

Texas Digital Library Hyku Pilot - 2018

Final Report

July 31, 2018

Courtney C. Mumma, Deputy Director, Texas Digital Library

Overview

In the spring of 2018, the Texas Digital Library engaged with several of its members in a pilot project to evaluate the Hyku repository application for managing cultural heritage content. While some of the TDL's members intend to launch their own instances of Hyku¹, Hyrax², or another variation of the Samvera³ community's suite of offerings, others may prefer a hosted solution. Many TDL members are facing probable transitions from repository applications such as ContentDM, DSpace (both hosted by TDL and standalone) and bepress Digital Commons. With this potential migration in mind, TDL intended to explore the feasibility and community benefits of hosting Hyku. The project was inspired by the DuraSpace community's work in piloting their HykuDirect service, including their contributions to the Hydra-In-A-Box Project determining the architectural needs for a hosted service, and TDL worked with the DuraSpace team and community to ensure that our potential service offerings would align well and support each other going forward.

[Hyku](#) is a bundled solution based on components built over the past several years by the [Samvera](#) community. Hyku and its dependencies can be configured for multiple hosts⁴, allowing deployment of Hyku-based services. Its features include, but are not limited to⁵:

- [IIIF](#) (International Image Interoperability Framework) [Image](#) and [Presentation](#) APIs for all deposited works
- Mediated deposit
- Internationalization
- Configurable User Interface, including a dashboard, collapsible sidebar, and static pages
- Consolidated Administrator and User functions
- Repository service status page
- Support for [RightsStatements.org](#) values
- Support for linked open entities in the [Geonames](#) controlled vocabulary for location metadata
- [ResourceSync](#) for content synchronization by aggregation services

¹ [Hyku](#) is an open-source repository application for managing cultural heritage content.

² [Hyrax](#) is an open-source, Samvera-powered repository front-end.

³ [Samvera](#) is an open-source repository software framework which offers flexible and rich user interfaces tailored to distinct content types on top of a robust back end. Samvera is also a community creating digital asset management solutions for Libraries, Archives, Museums and others.

⁴ [Docker Compose](#)-based multi-host configuration

⁵ For more information about these features, see project documentation in the DuraSpace Wiki at [Hyku](#).

- Multiple tenant management interface

Pilot

Rationale

The main goal of this pilot was to evaluate whether some of our TDL members found the Hyku application appealing enough to justify a migration from their current repository systems, and it also aimed to estimate the feasibility and benefits of a TDL-hosted Hyku service, including estimates of the labor and resources needed for installation, technical management, service support and training, migration from other systems, service management, and outreach. Inspired by DuraSpace's HykuDirect service pilot, the TDL Hyku Pilot was intended as a lightweight, open-ended pilot to investigate Hyku hosting service for TDL. The project team followed the Hyku Pilot [Plan](#) with little variation, except that the feasibility analysis was slightly limited by the small scale of the testing materials.

Many TDL members are facing potential transitions from repository applications such as ContentDM, DSpace (both hosted by TDL and standalone) and BePress Digital Commons. A TDL-hosted Hyku service solution could assist TDL current and potential members migrating from deprecated, outdated, or unwanted repository applications, and provide widespread and sustainable access to their digital heritage content over time.

The project team included TDL members using BePress, CONTENTdm and DSpace, in order to benchmark against other repository systems to the best of our ability. Every project team member had some familiarity with Hyku, and all had at least watched one tutorial video. Additionally, the University of Houston had previously done some limited testing.

The University of Houston is interested in the potential for Hyku as a cultural heritage repository for images, and highlighted their desire to integrate the Avalon A/V system with it based on work being done towards this end at Northwestern University Library, in partnership with Indiana University Library.⁶ The University of Texas at Austin is considering Hyku as a potential replacement for their DSpace repository, Texas Scholarworks, buoyed by their ongoing DAMS project. For Trinity University, consolidating multiple content repositories is a high priority, as they currently use ContentDM, BePress and a local DAMS marketing tool. The Texas Digital Library currently hosts DSpace repositories as a service for many of its member institutions. Its interest in this pilot project is to gather information about the appeal of a Hyku hosted repository

⁶ <https://www.ims.gov/grants/awarded/lg-70-17-0042-17>

service for its members, either as an add-on service for new repositories, or as a replacement for existing repositories on other platforms.

Roles and responsibilities

- Project sponsor: Texas Digital Library:
 - Set scope and timeline
- Project manager: Courtney Mumma, TDL Deputy Director
 - Creates, executes, and controls the project plan.
- Project team:
 - Identify requirements and risks, test functionality and quality, contribute recommendations.
 - Jane Costanza, Trinity; Colleen Lyon, UT Austin; Andrew Weidner, University of Houston; Todd Crocken, University of Houston; Kristi Park, TDL Director; Courtney Mumma, TDL Services Manager; Nicholas Woodward, TDL Senior Software Engineer
- Samvera Community (including anyone working with Hyku and Hyrax)
 - Project manager will keep the community informed and request feedback and project help when needed.
- DuraSpace Staff and Community

Project manager will keep the staff and community informed and request feedback and project help when needed.

Meetings

TDL hosted three virtual meetings to bring the project team together in early 2018. The Kickoff meeting on January 23, 2018, initiated the planning process and reviewed the components of the pilot. A second meeting on March 22, 2018, outlined the testing parameters, and was followed by an April 18, 2018 meeting, to check in at about the midway point of testing. The final meeting, to review the results of the pilot and this report, occurred on May 9, 2018.

Testing

Nick Woodward set up a multi-tenant Hyku pilot platform by the end of March so that the project team could test through April. The main test development site was located at <http://tdl.hyku-dev.tdl.org/>. Woodward walked the project team through setting up repositories and user accounts. TDL's role as a SuperAdmin user could set up new repositories, but that role does not automatically set up a DNS record, which requires a software engineer. In the TDL Hyku Dev instance, each pilot project team member had their own repository where Woodward set up administrative user accounts for each participant. The pilot repositories were named as sub-sub-domains describing the short name of the repository (/tdl, /uh, /utexas, /trinity).

Woodward experienced some technical challenges along the way. First, the TDL Hyku setup was different from the standard configuration to allow for nesting, which resulted in a need to create a new record for every repository, and an inability to create an HTTPS. While the latter would be a barrier for the medium or long term should TDL choose to move ahead with a Hyku hosting service, the team decided that it would not adversely impact the pilot. During the pilot, all users got a security warning asking them whether they want to proceed, and were assured that it was safe to do so and would not prevent them from using the site. If TDL launches a Hyku service, we will register a domain that will fix this issue.

Project team members were directed to review Using Your Hyku Repository draft [Documentation](#), especially focusing on the description of the differences between administrative sets, collections and works, as well as roles and permissions.

Courtney Mumma created a testing tool for the team to use to collect testing experiences and desired functionality. Starting from the Hyku Features list, found on the Duraspace Hyku wiki, <https://wiki.duraspace.org/display/hyku/Hyku+Features>, the project team was asked to prioritize their institutional requirements based on an outline of Hyku's features, capabilities and workflows. From the results, a testing [tool](#) was created using the framework provided by a November 2017 gap assessment [document](#) from the HykuDirect Pilot Project team, which partially employed the MoSCoW method to determine what a hosted Hyku service must have, could have and should have to facilitate a service launch.

The project team tested through to the end of April, logging their findings and recommendations in the testing tool. While the team originally planned to reach out to the greater TDL community to gather feedback, we shifted gears to instead include testing from other parties representative of users from each pilot institution when possible. Ultimately, three people tested from U of H, three from TDL, a few from UT Austin, and two from Trinity. TDL users were testing for basic functionality, so their findings are not included in the testing tool.

Member tester impressions

University of Houston testers had participated in testing with the DuraSpace Hyku Pilot group and essentially their responses to the system during this pilot were the same. The lack of bulk ingest, particularly of tab delimited with metadata, is the biggest reason they would not adopt Hyku as it currently is. Additionally, they found the Hyku user interface to be insufficient in terms of status alerts. For instance, U of H reported that it was unclear whether an item uploaded successfully. U of H also found the metadata faceting problematic and wanted an interface for creating and maintaining metadata profiles. For these reasons, U of H is likely to build a customized version of Hyrax instead, although their team remains committed to improving adoption of Hyku through the Bridge2Hyku project.

The UT Austin tester found that Hyku had improved since its release during Open Repositories in 2017, impressed especially by the image carousel, homepage messaging functionality (ex. could be used for downtime alerts), featured researchers and items. A few other librarians at UT Austin also had pleasant experiences with the Hyku UI. Despite improvements, they still found the interface confusing and not intuitive, though they admitted that might be due to being accustomed to using primarily DSpace. UT Austin struggled to envision what migration from DSpace would look like with Hyku admin sets and collections in a way that would be meaningful for users used to a DSpace hierarchy. The lack of batch upload is the biggest reason UT Austin would not be able to move to Hyku yet, but missing notifications about submissions needing approval were another blight.

The Trinity tester found the overall system a bit ‘clunky’ but did like some things. As users of Bepress/Digital Commons, they had some difficulty understanding the components of collection and admin sets, including struggling with a UI that didn’t satisfactorily indicate when updated settings in those areas were changed. The testers there were pleased with the granularity of permissions, but wished for simplified usability and clarity in the interface. They were concerned that they couldn’t do a citation-only record, but found upload of files easy, liked the visibility panel on the side and the ability to explicitly change the permissions on the individual records. They found the tabs at top were confusing as well as the workflow in the submission form itself. Because of their challenges, they believed turning Hyku in its current iteration over to staff in place of BePress would require better documentation and training on workflows and simplification of the set up for administrators. Additionally, Hyku would not provide the analytics they get from BePress, which is a high priority. Two Trinity public service librarians and one other staff member did some basic testing of the system and loved the interface in general, especially preferring the website and searching over BePress, but they didn’t like the item level display at all and had many of the same issues as the main tester.

Information about progress in the community to enhance, improve interactions and add functionality to Hyku are mapped to TDL pilot members’ requirements in the table below. The HykuDirect Pilot Project Gap Assessment (HDPPGA) ranking and links to any existing issues are included, as well as any notes on known progress of said development. Where GitHub issues existed prior to the TDL pilot, statements have been added to indicate that TDL members are also interested in the development. In parentheses, (S) means ‘should have’, (M) means ‘must have’ and (C) means ‘could have’, as defined by the MoSCoW method⁷, mirroring the HykuDirect pilot’s assessment methodology.

Category	Function/Enhancement/Interaction	Details	Institution(s)	HDPPGA*	Development progress?
Import/export	(S) Install ContentDM to Hyrax	Would require access to	U of H	(M) Bulk import API View Issue on GitHub	https://github.com/UnivLibrary/cdm_migr

⁷ MoSCoW method, Wikipedia entry: https://en.wikipedia.org/wiki/MoSCoW_method

	migration tool	ContentDM API.			ator The new B2H CDM Bridge tool should bypass this problem, and the Migrator can probably access hosted CDM without any issues. https://github.com/Bridge2Hyku/cdm-bridge/releases TDL will test this as part of the October 2018-February 2019 Bridge2Hyku project
				(S) Migration tooling and support: CONTENTdm, Digital Commons (bepress), DSpace View Issue on GitHub	The University of Houston IMLS-funded project Bridge2Hyku project (LG-70-17-0217-17) is working on this.
	(M) Batch ingest; Batch import/export		UT Austin, Trinity	(M) Bulk import API View Issue on GitHub	None.
				(M) Full content export and import (per tenant and for all tenants) View Issue on GitHub	None.
				(S) Batch import from Admin UI. Allows full definition of collections and works, including all metadata fields, in a batch operation. Hyku currently offers batch creation, but metadata must be manually entered and is shared across the batch (except for the work's title) View Issue on GitHub	None.
Dashboard	(C) Notifications	If people are submitting things that need approval it would really be better if notifications came via email. We could try to remember to login everyday and check it, but that	UT Austin, Trinity	(S) Extended Usage Analytics per tenant View Issue on GitHub	Work is in progress from the Samvera Analytics Working Group .

		seems like a workflow destined to fail			
				(C) Email Notifications View Issue on GitHub	None.
Works	(C) Adding keywords	It's rather annoying to have to click on Add Another for keywords. Would be nice if commas, semicolons, or enter could trigger a new keyword	UT Austin		
	(C) Uploading from the cloud	It would be really great if uploading from UT Box (and other systems) was possible	UT Austin		
	(S) Batch editing metadata	We do a lot of metadata editing that wouldn't be possible if we had to go one by one	UT Austin	(S) Batch metadata editing View Issue on GitHub	None.
	(M) Persistent links to works		Trinity	(M) Persistent (and consistent) shareable links to works and content items. Need is for relatively short links. View Issue on GitHub	None.
				(S) Persistent Identifier integration: DOI (Handle) View Issue on GitHub	Required for migration from systems which make use of these identifiers. Many policy questions around use of IDs. Required to be competitive with other IR solutions. There is a Samvera working group on DataCite DOIs
	(S) Configurable metadata		Trinity	(M) Configurable metadata fields display per tenant: Add/update/remove metadata fields; Rename field labels; Annotate fields; Define which metadata fields are displayed in works; Allow all metadata fields to be searchable/browseable View Issue on GitHub	The current process for customizing metadata fields and forms is documented here: http://samvera.github.io/customize-metadata-model.html

				(M) Configurable metadata forms per tenant: Define which fields are included and which are required on deposit forms (per work type); Define available values for dropdowns and selection lists View Issue on GitHub	See above
	(M) New type - journals	We host two journals and would need to see improvement	Trinity	(S) New Work Types: ETDs, Publications (articles, journals, books), and Datasets View Issue on GitHub	None.
Analytics	(M) User notifications, more robust reporting		Trinity		

Engaging with the Hyku Community

It is clear to the pilot team that should TDL move ahead with Hyku support for its members in the future, consistent involvement with the Samvera, Hyrax and Hyku service provider communities beginning at this early stage is essential. Even though the pilot's conclusions will not lead to immediate Hyku adoption, TDL will continue monitoring the community and reporting the status to our members. This is especially essential in terms of the requirements that pilot members identified as lacking in the current iteration of the system.

LDCX at Stanford

Early in the pilot, Courtney Mumma was able to attend LDCX 2018 at Stanford University. There, she was able to meet with representatives of the Samvera, Hyku, and Hyrax communities including Mike Giarlo and Hannah Frost of Stanford, and Tom Johnson of Digital Curation Experts. She brought questions from the TDL Hyku Pilot about the status of Hyku adoption, how the application and community might be supported in the future, and potential service providers. She learned that, at the close of the Hydra In a Box grant work, Stanford University would no longer continue to lead Hyku development or maintenance as they would not use the system locally. DCE will continue to work on Hyrax with Tom Johnson as the Technical Lead, but as of the end of LDCX, Hyku had no clear lead institution, developer, or governance.

Texas Conference on Digital Libraries

Nicholas Woodward, TDL's software developer and technical lead on the Hyku Pilot presented details of the project to the TDL community at TCDL in May 2018. The presentation served to inform and/or update the TDL member and greater Texas cultural heritage community, but failed to generate more interested partners beyond the original pilot partners.

Hyku Service Providers and Hyku Interest Group

DuraSpace convened a meeting of individuals and service providers in mid-April to discuss the future of Hyku development and support. The April 19th Hyku Service Providers Discussion would be the first of monthly meetings which still continue as of the publication of this report. A small community of Hyku implementers and piloters, like us, the Hyku Service Providers consist of colleagues from DuraSpace, Stanford, Notch8, Ubiquity Press, CoSector at the University of London, and the DCE Hyrax Technical Lead, Tom Johnson. The Hyku Interest Group discusses the Hyku repository application, particularly focusing on keeping its development aligned with Hyrax maturation. Group members are interested in working together to contribute to the maintenance and development of Hyku. The group has begun drafting a charter document to lead the formation of a Hyku Interest Group embedded in the Samvera Community. Nicholas Woodward, the TDL software developer lead for the pilot, presented the work of the group at Samvera Virtual Connect in July to a positive response⁸.

Outcomes

Feasibility and cost analysis of setting up a TDL-hosted Hyku service

One of the main goals of the pilot was estimation of the TDL service components needed in order to launch a service product. The testing environment did not include Amazon S3 backups, which are common to all active TDL member services. Additionally, very small items and collection sets were used to test for each institutional pilot partner, so scale tests were not possible. However, we were still able to collect some cost data for each service component based on testing.

For our four Hyku tenants, the TDL cost for storage and hosting on Amazon magnetic EBS storage for one year would amount to approximately \$2000USD if each tenant used up to 200 GB of storage. EBS volume storage pricing is available in Amazon's [documentation](#), and would change the cost to TDL should storage needs exceed that 200GB limit.

In addition to technological infrastructure, TDL would require dedicated staff time to manage the service, conduct training and outreach, continually engage and participate in the Samvera Hyku Interest Group and other community activities, and create a governance model that would include a member-composed steering committee and/or user group.

Assessment

The project team at TDL hope that this report will provide information and clarification of the Hyku project and functionality for the entire Texas cultural heritage community. At this time,

⁸ <http://hdl.handle.net/2249.1/87487>

however, it is unclear whether there is enough TDL member community interest to move forward with the creation of Hyku service. The following are some conclusions TDL has drawn about necessary feature development before we have a viable service. In general, the pilot members reacted neutrally to the system as it currently exists. They did not seem to think any of their institutions were anywhere close to ready to move to Hyku from their current systems, despite various levels of satisfaction with those legacy systems.

Without the following features, identified as 'must have' by our pilot team, Hyku is not ready for deployment as a TDL service offering:

- Bulk Import
- Bulk Export
- Migration tools for moving content from other platforms
- Persistent links to works and content items
- Configurable metadata fields per tenant
- New work types, especially journals
- Improved user notifications and reporting

Next Steps

Despite the unlikelihood of immediate adoption, TDL and its members will continue to monitor and assess the long-term viability of the Hyku service provider community and conduct additional investigatory work. First, TDL will share this report with the Samvera, Hyrax and Hyku communities in a public forum. Additionally, TDL continues to field questions from its member community and other Texas institutions about Hyku migration paths and support.

Beyond this pilot and final report, TDL intends to stay involved with the monthly Hyku Interest Group meetings and other Samvera community meetings and discussions. From discussions with the Hyrax, Samvera and Hyku communities, it is clear at this stage that Hyku will always trail Hyrax in terms of development and support, because Hyku has a layer of multitenancy code that must be updated to work with any updated Hyrax code.

In the spirit of staying closely involved with the ongoing Hyku project, TDL has joined the [Bridge2Hyku project](#). TDL has committed to project planning with the Bridge2Hyku team in September 2018. In late October, TDL will host Hyku for the Universities of Miami and Houston. The project will continue testing the migration toolkit developed by the University of Victoria for ContentDM migration to Hyku through January 2019, at which time the project partners with TDL's assistance will create and release their survey and aim to produce a survey report by February 2019.