

University of Houston

## **Bridge2Hyku Toolkit: Developing Migration Strategies for Hyku**

### **Abstract**

The University of Houston (UH) Libraries, in partnership and consultation with Indiana University at Bloomington (IUB) and 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), seeks \$249,103 in funding from an IMLS National Leadership/Project Grant to support the creation of the Bridge2Hyku (B2H) Toolkit. Focusing on general information and guides for migration as well as on specific content for migrating to the Hyku platform, the toolkit will help institutions to better understand their digital library ecosystems and how they can prepare for migration. The toolkit will include (1) B2H Github Organization to host a website and software repositories, (2) general migration guidance with specific information about installing, configuring, managing, and ingesting content into Hyku and discussions of data modeling and metadata mapping for compatibility with DPLA, (3) descriptive information about the migration software available for the CONTENTdm origin platform and (4) the software to assist users with migration to Hyku.

While Hyku will offer a robust, extensible platform for making digital objects accessible over the web, institutions cannot take advantage of the benefits the platform offers until they confront the difficulties of migrating legacy data from other systems. As the results of the Hydra-in-a-Box User Survey<sup>i</sup> demonstrate, many institutions want to migrate from their current DAMS, but they lack the staff expertise and resources to plan and execute a successful data migration. These problems are compounded by a lack of general guides, tips or hands-on tools to assist libraries in completing the DAMS migration lifecycle. The UH Libraries, in partnership with IUB, IUPUI, UVic and UM, and in consultation with community stakeholders Stanford, DuraSpace, and DPLA, will leverage a National Leadership Project Grant to build the B2H Toolkit. This toolkit will fill current gaps in the migration process by helping libraries to understand their own digital ecosystem and offer them strategies and tools for migrating their digital collections from CONTENTdm to Hyku and promoting them to DPLA. There is a potential for the toolkit to offer migration capabilities from other digital asset management systems in the future.

The B2H Toolkit plan and outcomes aim to establish a framework for sustainable data migration. The project will leverage IMLS' investment to build capacity for libraries and cultural heritage institutions in adoption of Hyku nationally and worldwide. It will engage and strengthen the open source Samvera Community (formerly Hydra Community) around Hyku by leveraging our collective expertise through strategic collaboration between UH, IUB, IUPUI, UVic, UM, Stanford, DuraSpace, and DPLA. This project will deliver national impact by addressing the following IMLS performance goals: (1) Broaden access and expand use of the Nation's content and collections, by helping libraries and cultural heritage institutions migrate their digital collections to the Hydra-in-a-Box platform and promote them to DPLA. (2) Improve management of the Nation's content and collections, by assisting institutions in migrating to a community supported, nationally recognized digital platform, which promises better management and shared expertise of institutions' digital assets.

This two year project will be divided into three phases. In phase one, the team will identify metadata and system requirements needed for cross-walking data from CONTENTdm to Hyku. Phase two will be dedicated to the B2H Toolkit development, documentation, and the creation of a B2H website. In phase three, the team will assess, improve and promote the developed toolkit. The B2H project team will track project activities using Agile project management methodology and assess project goals and outcomes at the completion of each phase. The team will also explore possibilities for incorporating its resources within the Samvera Community codebase. To make it sustainable, University of Houston Libraries will maintain the stewardship of the B2H Github Organization during and after the grant period. The project team will also encourage active participation from Samvera Community and make sure that it continues to be a long term essential resource for data migration to Hyku.

---

<sup>i</sup> Hydra-in-a-Box User Survey Final Report (2015): <https://purl.stanford.edu/jk292fy8802>

# Bridge2Hyku Toolkit: Developing Migration Strategies for Hyku

## Introduction

The University of Houston (UH) Libraries, in partnership and consultation with Indiana University at Bloomington (IUB) and 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), seeks \$249,103 in funding from an IMLS National Leadership/Project Grant to support the creation of the Bridge2Hyku (B2H) Toolkit. Focusing on general information and guides for migration as well as on specific content for migrating to the Hydra-in-a-Box platform, the toolkit will help institutions to better understand their digital library ecosystems and how they can prepare for migration. The toolkit will include (1) B2H Github Organization to host a website and software repositories, (2) general migration guidance with specific information about installing, configuring, managing, and ingesting content into Hyku and discussions of data modeling and metadata mapping for compatibility with DPLA, (3) descriptive information about the migration software available for the CONTENTdm origin platform and (4) the software to assist users with migration to Hyku.

## Statement of Need

Digital Asset Management Systems (DAMS) have evolved over time as the technologies that support them have been refined and user needs and expectations have shifted. Not surprisingly, libraries have come to reassess, select, and migrate to a new DAMS based on the changes in technology and user needs. Few studies have been done to understand how this reassessment process works. A research project by Stein and Thompson explored the motivations for migrating from one DAMS to another.<sup>1</sup> Their findings indicate that the reassessment and migration process often suggests that institutions are moving from proprietary systems, such as CONTENTdm, DigiTool, and Rosetta, towards open source solutions, including DSpace, Fedora, and Islandora. Furthering this trend, IMLS has invested funding in Stanford, the DPLA, and DuraSpace's collaborative effort to develop Hydra-in-a-Box, a turnkey, open source digital repository. Assuming this trend continues as the barriers to adopting open-source software decrease, the need for tools, tutorials, and documentation on how to migrate from a proprietary system to an open source system, such as Hyku, will understandably grow over time.

This grant proposal focuses on developing a general, extensible resource guide for digital collection migrations and specific tools and documentation for migration to Hyku from one of the most prominent proprietary DAMS on the market: CONTENTdm. This software is deployed widely across the globe by over 2000 organizations worldwide, including nearly 30 alone in the state of Texas. OCLC highlights that the range of institutions deploying CONTENTdm covers the full spectrum of library types and sizes, including academic, public, consortial, and special libraries. Library surveys validate many of OCLC's figures. Stein and Thompson, Myntti and Woolcott,<sup>2</sup> and the Hydra-in-a-Box<sup>3</sup> surveys all collect demographic information about survey participants (and/or their institutions) and which DAMS they currently use for providing access to digitized and born-digital content. In all of these surveys, CONTENTdm was either the first or second most-frequently used DAMS among survey participants.

---

<sup>1</sup> <https://uh-ir.tdl.org/uh-ir/handle/10657/1367>

<sup>2</sup> <https://catalogingunit.org.wordpress.com/2016/06/17/digital-asset-management-systems-used-by-catalogers>

<sup>3</sup> <https://purl.stanford.edu/jk292fy8802>

While many institutions continue to adopt CONTENTdm, there are also documented case studies where other organizations have decided to migrate away from the product, including the University of Houston,<sup>4</sup> the College of Charleston,<sup>5</sup> Texas Tech University,<sup>6</sup> the University of Utah,<sup>7</sup> and members of the Florida Council of State University Libraries.<sup>8</sup> These case studies document a variety of reasons for implementing a new DAMS, including the institution's desire for increased accuracy when querying the repository for objects as well as better prominence in search results; expanding the type and kinds of repository and user interface customizations available; engaging in a community of development and support that develops expertise across the profession; and investing library funds in this development as opposed to in fees for licensing and maintenance.

Many of the limitations documented in professional literature and white papers are noted in the Hydra-in-a-Box survey as desired expectations and functionality in the future Hyku product. While Hyku will offer a robust, extensible platform for making digital objects accessible over the web, institutions cannot take advantage of the benefits this platform offers until they confront the difficulties of migrating legacy data from other systems. As the results of the Hydra-in-a-Box user survey demonstrate, many institutions want to migrate from their current DAMS, but they lack the staff expertise and resources to plan and execute a successful data migration.<sup>9</sup> These problems are compounded by a lack of general guides, tips or hands-on tools to assist libraries in completing the DAMS migration lifecycle.

## Impact

The B2H Toolkit plan and outcomes aim to establish a framework for sustainable data migration. The project will leverage IMLS' investment to build capacity for libraries and cultural heritage institutions in adoption of Hyku nationally and worldwide. It will engage and strengthen the open source Samvera Community (formerly Hydra Project) around Hyku by leveraging our collective expertise through strategic collaboration between UH, IUB, IUPUI, UVic, UM, Stanford, DuraSpace, and DPLA.

The B2H project will deliver national impact by addressing the following IMLS performance goals: (1) Broaden access and expand use of the nation's content and collections. Serving as a bridge, the B2H Toolkit will fill current gaps in the migration planning and data migration processes by helping libraries and cultural heritage institutions, large or small, understand their own digital ecosystem and offer them strategies and tools to migrate their digital collections from CONTENTdm to the Hyku platform. With components connected to DPLA hubs, the Hyku platform offers the capability to promote institutions' digital content and collections to the DPLA discovery portal. Migrating digital content and collections to Hyku, a platform built for DPLA metadata aggregation,<sup>10</sup> is a first step toward publishing digital content in this national digital portal, especially for small- and medium-sized institutions without substantial technical expertise on staff. B2H will help to expand access to and use of the nation's digitized cultural heritage materials by helping institutions assess their collections, prepare for migration, and successfully execute data migrations to the Hyku platform. (2) Improve management of the nation's content and collections. Hyku, being developed by leaders in the cultural heritage digitization sector, offers robust features, systems interoperability, scalability, and linked data publishing. The B2H Toolkit will assist institutions in migrating to a community supported digital

---

<sup>4</sup> <https://ejournals.bc.edu/ojs/index.php/ital/article/view/9152/pdf>

<sup>5</sup> <http://journal.code4lib.org/articles/8327>

<sup>6</sup> <https://conferences.tdl.org/tcdl/index.php/TCDL/TCDL2013/paper/viewFile/582/268>

<sup>7</sup> [http://www.mwdl.org/events/DAMS\\_options.php](http://www.mwdl.org/events/DAMS_options.php)

<sup>8</sup> <https://islandora.pubwiki.fcla.edu/wiki/images/Islandora/6/67/DigitalLibraryPlatformWorkingGroupReport.pdf>

<sup>9</sup> <https://purl.stanford.edu/jk292fy8802>

<sup>10</sup> <http://www.dlib.org/dlib/may17/gore/05gore.html>

object delivery platform based on next generation web standards that promises better management capabilities than existing platforms and seeks to align content models and metadata delivery for easier data sharing and reuse.

The documentation and guidelines within the B2H Toolkit will also serve as crucial components for quicker and wider adoption of Hyku and the development of the open source software community. The B2H Toolkit intends to have the functionality to help migrate digital collections of various sizes and formats including image and text files. The B2H Toolkit will also adapt to offer AV migration strategies to Hyku for basic media functionality. Hyku AV functionality will evolve with the integration of Avalon, which will provide further enhancements to time-based media access. Since CONTENTdm is one of the most widely used digital platforms, the B2H Toolkit will be built initially to help CONTENTdm users to migrate to Hyku. The B2H Toolkit will have the potential to be further developed to help digital collections and content migration by other institutions using other open source or proprietary digital asset management systems.

The B2H Toolkit software will be published in a GitHub repository, thus will be widely open and useful to other institutions. The project team have assembled not only experts for development and testing of the toolkit but also advisory board members for input and suggestions. The project will be directed by Annie Wu, Head of Metadata and Digitization Services (UH), and Santi Thompson, Head of Digital Research Services (UH). The toolkit development and testing experts include: Sean Watkins, Lead Repository Developer (UH); Content Strategist (hired with grant funds at UH); Andy Weidner, Metadata Services Coordinator (UH); Anne Washington, Metadata Librarian (UH); Julie Hardesty, Metadata Analyst (IUB); Braydon Justice, Developer/Analyst Digital Scholarship (UVic); Dean Seeman, Head of Metadata (UVic); Mike Giarlo, Hydra-in-a-Box Technical Manager (SU); Christina Harlow, Repository Specialist (SU); Paul Clough, Digital Architect and Infrastructure Librarian (UM); and Elliot Williams, Digital Initiatives Metadata Librarian (UM). The project advisory board members include: Tom Cramer, Chief Technology Strategist & Associate Director DL Systems and Services (SU); Jon Dunn, Assistant Dean for Library Technologies (IUB); Hannah Frost, Product Manager of Hydra-in-a-Box (SU); Michele Kimpton, Executive Director, Digital Public Library of America; Debra Hanken Kurtz, CEO of DuraSpace; David Wilcox, Product Manager (DuraSpace); Erin Tripp, Business Development Manager (DuraSpace); Kristi Palmer, Associate Dean of Digital Scholarship (IUPUI); Chip Dye, Director of IT (IUPUI); and Lisa Goddard, Associate University Librarian, Digital Content & Strategy (UVic); Sarah Shreeves, Associate Dean for Digital Strategies (UM). The toolkit development and testing team, along with the advisory board members, brings together experts in digitization and metadata practice and standards, digital systems infrastructure and functionality, programming, metadata cross-walking, and data exchange in order to produce a data migration toolkit that will help to expand access to our digitized cultural heritage resources. The project team will present, publish, and promote the migration toolkit and engage libraries, cultural heritage institutions, and other communities in its use, assessment, and further development.

## **Project Design**

### **Performance Goals and Outcomes**

Our project aims to produce a community resource that will assist institutions considering a migration from their current digital content delivery system to the Hyku (Hydra-in-a-Box) platform. The B2H Toolkit will include general information about digital collection migration planning, specific information about modeling data for and migrating to Hyku, and software tooling for analyzing and extracting data from existing repository platforms. The project team has identified the following

performance goals and outcomes:

1. Identify migration needs within the LAM repository community
2. Formalize requirements and best practices for establishing Hyku repositories
3. Create a B2H website to assist with migration planning for Hyku
4. Develop and adapt software tooling to assist with migrations
5. Build community support for the shared B2H resources

With the assumptions that libraries and cultural heritage institutions would have strong interest in migrating from CONTENTdm to Hyku and would be willing to join and take advantage of a growing open source community, the B2H Toolkit will help libraries and institutions achieve their migration goals. While designing the project, the team is aware that the development of Hyku is ongoing. The team has included the Hyku development team members as our stakeholders to minimize any potential risks.

### Preliminary Work & Design

UH has already completed work on metadata remediation and CONTENTdm to Hyku migration workflow tooling. The Metadata Upgrade Project<sup>11</sup> was a multi-year effort to standardize and clean up the descriptive metadata in the UH digital collections that led to the recent development of a machine-actionable metadata application profile<sup>12</sup> based on the DPLA MAP<sup>13</sup>. The Bayou City DAMS implementation team has created a suite of workflow and data management tools -- including an identifier minter/resolver, a SKOS vocabulary manager, a digitization workflow tool, and a descriptive metadata workflow tool -- to ease the transition from CONTENTdm to Hyku.<sup>14</sup> UH has also developed the Hunting Ruby gem<sup>15</sup> that provides convenience methods to the CONTENTdm API<sup>16</sup> and the Buffalo Ruby library<sup>17</sup> that provides methods for mapping CONTENTdm metadata harvested with Hunting.

UVic has developed a `cdm_migrator` Ruby gem<sup>18</sup> that facilitates metadata crosswalking between CONTENTdm and Hyrax, the content management application upon which Hyku is built. The `cdm_migrator` exports CONTENTdm metadata as a CSV with Hyrax mappings and provides an interface for ingest into Hyrax after metadata remediation.

The University of Miami has developed expertise in CONTENTdm metadata transformation, recently completing a project to implement rights statements across all digital objects, and will serve as a regional aggregator for the DPLA hub in Florida. UM will participate in the Hyku pilot project in Fall 2017.

IUPUI has developed expertise around metadata remediation/aggregation issues in their work contributing collections to the Indiana DPLA service hub, Indiana Memory.<sup>19</sup>

---

<sup>11</sup> <http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/218>

<sup>12</sup> <https://vocab.lib.uh.edu/bcdams-map>

<sup>13</sup> <https://dp.la/info/developers/map/>

<sup>14</sup> <http://journal.code4lib.org/articles/12342>

<sup>15</sup> <https://github.com/uhlibraries-digital/hunting>

<sup>16</sup>

<https://www.oclc.org/support/services/contentdm/help/customizing-website-help/other-customizations/contentdm-api-reference.en.html>

<sup>17</sup> <https://github.com/uhlibraries-digital/buffalo/tree/master/lib>

<sup>18</sup> [https://github.com/UVicLibrary/cdm\\_migrator](https://github.com/UVicLibrary/cdm_migrator)

<sup>19</sup> <https://dplaind.wordpress.com>

The B2H team will also work with ongoing documentation and migration work in the open source digital collections community (e.g. Samvera Metadata Interest Group, Samvera CONTENTdm Migration Interest Group, Digital Library Federation Metadata Assessment Interest Group) to collect use cases, identify migration needs, and build communities of interest around shared problems in the digital collections migration space.

## **B2H Project Activities**

To meet our performance goals and outcomes, the project team will adopt a three phased approach. In phase one, the team will identify metadata and system requirements needed for crosswalking data from CONTENTdm to Hyku. Phase two will be dedicated to the B2H Toolkit development, documentation, and the creation of a B2H website. The B2H team will use the information gathered in Phase 1 to develop a general set of topics to consider while planning for a migration and create content on those topics for the B2H website. The B2H team will draw on the experience of its Hyku Metadata Advisors to create content that describes best practices for Hyku data and add that content to the B2H website. This part of the site will be useful for any organization tasked with planning a digital collections data migration to Hyku. Building on the software development and migration workflow planning efforts already underway at UH and UVic, the project team will test the B2H Toolkit in actual migration work, describe the applicability of the B2H tools for various use cases, identify gaps in the existing CONTENTdm to Hyku migration software suite, and coordinate development work around those unmet needs. UH, IUPUI, UVic, and UM will develop use cases for their migration, identify the appropriate intellectual and software tooling for their particular needs, apply those tools to an actual data migration, and write an account of their migration strategy and execution. The B2H team will collect existing CONTENTdm migration software, supplement existing documentation when needed, adapt existing migration software to the UH, UVic, IUPUI, and UM use cases, and develop new migration software if needed. Descriptions of and usage instructions for the software adapted and developed during testing will be made available on the B2H website and the code will be published in the B2H Github Organization. In phase three, the team will assess, improve and promote the developed toolkit. The project team will track project activities using Agile project management methodology and assess project goals and outcomes at the completion of each phase. The B2H team will engage other digital repository communities (e.g. DSpace, Digital Commons) in this work, providing a vehicle for collecting use cases and software specific to migration from those platforms to Hyku. The B2H team will also explore possibilities for incorporating its resources within the Samvera Community codebase.

### **Phase One**

Activities (Oct. 1, 2017 to March 31, 2018)

- Hire the Content Strategist. The Content Strategist will work at UH and will be primarily responsible for designing, building, and maintaining the B2H Toolkit website
- Gather and analyze information about the UH, IUPUI, UVic, and UM digital collections environment:
  - Content types
  - Metadata schema
  - Digital collection production workflows
  - Preservation considerations
  - Stakeholder considerations
- Gather information about Hyku requirements and limitations:
  - Local vs. Hosted platforms
  - Content types available
  - Metadata configuration and delivery
  - Metadata maintenance interfaces
  - Access controls

- Configuration and branding

#### Deliverables

- Collected data and use cases related to project partners digital collections environment (March, 2018)
- Collected data on requirements and limitations of Hyku (March, 2018)

### Phase Two

#### Activities (April 1, 2018 to March 31, 2019)

- Establish a B2H Github Organization<sup>20</sup> to host the B2H Toolkit software repositories and website, which contains:
  - Information about the B2H project and team
  - General migration guidance, with links to existing community documentation and projects
  - Specific information about installing, configuring, managing, and ingesting content into Hyku, including discussions of data modeling and metadata mapping for compatibility with DPLA
  - Descriptive information about the migration software available for the CONTENTdm origin platform
  - A blog for periodic updates from the project team
- Adopt, adapt, and develop migration software for the CONTENTdm origin platform and publish it in a B2H Github repository
- Test the Alpha B2H Toolkit's migration documentation and software against the pilot partners' digital collections
- Improve toolkit functionality:
  - Develop assessment tools such as survey to solicit feedback on the Alpha B2H Toolkit from the development team, pilot partners and community stakeholders
  - Analyze and review assessment feedback
  - Improve toolkit features and functionality based on feedback
- Publish pilot partners' migration stories on the B2H website as case studies

#### Deliverables

- Alpha toolkit release (including website with migration guidance, software repositories) (June, 2018)
- Alpha toolkit pilot testing feedback (December, 2018)
- Beta toolkit release (March, 2019)

### Phase Three

#### Activities (April 1, 2019 to Sept. 30, 2019)

- Solicit and analyze feedback on Beta B2H Toolkit and improve functionality
- Communicate the B2H Toolkit to libraries and the cultural heritage digitization community via conference presentations, workshops, and online training sessions with instructional materials:
  - Present our project at potential conferences such as Digital Library Federation (DLF) Forum, Open Repositories (OR), Code4Lib, Digital Initiative Symposium, Texas Conference on Digital Libraries (TCDL), Coalition for Networked Information (CNI)
  - Conduct in-person workshop as pre or post conference programs
  - Offer online webinars on this toolkit in collaboration with professional library associations such as Association of Library Collections and Technical Services (ALCTS), The Library Information Technology Association (LITA), and TDL
- Ensure sustainability of the toolkit:

---

<sup>20</sup> <https://github.com/blog/674-introducing-organizations>

- Engage the Samvera Community in future development and expansion of B2H
- Pursue opportunities for cross-sector partnership, sustainability modeling, and community relations

Deliverables:

- Conference presentations and workshops (based on conferences time)
- B2H Toolkit release (September, 2019)

## **Project Resources: Personnel, Time, Budget**

The Project Directors and Principal and Co-Principal Investigators, Annie Wu and Santi Thompson, will provide guidance and strategic direction for the B2H project. Their responsibilities include: executing the project as outlined in the funded proposal; carrying out the project's financial plan; reporting project progress to IMLS; maintaining an accurate record of project related expenses; complying with UH policies and procedures; and complying with all IMLS rules, regulations and terms and conditions. Annie Wu has rich experience directing UH Libraries digital initiatives, leading successful conversion of legacy data to meet the new Resource Description and Access bibliographic standard. She chairs the UH Libraries DAMS Task Force in evaluation, implementation and integration of digital access, preservation and archival systems. Annie has also worked with colleagues in developing policies and best practices on digital collection development and preservation. She has published and presented on information organization and discovery, digital initiatives, linked data, digital preservation policies and programs, and metadata practices and standards. Annie was also a member of the 2016-2017 cohort of Association of Research Libraries Leadership Fellows. Santi Thompson has experience developing policies, workflows, and services for digital research support and digital repositories. He has led library and consortial teams charged with developing complex strategies for digital preservation programming and access to research data. He publishes on the assessment of digital repository metadata, software, and content reuse and was the chair of the Texas Digital Library (TDL)'s Dataverse Implementation Team, which launched the Texas Data Repository - a statewide data repository for researchers at higher education institutions in Texas.

The Project Manager, Andy Weidner (UH), will coordinate collaboration between the B2H Toolkit development team and interested members of the community in order to drive toolkit development toward a timely completion. In Phase 1, he will work with the Content Strategist and Migration Strategists to collect and synthesize data about the pilot partners' collections. In Phase 2, he will oversee an Agile development cycle as the B2H team translates needs assessment data into website content, deploys the B2H Toolkit in data migrations from CONTENTdm to Hyku, identifies gaps in the CONTENTdm to Hyku migration software suite, develops new software to bridge those gaps, and publishes institutional migration stories as case studies for others to learn from. In Phase 3, he will organize B2H presentations and workshops and coordinate community engagement around improving and expanding the B2H Toolkit content. Weidner currently leads the Bayou City DAMS implementation team at UH, serves on the Dublin Core Metadata Initiative Infrastructure Advisory Committee, is a member of the Samvera Metadata and Samvera CONTENTdm Migration Interest Groups, and has published and presented extensively on metadata remediation strategies for CONTENTdm.

The Content Strategist (CS), to be hired by UH with grant funds, will build the B2H website and create original content by analyzing data collected from the pilot partners. The CS will generate workflow documentation established by the B2H development team and ensure that the user experience is consistent throughout the site. The CS will manage the B2H website through the Github repository. In Phase 1 the Content Strategist will work with the Migration Strategists to conduct interviews and surveys with pilot partners and other institutions. The CS will also assist with developing workflows, data models and use cases. In Phase 2, they will participate in the Agile development cycle and work



with developers to create documentation for migration software. The CS will also write blog posts regarding the status of the B2H Toolkit development. In Phase 3, the CS will deliver B2H presentations, assist in workshops and engage with the community by providing marketing materials for B2H Toolkit. The CS must have a strong work ethic and show they can work successfully with a collaborative team. The CS will have varied communication skills and exceptional writing ability in order to effectively address different audiences, channels, tones, and styles.

The Software Developers, Sean Watkins (UH), Braydon Justice (UVic), and Paul Clough (UM) will develop migration software to meet the needs of the B2H migration from CONTENTdm to Hyku. In Phase 2, the development team will participate in Agile development cycles and establish feature requests in the form of user stories to build the software tools. Developers will take advantage of any existing migration tools, such as CdmMigrator<sup>21</sup>, and contribute work back into those projects. Developers will be responsible for maintaining and managing Github repositories for any developed software. The developers will work with the Content Strategist in developing software documentation and supported content for the B2H toolkit website. In Phase 3, the developers will assist the Content Strategist in creating materials for workshops, presentations and marketing. Sean Watkins is on the Bayou City DAMS implementation team at UH where he leads the development for several digital repositories and software tools used in the DAMS workflows. His work includes leading a team to build a customized access interface for UH's digital library<sup>22</sup> that sits on top of CONTENTdm. He has developed software suites currently being used in the DAMS workflow which include a metadata editor, Brays<sup>23</sup> and preservation packager, Carpenters<sup>24</sup>. Paul Clough manages the digital collections infrastructure at the University of Miami Libraries, where his current work is focused on implementing and migrating to a new archival description and access system. His previous experience as a Librarian and software developer focused on book digitization, image collections, and the Avalon Media System. Braydon Justice is the lead ruby developer/analyst at the University of Victoria where he has developed their Spotlight instance with features like audio, video, and 3D object capabilities as well as image annotation. He is currently developing their iteration of Hyku and also created the CdmMigrator tool to facilitate metadata crosswalking from CONTENTdm to Hyku.

The Migration Strategists, Anne Washington (UH), Dean Seeman (UVic), and Elliot Williams (UM) will collect and synthesize pilot partners' collection assessment data for reporting and develop use cases for the Phase 2 backlog. They will develop general topics of consideration for migration and recommended practices, and they will plan and conduct migrations at their respective institutions from CONTENTdm to Hyku using the B2H Toolkit. After the migration, they will refine migration recommendations on the B2H website and publish case studies to the B2H blog. Anne Washington is on the Bayou City DAMS implementation team at UH where she leads the Data Migration Working Group. In her current and previous positions, she has worked closely on the development of metadata application profiles, metadata mappings, as well as metadata remediation and enrichment strategies. Elliot Williams manages and participates in the creation of metadata for unique digital materials at the University of Miami Libraries. In his current position, his work involves ingesting and reviewing new digital materials, leading metadata enhancement and remediation, collaborating in the design of user interfaces for discovery and access, and developing mappings to support metadata harvesting. Dean Seeman is the Head of Metadata at University of Victoria Libraries and is currently responsible for UVic's metadata migration from CONTENTdm to Hyku. He has overseen metadata production and strategy since 2009 within CONTENTdm and other systems, both at UVic and at Memorial University

---

<sup>21</sup> [https://github.com/UVicLibrary/cdm\\_migrator](https://github.com/UVicLibrary/cdm_migrator)

<sup>22</sup> <http://digital.lib.uh.edu/>

<sup>23</sup> <https://github.com/uhlibraries-digital/brays>

<sup>24</sup> <https://github.com/uhlibraries-digital/carpenters>

of Newfoundland. He has presented and published on cataloguing, metadata, linked data, and mass digitization.

The Hyku Metadata Advisors, Julie Hardesty and Christina Harlow, will provide feedback on metadata best practices focused on Hyku. Julie Hardesty is the Metadata Analyst at Indiana University Libraries and has been involved in the Samvera Community since 2012, co-facilitating the Samvera Metadata Interest Group<sup>25</sup>, participating in several working groups involving metadata issues, and working on the team developing Avalon Media System, a digital audio and video management and access system based on the Samvera software stack. Christina Harlow is the Repository Specialist at Stanford University Libraries where she is working on Hyrax<sup>26</sup> metadata, one of the core components of Hyku. She has extensive experience in data migrations to Fedora and has presented widely on metadata and linked data topics.

Hyku Functional Requirements Advisor, Mike Giarlo (Stanford University), will provide information about Hyku requirements and limitations. Mike is the Technical Manager of the Hydra-in-a-Box project, the product owner of Hyku, and a Software Architect at Stanford University. His primary role is the technical team lead for a distributed, multi-institutional development team working to enhance, bundle and deploy a feature-replete and “turnkey” Samvera-based digital repository solution. He has been an active software developer and architect within the Samvera Community since 2011.

The Advisory Board members will be consulted periodically during the toolkit development process to ensure alignment of the toolkit with the broader goals of the Hyku, DuraSpace, and DPLA communities.

## Timeline

Phase	Timeframe	Activities
Phase 1. Needs Assessment	Oct. 1, 2017 to March 31, 2018	<ul style="list-style-type: none"> <li>● Gather information about project partners’ digital collections</li> </ul>
		<ul style="list-style-type: none"> <li>● Gather information about Hyku requirements and limitations</li> </ul>
		<ul style="list-style-type: none"> <li>● Hire Content Strategist</li> </ul>
Phase 2. B2H Toolkit Development	April 1, 2018 to March 31, 2019	<ul style="list-style-type: none"> <li>● Establish a B2H Github Organization to host a website and software repositories as Alpha B2H Toolkit</li> </ul>
		<ul style="list-style-type: none"> <li>● Develop and adapt the Alpha B2H software</li> </ul>
		<ul style="list-style-type: none"> <li>● Test the Alpha B2H migration documentation and software</li> </ul>
		<ul style="list-style-type: none"> <li>● Improve Alpha Toolkit Functionality</li> </ul>
		<ul style="list-style-type: none"> <li>● Release Beta B2H Toolkit</li> </ul>
		<ul style="list-style-type: none"> <li>● Publish pilot partners’ migration stories</li> </ul>

<sup>25</sup> <https://wiki.duraspace.org/display/hydra/Hydra+Metadata+Interest+Group>

<sup>26</sup> <https://github.com/samvera-labs/hyrax>

Phase 3: Toolkit Improvement and Promotion	April 1, 2019 to Sept. 30, 2019	<ul style="list-style-type: none"> <li>● Improve Beta B2H Toolkit Functionality</li> </ul>
		<ul style="list-style-type: none"> <li>● Conduct conference presentations and training</li> </ul>
		<ul style="list-style-type: none"> <li>● Release B2H Toolkit</li> </ul>
		<ul style="list-style-type: none"> <li>● Implement B2H Toolkit sustainability</li> </ul>

**Budget:**

The project team requests \$249,103 in support of the above grant activities. The majority of the requested project fund will support the salary and fringe benefit of the future content strategist. Additional funding will support promotion and communication of the developed toolkit. This amount reflects: (1) \$120,000 for salaries and wages to develop the B2H Toolkit; (2) \$35,155 for fringe benefits; (3) \$25,000 for communication and conference travel to conduct presentations and workshops; and (4) \$10,000 for professional development and training. Additionally, the team seeks \$58,948 for indirect costs related to program administration, which are calculated based on UH’s federally negotiated rate (31%).

**Communication and Sustainability Plan**

The B2H Toolkit project team will disseminate the results of the project via six methods: (1) progress updates and final results will be shared on the team’s website and blog, and through the partners’ and stakeholders’ communication channels; (2) throughout the grant period, the project team will solicit feedback from pilot partners and advisory board members as well as from participants and attendees at conference presentations. This feedback will be incorporated into the final project outputs; (3) offer presentations on the B2H Toolkit to libraries and the cultural heritage digitization community via potential conferences such as Digital Library Federation Forum (DLF), Open Repositories (OR), Code4Lib, Digital Initiative Symposium, Texas Conference on Digital Libraries (TCDL), Coalition for Networked Information (CNI); (4) offer workshops and online training sessions to instruct potential users of this toolkit in collaboration with professional library associations such as Association of Library Collections and Technical Services (ALCTS), The Library Information Technology Association (LITA), and The Texas Digital Library (TDL); (5) the project team will deposit project code and documentation to a public GitHub repository; and (6) after the completion of the grant, the team will publish the methodology, results, and analysis in an article in an open access, peer reviewed journal.

To make it sustainable, University of Houston Libraries will maintain the stewardship of the B2H Github Organization during and after the grant period. This stewardship includes responsibilities such as administering team members, creating and managing repositories, reviewing community contributions, and creating and revising website content. The project team will also encourage active participation from the Samvera Community to make this toolkit a shared investment. Specifically, it will devise potential models for community/third party hosted support for future B2H Toolkit users. This will ensure that the toolkit continues to be a long-term, essential resource for data migration to Hyku.

## Schedule of Completion

Activities	2017			2018												2019									
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	
<b>Phase One</b>																									
Hire the Content Strategist	█	█																							
Gather information about project partners' digital collection		█	█	█	█																				
Gather information about Hyku requirements and limitations		█	█	█	█																				
Analyze collected data					█	█																			
<b>Phase Two</b>																									
Establish a B2H Github Organization and website							█	█	█																
Develop and adapt the Alpha B2H software							█	█	█	█	█	█	█	█	█	█	█								
Test the Alpha B2H migration documentation and software												█	█	█											
Develop and disseminate assessment tools																█									
Analyze and review assessment feedback																	█								
Improve toolkit functionality																	█	█							
Release Beta B2H Toolkit																		█							
<b>Phase Three</b>																									
Test and improve Beta B2H Toolkit functionality																		█	█	█	█	█	█	█	█
Conduct conference presentations and training												█	█	█	█	█	█	█	█	█	█	█	█	█	█
Release B2H Toolkit																									█
Implement B2H Toolkit sustainability strategy																							█	█	█

## DIGITAL PRODUCT FORM

### Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

### Instructions

You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

## PART I: Intellectual Property Rights and Permissions

**A.1** What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

All documentation and training materials produced as part of this project will be released under the Creative Commons Attribution 4.0 International license (CC BY 4.0). Software developed by this project will be under the copyright of the primary institution(s) that developed it and released under the Apache 2.0 license (see Part III). Both of these licenses align with the licensing requirements of the Samvera Community.

**A.2** What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

We will assert no additional ownership rights over new digital content other than that listed in A.1.

**A.3** If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

This project has no concerns regarding cultural sensitivities or privacy.

## Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

### A. Creating or Collecting New Digital Content, Resources, or Assets

**A.1** Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

**A.2** List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

**A.3** List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

## **B. Workflow and Asset Maintenance/Preservation**

**B.1** Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

**B.2** Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

## **C. Metadata**

**C.1** Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

**C.2** Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

**C.3** Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

## **D. Access and Use**

**D.1** Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

**D.2** Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

## **Part III. Projects Developing Software**

### **A. General Information**

**A.1** Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

The proposed activities for this project include the adoption, adaptation, and development of software to build a suite of data migration applications. The migration software suite will assist data migration personnel with the following tasks: 1) harvesting descriptive metadata from CONTENTdm, 2) crosswalking descriptive metadata to Hyku's schema, and 3) ingesting descriptive metadata and files into Hyku. CONTENTdm users who intend to migrate their data to Hyku are the primary audience for the software suite.

**A.2** List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

Because there are so many variables involved with planning and executing a migration, it can be difficult for institutions (especially those with limited technical expertise) to identify and/or develop applications to assist with data migration from one digital asset management system to another. What works for one institution may not work for another for a variety of reasons. This project aims to bring together a suite of modular software tooling designed to address the needs of varied migration use cases and document patterns for using each component. For example, the University of Houston has developed the Hunting Ruby gem. Hunting provides convenience methods to the CONTENTdm API that allow for flexible descriptive metadata harvesting and may be helpful for metadata analysis and remediation prior to migration. The University of Victoria has developed the `cdm_migrator` Ruby on Rails gem that may be plugged into the Hyku application. The `cdm_migrator` provides interfaces for metadata crosswalking from CONTENTdm and ingest into Hyku that non-technical personnel can easily use. This project will identify and address data migration gaps by modifying existing software, developing new software, and producing documentation to explain how to most effectively use the software.

### **B. Technical Information**

**B.1** List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

The migration software suite components will primarily consist of code written in Ruby, using Ruby Gems and Ruby on Rails conventions as appropriate. This aligns with the codebase of the Samvera Community. The project team will use other languages and tools as needed to communicate with APIs and data stores.

**B.2** Describe how the software you intend to create will extend or interoperate with relevant existing software.

The migration software suite components will harvest data from the CONTENTdm API and facilitate batch ingest into Hyku. The project will also investigate options for connecting to linked data vocabularies for authority control during migration.

**B.3** Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

Specific dependencies are listed in B.1. The migration software suite components will include a bundled distribution of all necessary dependencies and software.

**B.4** Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

Technical documentation will be developed iteratively alongside the migration software suite components. The software developers will be responsible for maintaining documentation within the code; the content strategist will lead other project personnel in the creation and editing of broader technical documentation through Github. This documentation will be made freely available.

**B.5** Provide the name(s) and URL(s) for examples of any previous software your organization has created.

University of Houston: <https://github.com/uhlibraries-digital/hunting>; <https://github.com/uhlibraries-digital/cedar>; <https://github.com/uhlibraries-digital/greens>  
University of Victoria: [https://github.com/UVicLibrary/cdm\\_migrator](https://github.com/UVicLibrary/cdm_migrator)  
Indiana University: <https://github.com/avalonmediasystem/avalon>  
Stanford: <https://github.com/samvera-labs/hyku>

## **C. Access and Use**

**C.1** We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

All code developed directly for the migration software suite will be released as open source under an Apache Software Foundation License, version 2.0. This license aligns with Samvera Community norms and will facilitate future integration into the Samvera Community codebase should we choose to follow that path.

**C.2** Describe how you will make the software and source code available to the public and/or its intended users.

The components of the migration software suite will have source code freely available for download from Github.

**C.3** Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: Bridge2Hyku

URL: <https://github.com/Bridge2Hyku>

## **Part IV: Projects Creating Datasets**

**A.1** Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

**A.2** Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?



**A.3** Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

**A.4** If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

**A.5** What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

**A.6** What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

**A.7** What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

**A.8** Identify where you will deposit the dataset(s):

Name of repository:

URL:

**A.9** When and how frequently will you review this data management plan? How will the implementation be monitored?