Digital Commons to Hyku: An Institutional Repository Migration at a Small Liberal Arts University

Johanna Meetz & Laura Baird

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.

Digital Commons to Hyku: An Institutional Repository Migration at a Small Liberal Arts University

Johanna Meetz
The Ohio State University

Laura Baird
Pacific University

ABSTRACT

Introduction Pacific University Libraries has had an institutional repository since 2009, when it selected Digital Commons to host a collection of theses and dissertations. Since then, the scope of the services has grown to include publishing open access journals as well as housing the books published by Pacific University Press—a library-born, hybrid, open access press. As our needs have changed, and with Elsevier's acquisition of bepress in 2017, the University migrated from bepress' Digital Commons platform to an open source Hyku platform hosted by Ubiquity Repositories.

Description of Program: As the first academic institution working with Ubiquity Repositories on migration and implementation, we were involved in the process of data extraction, normalization, mapping, ingest, and validation.

Lessons Learned: We learned the importance of having a mutual understanding of a platform's goals, data structure and mapping, and standards in implementation decisions.

Next Steps: As higher education continues to adapt to the changes brought by COVID-19, it has never seemed more important to utilize platforms that share the values of libraries worldwide. We hope that migrating to an open source platform will be a step toward more open scholarship, despite the current challenges and resource scarcity brought about by the pandemic.

INTRODUCTION

Like many institutions, Pacific University founded its institutional repository program over a decade ago. The institutional repository technologies, uses, and overall purpose have evolved from accepting deposits, providing access, and preserving scholarship related to Pacific University to include publishing journals, conference materials, and books. This development has happened alongside Pacific University's own transition from a traditional, liberal arts, undergraduate institution, to an institution with substantial growth in graduate, research-based, health professional programs in the 2000s. Research funding at Pacific University grew from $1,267,000 in 2009 to $7.7 million in 2017, making it the top funded private research university in the Pacific Northwest (Lang, 2019). As research has become part of Pacific University's institutional brand, increasing the visibility of our research has become part of the University mission, which has continued to be supported by the institutional repository, CommonKnowledge. In addition, Pacific University Libraries collaborated with faculty to create five scholarly journals, publish two others, and host an additional journal. Since founding Pacific University Press in 2016, 6 books have been published. These journals and books have allowed Pacific University and the Libraries to participate in the creation of scholarship as well as in its dissemination via the institutional repository.

This paper describes how Pacific University migrated from our existing repository, bepress' Digital Commons, to a new open source repository, Hyku, hosted by Ubiquity Repositories, which is part of Ubiquity Press. Although the Ubiquity Repositories' Hyku team had worked with cultural institutions, Pacific University was the first University to migrate to this platform. We hope that by describing our migration process, we will help other institutions who are considering their own migration from Digital Commons, or other platforms.

LITERATURE REVIEW

Over the past thirty years, institutional repositories have developed as a specialized type of digital asset management system. The desktop publishing advances in the late 1980s allowed digitization of text, graphics, and other files. In response to this growing number of born-digital and digitally-reformatted materials, the 1990s saw the creation of digital asset management systems ("What is Digital Asset Management?," 2020). The first disciplinary repository, arXiv, was released in 1991 (Jones, 2006). Similar e-print collections organized by discipline grew including EconPapers and CogPrints. As an alternative to subject-based repositories, institutional repositories were proposed in 1999 (Ginsparg et al., 1999).

The term, *institutional repository*, was first documented by the Scholarly Publishing for Academic Resources Coalition (Crow, 2002) as "digital collections capturing and preserving

the intellectual output of a single or multi-university community" (pp. 4). Several important defining characteristics were outlined for institutional repositories, including digital, institutionally defined, scholarly, cumulative and perpetual, open access, and interoperable. Similarly, Johnson (2002) focuses on institutional repositories as a demonstration of an institution's scholastic accomplishments and an alternative to the commercial scholarly publishing model. Later, Clifford Lynch (2003), director of the Coalition for Networked Information, argued that institutional repositories could help expand and diversify the forms of scholarship by preserving and hosting content. Institutions may have different purposes or motivations for institutional repositories. These may include control over intellectual output (Branin, 2005), assessment of intellectual output (Goodyear & Fyffe, 2006), or supporting teaching and learning activities and materials (Rogers-Urbanek, 2008).

The first institutional repository platform, EPrints (2000) was one of the most widely-used free, open access options. DSpace (2002) from MIT and HP Labs, and Digital Commons (2002) were released shortly after (Wikimedia Foundation, 2018). Consistent with the purpose of institutional repositories, DSpace and EPrints are open source platforms, and were widely adopted. By 2018, 31.8% of North American institutional repositories were hosted on DSpace and 6.3% on EPrints. Although not open-source, Digital Commons offered hosted software to 29.1% of these repositories (Luther, 2018). Nearly twenty years later, OpenDOAR, a global directory of open access repositories, lists 5,282 repositories from 128 countries ("Browse by Country," 2020).

There are pros and cons to both open source and proprietary institutional repository platforms. When proprietary platforms are chosen, it is often because the cost "of technical infrastructure…interface branding and customization; electronic publishing options" is built in (Corbett et al., 2016, pp. 4). In addition, proprietary platforms are generally better able to handle a quick implementation (Corbett et al., 2016, pp. 4). When an open source platform is chosen, it is often because of the high level of customization offered, as well as "extensibility, flexibility to ingest various formats, and interoperability" (pp. 5). In addition, though open source platforms are freely available, the cost to "build and support a repository based on open-source software" is significant (pp. 7). In fact, "the library must be able to *fully* support the repository—"adequate" support for such a significant and high investment resource is not enough (pp. 9). The number of positions and/or staff time needed at each library varies, but is much higher than what is needed to support a proprietary platform. Though there is still a worthwhile distinction to be made between open source and proprietary platforms, service models that may offer the best of both worlds are being developed. In fact, there are now "a number of commercial support services available for open source systems, ranging from hourly vendor support to full software-as-a-service offerings" (pp. 10).

With proprietary systems, the design, implementation, and outlook are at the discretion of the owner. In 2017, the nature of proprietary institutional repository platforms changed when Elsevier (2017) announced the acquisition of bepress. Citing Elsevier's past actions against open access, librarians expressed ethical concerns about the impact of this acquisition. Several announced plans to begin reviewing alternatives for possible migration (Royster et al., 2018). University of Pennsylvania (Penn) Libraries (2017) was the first bepress customer to announce their plan to investigate open source options to host Penn's institutional repository. Though Penn expressed concern about the effects of Elsevier's acquisition of bepress and acknowledges that "such changes are not in the best interests of the library community," this was only part of the reason they felt compelled to start investigating alternatives to bepress.

In general, rationale for migrating from one system to another varies. Stein and Thompson (2015) found that organizations migrate due to dissatisfaction with key functionality, predicted future needs like scalability and extensibility, and more. In their survey, nearly two-thirds of institutions surveyed were moving from proprietary to open-source digital asset management systems. Extensibility was a significant need identified by respondents, whether to allow institutions to develop their own functionality, use an API, or ensure that the code base is open source. In Penn's case, in the 13 years they have had a repository, they have added additional services such as "faculty-assisted submission, journal publishing, and digital project consultation" (Ferguson, 2018). Exploring other platforms will afford Penn the opportunity to evaluate which platforms are the best fit for the services they currently offer. As of 2020, Penn, as well as most other institutions, are still customers of bepress. A user survey from Stanford University Libraries suggests many institutions want to migrate from their current repository but lack the resources to do so (Frost & Gary, 2016).

DESCRIPTION OF PROGRAM

Background

Pacific University Libraries have been collecting scholarship in the institutional repository, CommonKnowledge, since June 2009. For the Libraries, the purpose of the institutional repository is to: (1) collect and make Pacific University's scholarship openly available, (2) steward Pacific University's scholarship, and (3) preserve the works it contains. Pacific University Libraries initially chose bepress because it was affordable, did not require hosting, and offered personal technical support. Over time, the number of scholarly activities and services supported by the institutional repository grew from faculty publications and theses and dissertations to include publishing conference materials, journals, and books. These additional services significantly changed the focus of the scholarly communication work done at Pacific University, and, by extension, the desired functionality of the repository platform. Also, the number of institutional repository

and publishing platform options had increased since 2009, while the cost of bepress had grown significantly. Like many institutions, though the overall budget of Pacific University, and of the Libraries in particular, has grown with time, it has not grown at the same rate over time, and there has been little growth in funding for resources and technology in particular. Much like Penn, as well as others, all of these factors, plus Elsevier's acquisition of bepress, caused Pacific University Libraries to begin reviewing alternative platforms in 2019.

The criteria used to identify and select a platform was limited. Since Pacific University is a smaller university supporting about 4,000 students, resources to host and support an in-house repository were unavailable, so a hosted solution needed to be found. In addition, the cost of the institutional repository and the journals the Libraries publish needed to be considered. Since Digital Commons does not charge its customers a fee per journal, that additional journal hosting cost was a factor. However, at the time of the migration, the Libraries did not intend to increase the number of journals it published or hosted—in fact, the number of journals was decreasing from seven to three. This reduction in journals made it possible for the migration to occur at a cost savings. Other bepress customers who may be interested in growing their journal publishing program will need to factor in this additional cost.

In 2019, a contract was signed with Ubiquity Press to migrate from Digital Commons to their hosted version of Hyku. Hyku is a turnkey repository application based on Hyrax, Fedora 4, and the Portland Common Data Model. Hyku's origins can be traced back to the Hydra Project (2008) which grew into Samvera (2017), an open source repository framework which allows a high degree of functionality and flexibility in implementation. While the Samvera framework and its components provide many options, some institutions preferred a stable code base to building a customized solution with Samvera. In 2017, Hyrax was developed to provide a common implementation of Samvera software (Samvera, 2020). Hyku, originally named Hydra-in-a-Box, is a customization of Hyrax designed to be scalable, hosted, and configurable with templates for deployment on Amazon Web Services (Frost, 2017). Although institutions may create an institutional repository based on Samvera, Hyrax, or Hyku, Hyku is intended to work as an out-of-the-box repository. Currently, two providers offer hosted Hyku solutions: Ubiquity Repositories by Ubiquity Press and HykuUP by Notch8 (Hyku, 2020). Though they share a common software base in Hyku, Ubiquity Repositories and HykuUP have some implementation differences, including user interface decisions, content types, and analytics interfaces.

Functionality, pricing, vendor relationships, support options, and other aspects factored into the choice of vendor. Pacific University Libraries chose to build on the existing relationship with Ubiquity, who were already publishing one of our journals in their hosted version of OJS (Open Journal Systems). Our migration team consisted of the vendor staff from Ubiquity

Repositories, the Scholarly Communication and Publishing Services Librarian, and the Systems & Applications Librarian. This core group worked to make the implementation decisions and execute the migration. Due to the small size of this group, many roles and responsibilities were shared or collaboratively executed, with primary focuses for each individual (see Figure 1). Other Pacific University employees contributed expertise for decisions about legal issues, policy, digital integration, and departmental processes.

The migration process started in the spring of 2019 and continued until April 2020, the go-live date for the new institutional repository. Institutional time commitment varied

Owner	Selected Responsibilities
Ubiquity Repositories	*Implementation* (primary)
	• Software hosting and configuration
	• Development
	• Vendor-side project management
	• Hyku expert
Scholarly Communication and Publishing Services Librarian	• Design & feature validation
	• Policy & practice decision-maker
	• Functional requirements for Pacific University
	• Institutional repository & scholarly communications expert
	• Migration scope
Systems & Applications Librarian	• Library-side project management
	• Data extraction & validation
	• Quality assurance & testing
	• Implementation (secondary)
	• Design & feature validation (secondary)
	• Policy & practice decision-maker (secondary)
	• Functional requirements for Pacific University (secondary)
Dean of the Libraries	• Contract negotiation
	• Review of site content/copy
	• Final legal and policy decisions (subject to administrative approval)
Pacific University Information Services (Information Technology)	• Domain name management

Figure 1. Migration Roles.

from 0 – 40 hours per week, beginning with 1-hour meetings every other week and growing to 3 two-hour meetings a week, with additional work outside the meetings.

Data Extraction

The first step of the migration was the identification and extraction of data from the existing system (see Figure 2). As with most repositories, bepress has multiple sets of data which work together to build the repository experience. These included works, the files and metadata submitted by depositors, collections (which contain works), collection-level metadata,

Type	Description	Status
COLLECTION	Organizational framework for the repository consisting of metadata and associated works	Migrated, with the exception of out of scope collections (i.e. journals migrated to another host)
Collection metadata	Information (title, description) about the collection	Migrated
Collection page	Text, code, or images from the collection landing page	Migrated, original links redirected
Collection visibility	Whether the collection is viewable by all users, institutional users, or only specific users	Migrated
Collection/navigation hierarchy	The relationship of the collection to other collections (i.e. Biology is part of the College of Arts & Sciences)	Extracted & revised
WORK	A digital object and the metadata describing the object	Migrated, with the exception of works in out of scope collections and draft or unapproved works
Work metadata (public)	Information (title, description) about the work which is visible to all users	Migrated, with the exception of bepress-specific data fields
Work metadata (private)	Information about the work which is not visible to all users (i.e. author's address, review notes)	Not migrated
Work visibility	Whether the work is viewable by all users, institutional users, or only specific users	Migrated

Figure 2. Data Extracted from the Institutional Repository. **(Figure continues on following page)**

Type	Description	Status
Work workflow status	Where the work is in the submission process (i.e. draft, submitted, approved)	Extracted, only submitted works migrated
Work embargo status	If placed under an embargo, the embargo date and the work's current visibility	Migrated
Collection membership	The collection the work belongs to	Migrated
Work file(s)	The digital object(s) or file(s) for the work	Migrated
File metadata	Information about the file like the file title and description	Partially migrated (title only)
File order	If there are more than 1 files, the order the files display	Migrated
File visibility	Whether the work is viewable by all users, institutional users, or only specific users	Migrated
USER DATA	User accounts and information about the repository's registered users	Not migrated
User metadata	Information about the user (i.e. username)	Not migrated
User contact information	Contact information for the user	Not migrated
User permissions	Privileges available to a user (ability to deposit works, approve works, etc.)	Not migrated
Relationships to works	Ownership of specific works	Not migrated
USAGE DATA	Information about how the institutional repository is used	Extracted, partially migrated
Views of works	Number of times a work has been viewed	Migrated
Downloads of works	Number of times works or files have been downloaded	Migrated
Viewer demographic information	Information about the repository websites' users	Not migrated
Other page views	Information about the views for other pages on the repository website	Not migrated

Figure 2. (continued)

navigational hierarchy information, user data (including relationships to works, contact information, and permissions levels), and usage data (or the views and downloads of works).

Migration may present an opportunity for data cleanup, if approached strategically. In our case, the majority of institutional repository users deposit once and do not use their accounts again, and because it was not a known workflow for Ubiquity, the migration team did not migrate user accounts. Over time, the institutional repository had amassed a variety of content types and collections including journals, books, conference schedules and programs, theses and dissertations, and other assorted works by faculty, staff, students, and departments within Pacific University. Some collections and their works, like the active journals, were migrated to other systems or institutions, or removed if they had no works. Our migration scope focused on the remaining works and collections and their usage data.

There were several options for data export from bepress including a CSV file for visible work metadata and usage statistics. They also provide an Amazon Web Services (AWS) S3 archive of all published works on the platform, which is accessible at all times to customers who also purchase the Archive service, or provided to customers when they decide to leave bepress (bepress, 2020). Although these two data sources include a wealth of information, some information was only available via the Digital Commons administrative dashboard. Typically, this dashboard data might include works in progress but not yet published, comments on works within the dashboard, and other details normally hidden from public view, but useful for staff and administrators. Due to CAPTCHA and other security restrictions on the administrative platform, this information was not available via typical web scraping techniques.

WORKS

Both the CSV extract and the AWS S3 archive had advantages and disadvantages as primary sources of data for the migration. The CSV extract did not contain all the metadata available on the AWS S3 archive. The AWS S3 archive, while more complete, did not necessarily contain information about whether data or files were public in the Digital Commons platform. The S3 archive was structured into folders for each collection, with sub-folders for each work. Each work folder contains a metadata.xml file describing the work and all files associated with the work. For works with multiple files, data in the AWS S3 archive might not be sufficient to determine which files should be visible (i.e. published, final file) and which files should be hidden or deleted (i.e. drafts, superseded or unpublished). For metadata fields, like author email or address, fields could appear in the AWS S3 metadata, but be hidden to the public in the Digital Commons archive. Both of these common cases did not have explicit markers to determine which content should be visible in the migrated product.

Since it had the more complete dataset for extraction, we chose to work with the S3 archive. In addition to the metadata completeness, since our scoped works included different visibility levels, and CAPTCHA settings prevented file scraping, we needed the S3 archive to transfer files which would otherwise require a login to the Digital Commons platform. Once the data source was selected, we worked with bepress to schedule a push to the S3 archive to ensure completeness after our deposit cutoff date for the Digital Commons platform. We also requested that they adjust privileges to the S3 archive so we could access all works, including those with private or institutional visibility settings. We then coordinated with the new vendor, Ubiquity Repositories, to create an AWS S3 bucket and IAM (Identity & Access Management) privileges to ensure they had access to the metadata and files for works.

COLLECTIONS

Since CommonKnowledge has a smaller number of collections, we used a combination of harvesting sources and techniques to obtain the data. Since each collection had a corresponding S3 folder with the same three letter code as the collection URL, we used the S3 folder list as a starting point for a list of collections, which included 121 entries. We then used the Digital Commons public website to identify the visibility of the collection, the description or landing page content of the collection, and the hierarchy of the collection. Library faculty went through the list manually to identify those which were superseded, duplicate, or out of scope of the migration. We had 121 collections, and ended up with 59 collections either migrated as they were or combined with others.

USAGE DATA

We used the built in bepress tool to export work usage data. Due to the dynamic nature of usage data, we did several test exports, then a final data extraction the day before usage data was imported into the new Ubiquity Repository. For each work, the download total was extracted and summed, to be imported into the new system.

Initial Analysis & Normalization

Once we had data sets for the migration, our next task was to normalize them so that we could use the same import process for the entire data set. The work data had a high degree of variability, due to the customization of work submission processes for each collection and department at Pacific University. In collaboration with Ubiquity Repositories, we repackaged the data from the S3 archive xml files for each work and identified individual fields used to describe our works in bepress, resulting in over 130 unique fields. In some cases, due to the submission creation workflow, works from different collections had different field

names, but similar meaning. For example, volume, volnum, and volume_issue had been used to describe the volume of a journal in different collections over the history of the repository. Manual analysis of potential duplicate fields was performed by library faculty, using works with different fields and the front-end display on the Digital Commons platform to determine if fields could be merged and, if so, how. Generally, analyzed fields were treated in one of four ways. The majority of fields were not changed at all. Some single-value fields were merged, keeping the non-null value and discarding the null values, as in the case of the volume field. Multi-value fields like subject, keyword, and notes or comments had each value added to the field (i.e. category and subject were combined, keeping each value of either field). Lastly, some fields were discarded from the data set, especially those specific to bepress. For usage data and collections, the import datasets were fairly uniform in structure and fields.

Data Mapping

WORKS

With our normalized data sets, we were able to consider how to map or crosswalk the data structures in Digital Commons to those in Ubiquity Repositories. Although both platforms provide many of the same data and functionality, they have fundamentally different data structures and user workflows. Digital Commons is structured around collections, with the option of customized work submission forms for each collection (for example, undergraduate capstones). This means that works within a collection are likely to have similar metadata. In contrast, Ubiquity Repositories is structured around work types (for example, scholarly articles), which means that comparable works are likely to have similar metadata. The submission form for a work is based on the work type, with different metadata fields available by work type. For CommonKnowledge, this transition presented a challenge as we had several collections with mixed work types (i.e. the Optometry collection contained submissions from faculty which could be articles, handbooks, and other formats at their discretion). In addition, although we had over a hundred collections in Digital Commons, we knew that the new repository would have far fewer work types. As a result, we had to re-examine the submission workflow and determine which fields were the most important for each work type.

In addition, there were four other key implementation differences. First, Ubiquity Repositories and Hyku use the same metadata fields for the submission form, public web page, and administrative views of the work. Unlike Digital Commons, there are no fields available for administrators only. Private metadata fields cannot be used to describe a work, nor can the platform restrict access to special metadata fields to administrators at this time. As a result,

all data we wanted to migrate had to be mapped to a field in the submission form for its work type. We also had to decide whether to keep any private data, with the understanding that it would be available to users in the submission form and could display on the public interface, depending on the work's overall visibility setting.

With the assistance of Ubiquity Repositories, we created ten (10) work types, extending the two basic Hyku work types. These include:

- Article
- Book
- Image
- Thesis or dissertation
- News clipping
- Book chapter
- Media
- Presentation
- Uncategorized
- Text work

Most metadata fields applied to every work, including

- Title
- Creator name
- Creator name type
- Contributor
- Resource type
- Institution
- Abstract
- Date published
- Department
- Additional links
- Language
- License

- Rights statement
- DOI
- Keyword
- Subject
- Additional information

However, some fields like ISSN, ISBN, and duration were only available for articles, books, and media respectively.

Secondly, visibility options varied slightly between the platforms in visibility granularity, authentication methods, and overall definitions (see Figure 3). Generally, bepress works, as configured in CommonKnowledge, had public metadata upon approval, with varying visibility for the files. In contrast, Hyku access control generally applies to the full work (metadata and files) with the option to set a different visibility for specific files. This meant that embargoed works would not be discovered by site users, as the work page would not be available until the release date. In CommonKnowledge, most theses and dissertations begin as embargoed works, but may be available by special request to researchers during the embargo period. To allow this access, embargoed works in Digital Commons were usually migrated as public works in Hyku, with embargoed files inheriting the embargo date from the Digital Commons

Visibility Option	Digital Commons	Hyku
Granularity	Focused on file visibility	Includes both metadata and file visibility
Authentication Methods	IP address, or email domain, or specific user email accounts	Registered users (login)
Public Visibility	Viewable by all site visitors	Viewable by all site visitors
Institutional Visibility	Viewable by IP address, email domain (signed in users), or specific user email addresses	Viewable by all registered users
Private Visibility	Viewable only by administrators or authorized users	Viewable only by administrators or authorized users
Embargo	Viewable by a set of users after a specific date	Viewable by a set of users after a specific date, may specific starting visibility and ending visibility
Lease		Viewable by a set of users until a specific date

Figure 3. Access Control or Visibility Options.

embargo date. Likewise, Digital Commons institutionally accessible works also migrated as Hyku public works with institutionally accessible files. Selected works which had not completed the bepress submission workflow with final approval were migrated as private Hyku works, so they are visible only to Hyku administrators and their status was noted in Hyku.

Thirdly, the works' file metadata was less straightforward. Since we were using the Amazon S3 archive as a data source, we had some information about supplemental files from bepress, which were listed in each work's metadata.xml. Each supplemental file name in S3 started with a number underscore and followed with the truncated source file name. With some scripting and interpretation, the supplemental files could be identified from file structure and metadata.xml contents. However, if there were multiple non-supplemental files attached to a work, we relied on knowledge of the Digital Commons ingest process, which created a stamped.pdf file for the majority of works, to determine the main or primary file. This file would be the main file in the Digital Commons work page, available via a download button. For public works without a stamped.pdf, and multiple files without the supplemental file prefix, an automated program compared the file in Amazon S3 to the file on the Digital Commons webpage to find a match. For works in this situation with viewing restrictions, the file was manually downloaded and compared. After the primary files and supplemental files for each work were identified, their Amazon S3 location was added to a new CSV file containing the transformed and mapped repository metadata, with files in display order. The visibility of each work and file was also added to this CSV file for ingest.

Lastly, some collection information had to be encoded into the works themselves. In Digital Commons, the front-end display of collections communicated hierarchy, so that collections appeared to be nested (i.e. College of Arts & Sciences > Faculty Scholarship (CAS) > Department of Biology, see Collections for more). In contrast, Ubiquity Repository collections display with a limited number of hierarchical levels. This necessitated that all collections be rearranged, and some were also combined. This change in structure also meant that data implicit in the former collection structure (like department affiliation) had to be captured and displayed in the work forms and work metadata, instead of, or in addition to, the collection.

COLLECTIONS

Although the data for the collections was fairly similar, the transition from a hierarchical display of collections to a flat structure was a challenge. Formerly, the Collections mirrored the institutional structure of Pacific University, which is composed of the Colleges of Arts & Sciences, Business, Education, Health Professionals, and Optometry. Each College has departments and programs, which may each have their own or multiple collections. Typically,

student works and faculty or professional works were separate collections under a department or program. In addition, collections might be collaborations between departments or Colleges, or might represent multi-institution collaborations and scholarly pursuits outside of the University structure, such as conferences or journals. With the assistance of Ubiquity Repositories, we were able to customize a display of collections. Though the Hyku collections kept their flat structure, they displayed as nested on the institutional repository website in three levels. The first level includes Scholarship, Student Scholarship, Pacific University Press, and Bee Tree Books, followed by the College or Department, and the last level representing the Hyku collection. Despite this accommodation, we still had to combine several collections to fit the platform.

In addition to the display hierarchy, we had to rename all the collections for clarity and uniqueness. In contrast to Digital Commons, which starts the deposit process on a collection page, and then proceeds to a collection-based process with the collection pre-selected based on the entry page, Ubiquity Repositories has a drop-down selection of collections on the work-type-based deposit page. This display does not emphasize the path of the collection to the same extent (i.e. *Faculty Scholarship > College of Arts & Sciences > Biology* versus *Biology*) so a new naming convention was formed for collections. Since we wanted to emphasize faculty versus student scholarship, most departments, schools, and Colleges were assigned two sub-collections (i.e. *Biology* and *Biology - Student Scholarship*). Due to the limited hierarchy display in Ubiquity Repositories, we decided on a single student scholarship collection. Within this collection, the new field, Resource Type, is used to distinguish between theses, dissertations, capstone projects, and other student works.

Data Ingest

While Pacific University was involved in validation of the process and of the ingested data, it was run by Ubiquity staff on Ubiquity servers. A repository metadata CSV file, created from the extracted and transformed metadata was used for Ubiquity Repository's ingest process. Collections were added to the repository first, to allow works to relate to an existing collection. To import works, Ubiquity Repository used a custom importer with OpenRefine to transfer the metadata and files into the new repository platform. Since our migration, newer versions of Hyku have been released which include built-in import processes which ingest data from OAI-PMH, CSV, Bagit, or XML data. Ubiquity Repositories used a separate import process transferred the usage data to the new repository, after a complete and valid migration of the work data. For each work, the existing usage data was imported into a database and mapped to the work's identifier number in Hyku.

Data Testing & Validation

After the ingest, the data migration was tested by comparing corresponding Ubiquity Repository and Digital Commons pages. Rather than testing the display of each work, our test plan selected a set of representative works for each work type, migrated collection, merged field, new field, and file visibility case. For each representative work, each metadata field and file was checked against the source data to confirm the data migration worked as intended. When migration errors were found, they were fixed in a small test import, and the data set was re-imported for further testing. After a migration passed all the initial testing, an additional three random works from each collection were compared, considering the data mapping. As a further test, volunteers from the library staff were invited to test specific collections and note any irregularities between the new and old platforms. Once the display and data mapping were validated, a final check was made to ensure that all works migrated to the new platform by comparing the number of works in collections prior to and after the migration.

LESSONS LEARNED

Although there were many lessons from our migration process, the top three were (1) to form a consensus on specifications and priorities, (2) to standardize data and data practices where possible, and (3) plan for the full scope of the project.

Specifications & Priorities

Working with a vendor on a new platform, during the development process afforded us many opportunities to clarify our priorities and the platform functionality. In many cases, as the first higher education client for Ubiquity Repositories, we advocated for features and workflows to support common practices within libraries and higher education institutions. Due to the different organizational backgrounds of Ubiquity Repositories and Pacific University Libraries, initial priorities and assumptions sometimes differed and had to be collaboratively negotiated. Both parties brought valuable perspectives to the discussion, which would have been helpful to share at the beginning, so that that knowledge base would be available to all involved throughout the project. This was especially true with beta releases and phased implementation, as the vendor collaborated with us to determine which features to prioritize for release.

For example, while Pacific University's institutional repository, CommonKnowledge, primarily works to make scholarship related to Pacific University openly available, that goal cannot always be accomplished. Like most repositories, CommonKnowledge holds some items that need a temporary, or even a permanent, embargo. This is relatively common with theses or

dissertations that students wish to publish in the future. At the time of our migration, Ubiquity's platform did not have a way to make any items private. We worked to advocate for this option, as it is a necessary part of our work, even if it is not our primary goal. Without discussions and input from both sides, it would have been difficult for Ubiquity to discover that privacy is important to Pacific University Libraries. Although the project focused on migration of our existing data, due to our priorities, it was important to us during the data mapping and normalization to ensure that the data structures in the new platform would facilitate a clear and efficient deposit workflow. Our focus on appropriate visibility for works, authors, and files added complexity and time to the migration project, but it was an essential step for our institution.

Even though our specifications and priorities evolved through negotiation and discussion, having clear priorities, agreed upon ahead of time, could facilitate a faster migration and more efficient communication during the project from design to testing to triage. A priority framework can also help expedite decisions about which issues to address first, or clarify which configuration options will best serve the repository users. Many of our priorities emerged during discussions in the development or testing phases, but may have been useful earlier in the process.

Standardization of Data & Data Practices

Another lesson from the migration is the importance of data standards or standard data practices. Although systems may support custom data, during the migration process, all customizations require unique solutions or resolutions. For example, CommonKnowledge had customized many of the submission forms for various collections, and added additional metadata fields that did not easily fall within either bepress' norms or the norms of any other schema, like Dublin Core. Each of these unique instances needed to be addressed manually. Should the information in the custom fields be retained? If so, where might it fit within Hyku's schema, Dublin Core? In this case, some of the easily customizable features bepress offers its customers proved difficult in the long run, adding time to the mapping and development portion of the project.

Following data standards may help with regular data input/output activities such as optimizing contents for search engines or exposing data to external indices, in addition to expediting future migrations. When possible, using documented data standards or vocabularies, especially those with a maintaining agency, can make the data more interoperable and semantic. When a platform is built, these standards may also cut development time, since they may be used rather than designing a custom solution.

Plan for the Full Scope

Initially, the project was described as a migration project. However, due to fundamental differences in the platform design, migration required deeper analysis of the digital objects, digital structures, and digital workflows on each platform. We have yet to fully see how the changes in workflow will impact use and perception of CommonKnowledge. In addition, as the platform features were designed and built during the migration process, specifications and decisions went through multiple iterations before they were ready for release on the live platform. This process required lengthy discussions with stakeholders in different areas, above and beyond the staff time planned for the migration.

Although this article focuses on the migration process, transitioning from one institutional repository to another involves much more than the data migration. Those planning a similar migration may want to factor in an initial needs assessment for the institutional repository, a data analysis and design phase with stakeholders, an interoperability assessment, the data migration, a training and marketing program, and much more. In addition to the technical obstacles we faced and overcame during the migration, we also had to address the human dimension of the migration, an important part of the project scope. Bepress offered us and our users an abundance of documentation, training materials, and general support, which is not a strong point of our new institutional repository and publishing platforms. As a result, much more hands-on training was needed, particularly for our journal editors, to successfully transition to the new platforms. This involved creating documentation and instructional materials, personal training sessions, and email correspondence.

In addition, users have encountered some challenges when using the platform. The top challenges may be a result of the changed interface and included: (a) forgetting to attach files when submitting a work, (b) difficulty correcting or fixing issues with submitted works, (c) account or sign-in issues, and (d) moderation/approval workflow issues. Work is ongoing in terms of addressing some of these design and interface concerns either through increased documentation and training or interface updates, and a second update is planned for the spring of 2021 to implement improvements based on user feedback.

Organizations considering a migration should think of the migration as one part of a larger process. While the migration may require specialized knowledge and expertise, it represents a fraction of the work and time involved in transitioning from one system to another. Before the migration, priority framing, stakeholder discussions, and evaluation of vendors is essential and may require support from staff with a variety of roles. Post-migration, training, support, documentation, and outreach may involve as much, or more, institutional labor as the migration itself, over a longer period of time.

NEXT STEPS

Pacific University's goals for this repository migration were simple: to continue to offer the same basic functionality on the front and back ends and to reduce costs. These goals were largely dictated by necessity and practical concerns rather than by a set of specific requirements or philosophical ideals. In the end, migrating away from a platform now owned by Elsevier was a bonus rather than a main motivator. For us, Elsevier's acquisition of bepress made it easier to move away from the platform because we could reduce our overall reliance on a single vendor. To that end, Pacific University Libraries has seen significant cost savings, and is pleased to be using an open source platform with Ubiquity's support.

This migration coincided with a significant staffing changes and the COVID-19 pandemic, limiting outreach opportunities. In January of 2020, the Scholarly Communication and Publishing Services Librarian left Pacific University. Though the migration had been scheduled to be complete by that time, as often happens, it was delayed. The Systems & Applications Librarian completed the migration with significant platform functionality updates in April 2020. As of October 2020, the Scholarly Communication and Publishing Services Librarian position is still vacant, due to a temporary budget reduction in the library. During the same time period, Pacific University, like many institutions, pivoted from in person to online operations, and then to a hybrid environment in response to the pandemic This further reduced the staffing and institutional resources for outreach, both inside the Libraries and within the University at large. In a mostly digital or distanced environment, outreach presents a challenge. Email announcements, digital media, social media, and other digital tools may help, but may also be deferred or ignored by busy faculty, staff, and students adjusting to a new educational paradigm.

CONCLUSION

Despite the fact that many librarians have expressed the desire to migrate away from bepress after it was acquired by Elsevier in 2017, it appears most have not. bepress reports that it currently has over 600 customers—and growing (bepress, 2020). This is more than the 500 reported at the time of its acquisition by Elsevier (Ferguson, 2018). Although a panel of librarians expressed concern about the acquisition (Royster et al., 2018), a survey of their institutional repositories indicates that they are all still using Digital Commons. Penn Libraries is also still searching for the appropriate home for its institutional repository (Wipperman, 2019).

As budgets continue to tighten at Pacific University, and at most colleges and universities, and as we face an uncertain future, the cost savings provided by the migration have become more significant.

In many ways COVID-19 has highlighted the importance of open access and the vital role institutional repositories and library publishing programs play in making information available to all. In fact, preliminary evidence suggests that open publishing and pre-print use may have increased during the pandemic (Callaway, 2020). It has never seemed more important to use platforms that align with those values. Unfortunately, the pandemic has likely reduced the resources and capacity available for large migration projects. Time will tell whether these trends continue.

WORKS CITED

bepress. (2020). Customers. Retrieved from https://bepress.com/products/digital-commons/why-digital-commons/customers/

bepress. (2020). Getting started with bepress Archive. Retrieved from https://bepress.com/reference_guide_dc/getting-started-bepress-archive/

Branin, J. (2005). Institutional repositories. In M. A. Drake (Ed.), *Encyclopedia of Library and Information Science* (2nd ed., pp. 237–248). Boca Raton: Taylor & Francis Group. Retrieved from https://kb.osu.edu/dspace/bitstream/1811/441/1/inst_repos.pdf

Browse by Country. (2020). OpenDOAR. Retrieved August 25, 2020 from https://v2.sherpa.ac.uk/view/repository_by_country/

Callaway, E. (2020). Will the pandemic permanently alter scientific publishing?. *Nature 582*, 167–168. https://doi.org/10.1038/d41586-020-01520-4

Christine L. Ferguson (2018) Elsevier, bepress, and a Glimpse at the Future of Scholarly Communication, *Serials Review, 44*(1), 51–56. https://doi.org/10.1080/00987913.2018.1434379

Corbett, H., Ghaphery, J., Work, L, & Byrd, S. (2016). Choosing a Repository Platform: Open Source vs. Hosted Solutions. In B. B. Callicot, D. Scherer, & A. Wesolek (Eds.), *Making Institutional Repositories Work* (pp. 3-14). Purdue University Press. Retrieved from http://scholarscompass.vcu.edu/libraries_pubs/33

Crow, R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper.* SPARC.

Elsevier. (2017). Elsevier acquires bepress, a leading service provider used by academic institutions to showcase their research [Press release]. Retrieved from https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-bepress,-a-leading-service-provider-used-by-academic-institutions-to-showcase-their-research

Ferguson, C. L. (2018). Elsevier, bepress, and a Glimpse at the Future of Scholarly Communication, *Serials Review, 44*(1), 51–56. https://doi.org/10.1080/00987913.2018.1434379

Frost, H. & Gary, G. (2016). The Hydra-in-a-Box Survey on Digital Repositories. Retrieved October 13, 2020 from https://purl.stanford.edu/jk292fy8802

Frost, H. (2017). Hyku and Hyrax: How are they related and how are they different?. Lyrasis. Retrieved October 13, 2020 from https://wiki.lyrasis.org/pages/viewpage.action?pageId=85530575

Ginsparg, P., Luce, R., & Van de Sompel, H. (1999, October 21-22). First meeting of the Open Archives initiative. Open Archives. Retrieved from https://www.openarchives.org/news/ups1-press.htm

Goodyear, M., & Fyffe, R. (2006). Institutional repositories: An opportunity for CIO campus impact. *EDUCAUSE Review 41*(2), 10–11.

Johnson, R. K. (2002). Institutional Repositories: Partnering with faculty to enhance scholarly communication. *D-Lib Magazine*, *8*(11). https://doi.org/10.1045/november2002-johnson

Jones, R. E., Andrew, T., & MacColl, J. (2006). *The Institutional Repository*. Chandos Publishing.

Lang, J. (2019, January 29). Pacific University Leads Region in Research Spending Among Private Universities. Retrieved October 15, 2020 from https://www.pacificu.edu/about/media/pacific-university-leads-region-research-spending-among-private-universities

Luther, J. (2018). The Evolving Institutional Repository Landscape. Choice. Retrieved from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1071&context=scholcom

Lynch, C. A. (2003). Institutional Repositories: Essential infrastructure for scholarship in the Digital Age. *portal: Libraries and the Academy, 3*(2), pp. 327–336. https://doi.org/10.1353/pla.2003.0039

Penn Libraries. (2017). Penn Libraries to End Partnership with bepress [Press release]. Retrieved from https://beprexit.wordpress.com/official-statement/

PKP (2020). OJS Hosting Plans and Prices. PKP Publishing Services. Retrieved from https://pkpservices.sfu.ca/ojs-hosting-plans-prices/

Rogers-Urbanek, J. P. (2008). Closing the repository gap at small institutions. *portal: Libraries and the Academy, 8*(1), 91–94.

Royster, P., Weaver, R., Billings, M., Fitzsimmons, P., & Fishel, T. (2018, May 22). *Digital Commons Users Discuss the bepress Acquisition* [Panel presentation]. Library Publishing Forum. Retrieved from https://digitalcommons.unl.edu/scholcom/84/

Hyku. (2020). *FAQ*. Retrieved October 13, 2020 from https://hyku.samvera.org/faq.html

Samvera. (2020). *FAQ*. Retrieved October 13, 2020 from https://samvera.org/samvera-community-sourced-software/faq/#01

Stein, A. & Thompson, S. (2015). Taking control: identifying motivations for migrating library digital asset management systems. *D-Lib Magazine*, *21*(9/10). https://doi.org/10.1045/september2015-stein

What is digital asset management?. (2020). Retrieved from https://www.ibm.com/topics/digital-asset-management

Wikimedia Foundation. (2018). Institutional Repository. In New World Encyclopedia. Retrieved October 13, 2020, from https://www.newworldencyclopedia.org/entry/Institutional_repository

Wipperman, S. (2019, January 23). SCRIP Organizational Profiles. *Teaching, Research & Learning, Penn Libraries.* Retrieved from https://penntrl.wordpress.com/author/swipperman/