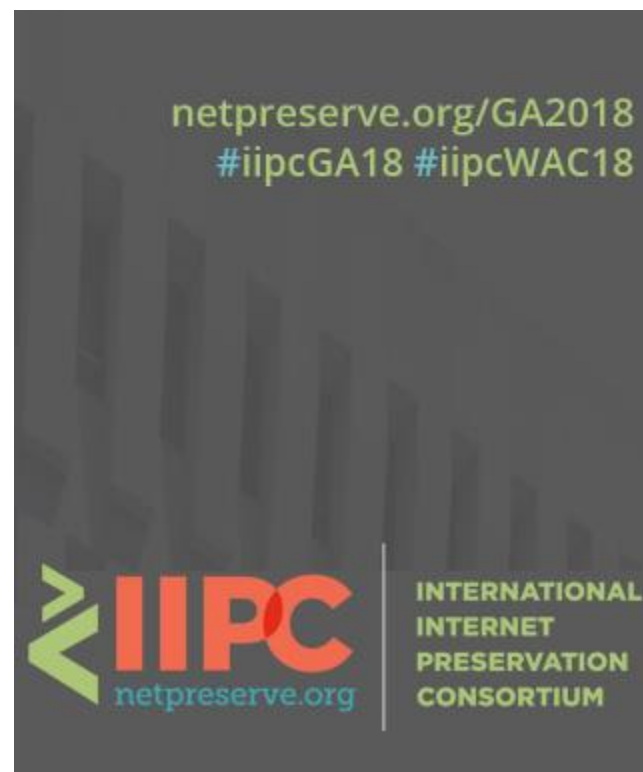
- Te Mawhai (the spiderweb) / Rachael Kai-Mahuta
- Archiving the Future / Wendy Seltzer
- [Addressing our many solitudes](#) / (Corey Davis), Jeremy Heil, Nich Worby
- [Preserving the public record vs “the right to be forgotten”](#) / Nicola Bingham
- [Arquivo.pt: taking a web archive to the next level](#) / João Gomes
- [WARC file format workshop](#) / Sara Aubry
- [Sifting needles out of \(well-formed\) haystacks](#) / Thib Guicherd-Callin
- [Collaborative, selective, contemporary](#) / Regan Murphy Kao, Yan Long, Nicholas Taylor, Zhaohui Xue





takeaways

- potential for broader impact + participation
- ambivalence about community investment in tools
- IA is (even more) core infrastructure for web archiving
- more consortial initiatives
- more concern w/ ethics

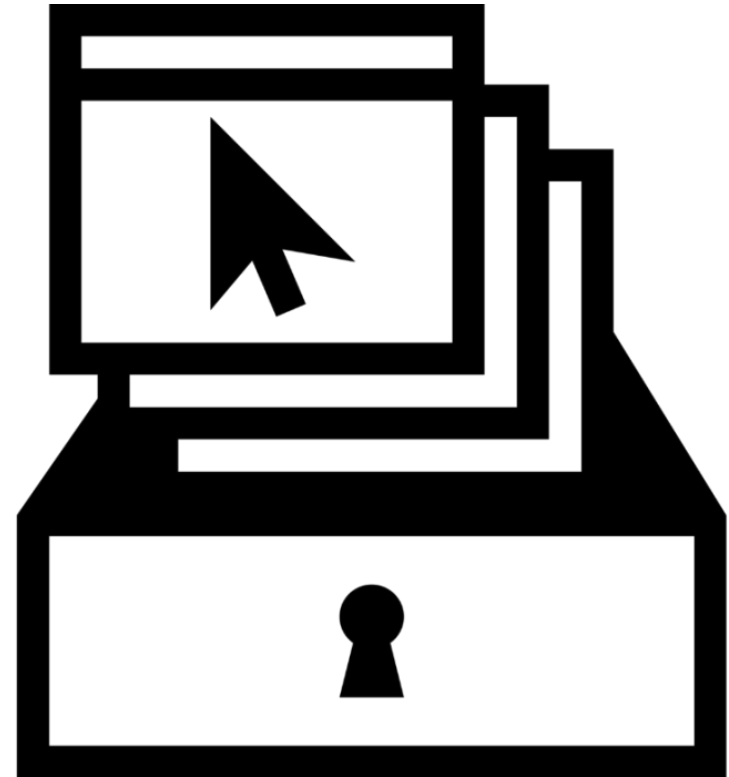




LOCKSS + web archiving

aligning w/ web archiving via LAAWS

- **common challenges:**
 - JavaScript link obfuscation
 - AJAX
 - high-fidelity replay
- **common approaches:**
 - JavaScript pre-execution
 - multi-crawler framework
- **common technologies:**
 - Open Wayback / PyWb
 - Web Archiving Proxy
 - WARC data format



IMLS WASAPI grant project

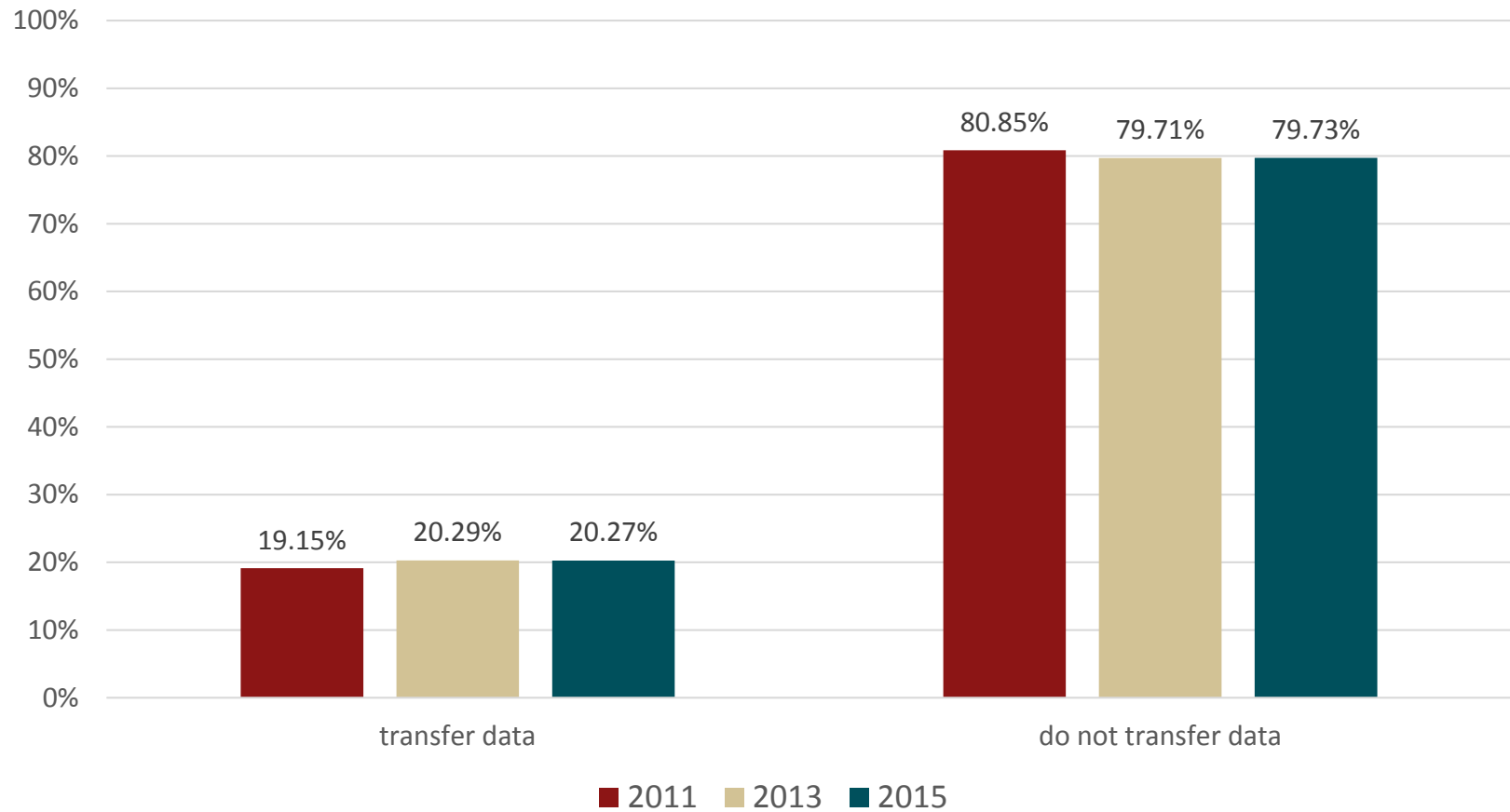
National Digital Platform Projects funded in August 2015

Systems Interoperability and Collaborative Development for Web Archiving

(LG-71-15-0174-15): The Internet Archive, working with partner organizations University of North Texas, Rutgers University, and Stanford University Library will undertake a two-year research project to explore techniques that can expand national web archiving capacity in several areas.



WARC download from Archive-It



WASAPI data transfer APIs

- API specification
- Archive-It + LOCKSS implementations
- use by Archives Unleashed Cloud
- client utilities
- documents + presentations
- Google Group + Slack

The screenshot shows the GitHub repository page for 'WASAPI-Community / data-transfer-apis'. The repository has 37 stars and 11 forks. The README file is selected, showing the following content:

WASAPI Data Transfer APIs

This is the public repository for work on the Web Archiving Systems API (WASAPI) data transfer APIs. The intention for these APIs is to provide a standardized mechanism for export and import of web archive data (and perhaps, ultimately, derivative data and capture metadata) between diverse systems for preservation, replication, research use, data delivery, and other purposes. Design and development is being carried out by project partners on the Institute of Museum and Library Services-funded National Leadership Grant, I-0-71-13-0174, "Systems Interoperability and Collaborative Development for Web Archiving" (SDCI) in consultation with a technical working group and based on feedback from the web archiving community.

Clients & Utilities

- Stanford University Digital Library Systems and Services: wasapi-downloader, <https://github.com/ru-dl/wasapi-downloader>
- UNT Libraries: py-wasapi-client, <https://github.com/unt-libraries/py-wasapi-client>
- LOCKSS: LAAWS Live Demo, <http://demo.laaws.lockss.org/>
- Rutgers University: Location Mapper, <https://github.com/jmw620/LocationMapper>

Documents

- 2017-12-12: "WASAPI Document Repository in archives.org"
- 2017-04-23: "National Symposium on Web Archiving Interoperability: Agenda & Presentation Links"
- 2017-03-02: "National Symposium on Web Archiving Interoperability: Trip Report"
- 2017-03-30: "IMLS Year 1: Interim Performance Report Narrative"
- 2016-11-29: "Archive-It 2016 State of the WARC Report"
- 2016-08-19: "WASAPI Survey on Data Transfer APIs"
- 2015-09-15: "Systems Interoperability and Collaborative Development for Web Archiving"

Presentations

- 2017-12-11: "Web Archiving Systems API (WASAPI) Project Updates and Data Transfer APIs Specifications and Demonstrations" at IIPC/RESAW Web Archiving Week
- 2017-04-27: "Stanford Web Archiving Work Cycle Inception Deck for Automating Web Archive Crawl Download and Accessioning"
- 2016-08-03: "WASAPI: Web Archiving Systems Application Programming Interfaces" at the SAA Web Archiving Roundtable Meeting at Archives Records 2016
- 2016-06-14: "WASAPI Web Archive Data Transfer APIs" at Archives Unleashed 2.0
- 2016-05-26: "Systems Interoperability and Collaborative Development for Web Archiving - Filling Gaps in the DMLS National Digital Platform" at Texas Conference on Digital Libraries
- 2016-04-12: "Building API-Based Web Archiving Systems and Services" at the International Internet Preservation Consortium General Assembly
- 2016-04-05: "Building National Web Archiving Capacity" at the CNE Spring 2016 Meeting

Meeting Notes

- 2016-12-13: WASAPI Technical Working Group
- 2016-03-30: WASAPI Technical Working Group

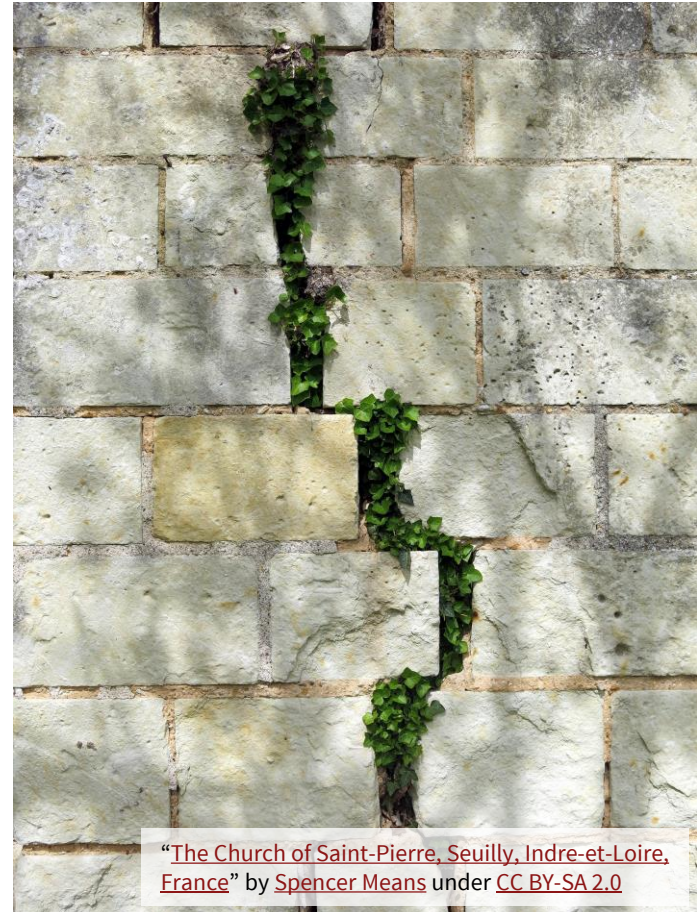
How to connect

- Join or send a message to the WASAPI community Google Group.
- Join the WASAPI Slack.

© 2017 GitHub, Inc. [Terms](#) [Privacy](#) [Security](#) [Status](#) [Help](#) [GitHub: data-transfer-apis](#)

Ivy Plus (web archive) preservation network

- complement **collaborative** web archiving initiative
- enable **consortial custody** of Archive-It WARC's
- **pilot network** w/ 5 nodes (4 institutions)
- **seeking buy-in** for continuing service
- **potentially expand** to preserve other, shared web archive + non-web digital collections



“The Church of Saint-Pierre, Seully, Indre-et-Loire, France” by Spencer Means under [CC BY-SA 2.0](https://creativecommons.org/licenses/by-sa/2.0/)



Questions

“Any Questions?” by [Matthias Ripp](#) under [CC BY 2.0](#)