DLP Technical Implementation: Frequently Asked Questions

Prepared by: Emily Porter, Digital Library Program Coordinator

Last revised: April 2020

The following questions and answers are provided to clarify the DLP Implementation Phase 1 Business Case.

When will the new repository be finished?

Our first phase of implementation is underway, building products for discovery (Lux) and staff curation (Curate) activities. We aim to launch version 1 of both products by mid-2020, with a beta preview of the Lux product to be released in April, 2020. After that point, we will plan for and prioritize additional work and integrations related to additional migrations, system integrations, and the Self Deposit application. Software development will be ongoing to provide enhancements and stability over time. Content migrations will also be ongoing, as we have a very large amount of material to ingest.

When will [specific application] be finished?

A roadmap is developed for each Cor product. Each product will have an initial version 1 focused on minimum viable product features drawn from our Discovery Phase requirements, followed by subsequent releases to add additional features. We plan to launch version 1 of Lux and Curate by mid-2020.

What does “Cor” mean vs. “DLP” or “Emory Digital Collections”?

DLP is the abbreviated name for the ongoing Digital Library Program initiative which includes the repository technology as well as broader policy and operational processes. DLP was also a department name within the Library Technology & Digital Strategies Division, but it was re-organized in 2018 to become part of the Digital Product Strategy team.

Emory Digital Collections is the name for the public-facing search and discovery interface (the internal product name is Lux).

Cor (Latin for “heart”, inspired by the Emory motto: “the wise heart seeks knowledge”) is the internal name for the Emory Preservation Repository platform built as part of the larger DLP project. Cor includes multiple, inter-related software applications based on the Samvera (previously known as Hydra) digital repository technology framework, as well as preservation services and storage.

The Cor product suite will ultimately contain three distinct user facing applications:

1. Curate: Library staff application supporting management, preservation, and ingest
2. Lux: Consumer/front-end application providing search, discovery, and delivery
3. Self-Deposit: provides expanded mediated submission functionality and additional submission workflows for campus submitters

What do you mean by “user facing products”?

- When will the new repository be finished?
- When will [specific application] be finished?
- What does “Cor” mean vs. “DLP” or “Emory Digital Collections”?
- What do you mean by “user facing products”?
- How does my current application fit into/relate to the new product suite?
- When will [specific feature] be available?
- Why are you only ingesting certain formats (text and images) right now?
- When will you finish migrating all the content we have committed to preserving?
- How can I get my (new) material into the repository?
- Who will have access to the Curate application?
- Who will have access to the Lux application?
- Who will have access to the Self Deposit application?
- How does the new ETDs application relate to the Cor Self Deposit application?
- When will our new digital repository material be included in discoverE/Primo?
- When we will send our new digital repository material to the Digital Library of Georgia, DPLA, HathiTrust, etc.?
- When will you integrate with [specific Emory system or external system]?
- Who is developing the new software?
- I’m a DLP Discovery Phase Working Group alumni/product owner/service manager with ties to the repository. How can I contribute to the development process ahead?
- How will you keep stakeholders informed during the technical development work?
We mean Emory community/visitors and Library staff users who will interact with the repository through a web application interface (Lux, Curate, and Self Deposit specifically). The Samvera technology framework incorporates additional specialized applications, utilities, integrations, and components. Not all are intended for, or will be accessible to, the general public or Emory or Library Staff – usage is limited to technology teams supporting the repository, such as LTDS and LITS Middleware or LITS Infrastructure.

How does my current application fit into /relate to the new product suite?

The new DLP product suite (Cor) is not an exact 1:1 recreation of the previous Fedora repository applications or other digital asset management products supported by the Emory Libraries.

Because the three products are closely integrated, all material ingested to the Cor repository

- will be available to end users via the same front-end discovery application (with visibility based on rights/permissions),
- will utilize the same Fedora instance
- will receive preservation and extended management functions through Curate, even if they are submitted via the Self Deposit application

The following is an approximation of where selected Emory Libraries' applications' current features most closely map to in the new environment's user facing products:

<table>
<thead>
<tr>
<th>System</th>
<th>Discovery (&quot;Lux&quot;)</th>
<th>Staff Curation (&quot;Curate&quot;)</th>
<th>Self-Deposit (name TBD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The &quot;DAMS&quot;</td>
<td>Y^</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Dataverse</td>
<td>Y^</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Digitized Books</td>
<td>Y^</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ETDs</td>
<td>Y^</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>The Keep</td>
<td>Y^</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUNA</td>
<td>Y^</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenEmory</td>
<td>Y^</td>
<td>Y</td>
<td>Y^</td>
</tr>
</tbody>
</table>

^ All material (depending on access controls/rights) can be viewed in the Discovery application.

** If mediated self-deposit submissions are needed vs. automated /harvested submissions

When will [specific feature] be available?

Each product will receive its own roadmap and backlog for development, based on the approved requirements identified in the DLP Discovery Phase, and each product will have minimal viable product features identified for their version 1 releases. Additional features will be assessed and queued based on future needs once the initial products are in production.

Why are you only ingesting certain formats (text and images) right now?
We prioritized the first phase of work around the top 5 initial collections identified by the Digital Collections Steering Committee. The top 5 Phase 1 collections approved included digitized books and still images. Additionally, there are enhancements in development for audiovisual material in the Samvera and IIIF technical communities that we want to wait for. By the time we release v.1 of the repository, we anticipate these enhancements will be complete. Material that does not contain Restricted Data will be queued for migration like other types of materials: we must also develop a strategy for accommodating material which may contain sensitive/restricted data in compliance with Emory LITS security requirements. Our project is also working in parallel with Emory’s migration of hosting infrastructure to Amazon Web Services (AWS), so we are monitoring the potential risk of having to re-migrate our repository data once their service offering is in production. This will be particularly significant for very large files (audiovisual, disk images, etc.).

When will you finish migrating all the content we have committed to preserving?

We have currently identified close to 200TB worth of digital content queued for ingest and are working with the Digital Collections Steering Committee to prioritize this queue. Migration of legacy content will be an ongoing effort but we do not yet have enough information to estimate the exact duration. In the first phase of work, we are migrating 5 collections and will use this experience to help us estimate future migrations.

How can I get my (new) material into the repository?

The Digital Collections Steering Committee and its subgroups provide policy and a selection process for reviewing and prioritizing material to be ingested into the repository. New material to be ingested into the repository must be reviewed by this group.

Who will have access to the Curate application?

Curate use will be restricted to designated Library staff who manage digital collections in the repository, but we will monitor needs over time.

Who will have access to the Lux application?

The repository “front door” as a resource will be broadly available to the Emory community and the world. Access to some repository material will be restricted to selected Emory users or groups, however.

Who will have access to the Self Deposit application?

The self-deposit application will be available to selected Emory campus users, based on approved deposit workflows and agreements. These users will be distinct from those who use the Curate application, however, and their access will be limited to depositing material as opposed to managing or preserving it long-term.

How does the new ETDs application relate to the Cor Self Deposit application?

The new ETD application was rebuilt in 2017-2018 in tandem with DLP activities in order to provide much needed stability for that service, which is used by multiple graduate schools on campus, and is tied to students’ ability to graduate. We have not fully planned the implementation of the Self Deposit application, but we anticipate incorporating or expanding the functionality of the ETD application, which already utilizes the Samvera framework. Ultimately, we will incorporate ETDs into our larger Cor infrastructure and ensure those materials receive the same preservation and discovery features as other materials in the repository.
When will our new digital repository material be included in discoverE/Primo?

Integration with discoverE/Primo is a planned feature, but the version of Fedora we are implementing does not support the use of OAI-PMH, the export protocol we previously used to share repository data with Primo. The implementation team will continue to monitor developments and technical options to ensure repository data is discoverable as part of the Libraries’ larger discovery strategy.

When will we send our new digital repository material to the Digital Library of Georgia, DPLA, HathiTrust, etc.?

As the repository has been in transition, we have delayed starting up new dissemination efforts which are subject to break as our technology changes. As we move to the Samvera framework, we can utilize more standards-based approaches to distributing our content and contributing to new shared collections. Additionally, following policy and procedures established in 2018 by the Digital Collections Steering Committee, distribution of Emory digital collections material to third party services must be formally established moving forward. Currently, we have approved an ongoing export of material to HathiTrust, and that feature is on the roadmap.

When will you integrate with [specific Emory system or external system]?

In v.1 of functionality we will support Shibboleth and integrate with selected metadata authority data providers. We have longer-term plans to integrate with the Emory Libraries’ discovery platform and metadata systems such as Alma and ArchivesSpace. We will also build integrations with campus systems such as Emory Shared Data when we begin to migrate collections that utilize that data.

Who is developing the new software?

The Library Technology & Digital Strategies team (Software Engineering, Digital Product Strategy, Web & UX teams, and Core Systems) will work with LITS Middleware and Infrastructure teams. We are also consulting with the Data Curation Experts group, who worked with us on the 2017 ETD re-development project, to develop customized bulk import tools to support our complex migration needs.

I’m a DLP Discovery Phase Working Group alumni/product owner/service manager with ties to the repository. How can I contribute to the development process ahead?

Specific stakeholders were requested to serve on project teams relative to the scope of the phase one implementation effort and the first collections being migrated. As the development teams work on features and migrate initial content, we may also request additional feedback through periodic demos, wireframes, user story requests, recruiting for user testing, and/or feature or migration testing.

How will you keep stakeholders informed during the technical development work?

Each product will have a product manager, a product roadmap, and an ongoing communications plan. We will also provide updates through larger Library Technology & Digital Strategies and Emory Libraries communications plans. Individuals from the Emory community can also sign up for updates through the DLP-UPDATES list-serv by contacting Emily Porter.