

The Hydraulics Project

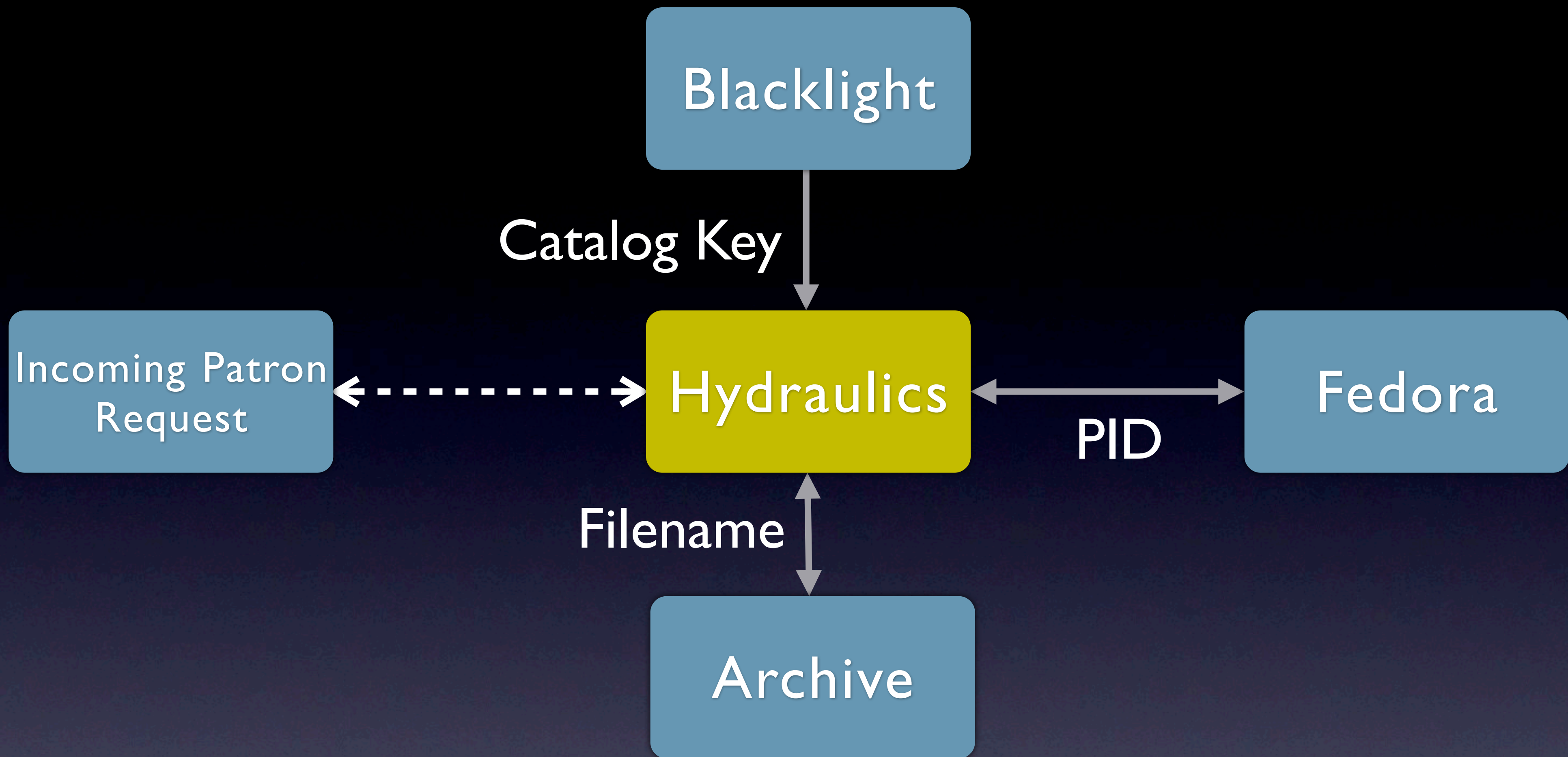
Empowering Communities to Build a Digital Library Utilizing Fedora and an Event-Driven Service-Oriented Messaging Framework

Andrew Curley
University of Virginia Library
Open Repositories 2011

What is Hydraulics?

End-to-end digitization workflow management tool that integrates:

- **Request module**
- **Management system for digital production**
- **Archive management workflows**
- **Patron delivery workflows**
- **Fedora ingestion workflows**



Hydraulics as Connection

Hydraulics serves as the canonical source by facilitating communication with Blacklight, Fedora and an archival filesystem.

Hydraulics vs. Tracksys

Hydraulics

- Generalized, open-sourced, derived from Tracksys

Tracksys

- UVA implementation, tied to local practices and needs

**Hydraulics : Tracksys ::
Blacklight : VIRGO**

Hydraulics Stack

Ruby on Rails

- **ActiveMessaging gem**

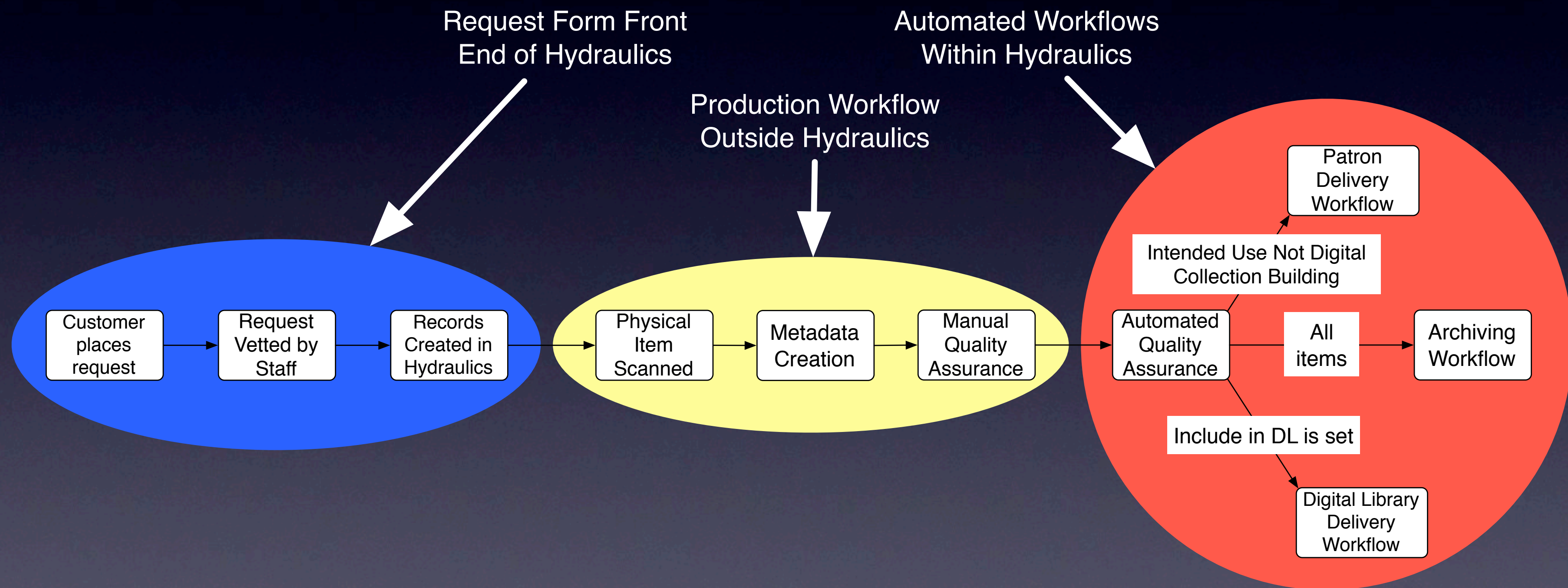
Fedora

- **3.4.2 REST API Compatible**

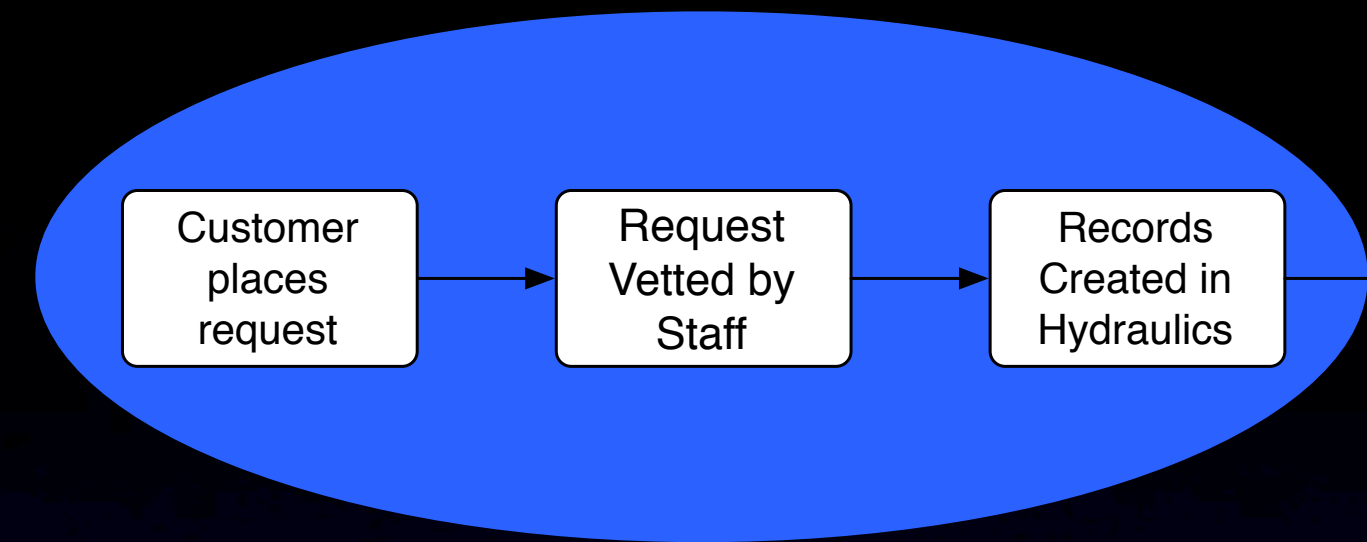
Solr

ActiveMQ

The Hydraulics Model



Request Module



Public facing request form integrates with underlying database

Distinguish internal and external patrons

Facilitates engagement between Special Collections and Preservation/Conservation staff and patron

Allows librarians to link digital objects with canonical metadata

Fee and billing management

Request Form: User Information

Integration with local LDAP

Populate with existing information

Request can be made on behalf of
another person

Billing address information

Digitization Services Request Form

** indicates required field*

Are you submitting this request for someone else? [?]

Contact Information

* Email:

Send a copy of request confirmation to: [?]

Address

* First Name:

* Last Name:

Organization:

* Address 1:

Address 2:

* City:

State: [?]

* Country: [?]

Postal Code:

Phone:

different billing address

How did you hear about our services?

How did you hear about our services?

different billing address

Request Form: Bibliographic Info

Multiple free text fields

Flexible, non-mapped fields

Intended to guide staff in locating
material

Based on years of experience
handling patron request of Special
Collections materials

Items to Digitize Add another item

Item 1

Is this item owned by U.Va. Library? Yes No

For Library-owned material, please supply one or more of the following identifiers:

Call/accession number:

Copy number:

Volume:

Issue:

Location:

Title:

Name of the creator (author, photographer, etc.) of this item:

Year:

Describe this item:

If you saw this item on the web, please provide the URL:

How many pages/images will be digitized? or all

Special Instructions:

How did you hear about this resource?

How did you hear about this resource?

Request Form: Intended Use

Intended uses govern type of deliverable given to patron

Guide user in choosing their deliverable type rather than explicitly asking for technical specifications

Two kinds of deliverables:

- 300dpi watermarked JPEG
- Highest possible dpi TIF

Deliverable Information


Based on your chosen intended use, we will provide an explanation of the deliverable will appear.

* What is the intended use for this item? Online Exhibit

Format: JPG
Resolution: 300dpi

Your images will include a copyright statement. The text states:
"Under 17USC, Section 107, this single copy was produced for the purposes of private study, scholarship, or research. No further copies should be made. Copyright and other legal restrictions may apply. Special Collections, University of Virginia Library."

Classroom Instruction
 Digital Archive
 GIS Processing
 Online Exhibit
 Personal Research
 Physical Exhibit
 Presentation
 Print Publication (academic)
 Print Publication (non-academic)
 Sharing with Colleagues
 Web Publication



Under 17USC, Section 107, this single copy was produced for the purposes of private study, scholarship, or research. No further copies should be made. Copyright and other legal restrictions may apply. Special Collections, University of Virginia Library.

[Add another item](#)

I understand and agree that I will only use the digitized resources as I indicate above in my request. I also understand it is my responsibility to determine who owns the copyright of these materials unless otherwise noted.

[Submit](#) [Cancel](#)

[Submit](#) [Cancel](#)

I understand and agree that I will only use the digitized resources as I indicate above in my request. I also understand it is my responsibility to determine who owns the copyright of these materials unless otherwise noted.

Request Approval: Bibliographic Records

Integration with Blacklight

Association accomplished by catalog key
and/or barcode

Barcode is only unique value!

Bibliographic records must be “item”
records; local practice is to make
“manifestation” records, with “item”
information in MARC 999 field

Title
Novvelles inventions povr bien bastir et a petits fraiz, trovvees n'Agveres .

Citation

Description

Series title

Creator name
L'Orme, Philibert de, 1515?-1570

Creator name type
personal

Catalog key
u1938996 [i](#) [Get metadata values from U.Va. Library catalog](#)

Title control
AJL6724

Barcode
X030078580 [i](#) [Get metadata values from U.Va. Library catalog](#)

Call/accession number
NA2517 .D4 1576

NA2517 .D4 1576
Call/accession number

Request Approval: Routing Slip

Printed by staff after order vetting is complete

Provides workflow template and order metadata to production staff

Allows for notes or problems to be passed along throughout production

Allows production coordinators to manage priorities at a glance



1 of 13 units

Call Number: NA2517 .D4 1576

Location: SC-STKS-F

Equipment: _____

Order Information

Customer Name: Andrew Curley

Order Number: 1265

Due Date: 12/1/2007

Project Name: Curley 1

Customer Status: Staff

(4 wks = 12/8/2007)

Unit Information

Unit Number: 305

Pages to be scanned: _____

Files from previous unit?: _____

Special Instructions / Pages to be scanned:

This unit represents an item that Andrew has added to the order system after the fact to represent items already in process.

Items to scan: Entire Book

Production Workflow

Materials Ready: _____

1. In process scanning: _____

2. Crop and Rotate: _____

3. Process: _____

4. Scan Covers: _____

5. Build Catalog: _____

6. 1st QA: _____

7. Create Metadata: _____

8. Rescans and Corrections: _____

9. 2nd QA: _____

10. Final QA 1: _____

11. Final QA 2: _____

12. Finalized: _____

13. Delivered/Archived: _____

Deliverables

Intended Use: Digital Collection Building

Location: _____

Transcription Format: _____

Resolution: _____

Format: _____

Transcription Format: _____

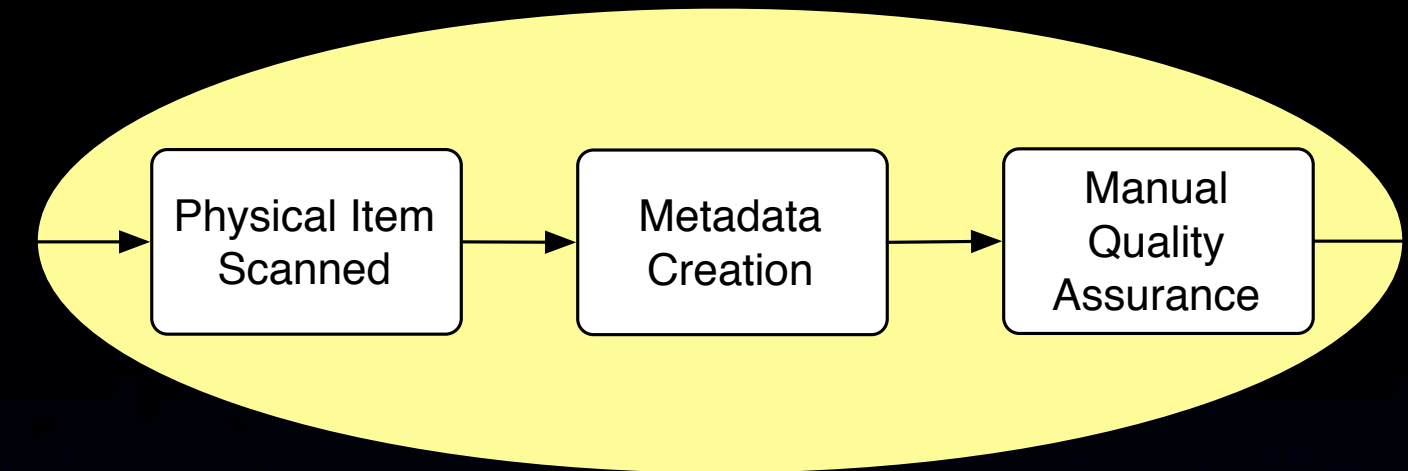
Location: _____

Intended Use: Digital Collection Building

Format: _____

Resolution: _____

Production Workflow



Workflows local to institutions, based on available equipment, computing and software

UVA Digitization Services uses:

- **Phase One Capture One DB**
- **MS Expression Media (soon-to-be Phase One Media Pro, also formerly Iview)**

Expression Media used for quality assurance and metadata entry

Production Workflow: Metadata Entry

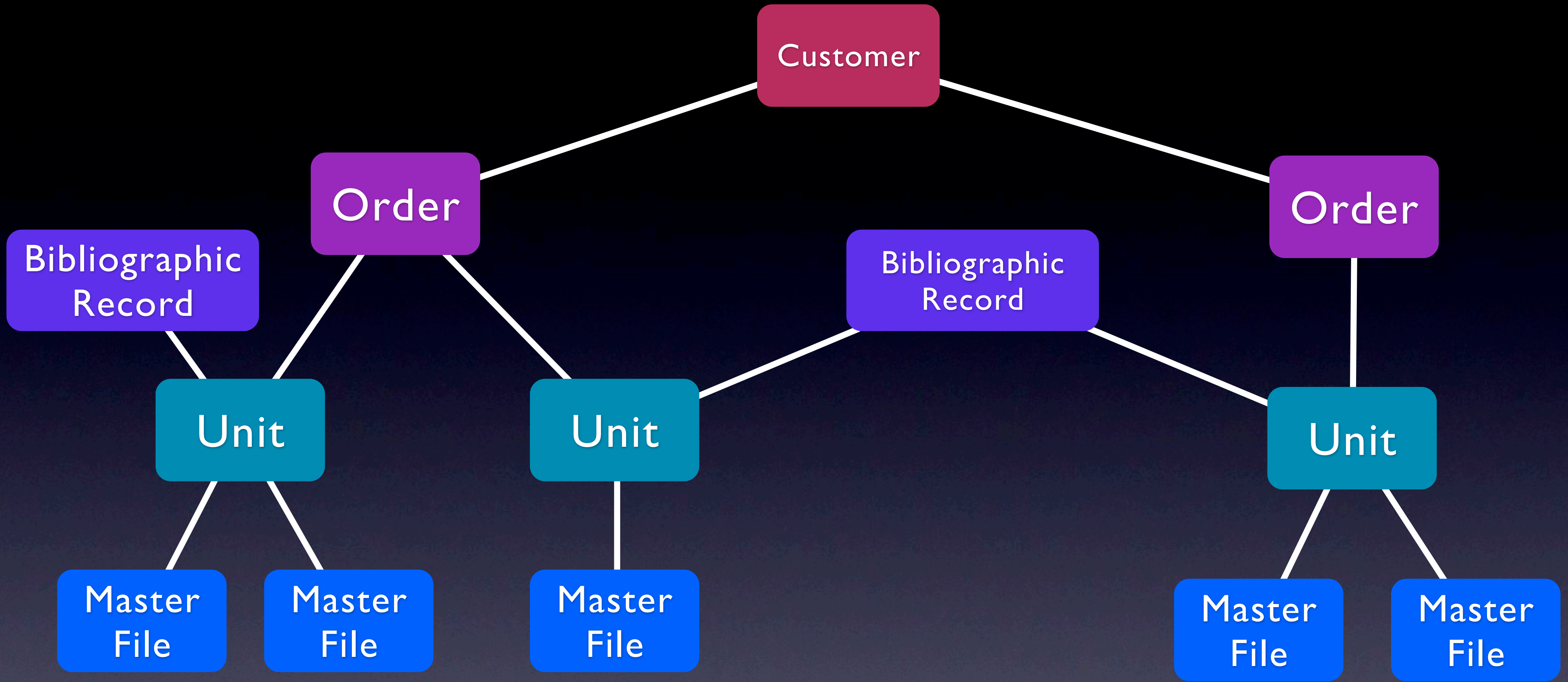
The screenshot displays a software interface for digital asset management. The main window shows a thumbnail of a historical document page with the title "NOUVELLES INVENTIONS POUR BIEN BASTIR ET A PETITS FRAIZ, TROUVES". The document is framed by an ornate border. Below the thumbnail, a metadata entry form is visible, containing the following information:

- File Name: 000000305_0010.tif
- Title: Title Page
- Caption: -
- Author: L'Orme, Philibert de
- Source: Nouvelles inventions pour bien bastir et a petits fraiz, trov...
- Credit: NA2517 .D4 1576

The interface also includes a left sidebar with "Catalog Fields" and "Catalog Folders" sections. The "Catalog Fields" section lists metadata fields such as Author, Credit, and Source, each with a value and a count of 122. The "Catalog Folders" section lists folders like "digiserv-production" and "DSSR". The top of the window shows a search bar and various navigation icons. The bottom status bar indicates "Catalog 14.7 MB | Media 7.1 GB" and "122 Items in list".

Production Workflow: Metadata Export

```
<MediaItem>
  <AssetProperties>
    <Filename>000001483_0134.tif</Filename>
    <Filepath>000001483:000001483_0134.tif</Filepath>
    <UniqueID>427</UniqueID>
    <Label>0</Label>
    <Rating>0</Rating>
    <MediaType>TIFF</MediaType>
    <FileSize unit="Bytes">48510315</FileSize>
    <Created>2008:08:27 15:08:47</Created>
    <Modified>2008:11:13 15:29:34</Modified>
    <Added>2008:08:27 15:18:45</Added>
    <Annotated>2008:11:14 15:55:28</Annotated>
  <ThumbnailSource>000001483_0134.jpg</ThumbnailSource>
</AssetProperties>
<MediaProperties>
  <Width unit="Pixels">3320</Width>
  <Height unit="Pixels">4862</Height>
  <Resolution unit="DPI">600</Resolution>
  <Depth unit="Bits">24</Depth>
  <ViewRotation>1</ViewRotation>
  <SampleColor>R:DO G:DO B:C0</SampleColor>
  <Pages>1</Pages>
  <ColorSpace>RGB </ColorSpace>
  <Compression>65537</Compression>
  <PrimaryEncoding>TIFF (Uncompressed)</PrimaryEncoding>
  <ColorProfile>Adobe RGB (1998)</ColorProfile>
</MediaProperties>
<AnnotationFields>
  <Headline>119</Headline>
  <Author>Phillips, David Graham</Author>
  <Credit>Taylor 1917 .P55 S8</Credit>
  <Source>Susan Lenox, her fall and rise, with a portrait of the author</Source>
  <Location>SC-STKS</Location>
</AnnotationFields>
<MetaDataFields>
  <Maker>Phase One</Maker>
  <Model>P 45+</Model>
  <Software>Adobe Photoshop CS2 Macintosh</Software>
  <SourceURL></SourceURL>
  <ExifVersion>2.2</ExifVersion>
  <CaptureDate>2008:08:22 13:29:23</CaptureDate>
  <ISOSpeedRating>50</ISOSpeedRating>
  <ExposureBias>+0.0</ExposureBias>
  <ExposureTime>1/15</ExposureTime>
  <Aperture>f11.0</Aperture>
  <FocalLength>120.0</FocalLength>
</MetaDataFields>
</MediaItem>
```

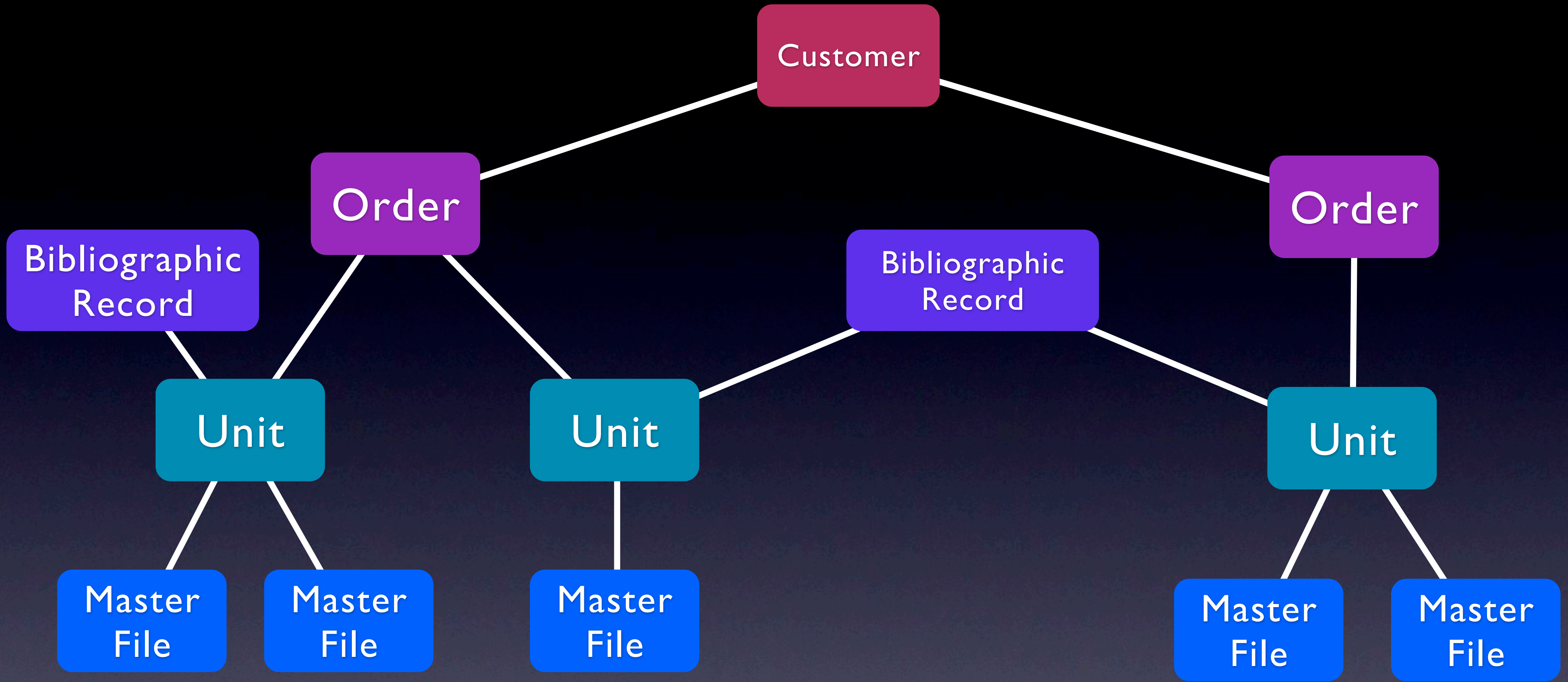



Hydraulics Core Database Layout

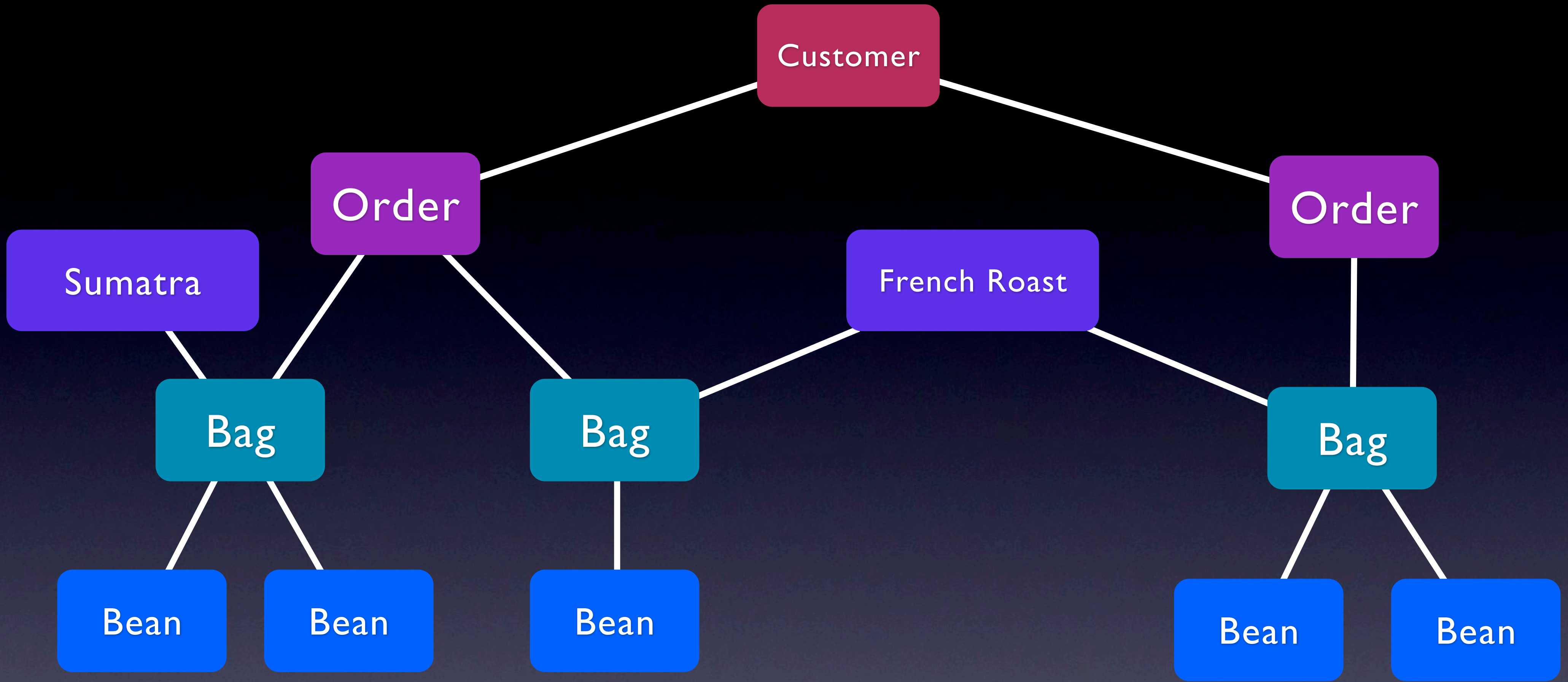


<http://www.flickr.com/photos/niallkennedy/1780421312/>

Units As Coffee Beans



Hydraulics Core Database Layout

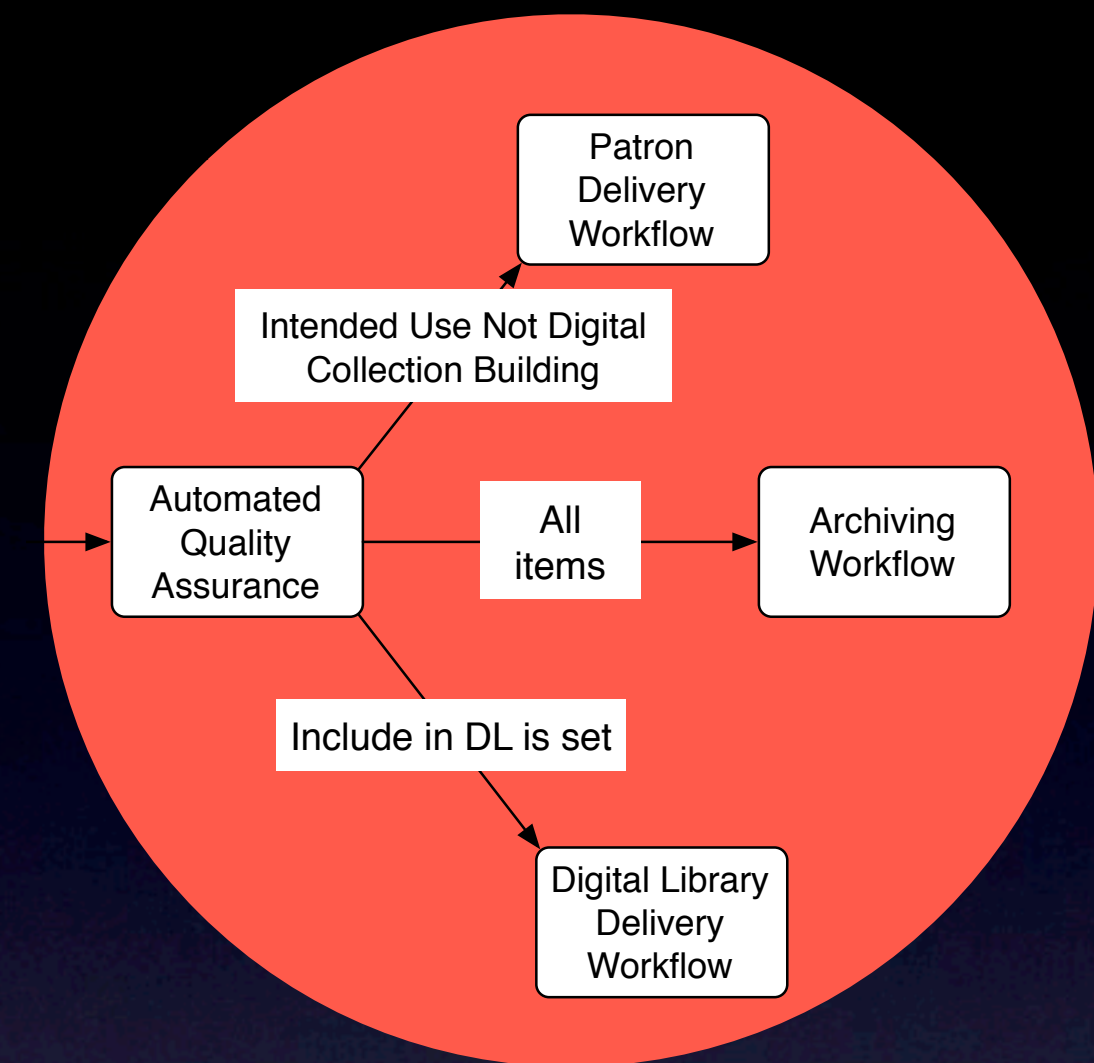


Hydraulics Core Database Layout (Using Coffee Analogy)

Finalization Workflows

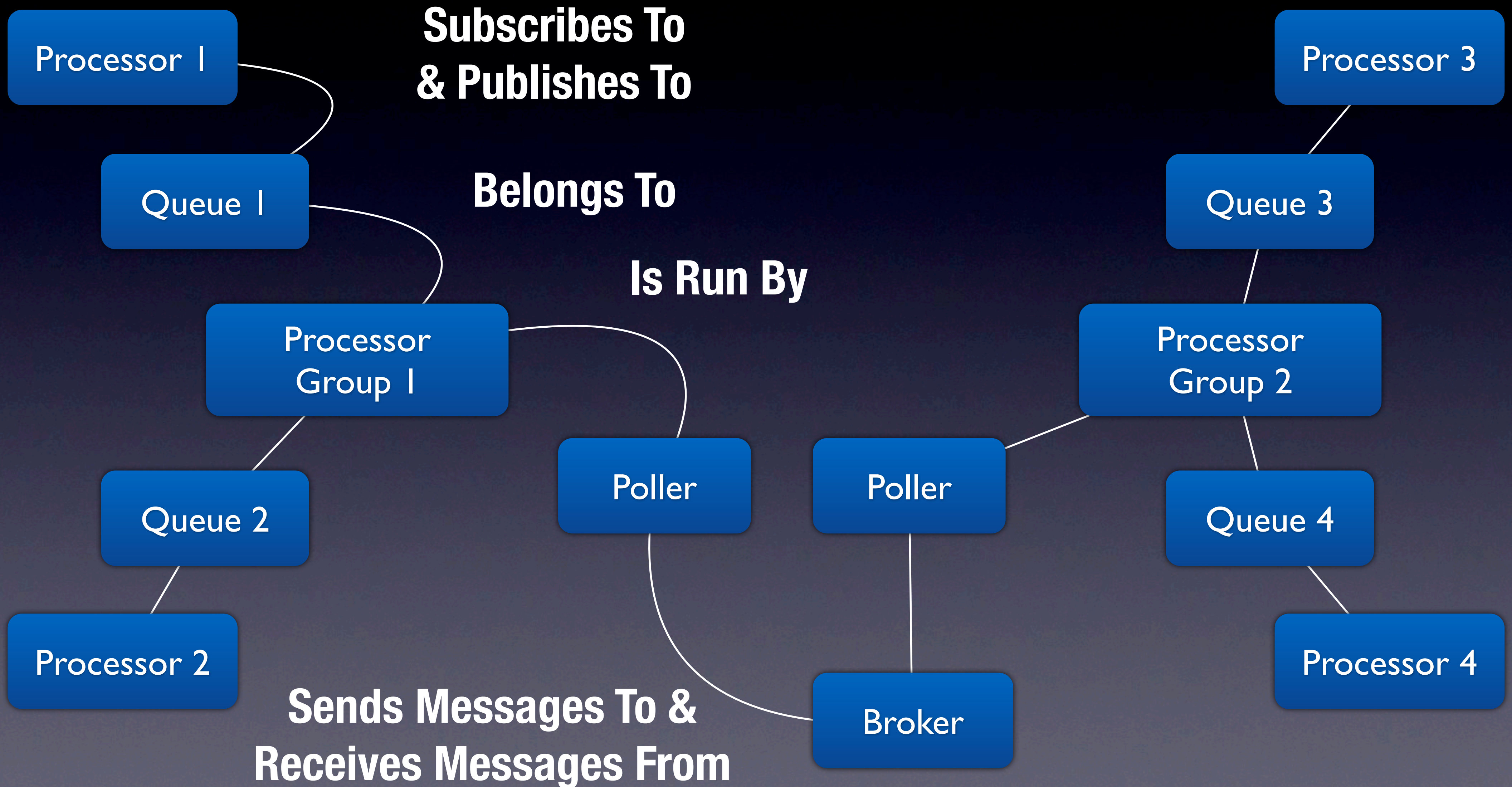
Three distinct workflows:

1. Archiving
2. Patron Delivery
3. Digital Library Delivery

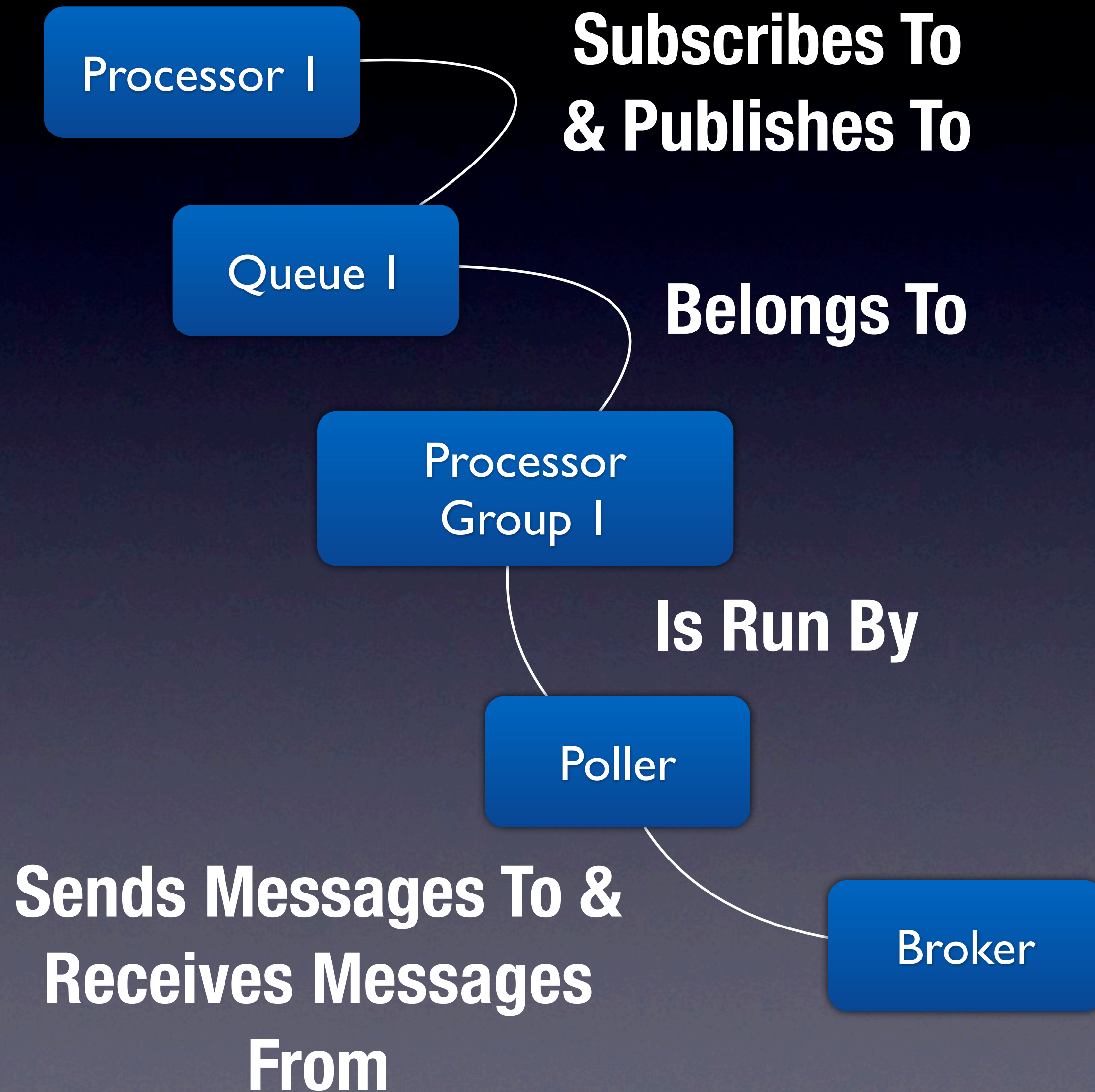


Decisions that govern what workflows each Unit undergoes are data-driven

Messaging in Ruby on Rails



Messaging in Ruby on Rails



- **ActiveMessaging gem**
 - **Creates Processor class**
 - **Defines queues in config/messaging.rb**
 - **Controls processor message handling**
- **Daemons gem**
 - **Runs Ruby processes in background**
 - **Similar to Linux 'service' command**
- **ActiveMQ**
 - **Messaging broker implementing JMS**
 - **Message Persistence**

Logging Messages in Hydraulics: AutomationMessageClass

Write 'success', 'failure' and 'error' messages to database

Associated to an Order, Unit, Master File or Bibliographic record

Error messages are flagged; display on administrative home page; contain both stack trace and diagnostic message

Provide persistent audit trail of completed work

Tracksys has processed 2.2 million messages as of 6/10/2011

Finalization Workflows: Archiving

All Units archived to HSM for long-term preservation

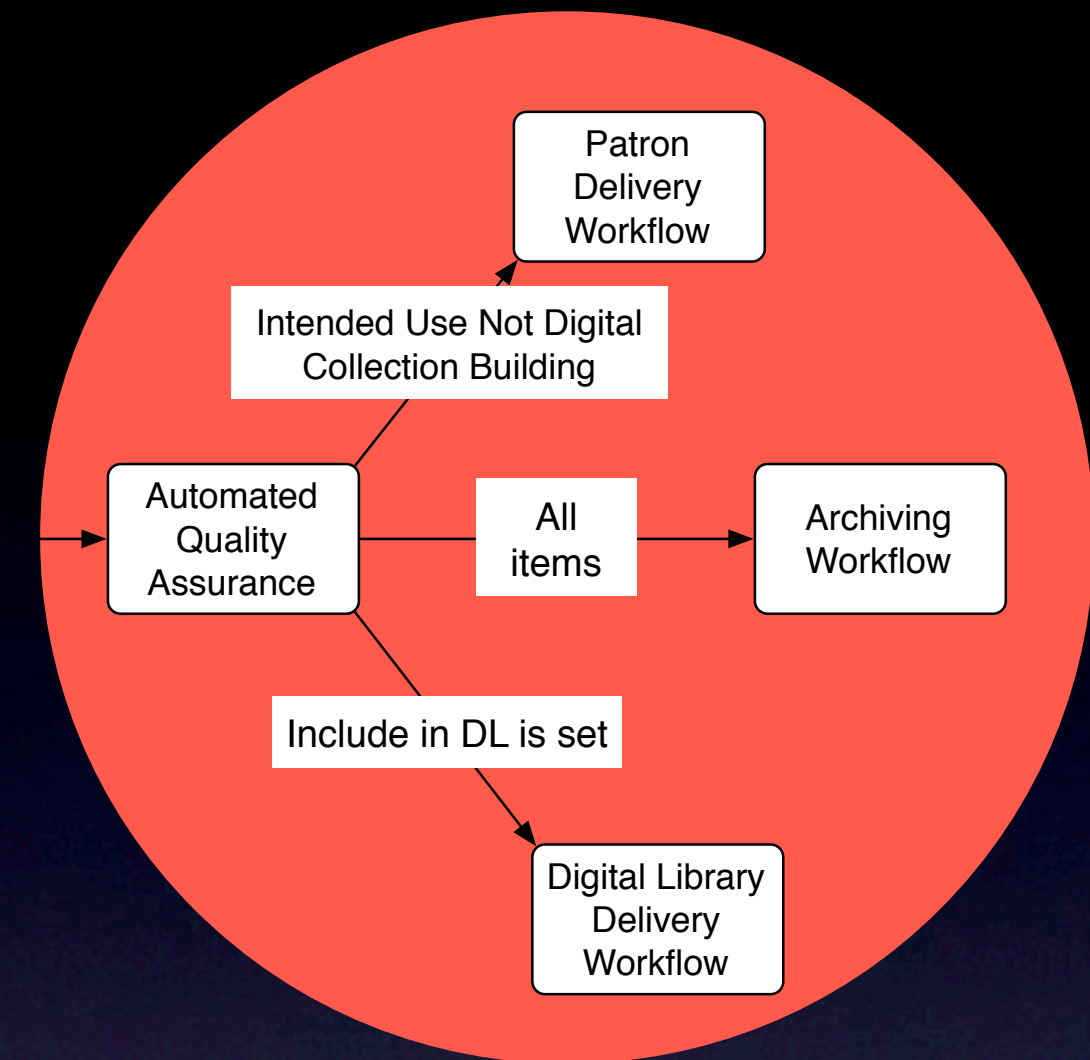
- **Quantum Stornext**

Checksums created

File paths systemized

One-click download of files for future redelivery

Able to ingest content into Fedora repository directly from archive



Finalization Workflows: Patron Delivery

Deliverable images created on a Unit-by-Unit basis

- **Type determined by intended use**

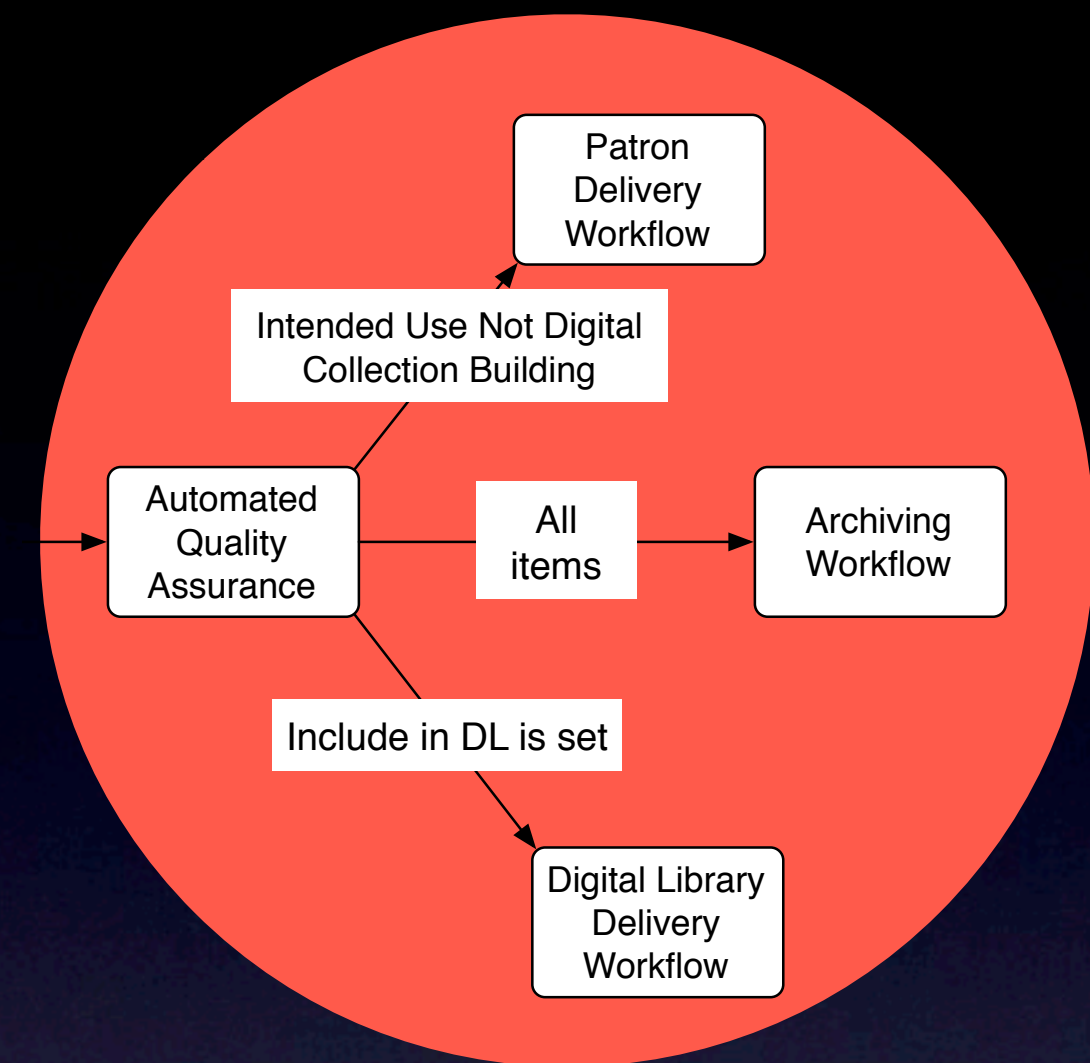
Manifest of digital objects and invoice (PDF-format)

Zip archive of deliverables made web accessible

Email sent to customer providing URL for image pickup

Optional DVD pickup; requires staff override

Automated server cleanup



Finalization Workflows: Digital Library Delivery

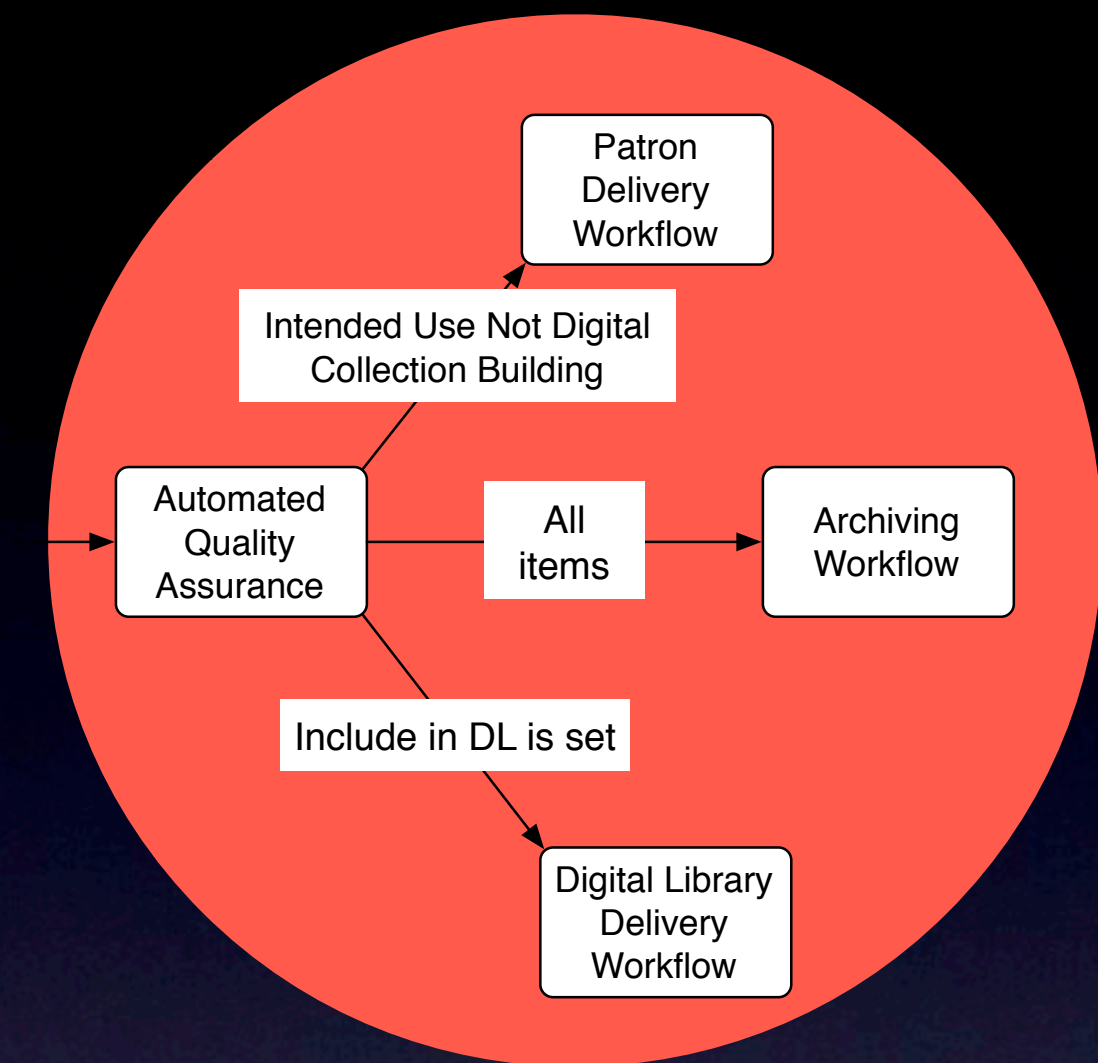
Atomistic model, datastreams match Hydra guidelines

Fedora Objects will be created for:

- Bibliographic Records
- Master File

Object creation through API calls; No FOXML

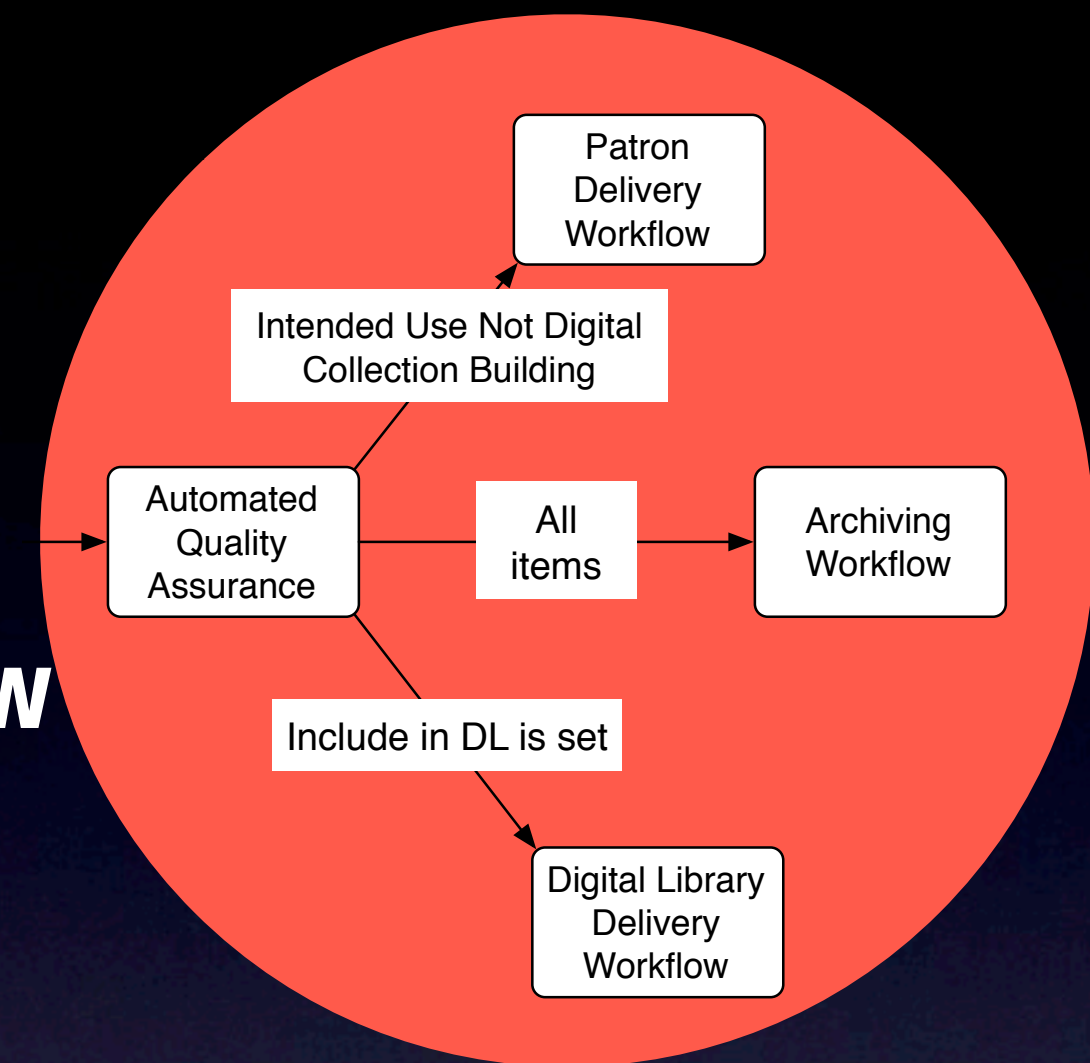
Flexible access rights and 'discoverability' set at Unit, inherited by Master File and Bibliographic objects.



Finalization Workflows: Digital Library Delivery

Datastreams of bibliographic record object:

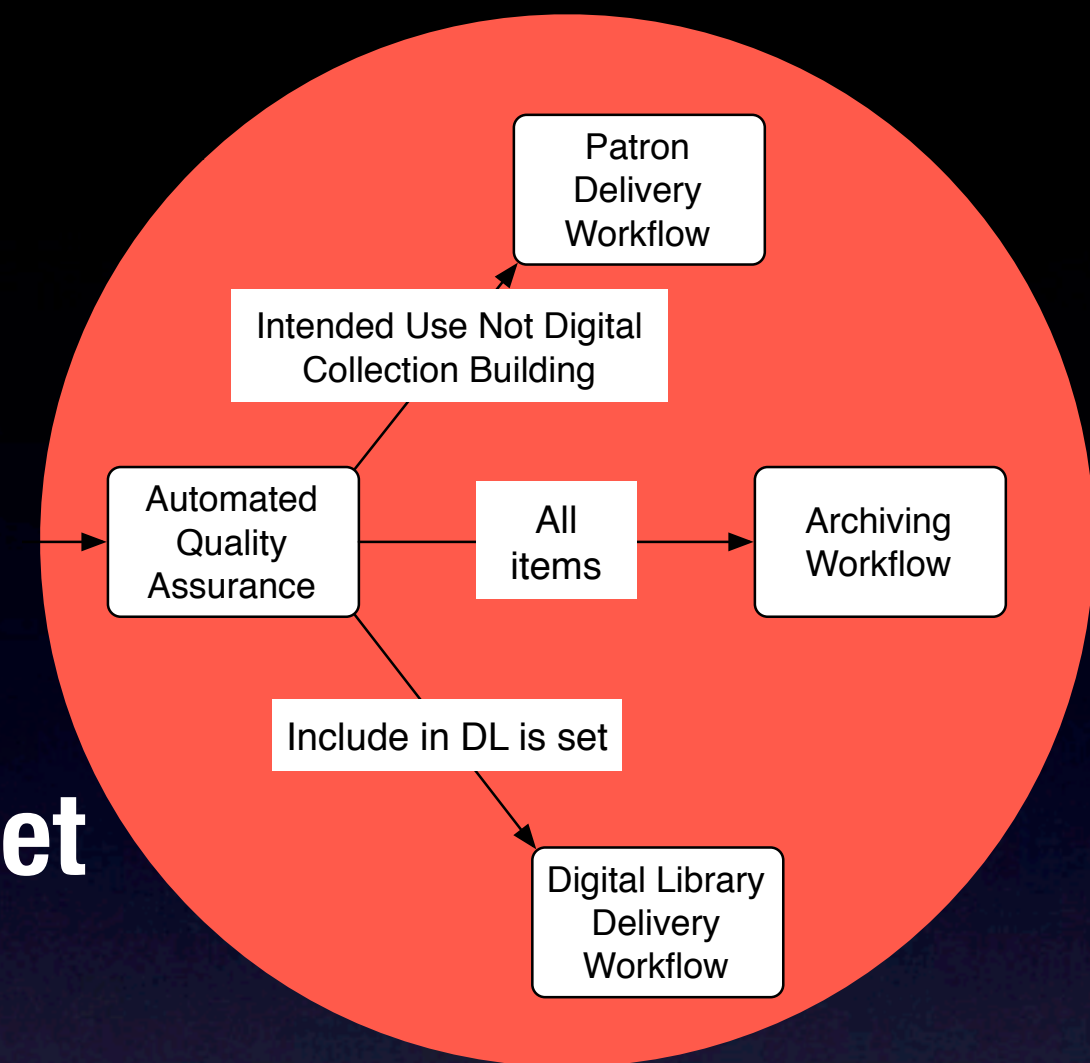
- **MARC - External reference to Blacklight MARC XML view**
- **descMetadata - Transformation of MARC XML to MODS. LC stylesheet used, with local modification.**
- **DC - Dublin Core generated by MODS-to-DC LC stylesheet**
- **solrArchive - Record of solr <add> doc. Generated by custom and parameterized MODS-to-Solr XSLT**
- **POLICY and rightsMetadata - XACML policy**
- **RELS-EXT and RELS-INT**

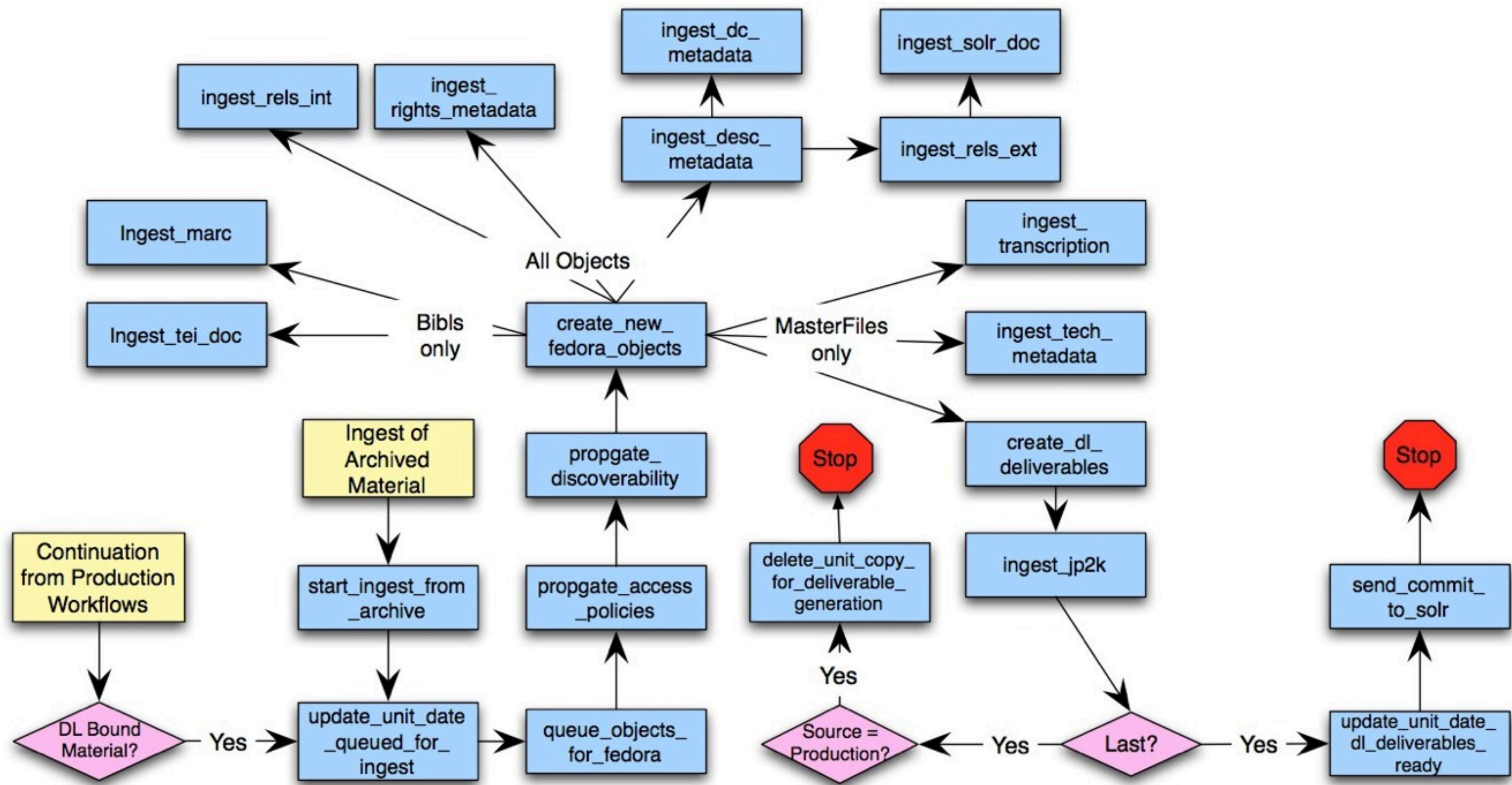


Finalization Workflows: Digital Library Delivery

Datastreams of Master File objects

- descMetadata - MODS created from data in Hydraulics
- DC - Dublin Core generated by MODS-to-DC LC stylesheet
- solrArchive - Record of solr <add> doc. Generated by custom and parameterized MODS-to-Solr XSLT
- POLICY and rightsMetadata - XACML policy
- RELS-EXT and RELS-INT
- content - Binary JPEG2000
- technicalMetadata - MIX generated from Hydraulics.





Fedora Ingestion Workflow

Demonstration of Ingestion

Empowering Communities: Production

Automated delivery and quality assurance saves staff time; more focus on digital production

More efficient access to preservation archive

Easier redelivery of already digitized content; less harm to material

Empowering Communities: Selection

Subject specialist, librarians and scholars do the selection of content by placing requests

Demand-driven repository development

Newly accessed collections can be added to digitization queue shortly after cataloging

Flexibility to select content not originally requested for the digital library

Empowering Communities: Object Creation & Maintenance

By brokering all ingestion through a messaging service:

- **Metadata changes are now effectively done with a click of a button**
- **Policy/Access rights can be changed swiftly in response to IP violations or concerns**
- **Poor quality images can be replaced as soon as they are rescanned and re-archived**
- **New architectures can be instantiated quickly**

Additional Functionality

The following topics were not covered in this presentation, but are part of Hydraulics:

- 1. EAD metadata: Associating Master Files with Component**
- 2. Non-EAD MSS metadata: Associating Master Files with Box/Folder-level information provided by archivists at the time of digitization**
- 3. Associating transcription of images with Master File records; methods of ingesting and indexing this content**
- 4. User access and privileges**
- 5. Extensibility of Master File class beyond TIF images**
- 6. Statistical analysis of production metrics (for you managers out there....)**
- 7. Hand-crafted MODS, RDF, SOLR can be provided to ingestion workflow for all object types**

Next Steps

- 1. Community outreach and engagement**
- 2. Possible integration with the Hydra initiative**
- 3. Redesign code for Rails 3**
- 4. Expand Master Files to handle A/V material**
- 5. Develop and release app as Rails plugin**

Andrew Curley

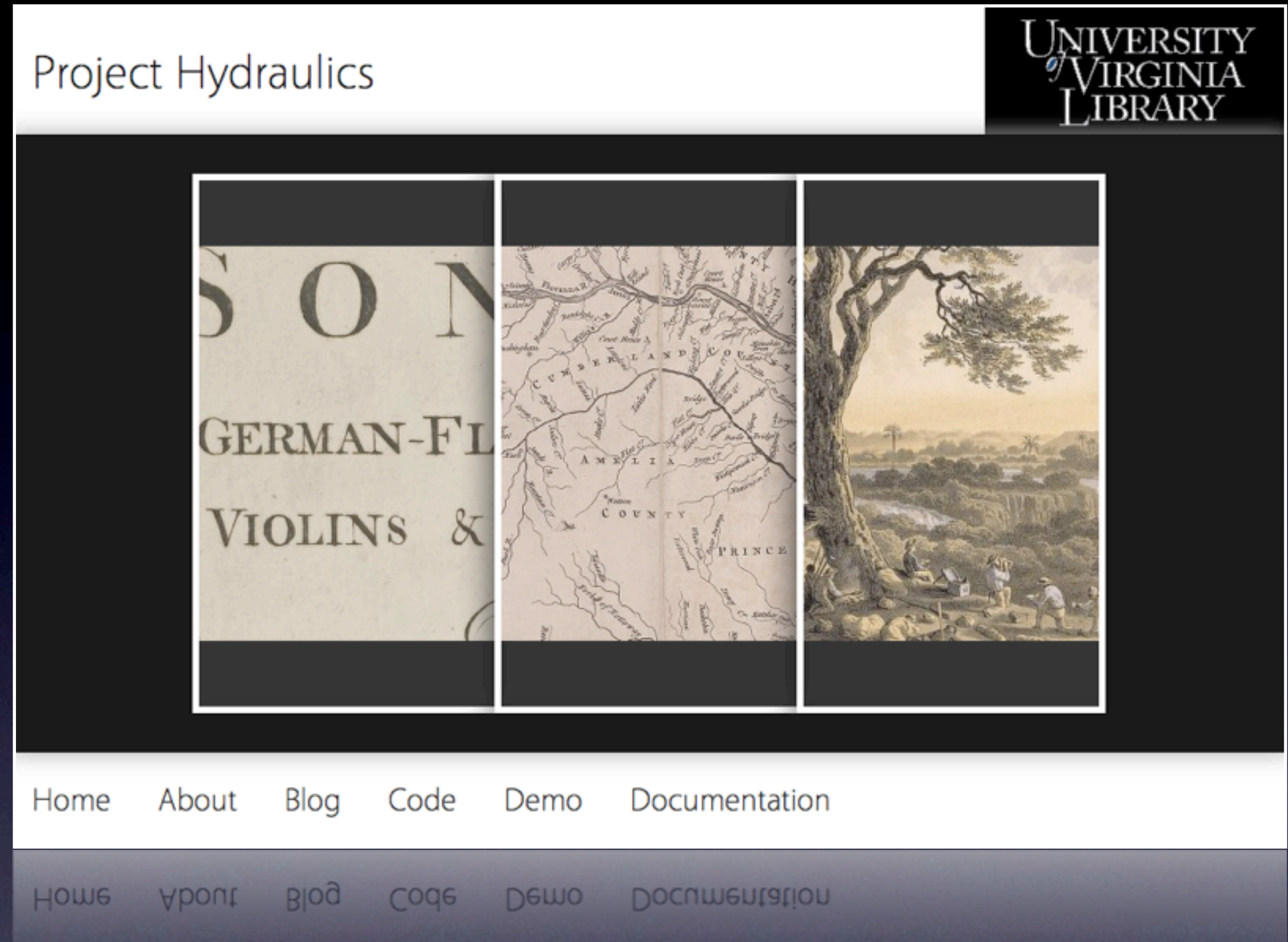
University of Virginia Library

Digital Curation Services

andrew.curley@gmail.com

@andrewcurley

#projecthydraulics



<http://projecthydraulics.org/>
<http://demo.projecthydraulics.org/request>
<http://demo.projecthydraulics.org/admin>

<https://github.com/uvalib/hydraulics>