

# Approaching MVP<sup>2</sup>

Emory's Discovery and Design Process for  
Minimum Viable Product-Suite + Preservation

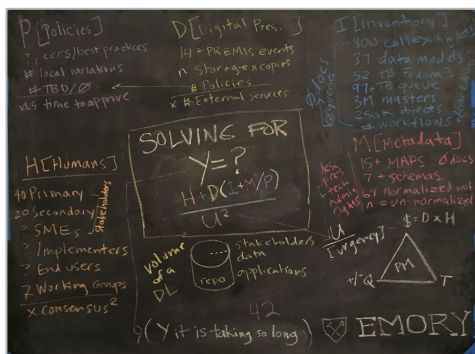
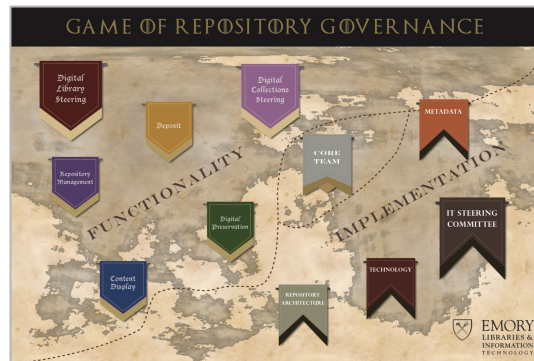
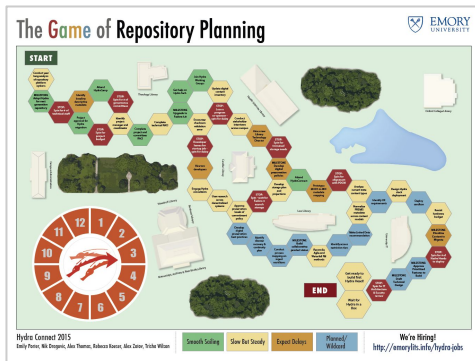


EMORY  
UNIVERSITY

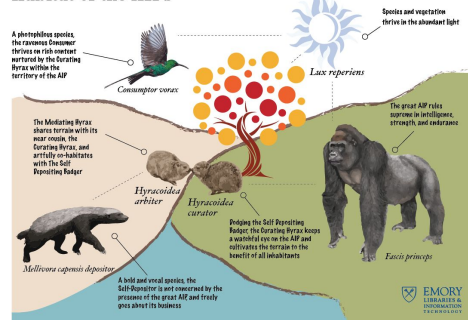
<https://goo.gl/iX3sGD>

Emily Porter  
Rosalyn Metz  
Nik Dragovic  
Collin Brittle

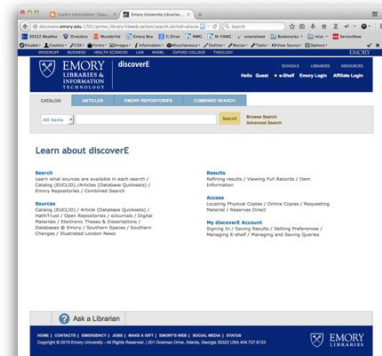
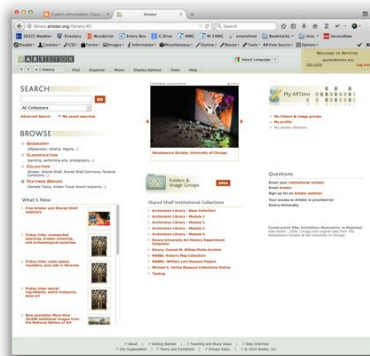
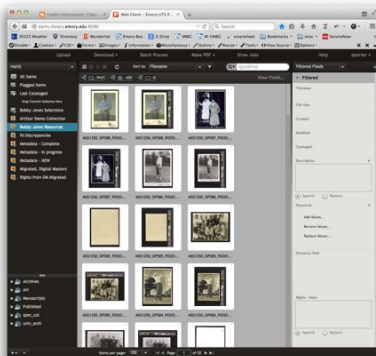
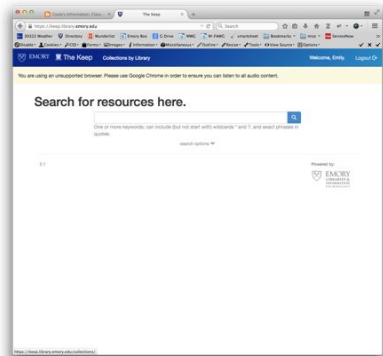
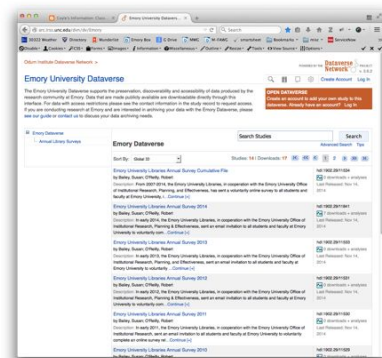
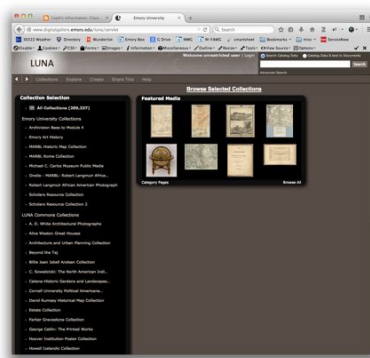
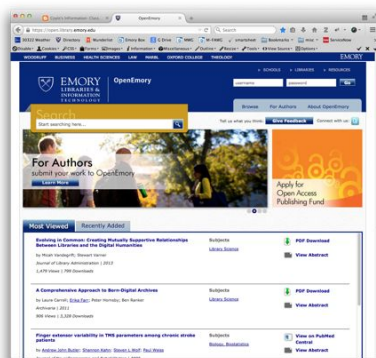
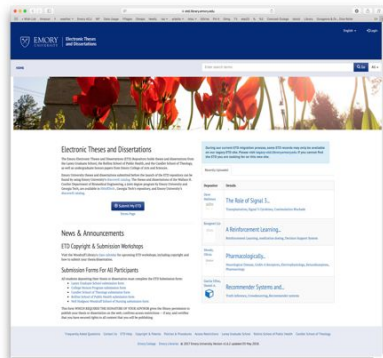
# You may remember us from



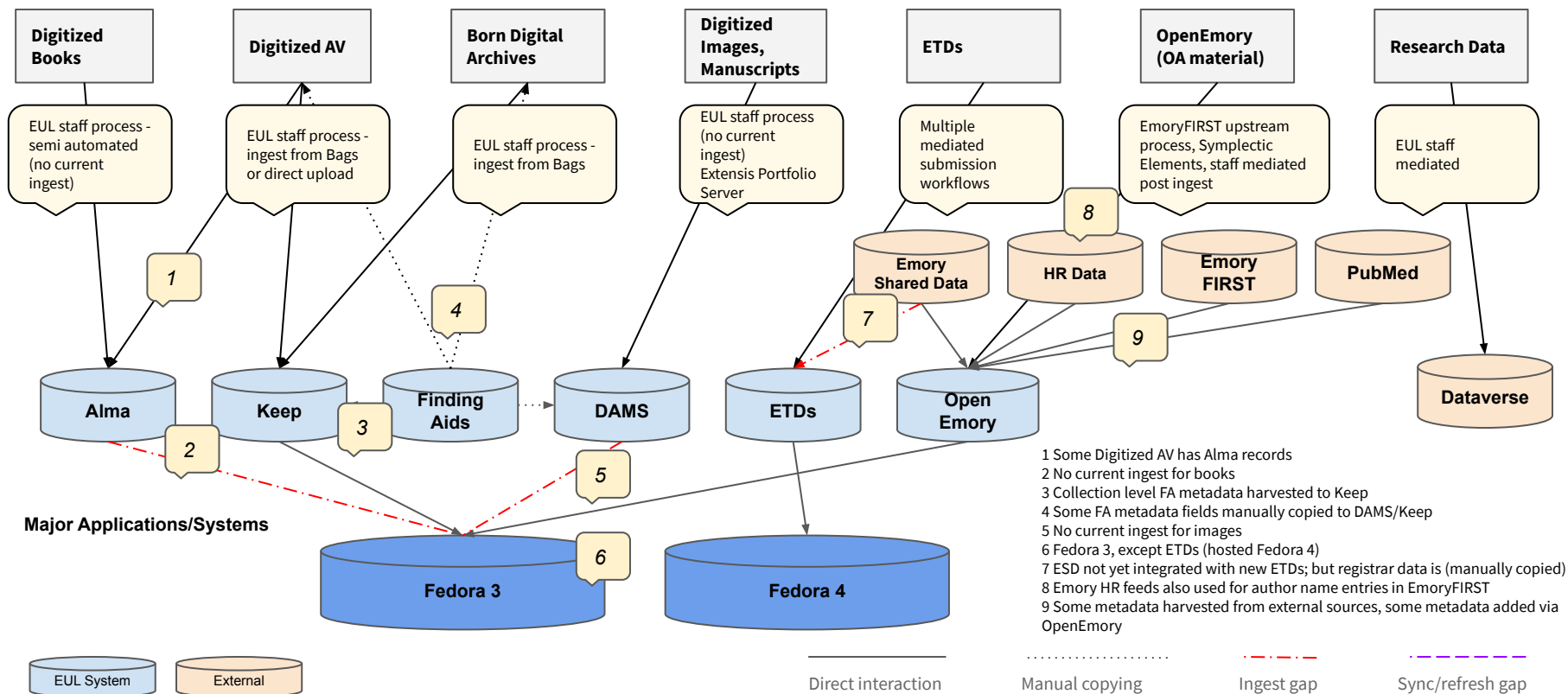
## Habitat of the AIPs



# Our users can't find us



# Our workflows don't all work (or flow)



# Problems to solve

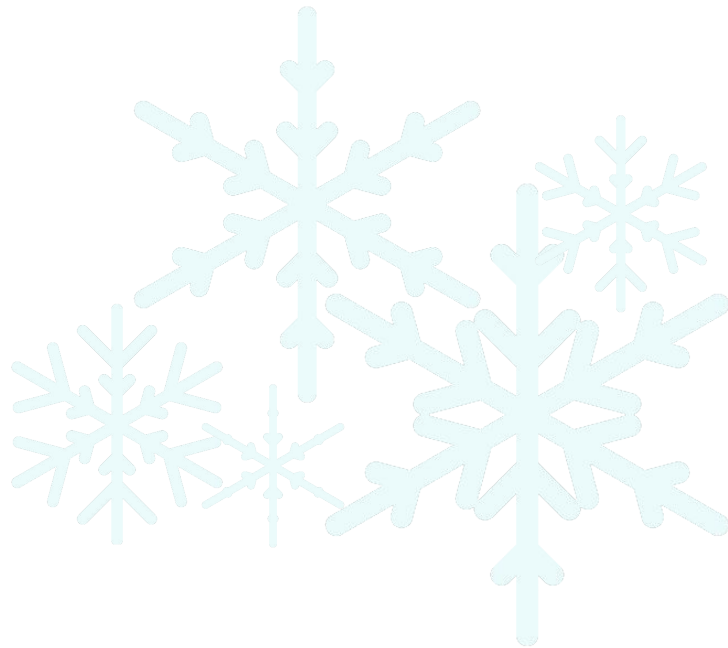
Digital Preservation (aka Long Term Access)

Discovery

Scalability

Culture of silos

***Opportunity: re-architect our repository infrastructure from the ground up - not a 1:1 to current state***



Workin' on it...



Tracy Baim/Windy City Times

# Timeline



# Discovery

What do we need to know in order  
to design a sustainable long-term  
solution?

Nik Dragovic

---

# How we worked

## Who we talked to

- 50+ stakeholders
  - Digital Archives
  - University Archives
  - Scholarly Communications Office
  - Digitization & Digital Curation
  - Metadata Services
  - Collection Management
  - Research Engagement Services
  - Access Services
  - Software Engineering
  - Middleware
  - Enterprise infrastructure
  - Additional Libraries & campus stakeholders
- 20 end users/consumers

## How we interacted

- Formal project management structure
- Governance, scope, charters
- Lots of documentation ([project wiki](#))
- Working Groups (loosely based on OAIS), Core Team, Steering Group
- Digital Collections Steering Committee
- (Slowly... 1.5 years)

# User & Stakeholder Research

## Product and content owner interviews

- 11 sessions: current state product owners
- 9 sessions: current state content owners (in queue for repository ingest)

## User research (end-user/consumers)

- 20 sessions: semi-structured discovery interface reviews
- 4 user segments (undergrad, grad, faculty, library/digital scholarship staff)

## User profiles

- 21 profiles created (related to Content Display, Deposit, Repository Management)

# Functional Requirements Working Groups

## Groups

- Content Display
- Deposit
- Repository Management
- Digital Preservation

## Activities

- Drafted and refined user profiles
- Documented current state processes
- Wrote user stories
- Assessed Samvera product features
- Provided local use cases
- Identified gaps/new feature needs
- Reviewed preservation standards and best practices for local implementation

# Implementation Working Groups

## Groups

- Metadata
- Repository Architecture
- Technology

## Activities

- Inventoried and normalized legacy data models and metadata
- Reviewed Samvera stack & related technologies (PCDM, IIIF, BagIt)
- Development and deployment best practices
- Local IT requirements (authentication, AWS, etc.)

# Operations & Service Planning

## Staffing and re-organizing

- New roles: dedicated program staff
- Software engineering team restructured and dedicated to project
- Additional DevOps team staffing

## Policy development and updates - gatekeeping the repository

- Digital Collection Development Policy (revised)
- 3rd Party Dissemination Policy (new)
- Digital Preservation Policy (revised) and Digital Preservation Strategy (bit-level)
- Retention Policy (new)

## Mission and Vision for the repository as a whole

# Technical Design

How many products do we need to  
build, and what will each do?

Collin Brittle

---

# How we worked

## Who we talked to

- Ourselves! (Working Group leads)
- Software Engineers
- Core Systems/Application Support Team
- Web & UX Team
- Assistant Director, Library Technology
- Middleware Team
- & DCE

## How we interacted

- Loosely structured
- Rolling workshops/"lock-downs"
- Requirements roundup
  - Solution goals/drivers
  - Key integrations
  - Design principles
  - User profiles
- Whiteboarding & scenarios
- Impact criteria - pros and cons
- (Quickly! 3 months, in tandem with closing of Discovery Phase)

# Goals & Drivers

A rare opportunity to redesign the whole ecosystem...

Preservation focus (long term scalability vs. short term/niche solutions)

Abstraction

- Generic AIP/models and viewers
- Content/format agnostic

Re-conceiving our siloed applications as workflows and collections in shared space

Migrating legacy content

# Key Outcomes

System architecture

Samvera-based product suite

Prioritized products and functionality

Business case proposal for implementation

# How Many Hyraxes?

Workshop Agendas and Schedules - ...

Integrations List - Google Sheets

Ground Rules

How Many Hyraxes? - Google

 **How Many Hyraxes?** ☆ 📁

File Edit View Insert Format Data Tools Add-ons Help

All changes saved in Drive

    100% ▾ \$ % .0 .00 123 ▾ Arial ▾ 10 ▾ **B** *I* S A      

*fx*

|    | A      | B  | C       | D     | E   | F         | G    | H |  |
|----|--------|----|---------|-------|-----|-----------|------|---|--|
| 1  | Collin | Zo | Rosalyn | Emily | Nik | Elizabeth | Doug |   |  |
| 2  | 2      | 2  | 2       | 1     | ?   | ?         | ?    |   |  |
| 3  |        |    |         |       |     |           |      |   |  |
| 4  |        |    |         |       |     |           |      |   |  |
| 5  |        |    |         |       |     |           |      |   |  |
| 6  |        |    |         |       |     |           |      |   |  |
| 7  |        |    |         |       |     |           |      |   |  |
| 8  |        |    |         |       |     |           |      |   |  |
| 9  |        |    |         |       |     |           |      |   |  |
| 10 |        |    |         |       |     |           |      |   |  |
| 11 |        |    |         |       |     |           |      |   |  |
| 12 |        |    |         |       |     |           |      |   |  |



?

# Samvera Suite: Another View

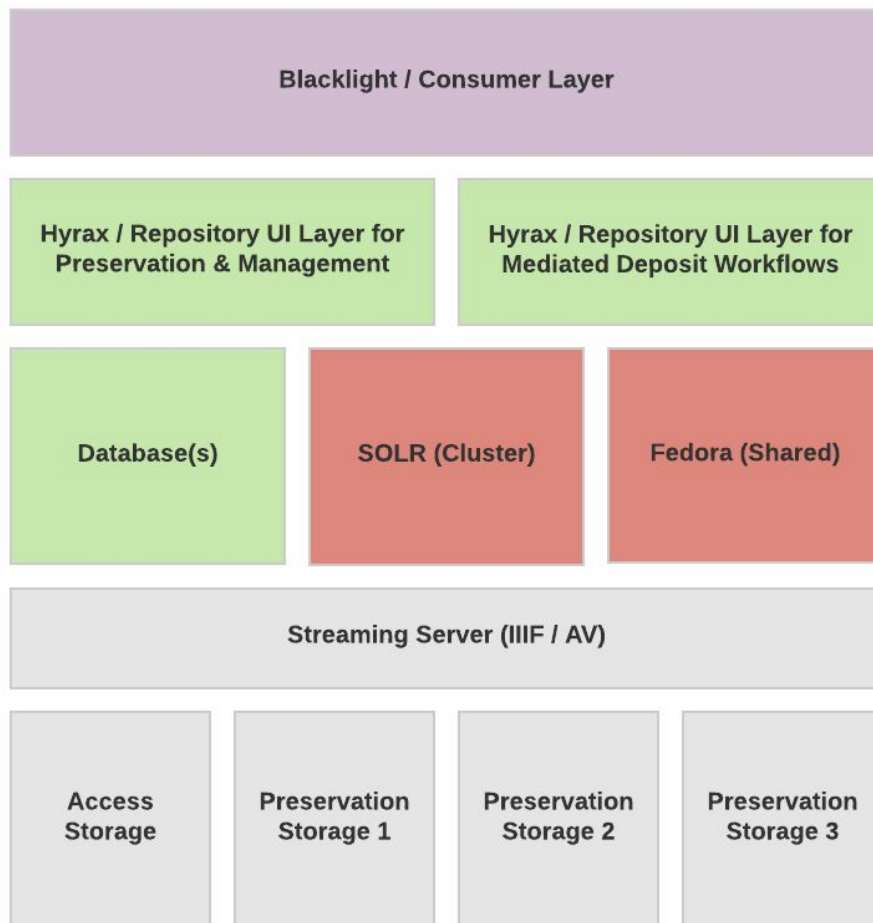
**Custom Blacklight UI** for discovery and delivery

## 2 Hyraxes

- Library Staff vs. Self Depositor use cases
- Shared Fedora, SOLR
- Shared & distinct DB components
- Compatible content models, metadata
- Distinct features/access/workflows

## Preservation & Storage

- Preservation actions & audits
- Externally stored binaries
- Export to 3rd party services



# Content Types vs. Presentation Types

Primary “content types” - we realized they are more about viewers/display/UI needs than they are about data models - the data structure can follow a more abstract pattern

| Type        | Viewer/Interaction                 |
|-------------|------------------------------------|
| Image       | Universal/generic viewer           |
| Text        | Universal/generic viewer           |
| Audio       | Universal/generic viewer           |
| Video       | Universal/generic viewer           |
| Disk Images | TBD; generally staff-only access   |
| Binaries    | Basic file preview only + download |

For our repository front-end UI, we are prioritizing basic use cases:

- Search
- Browse
- Preview
- Download
- Share/Cite

...and offloading specialized format viewer needs to other tools/services

# Digital Object Pattern

## Metadata (Actionable/Indexed)

Descriptive/Rights metadata

Technical metadata (File-level)

Preservation Events/Workflows

Administrative metadata

Structural metadata (PCDM)

*Metadata converted to  
RDF for Hyrax/Fedora 4*

*(May be re-serialized as  
binaries for further  
preservation)*

## Content Files (Primary and Supplemental)

Content file 1 (binary)

Content file 2 (binary)

Content file 3 (binary)

... + additional

... + additional

*The content itself:  
relationships provided in  
structure/metadata  
(PCDM + File Use)*

## Supplemental Preservation Files (Metadata or Administrative Info)

Source Metadata (binary file)

Desc. Metadata record (binary file)

Supplemental PREMIS (binary file)

License/agreement (binary file)

METS (binary file)

*Variable supplemental info  
stored as files (not  
system-readable);  
users can view or download*

*Complex metadata stored as  
binaries (MARC, DDI, etc.)*

# Finding MVP for v. 1.0

What's critical for our first release?

Emily Porter

---

# How we're working

## Who we're talking to

- Ourselves! (Library Technology team)
- Middleware, Storage teams
- Project Management Office
- Collection Managers
- Content curators/SMEs
- Public service desks/discovery SMEs
- Other consultants

## How we're interacting

- Just starting!
- Scrum roles for product development
- Feature development process and tools: epics, acceptance criteria, issues
- Resuming formal project management
- Project/org. roles for policy & migrations
- Product configurations & permissions
- Collection structures
- Migration plans and methods
- Consumer UI/discovery needs

# Write, prioritize, rinse, repeat

- Prioritized products: Curation and Consumer/Discovery (Mediated Deposit app later)
- Backlog and features:
  - Mined the requirements from Discovery Phase
  - Created product suite backlog of epics/stories (275+), tied to our user profiles
  - Categorized stories by product and feature segment
  - Targeting critical features for MVP > roadmap for V.1.0

# V.1 Preservation Features

Support for Emory AIP convention and  
Preservation Policies

Workflow structures: Accessioning, Ingest,  
Versioning, Decommissioning, Deletion

Workflow/Events metadata  
(common properties for all actions)

Persistent identifiers

Tombstones/decommissioning

Tombstones/deletion

Message digest calculation (multiple)

Fixity check

Format identification (FITS)

Validation (FITS)

Metadata extraction (technical/FITS)

Metadata modification

Virus check

Quarantine

Unquarantine

Policy assignment event  
(rights, license, visibility)

Normalization (for selected workflows)

Replication (2- copy: 3-copy to come ASAP)

# Roadmap: Curation & Discovery Themes

1. Discovery, Search and Browse
2. Web Accessibility/Section 508
3. Branding
4. Mobile-friendly, responsive design
5. Citation Building and Export
6. Flexible File Format Support
7. Complex Digital Object Support
8. Support for Relating Content
9. File Previews and Download Options
10. Digital Preservation, Storage & Monitoring
11. Persistent Identifiers and URLs
12. Library Staff Ingest & Migration Utilities
13. Metadata Management
14. Content Dissemination/Export
15. Access Controls and Embargoes
16. Rights Management
17. Digital Asset and Digital Collection Management
18. Submission workflow management
19. User & Permissions Management
20. Reporting and Analytics
21. Data/Security Compliance
22. Streaming Media and IIIF Support

# Minimum Viable Migration

Target: minimum of **5** collections for v.1 launch (hundreds in queue)

Prioritization from our Digital Collections Steering Committee, based on:

- Publicly accessible
- High impact
- Completeness/readiness of collections

Pilot one or more migration tools for legacy systems

# Questions?

Thank you!

## Presentation

<https://goo.gl/iX3sGD>

## Project Wiki

<https://wiki.service.emory.edu/display/DLPP/>

---