

# Web Archiving and Digital Libraries (WADL) 2022

Martin Klein

Los Alamos National Laboratory  
Research Library  
<http://orcid.org/0000-0003-0130-2097>  
[mklein@lanl.gov](mailto:mklein@lanl.gov)

Zhiwu Xie

Virginia Tech  
Blacksburg, VA USA  
<http://orcid.org/0000-0002-2702-3806>  
[zhiwuxie@vt.edu](mailto:zhiwuxie@vt.edu)

Mat Kelly

Drexel University  
Philadelphia, PA USA  
<https://orcid.org/0000-0002-0236-7389>  
[mrk335@drexel.edu](mailto:mrk335@drexel.edu)

Edward A. Fox

Virginia Tech  
Blacksburg, VA USA  
<http://orcid.org/0000-0003-1447-6870>  
[fox@vt.edu](mailto:fox@vt.edu)

## ABSTRACT

The 2022 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, including creation/authoring, uploading/publishing, crawling, indexing, exploration, and archiving. It will also explore areas such as archiving processes and tools for “non-traditional” resources such as social media, scholarly and government datasets, 3D objects, and digital online art.

## CCS CONCEPTS

• Information systems → Digital libraries and archives;

## KEYWORDS

Web Archiving, Digital Libraries, Community Building, Digital Preservation

### ACM Reference Format:

Martin Klein, Mat Kelly, Zhiwu Xie, and Edward A. Fox. 2018. Web Archiving and Digital Libraries (WADL) 2022. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

WADL 2022 will continue the WADL tradition to provide a forum and collaboration platform to bring together international leaders from academia, industry, and government, who focus their efforts on the challenge of designing and implementing concepts, tools, and standards in the realm of web archiving. 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 a diverse community of people integrating web archiving with digital libraries,
- help attendees learn about useful methods, systems, and software in this area,
- help define and outline future research and practice in this area, to enable more high-quality web archiving,
- 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 2020 [5] was held as a virtual event<sup>1</sup>, in conjunction with JCDL 2020. We received very positive feedback from participants and are therefore strongly encouraged to continue the workshop in 2022. Thus far, two previous WADL meetings resulted in the publication of a special issue in the IEEE TCDD Bulletin such as in 2016 [3] and 2015 [2] and the workshop organizers are motivated to revitalize this effort with WADL 2022. Other workshop proceedings will be openly accessible from VTechWorks, Virginia Tech’s institutional repository.

A previous 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<sup>2</sup> as part of the JCDL 2013 workshop program.

An earlier similar workshop at a prior JCDL conference took place in Ottawa in 2011<sup>3</sup>, partly as a result of the emergence of a cooperative to explore web archiving<sup>4</sup>. Broader in scope but related are the annual General Assembly meetings of the International Internet Preservation Consortium (IIPC)<sup>5</sup>, for which the lead author of this proposal served as the general program chair in 2020 and 2021. In addition, various sponsored programs have connected,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

<sup>1</sup><https://fox.cs.vt.edu/wadl2020.html>

<sup>2</sup><http://www.ctrnet.net/sites/default/files/JCDL2013WorkshopWebArchiving20130603.pdf>

<sup>3</sup><http://infolab.stanford.edu/wac/>

<sup>4</sup><https://cs.harding.edu/wag2011/>

<sup>5</sup><https://netpreserve.org/>

like a closely related initiative<sup>6</sup> funded by the Andrew W. Mellon Foundation.

### 3 WORKSHOP TOPICS

WADL 2022 will cover all topics of interest, including but not limited to:

- Special event archiving and collection building
- Crawling of dynamic, online art, and mobile content
- Social media archiving
- National perspectives of web archiving
- Archival standards, protocols and systems, tools
- Archival metadata, description, classification
- (machine-friendly) discovery of archived resources
- Interoperability of web archiving systems
- Extraction and analysis of archival records
- Community building
- Diversity in web archives
- Ethics in web archiving

### 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 around 30-35 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, History, Journalism, etc.

#### 4.2 Format and Duration

We propose a full-day workshop. We anticipate invited speakers, presentations of selected papers and posters, as well as demonstrations and panels. In synchronization with JCDL conference organizers, we propose to prepare a hybrid workshop, which allows for safe in-person as well as online participation. Realistically, we anticipate last minute cancelations and changes to the program but we hope that a hybrid workshop model will accommodate such scenarios. We plan to schedule sessions with pre-recorded and live conference-style 10-15 minute presentations, while leaving plenty of room for discussion. In addition, we plan to have a small poster session where virtual and in-person attendees can present early-stage work.

#### 4.3 Special Requirements

We have an international program committee of about 15 people, in addition to the four co-chairs. If this proposal is accepted, we will need to rely on JCDL conference organizers to provide the technology to facility a hybrid event. As done for previous WADL events, the organizers will evaluate venues to offer an opportunity to publish invited contributions that were presented in their preliminary stages at the workshop. In the past, the workshop has also led

to a call for contributions for a special issue of IJDL. The co-chairs intend to invest this effort again and edit an IJDL special issue on web archiving following WADL 2022.

### 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 and leads the Prototyping and Development Team. He was program chair of DL 2014, poster chair of JCDL 2015, and panel co-chair of JCDL 2016. In addition, he has been co-chair of WADL since 2015. 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. 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 “Continuing Education to Advance Web Archiving” project, which grew out of discussions from prior WADL workshops. His research interests focus on library cyberinfrastructure, data management, digital preservation, and web archiving, with application areas ranging from ecological systems to smart building. He was a co-chair of WADL from 2015-2018. More information can be found at: <http://orcid.org/0000-0002-2702-3806>.

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) as well as a member of the ACM Publications Board and co-chair of its Digital Library Committee. 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 127 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-tedl.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.
- [4] Matthew Weber, David Lazer, and Kris Carpenter Negulescu. 2014. WIRE 2014 Workshop - Working with Internet Archives for Research. <http://wp.cominfo.rutgers.edu/nsfa/>

<sup>6</sup><https://library.columbia.edu/collections/web-archives.html>

- [5] Zhiwu Xie, Martin Klein, and Edward A. Fox. 2020. *Web Archiving and Digital Libraries*. Association for Computing Machinery, 583–584. <https://doi.org/10.1145/3383583.3398509>