Web Archiving and Digital Libraries (WADL)

Martin Klein
Los Alamos National Laboratory
Research Library
Los Alamos, NM, USA
http://orcid.org/0000-0003-0130-2097
mklein@lanl.gov

Zhiwu Xie
Virginia Tech
University Libraries
Blacksburg, VA 24061 USA
http://orcid.org/0000-0002-2702-3806
zhiwuxie@vt.edu

Edward A. Fox
Virginia Tech
Dept. of Computer Science
Blacksburg, VA 24061 USA
http://orcid.org/0000-0003-1447-6870
fox@vt.edu

ABSTRACT

The 2018 edition of the Workshop on Web Archiving and Digital Libraries (WADL) will explore the integration of Web archiving and digital libraries. The workshop aims at addressing aspects covering the entire life cycle of digital resources and will also explore areas such as community building and ethical questions around web archiving.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives**;

KEYWORDS

Web Archiving, Community Building, Digital Preservation

1 INTRODUCTION

Our understanding of the past will, to a large extent, depend on our success with Web archiving. WADL 2018 will bring together international leaders from industry, government, and academia, who focus their efforts on this important challenge. Together, we will explore the integration of Web archiving and digital libraries, over the complete digital resource life cycle: creation/authoring, uploading, publishing on the Web, crawling/collecting, compressing, formatting, storing, preserving, analyzing, indexing, supporting access, etc. The objectives of this workshop are to:

- continue to build the diverse community of people integrating Web archiving with digital libraries;
- help attendees learn about useful methods, systems, and software in this area;
- help chart future research and practice in this area, so more and higher quality Web archiving occurs;
- produce an archival publication that will help advance technology and practice; and

- promote synergistic efforts including collaborative projects and proposals.

2 RELATED WORK

The most recent related workshop, WADL 2017[1] was held in conjunction with JCDL 2017. The event was considered a great success by all participants, who voted in favor of a continuation of the workshop in 2018. Previous WADL meetings resulted in the publication of a special issue in the IEEE TCDL Bulletin such as in 2016 [3] and 2015 [2]. The workshop organizers are strongly motivated to continue this tradition with WADL 2018.

An earlier workshop, WIRE [4], focused on research of archival holdings and on making use of archives that preserve the web. The first workshop on Web Archiving and Digital Libraries, WADL 2013, led to a summary [1] after a group responded to the call for meeting[2] as part of the JCDL 2013 workshop program.

An earlier similar workshop at a prior JCDL conference took place in Ottawa in 2011[3], partly as a result of the emergence of a cooperative to explore Web archiving[4]. Broader in scope but related are the annual General Assembly meetings of the International Internet Preservation Consortium (IIPC)[5]. In addition, various sponsored programs have connected, like a closely related initiative[6] funded by the Andrew W. Mellon Foundation.

3 WORKSHOP TOPICS

WADL 2018 will cover all topics of interest, including but not limited to:

- Special Event Archiving and Collection Building
- Crawling of Dynamic and Mobile Content
- Social Media Archiving
- Archival Standards, Protocols and Systems, Tools
- Archival Metadata, Description, Classification
- Discovery of Archived Resources
- Extraction and Analysis of Archival Records
- Community Building
- Diversity in Web Archives
- Ethics in Web Archiving

---

[1] http://fox.cs.vt.edu/wadl2017.html
[2] http://www.ctrnet.net/sites/default/files/JCDL2013WorkshopWebArchiving20130603.pdf
[3] http://infolab.stanford.edu/wac/
[4] http://cs.harding.edu/wag2011/
[5] http://netpreserve.org/
[6] https://library.columbia.edu/bts/web_resources_collection.html

4 WORKSHOP LOGISTICS

4.1 Audience and Attendees

Over the past few years, we have observed that the community interested in this topic is growing. Based on interest in recent years, we expect to have 20-30 attendees, including a solid representation of students. We will advertise the workshop and actively solicit submissions. All submissions will be peer-reviewed by the program committee and accepted contributions will be compiled into the WADL program. We anticipate the program to include aspects from multiple disciplines such as Computer Science, Library and Information Science, Web Science, Social Sciences, etc.

4.2 Format and Duration

We propose a full-day workshop, possibly split across two days. We anticipate invited speakers, presentations of selected papers and posters, as well as demonstrations and panels.

4.3 Special Requirements

We have an international program committee of about 15 people, in addition to the three co-chairs. We will consider running a WebEx teleconference during the workshop so that those unable to attend at the last moment will still be able to be involved. We plan to have a small poster session in addition to the usual type of activities. As in 2015 and 2016, we expect to have a special issue of IEEE TCDL Bulletin as an opportunity to publish invited contributions. In the past, the workshop has led to a call for contributions for a special issue of IJDL. The co-chairs fully intend to invest this effort in a special issue again in 2018.

5 BIOGRAPHICAL INFORMATION

Martin Klein holds a Ph.D. in Computer Science from Old Dominion University. He currently is a scientist in the Research Library at Los Alamos National Laboratory. He was program chair of DL 2014, poster chair of JCDL 2015, panel co-chair of JCDL 2016, and conference co-chair of iPres 2016. In addition, he also was co-chair of WADL 2015, 2016, and 2017 and the General Assembly of the International Internet Preservation Consortium (IIPC) in 2017 and 2018. He is an editorial board member of the International Journal on Digital Libraries (IJDL), guest editor for the Bulletin of IEEE Technical Committee on Digital Libraries, and served as a board member of the Web Archiving Collaboration at Columbia University. Martin Klein is the lead editor of the ANSI/NISO Specification Z39.99 and has published numerous journal/magazine articles and conference/workshop papers. More information can be found at: http://orcid.org/0000-0003-0130-2097.

Zhiwu Xie is a professor at Virginia Tech Libraries. He leads the IMLS-funded "Developing Library Cyberinfrastructure Strategy for Big Data Sharing and Reuse" (LCI) project, the Goodwin Hall Living Lab data management system, IMLS ETDplus, and VTechData, among others. He is deeply involved in Fedora/Hydra, APTrust, PREMIS, ResourceSync, and AltMetrics Data Quality. His research has received supported from Mellon, IBM, Amazon, and NSF XSEDE. He was a co-chair of WADL 2015, 2016, and 2017. His website is at http://scholar.lib.vt.edu/staff/zxie/.

Edward Fox holds a Ph.D. and M.S. in Computer Science from Cornell, and a B.S. from M.I.T. Since 1983 he has been at Virginia Tech, where he serves as Professor. He directs VT's Digital Library Research Laboratory and the Networked Digital Library of Theses and Dissertations. He was a member of the Board of CRA (the Computer Research Association). He was chair of the IEEE Technical Committee on Digital Libraries, and earlier was chair of ACM SIGIR. He was chair of the steering committee for JCDL, and is on the international advisory committee for ICADL. He has been (co-)Principal Investigator on over 120 research grants/contracts. He has taught over 80 tutorials and has given 66 keynote/distinguished/international invited talks. He has (co-)authored 18 books, 120 journal/magazine articles, 49 books chapters, 211 refereed conference/workshop papers, 73 posters, and over 150 other publications/reports, plus over 300 additional talks. Fox was editor for IR and DL for ACM Books. He was Co-Editor-in-Chief for ACM JERIC, and is on the boards of IJDL, JEMH, JIIS, J. UCS, Multimedia Tools & Applications, and PeerJ CS. He is a Fellow of ACM and of IEEE. His website can be found at: http://fox.cs.vt.edu.

REFERENCES

[1] Edward A. Fox and Mohamed M. Farag. 2013. Report on the Workshop on Web Archiving and Digital Libraries (WADL 2013). *ACM SIGIR Forum* 47, 2 (2013), 128–133. http://sigir.org/files/forum/2013D/p128.pdf

[2] Edward A. Fox, Zhiwu Xie, and Martin Klein. 2015. Web Archiving and Digital Libraries 2015 (WADL 2015) Overview. *Bulletin of IEEE Technical Committee on Digital Libraries* 11, 2 (2015), 1–2. http://www.ieee-tcdl.org/Bulletin/v11n2/papers/intro.pdf

[3] Edward A. Fox, Zhiwu Xie, and Martin Klein. 2017. Web Archiving and Digital Libraries (WADL) 2016: Highlights and Introduction to this Special Issue. *Bulletin of IEEE Technical Committee on Digital Libraries* 13, 1 (2017), 1–2. http://www.ieee-tcdl.org/Bulletin/v13n1/papers/intro.pdf

[4] Matthew Weber, David Lazer, and Kris Carpenter Negulescu. 2014. WIRE 2014 Workshop - Working with Internet Archives for Research. (2014). http://wp.comminfo.rutgers.edu/nsfia/