A tale of two migrations: a medical library case report

Lisa Palmer, Tess Grynoch, Sally Gore
Lamar Soutter Library, UMass Chan Medical School, Worcester, MA
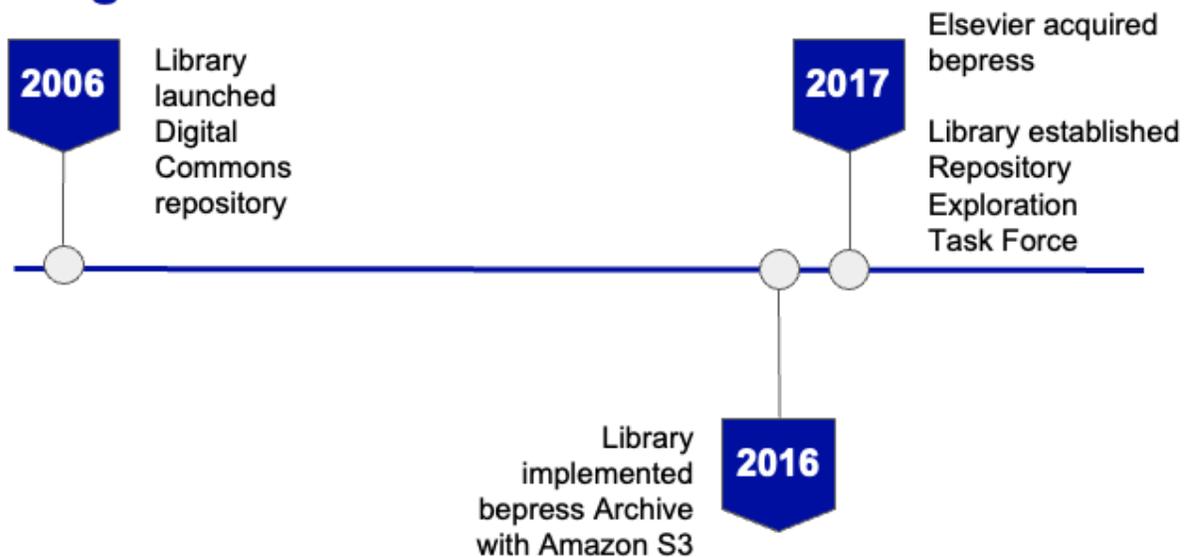
Lamar Soutter Library
Education • Research • Health Care
Empowering the future. Preserving the past.

UMass Chan
MEDICAL SCHOOL

Hello, I'm Lisa Palmer. We recently completed two migrations at UMass Chan Medical School. These migrations were from the bepress Digital Commons platform to Open Repository, a DSpace solution from Atmire, and Janeway, a publishing platform. My colleagues Tess Grynoch and Sally Gore are the other members of our library migration team and helped prepare this lightning presentation.

Background

2006 — Library launched Digital Commons repository

2016 — Library implemented bepress Archive with Amazon S3

2017 — Elsevier acquired bepress

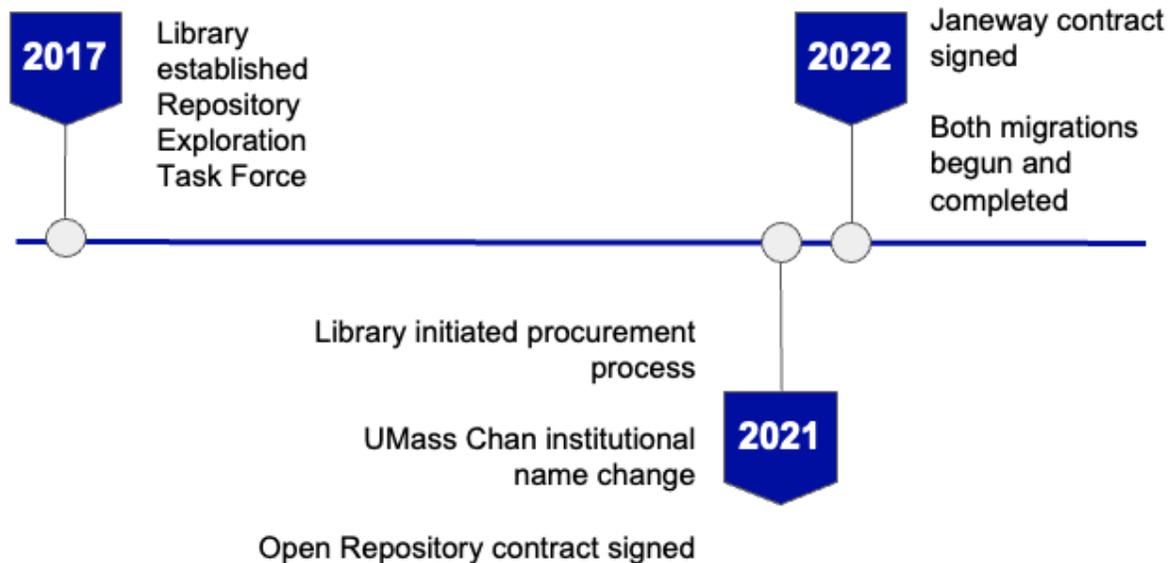Library established Repository Exploration Task Force

2

UMass Chan was an early adopter of Digital Commons. As seen in this timeline, we launched our eScholarship repository in 2006 and were their second medical school customer. In 2016 we implemented the bepress Archive service, which works with Amazon's Simple Storage Service (S3) to provide an archive of IR content and metadata.

In August 2017, bepress was acquired by Elsevier. This raised several questions for us, most notably future affordability and our ability to maintain our commitment to open access. That December, library management approved a task force to assess needs and make recommendations. Two key factors driving our work were (1) we wanted to support sustainable, community-led open source platforms and (2) our library does not have the capacity to manage a locally hosted repository. Thus, we focused on hosted, turnkey solutions that do not require substantial library staff resources for development or system administration.

eScholarship@UMassChan utilized Digital Commons' repository and publishing features. Based on this, we developed campus-specific functional requirements, performed an environmental scan to identify options, evaluated numerous platforms, narrowed our choices and met with providers. After several years (and a global pandemic), in June 2021 the library initiated the procurement process for Open Repository and for Janeway. Legal approvals and organizational changes, including our medical school changing its name in September 2021, proved to be substantial hurdles. We finally signed a contract with Open Repository in November 2021 and with Janeway in February 2022.

Migration

- Items: 25K+
- Collections: 200+
- Four active peer-reviewed journals and one serial textbook migrated to Janeway
- Metadata sources: Amazon S3 archive, various csv reports from Digital Commons, output from the Digital Commons API and OAI feeds
- March-September 2022: Weekly virtual meetings with library implementation team, UMass Chan Information Technology, service providers
- Deadline: Digital Commons subscription ends mid-September 2022

4

After an onboarding period, migration activities began in earnest in February 2022, with a library implementation team of four people. Four peer-reviewed journals and one textbook were migrated to Janeway, and everything else to Open Repository. We utilized various metadata sources which are listed on the slide. From March through September, the library team and representatives from UMass Chan Information Technology met weekly with each provider. Given our remote working structure and the fact that both providers are located in Europe, this wasn't the easiest accomplishment, but we worked hard and successfully met our September deadline, when our Digital Commons subscription ended.

Challenges

1. Time zones
2. Security (Single Sign On)
3. DNS and URL redirects
4. Digital Commons data
   a. Geographic usage data not migrated
   b. In-process journal manuscript files and peer reviews could not be bulk exported
   c. Metadata for hidden supplementary files not included in bepress Archive
   d. Collection filtering information not included in bepress Archive metadata
   e. Community / collection logos and descriptions saved manually

5

Our migration had several challenges that caused delays, in addition to pandemic-related work disruptions and staff shortages:

1. Finding a meeting time and timely correspondence were challenging across multiple time zones.
2. The repository had not used Single Sign On (SSO) technology previously to authenticate and login securely. We implemented it with Shibboleth for Open Repository and OIDC for Janeway. UMass Chan IT figured out how to test SSO for both systems in a secure way.
3. To ensure Digital Commons URLs redirected accurately to both new platforms, a chain of redirects needed to be set up and the DNS (domain name system) implementation timed carefully.
4. We also experienced issues with migrating Digital Commons data, some of which are listed on the slide. We were unable to to programmatically export granular usage data with geographic information, such as download counts for every item for every month broken down by country of origin. We decided to migrate usage statistics without the geographic data. For the journals, files for manuscripts in process and their peer reviews could not be bulk exported.

We had to download all files individually and organize them for the Janeway developers. Supplementary files hidden for display are included in the bepress Archive, but unfortunately their metadata is not included in that metadata file. This made it very difficult to identify these files programmatically and make sure they did not display in the new repository. The collection filtering information used to map items to display in multiple collections was not in the bepress Archive metadata. I am rebuilding these virtual collections in the new repository. Community and collection logos and descriptions are not available for migration. We manually saved the images and html configuration information as text files.

This was our experience … your mileage may vary and things may be different for future migrations. Bepress was very helpful in answering my questions and assisting when possible.

Lessons learned

1. Doing two migrations at once is difficult
2. Build in more time to the migration schedule for troubleshooting and customizing
3. Preparing for migration
   a. Clean up and standardize metadata
   b. Document metadata for each collection, including custom fields
   c. Ask journal editors to review and clean up the reviewer list
4. Bring local IT in early on

6

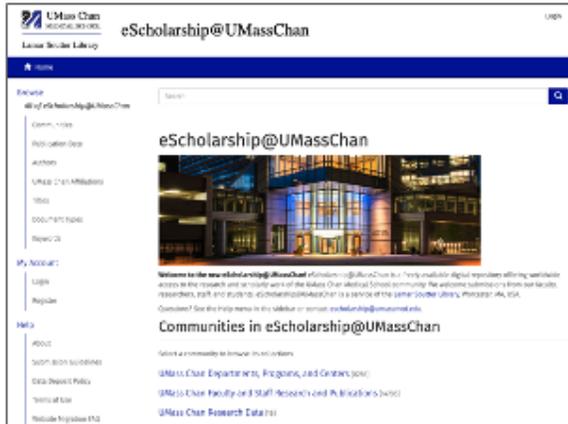We learned many lessons from this experience but I will focus on a few.

1. Doing two migrations at once was difficult. No kidding, right? Our plan was to do the Janeway migration first. Factors described earlier made this impossible. As a result, resources were stretched, URL redirects were more complicated, and everything was more stressful!
2. In terms of your migration schedule, build in much more time than you think you'll ever need.
3. Prepare by cleaning up records as needed. Make your submission forms and use of metadata fields as consistent as possible. Standardization pays off when you're mapping metadata to the new system. Document any custom fields. For active, peer-reviewed journals, ask your editors to clean up the reviewer list before the migration if possible.
4. Finally, bring local IT in early on, if needed. Our weekly meetings with IT and the service providers were critical and helped with communication and relationship building.
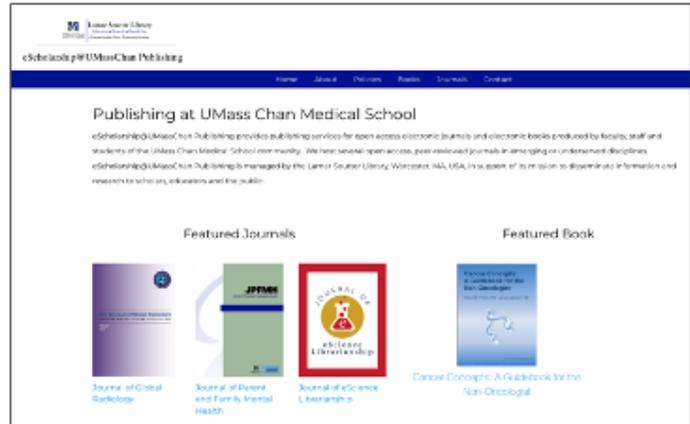
The screenshots on this slide are of the home pages for our new repository and publishing systems: repository.escholarship.umassmed.edu and publishing.escholarship.umassmed.edu. The repository will migrate to the DSpace 7 version of Open Repository in 2023, so there is no rest for the weary. We have learned a lot and look forward to a smooth migration.

Thank you! I look forward to any questions. Or please feel free to reach out to me at Lisa.Palmer@umassmed.edu.