

# **Bridge2Hyku Toolkit Pilot Test Report**

## **By B2H Project Team**

### **Introduction**

The University of Houston (UH) Libraries, in partnership and consultation with Indiana University at Bloomington (IUB), Indiana University-Purdue University Indianapolis (IUPUI), University of Victoria (UVic), University of Miami (UM) and primary community stakeholders including Stanford University, DuraSpace, and the Digital Public Library of America (DPLA), was awarded an IMLS National Leadership Grant (LG-70-17-0217-17) to support the creation of the Bridge2Hyku (B2H) Toolkit for sustainable data migration from CONTENTdm (CDM) to Hyku. Hyku (formerly called Hydra-in-a-Box) is an open source digital repository developed by Stanford University, Duraspace and the Digital Public Library of America (DPLA). The B2H Toolkit aims to help institutions understand their digital library ecosystems and provide software and guidance for successful migration to Hyku. This two year grant project is divided into three phases. In phase one, the team identified metadata and system requirements needed for crosswalking data from CONTENTdm to Hyku. Phase two is dedicated to toolkit development, documentation, and the creation of a B2H website. In phase three, the team will assess, improve and promote the developed toolkit.

To wrap up phase two, the B2H project team at UH and UVic partnered with the Texas Digital Library (TDL) to conduct the B2H Toolkit Pilot Test. TDL is a consortium that provides open source digital repositories to higher education institutions in Texas. Hyku was developed as a Hyrax/Samvera instance with consortial use in mind. Planning began in the late summer of 2018. The pilot test occurred between November 2018 and February 2019. Participants tested against a rubric to provide structured feedback on functionality, pain points, and any missing features.

### **B2H Toolkit Pilot Test Planning**

In September 2018, the UH B2H team partnered with developers from TDL and UVic to generate user stories to address customizations required for a consortial Hyku instance with three tenants<sup>1</sup>. The group also determined the migration paths worth attempting. To test these migration paths, the pilot included a testing rubric modeled after the one used in the TDL Hyku Pilot earlier in 2018<sup>2</sup>. This allowed the Bridge2Hyku pilot test to align with feature sets and requirements collaboratively developed by the TDL Hyku pilot institutions.

### **B2H Toolkit Pilot Test Processes**

The pilot test activities included:

---

<sup>1</sup> A tenant is a institution that has a unique repository layer within a consortial platform.

<sup>2</sup> The report for the TDL Hyku Pilot is available here:

[https://drive.google.com/open?id=1x85ddpd10W7jLp3ZgBL\\_tm00lrs93U-V](https://drive.google.com/open?id=1x85ddpd10W7jLp3ZgBL_tm00lrs93U-V)

- Customizing Hyku for consortial needs
- Confirming the B2H Toolkit's compatibility with several migration paths
- Identifying and prioritizing additional features to develop in Phase 3 of the Bridge2Hyku grant

### *Hyku Customization*

The team spent 3 sprints in October, November, and December of 2018 customizing Hyku for consortial needs and specific participating institutional issues. They were able to customize Hyku for University of Houston's metadata schema and University of Victoria's schema, but were unable to successfully edit filesets, which would have allowed different metadata profiles on each Hyku instance. These customizations were done on TDL's GitHub.<sup>3</sup> The unfinished issues surrounding file sets were reversed to create a stable Hyku platform for the migration testing to begin.

### *B2H Tools' Compatibility*

Pilot testing focused on the B2H tools: CDM Bridge and HyBridge, which were developed for digital collections data migration.<sup>4</sup> The pilot test aimed to confirm the B2H tools' functionality and compatibility with several migration paths:

- local CDM to local Hyku
- local CDM to hosted Hyku
- hosted CDM to hosted Hyku
- hosted CDM to local Hyku

With CDM Bridge's ability to pull collections from any CONTENTdm (CDM) instance when provided the correct credentials, we were able to test local and hosted CDM instances without any difficulties. CDM Bridge's crosswalking functionality allowed UH and UVic to move metadata into their customized schemas.

The HyBridge gem was added to TDL's instance of Hyku and has also been added to UH's local Samvera instance, confirming the gem's ability to be used with Hyrax and Hyku in both hosted and local instances.

There were some issues when we initially started testing the migration workflow. Our TDL partner, Nicholas Woodward, Senior Software Engineer, worked with Sean Watkins, Lead

---

<sup>3</sup> <https://github.com/TexasDigitalLibrary/hyku>

<sup>4</sup> CDM Bridge and HyBridge are available on the B2H GitHub repository: <https://github.com/Bridge2Hyku>

Developer at UH, to optimize the server settings for a production-level use of Hyku's ingest framework. Woodward shared the issues encountered:

“We ran into issues early on with the system resources being quickly exhausted during the derivative generation process. It was primarily overload on the CPUs that would cause the Rails application to become unresponsive for users. With some help from Sean, we were able to determine the optimum number of threads to run when creating derivatives after some trial and error. The solution was fewer processes running for a longer period of time”

After successful modifications to the server and job configurations, TDL migration testers also found issues surrounding the location and rights metadata fields. These fields use URIs from [GeoNames](#) and [Rightsstatements.org](#), respectively. Hyku expects the values of these fields to be URIs, but neither CDM Bridge nor HyBridge were configured to supply URIs or validate field input. Through troubleshooting, the UH team was able to identify the format required for these field values prior to HyBridge ingest, and what warning messages should appear if these errors occur.

### *B2H Tool Enhancement*

The viability of the B2H tools allowed for the testers to search for pain points and specific issues within the process. For issues that required small development enhancements, the UH Libraries Project Team made updates to CDM Bridge and HyBridge after the initial alpha release<sup>5</sup> of the toolkit, including some that were found by early adopters of the applications:

<i>Feature</i>	<i>Product</i>	<i>Source of feedback</i>
Ability to map Item/File-Level Metadata	CDM Bridge	IUPUI testing, feedback from presentation and demonstrations
Backend modifications required for working with larger collections	CDM Bridge	Bug reports submitted by UNC-Chapel Hill
GeoNames issues (location field requires URI)	HyBridge	B2H Pilot Testing

---

<sup>5</sup> The initial alpha release was September 10th, with HyBridge 0.1.0 and CDM Bridge 0.5.0 creating a functional migration workflow.

Warning messages for invalid Rightsstatement.org URIs (rights field requires URI)	HyBridge	B2H Pilot Testing
Hyrax compatibility updates (needed to work with both Hyku and Hyrax instances)	HyBridge	B2H Pilot Testing
Ingest status headers (Ready/Processing/Complete)	Hybridge	B2H Pilot Testing
Warning messages for invalid work types	Hybridge	B2H Pilot Testing

**Recommendations and Next Steps**

With the completion of the B2H tools enhancement during pilot testing period, the team identified one additional feature to implement: specifying a destination collection for content to be ingested using the HyBridge migration tool. This feature is prioritized for the April development sprint by the B2H Project Team. After completion of this enhancement, the project team will announce the B2H Toolkit beta release.