



## Northeast Institutional Repository Day (#nird23)

### NIRD 2023 Program

November 30, 2023 - December 1, 2023

*Schedule and presentations subject to change*

*Session times include Q&A*

### Thursday November 30, 2023, all times in ET

**1:00 - 1:20 pm - Introduction**

**1:20 - 1:50 pm - Lightning Talks**

#### **Rethinking Institutional Repositories**

*Josh Cromwell, University of Southern Mississippi*

Over the past twenty years, institutional repositories (IRs) have become commonplace across most colleges and universities. While IRs were originally conceived as a means to collect and disseminate faculty scholarship, in recent years it has become apparent that this may not be the most effective use case for the modern IR. In light of this changing landscape, how should IR managers think about the IR today? This session will provide an overview of the forthcoming book *Rethinking Institutional Repositories*, published by ACRL, which seeks to answer this question through contributions from IR managers at a wide range of institutions. The session will also briefly highlight several case studies from the book that provide practical suggestions for managing the modern IR, developing innovative projects and use cases for the IR, and using the IR as a means to highlight and showcase diverse voices and viewpoints and to provide an inclusive platform for all members of the community.

#### **Surveying and Editing the Metadata of Our Marathon: The Boston Bombing Digital Archive**

*Anna Ryerson, Northeastern University*

In this talk, I will discuss my experience surveying and editing the metadata of a large crowdsourced public history archive. The Our Marathon collection includes nearly 8,000 items, with materials ranging from letters to collages to oral histories and other first-person accounts collected in the wake of the 2013 Boston Marathon bombing. Along the way, collaborations were established between Northeastern University and the NPR radio station, WBUR, the Boston Globe, and the Boston Public Library. This archive bears some resemblance to other projects that used crowdsourced materials in response to a public trauma, such as the September 11 Digital Archive and the Hurricane Digital Memory Bank relating the experiences of Katrina and Rita.

I added to and edited the Metadata Object Description Schema (or MODS) records from this collection, in order to clarify the copyright status, associated names and subjects of these materials, as well as the languages used in certain items, so as to improve discoverability for researchers viewing the collection through Northeastern University Library's Digital Repository Service. One of the biggest issues with these records initially was their lack of standardization and authorities, and in order to address these problems I needed to develop new ways of searching and surveying this collection.

In working with this collection, I have realized that it presents some challenges that are perhaps unique to such a large, crowdsourced response to a shared trauma. Because this is a kind of memorial, with a goal of both community building and healing, it is important for users to be able to access this material on their own terms. Yet the large number of items in this collection require organization to allow for meaningful access.

## **1:50 - 2:00 pm - Break**

## **2:00 - 2:50 pm - Presentations**

### **Curating Audiovisual Data in Data Repositories**

*Lauren Phegley, University of Pennsylvania; Meg Jerrild, ICPSR; Madina Grace, Duke University*

This presentation reviews the practices of curating audiovisual data for submission into data repositories. As part of our Data Curation Network training program, we decided to write a primer on the topic of curating audiovisual data. Audiovisual materials are not a common form of research data in all fields, but is a burgeoning data type especially in the social sciences. Audiovisual data curation processes are not well documented, which motivated us to write a foundational and accessible guide for curators. We interviewed multiple experts in this field in order to learn more about their needs, challenges, and existing procedures. Our hope for this guide and presentation is to encourage further exploration of this fast developing topic. This presentation will cover our investigation of curation workflow, ethical issues, technical concerns, documentation, metadata, and special considerations.

### **Gateway to the Third Dimension: Extending IRs to Support 3D Model Data**

*Bryan J. Brown and Dave Rodriguez, Florida State University*

For the last 2 years, Florida State University Libraries' Technology and Digital Scholarship department has been exploring ways to incorporate 3D model data into DigiNole, FSU's digital repository built on the open source Islandora 7 platform. In August 2023 we finally reached a major milestone with a publicly viewable demo open for review and testing by internal stakeholders. This represents the culmination of much collaborative work between 3D modeling subject experts and developers, and a lot of hard lessons learned along the way that we are ready to share with others trodding a similar path. Join us for a summary of our journey into the unfamiliar realm of 3D modeling, and learn how you can extend your IR to handle 3D model data as well!

This session covers the specific requirements for supporting 3D model data as scholarly research outputs stored in an institutional repository, and is aimed at an audience familiar with standard scholarly content in an institutional repository but with little to no prior knowledge of 3D modeling. Topics covered will include the production and use of scholarly 3D model data by students and faculty, unique metadata elements for 3D model data, how to create a 3D model test suite, tips for integrating a 3D model viewer like the Online 3D Viewer (<https://3dviewer.net/>) into your repository, an overview of the most popular 3D model file formats, and a discussion about the complexities of building a system that creates a consistent user experience for ingesting and displaying 3D models with surprisingly inconsistent data structures.

## **2:50 - 3:00 pm - Break**

## **3:00 - 4:00 pm - Presentations**

### **Approaching Accessibility for Your IR**

*Christine Anne George, Mariah Lewis, and John Schriener, NYU Law Library*

Accessibility standards are commonplace. While this progress is something to be championed, it can leave an institutional repository in a difficult situation. How do you uphold accessibility standards when you are not in creating the materials that are being added to your IR? This session will start by looking at the policy and potentially political considerations, as well as the practical aspects of implementing the policy. How do you implement the standards? What stakeholder buy in do you need? Is there technology that can help? Are you actually able to acquire the technology? Who is going to pay for the technology? Will AI truly save us all? And, because of the inevitable way that things like this usually go, how do you formulate Plans B – D just in case things do not go according to plan/take longer to implement? This

presentation is coming from librarians who are currently working through this process and will outline what we've done so that others (hopefully) don't have to experience the same. Commiseration and comments from the audience will be highly encouraged.

## **Friday December 1, 2023, all times in ET**

### **10:00 - 10:50 am - Lightning talks**

#### **Agility in Changing Institutional Repository Platforms**

*Peter Whiting, University of Southern Indiana*

The purpose of this presentation is that it is perfectly okay to instill flexibility to change an institutional repository (IR) platform that will improve the (IR) user experience. New to the IR landscape in 2019 this was the first IR at this academic institution. Even with an ambitious focus to make this a shiny IR resource it fell short in its mission. It was time to go back to the drawing board in 2022, post-Covid pandemic, to search the IR landscape for a new generation IR. The goal was to have an abundance of modern features in the IR that would survive past the expiration date of staleness. Agility to change is necessary and making the change in a timely matter can benefit both library users of the IR and the library staff overseeing the IR. After four years with an IR platform the change in 2023 to a new IR platform that became a successful launching pad for new IR experience.

#### **Second Time Is the Charm: A DSpace 7 Migration Story**

*Erica Johns, Robert Wilson, and Brandon Kowalski, Cornell University*

Members from Cornell University Library's IT department share lessons learned from two attempts to migrate their DSpace 6.3 repository to the new DSpace 7.6 infrastructure. Learn about performance tuning, customizations, streamlined and secure access, containerization, and more!

#### **Kwalk: A Simple Program to Crosswalk Metadata for Repository Uploads**

*Kirsten Vallee, University of Chicago*

University of Chicago's Center for Digital Scholarship has been utilizing this program to better edit metadata for batch upload to Knowledge@UChicago. There are plans to share this software in the future as it is platform agnostic and has a potential wide range of use cases.

Suppose you need to upload 1,000 items to TIND from a source like Lens.org or PLOS journals. You obtain informal metadata for the items by you or another person creating the spreadsheet from scratch, exporting the data, or web scraping each individual record. You might need to do the following after obtaining the data:

- Rename all the fields in the from the invented field names to TIND's field names
- Add some fields that are missing
- Leave out some fields you don't want
- Combine several fields into one field
- Modify the values of date formats or author names in a programmatic way
- Generate syntactically correct upload URLs from a simple filename field

Kwalk is a program that lets us write a simple crosswalk that we can apply to each batch of metadata as we receive it and have multiple crosswalks for multiple projects as we work on them in an intermixed fashion. The program allows us to apply special functions to modify date formats, combine literal and field name text, generate uniform upload URLs, and much more.

#### **Increase Discoverability of IR Works by Utilizing SEO (Search Engine Optimization) Tools**

*Yan Yu and Beth Klein, University of Notre Dame Kresge Law Library*

The University of Notre Dame Law School utilizes Digital Commons to maintain its institutional repository, known as NDLScholarship, which is overseen by a small team within the Kresge Law Library at the Law School. Upon conducting a content review of the repository in collaboration with the Bepress consultant, it became evident that the SEO tools and features are overlooked. The new manager initiated a project to enhance the discoverability of the repository's content, leveraging Digital Commons' built-in SEO tools and features. This presentation intends to outline the team's approach, including the methods employed to gather and prioritize metadata, as well as the creation of descriptive page titles and introductory text to be incorporated at various structural levels such as the site homepage, community page, and individual publication items.

**10:50 - 11:00 am - Break**

**11:00-11:50 am - Presentations**

**Making Migration Less Mysterious: Developing a Migration Plan for ScholarWorks@UMassAmherst**

*Erin Jerome, University of Massachusetts Amherst*

In January 2023, after years of environmental scans, interviews with stakeholders and other IR managers, and platform investigations and pilots, the UMass Amherst Libraries made the decision to migrate its IR from bepress' Digital Commons to a combination of Janeway and Atmire-hosted DSpace 7.x. We all love a good migration presentation, but for most of us, migration remains a mysterious process that's difficult to envision. In this talk, I will walk through my process of creating a migration plan for our rather large and unwieldy IR – from interviews with IR managers who have been through migration, the beginning stages of data cleanup and standardization, and the fun--Excel column limits! Items uploaded multiple times!--discoveries made along the way. I will also discuss how the cleanup and discoveries are shaping our IR policies moving forward.

**Signs and Wonders: Integrating Multiple Systems to Digitize the Deaf Catholic Archives**

*Lisa Villa, College of the Holy Cross; Lenora Robinson, Deaf Catholic Archives/College of the Holy Cross*

The Archives and Distinctive Collections at the College of the Holy Cross is in the second half of a two-year CLIR "Digitizing Hidden Collections" grant project to preserve and provide access to key components of the Deaf Catholic Archives (DCA). Ideally, digital content would be accessible through both the institutional repository as well as the archival finding aids.

Wonder how we did it? The project utilized three systems (Digital Commons, ArchivesSpace and Google) to efficiently upload thousands of items to our repository and link them to their respective finding aids.

This presentation will discuss how we worked to build the repository structures and then developed workflows to create and populate an online collection that allows multiple access points for a large, complex, and growing archive. As this work continues, the next phase of the grant will use another tool to create exhibits using our IR, thus facilitating additional access, promotion and outreach efforts. These efforts, as well as the creation of metadata and description, will intentionally include input from members of the Deaf community here at Holy Cross and in partnership with others.

Our experience offers suggestions for how to build out a large collection with several structures that require different technical treatments. We have already begun adapting them for other digitization projects. Though we used specific platforms, this presentation will demonstrate how different applications can be harnessed for a large, long-term project.

**11:50 am - 12:00 pm - Wrap up**