

# Report on the Digital Library Federation 2015 Forum

26-28 October 2015, Vancouver, BC

Jody L. DeRidder

Under the new leadership of [Bethany Nowwiskie](#), the Digital Library Federation (DLF) is beginning to stretch its wings and cross new boundaries in terms of focus, inclusion, perspectives, and community engagement. The sold-out 2.5 day forum was preceded by [multiple community meetings, pre-conferences and workshops](#), and was held for the first time outside the U.S. (in Vancouver, B.C.). Shortly before the forum, Charles Henry, President of CLIR (Council of Library and Information Resources, the parent organization of DLF) [argued for an international effort to develop a Digital Library for the Middle East](#), to capture images of important cultural heritage content before it is destroyed. He had recently represented CLIR and DLF at a symposium organized by the Asia Society, Middle East Institute, Antiquities Coalition, and UNESCO, entitled "[Culture under Threat: the Security, Economic, and Cultural Impact of Antiquities Trafficking and Terrorist Financing.](#)" Also shortly before the forum, the [National Digital Stewardship Alliance \(NDSA\)](#) announced that it had selected DLF to serve as NDSA's institutional home starting in January 2016, in order "to broaden and catalyze the information stewardship community to safeguard permanent access to the world's scientific evidence base, cultural heritage, and public record."

## Contents

Opening Plenary.....	1
Sessions.....	2
Access.....	2
Preservation .....	3
Assessment .....	3
Closing Plenary.....	4
Conclusions .....	4

## Opening Plenary

The opening plenary began with an address about retaining access to disappearing cultures, by Larry Grant, adjunct professor in the University of British Columbia First Nations Languages Program and the Language and Culture Consultant for Musqueam First Nation, where he is an elder. He was followed by [Safiya Noble](#) from the Department of Information Studies at the University of California, Los Angeles, who is currently writing a book on Google and information bias (forthcoming, NYU press). She focused on the gender and racial biases in the algorithms used by Google to generate auto-suggestions and result lists, and noted that 73% of people using the search engine believe that the top results are

trustworthy<sup>1</sup>. She also cited an article<sup>2</sup> indicating how the results of Google searches impact the outcome of political races. The opening plenary closed with her call for librarians to join the “[critical librarianship](#)” movement to bring social justice principles into our work.

The [schedule for the forum](#) provided up to 4 choices of sessions during each time slot, with a variety of genres: presentations and panels, workshops, working sessions or series of snapshots (7-minute updates) and project updates (20-minute presentations) as well as a poster session accompanied by a lightning round. Many of the presentations focused on improving user access and usability; several addressed digital scholarship; some were digital humanities or research related, and a scattered few addressed digital preservation issues. Most sessions were presented by librarians, archivists, researchers and technologists, with a sprinkling of administrators. A handful of efforts and tools that may be of interest are mentioned here, broadly categorized as “access”, “digital preservation”, and “assessment.”

## Sessions

### Access

[BigDIVA](#) (Big Data Infrastructure Visualization Application) is a new tool developed at Texas A&M that enables researchers to decide for themselves which results are relevant, rather than relying on the search and page ranking algorithms.

[Mirador](#) is an open-source, web based, multi-window image viewing platform with the ability to zoom, display, compare and annotate images from around the world, leveraging the [International Image Interoperability Framework \(IIIF\)](#). At its essence, the latter is a method of implementing low-level APIs in software that enables sharing of content. Eleven national libraries and a host of other institutions have already committed to implementation, as well as some aggregators.

[Bibliopedia](#) provides a platform for researchers to organize, visualize, share and search archives without experience in metadata or data visualization.

In a Mellon-funded [Early Modern OCR Project \(eMOP\)](#), Texas A&M has developed an open-source tool called [FrankenPlus](#) for training Tesseract OCR to read handwritten materials, for full text indexing.

[Spotlight](#) is an open-source Stanford application that overlays a SOLR index (such as ours in Acumen) to enable non-technical people to develop exhibits of content with search and retrieval, themes, and browsing.

---

<sup>1</sup> Tarleton Gillespie. 2014. “The Relevance of Algorithms.” In [Media Technologies](#), ed. Tarleton Gillespie, Pablo Bockzkowski and Kirsten Foot. Cambridge MA: MIT Press). [Available from [http://mixedrealitycity.org/readings/Gillespie\\_TheRelevanceofAlgorithms.pdf](http://mixedrealitycity.org/readings/Gillespie_TheRelevanceofAlgorithms.pdf) ]

<sup>2</sup> Robert Epstein and Ronald E. Robertson. 2013. “Democracy at Risk: Manipulating Search Rankings Can Shift Voter’s Preferences Substantially Without Their Awareness.” Summary of a paper presented at the 25<sup>th</sup> annual meeting of the Association for Psychological Science, Washington D.C., May 2013. [Available from [http://aibr.org/downloads/EPSTEIN\\_and\\_Robertson\\_2013-Democracy\\_at\\_Risk-APS-summary-5-13.pdf](http://aibr.org/downloads/EPSTEIN_and_Robertson_2013-Democracy_at_Risk-APS-summary-5-13.pdf) ]

Seeking to optimize research data for discovery, reuse and citation, [Lagotto](#) can be implemented to allow users to track events around research articles and other scholarly outputs.

In an effort to include linked data in EAD finding aids, Emory University has developed an open source add-on ([namedropper](#)) for the Oxygen XML editor which helps identify names in finding aids for tagging.

### Preservation

At Columbia<sup>3</sup> they are developing workflows for managing large quantities of incoming digital content (which has no attendant metadata) using [FRED \(Forensic Recovery of Evidence Device\)](#) to access it, [Archivematica](#) (to test content, extract technical metadata, and generate bagIt archives with METS files), [Hydra](#) (for delivery) and [archivelt](#) (to generate finding aids for content and provide web access). They are leveraging the [Portland Common Data Model](#) for repository content. To navigate, they use [ORE \(Object Reuse and Exchange\)](#) and the [NepoMuk](#) ontology to provide a linked data representation of the content as it was organized within the file system structure. Lessons learned include:

- Survey the content before acquiring; costs are significant, and you need to cost it out before starting to do anything.
- If you can negotiate with the content creators in advance, do so, and get them to organize content.
- Document commitments and deliverables in advance, as the scope creep is amazing.

[ePADD](#) is a software package developed at Stanford to support archival processes around the appraisal, ingest, processing, discovery and delivery of email archives.

COPPUL (Council of Prairie and Pacific University Libraries) is piloting a cloud-based preservation service using [Archivematica](#). COPPUL manages the partnership and seeds the one-time setup costs, Artifactual Systems (developers of Archivematica) provides account administration, installation, server administration and technical support, and the University of British Columbia library provides fee-based server hosting and digital object storage service.<sup>4</sup>

### Assessment

In the Digital Library Assessment Interest Group presentation that I moderated, we reported on our white papers currently available for review on [user studies](#), [analytics](#), and [citations](#), and the progress made in developing a [digitization project cost calculator](#). In the working lunch that followed, we engaged the participants in determining our next steps, and forming working groups for this next year on assessment of metadata, further analytics development, expanded cost assessment, content reuse and developing guidelines for user studies.

---

<sup>3</sup> I have requested slides from the presentation.

<sup>4</sup> I have requested the slides from this presentation.

## Closing Plenary

The closing plenary, entitled “Capacity and Community: Setting Agendas for #ourDLF” demonstrated clear efforts to increase inclusiveness of cultures, perspectives, and types of content and to build community approaches to solving problems. Prior to the conference, [DLF members were asked to contribute ideas](#) for collaborative work and the future directions of DLF. [Ten members of the community](#) were asked to share their thoughts on these as the panel (of which I was one), and then members of the audience were asked to contribute as well. The result was a wide array of perspectives, calls to action, deeply personal revelations and expressions of gratitude, upon which DLF leadership intends to build greater community engagement in the year to come.

## Conclusions

The increasing interest and energy in the Digital Library Federation Forum and its community development is exciting, as it offers the opportunity to leverage collaborative efforts to develop solutions to common problems. The resources and ideas shared at the forum are potentially very useful. For example, I suggest we implement [Spotlight](#) over Acumen to highlight content and enable researchers to develop exhibits, and also implement [International Image Interoperability Framework \(IIIF\)](#) APIs over our images to provide broader access and usability of our content. We may also be able to leverage the [FrankenPlus](#) software to improve our extraction of searchable text from handwritten documents.

Columbia’s experiences in managing incoming digital content may help guide us in developing our own policies and procedures, and the cloud-based preservation service experiment may provide suggestions for how we may improve upon ADPNet services and functionality.

[BigDIVA](#), [Bibliopedia](#), and [Mirador](#) may all be useful for our researchers and our digital humanities efforts, and our combined progress towards developing best practices and guidelines in digital library systems will assist us in improving our return on investment.

Our participation in the Digital Library Federation is fertile ground for developing useful collaborations to further our interests.