

After Fedora: Linked Data and Ethical Design in the Digital Library

KATE DOHE

Introduction

A common application of linked data technology within the library community is within digital library projects, many of which are deep into their second decade. For nearly as long, practitioners have raised implementation concerns about linked data in digital projects:

- that transforming and maintaining linked data requires expensive programming expertise,
- that the application stack is complex and fragile with many interdependencies, and
- that the maintenance communities are often made up of only a handful of qualified volunteers.

Such technical issues present very real ethical dilemmas for digital library practitioners – is the cost of implementing linked data systems so high as to be inaccessible to all but the wealthiest organizations? Do the decision makers in these open source projects accurately reflect the needs of a variety of libraries and cultural heritage organizations? Is the meticulous nature of designing around linked data worth the inevitable slowdowns in making digital content accessible? Is the level of effort of large-scale migration to linked data and maintenance over time actually sustainable in cash-strapped academic libraries? On balance, can

these applications meet the needs of users as they evolve over time? Are these projects and their user communities truly “open” and egalitarian? These questions took on new urgency in the digital library community in 2015, when the newly-released Fedora 4 repository application implemented native linked data architecture,¹ and initiated a sea change in the digital library application landscape.

Any conversations about the practical applications of linked data are inevitably shaped by the design, features, and functionality of the systems that store and serve that data to end users. Many aspects of digital librarianship and its technologies are not widely understood within the profession at large, and as a result practitioners rely heavily on community-driven professional networks for information sharing, open-source application development, and informal support and consulting. In particular, this chapter will explore digital repository management and the small, semi-formal practitioner communities built around specific systems, and how implementation decisions about open source technical projects might be made by a handful of individuals. Repository software is frequently difficult to explain from both a usability and infrastructure perspective, especially when those systems are rarely seen or directly accessed by end users. The complexity and breadth of our professional mission—to make as much of our most unique and valuable content discoverable and available to anyone in the world, preserve it for future generations, and deliver an inclusive end user experience for both novice browsers and expert researchers—means our digital infrastructure has vastly expanded to meet our requirements. However, the nuances of the digital library ecosystem requires both individual and institutional participants to possess a level of technical expertise that often functions in an exclusionary way, limiting meaningful

1 Cramer, Tom. “Fedora Update at CNI 2013 Fall Meeting.” 12/10/2013. <https://www.slideshare.net/Tom-Cramer/2013-12-cni-fedora-updatebuilding-community-building-software>.

participation within these communities to only those with time, resources, training, and experience.

Systems and application design is itself an expression of values by the people and organizations who build and maintain these products,² and consequently, the choices and practices of those communities directly influence the creators and consumers of linked data. This chapter will explore the landscape of linked data applications in digital libraries, with particular focus on the Fedora Commons community and related projects after the move to linked data. The project is more than middleware repository software—it's a community with its own values and relationships and ethics and, ultimately, it will be those ethics that shape the project's future long after the shift to Linked Data Platform (LDP) specification³ shook its foundations. More urgent for the digital library community, and Fedora in particular, is the extent to which its community seeks to expand beyond only the wealthiest research libraries and institutes. "When maintenance work is enacted without critical interrogation of the power structures within which its objects are embedded, that work can lead to the maintenance of harmful systems such as patriarchy or white supremacy," the authors of "Information Maintenance as a Practice of Care" explain,⁴ and embracing an inclusive, feminist, critical perspective to complement technical expertise will be vital to the survival of digital repository projects in the long term.

Fedora, A Personal Story

It wasn't working.

2. Sasha Costanza-Chock, "Design Values: Hard-Coding Liberation?" in *Design Justice*, 2020, <https://design-justice.pubpub.org/pub/3h2zq86d/release/1>.

3. Andrew Woods, "Linked Data Platform - Fedora 4.7.5 Documentation - LYRASIS Wiki," Wiki. LYRASIS Wiki., February 16, 2018, <https://wiki.lyrasis.org/display/FEDORA475/Linked+Data+Platform>.

4. "Information Maintenance as a Practice of Care: An Invitation to Reflect and Share | Maintainers," 2019, accessed December 20, 2021, <https://themaintainers.org/information-maintenance-as-a-practice-of-care-an-invitation-to-reflect-and-share/>.

In early 2017, I stared at a whiteboard in one of my library's nondescript conference rooms, looking at the options identified by our loose band of developers and librarians as they worked on our first digital collection destined for our new Fedora 4 repository. I was less than a year into my role as department head for Digital Initiatives at the University of Maryland Libraries, still getting my feet under me and my head around our portfolio, and pushing hard to deliver on the promise of our brand new technology stack to get our digitized content accessible at scale with our first collection of digitized student newspapers. We had a terabyte of files, and 3,500 newspaper issues dating from 1911 ready for ingest. Our team had built the first version of our command-line interface (CLI) loader, defined the content models, and run our initial test loads. Then we began to encounter significant performance problems when the rubber met the road and we attempted to load the whole batch.⁵ I expected this, given that we were among the first universities to put Fedora 4 into production, and I knew from experience that life on the bleeding edge inevitably involves some bumps. What I didn't expect, at the bottom of the whiteboard, was the half-joking final option to "Put the collection in DSpace," in reference to our venerable institutional repository.

We didn't throw our hands up and load the collection in DSpace, which is a great solution for published products but often a poor one for complex digital collections. I am fortunate enough to work at an institution with the resources and talent to tackle hard problems, and in what felt to me like an 11th hour save, my team came up with a solution that allowed us to load the content by temporarily turning off indexing during content ingest. I came away from the experience with a more skeptical view of the Fedora platform and its implementation of the LDP specification,

5. UMD Libraries, Digital Systems & Stewardship Division, "Stew of the Month: February 2017," Digistew (blog), March 13, 2017, <https://dssumd.wordpress.com/2017/03/13/stew-of-the-month-february-2017/>.

but a continued commitment to the project. After all, I reasoned, we were early adopters, and this was to be expected as part of the learning process.

Five years later, my institution remains one of fewer than twenty-five self-reported institutions that went into production with Fedora 4 for a major digital library initiative.⁶ While this number is certainly an undercount of Fedora 4-based projects, it is an illustrative indicator of a broader challenge encountered in the digital library community in recent years. For many institutions, Fedora 4 wasn't working. Other established digital repository programs also encountered considerable complexities when attempting to move forward with digital library migration projects. According to an IMLS grant statement of need by Duraspace in 2019,⁷ nearly two thirds of the Fedora community (estimated at 240 institutions, the overwhelming majority of which are academic institutions and research libraries) were still using older versions of Fedora four years after its release. Fedora 4, as the first version of the Fedora digital repository platform to implement linked data, initiated a rift in the digital library community. Application communities split, debates erupted in the community over the validity of the project, and many of my digital librarian peers and I found ourselves on what Ruth Tillman termed "The Repository Ouroboros"⁸—eternally planning a migration that never quite gets behind us. Over dinners, in the back rooms at conferences, during breakout sessions—the places where the Fedora community is a *community*—we engaged in sustained conversations about what we ask of our software projects, what we are obligated to deliver, and whether the technologies we

6. "Fedora 4 Deployments - Fedora Repository," DuraSpace Wiki, June 11, 2018., [https://wiki.duraspace.org/display/FF/Fedora 4 Deployments](https://wiki.duraspace.org/display/FF/Fedora+4+Deployments).

7. LYRASIS, "Fedora Migration Paths and Tools: A Pilot Project," Institute of Museum and Library Services, 2020, <http://www.imls.gov/grants/awarded/lg-246264-ols-20>.

8. Ruth Kitchin Tillman, "Repository Ouroboros." Ruth Kitchin Tillman, July 30, 2019, <https://ruthtillman.com/post/repository-ouroboros/>.

make and champion are fulfilling our ethical standards. In short, whether they're *working*, both for us and our users.

Many of the events and decisions outlined in this chapter are described in publicly available documentation and contemporaneous sources, and involve a small group of committed software developers, digital repository librarians, and library administrators. This community is extremely technical and frequently reflects the resources and needs of North American, predominantly white research institutions. Often, the work is poorly understood and intimidating to other librarians, but the values and choices of the digital library community have outsized impacts on the accessibility, usability, durability, and discoverability of our digital cultural heritage materials. Over the course of this chapter, I hope to shed light on what happened, why, and what an ethically-considered path forward for the digital library community might be. I am myself an outside observer on the events of this chapter; while my institution's experience with Fedora 4 informs my own perspective, and UMD Libraries have a long history of engagement with the Fedora community, I was not a formal member of the project's technical or leadership groups at the time (in other words, I was not in the room where it happened, but I was loitering in the hallway).

Fedora, The Product

Running Fedora can be a complicated business. Explaining the entire ecosystem of Fedora, its architecture, and its full history is outside the scope of this chapter. Instead, this chapter focuses on the intersection of product decisions related to linked data and the impact those choices have on Fedora adopters and the digital library community at large. For those outside this practitioner community, the differences between repository architectures are obscure and highly technical, but they have significant impacts on the ability of a given institution to preserve and facilitate access to their digital materials.

Some digital repository solutions (such as Veridian, ContentDM, and Tind, among others), are commercial, closed-source, all-in-one applications, using either the Software-as-a-Service business model or licensed and self-hosted. However, most digital library products have their roots in the open source software community, and numerous commercial services implement open standards and software as part of their service offerings, like services such as Tind that implement the open International Image Interoperability Framework (IIIF) specification. Openness, whether in source code or digital content, is commonly understood as a state of egalitarian access—a removal of the financial, social, and political barriers to knowledge and tools that any person may consume, use, and transform. The preamble to the 2007 GNU GENERAL PUBLIC LICENSE,⁹ one of the most widely used open source code licenses, references “freedom” eight times in as many paragraphs, positioning open source code as a type of natural right. Within this framework, code is more than a set of instructions for a computer; it is a form of communication—one that transmits and manifests values and ideas from programmers to users. Fedora is one of these open source projects.

Fedora is not an all-in-one or “turnkey” digital library application like other open source projects such as Dspace,¹⁰ which, by default, provides adopters with a public user interface for create/read/update/delete functions, a database, and administrative services to manage metadata and binaries.¹¹ Rather, Fedora is a middleware repository solution designed to manage large or highly complex digital content libraries, without its own public interface. In order to fulfill organizational goals of presenting content via user interfaces, institutions that adopt Fedora for this purpose need to also make choices about the public and staff application layers. This approach facilitates flexibility (the F in

9. “The GNU General Public License v3.0 - GNU Project - Free Software Foundation,” n.d., accessed November 23, 2021, <https://www.gnu.org/licenses/gpl-3.0.html>.

10. <https://dspace.lyrasis.org>.

11. A binary in this context is a non-text computer file like an image, PDF, or audio recording.

the old FEDORA acronym) but shifts the development and maintenance burden increasingly to personnel at local institutions.

As a result, the “Fedora ecosystem” can be categorized by a handful of prominent application choices:

- Islandora,¹² which uses Drupal as the foundation for the public and administrative interfaces and provides a turnkey digital library solution for adopters.
- Samvera¹³ (formerly Hydra), which is a set of community-developed Ruby on Rails gems that can create, update, delete, and display data housed in a Fedora repository. Some fully packaged applications do exist that are built atop Samvera, although their use cases are somewhat specific (Avalon¹⁴ for audiovisual materials, Hyku¹⁵ primarily for institutional repositories).
- Custom applications, which may pick and choose services such as search and discovery interfaces, image browsers and servers, or preservation management tools based on local needs and resources available. This group encompasses a high percentage of Fedora’s international users.

These components of the Fedora portfolio are not mutually exclusive. For example, at UMD we run Avalon for audiovisual materials, as well as fully customized interfaces for our primary Fedora repositories. However, increasing the complexity of the ecosystem also inherently raises annual costs for management, maintenance, and feature development. Thus, “Fedora” is a platform with a wide range of applications and services built atop it, with their own requirements and methods for interacting with Fedora repository data. Notably, while the user communities for these systems overlap with each other and have many common institutional members, developers, and leadership, few formal agreements or communication channels exist regarding

12. <https://www.islandora.ca/>.

13. <https://samvera.org/>.

14. <https://www.avalonmediasystem.org/>.

15. <https://hyku.samvera.org/>.

coordination of efforts and strategic direction. Significant changes to Fedora's design—like the move to linked data—inevitably necessitate substantial updates to applications relying on it, which in turn impacts locally-maintained custom features or functionality.

Fedora 4, The Event

At the 2014 Coalition for Networked Information (CNI) Fall Meeting, representatives from Stanford University, UCSD, and Penn State presented on the Fedora 4 project as the earliest adopters and founding members of the initiative.¹⁶ Within only five years, Stanford publicly announced in a technical analysis document¹⁷ that they had not adopted Fedora 4 and had no plans to do so. While Stanford has been uniquely transparent and detailed in their assessment, they were hardly the only university library in the community to shelve their Fedora 4 plans. What ethical and architectural issues surfaced in those five years to bring the most enthusiastic early supporters to this point? What impact did system design choices have within the digital repository community?

The initial announcement of Fedora 4 was met with enthusiasm, particularly for its embrace of the World Wide Web Consortium's (W3C) Linked Data Platform (LDP) specification.¹⁸ In order to align with this specification, Fedora's development and leadership governance groups made four key choices, spelled out in its API Specification,¹⁹ which define how Fedora must use linked data to internally manage repository resources, version

16. "Fedora 4 Early Adopters," CNI: Coalition for Networked Information (blog), December 6, 2014, <https://www.cni.org/topics/repositories/fedora-4-early-adopters>.

17. "Stanford and Fedora (Public-Facing) — 2019," n.d., Google Docs, accessed November 23, 2021, https://docs.google.com/document/d/1uFxZ5MwVY7qVEHTYrFsNHA5WbMe_1iTcmTTlxACbwQo/

18. "Linked Data Platform 1.0 | W3C Recommendation 26 February 2015." <https://www.w3.org/TR/ldp/>, accessed November 19, 2021.

19. "Fedora API Specification," n.d., accessed November 27, 2021, <https://www.w3.org/TR/fedora-api/>.

those resources, facilitate access, and issue notifications. On the face of it, these changes deliver on the promise of linked data—complex resource management at scale, with open standards, to power the next generation of digital library technology! In reality, these design choices created significant challenges in local implementations. First, the use of LDP containers for digital resource management represents a considerable increase in data model complexity from previous digital library implementations. A framework known as the Portland Common Data Model (PCDM) provides repository managers with a flexible method for representing these containers, but in practice this makes it both more challenging to migrate legacy content from older versions of Fedora (or other repository architecture that may require more custom development) and introduces new user experience challenges when representing even fairly simple digital objects.

For example, at the University of Maryland, our first Fedora project with student newspapers highlights the challenges for the end user with this approach. In our model for newspapers,²⁰ which is an implementation of the National Digital Newspaper Program (NDNP) specification, an “Issue” of a newspaper has member pages that have image files associated with them, as well as Optical Character Recognition annotation text stored as XML. In addition, the conceptual model of the “Issue” also has member “Articles,” which may appear on noncontiguous pages. While Issues and Articles are searchable by end users, as well as Annotations that belong to a given page, there are no links in the software between Articles and the Annotation text, or even Articles and pages. A common pain point for users is selecting a relevant-looking “Article” from search results, and then being deposited onto the first page of the newspaper issue without any indication where a given article may appear. This is a data-model

20. Abdul Rasheed, Mohamed Mohideen, Peter Eichman, Joshua Westgard, Kate Dohe, Xiaoyu Tai, and Ben Wallberg, “Newspapers Re-Issued: Developing a Custom IIIF Based Newspaper Viewer at the University of Maryland Libraries,” March 2017, <https://doi.org/10.13016/M21847>.

limitation with no easy solution, but a very real problem to end users all the same.

The prospects for transforming and migrating our tens of thousands of books, diaries, postcards, audiovisual materials, and other content types, many of which have custom access rules, from our considerably older Fedora 2 repository into linked data represented a daunting migration project. Moreover, the project represented a considerable institutional investment in content that, to many external eyes, was already available and safely housed in a repository. Like many American research libraries, Maryland has far more desired digital projects and content in urgent need of digitization than we have capacity and funding to execute. The pressures of releasing new content—particularly in a time period when we were pursuing more mass-digitization funding initiatives—squeezed expensive and less visible maintenance projects onto the back burner. However, such projects are vital to averting catastrophic outcomes like security exploits that threaten our digital content, users, and institutional systems; continuously kicking the can down the road has hard consequences.

Another common complaint from the Fedora community revolved around performance problems with the repository, specifically regarding its database's performance at ingest time. Fedora opted to implement a data store (the underlying persistent storage layer of the repository) called ModeShape in the production release of Fedora 4, which managed the storage of repository files. Furthermore, Fedora 4 did not include its own search and query services, choosing instead to allow for external implementations of these services. Some readers may be familiar with traditionally structured, relational databases for web applications like PostgreSQL or MySQL. To support storage and querying of linked data resources, applications can use a type of database called a triplestore—as in linked data triples—and Fedora 4 opted to support this as an external service, as well as external search index services like Solr Triplestores are newer technology

than relational databases, are not as widely utilized, and can be prone to performance problems that require specialized expertise to troubleshoot. In the community, performance problems related to these architectural decisions became known as a “many members” problem, in which ModeShape’s performance drops off steeply once there are about a thousand direct children of a node, such as a digital collection with a large number of objects or a book with many scanned page images. While this specific issue was resolved by the Fedora technical community, the problem surfaced again with the addition of shortcuts in Fedora to see all children nested under more hierarchical structures. This problem was also exacerbated by technical recommendations early in Fedora 4’s development to adhere to Semantic Web “Cool URI” principles²¹ and keep repository structures as flat as possible.²²

What do these architectural issues mean in practical terms? Loading content—particularly lots of it or more complicated content types—into Fedora 4 could encounter timeouts, slow response times, and “lost” actions, depending on the methods, services, and applications in use by a local institution. Complicated object modeling choices could complicate user experience refinements. Migration analysis from older versions of Fedora would require a significant amount of technical expertise at individual institutions, spanning developers, curatorial staff, technical librarians, metadata specialists, and other key stakeholders in a months-long focused effort.

These problems are all surmountable, and no complex application is free from implementation problems, but they do require highly specialized talent to solve:

- Software developers with deep expertise in back-end languages like Java (which Fedora is written in), *and* a

21. W3C Interest Group, Danny Ayers, and Max Völkel, “Cool URIs for the Semantic Web,” Worldwide Web Consortium (W3C), March 12, 2008, <https://www.w3.org/TR/cooloris/>.

22. From consultation with Peter Eichman, UMD Libraries Senior Software Developer and former Fedora Committer.

strong understanding of cultural heritage requirements. At the University of Maryland, we attempted to hire a contract developer to assist with this work prior to my arrival, who “spent several months attempting to learn a digital library toolkit. Ultimately the developer left before accomplishing any significant milestones.”²³

- Functional specialist librarians with experience in linked data, system dependencies, user experience and discovery analysis, repository design, metadata transformation processes, and institutional knowledge of the content and collections being migrated.
- Project and product management experts to oversee and coordinate this level of large-scale technical migration work.

This depth of talent is frequently only found at the wealthiest universities and research institutes in the country. Even at those institutions, as at UMD, a fairly wealthy research library, it is common to find one person occupying multiple roles in this process—spreading themselves very thin across intellectually demanding work and risking significant migration delays in the event of one personnel departure. It is important to note as well that, while digital initiatives often loom large in those wealthy library strategic plans, the site visitors and users of those projects are dwarfed by less-flashy core library services like reference, access, lending, maintenance, and instruction, creating cognitive dissonance in personnel who need to support all those services and organizational priorities that may not reflect on-the-ground realities. Moreover, the extremely technical nature of that talent means that the primary participants in these conversations are developers and IT managers rather than librarians (though many in the Fedora community are both). The level of expertise required

23. Kate Dohe, Babak Hamidzadeh, and Ben Wallberg, “Doing More, With More: Academic Libraries, Digital Services, and Revenue Generation,” Ithaka SR Issue Brief, January 24, 2019, p. 11, <https://doi.org/10.18665/sr.310917>.

for even accessing the conversation, let alone contributing to it or supporting implementation, functions as an exclusionary barrier to entry.

In this scenario, “Fedora” conceptually becomes a black box to outside observers who nevertheless rely on the repository, one which appears to them (rightly or wrongly) to have substantial problems. In early 2019, I provided a departmental update at a library-wide meeting on our digital collections’ efforts, and the first question from one of my colleagues after I finished my presentation was “Why is Fedora terrible?” Is the question pointed and reductive? Certainly. If the question is born from a lack of understanding, then it is imperative to contemplate why this is so difficult for even technical librarians to grasp and explain. It reflected broader frustration with the way that the complexity of our digital collections stack gets in the way of the work of our curators and digitization personnel, frustrations I’d heard echoed at conferences and in conversations with my peers at other institutions.

Fedora, Fractured

As described earlier, much of the broad Fedora user community can be further bifurcated by the two primary application communities that rely on the Fedora repository for metadata and binary management services—Hydra (at the time) and Islandora. Fedora 4’s launch coincided with a series of additional technical changes in both communities, which further complicated the digital library landscape.

The version of Islandora available at the time of Fedora 4’s release used Drupal 7 for its public and administrative management interfaces. Drupal 7 had been plagued by multiple high-profile security exploits demanding urgent IT group attention, notably the 2014 Drupalgeddon exploit that allowed anonymous attackers to compromise administrator accounts in

Drupal,²⁴ and Islandora community members were advocating for the next version of Islandora to use the more recently released Drupal 8 instead, *in addition* to changing the repository architecture to Fedora 4. This raised the complexity of the project substantially, making it effectively a completely new application. Islandora CLAW was the name of the in-development Drupal 8/Fedora 4 project, which formally began in June 2016.²⁵ Islandora 8, the final production implementation with Drupal 8 and Fedora 5 support, was released four years later in summer of 2020.

Facing legal action, the Hydra Project changed their name to Samvera in 2017.²⁶ While this was happening, substantial project and code changes were occurring with a number of Samvera's "core" community gems (gems are code packages for Ruby on Rails developers to install and implement without needing to create application functions completely from scratch). This level of "churn" made it time-consuming and expensive for personnel at the local level to keep up with upstream changes as numerous gems and projects consolidated, then split apart, dropped functionality, introduced new functionality, upgraded underlying versions of Rails, and re-merged in a dizzying fashion.²⁷

One of those core community gems central to implementation of Samvera is ActiveFedora, which enables creation and management of objects within a Fedora repository. ActiveFedora is itself modeled on a widely used and foundational Ruby on Rails method called ActiveRecord, which is designed for interaction

24. "Drupal Vulnerability: Mass Scans & Targeted Exploitation | Volexity," accessed May 1, 2022, <https://www.volexity.com/blog/2014/10/16/drupal-vulnerability-mass-scans-targeted-exploitation/>.

25. "Introduction to Islandora CLAW - Islandora CLAW," n.d., accessed November 28, 2021, <https://clawkim.readthedocs.io/en/latest/user-documentation/intro-to-claw/>.

26. Christopher Awre and Richard Green, "From Hydra to Samvera: An Open Source Community Journey," *Insights* 30, no. 3 (2017): 82–88, <https://doi.org/10.1629/uksg.383>.

27. Jonathan Rochkind, "Performance on a Many-Membered Sufia/Hyrax Show Page," Bibliographic Wilderness (blog), October 19, 2017, <https://bibwild.wordpress.com/2017/10/19/performance-on-a-many-membered-sufiahyrax-show-page/>.

with traditional relational databases. Over the course of 2016 and 2017, it became apparent in test deployments at numerous institutions²⁸ that ActiveFedora exhibited considerable performance problems due to the complex modeling changes required by Fedora 4 and its linked data implementation, which substantially complicated maintenance, migration, and upgrade choices at institutions that had heavily invested in Samvera. The change in modeling turned simple steps for item creation or updating into a long sequence of tasks. An institutional investment in Samvera almost inevitably requires expensive institutional investment in Ruby on Rails developers, therefore restricting community participation and decision-making to the research libraries and cultural heritage organizations that can afford such talent. In response to these technical pressures, a handful of institutions led by Princeton University developed a new gem called Valkyrie,²⁹ which crucially allowed technical teams to choose from a variety of backend database options, of which Fedora 4 was only one of several. In practical terms, this meant it was possible for developers and institutions to run a Samvera-based digital library application *without the Fedora repository underpinning it*, thus backing away from a commitment to linked data and, in many ways, the philosophical concept of repositories in favor of straightforward Rails web applications.

Even the Valkyrie project leaders understood the implications these options would have on the digital library community at large. “Using Fedora has been *the* defining aspect of the Samvera community, with Samvera defined in relation to Fedora,” wrote Esme Cowles, Assistant Director of Information Technology at Princeton University Library, in his remarks at Open Repositories in 2018. “As using different backends becomes a core part of the Samvera stack, it means we should rethink Samvera, and

28. Adam Wead, “Benchmark Testing with Valkyrie and Hyrax,” @awead, August 29, 2017, <https://awead.github.io/fedora-tests>.

29. “Princeton University Library Systems by Pulibrary,” n.d., accessed November 27, 2021, <https://pulibrary.github.io/2017-07-06-valkyrie>.

how it relates to other technologies. And like the renaming from Hydra to Samvera last year, it gives us an opportunity to think about who we are and who we want to be.”³⁰

Valkyrie promised choices, but who has the power and depth of expertise to choose? The question of whether to build a digital library on top of a SQL database or a linked data platform may seem like a deeply technical decision for library IT departments to argue about, and in many ways it is. However, this seemingly obscure, highly localized decision had very serious impacts across the community and posed a risk to the financial stability of the Fedora project overall. Fedora’s software is open source and free for anyone to use, but Fedora (as well as DSpace, Samvera, and many other open digital library projects) is funded through membership fees by participating institutions, with several tiers of membership ranging from \$500 to \$20,000 per year. Importantly, those membership fees often function to define who may participate in governance and leadership discussions for the project, with higher membership tiers having more “seats at the table.”³¹ Furthermore, Fedora relies upon developers and technical personnel at member universities to maintain, test, and update the system’s code base and documentation; some of these contributors become formal “committers” on the project.³² As noted earlier, participation in Samvera necessitates considerable resources to hire or contract expensive software development talent, and Samvera community members disproportionately occupied higher financial tiers of membership within Fedora and participation as committers. Suddenly, their contributions could not be relied upon.

30. Esme Cowles, “Valkyrie, Reimagining the Samvera Community,” Princeton University Library Newsletter, 2018, accessed November 23, 2021, <https://library.princeton.edu/news/digital-collections/2018-06-05/valkyrie-reimagining-samvera-community>.

31. “Membership Value and Benefits,” accessed May 1, 2022, <https://www.lyrasis.org/programs/Pages/membership-value-and-benefits.aspx>.

32. “Fedora Committers - Fedora Repository - LYRISIS Wiki,” December 13, 2021, <https://wiki.lyrasis.org/display/FF/Fedora+Committers>.

If enough Samvera institutions withdrew their financial and technical support for Fedora, that certainly would have very real implications for the *entire* Fedora community, regardless of their participation in Samvera, and could potentially sink the entire community. This is a direct function of the economic privilege of those organizations, with wealthy, predominantly white institutions making technical choices and elevating use cases that may not serve cultural heritage institutions without those considerable resources in hand.³³ While this may come across as alarmist, the similarly membership-driven Digital Preservation Network collapsed in late 2018, partly due to a large simultaneous exodus among institutional members (the University of Maryland Libraries was one of those departing members).³⁴ Indeed, between 2018 and 2021, former “Platinum” Fedora members Northwestern University, Princeton University Library, Stanford University Libraries, the University of Michigan Library, the University of New South Wales, the University of Oxford, and the University of Toronto all dropped or reduced their financial contributions to Fedora—though it must be noted that not all of these institutions are Samvera institutions, and the financial effects of the pandemic are of course also a variable in this situation.

What would the end of Fedora mean? This is a question that has kept me up on more than one evening (including now, while I write this chapter), as I contemplate the implications for my university library. We would be left with not merely one, but *two* unsupported, high volume digital collections repositories in urgent need of migration. We would likely embark on a costly research project to identify an alternative digital strategy and potentially not find any other solutions that fit our available resources, expertise, and specialized user requirements. There

33. Kate Dohe, “Linked Data, Unlinked Communities,” Lady Science, 2018, <https://www.ladyscience.com/essays/linked-data-unlinked-communities>

34. David Rosenthal, “Digital Preservation Network Is No More,” 2019, accessed December 12, 2021, <https://blog.dshr.org/2019/01/digital-preservation-network-is-no-more.html>.

are few systems that offer the long-term preservation, complex object management, scalability, and metadata functionality that digital repositories like Fedora provide—features that support our Libraries’ central mission to preserve and make accessible our unique digital content. At UMD, I could not shake the impression that the “choice” to implement a Valkyrie-based Samvera system *without* Fedora meant deprioritizing permanence and authenticity in favor of theoretically easier development and short-term content management—a decision that prompted me to reflect on my own professional values at length in discussions with my team. With any new system we chose, we would be staring down considerable tradeoffs in functionality, an increase in costs, a significant hit to staff morale, and no certain ability to make new digital collections content available in the interim. The University of Maryland Libraries is an unambiguously *wealthy* research library, and the prospective failure of the Fedora project could have been catastrophic for our own program. Smaller institutions without our staffing levels and in-house technical expertise would have found themselves with even fewer options.

And yet, could I say with confidence that our Fedora-first strategy was working for us? Our Fedora 4 repository, in and of itself, continues to be stable and perform well. Our internally created and maintained command-line tools³⁵ for adding new content in batches are fairly reliable, and while we encounter occasional quirks with the public user interface and have challenging strategic questions regarding search and discovery for end users, the overall experience for patrons and researchers is at least functional. However, in 2018, ingesting and managing content was still a cumbersome process that relied heavily on our most technical (and overextended) personnel, and we lacked a web-based administrative interface for curators and digitization specialists to create and update content. The pressure for this small team to work on releasing new content—much of which

35. <https://github.com/umd-lib/plastron>.

had been trickling through project queues for years, to the vocal frustration of users, donors, curators, and administrators—meant that migration plans kept being pushed back, especially as it seemed apparent that we would undertake our move from Fedora 2 to 4 with little community support compared to the attention paid to institutions moving from Fedora 3, or who had previously adopted one of the primary application platforms. The fractures in the Samvera community over its commitment to Fedora did not fill us with confidence in the long-term ability of the project to meet our needs as Fedora adopters. Islandora CLAW was still in development, with no clear release timeline, and UMD Libraries had only limited experience with Drupal at that time. With no community to join, we elected to build our own staff interface for Fedora 4 called Archelon.³⁶ This approach allowed us to prioritize our own articulated stakeholder needs, and I am proud of this application: proud of the locally inclusive way we built it, and certainly proud of my colleagues. Yet Archelon is also undeniably expensive, far from feature complete, and continues to have reliability and interface design issues that prevent it from being used by our curators and digitization personnel for their activities or for batch ingest.

I am uncertain whether it's working, or if it's only the best of bad choices.

Fedora's Future

2018 and 2019 were pivotal years for the Fedora community in multiple ways. As the Valkyrie project was debuting at the first international digital repository conference after its production launch, Fedora community leadership met at Open Repositories to discuss the strategic future of the product and plan for

36. <https://github.com/umd-lib/archelon>.

Fedora 6.³⁷ Notes and documents from this meeting,³⁸ and a subsequent leadership meeting in April 2019 at CNI in St. Louis, reflect the anxieties of project leaders. How can Fedora better support migrations? What is the value-add of Fedora to a digital library program? How can the community better articulate this to library professionals outside the development groups? What, fundamentally, are the values of the Fedora community, and could it be made more inclusive at both an institutional and individual level? These conversations coincided with the initial design and planning for Fedora 6, as well as a merger between Fedora's parent organization Duraspace and LYRASIS. The community's struggles to move to Fedora 4 were self-evident by this point, and the future of the repository rode on the choices and values of project developers and governance.

Fedora 6 redoubled the project's commitment to digital preservation services, with the replacement of the problematic ModeShape with the community-driven Oxford Common File Layout (OCFL) specification.³⁹ OCFL is a software-independent standard for digital repository file layouts that ensures data are stored in a transparent, human-parse-able, and predictable manner independent of any particular software, and can be reconstructed without the original software. Removing the requirement for a repository application to define and manage file layout has substantial benefits for both digital preservation and migration planning—theoretically reducing the staff resources necessary to move content between repositories *and* making the content

37. Carol Minton Morris, "NOW AVAILABLE: Fedora 5.0.0," Duraspace.org, December 20, 2018, <https://duraspace.org/now-available-fedora-5-0-0/>. Fedora 5, released in 2018, served largely as a technical alignment release. As Daniel Bernstein, the technical lead for the project, stated in all-caps, "The move from Fedora 4.x to 5.x DOES NOT represent the same scope of change."

38. "2018-06-03 - OR2018 Fedora Leadership Group Meeting - Fedora Repository - LYRASIS Wiki," n.d., accessed December 12, 2021, <https://wiki.lyrasis.org/display/FF/2018-06-03+-+OR2018+Fedora+Leadership+Group+Meeting>.

39. <https://ocfl.io/>.

repository safer in the long term. OCFL also emphasizes transparency in design and implementation—lowering some of the highest technical barriers to understanding and participating in the digital library community. That said, OCFL is still a new specification—achieving 1.0 status in 2020—and the choice to implement it in Fedora 6 represented a gamble on the success of the spec. Additionally (and critically), OCFL’s community leaders draw disproportionately on representatives from wealthy, predominantly white research institutions,⁴⁰ and does not resolve the harder and more fundamental questions about who gets to drive community infrastructure and how the barriers to participation might be lowered. With Fedora 6’s recent (as of the time of this writing) release in 2021, it’s still too early to say whether this implementation change and the community focus on easing the migration path will push Fedora forward.

What does it mean for Fedora to work? The project is more than middleware repository software—it is a community, with its own values, relationships, and ethics; ultimately, it will be those ethics that shape the project’s future, long after the shift to linked data shook its foundations. More urgent for many digital library projects and networks, and Fedora in particular, is the extent to which these communities seek to expand beyond only the wealthiest research libraries and institutes. If Fedora continues to only be a project for technical personnel at American R1s and elite private universities to engage with meaningfully, then that choice and its consequences should be made intentionally—not as a byproduct of neglect. Yet such a choice would be a disservice to the digital library community—a reinforcement of the elitism and stratification that led to Fedora being perceived as and treated like a status symbol for libraries—a complicated nuisance and not an essential piece of infrastructure. A future without robust and open

40. Andrew Hankinson, Donald Brower, Neil Jefferies, Rosalyn Metz, Julian Morley, Simeon Warner, and Andrew Woods, “The Oxford Common File Layout: A Common Approach to Digital Preservation,” *Publications* 7, no. 2 (June 2019): 39, <https://doi.org/10.3390/publications7020039>.

source digital repositories would ultimately shift the emphasis of digital libraries away from *permanence*—the idea that access to our digital materials should survive us; that our work is more than just building websites for today with no consideration of long-term maintenance; that our content is as ephemeral and fragile as the rest of the web. Digital library infrastructure is inextricable from its user and adopter communities, and for Fedora to work, the rest of us in this community must as well.