

Web Archiving and Digital Libraries (WADL) 2023

Mat Kelly
Drexel University
Philadelphia, PA, USA
mkelly@drexel.edu

Zhiwu Xie
Virginia Tech
Blacksburg, VA, USA
zhiwuxie@vt.edu

Brenda Reyes Ayala
University of Alberta
Edmonton, AB, Canada
reyesaya@ualberta.ca

Edward A. Fox
Virginia Tech
Blacksburg, VA, USA
fox@vt.edu

ABSTRACT

The 2023 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

1 INTRODUCTION

WADL 2023 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 2022 [4] was held as a virtual event¹, in conjunction with JCDL 2022. We received very positive feedback from participants and are therefore strongly encouraged to continue the workshop in 2023. 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 2023. Other workshop proceedings will be openly accessible from VTech-Works², Virginia Tech’s institutional repository.

A previous workshop, WIRE [5], 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 as part of the JCDL 2013 workshop program.

An earlier similar workshop at a prior JCDL conference took place in Ottawa in 2011³, partly as a result of the emergence of a cooperative to explore web archiving⁴. 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 initiative⁶ funded by the Andrew W. Mellon Foundation.

3 WORKSHOP TOPICS

WADL 2023 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, and
- Ethics in web archiving.

¹<https://fox.cs.vt.edu/wadl2022.html>

²<https://vtechworks.lib.vt.edu/>

³<http://infolab.stanford.edu/wac/>

⁴<https://cs.harding.edu/wag2011/>

⁵<https://netpreserve.org/>

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

4 WORKSHOP ORGANIZATION AND 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. We plan to schedule sessions with pre-recorded and live in-person 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 facilitate a hybrid event. As was 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 2023.

5 BIOGRAPHICAL INFORMATION

Mat Kelly is an assistant professor at Drexel University's College of Computing and Informatics. He holds a Ph.D. in Computer Science from Old Dominion University. He is a Program Chair for JCDL 2023, was Publicity Chair for both JCDL 2018 and JCDL 2019, and Virtual Meeting Support Chair for JCDL 2021. He is a managing editor of IJDL. He leads the IMLS-funded "Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements". More information can be found at: <https://matkelly.com>.

Brenda Reyes Ayala holds a PhD in Information Science from the University of North Texas. She is currently an Assistant Professor at the School of Library and Information Studies at the University of Alberta. She is a Program Committee Member of JCDL and was Volunteer Chair of JCDL in 2018. She leads the project "Preserving our digital heritage: Improving the detection of quality problems in web archives", which is funded by the Social Sciences and Humanities Research Council of Canada. In the past she has worked as a web archivist at the University of North Texas Libraries and the National Library of Spain. More information can be found at <https://reyesayala.github.io>.

Zhiwu Xie is a professor at Virginia Tech Libraries. He led 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 has been a WADL co-chair since 2015. 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 is Chairman of 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 was Co-Editor-in-Chief for ACM JERIC, and is on the boards of J-ETD, JIIS, and IJDL. 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. <https://bulletin.jcdl.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. <https://bulletin.jcdl.org/Bulletin/v13n1/papers/intro.pdf>
- [4] Martin Klein, Mat Kelly, Zhiwu Xie, and Edward A. Fox. 2022. *Web Archiving and Digital Libraries*. Association for Computing Machinery, 1–2. <https://doi.org/10.1145/3529372.3530920>
- [5] Matthew Weber, David Lazer, and Kris Carpenter Negulescu. 2014. WIRE 2014 Workshop - Working with Internet Archives for Research. <https://lazerlab.net/event/wire-workshop-working-internet-archives-research>