

Case Report

The Ecosystem of Repository Migration

Juliet L. Hardesty *  and Nicholas Homenda

Herman B Wells Library W501, 1320 East Tenth Street, Indiana University, Bloomington, IN 47405, USA;
nhomenda@indiana.edu

* Correspondence: jlhardes@iu.edu

Received: 19 December 2018; Accepted: 1 March 2019; Published: 5 March 2019



Abstract: Indiana University was an early adopter of the Fedora repository, developing it as a home for heterogeneous digital library content from a variety of collections with unique content models. After joining the Hydra Project, now known as Samvera, in 2012, development progressed on a variety of applications that formed the foundation for digital library services using the Fedora 4 repository. These experiences have shaped migration planning to move from Fedora 3 to Fedora 4 for this large and inclusive set of digital content. Moving to Fedora 4 is not just a repository change; it is an ecosystem shift. End user interfaces for access, management systems for collection managers, and data structures are all impacted. This article shares what Indiana University has learned about migrating to Fedora 4 to help others work through their own migration considerations. This article is also meant to inspire the Fedora repository development community to offer ways to further ease migration work, sustaining Fedora users moving forward, and inviting new Fedora users to try the software and become involved in the community.

Keywords: Fedora/Samvera migration; project planning; service management; digital library services

1. Introduction

When faced with the need to update digital repository software, the technical details of moving data from one system to another can be identified and outlined. This core activity, however, tends to have a ripple effect that encompasses not only the way the new digital repository software stores objects and metadata, but also how those objects are managed and accessed. Online services for collection management and search and discovery activities are impacted. The entire ecosystem of digital collections must be considered in order to adjust for migrating to a new repository.

More specifically for this case report, what began as a need to move from version 3 of the Fedora digital repository to Fedora 4 at Indiana University (IU) became an endeavor addressing collection management systems, online access services, legacy boutique sites associated with grant projects, and current community development efforts involving the Samvera software stack. Samvera is a major repository framework for academic and cultural heritage institutions in North America, offering advanced capabilities for digital collection preservation and access [1]. It powers research tools like Deep Blue Data at University of Michigan and Duke University's digital repository [2,3].

Although the topic of digital repository migration has been studied in the literature and is available in numerous case studies, there is a gap in examining the decision-making process used to formulate a migration strategy for the Fedora repository that affects a broad ecosystem of digital repository collections, applications, and stakeholders. In working through the full ramifications of migrating to a new repository system using a case that has extensive digital content already, this article shares what IU has learned about migrating to Fedora 4 to help others develop their own migration considerations. This is also meant to inspire the Fedora repository development community to offer

ways to further ease migration work, sustaining Fedora users moving forward and inviting new Fedora users to try the software and become involved in the community. This knowledge could be useful for repository developers looking for ways to ease the migration path, and repository managers and librarians aiming for a better understanding of what it means to migrate and what sorts of decisions are required.

2. Literature Review

Digital repository migration has been studied and reported in the literature with several commonalities throughout: customizing existing migration tools and workflows, the difficulties of metadata mapping, and the need to consider new applications for description, preservation, and access when migrating repositories. A recent study of digital repository managers surveyed the benefits and challenges of migration projects, finding that metadata normalization, skill and knowledge enhancement, and service improvements were the most widely-reported benefits. Survey responses indicated common challenges of building user trust and negotiating with stakeholders about features, workflows, and priorities associated with migration projects [4].

Most digital repository platforms offer documentation, guides, and training sessions for members of their user communities to perform migration or “upgration” operations using specific tools and workflows developed for these purposes. Islandora, for example, provides official documentation for migrating between versions of its software and suggests migration in stages or addressing individual components of the software stack [5]. Additional tools have been developed to import content migrating from different repositories, such as the Move to Islandora Kit, which formats CONTENTdm, CSV, and OAI-PMH collections into Islandora ingestion packages [6]. Deng and Dotson’s case study migrating digital collections into Islandora from various sources reports on challenges encountered with metadata mapping, especially when moving from many to fewer fields across different collections, and notes the need for human intervention using tools outside of the migration pathway to rectify metadata issues [7].

CONTENTdm appears frequently in recent literature about digital repository migration, with most studies reporting how their institutions migrated from CONTENTdm systems to open source repositories. In 2013, Gilbert and Mobley presented the methodology employed by the Lowcountry Digital Library to migrate from CONTENTdm to a Fedora 3 repository with a custom front-end integration of Drupal and Blacklight. Primary challenges included aligning local practice with open source platforms and institutionally-specific customizations available for reuse, as well as normalizing legacy metadata [8]. At the University of Utah, migration from CONTENTdm to a homegrown system was aided by custom ingestion and metadata management tools, and metadata mapping and standardization was a major component of their efforts [9]. Staff at the University of Houston built a complex network of interconnected applications to create a custom software stack featuring ArchivesSpace, Archivematica, and Hyku in order to migrate from CONTENTdm, relying on back-end and front-end applications to be developed to meet collection and user needs [10]. Similarly, staff at the University of Oregon noted metadata mapping, cleaning, and normalization as a major part of their migration work from CONTENTdm to Hydra (now Samvera) software [11]. Most recently, the Bridge2Hyku project announced a 1.0 release for its CONTENTdm metadata exporting tool, CDM Bridge [12].

Numerous resources are available for those interested in migrating to Fedora or between versions of Fedora repositories. Duraspace, the organization that manages development and support of the open-source Fedora repository software, offers specialized and generalized workshops for a broad user community, helping repository managers develop migration plans and strategize data and functionality migration. Recent workshops have focused on migrating to Fedora version 4, with Fedora 3 representing the majority of repositories from which people are migrating [13]. On the Duraspace wiki, a technical guide to “upgration” from Fedora 3 to 4 helps repository managers plan data migration and offers accounts of pilot institutions who completed migrations with Hydra (now

Samvera), Islandora, and custom front-end applications [14]. This site also offers guidance on the migration-utils pluggable migration tool to aid these endeavors [15]. Additionally, Armintor (one of Fedora 4's developers) discusses the technical aspects of migration using the FedoraMigrate gem in a recent workshop [16].

The Samvera community is heavily interested in Fedora repository migration due to the software community's reliance on Fedora as a back end for its repository front-end and administrative applications, including Avalon and Hyrax [17,18]. Migration efforts at the University of Cincinnati detail the process for upgrading Samvera repository software from Sufia 7 to Hyrax, which requires a Fedora version upgrade. This project required addressing several components of the Hyrax software stack, including individual gems and upgrading Ruby itself [19]. The Samvera MODS to RDF Working Group issued a white paper in 2018 offering recommendations for mapping MODS metadata fields to RDF predicates, which is necessary for migrating objects with descriptive metadata to make full use of the Fedora 4 repository [20,21].

Indiana University has been involved in the Samvera (formerly Hydra) community since 2012 and is actively developing applications with the Samvera community for general use and local digital library services. In addition to work on Avalon, Indiana University has developed Samvera-based repository applications for digital objects such as digitized biological specimens. Halliday and Hardesty reported on the early steps in this exploratory process in 2015, documenting the customization needed for the Sufia software to store Darwin Core metadata for the Indiana University Center for Biological Research Collections, pointing to needs for batch ingestion workflows, hierarchical organization of works and collections, and flexible descriptive metadata fields to support multiple metadata application profiles in the same repository [22]. In 2017, Hardesty updated the Samvera community on Indiana University's efforts with work completed to date on its migration from Fedora 3 to Fedora 4 and the recently-launched Pages Online service built using Samvera software [23,24]. Additional topics during that panel discussion reinforced the challenges of metadata mapping and normalization when migrating digital content in various repository systems.

3. Background

Indiana University (IU) was an early adopter of the Fedora repository, serving in 2003 as one of the initial implementation partners on the project led by the University of Virginia and Cornell University. At the time, IU's repository was developed as a home for heterogeneous digital library content from a variety of collections with unique content models, all contained in a single Fedora instance. This model also involved building a number of boutique digital project sites for these different collections in a one-to-one relationship [25]. The IU Libraries now maintain over 430 digital collections, with the majority residing in the Fedora repository. There are not, however, 430 boutique web sites. IU's online model shifted over time to offer various format-based services with web sites that allow for collection management and access.

In 2012 the Indiana University Libraries joined the Hydra Project, now known as Samvera, aiming to collaboratively develop open source repository software that interacts with the Fedora repository. As this development progressed, service managers at the Indiana University Libraries saw repository solutions emerge that would facilitate migration from legacy digital library services to new applications on the Fedora 4 repository. A group of service managers from the Library Technologies, Digital Collections Services, and Scholarly Communication departments convened in 2014 to outline a vision statement for service management of then-Hydra applications on the Fedora 4 repository. The group's report envisioned a single Fedora 4 repository with a variety of web applications serving different types of content through individual Hydra applications [26]. Content would be ingested into Fedora and managed by collection managers through a single administrative interface rather than separate interfaces for each application, maximizing usability and minimizing confusion about the content model divisions of separate digital library services.

Since then, the Hydra community has transitioned to the Samvera community and the software stack focus has shifted from custom Hydra Heads to Hyrax, an application envisioned to eventually be capable of managing all of the different kinds of digital content for a single institution. To date, Indiana University's Samvera development has included work with Northwestern University on the Avalon Media System, which has transitioned to using Fedora 4 [27]; with Indiana University Purdue University Indianapolis (IUPUI) on Pumpkin [28], a digital object page turning application using Fedora 4; and other projects still in proof of concept stage but using Fedora 4 (Imago, a project for biological specimens and Phydo, a digital preservation repository for audiovisual content). Other collections were initially stored in Fedora 3 and have already been migrated to other services (an image collection to Shared Shelf and an online journal to Open Journal Systems). The rest of IU's digital collections are in a single instance of Fedora 3 with multiple end user access sites and collection management sites. The sites cannot be modified to work with Fedora 4—the metadata for digital objects will need to be serialized as RDF statements and the data model identification is a completely new system. This means that new end user access interfaces will be required. Decisions must be made on whether to keep the access sites separated or combine them into a single end user access system. The same considerations are needed for the collection management sites: can they be combined into a single management interface or do they still require separate systems for management?

4. Examples

4.1. Archives Online

As an example of an application and service using the Fedora 3 repository, Archives Online at Indiana University provides access to archival collections and associated digitized collection content. Launched in 2007, Archives Online publishes and provides access to archival collections encoded in the Encoded Archival Description XML standard. This service currently offers over 120,000 multi-page digitized items [29]. Libraries staff digitize materials from these archival collections and deposit digitized TIFF files and accompanying spreadsheets of descriptive metadata into server drop boxes that trigger automatic derivative and checksum generation, deposit into Fedora, creation of METS files, and issuing persistent URLs (PURLs). These digital items are publicly viewable through a legacy application called METS Navigator that offers limited page turning and “search within” functionality based on underlying plain text generated by optical character recognition (OCR) during the derivative generation stage, as seen in Figure 1. Image derivatives are PDFs and JPEG access copies in fixed dimensions regardless of the original image's size. Since this service was developed to support digitized collections material, it only accepts TIFF files and cannot currently display other image file formats or process non-TIFF born-digital collections content. Moreover, Archives Online does not offer the ability to search across digitized items within a collection or among all digitized content. The Archives Online service has over 2200 finding aids encoded in EAD XML but those files are stored and managed externally from Fedora.

4.2. Image Collections Online

Another digital collection service using the same Fedora 3 repository is Image Collections Online [30]. As seen in Figure 2, Image Collections Online (ICO) was launched in 2012 as a way to explore digital image collections at Indiana University. In the years since, 20 libraries, archives, museums, and research projects have uploaded, described, and made openly available 44 collections totaling over 70,000 images to this service. ICO was developed by the IU Libraries from scratch beginning in the mid 2000's to host individual websites for image collections. These individual websites became difficult to manage, so a combined service was launched to provide a degree of collection-specific description and repository-specific information while also offering searching and browsing across all images in the service. The IU Libraries chose to allow flexibility in descriptive metadata fields and their labeling while eventually requiring new collections to contain a core set of

metadata fields at minimum. Figure 3 depicts the metadata entry form used by collection managers and administrative users. Image derivatives and metadata are stored in the same Fedora 3 repository. ICO is now a legacy application that is no longer in development, meaning that substantial improvements would need to be incorporated in a successor to this service. Moreover, the application is experiencing stability problems due to its age and technical bugs are becoming more difficult to fix.

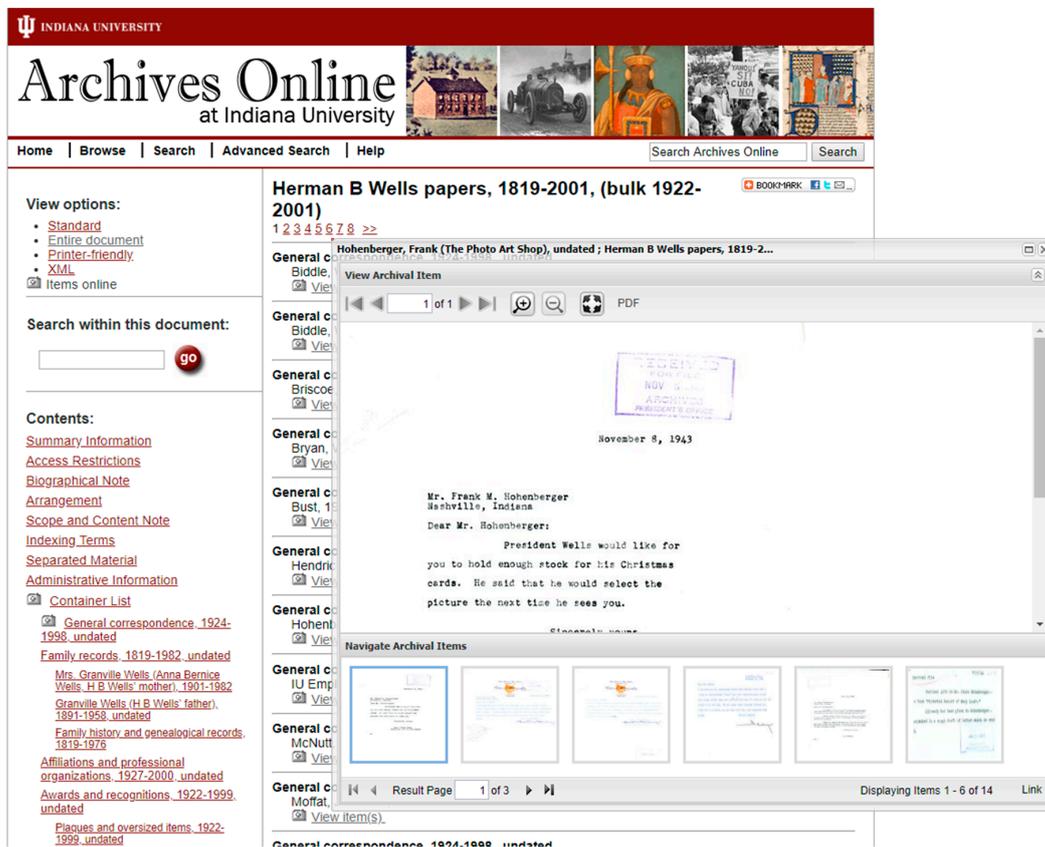


Figure 1. Screen shot of Archives Online service with viewer for digitized archival items.

4.3. Pages Online

Pages Online is a service developed collaboratively by Indiana University Bloomington and Indiana University Purdue University Indianapolis (IUPUI) for digitized multi-page items with a user-friendly administrative interface, as seen in Figure 4. It uses customized Samvera software codenamed Pumpkin, which is based on the Plum application developed by Princeton University [31]. Plum was built using then-Hydra's Curation Concerns software and the Fedora 4 repository, among other software stack components. The Pages Online service launched in 2017 as a pilot for a Samvera and Fedora 4-based digital object service and repository, and its stability, increased performance, and ease of use have been encouraging. Efforts are currently underway to rebase its code on the Hyrax software and expand its use cases to include digital image collections, which will make its set of features able to account for the majority of collections currently in Fedora 3 at IU. This service gives collection managers an administrative interface where they can upload, describe, order, structure, and select access levels for their materials, which was not possible in previous applications or services to this degree. Additionally, the public-facing interface allows for searching, faceted browsing, and the IIIF Universal Viewer to discover and interact with digital objects, as seen in Figure 5 [32]. This increased ease of use helps make the case to stakeholders for migrating to Fedora 4 with a new and improved front end application.

The screenshot shows the 'Image Collections Online' interface. At the top, there is a search bar and a navigation menu. The main heading is 'Lilly Library > The Jerry Slocum Mechanical Puzzle Collection'. Below this, there is a large image of the puzzle box and its components (a hack saw blade and a compass). To the right of the image is a metadata table:

Title	The Niagara Puzzle, (with File & Compass)
Acquisition number	000630
Puzzle category	6.2 Rolling ball puzzles
Manufacturer/Publisher	R. Journet
Place of Origin	England
Date	circa 1914
Notes	Roll all the balls from the bank into the whirlpool (hole). The hack saw blade and the compass, along with a map, were hidden inside Journet puzzles that were sent to British soldiers in German prison camps during WWI to help them escape.
Has supplemental materials	false
Location	out on display - A World of Puzzles

Below the metadata table is a link: [contact the curator](#). At the bottom of the page, there are terms of use, a bookmark link, and copyright information.

Figure 2. Screen shot of Image Collections Online service.

The screenshot shows the administrative interface for editing an item. The breadcrumb trail is 'The Jerry Slocum Mechanical Puzzle Collection > Edit Item LL-SLO-000630'. Below this is a navigation menu with options: Search, Browse, Vocabulary Sources, Batches, Export/Import, and Manage. The main content area is divided into two columns. The left column contains various metadata fields with input boxes and dropdown menus:

- Title: The Niagara Puzzle, (with File & Compass)
- Acquisition number: 000630
- Puzzle category: 6.2 Rolling ball puzzles
- Designer: (empty)
- Manufacturer/Publisher: R. Journet
- Seller: (empty)
- Place of Origin: England
- Date: 1914
- Size: (empty)
- Longest Dimension: (empty)
- Notes: Roll all the balls from the bank into the whirlpool (hole). The hack saw blade and the compass, along with a map, were hidden inside
- Instructions: (empty)
- Solutions: (empty)
- Has supplemental materials:
- Location: Out on Display
- Exhibitions: A World of Puzzles
- Public services use history: (empty)

The right column contains an 'Item Preview' section with a thumbnail of the puzzle box and a list of actions: download screen size image, download large size image, download master image, and Manage Files. Below this is a 'Transform metadata to...' section with a dropdown menu set to 'MODS'. At the bottom is a 'Modification History' section with a list of dates and times:

- Apr 25, 2013
- Apr 22, 2013
- Mar 14, 2013
- Mar 25, 2012
- Dec 8, 2011
- Dec 5, 2011

Figure 3. Screen shot of the administrative interface of Image Collections Online, with custom fields for descriptive metadata for each collection.

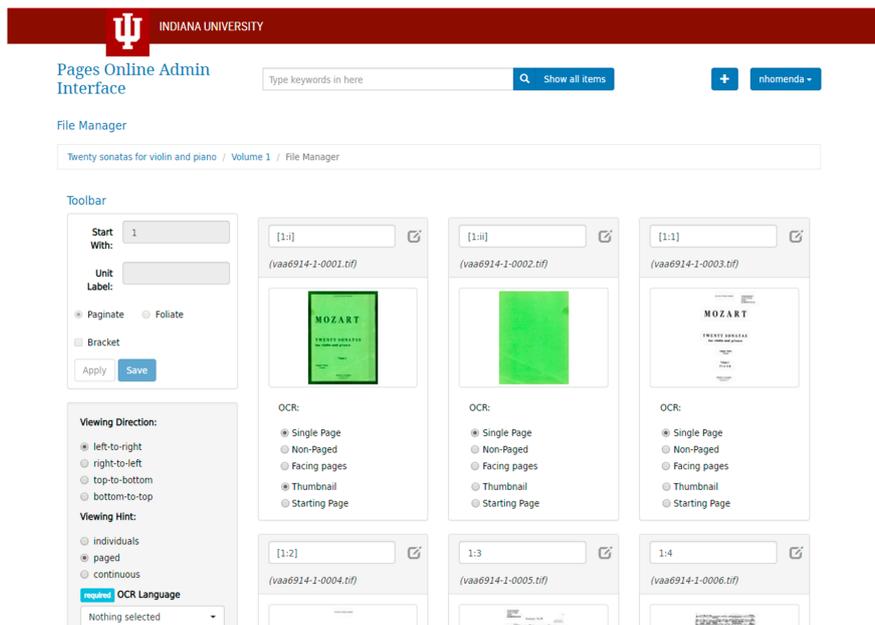
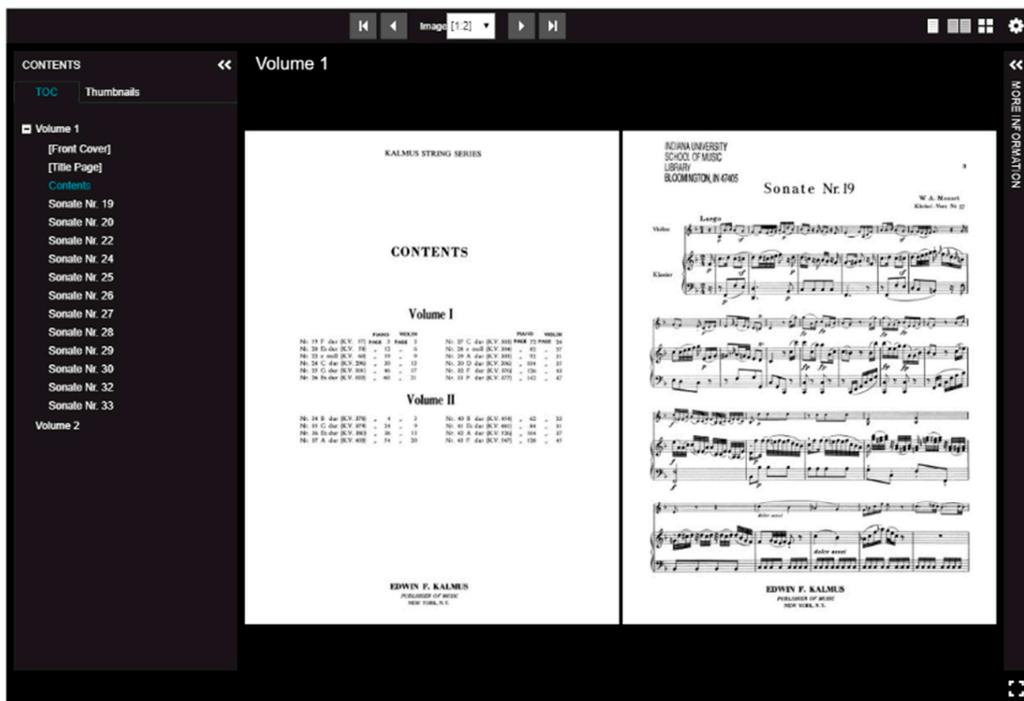


Figure 4. Indiana University’s Pages Online administrative interface, allowing collection managers to describe, paginate, order, structure, and set access levels for uploaded items.



[View record in IUCAT](#) [Open large viewer](#)

Figure 5. Item viewed in the Universal Viewer in Indiana University’s Pages Online.

5. Migration in Stages

When IU migrated a multi-page musical score collection to Pages Online, the migration occurred in stages. This collection was not previously stored in Fedora so the parameters are outside of the large migration from Fedora 3 to 4 but this migration example shows a workable methodology nonetheless. Beginning with items characterized as simplest and easiest, items were migrated in batches as much as possible. The main goal was to avoid failures based on unique characteristics, such as a non-existent library catalog record or master files in an unexpected location. Approximately 250,000 pages of

content were migrated in this way with the bulk of the migration occurring in about a week and the entire migration happening over two months. Quality control checks were done on a per batch basis since it was easier to review each batch for the peculiarities that brought that batch of items together.

Figure 6 shows a list of the increasingly intense migration stages that were encountered. This staged migration involved on one end the “smooth cases” where items had everything in an optimal state (the top of the list, least intense): structural XML was available and in the most common format (YAML); master images were available and file naming was conventional; a single library catalog record existed for the item; and the item was not overly complex or multi-volume. On the other end (the bottom of the list, most intense) were extreme edge cases that often required individual migration activity. Some items had to be reconstructed by hand, with missing pages re-digitized, structure reimaged, catalog records created or updated, and file naming normalized.

-
- a. “Smooth Cases” we have clear XML structure and master images
 - b. Accompanying materials- smaller subset with some peculiarities
 - c. Multi-volume works
 - d. Multiple library catalog records per item
 - e. No XML available
 - f. Very complex/incomplete XML available
 - g. Extreme edge cases

Figure 6. Pages Online migrated from Variations to Fedora 4 in stages.

Migrating this sheet music collection provided a test run of the concept of migration in stages on a larger scale. Migrating the entirety of all collections or even all items within a single repository instance is not a feasible approach when considering the various ways a migration impacts the access and management systems involved. Starting with a relatively simple collection within the range of collections that require migrating helps in the same way that selecting common items to migrate together in sets within a single collection moves the process forward in reliable steps. Grouping by anticipated exceptions or problems will make quality control checks and diagnosing problems easier. The success of migrating musical scores content to Pages Online provides a model for migrating other digital collections, like Archives Online and Image Collections Online, to the next generation of IU’s online digital library services featuring Fedora and Samvera software. It is possible that the content models and services of Image Collections Online and Archives Online can be managed with Pages Online, making it a potential access and management home for all of these services in the future and further easing the migration strategy for those collections.

6. Broader Migration Considerations

IU experienced changes over time with collection access and management moving from individual web applications to service models for aggregated collection management and access. In combination with developments and progress in the Fedora and Samvera open source communities, IU saw both the need to update older applications and a possible model for a Samvera-based service to manage its Fedora content.

Generally speaking, the first migration consideration might be a determination of where to migrate based on a comparison of different repository systems. For Indiana University, however, this was not the case due in large part to the depth of experience with Fedora as a repository system (since 2003), the work contributed within the Fedora open source community, and deep involvement in the Samvera community to develop an open source collection access and management solution that uses Fedora. To consider a different repository system for migration further complicates what is already a complex migration process. Additionally, Fedora 4 was a restructuring of the Fedora repository system, aligning it more with Linked Data standards like Linked Data Platform and offering ways to manage large collections of large sizes with a great deal of flexibility. There was no dissatisfaction with

Fedora 3 or Fedora as a repository system. There was the desire to continue using improved repository software for digital object management and move forward with the open source communities and their work.

Migrating a digital repository from Fedora 3 to Fedora 4 is not simply a matter of moving digital objects to a new version of the repository software. Understanding associated content and the online context (or service) through which access and management is handled impacts decisions about how this content should be stored in the repository. Due to the complexity and variety of IU's collections, running a single query or looking at a single dashboard was not enough to construct a complete inventory but understanding the overall collection landscape was important to making decisions that affect how digital object migration was handled from Fedora 3 to Fedora 4.

IU Libraries staff compiled inventories for all digital collections within the existing Fedora 3 repository as well as legacy collections that could potentially be included in the new Fedora 4 repository. Producing these inventories required a variety of methods since the applications and storage systems associated with these collections all provided slightly different views of their collection contents. Moreover, changes in content modeling, descriptive metadata, and file naming practices over the past 20 years had resulted in inconsistencies that would need to be identified and normalized. Searching using Fedora 3's default user interface returns results slowly with limited browsing and export capabilities, so SPARQL queries were used first to return Fedora-specific collection information. This inventory was combined with manifests of master files stored in IU's Scholarly Data Archive (SDA), a tape archive system for long-term storage. Finally, this information was triangulated with institutional knowledge and auxiliary collection management databases and application reports from collection managers to obtain a comprehensive inventory of digital collections suitable for migration.

Inventorying digital collections is not necessarily easy, but it is necessary. Identifying all collections in their contexts, within Fedora and outside Fedora, was required to incorporate considerations that encompassed the online services that provide access to digital objects in line to be migrated to Fedora 4. These inventory sheets are useful not only for migration but for problem diagnosis post-migration. Collection owners often refer to their items in particular ways that generally do not reference a Fedora-specific identifier and a complete inventory can help narrow down problem items to review. Based on complete inventories, content models can also be reviewed and adjusted during migration if objects are to be structured or described differently. This can be due to changes in software used with a new repository like Fedora 4.

After compiling the collections inventory, developers and service managers gathered together in what was termed a "Repository Retreat" to review and consider the best direction to take that incorporates a complete migration to Fedora 4. The discussion involved sharing all of the evidence from the collection inventories as well as establishing and confirming definitions, assumptions, principles, and goals for this migration (for example, that this is not just about moving digital objects from one version of Fedora to another—this requires rethinking the services and management models that have gathered over time). These technical considerations also needed to keep in mind a connection to higher-level strategies and policies, keeping the teaching, learning, and research mission of the digital collections and the IU Libraries in focus [33].

The main activity of the retreat was to compare and evaluate the collection information along with the access and management features considered important to the services currently provided (see Figure 7). The group agreed on the assumption that the Hyrax software stack would be the likely candidate for any new online services so the current features list was also compared to the available Hyrax features. This helped to identify what was important as well as what was possible.

Process/Tool/Feature	Pages Online	MCO	ICO	Archives Online	Hyrax 2.1
Name includes "Online"	X	X	X	X	-
Derivative creation	X	X	X	X	X
Z39.50 MARC metadata ingest	X	X	X	X	-
IIIF	X	X	X	X	-
Access controls	X	X	X	X	X
Universal Viewer	X	X	X	X	X
Dropbox to upload content via web browser	X	X	X	X	-
Can edit item structure via the web browser	X	X	X	X	X
Master File Download	X	X	X	X	X
Batch Upload (spreadsheets not included as strength)	X	X	X	X	X (but some metadata)
Scalable image server	X	X	X	X	X
Batch Download	X	X	X	X	-
Unit/Collection-level Branding	X	X	X	X	X
Items can be grouped by collections	X	X	X	X	X
Bulk actions achievable in-browser (delete, access control changes, publish, move between collections)	X	X	X	X	X
Faceted Browsing	X	X	X	X	X
Collections can be grouped by units	X	X	X	X	X
Can create ad-hoc groups	X	X	X	X	X
ADSL/DAP integration	X	X	X	X	X
CAS integration	X	X	X	X (Submit only)	X
CAS guest accounts supported for access by external users	X	X	X	X	X
IP-based access controls	X	X	X	X	X
Player is embeddable outside of the application	X	X	X	X	X
LTI integration for Canvas	X	X	X	X	X
File versioning	X	X	X	X	X

Figure 7. Screen shot of the Process/Tool/Feature comparison across current digital collection services and Hyrax software.

The comparison showed that there were common processes, tools, and features across all current services. While some custom services encompassed a more complete list of the 43 identified processes, tools, and features, Hyrax has the capability to handle 27 of these features already, or 63%, lending support to the idea of using that software for collection management and access.

7. Discussion

These experiences have shaped how Indiana University is planning our migration from Fedora 3 to Fedora 4 for this large and inclusive set of digital content. Using Fedora 4 means descriptive and structural metadata will move from XML files to properties on objects and relationships between objects. Descriptive metadata managed with complex hierarchical standards such as MODS XML will become simple statement properties. Structural metadata defining ordered pages was stored as METS XML. An analysis of the types of structures defined in METS across IU’s collections showed that these connections can be reworked as relationships connecting objects in Fedora 4 (see Figure 8).

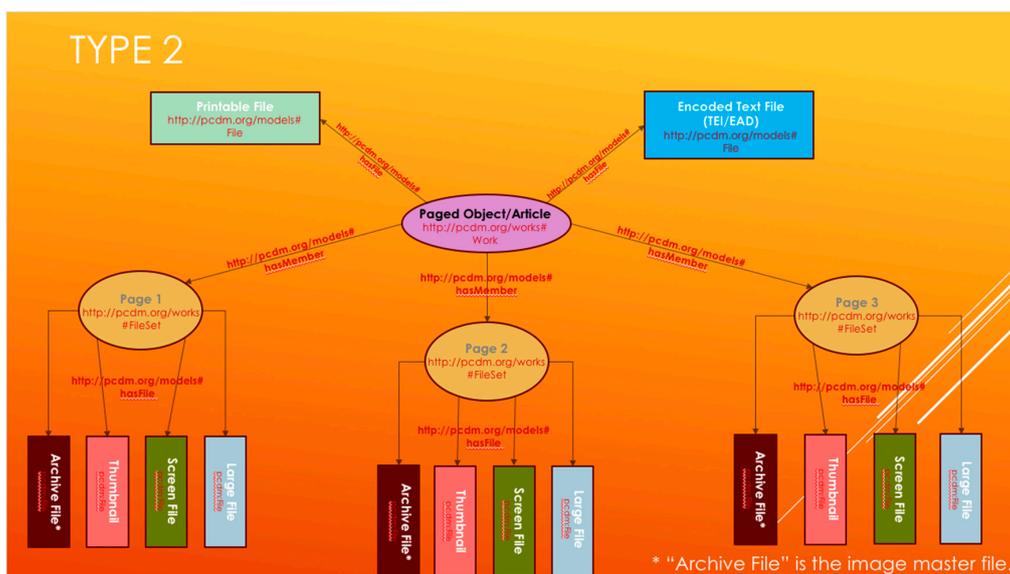


Figure 8. One of the more common types of structure defined in IU’s METS is shown mapped as relationships that can be defined in Fedora 4. Diagram created by Madison Chartier, a graduate student in the School of Informatics, Computing, and Engineering at Indiana University.

There are noted cautions around moving metadata to a new schema [34]. It is a change that inevitably ends with different metadata but the changes here also open possibilities that have so far

eluded IU's digital collection service efforts. Storing descriptive metadata as individual properties in the digital repository allows for easier updates, meaning a single property can be edited without the need to rewrite an entire XML file with every "save" action. Using RDF for these properties means Linked Data is more easily incorporated for enhanced connections and information about subjects and names and connections to other resources. Additionally, the defined structures can be more easily shared for personal research organization, such as playlists, book bags, and alternative data sets.

This new form of metadata and new way of structuring object relationships means that access and management systems have to change to continue working. As mentioned previously, this case report includes format-based services and older collection-specific boutique sites that offer collection management and end-user access. Using Hyrax for collection management and access will streamline these services across all collections, reducing the number of sites and the effort required to maintain and support these online services.

Additional open source efforts can enhance and further streamline end-user viewing of digital objects. The International Image Interoperability Framework (IIIF) is "a set of shared application programming interface (API) specifications for interoperable functionality in digital image repositories" [35]. The content to be migrated from Fedora 3 are digitized two-dimensional objects—photographs and documents (single and multi-page). The entirety of IU's digital content also includes time-based media and even 3D digital objects. Using an IIIF-enabled viewer means a single viewer can be capable of showing these different media types. For migration, this also means that static derivatives currently stored in Fedora 3 do not need to be migrated. Instead, new derivatives will be generated to work in an IIIF-enabled viewer. Recently, the Samvera Community has released Hyrax 2 that includes an IIIF server as well as the ability to use Universal Viewer [36].

Discussing needs, desired outcomes, and possibilities is all necessary when migrating data, regardless of how simple the migration might appear. Knowing what a choice is and is not (assumptions) is a good place to begin. Once the desired outcomes are established, learning what is possible and having the resources to test features, even in a small way, can help outline a path to reach those desired outcomes. For IU, using Fedora 4 was not a choice, it was an assumption. Using Hyrax was a desired outcome but was uncertain without feature comparison and testing software capabilities.

Migrations present opportunities and challenges to collection managers and developers. Collection managers have the opportunity to make collections more "future-proof" by normalizing content models and metadata. This is also an opportunity to re-envision collection management within new applications and add desired features that old systems lacked. There are also challenges, however, such as maintaining unsupportable or one-off features that are important from an old system as well as making the case to stakeholders who are happy with the old system functionality. For Fedora developers, migration activity could be aided by making it easier to retrieve broad views of what is in Fedora and improve the views of migration reporting from within that digital repository system.

IU sees this process as the best way to approach migrating all digitized content to Fedora 4. Collections within the Fedora 3 instance as well as external collections all need to end in Fedora 4. Along the way, the collections considered the "smoothest cases" will be migrated first with end-user and collection management access implemented using Hyrax. As the migration stages occur, access and management features might dictate whether or not multiple instances of Hyrax are needed. Ideally, the number of systems and repository instances will be significantly reduced while keeping as much feature parity as is reasonable for long-term access and management maintenance.

8. Conclusions

Repositories do not exist in a vacuum. Migrating to a repository system is not just about how to move the data. All of the systems making use of that data can be impacted as well. Moving to Fedora 4 is not just a repository change for Indiana University; it is an ecosystem shift. End user interfaces for access, management systems for collection managers, and data structures are all impacted. As IU's knowledge and experience has grown with the Samvera community, considerations for using

Fedora 4 have evolved to also encompass the services and applications being providing. Fedora 4 is necessitating a need to change online services. Additionally, there are service needs that necessitate changing repository structure and workflow.

This case report revealed IU's process for planning migration to Fedora 4. The full migration is still in progress and is not complete. The questions asked, decisions made, and the impacts of those decisions are shaping the plan and timeline for Fedora 4 migration. This process directly relates to integrating between systems, sustaining the use of Fedora, and offering a methodology that can be reproduced for others to try.

Author Contributions: Both authors contributed equally to this research.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Samvera. Available online: <http://samvera.org> (accessed on 14 December 2018).
2. Deep Blue Data. Available online: <https://deepblue.lib.umich.edu/data> (accessed on 14 December 2018).
3. Duke Digital Repository. Available online: <https://repository.duke.edu> (accessed on 14 December 2018).
4. Tripp, E. Reframing Open Source Repository Upgrades. In Proceedings of the Digital Library Federation Forum, Pittsburgh, PA, USA, 23 October 2017; Available online: <https://osf.io/36pmc/> (accessed on 14 December 2018).
5. Islandora Documentation: Manual Migration/Upgrade. Available online: <https://wiki.duraspace.org/pages/viewpage.action?pageId=69833583> (accessed on 14 December 2018).
6. Move to Islandora Kit GitHub Repository. Available online: <https://github.com/MarcusBarnes/mik> (accessed on 14 December 2018).
7. Deng, S.; Dotson, L. Metadata migration to Islandora: Is there an easy way? In Proceedings of the Metadata Interest Group Meeting, American Library Association Annual Conference 2014, Las Vegas, NV, USA, 29 June 2014; Available online: <http://stars.library.ucf.edu/ucfscholar/14/> (accessed on 14 December 2018).
8. Gilbert, H.; Mobley, T. Breaking Up with CONTENTdm: Why and How One Institution Took the Leap to Open Source. *Code4Lib Journal*. 17 April 2013. Issue 20. Available online: <http://journal.code4lib.org/articles/8327> (accessed on 14 December 2018).
9. Neatrou, A.; Myntti, J.; Brunsvik, M.; Maringanti, H.; McBride, B.; Witkowski, A. A Clean Sweep: The Tools and Processes of a Successful Metadata Migration. *J. Web Librariansh.* **2017**, *11*. [CrossRef]
10. Weidner, A.; Watkins, S.; Scott, B.; Krewer, D.; Washington, A.; Richardson, M. Outside the Box: Building a Digital Asset Management Ecosystem for Preservation and Access. *Code4Lib Journal*. 20 April 2017. Issue 36. Available online: <http://journal.code4lib.org/articles/12342> (accessed on 14 December 2018).
11. Simic, J.; Sato, L.; Wick, R.; Mellinger, M. *Case Study: Oregon Digital Migration from CONTENTdm to Hydra for Digital Collections*; Hydra Connect: Boston, MA, USA, 5 October 2016; Available online: <https://docs.google.com/presentation/d/1nyg2-1n2vLMhH2zF9yonly9T23CM78mvPkNL2LoCONc/edit#slide=id.p4> (accessed on 14 December 2018).
12. Bridge2Hyku: An Early Present for CONTENTdm Users. Available online: <https://bridge2hyku.github.io/updates/early-present/> (accessed on 14 December 2018).
13. Armintor, B. *Migration to F4: Concepts and Planning*; Fedora Camp NYC: New York, NY, USA, 28–30 November 2016; Available online: <https://docs.google.com/presentation/d/17OKV00CxaIzUBU-QnIJ101h7beCEHdx9Cy7cZGFAAok/edit?usp=sharing> (accessed on 14 December 2018).
14. Fedora 3 to 4 Upgration. Available online: <https://wiki.duraspace.org/display/FF/Fedora+3+to+4+Upgration> (accessed on 14 December 2018).
15. Migration Utilities GitHub Repository. Available online: <https://github.com/fcrepo4-exts/migration-utils> (accessed on 14 December 2018).
16. Armintor, B.; Woods, A. Migrating from Fedora 3 to 4: Now with More Hydra. In Proceedings of the Code4Lib Conference, Philadelphia, PA, USA, 7–10 March 2016; Available online: <http://2016.code4lib.org/slides/Workshop-Fedora4-Hydra-Migration.pdf> (accessed on 14 December 2018).

17. Avalon Media System. Available online: <http://www.avalonmediasystem.org/> (accessed on 14 December 2018).
18. Hyrax. Available online: <http://hyr.ax/> (accessed on 14 December 2018).
19. Horton, G.; Scherz, T. *Our Journey from Sufia 7 to Hyrax*; Samvera Connect: Evanston, IL, USA, 9 October 2017; Available online: <https://drive.google.com/file/d/12IDsGLYahbZ0tRxvhy1okb4167frQycK/view> (accessed on 14 December 2018).
20. Samvera MODS and RDF Descriptive Metadata Subgroup. Available online: <https://wiki.duraspace.org/display/samvera/MODS+and+RDF+Descriptive+Metadata+Subgroup> (accessed on 14 December 2018).
21. Samvera MODS to RDF Working Group. MODS to RDF Mapping Recommendations v.0.1. May 2018. Available online: <https://goo.gl/SGCfev> (accessed on 14 December 2018).
22. Halliday, J.; Hardesty, J.L. Making the SHiFt: Using Sufia with Hydra/Fedora for collection management and access. In Proceedings of the Open Repositories, Indianapolis, IN, USA, 11 June 2015; Available online: http://program.or2015.net/halliday-making_the_shift-99_b.pptx (accessed on 14 December 2018).
23. Pages Online. Available online: <https://pages.dlib.indiana.edu> (accessed on 14 December 2018).
24. Gum, J.; Collier, A.; Ballinger, L.; van Tuyl, S.; Friesen, J.; Hardesty, J.L. *So You Want to Migrate From DSpace/CONTENTdm/Digital Commons/Other to Samvera?* Samvera Connect: Evanston, IL, USA, 9 October 2017; Available online: https://drive.google.com/open?id=1cMvdMZg_vKFxp4L3KptRjl37XUtHEOA (accessed on 14 December 2018).
25. See the Charles W. Cushman Photograph Collection as an Example of an Early Boutique Digital Collection. Available online: <http://www.dlib.indiana.edu/collections/cushman> (accessed on 14 December 2018).
26. Dalmau, M.; Hardesty, J.L.; Homenda, N.; Konkiel, S.; Laherty, J. Vision Statement for the Indiana University Bloomington Libraries' Repository and Related Services. 28 March 2014. Available online: <https://docs.google.com/document/d/1UypJHyEK1u1p8uXI9Nry58k6W2vZMw6khSQ2qjgYwj0/edit?usp=sharing> (accessed on 14 December 2018).
27. Avalon Media System: Avalon Features—Release 6. IU Service is Media Collections Online (<http://media.dlib.indiana.edu/>). Available online: <http://www.avalonmediasystem.org/avalon-features-release-6>. (accessed on 14 December 2018).
28. Pumpkin: Hydra Head GitHub repository. Available online: <https://github.com/IU-Libraries-Joint-Development/pumpkin> (accessed on 14 December 2018).
29. Archives Online at Indiana University. Available online: <http://www.dlib.indiana.edu/collections/findingaids> (accessed on 14 December 2018).
30. Image Collections Online. Available online: <http://www.dlib.indiana.edu/images> (accessed on 14 December 2018).
31. Plum: Hydra Head to Support Digitization Workflows. Available online: <https://github.com/pulibrary/plum> (accessed on 31 January 2019).
32. Universal Viewer (UV). Available online: <https://universalviewer.io> (accessed on 14 December 2018).
33. About IU Libraries. Available online: <https://libraries.indiana.edu/about-iu-libraries> (accessed on 14 December 2018).
34. Kleppmann, M. Rewriting (migrating) data into a new schema is certainly possible, but it's an expensive thing to do on a large dataset, so most databases avoid it if possible. In *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, 2017; p. 130.
35. IIF Frequently Asked Questions (FAQs). Available online: <https://iif.io/community/faq/> (accessed on 14 December 2018).
36. Hyrax Management Guide: Image Server. Available online: <https://github.com/samvera/hyrax/wiki/Hyrax-Management-Guide#image-server> (accessed on 14 December 2018).

