OSSArcFlow

Investigating, Modeling, and Synchronizing a Range of Archival Workflows for Born-Digital Content

ABOUT THE PROJECT

The Educopia Institute, in collaboration with the University of North Carolina at Chapel Hill School of Information and Library Science (UNC SILS), LYRASIS, and Artefactual, Inc., launched the Open Source Software Archival Workflow (OSSArcFlow) research project in July 2017. This two year project draws together 12 partner institutions from different cultural heritage sectors to investigate, synchronize, and model the use of OSS tools in born-digital workflows. The project is funded by the Institute for Museum and Library Services (IMLS) and aims to increase institutional capacity for curating born-digital content.

AS-IS WORKFLOWS

In the fall of 2017, the project team worked with partners at each institution to mockup a visual representation of their current workflow activities. Representing a “snapshot in time,” these documents show how a diverse group of institutions are using OSS tools in their workflows to curate born-digital content. They also provide an essential starting point for synthesizing and comparing both the gaps and overlaps that currently exist between common OSS tools and environments.

EDUCOPIA
INSTITUTE

As-Is Digital Curation Workflow, August 2017

Woodruff Library, Atlanta University Center

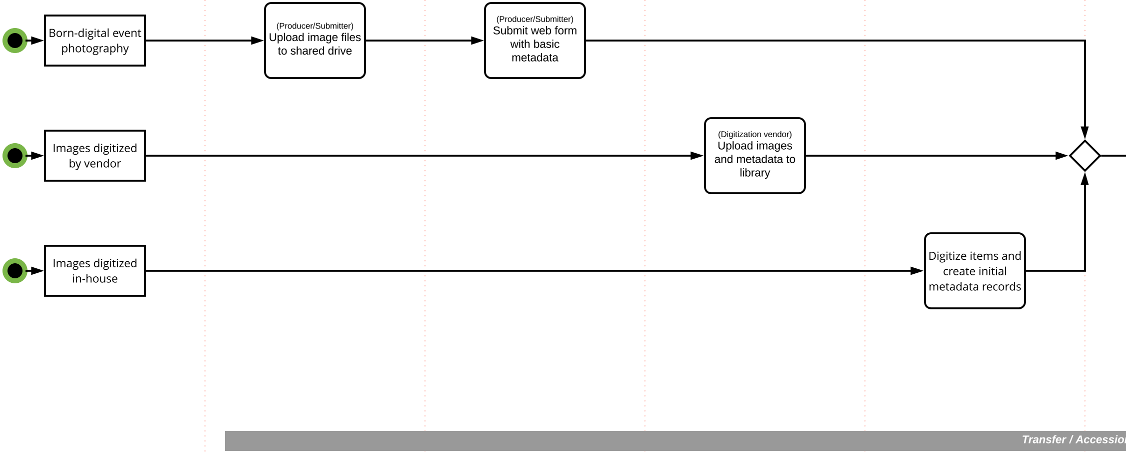
OssArcFlow Project

https://educopia.org/research/ossarcflow

ArchivesSpace

archivematica

BitCurator



Born-digital event photography

(Producer/Submitter) Upload image files to shared drive

(Producer/Submitter) Submit web form with basic metadata

Images digitized by vendor

(Digitization vendor) Upload images and metadata to library

Images digitized in-house

Digitize items and create initial metadata records

Transfer / Accessio

Metadata and Digital Resources Librarian

Digital Services Department staff

Processing archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

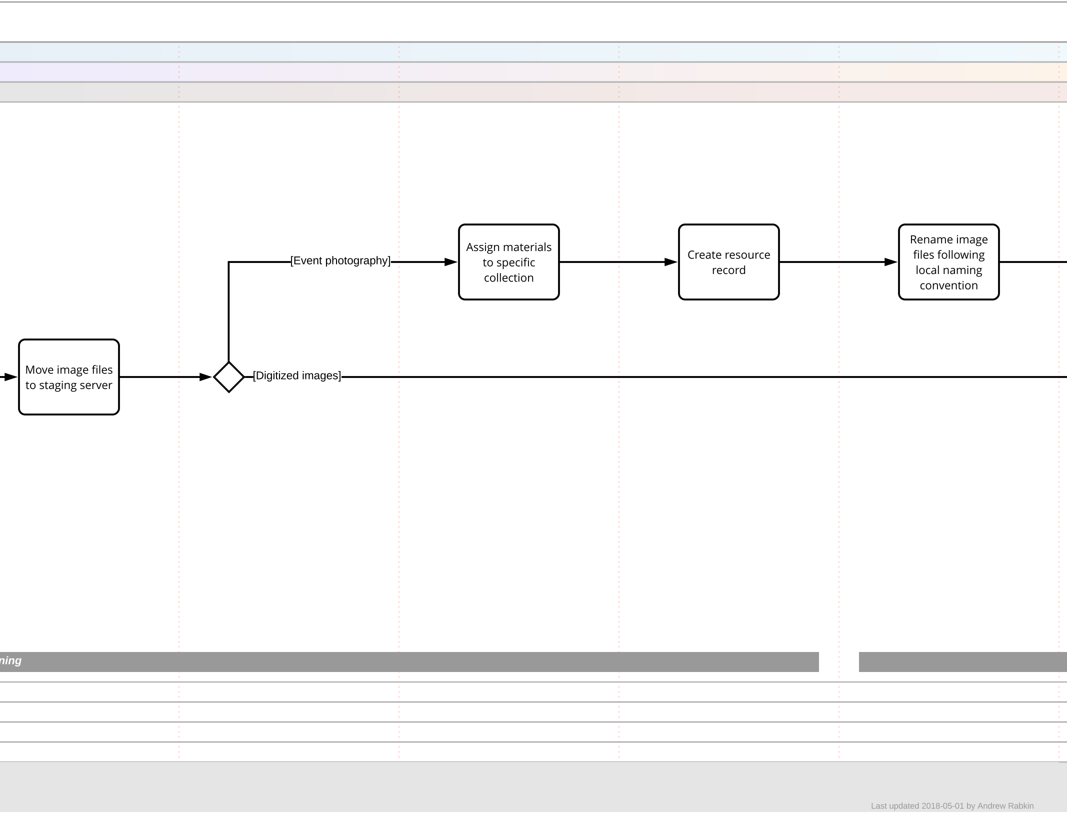
Action

Connection

Object / Data

Delay

Pain point


Move image files to staging server
[Event photography]
Assign materials to specific collection
Create resource record
Rename image files following local naming convention
[Digitized images]
ning

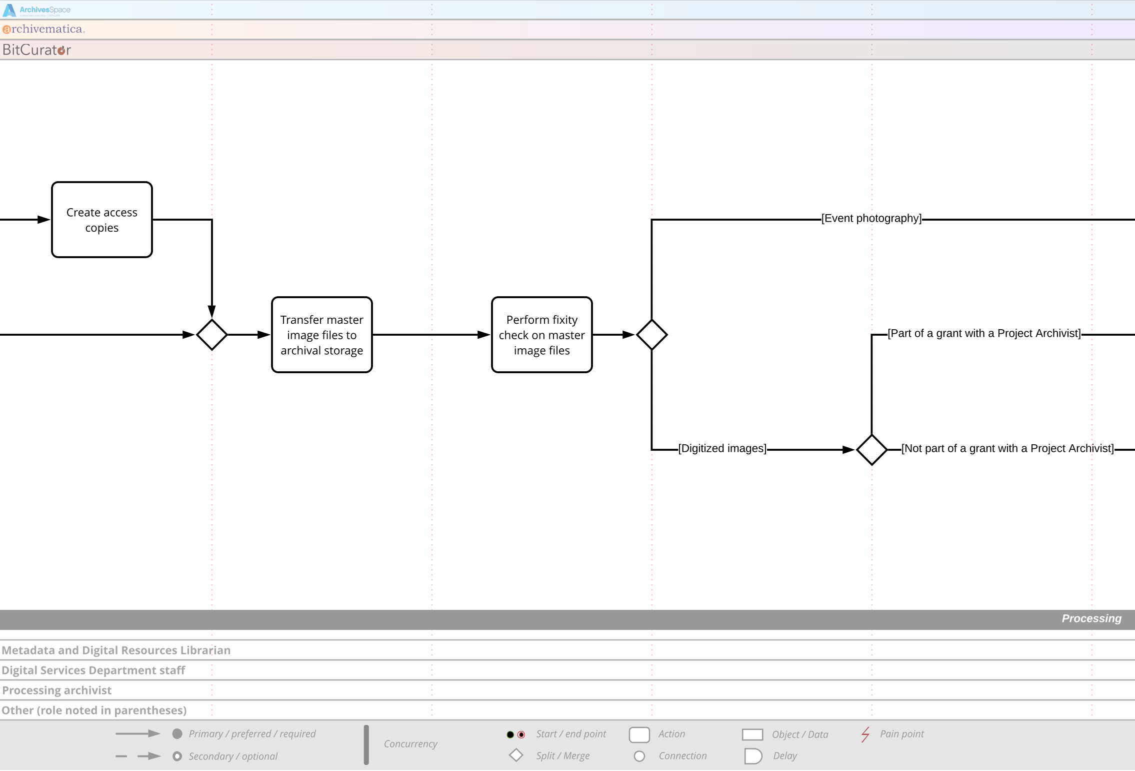
Last updated 2018-05-01 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Woodruff Library, Atlanta University Center

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Create access copies

Transfer master image files to archival storage

Perform fixity check on master image files

[Event photography]

[Part of a grant with a Project Archivist]

[Digitized images]

[Not part of a grant with a Project Archivist]

Processing

Metadata and Digital Resources Librarian

Digital Services Department staff

Processing archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

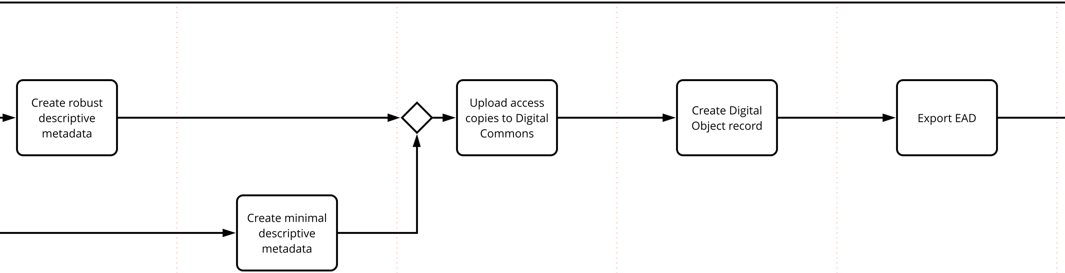
Action

Connection

Object / Data

Delay

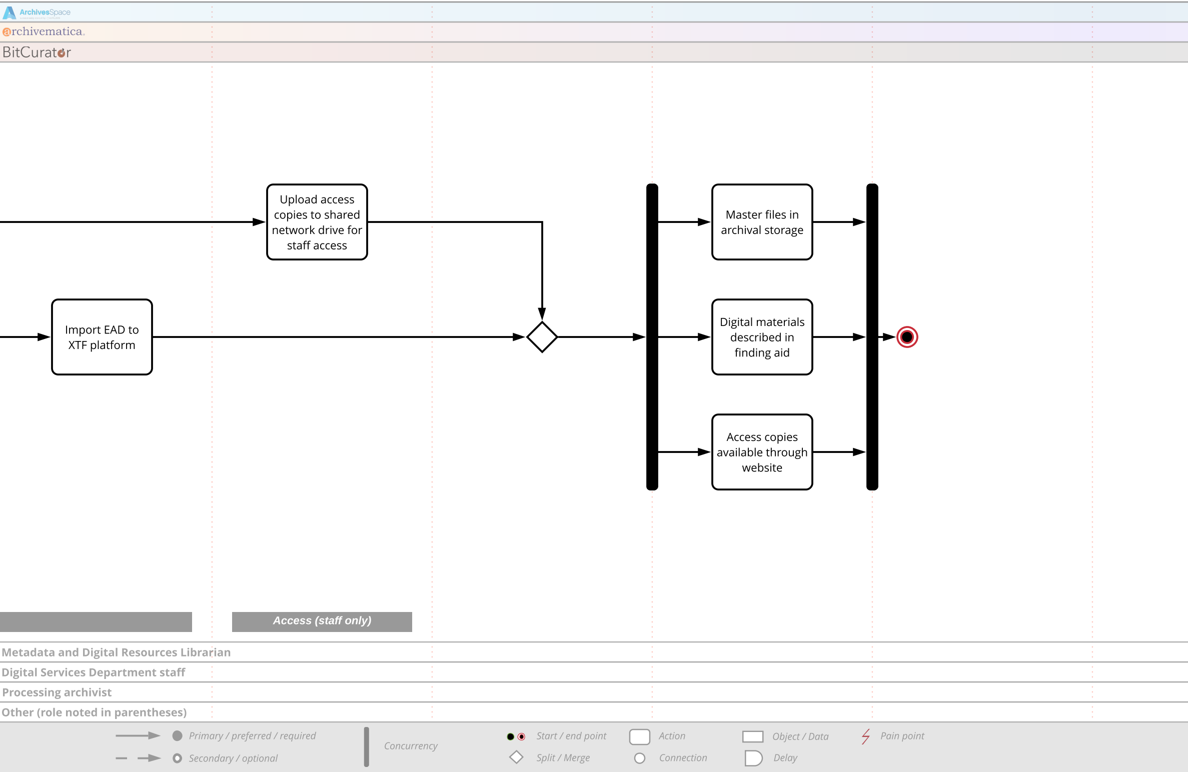
Pain point


Create robust descriptive metadata
Upload access copies to Digital Commons
Create Digital Object record
Export EAD
Create minimal descriptive metadata

Last updated 2018-05-01 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017 Woodruff Library, Atlanta University Center ***OssArcFlow Project***

[*https://educopia.org/research/ossarcflow*](https://educopia.org/research/ossarcflow)


ArchivesSpace
archivematica
BitCurator
Upload access copies to shared network drive for staff access
Import EAD to XTF platform
Master files in archival storage
Digital materials described in finding aid
Access copies available through website
Access (staff only)
Metadata and Digital Resources Librarian
Digital Services Department staff
Processing archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point

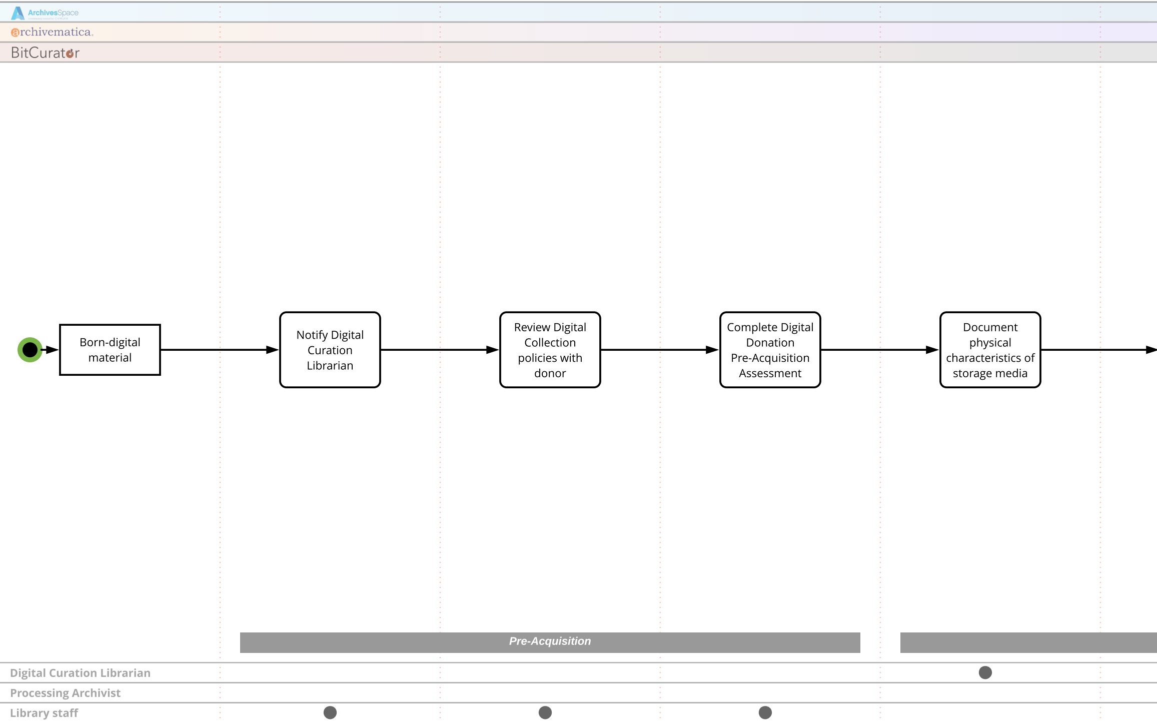
Last updated 2018-05-01 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

District of Columbia Public Library

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Born-digital material

Notify Digital Curation Librarian

Review Digital Collection policies with donor

Complete Digital Donation Pre-Acquisition Assessment

Document physical characteristics of storage media

Pre-Acquisition

Digital Curation Librarian

Processing Archivist

Library staff

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

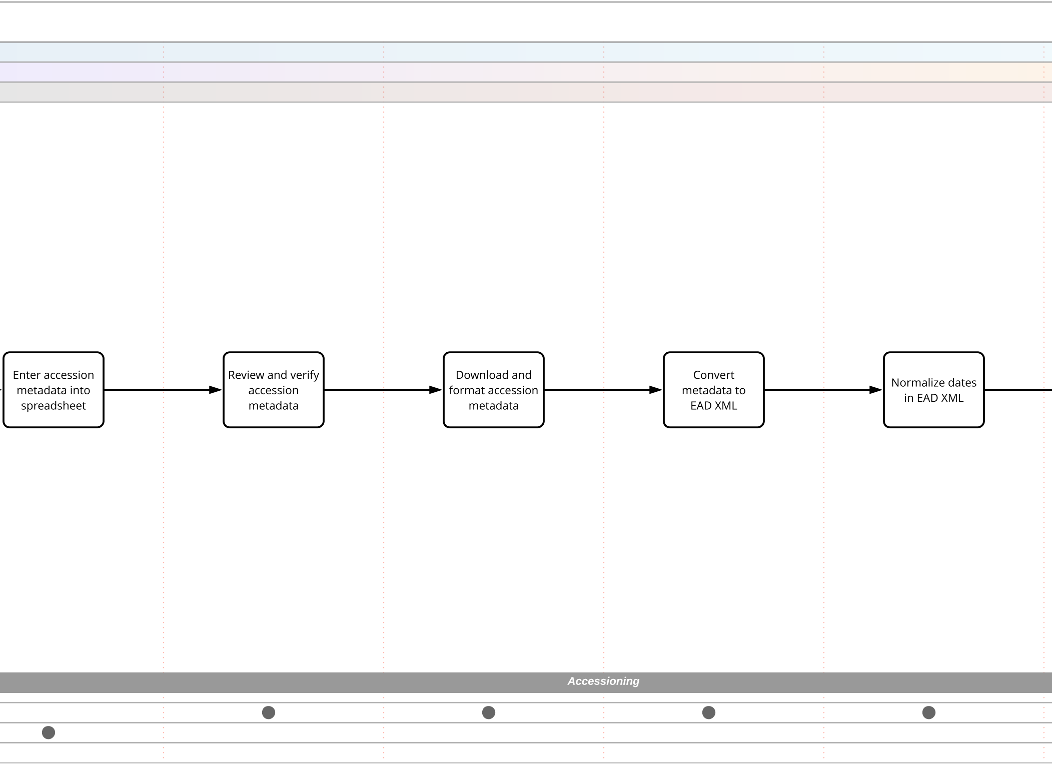
Action

Connection

Object / Data

Delay

Pain point

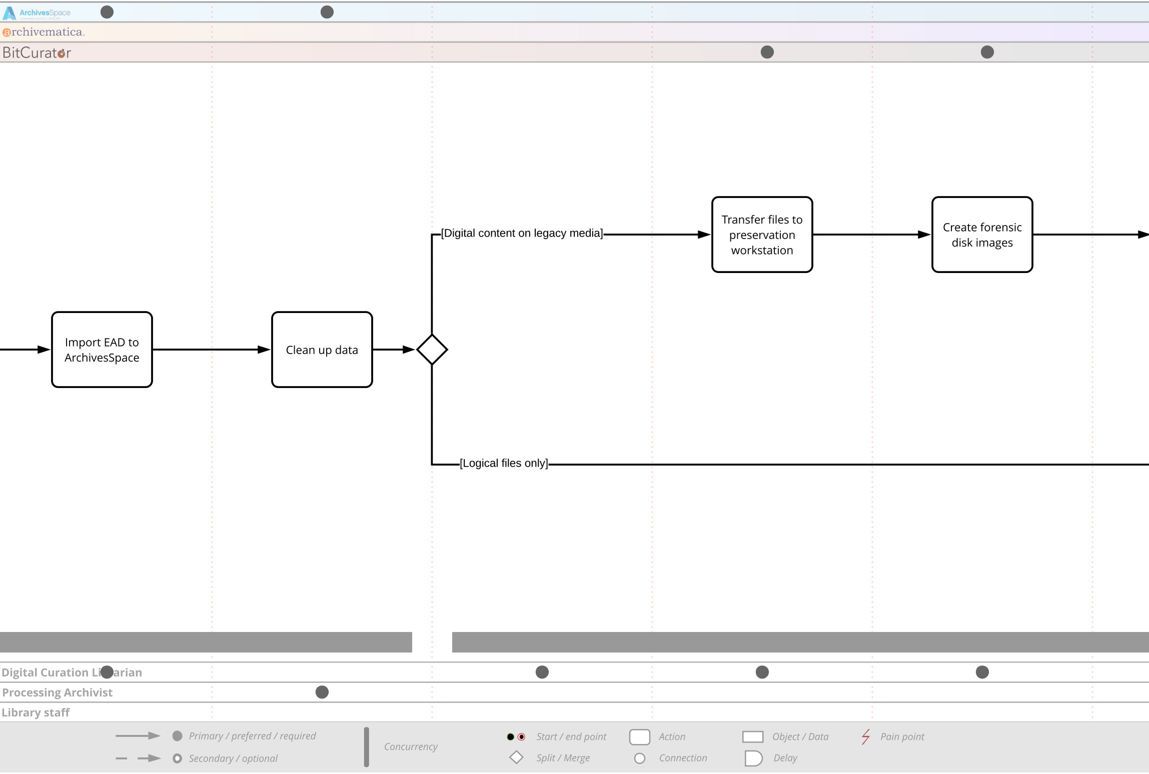

Enter accession metadata into spreadsheet
Review and verify accession metadata
Download and format accession metadata
Convert metadata to EAD XML
Normalize dates in EAD XML
Accessioning

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

District of Columbia Public Library

OssArcFlow Project
https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica.

BitCurator

Import EAD to ArchivesSpace

Clean up data

[Digital content on legacy media]

Transfer files to preservation workstation

Create forensic disk images

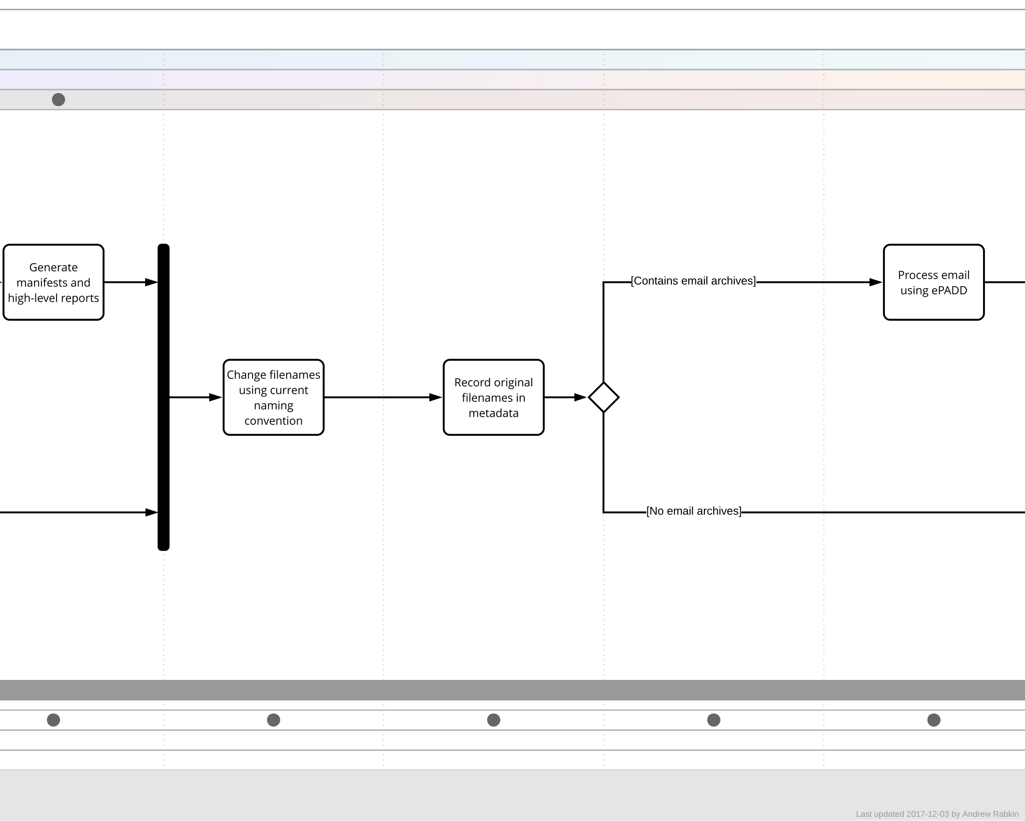
[Logical files only]

Digital Curation Librarian

Processing Archivist

Library staff

Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


Generate manifests and high-level reports
Change filenames using current naming convention
Record original filenames in metadata
[Contains email archives]
Process email using ePADD
[No email archives]

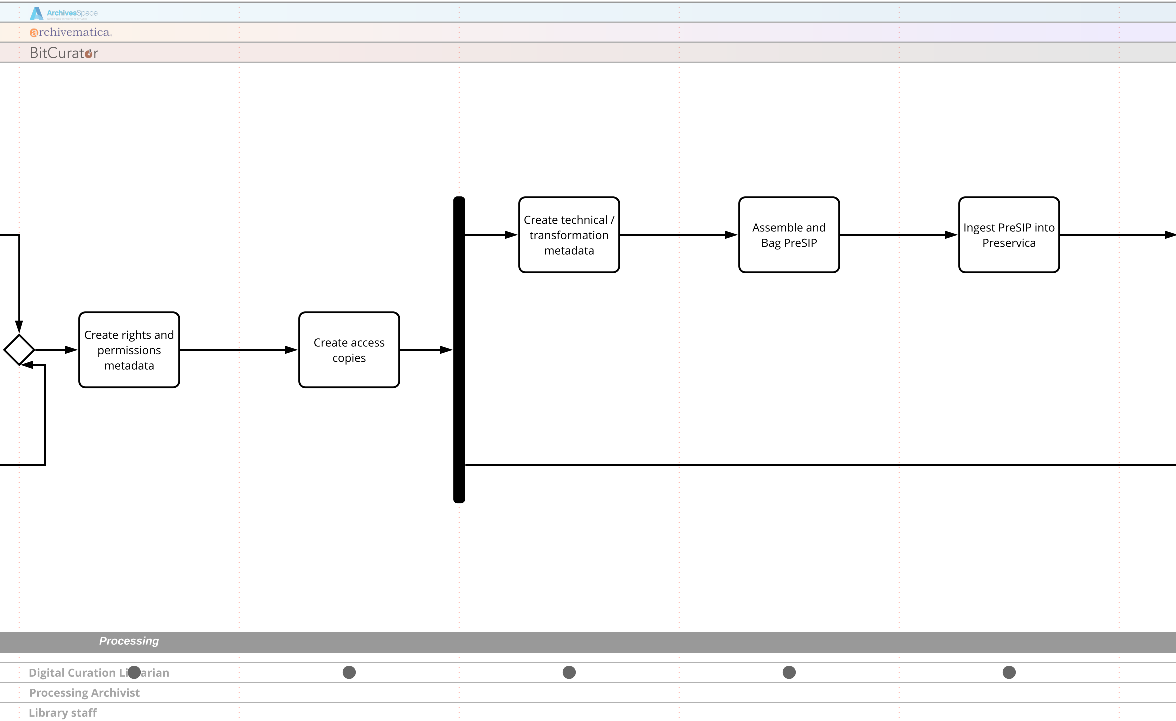
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

District of Columbia Public Library

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
Create technical / transformation metadata
Assemble and Bag PreSIP
Ingest PreSIP into Preservica
Create rights and permissions metadata
Create access copies
Processing
Digital Curation Librarian
Processing Archivist
Library staff

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

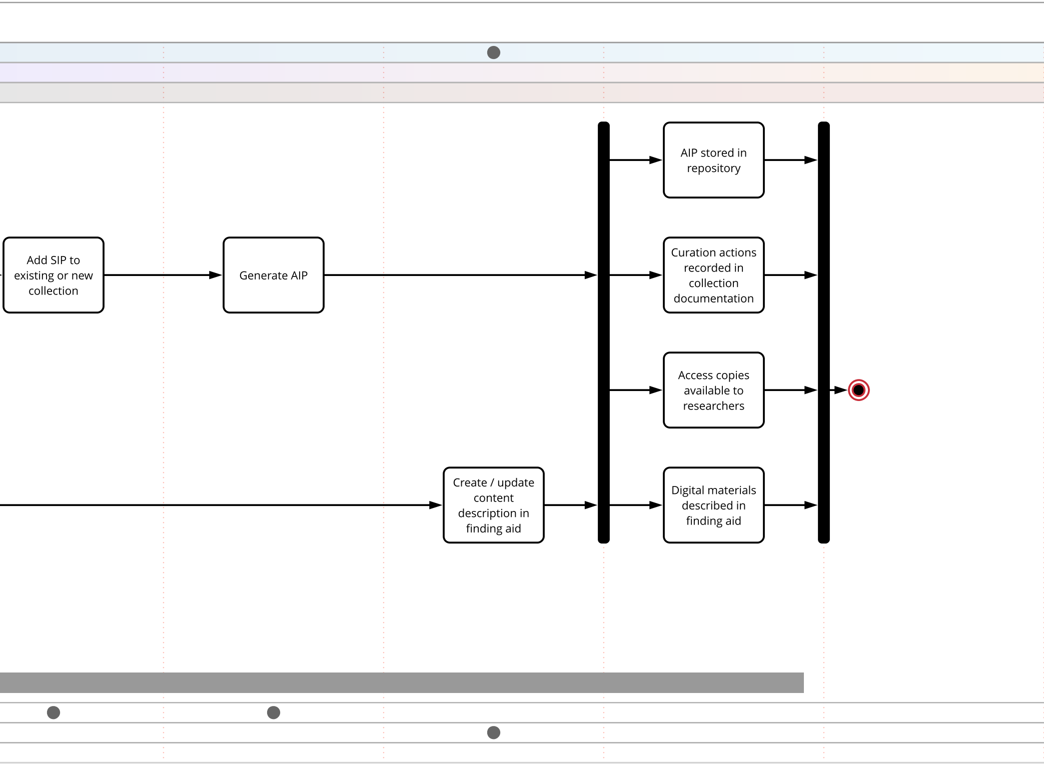
Action

Connection

Object / Data

Delay

Pain point


Add SIP to existing or new collection
Generate AIP
Create / update content description in finding aid
AIP stored in repository
Curation actions recorded in collection documentation
Access copies available to researchers
Digital materials described in finding aid

Last updated 2017-12-03 by Andrew Rabkin

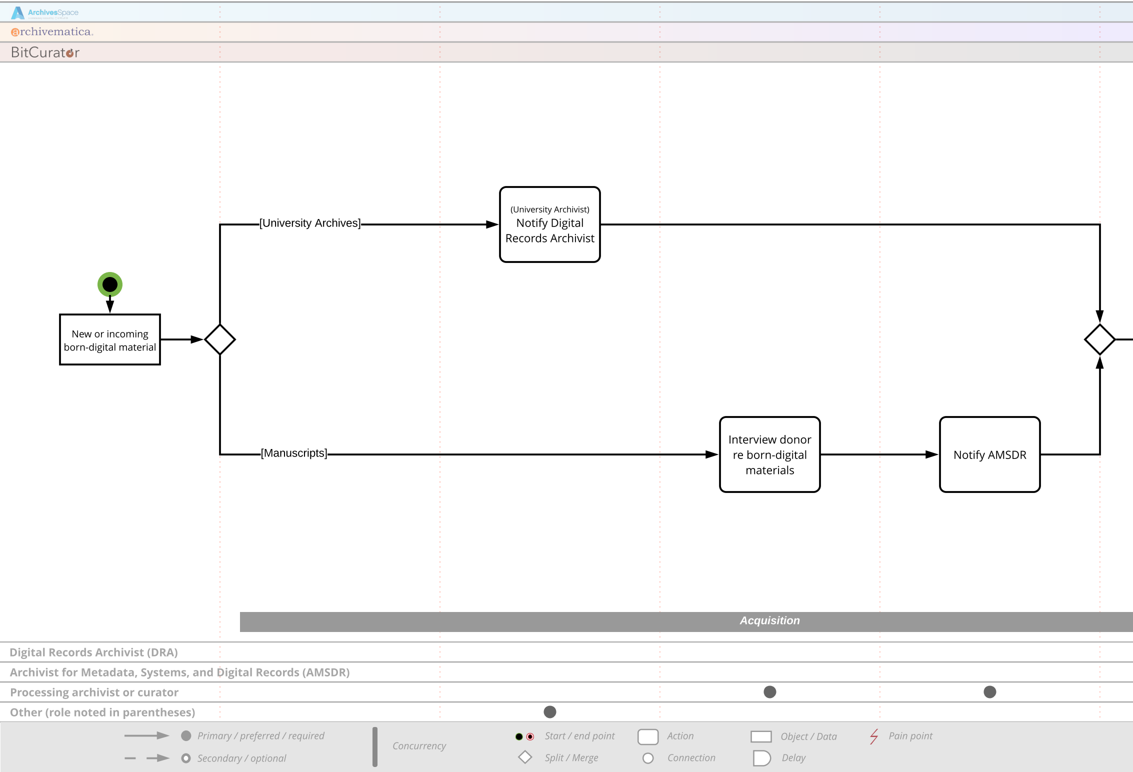
As-Is Digital Curation Workflow, August 2017

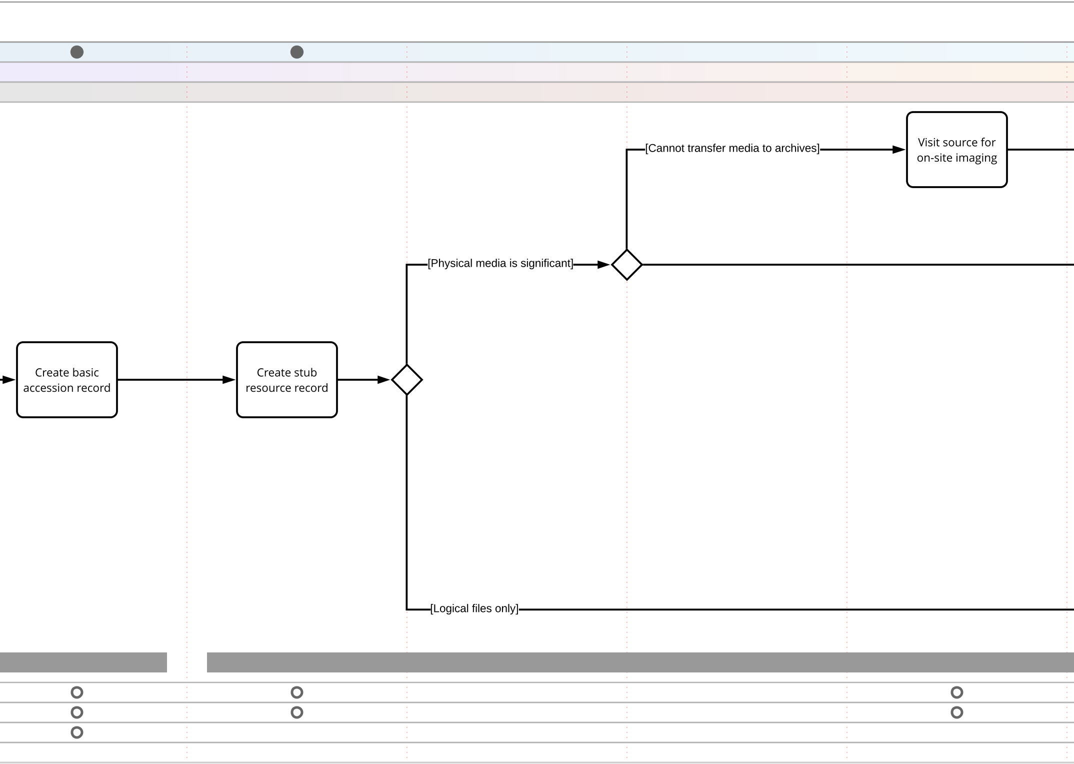
Born-digital content

Duke University

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
New or incoming born-digital material
[University Archives]
(University Archivist) Notify Digital Records Archivist
[Manuscripts]
Interview donor re born-digital materials
Notify AMSDR
Acquisition
Digital Records Archivist (DRA)
Archivist for Metadata, Systems, and Digital Records (AMSDR)
Processing archivist or curator
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


[Cannot transfer media to archives]
Visit source for on-site imaging
[Physical media is significant]
Create basic accession record
Create stub resource record
[Logical files only]

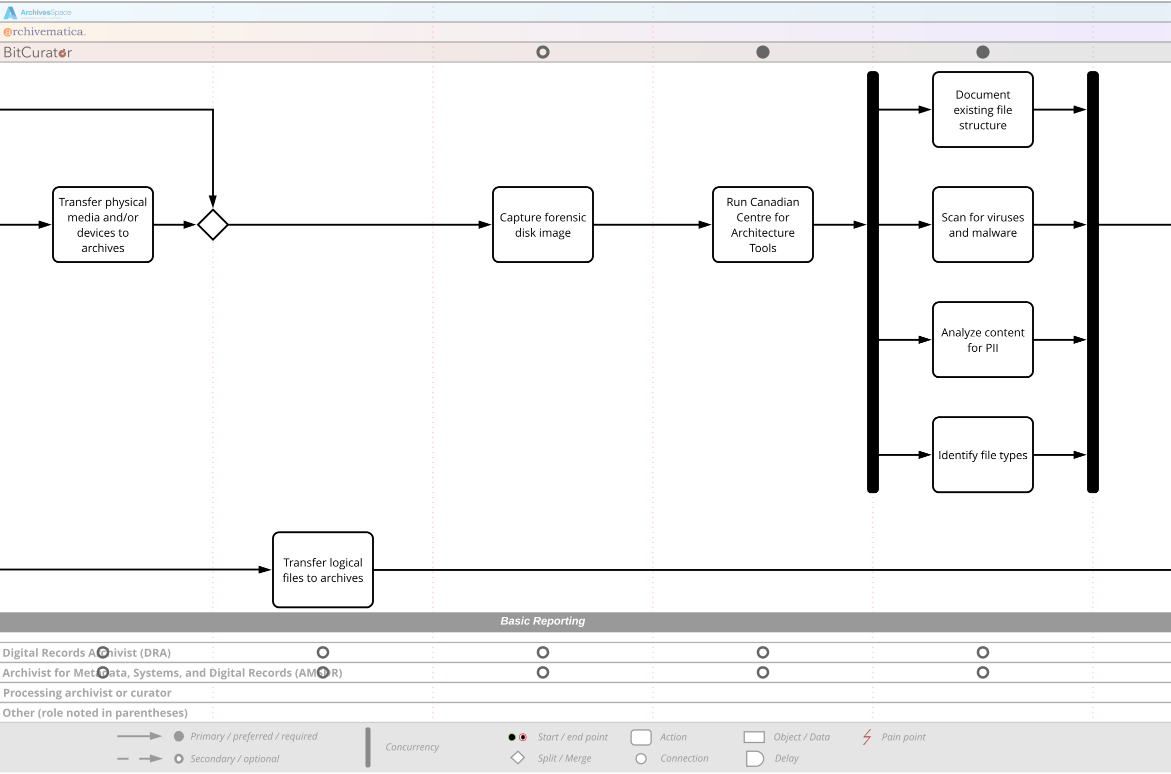
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Duke University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Transfer physical media and/or devices to archives

Capture forensic disk image

Run Canadian Centre for Architecture Tools

Document existing file structure

Scan for viruses and malware

Analyze content for PII

Identify file types

Transfer logical files to archives

Basic Reporting

Digital Records Archivist (DRA)

Archivist for Metadata, Systems, and Digital Records (AMSDR)

Processing archivist or curator

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

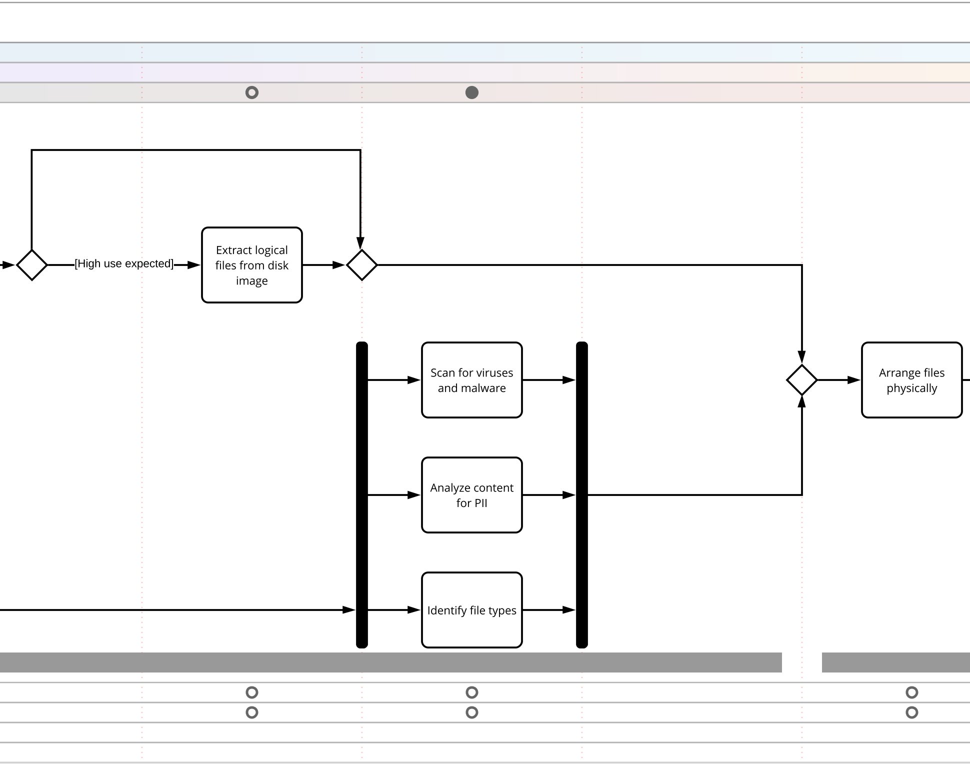
Action

Connection

Object / Data

Delay

Pain point


[High use expected]
Extract logical files from disk image
Scan for viruses and malware
Analyze content for PII
Identify file types
Arrange files physically

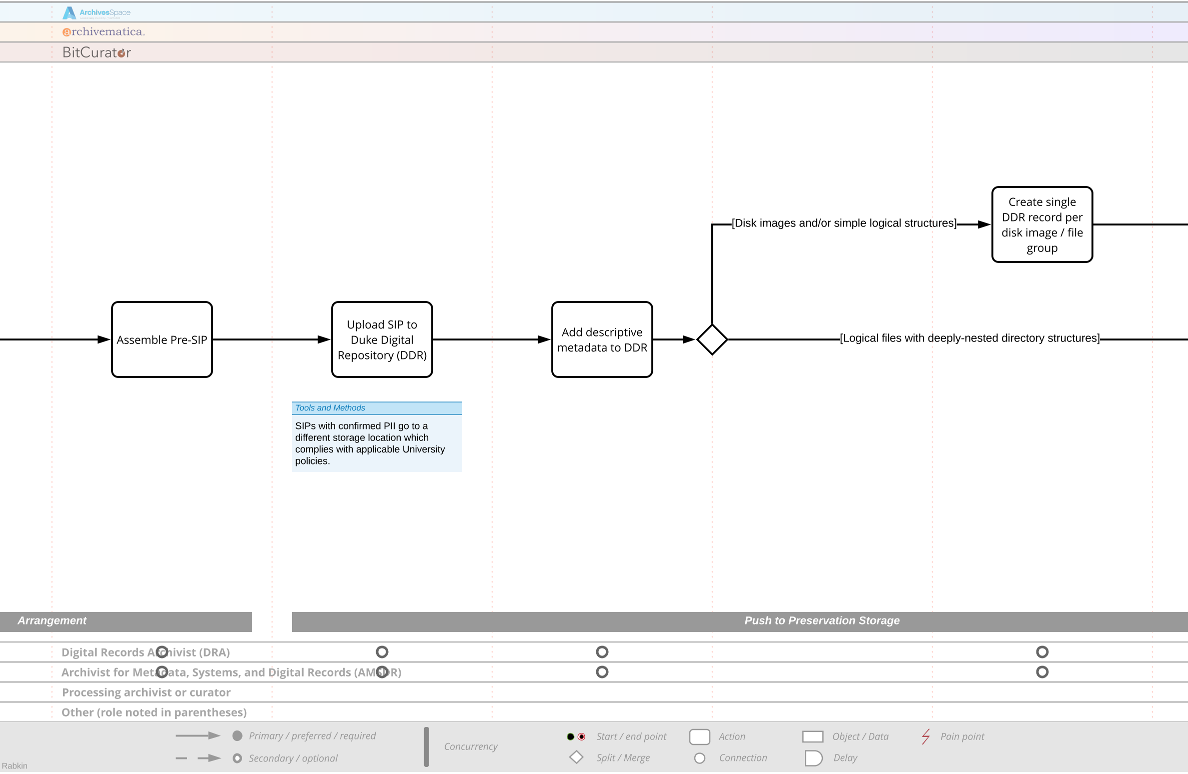
Last updated 2017-12-03 by Andrew

As-Is Digital Curation Workflow, August 2017

Duke University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Assemble Pre-SIP

Upload SIP to Duke Digital Repository (DDR)

Add descriptive metadata to DDR

[Disk images and/or simple logical structures]

Create single DDR record per disk image / file group

[Logical files with deeply-nested directory structures]

Tools and Methods

SIPs with confirmed PII go to a different storage location which complies with applicable University policies.

Arrangement

Push to Preservation Storage

Digital Records Archivist (DRA)

Archivist for Metadata, Systems, and Digital Records (AMSDR)

Processing archivist or curator

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

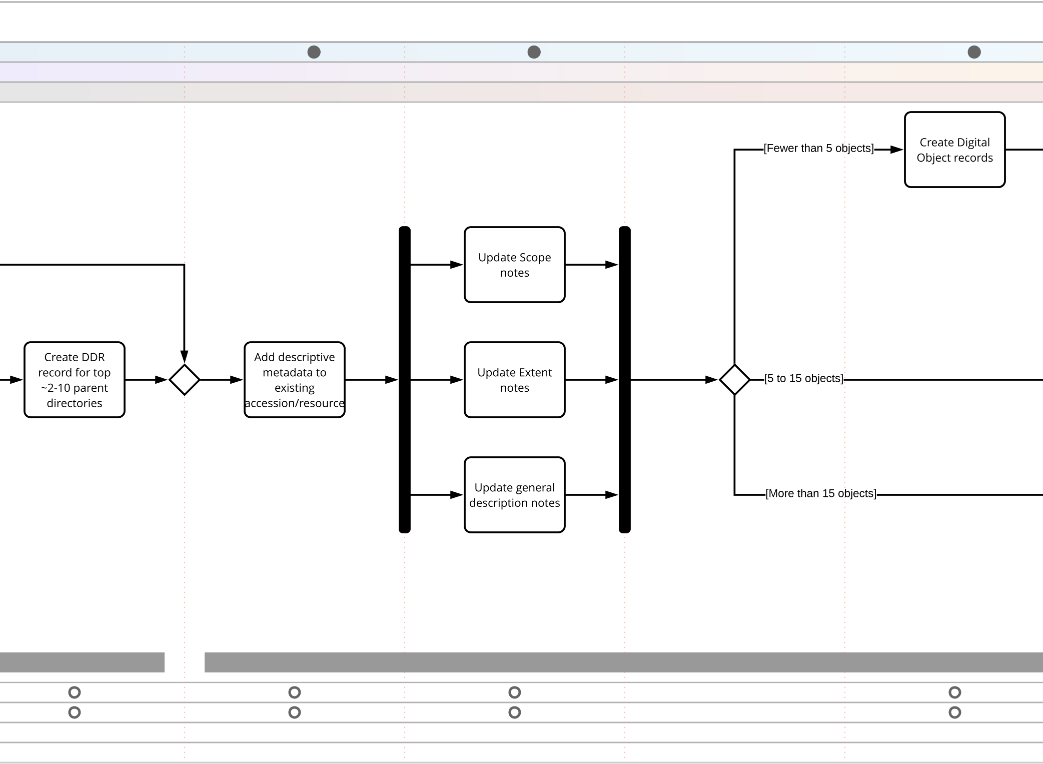
Connection

Object / Data

Delay

Pain point

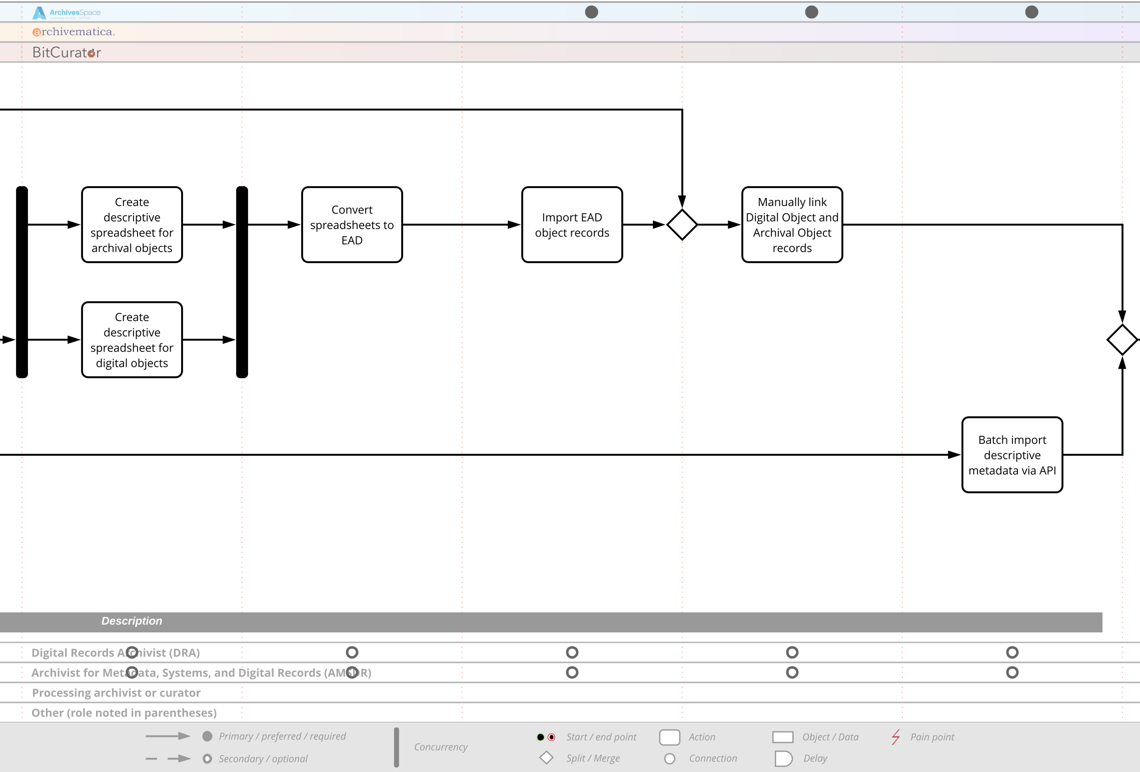
Rabkin

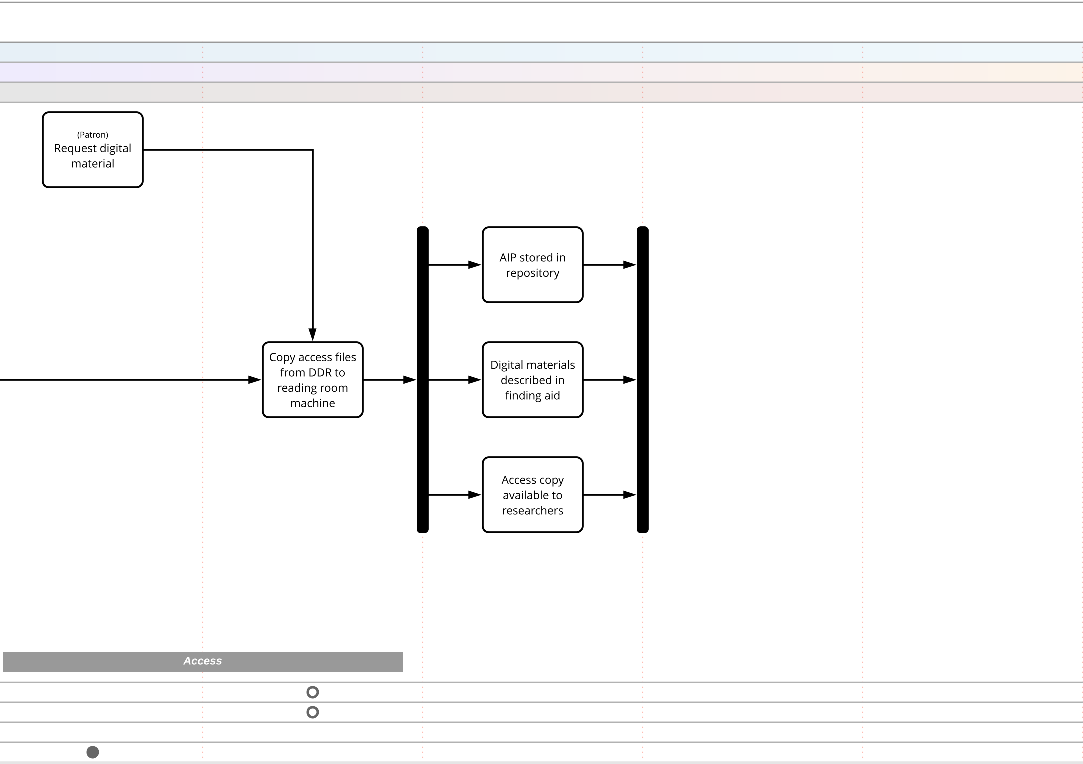

Create DDR record for top ~2-10 parent directories
Add descriptive metadata to existing accession/resource
Update Scope notes
Update Extent notes
Update general description notes
[Fewer than 5 objects]
Create Digital Object records
[5 to 15 objects]
[More than 15 objects]

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017 Duke University ***OssArcFlow Project***

[*https://educopia.org/research/ossarcflow*](https://educopia.org/research/ossarcflow)


ArchivesSpace
archivematica.
BitCurator
Create descriptive spreadsheet for archival objects
Create descriptive spreadsheet for digital objects
Convert spreadsheets to EAD
Import EAD object records
Manually link Digital Object and Archival Object records
Batch import descriptive metadata via API
Description
Digital Records Archivist (DRA)
Archivist for Metadata, Systems, and Digital Records (AMSDR)
Processing archivist or curator
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


(Patron)
Request digital material
Copy access files from DDR to reading room machine
AIP stored in repository
Digital materials described in finding aid
Access copy available to researchers
Access

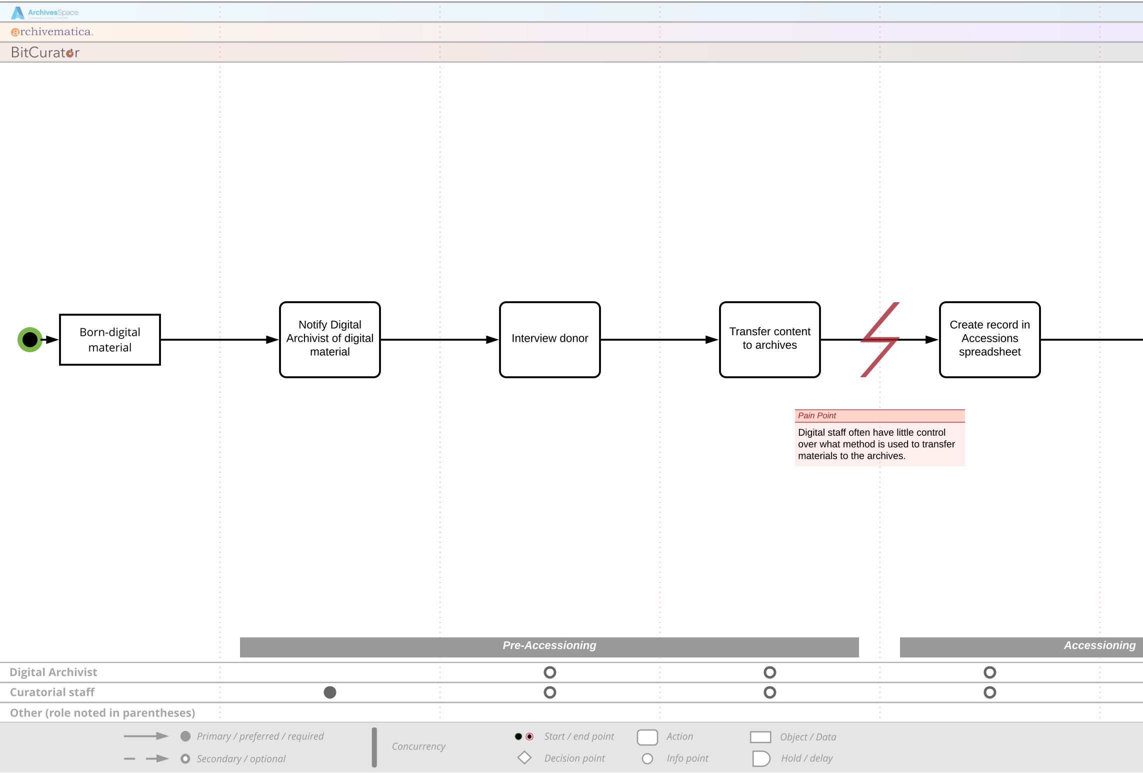
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace
archivematica.
BitCurator

Born-digital material → Notify Digital Archivist of digital material → Interview donor → Transfer content to archives → Create record in Accessions spreadsheet

Pain Point

Digital staff often have little control over what method is used to transfer materials to the archives.

Pre-Accessioning

Accessioning

	Born-digital material	Notify Digital Archivist of digital material	Interview donor	Transfer content to archives	Create record in Accessions spreadsheet
Digital Archivist			○	○	○
Curatorial staff		●	○	○	○
Other (role noted in parentheses)					

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

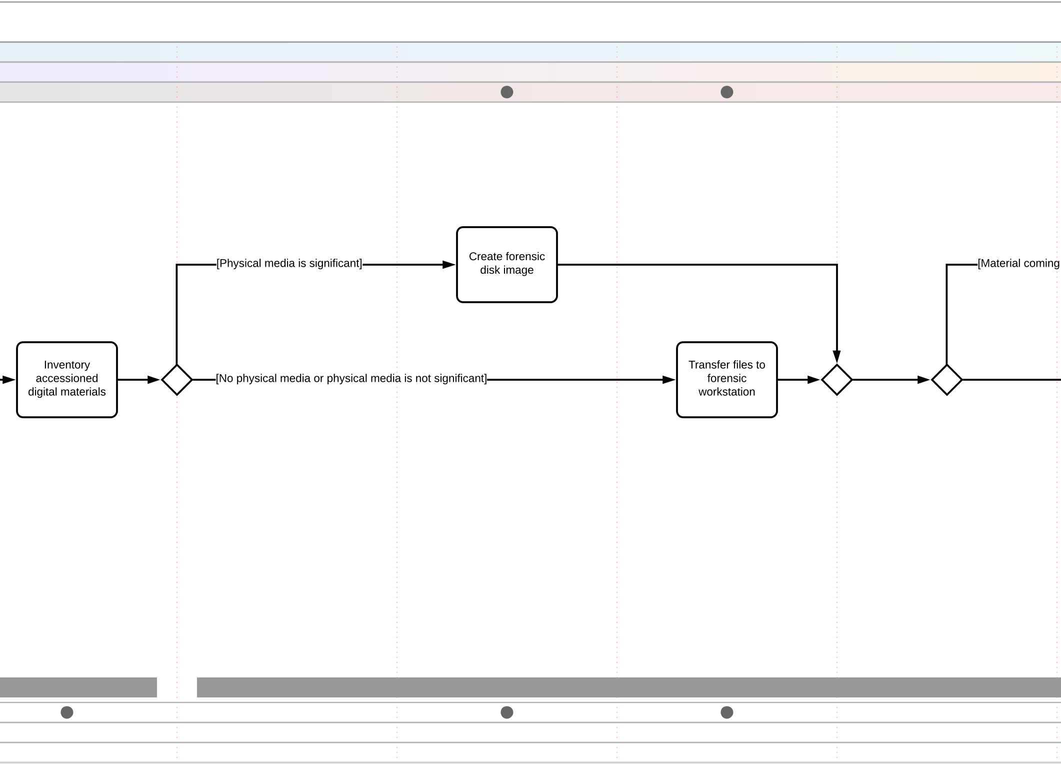
Decision point

Action

Info point

Object / Data

Hold / delay


Inventory accessioned digital materials
[Physical media is significant]
Create forensic disk image
[No physical media or physical media is not significant]
Transfer files to forensic workstation
[Material coming

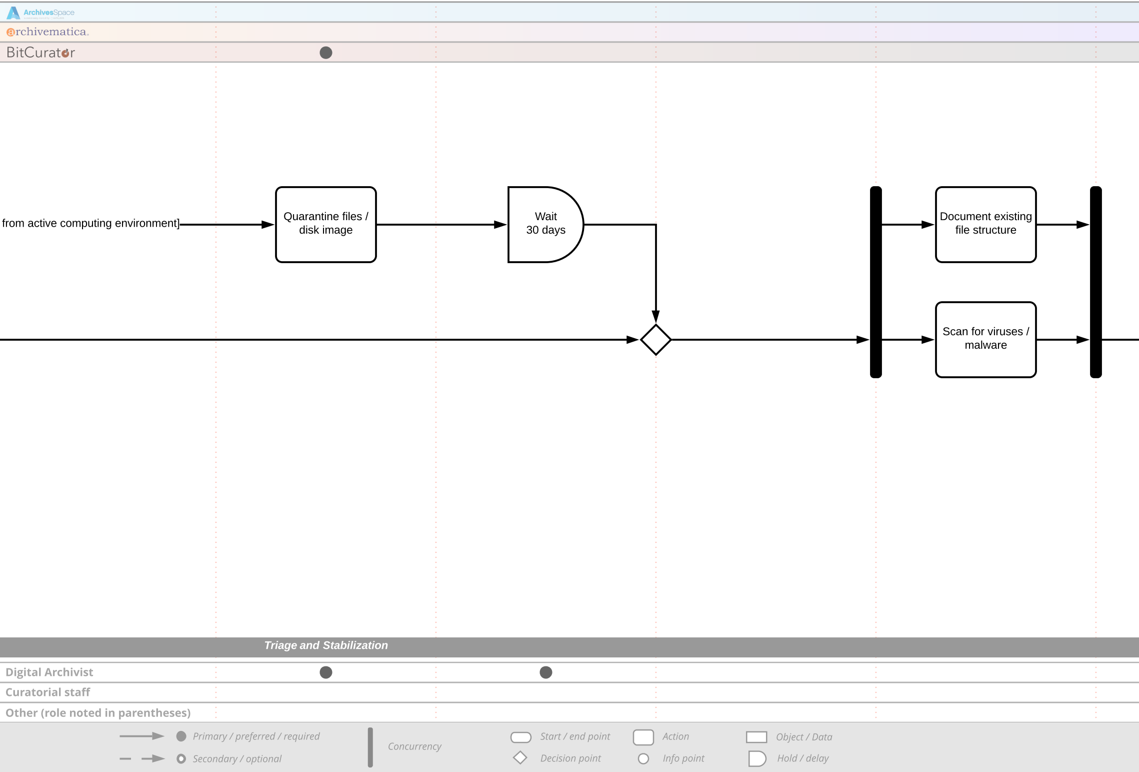
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

from active computing environment]

Quarantine files / disk image

Wait 30 days

Document existing file structure

Scan for viruses / malware

Triage and Stabilization

Digital Archivist

Curatorial staff

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

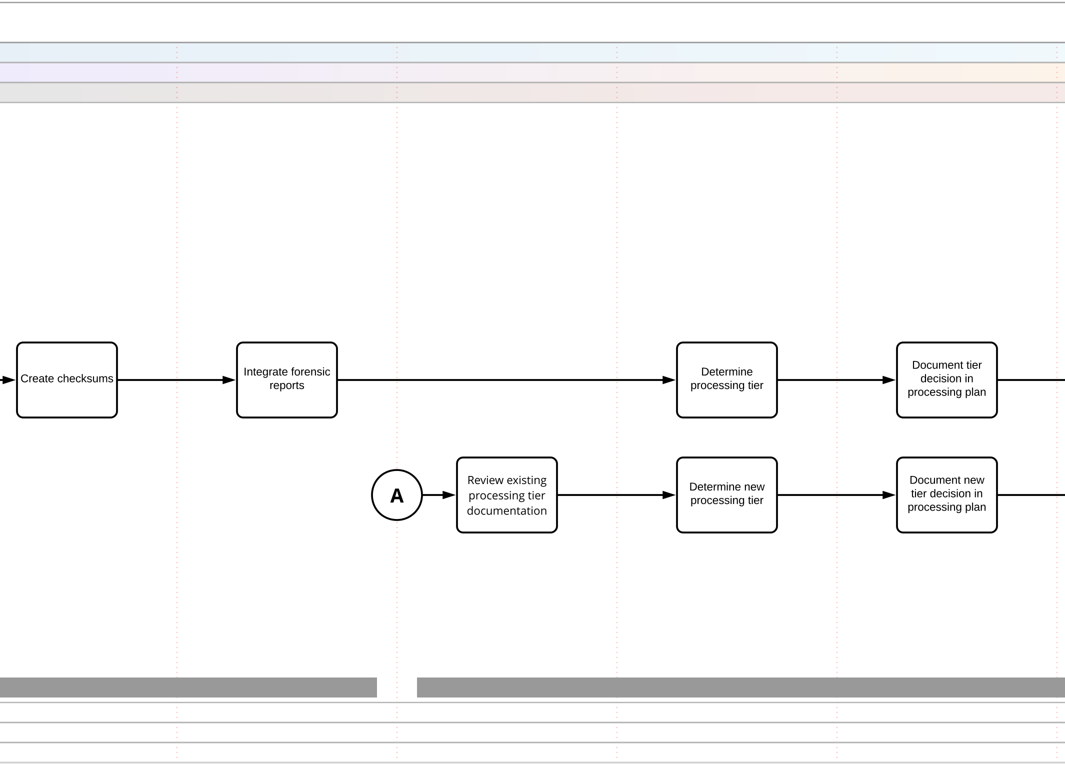
Decision point

Action

Info point

Object / Data

Hold / delay


Create checksums
Integrate forensic reports
Determine processing tier
Document tier decision in processing plan
A
Review existing processing tier documentation
Determine new processing tier
Document new tier decision in processing plan

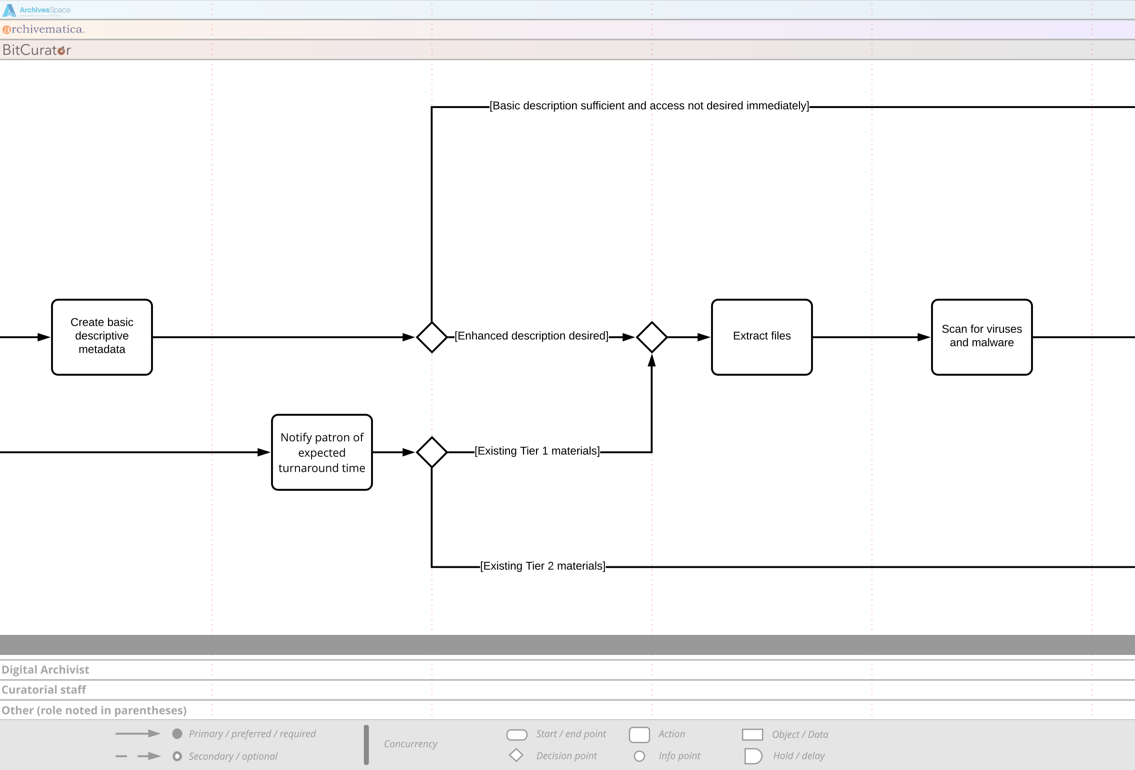
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica.

BitCurator

Create basic descriptive metadata

[Basic description sufficient and access not desired immediately]

[Enhanced description desired]

Extract files

Scan for viruses and malware

Notify patron of expected turnaround time

[Existing Tier 1 materials]

[Existing Tier 2 materials]

Digital Archivist

Curatorial staff

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

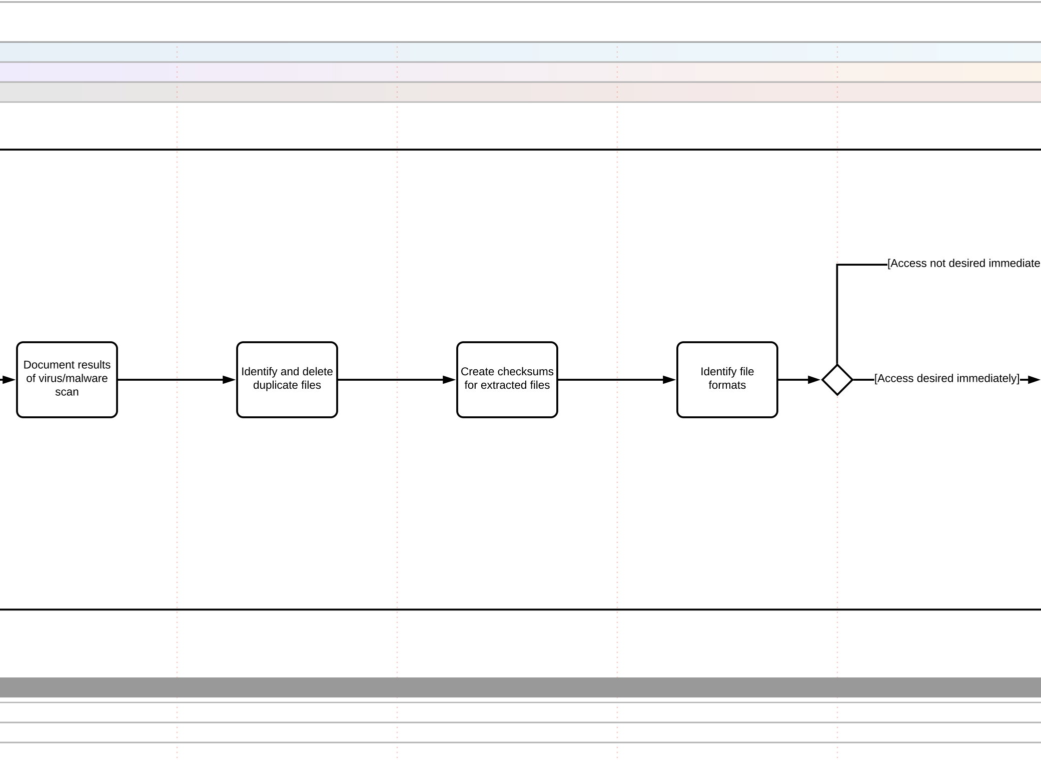
Decision point

Action

Info point

Object / Data

Hold / delay


Document results of virus/malware scan
Identify and delete duplicate files
Create checksums for extracted files
Identify file formats
[Access not desired immediate
[Access desired immediately]

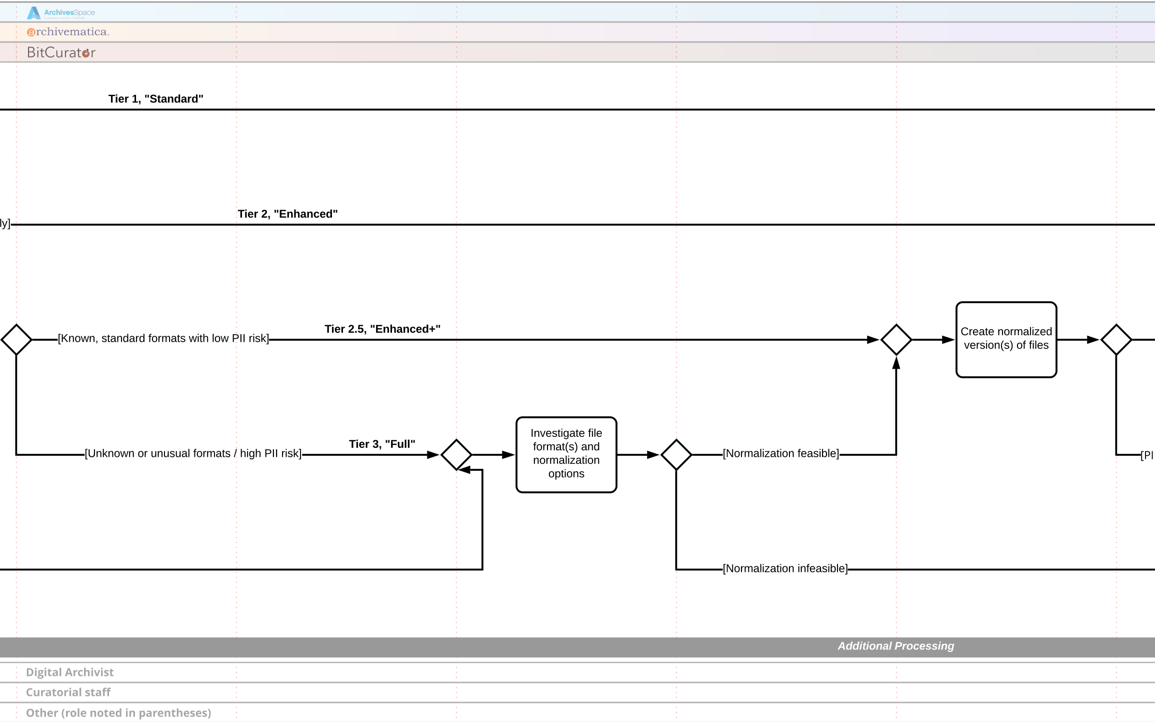
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Tier 1, "Standard"

Tier 2, "Enhanced"

ly]

[Known, standard formats with low PII risk]

Tier 2.5, "Enhanced+"

Create normalized version(s) of files

[Unknown or unusual formats / high PII risk]

Tier 3, "Full"

Investigate file format(s) and normalization options

[Normalization feasible]

[PI

[Normalization infeasible]

Additional Processing

Digital Archivist

Curatorial staff

Other (role noted in parentheses)

→ ● *Primary / preferred / required*

⇢ ○ *Secondary / optional*

| *Concurrency*

▭ *Start / end point*

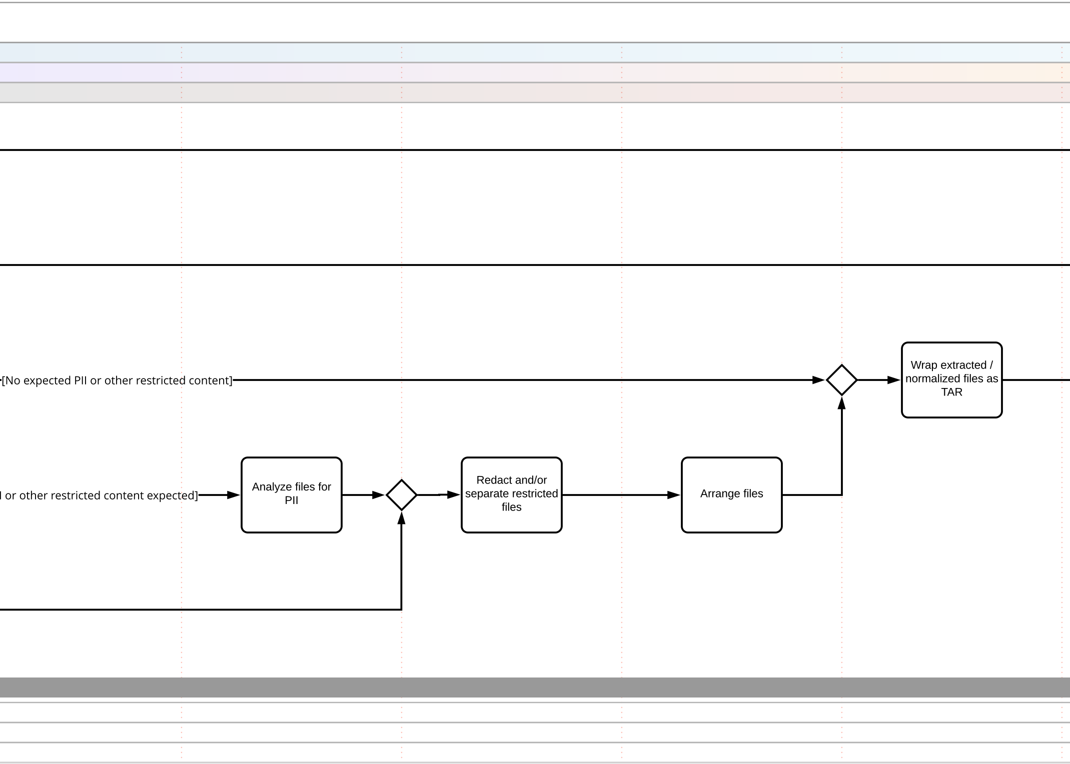
◇ *Decision point*

▢ *Action*

○ *Info point*

▭ *Object / Data*

D *Hold / delay*


[No expected PII or other restricted content]
I or other restricted content expected]
Analyze files for PII
Redact and/or separate restricted files
Arrange files
Wrap extracted / normalized files as TAR

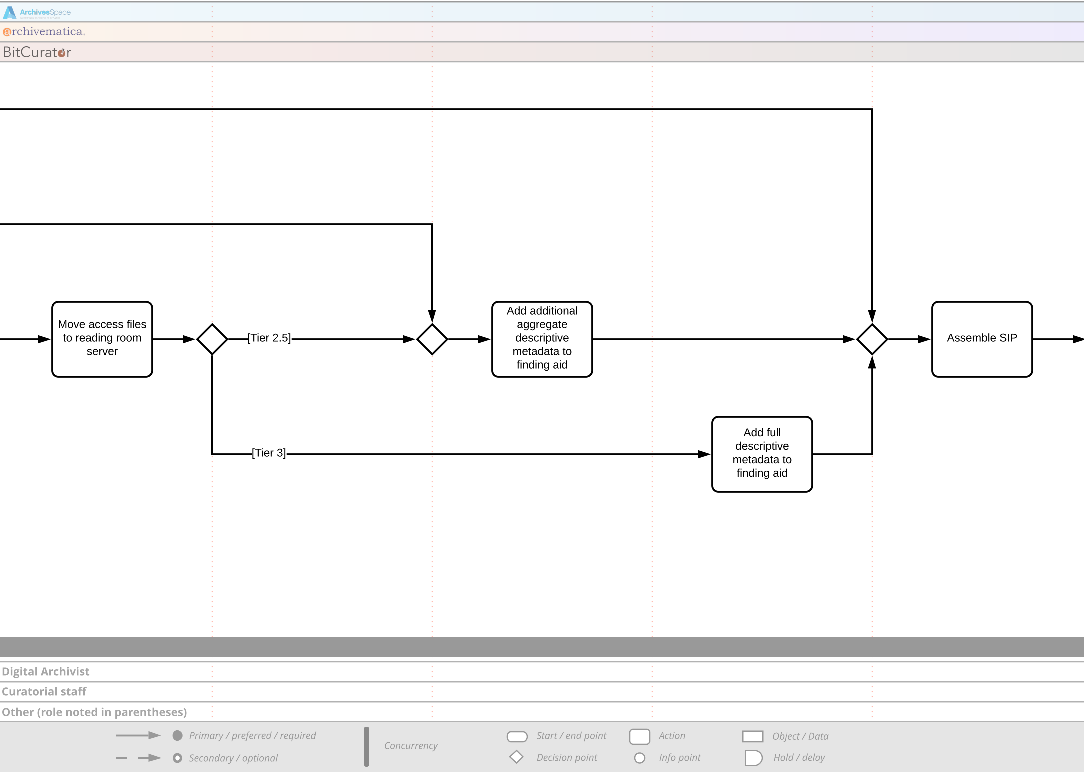
Last updated 2018-05-02 by Andrew Rabkin

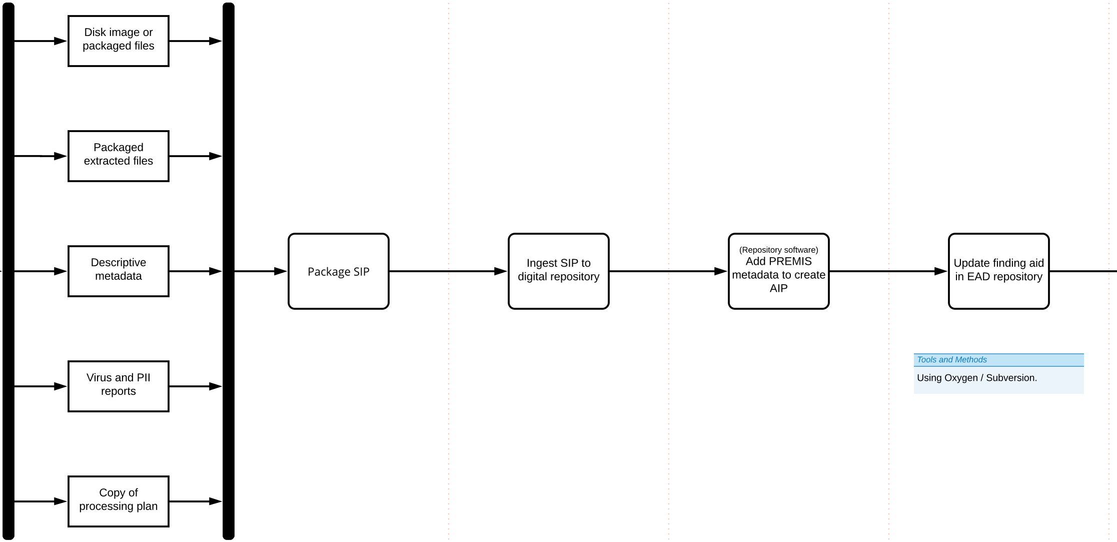
As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
Move access files to reading room server
[Tier 2.5]
Add additional aggregate descriptive metadata to finding aid
[Tier 3]
Add full descriptive metadata to finding aid
Assemble SIP
Digital Archivist
Curatorial staff
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Decision point
Action
Info point
Object / Data
Hold / delay


Disk image or packaged files
Packaged extracted files
Descriptive metadata
Virus and PII reports
Copy of processing plan
Package SIP
Ingest SIP to digital repository
(Repository software) Add PREMIS metadata to create AIP
Update finding aid in EAD repository

Tools and Methods

Using Oxygen / Subversion.

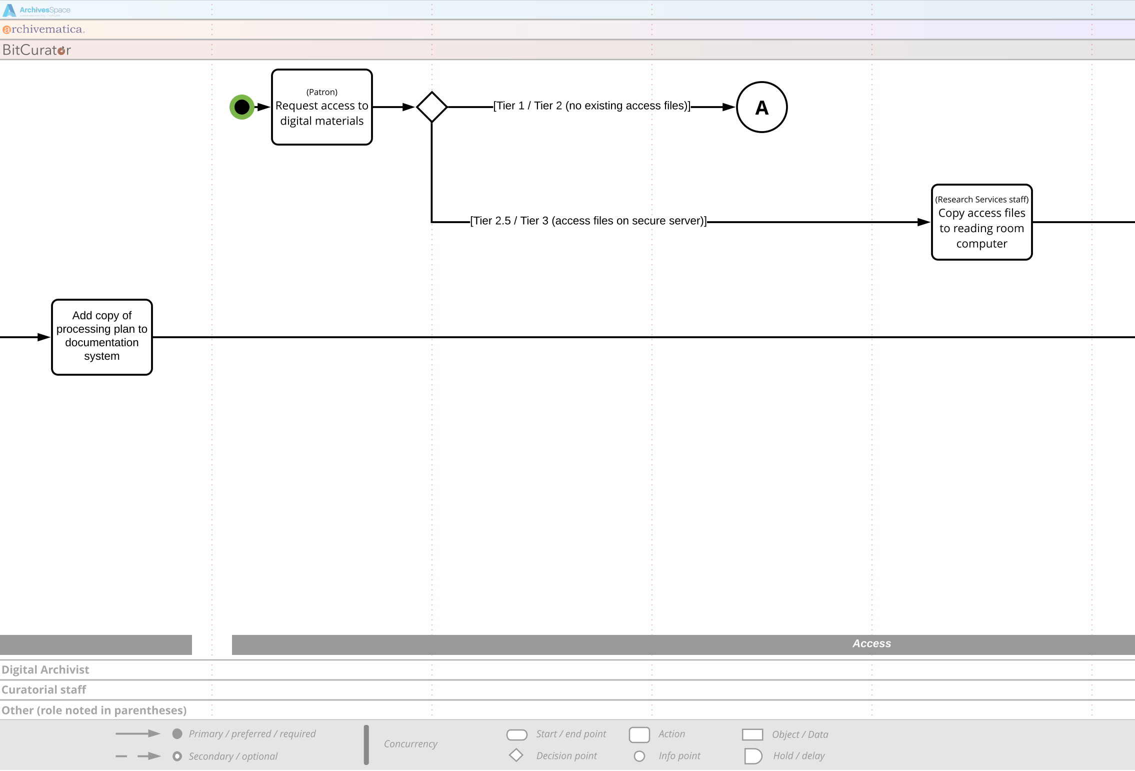
Last updated 2018-05-02 by Andrew Rabkin

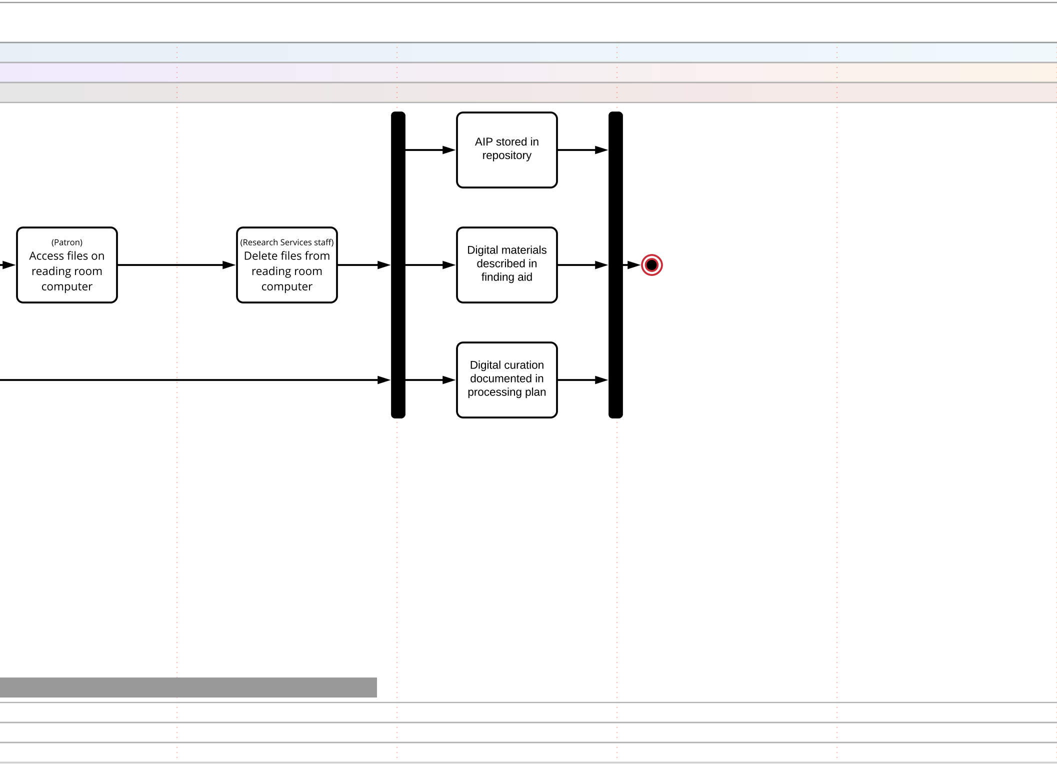
As-Is Digital Curation Workflow, August 2017

Emory University

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica.
BitCurator
(Patron) Request access to digital materials
[Tier 1 / Tier 2 (no existing access files)]
A
[Tier 2.5 / Tier 3 (access files on secure server)]
(Research Services staff) Copy access files to reading room computer
Add copy of processing plan to documentation system
Access
Digital Archivist
Curatorial staff
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Decision point
Action
Info point
Object / Data
Hold / delay


(Patron)
Access files on reading room computer
(Research Services staff)
Delete files from reading room computer
AIP stored in repository
Digital materials described in finding aid
Digital curation documented in processing plan

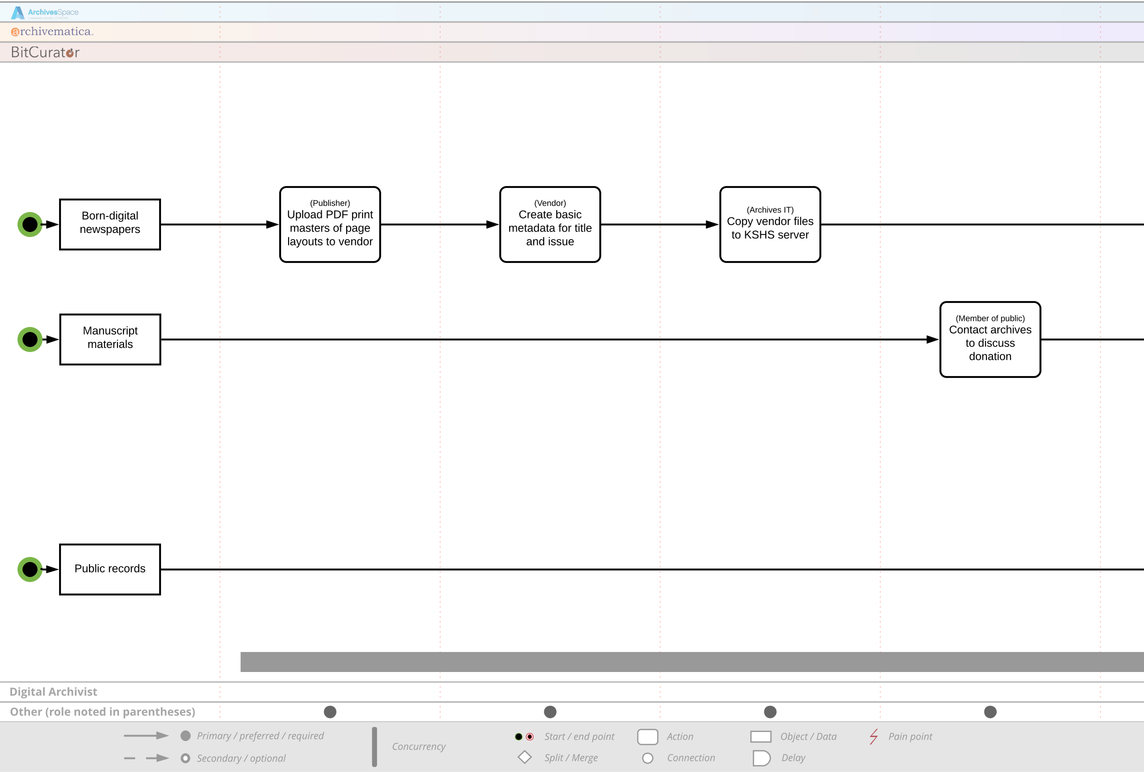
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Kansas Historical Society

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Born-digital newspapers

(Publisher)
Upload PDF print masters of page layouts to vendor

(Vendor)
Create basic metadata for title and issue

(Archives IT)
Copy vendor files to KSHS server

Manuscript materials

(Member of public)
Contact archives to discuss donation

Public records

Digital Archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

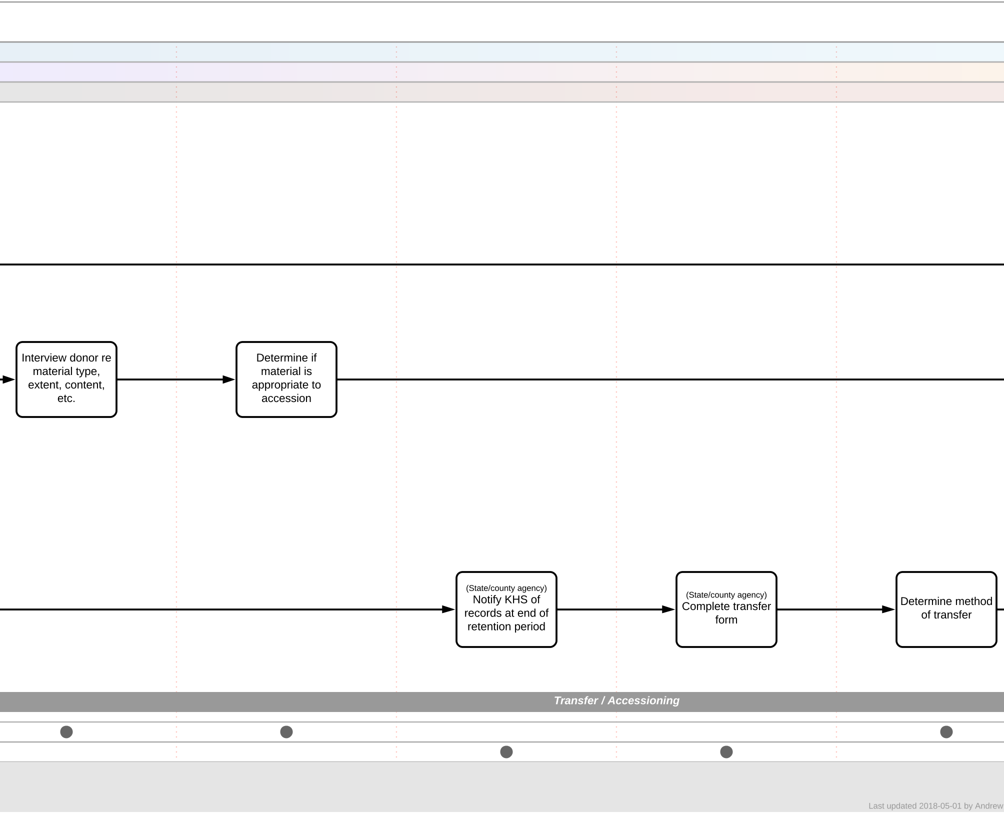
Action

Connection

Object / Data

Delay

Pain point


Interview donor re material type, extent, content, etc.
Determine if material is appropriate to accession
(State/county agency) Notify KHS of records at end of retention period
(State/county agency) Complete transfer form
Determine method of transfer
Transfer / Accessioning

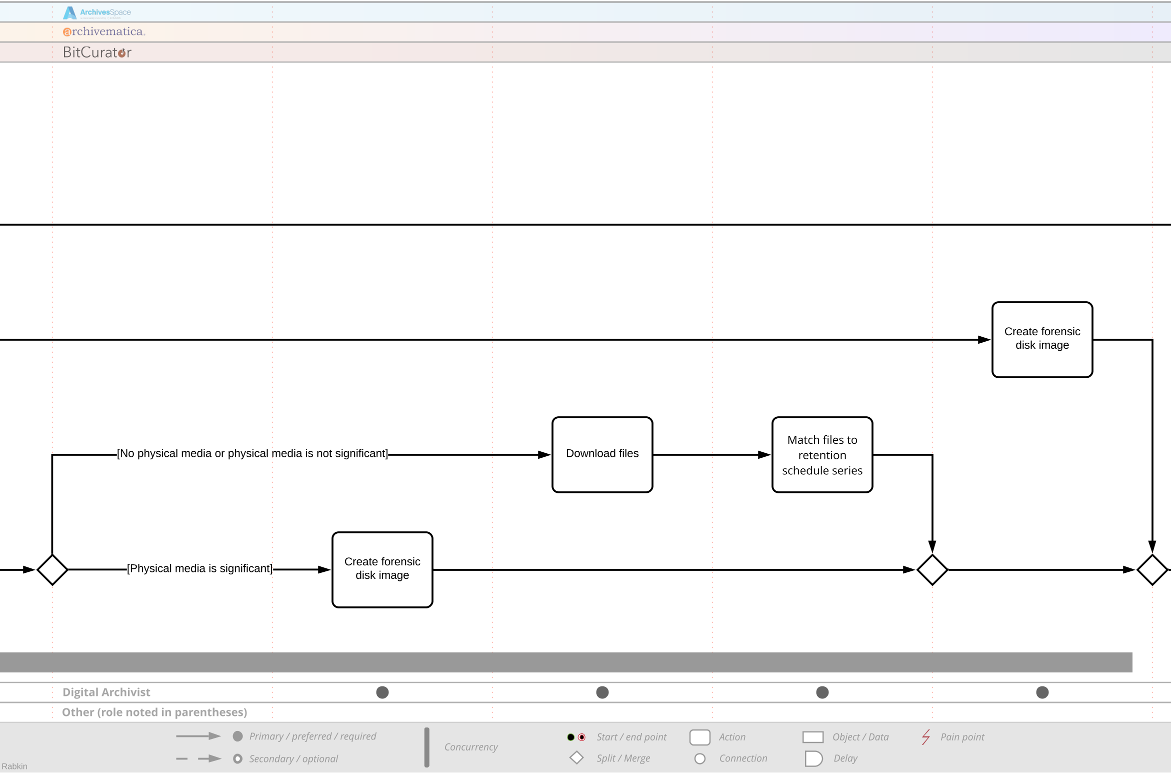
Last updated 2018-05-01 by Andrew

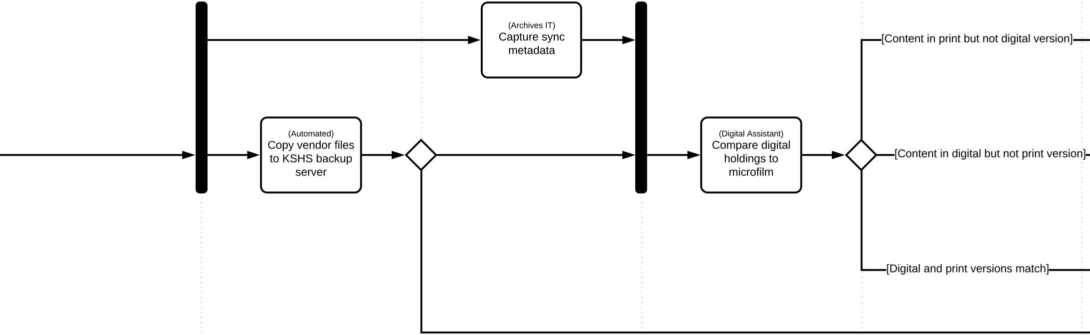
As-Is Digital Curation Workflow, August 2017

Kansas Historical Society

OssArcFlow Project

https://educopia.org/research/ossarcflow

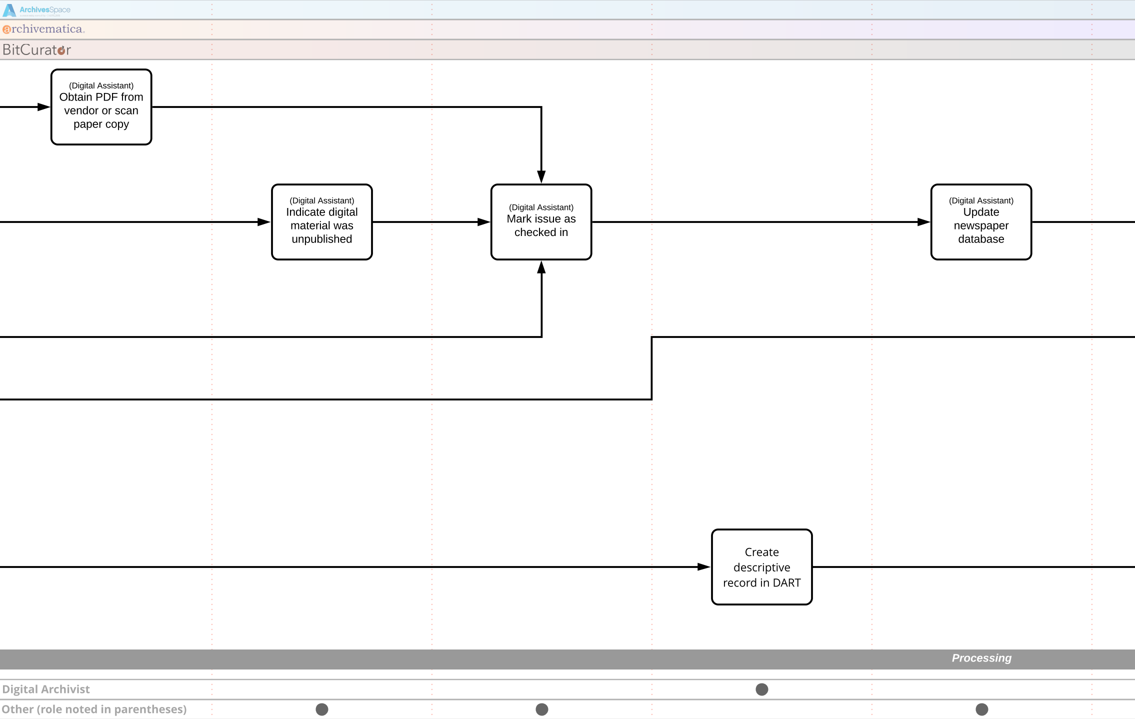

ArchivesSpace
archivematica.
BitCurator
[No physical media or physical media is not significant]
[Physical media is significant]
Create forensic disk image
Download files
Match files to retention schedule series
Create forensic disk image
Digital Archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point
Rabkin


(Archives IT)
Capture sync metadata
(Automated)
Copy vendor files to KSHS backup server
(Digital Assistant)
Compare digital holdings to microfilm
[Content in print but not digital version]
[Content in digital but not print version]
[Digital and print versions match]

Last updated 2018-05-01 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017 Kansas Historical Society ***OssArcFlow Project***

https://educopia.org/research/ossarcflow



ArchivesSpace
archivematica
BitCurator

(Digital Assistant) Obtain PDF from vendor or scan paper copy

(Digital Assistant) Indicate digital material was unpublished

(Digital Assistant) Mark issue as checked in

(Digital Assistant) Update newspaper database

Create descriptive record in DART

Processing

Digital Archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

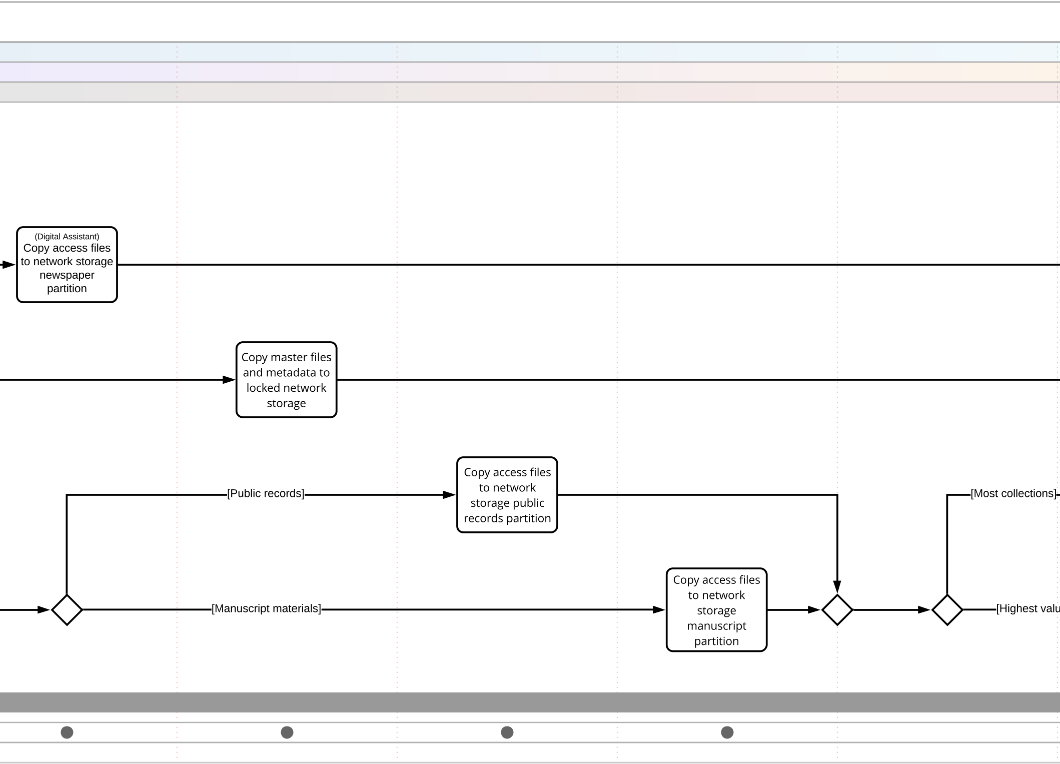
Action

Connection

Object / Data

Delay

Pain point


(Digital Assistant)
Copy access files to network storage newspaper partition
Copy master files and metadata to locked network storage
[Public records]
Copy access files to network storage public records partition
[Most collections]
[Manuscript materials]
Copy access files to network storage manuscript partition
[Highest valu

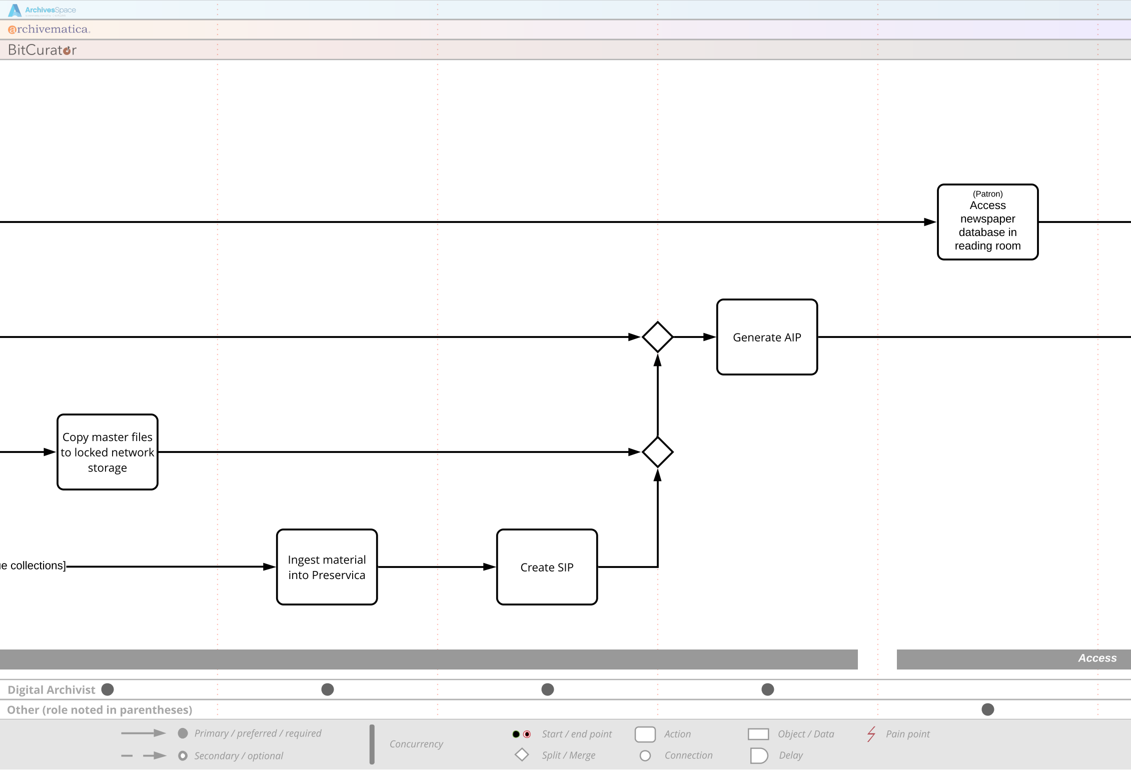
Last updated 2018-05-01 by Andrew Rabkin

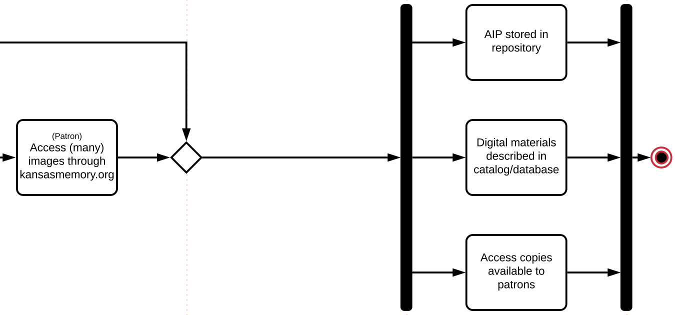
As-Is Digital Curation Workflow, August 2017

Kansas Historical Society

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica.
BitCurator
(Patron) Access newspaper database in reading room
Generate AIP
Copy master files to locked network storage
e collections]
Ingest material into Preservica
Create SIP
Access
Digital Archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


(Patron)
Access (many) images through kansasmemory.org
AIP stored in repository
Digital materials described in catalog/database
Access copies available to patrons

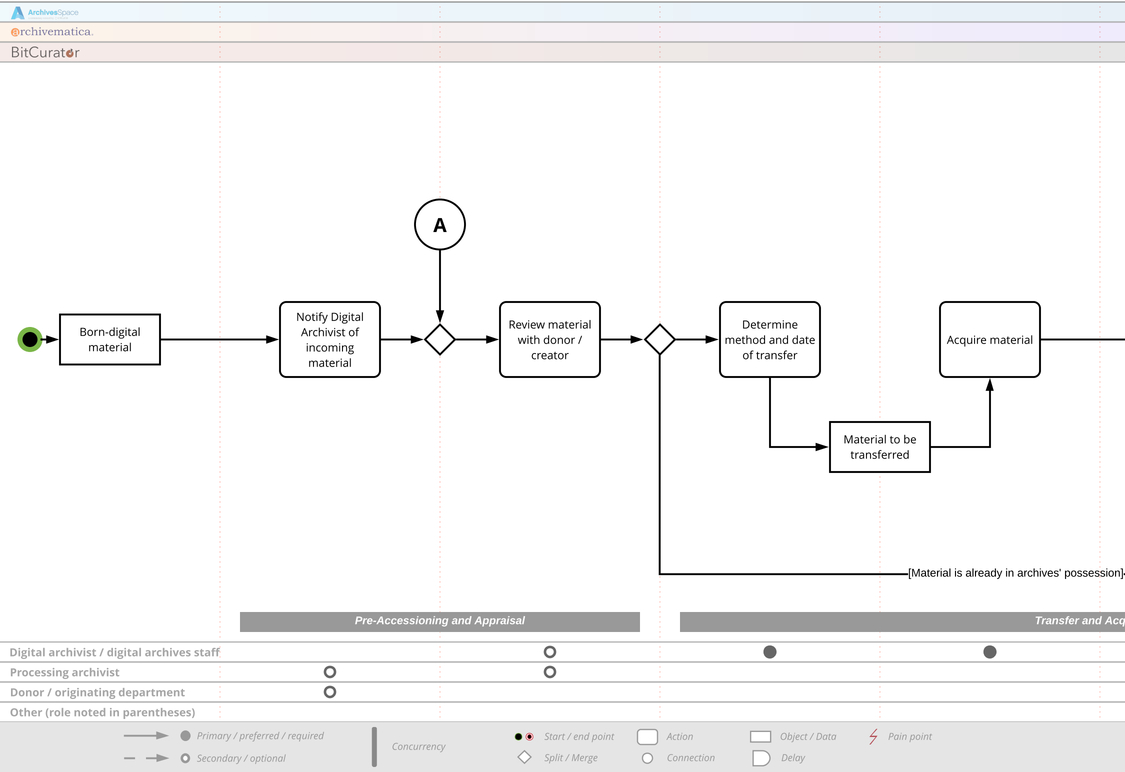
Last updated 2018-05-01 by Andrew Rabkin

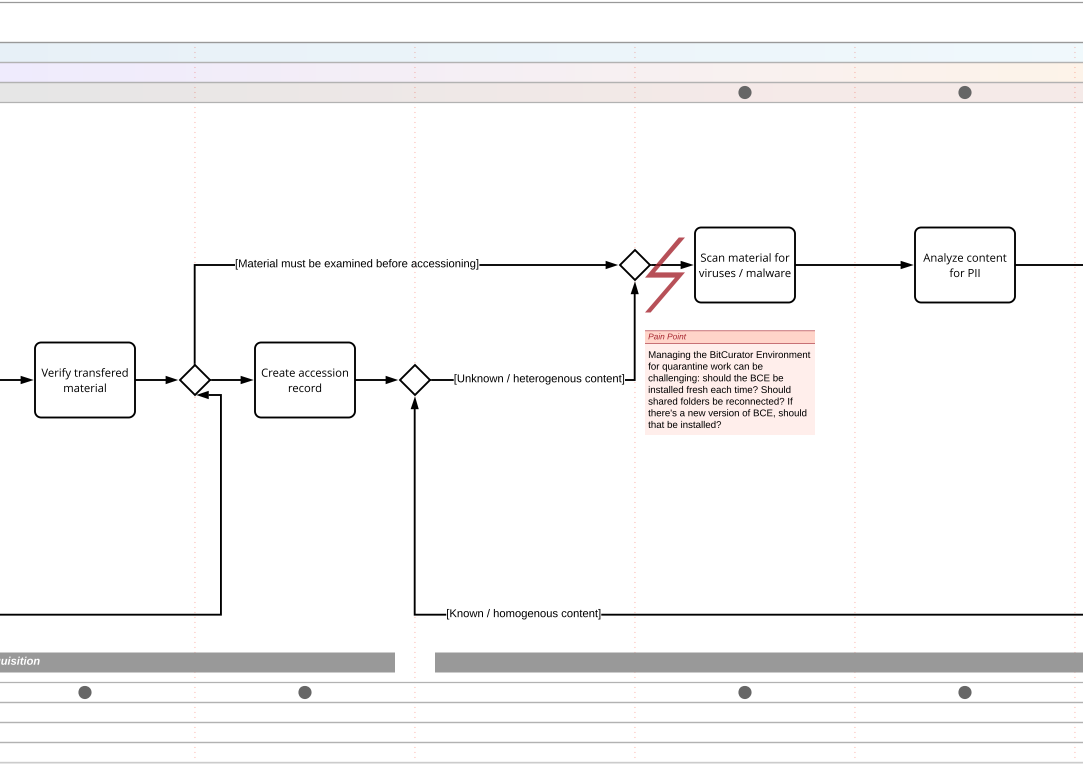
As-Is Digital Curation Workflow, August 2017

Massachusetts Institute of Technology

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
A
Born-digital material
Notify Digital Archivist of incoming material
Review material with donor / creator
Determine method and date of transfer
Acquire material
Material to be transferred
[Material is already in archives' possession]
Pre-Accessioning and Appraisal
Transfer and Acq
Digital archivist / digital archives staff
Processing archivist
Donor / originating department
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


[Material must be examined before accessioning]
Scan material for viruses / malware
Analyze content for PII
Verify transfered material
Create accession record
[Unknown / heterogenous content]
[Known / homogenous content]
uisition

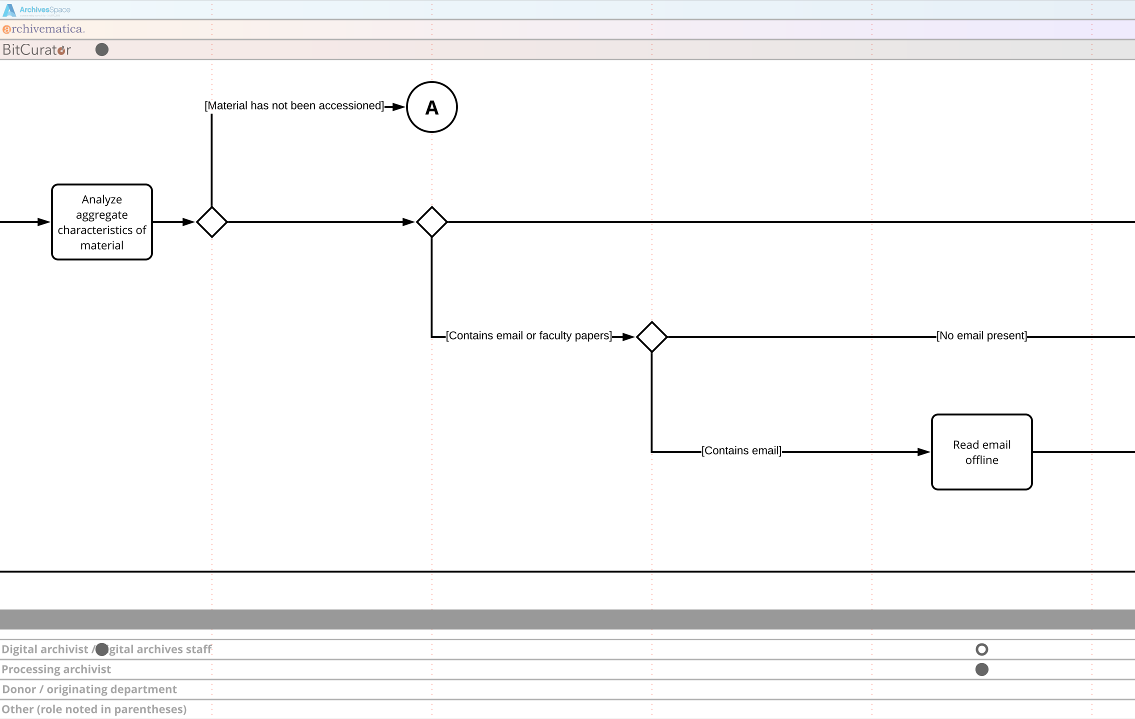
Pain Point

Managing the BitCurator Environment for quarantine work can be challenging: should the BCE be installed fresh each time? Should shared folders be reconnected? If there's a new version of BCE, should that be installed?

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017 Massachusetts Institute of Technology ***OssArcFlow Project***

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
[Material has not been accessioned]
A
Analyze aggregate characteristics of material
[Contains email or faculty papers]
[No email present]
[Contains email]
Read email offline
Digital archivist / digital archives staff
Processing archivist
Donor / originating department
Other (role noted in parentheses)

● *Primary / preferred / required*

○ *Secondary / optional*

Concurrency

Start / end point

◇ *Split / Merge*

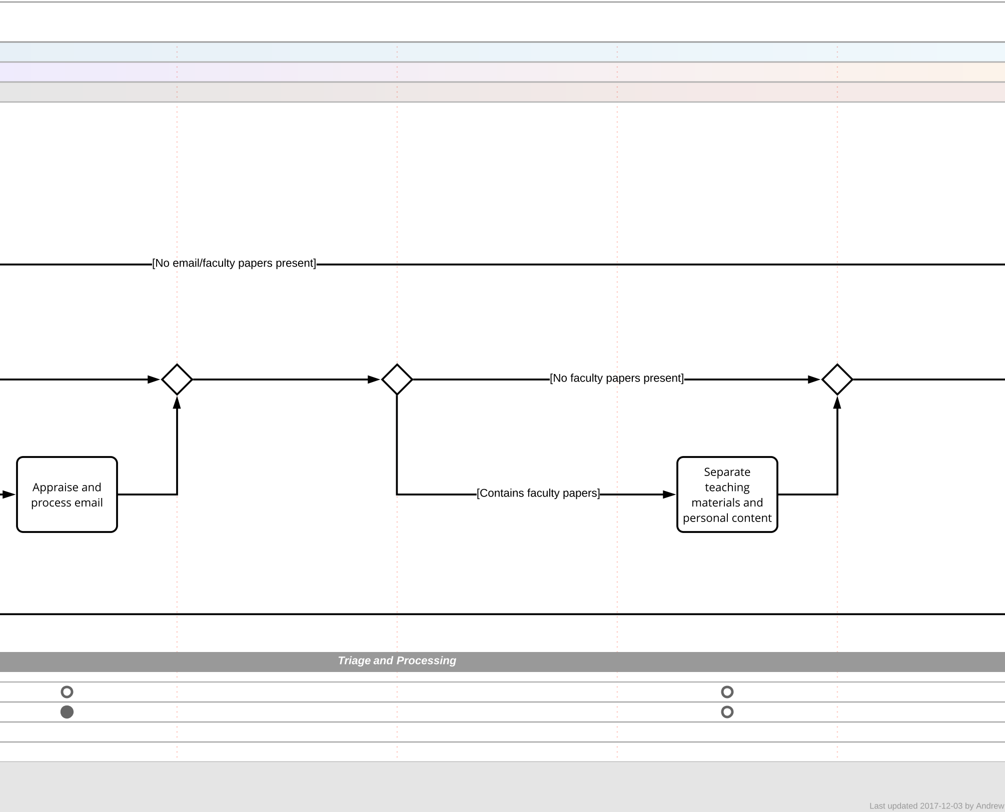
Action

○ *Connection*

Object / Data

Delay

Pain point


[No email/faculty papers present]
[No faculty papers present]
Appraise and process email
[Contains faculty papers]
Separate teaching materials and personal content
Triage and Processing

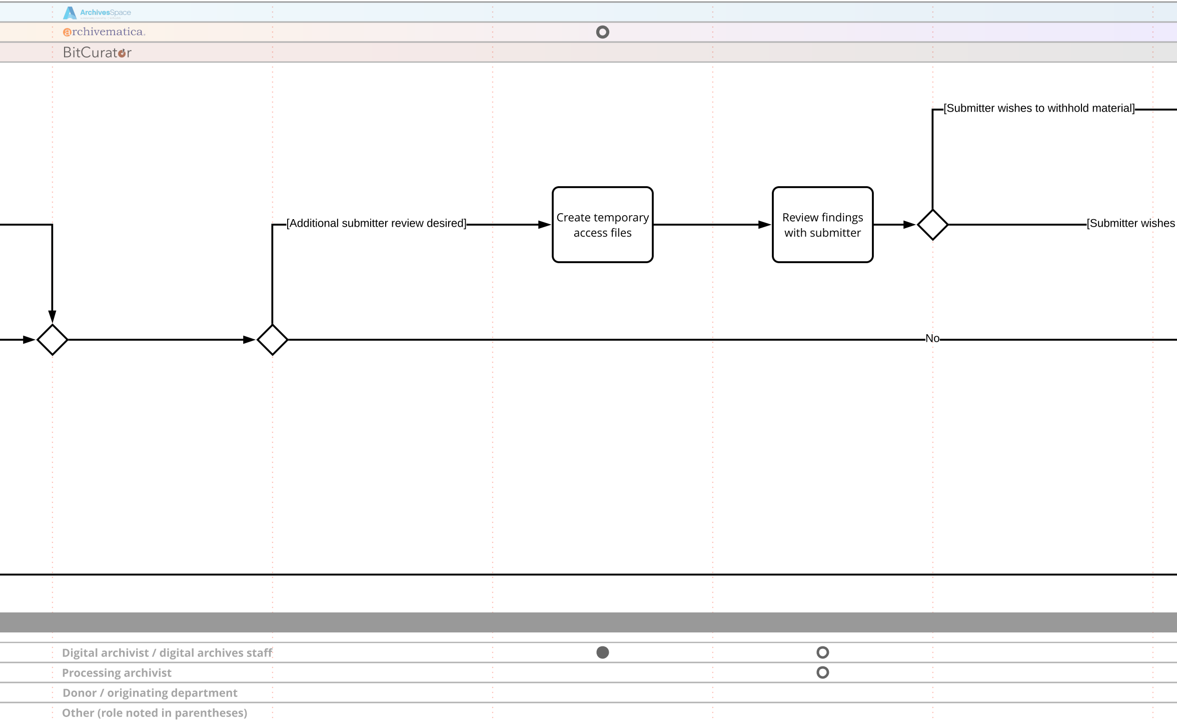
Last updated 2017-12-03 by Andrew

As-Is Digital Curation Workflow, August 2017

Massachusetts Institute of Technology

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica.
BitCurator
[Submitter wishes to withhold material]
[Additional submitter review desired]
Create temporary access files
Review findings with submitter
[Submitter wishes
No
Digital archivist / digital archives staff
Processing archivist
Donor / originating department
Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

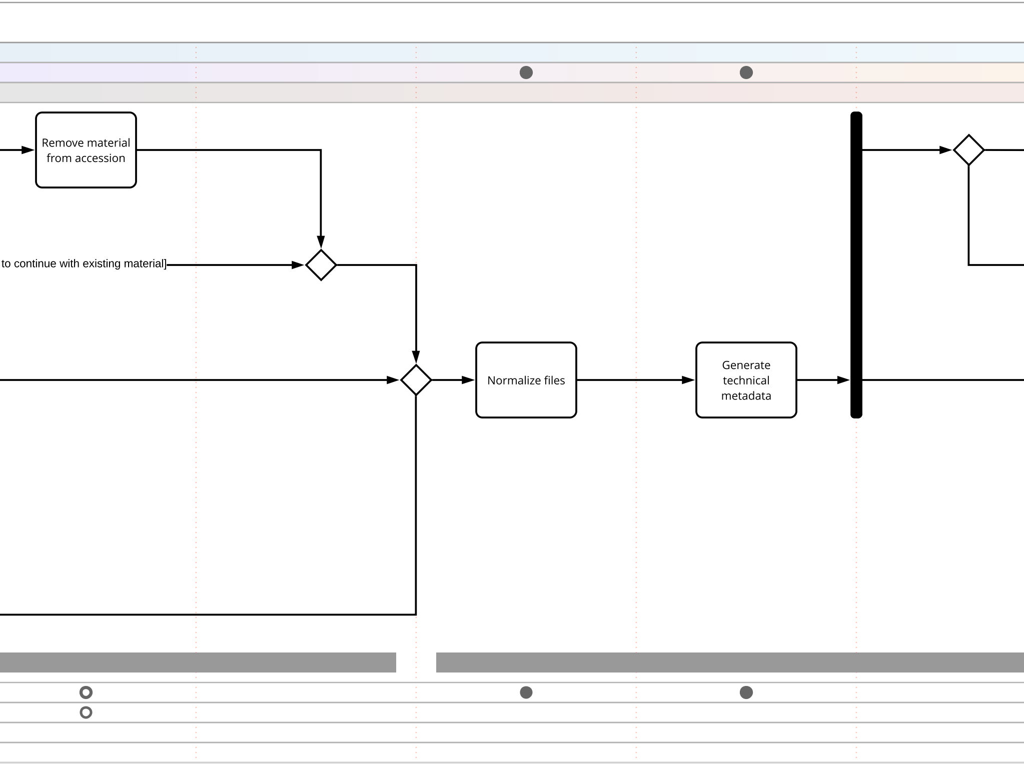
Connection

Object / Data

Delay

Pain point

Rabkin


Remove material from accession
to continue with existing material]
Normalize files
Generate technical metadata

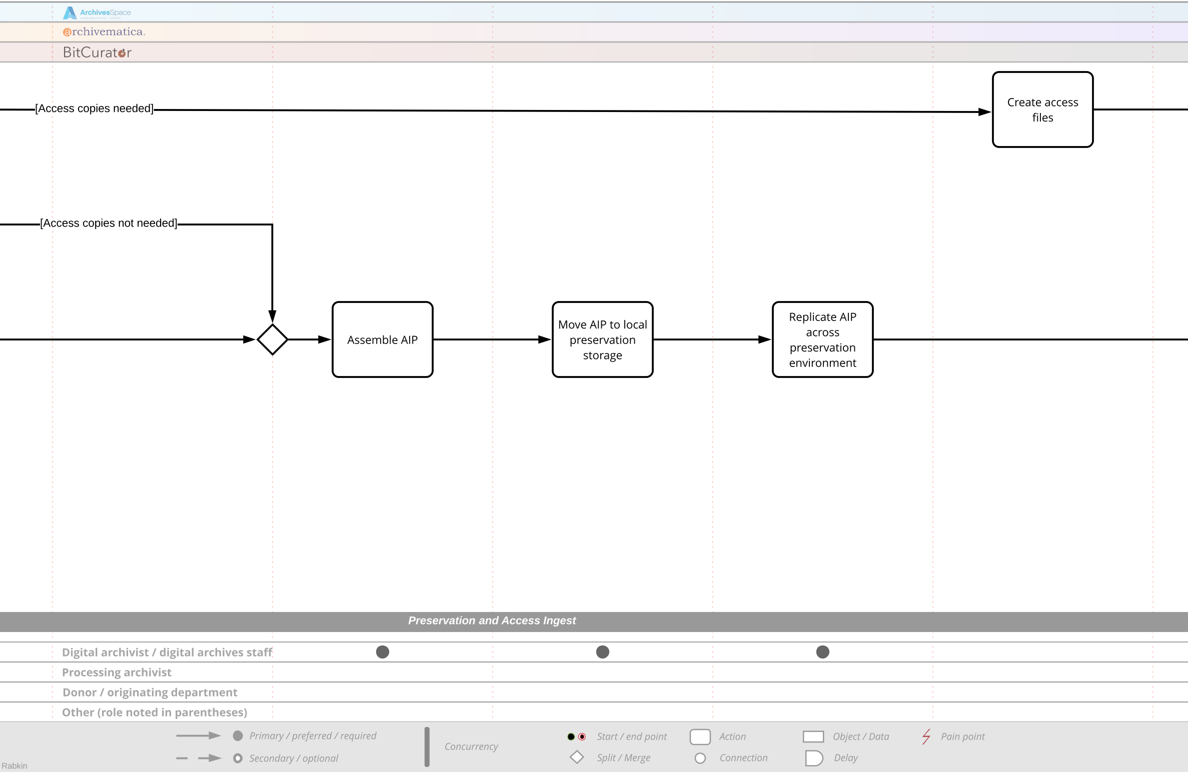
Last updated 2017-12-03 by Andrew

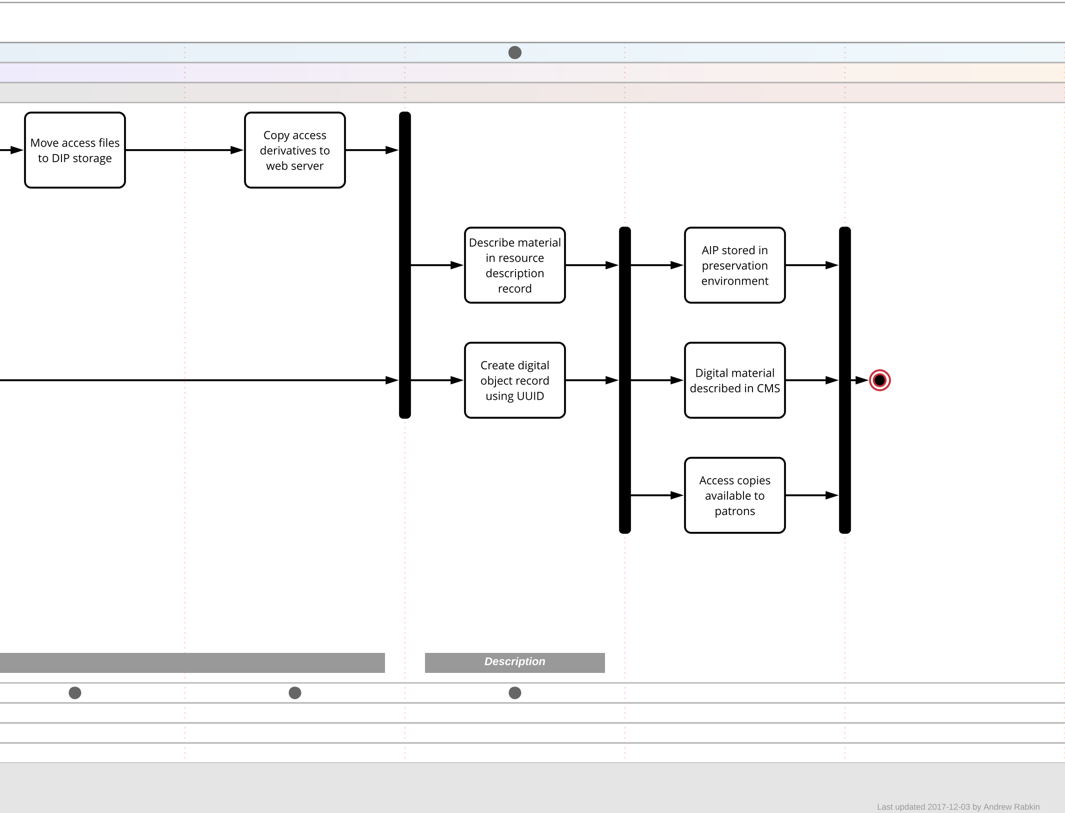
As-Is Digital Curation Workflow, August 2017

Massachusetts Institute of Technology

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
[Access copies needed]
Create access files
[Access copies not needed]
Assemble AIP
Move AIP to local preservation storage
Replicate AIP across preservation environment
Preservation and Access Ingest
Digital archivist / digital archives staff
Processing archivist
Donor / originating department
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point
Rabkin



Move access files to DIP storage
Copy access derivatives to web server
Describe material in resource description record
Create digital object record using UUID
AIP stored in preservation environment
Digital material described in CMS
Access copies available to patrons
Description

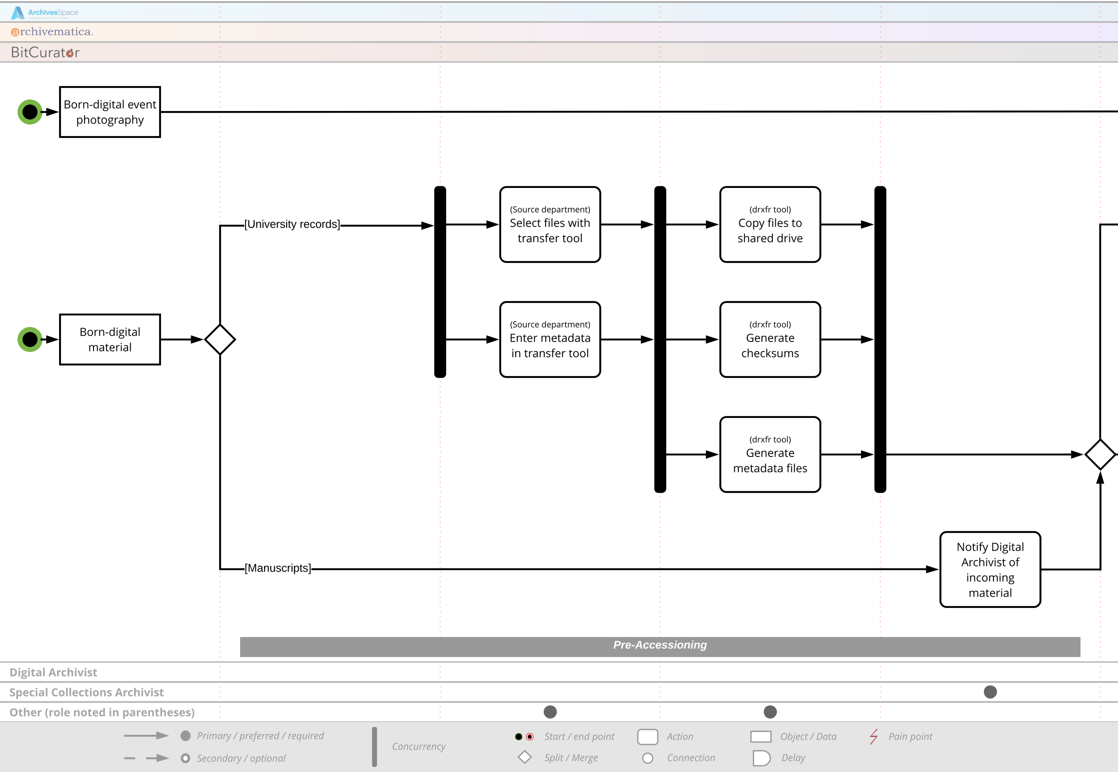
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Mount Holyoke College

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Born-digital event photography

Born-digital material

[University records]

(Source department) Select files with transfer tool

(Source department) Enter metadata in transfer tool

(drxfr tool) Copy files to shared drive

(drxfr tool) Generate checksums

(drxfr tool) Generate metadata files

[Manuscripts]

Notify Digital Archivist of incoming material

Pre-Accessioning

Digital Archivist

Special Collections Archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

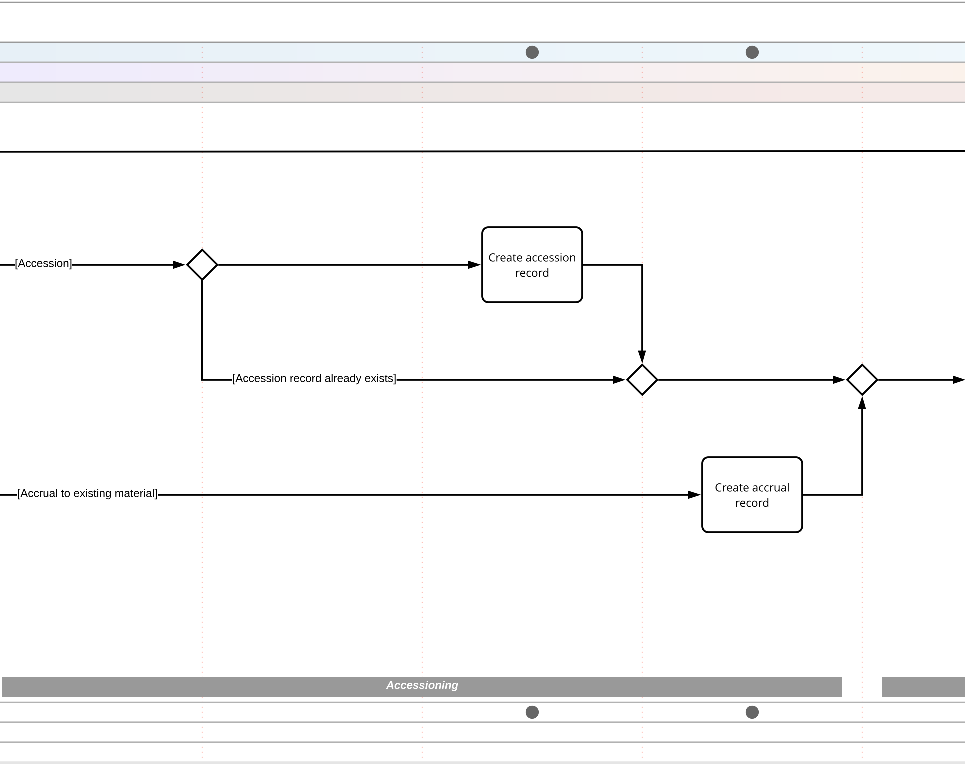
Action

Connection

Object / Data

Delay

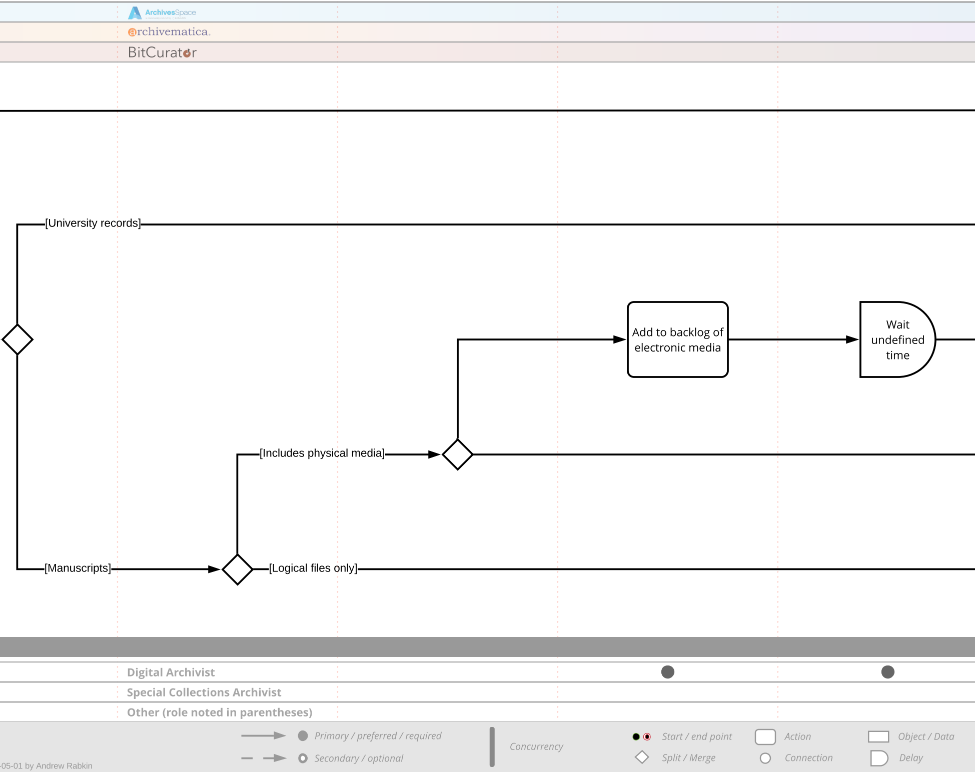
Pain point


[Accession]
Create accession record
[Accession record already exists]
[Accrual to existing material]
Create accrual record
Accessioning

Last updated 2018-

As-Is Digital Curation Workflow, August 2017

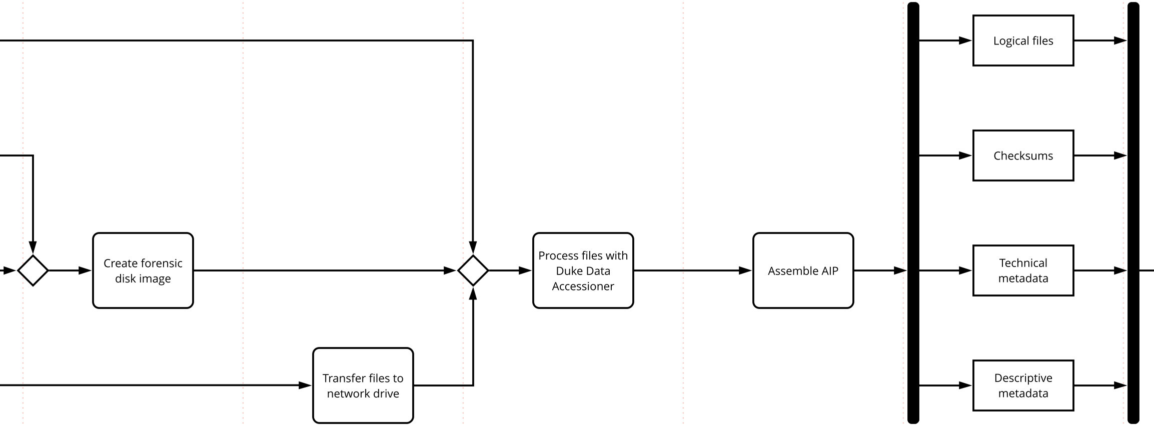
Mount Holyoke College


ArchivesSpace
archivematica.
BitCurator
[University records]
[Manuscripts]
[Includes physical media]
[Logical files only]
Add to backlog of electronic media
Wait undefined time
Digital Archivist
Special Collections Archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay

-05-01 by Andrew Rabkin

OssArcFlow Project

https://educopia.org/research/ossarcflow

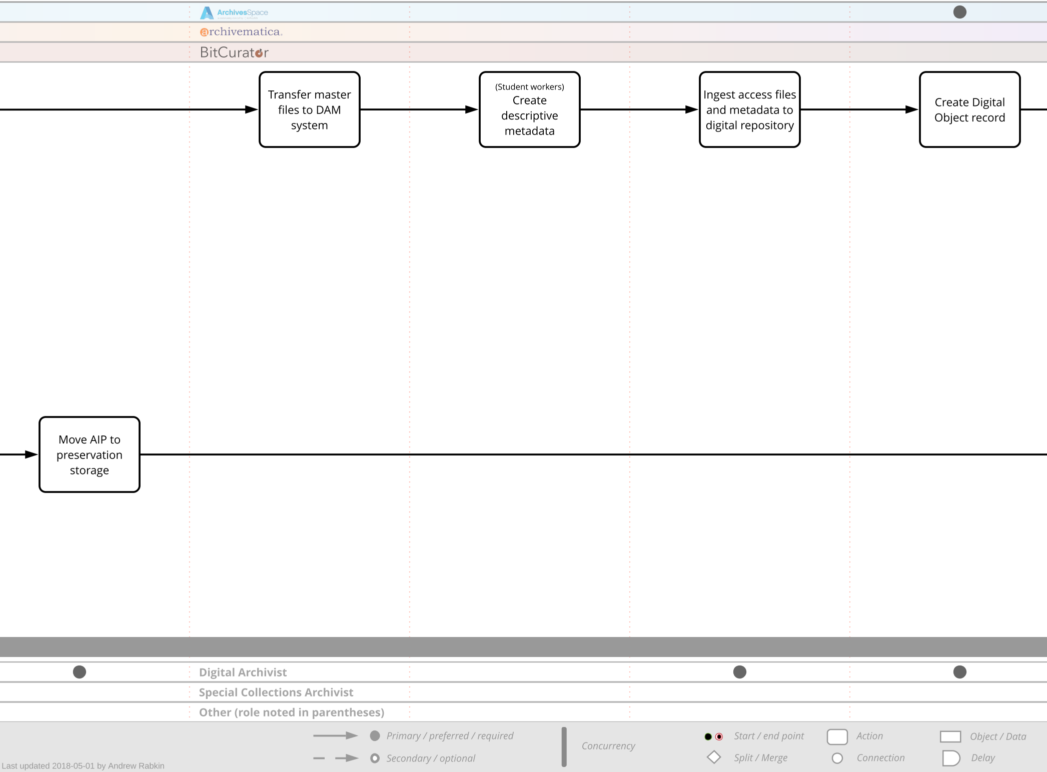

Create forensic disk image
Transfer files to network drive
Process files with Duke Data Accessioner
Assemble AIP
Logical files
Checksums
Technical metadata
Descriptive metadata

Processing

ϟ Pain point

As-Is Digital Curation Workflow, August 2017

Mount Holyoke College



ArchivesSpace

archivematica

BitCurator

Transfer master files to DAM system

(Student workers) Create descriptive metadata

Ingest access files and metadata to digital repository

Create Digital Object record

Move AIP to preservation storage

Digital Archivist

Special Collections Archivist

Other (role noted in parentheses)

● *Primary / preferred / required*

○ *Secondary / optional*

Concurrency

Start / end point

◇ *Split / Merge*

Action

○ *Connection*

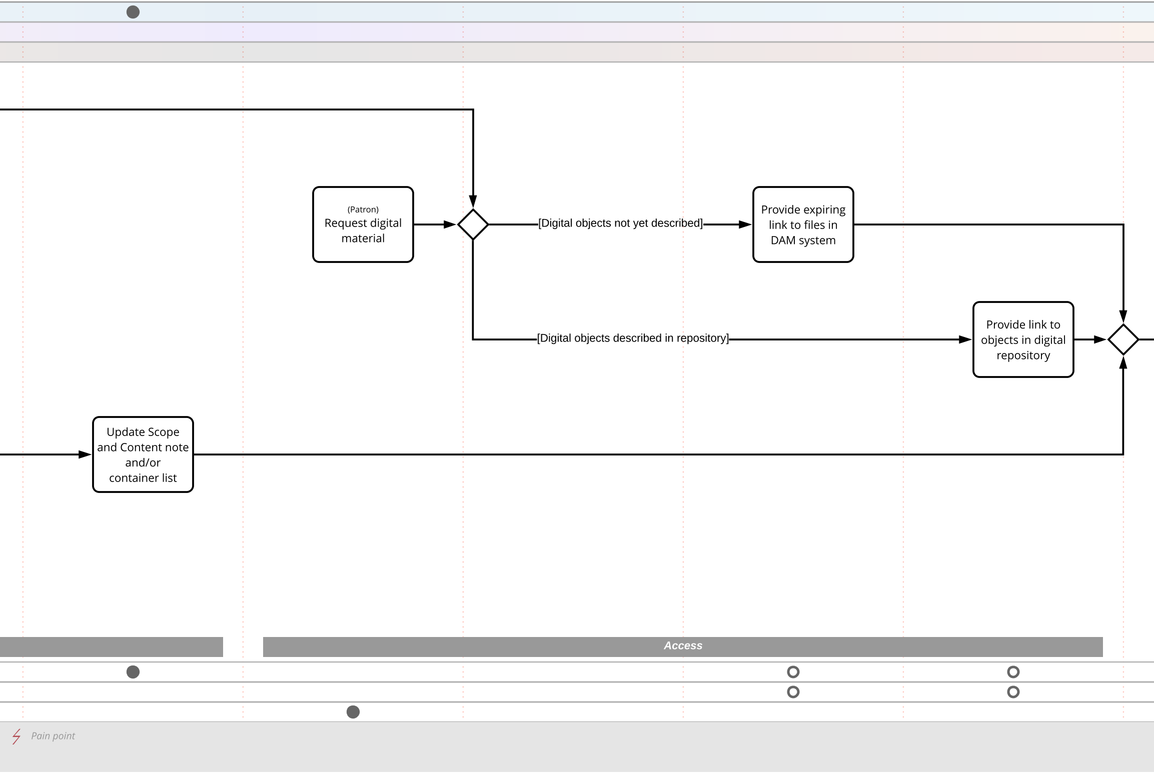
Object / Data

Delay

Last updated 2018-05-01 by Andrew Rabkin

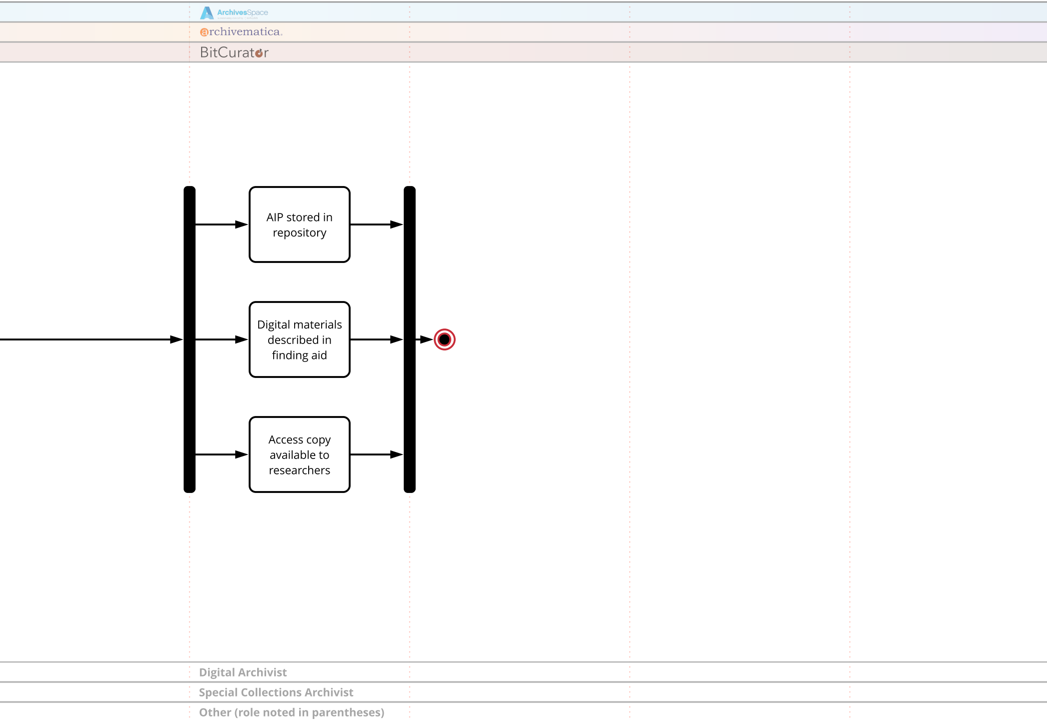
OssArcFlow Project

https://educopia.org/research/ossarcflow


(Patron)
Request digital material
[Digital objects not yet described]
Provide expiring link to files in DAM system
[Digital objects described in repository]
Provide link to objects in digital repository
Update Scope and Content note and/or container list
Access
Pain point

As-Is Digital Curation Workflow, August 2017

Mount Holyoke College


ArchivesSpace
archivematica
BitCurator
Digital Archivist
Special Collections Archivist
Other (role noted in parentheses)

AIP stored in repository

Digital materials described in finding aid

Access copy available to researchers

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

Connection

Object / Data

Delay

Last updated 2018-05-01 by Andrew Rabkin

OssArcFlow Project

https://educopia.org/research/ossarcflow

⚡ *Pain point*

Last updated 2018-05-01 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

New York Public Library

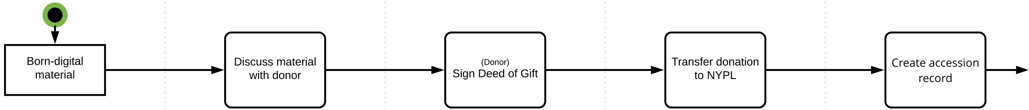
OssArcFlow Project

https://educopia.org/research/ossarcflow

ArchivesSpace

archivematica

BitCurator

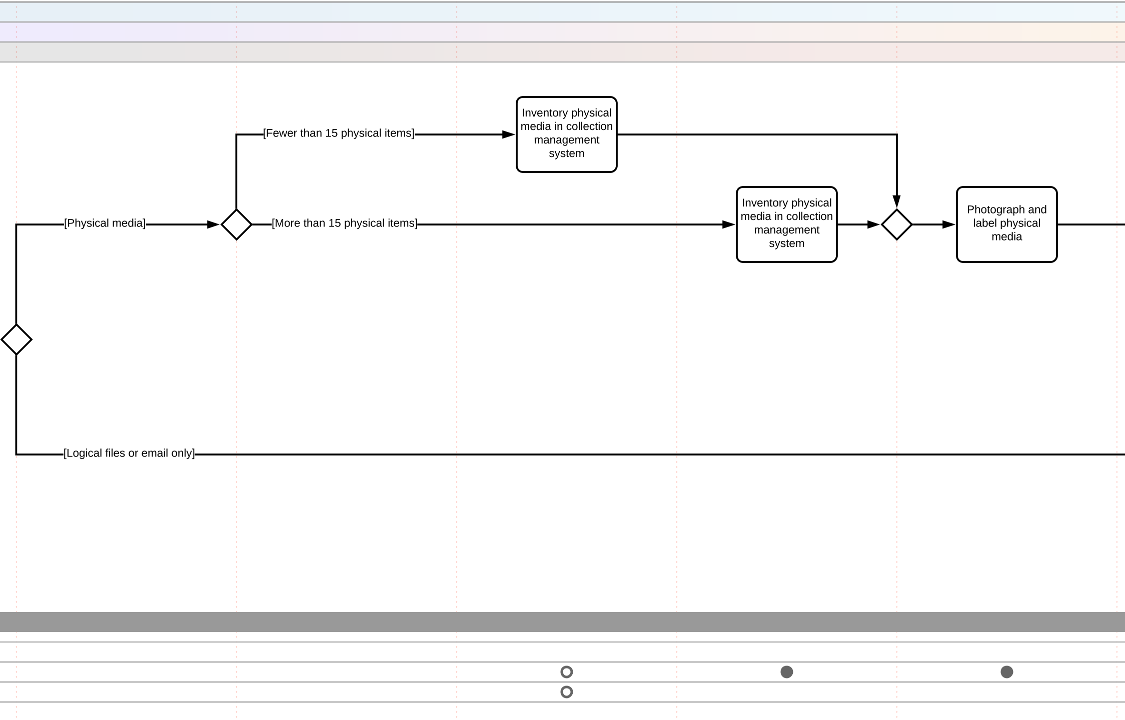


Born-digital material → Discuss material with donor → (Donor) Sign Deed of Gift → Transfer donation to NYPL → Create accession record →

Pre-Accessioning

	Discuss material with donor	Sign Deed of Gift	Transfer donation to NYPL	Create accession record
Curatorial staff	●		○	
Digital archives staff	○		○	○
Processing staff				○
Other (role noted in parentheses)		●	○	

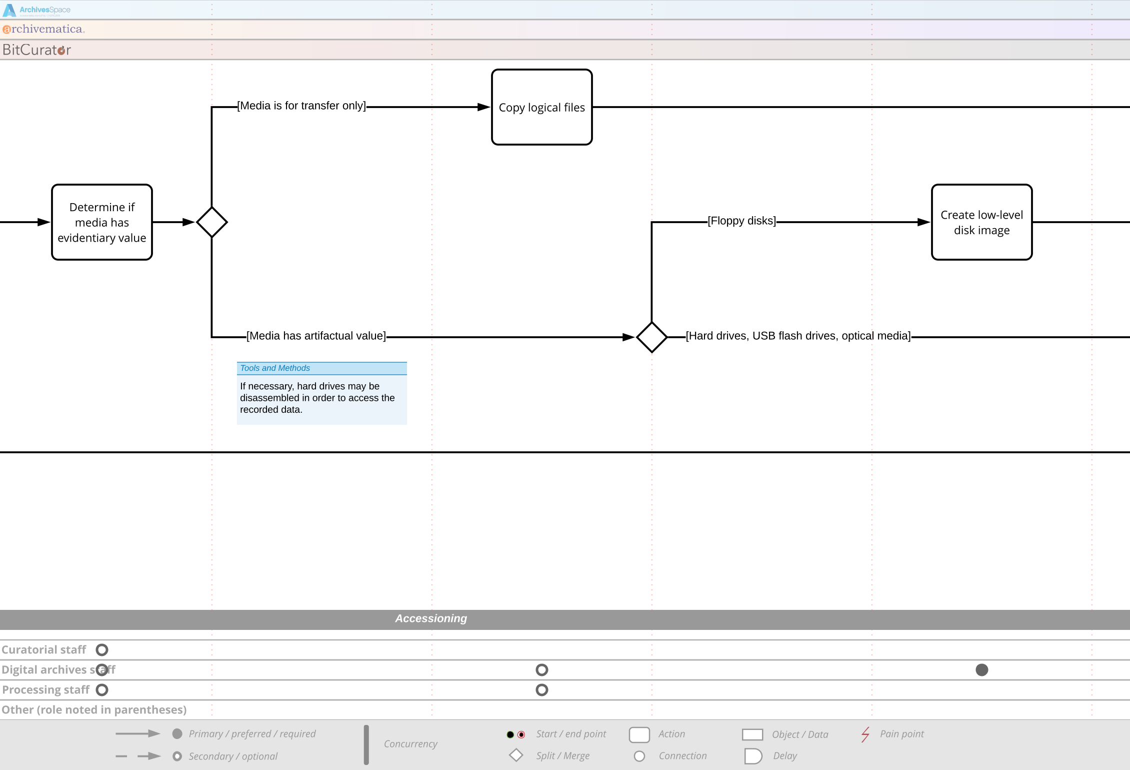
→ ● *Primary / preferred / required*
⇢ ○ *Secondary / optional*
Concurrency
● ◉ *Start / end point*
◇ *Split / Merge*
▢ *Action*
○ *Connection*
▭ *Object / Data*
◗ *Delay*
ϟ *Pain point*


[Fewer than 15 physical items]
Inventory physical media in collection management system
[Physical media]
[More than 15 physical items]
Inventory physical media in collection management system
Photograph and label physical media
[Logical files or email only]

Last updated 2018-05-02 by Andrew Rabkin

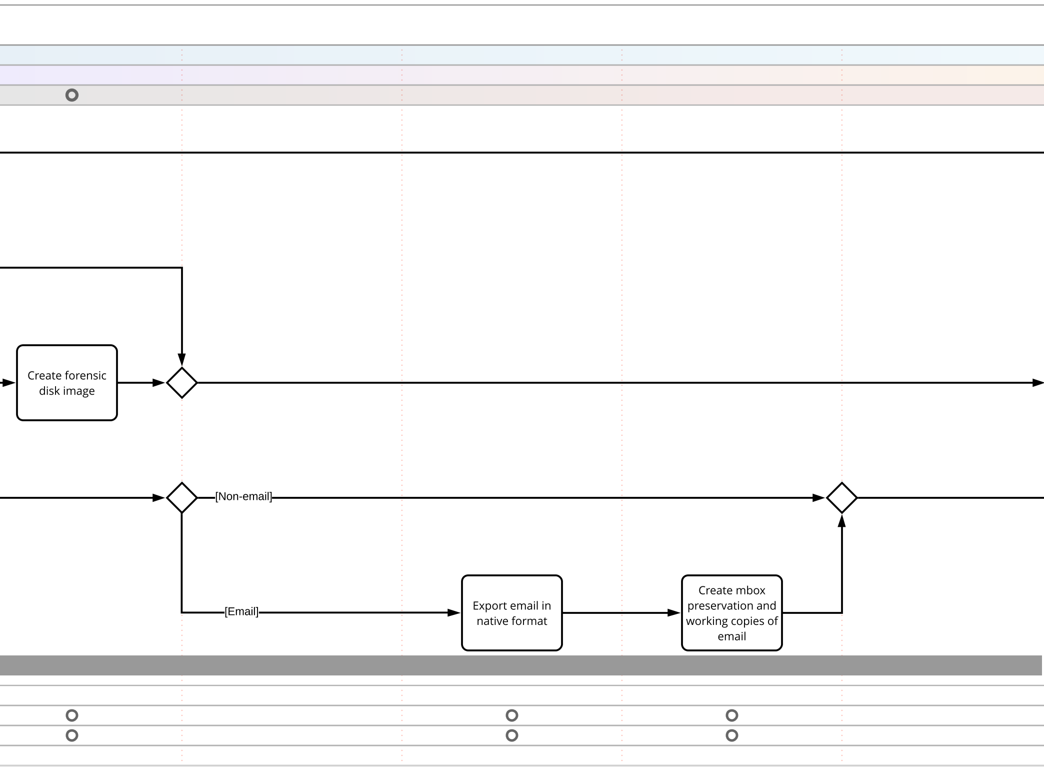
As-Is Digital Curation Workflow, August 2017 New York Public Library ***OssArcFlow Project***

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
Determine if media has evidentiary value
[Media is for transfer only]
Copy logical files
[Media has artifactual value]
[Floppy disks]
Create low-level disk image
[Hard drives, USB flash drives, optical media]
Accessioning
Curatorial staff
Digital archives staff
Processing staff
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point

Tools and Methods

If necessary, hard drives may be disassembled in order to access the recorded data.


Create forensic disk image
[Non-email]
[Email]
Export email in native format
Create mbox preservation and working copies of email

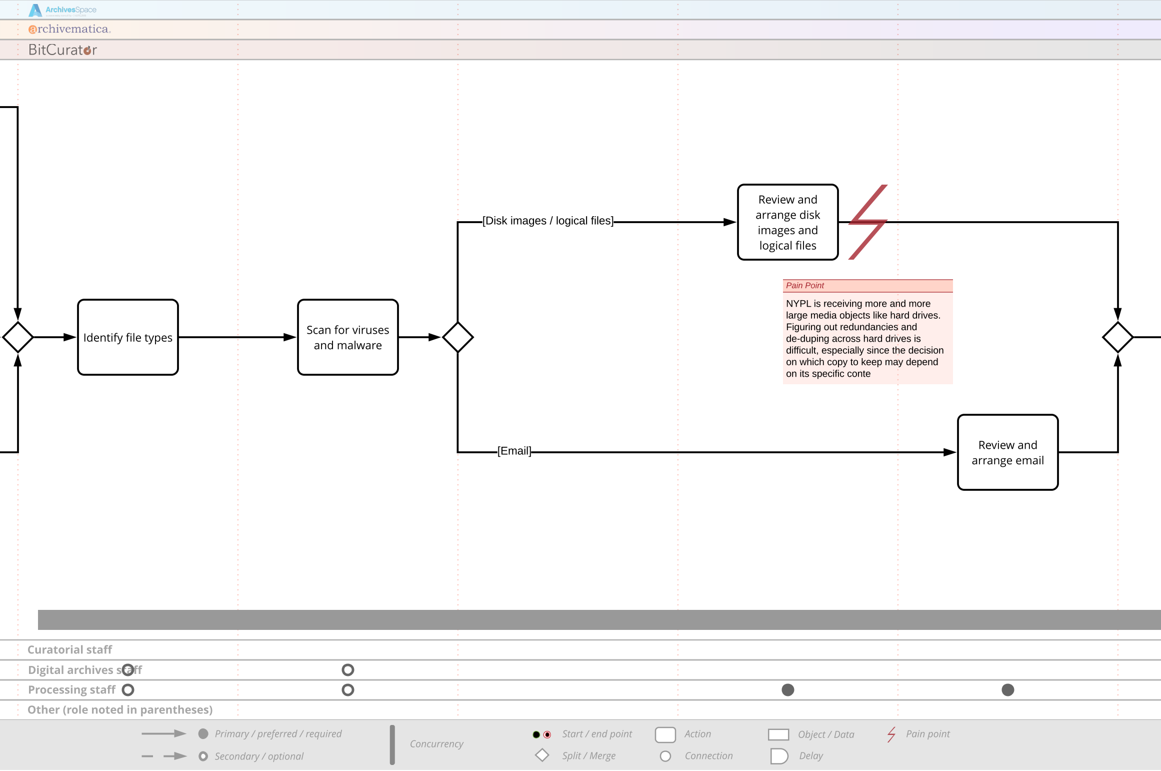
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

New York Public Library

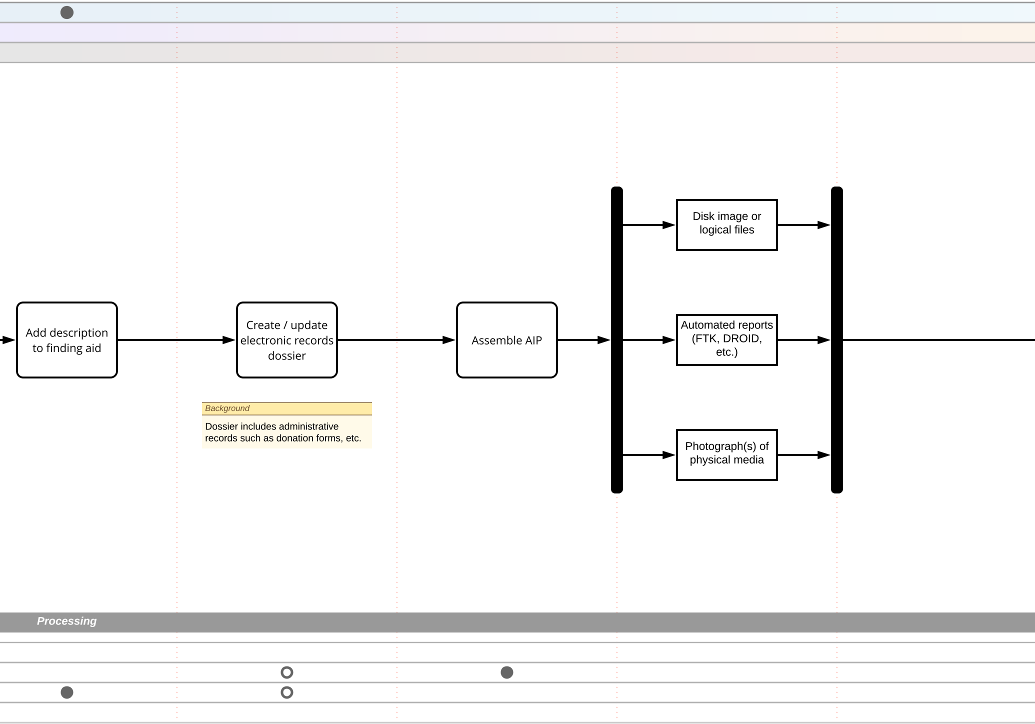
OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
Identify file types
Scan for viruses and malware
[Disk images / logical files]
Review and arrange disk images and logical files
[Email]
Review and arrange email
Curatorial staff
Digital archives staff
Processing staff
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point

Pain Point

NYPL is receiving more and more large media objects like hard drives. Figuring out redundancies and de-duping across hard drives is difficult, especially since the decision on which copy to keep may depend on its specific conte


Add description to finding aid
Create / update electronic records dossier
Assemble AIP
Disk image or logical files
Automated reports (FTK, DROID, etc.)
Photograph(s) of physical media
Processing

Background

Dossier includes administrative records such as donation forms, etc.

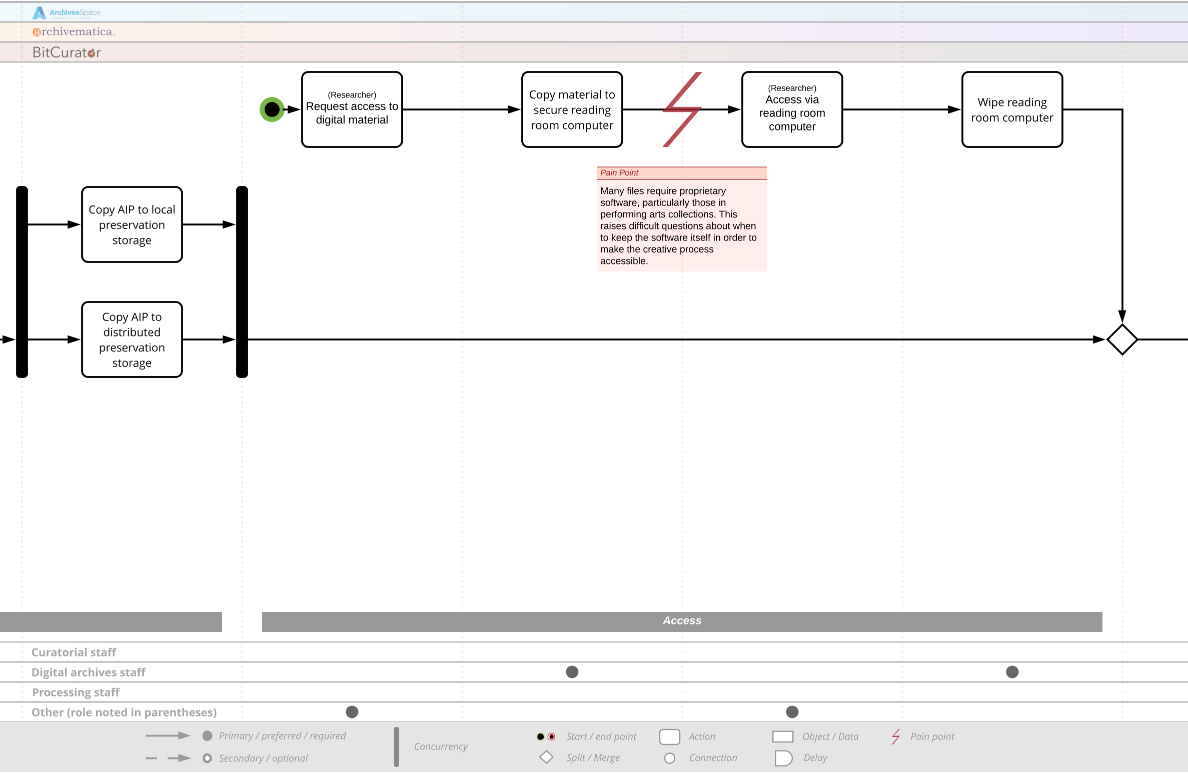
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

New York Public Library

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

(Researcher)
Request access to digital material

Copy material to secure reading room computer

(Researcher)
Access via reading room computer

Wipe reading room computer

Pain Point

Many files require proprietary software, particularly those in performing arts collections. This raises difficult questions about when to keep the software itself in order to make the creative process accessible.

Copy AIP to local preservation storage

Copy AIP to distributed preservation storage

Access

Curatorial staff

Digital archives staff

Processing staff

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

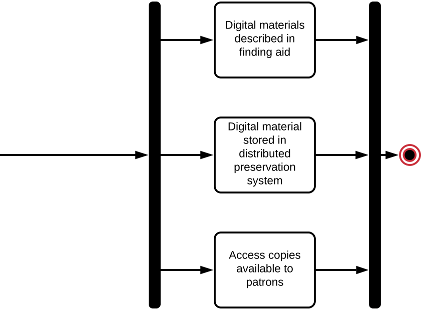
Action

Connection

Object / Data

Delay

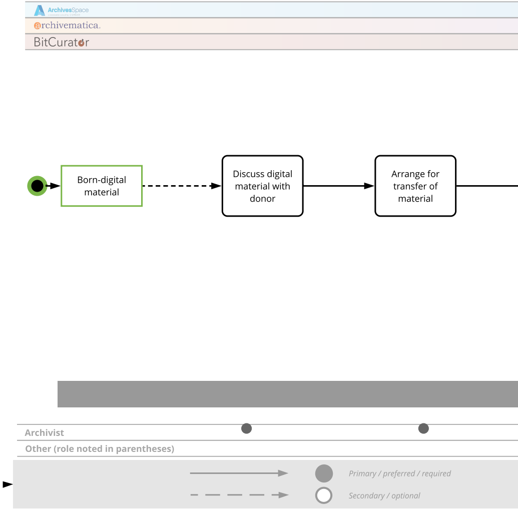
Pain point


Digital materials described in finding aid
Digital material stored in distributed preservation system
Access copies available to patrons

Last updated 2018-05-02 by Andrew Rabkin

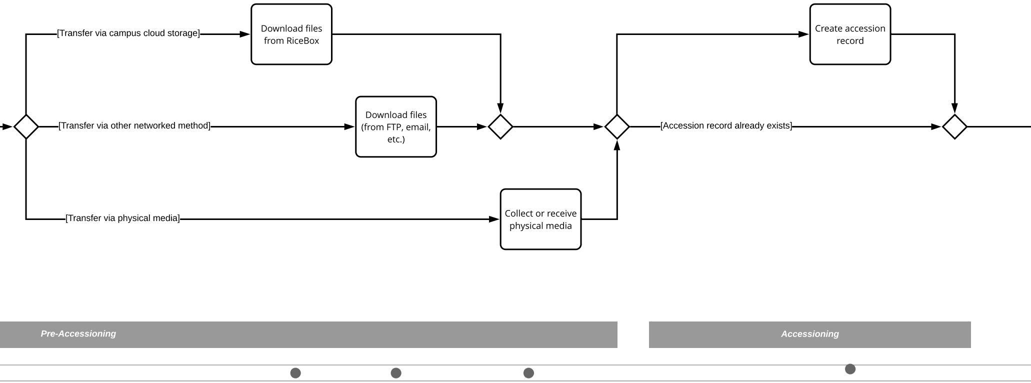
As-Is Digital Curation Workflow, August 2017

Rice University


ArchivesSpace
archivematica.
BitCurator
Born-digital material
Discuss digital material with donor
Arrange for transfer of material
Archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional

OssArcFlow Project

https://educopia.org/research/ossarcflow


[Transfer via campus cloud storage]
Download files from RiceBox
[Transfer via other networked method]
Download files (from FTP, email, etc.)
[Transfer via physical media]
Collect or receive physical media
Create accession record
[Accession record already exists]
Pre-Accessioning
Accessioning

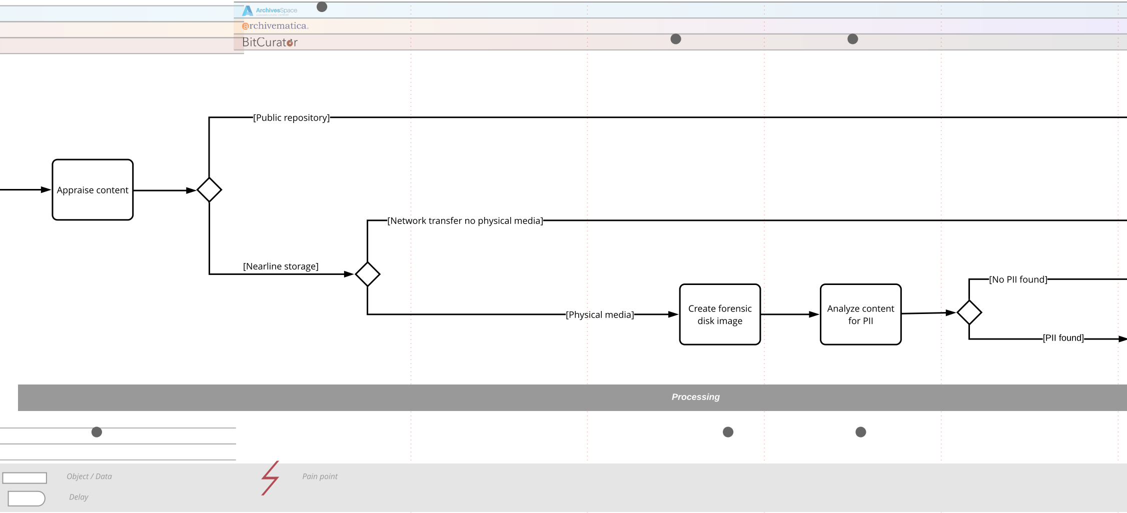
Concurrency
Start / end point
Split / Merge
Action
Connection

As-Is Digital Curation Workflow, August 2017

Rice University

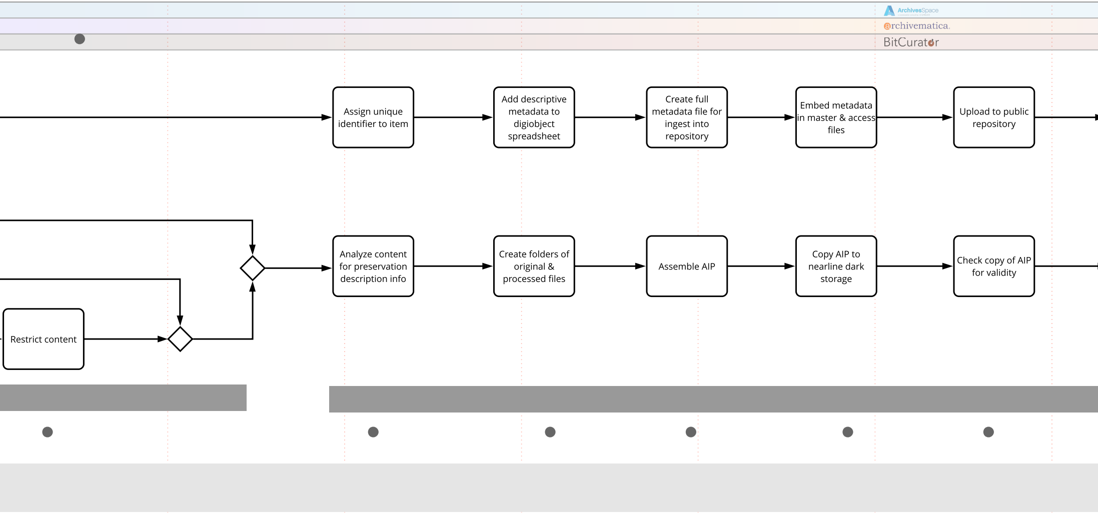
OssArcFlow Project

https://educopia.org/research/ossarcflow



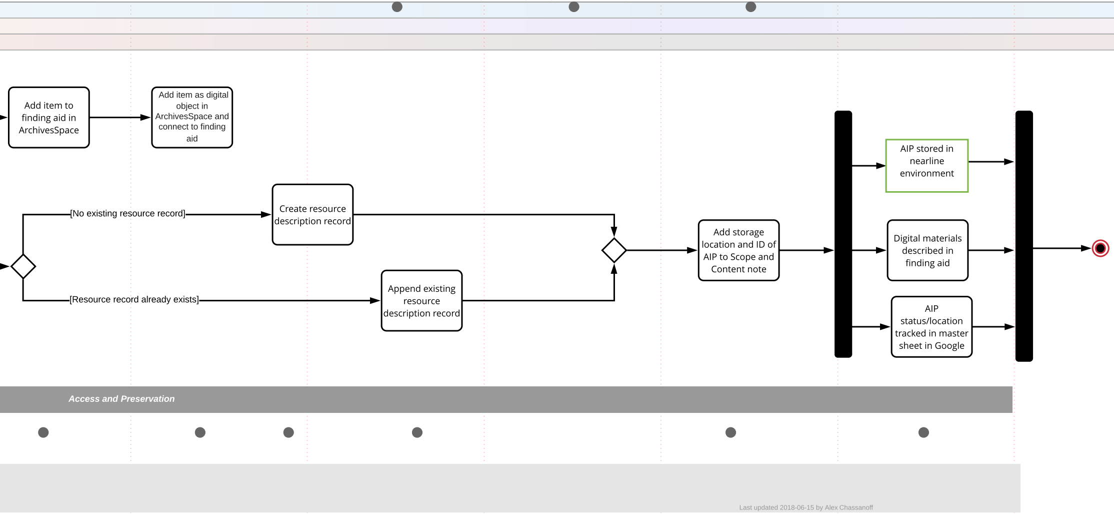
ArchivesSpace
archivematica.
BitCurator
Appraise content
[Public repository]
[Nearline storage]
[Network transfer no physical media]
[Physical media]
Create forensic disk image
Analyze content for PII
[No PII found]
[PII found]
Processing
Object / Data
Delay
Pain point

As-Is Digital Curation Workflow,


ArchivesSpace
archivematica.
BitCurator
Assign unique identifier to item
Add descriptive metadata to digiobject spreadsheet
Create full metadata file for ingest into repository
Embed metadata in master & access files
Upload to public repository
Analyze content for preservation description info
Create folders of original & processed files
Assemble AIP
Copy AIP to nearline dark storage
Check copy of AIP for validity
Restrict content

August 2017 Rice University ***OssArcFlow Project***

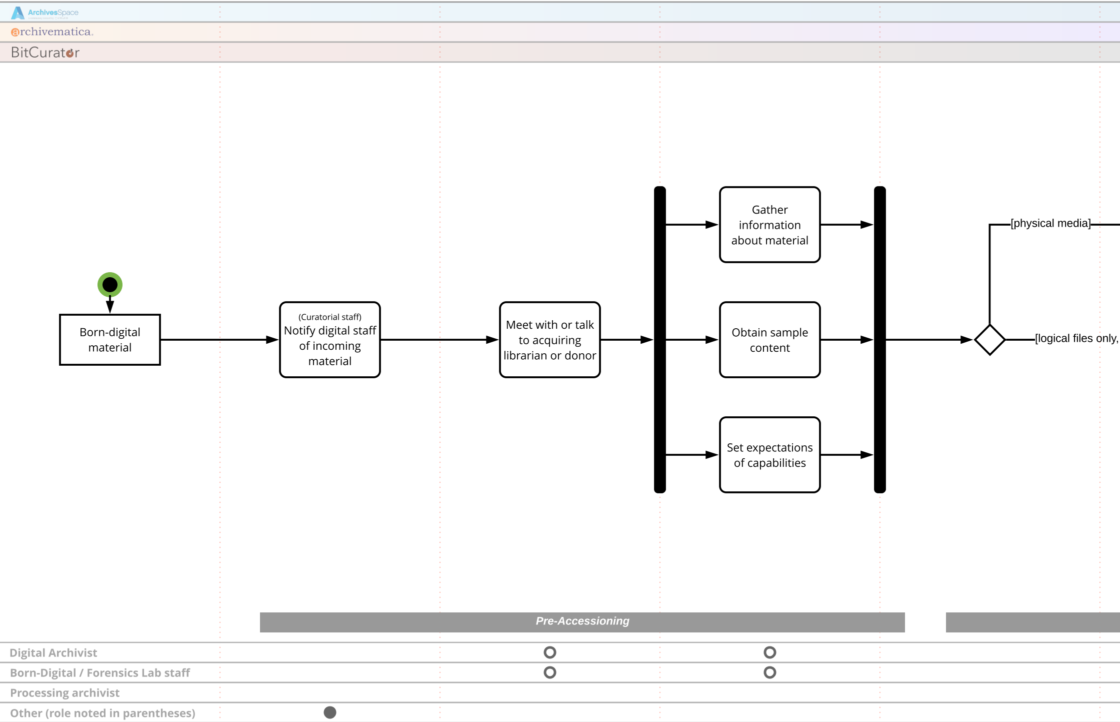
https://educopia.org/research/ossarcflow


Add item to finding aid in ArchivesSpace
Add item as digital object in ArchivesSpace and connect to finding aid
[No existing resource record]
Create resource description record
[Resource record already exists]
Append existing resource description record
Add storage location and ID of AIP to Scope and Content note
AIP stored in nearline environment
Digital materials described in finding aid
AIP status/location tracked in master sheet in Google
Access and Preservation

Last updated 2018-06-15 by Alex Chassanoff

As-Is Digital Curation Workflow, August 2017 Stanford University

OssArcFlow Project
https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Born-digital material

(Curatorial staff)
Notify digital staff of incoming material

Meet with or talk to acquiring librarian or donor

Gather information about material

Obtain sample content

Set expectations of capabilities

[physical media]

[logical files only,

Pre-Accessioning

Digital Archivist

Born-Digital / Forensics Lab staff

Processing archivist

Other (role noted in parentheses)

→ ● *Primary / preferred / required*

⇢ ○ *Secondary / optional*

| *Concurrency*

● ◉ *Start / end point*

◇ *Split / Merge*

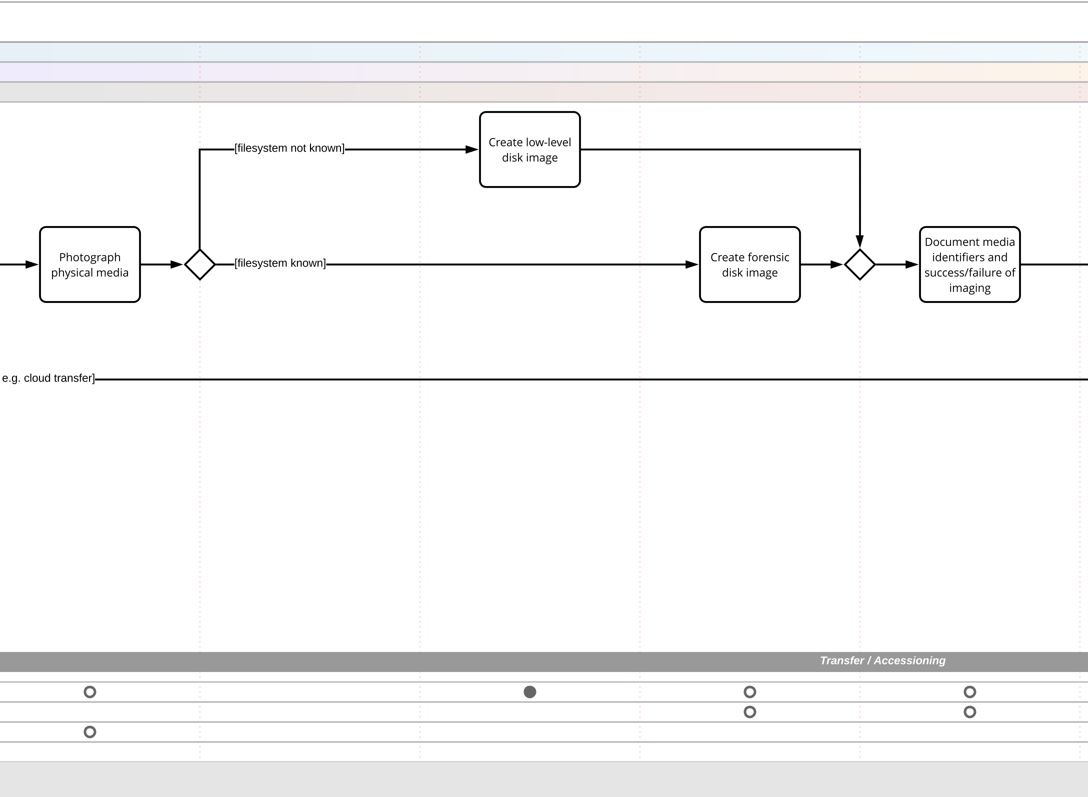
▢ *Action*

○ *Connection*

▭ *Object / Data*

◗ *Delay*

ϟ *Pain point*


Create low-level disk image
[filesystem not known]
Photograph physical media
[filesystem known]
Create forensic disk image
Document media identifiers and success/failure of imaging
e.g. cloud transfer]
Transfer / Accessioning

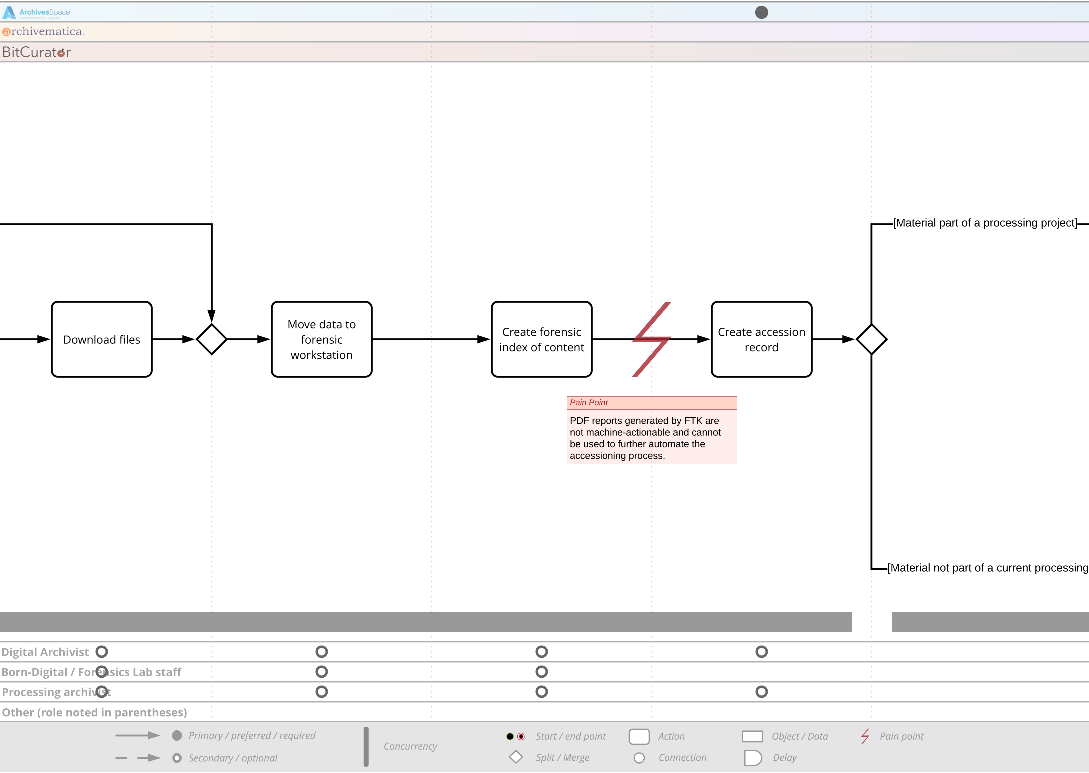
Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

Stanford University

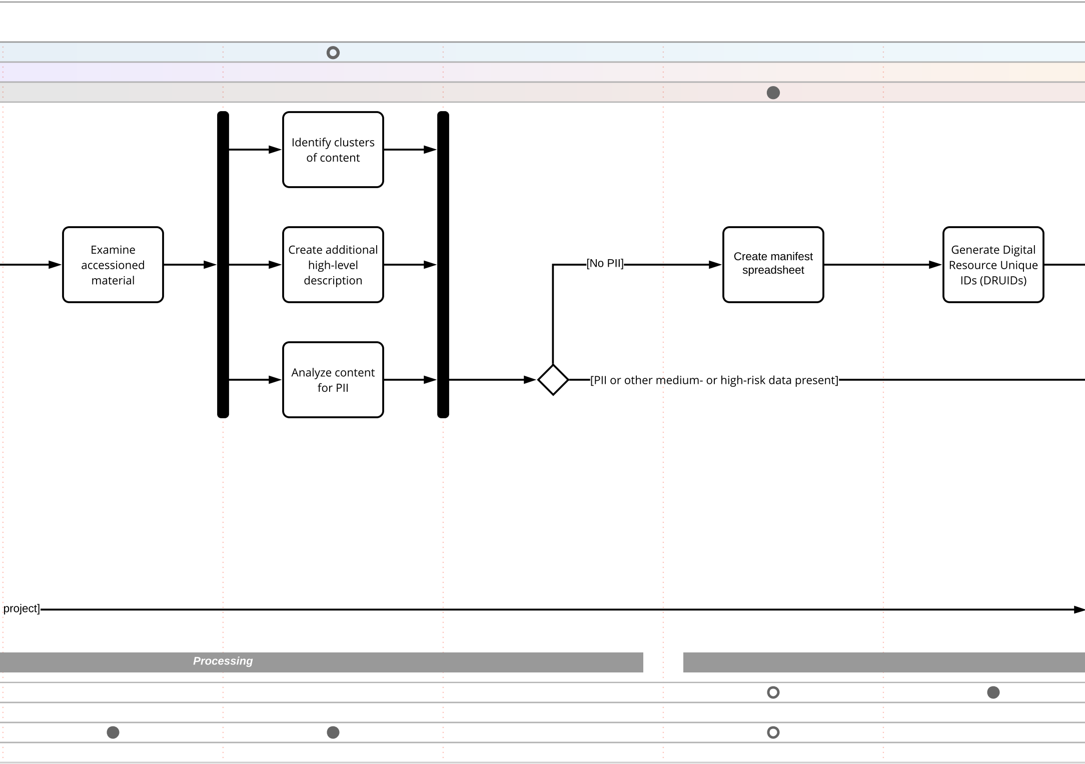
OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
Download files
Move data to forensic workstation
Create forensic index of content
Create accession record
[Material part of a processing project]
[Material not part of a current processing
Digital Archivist
Born-Digital / Forensics Lab staff
Processing archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point

Pain Point

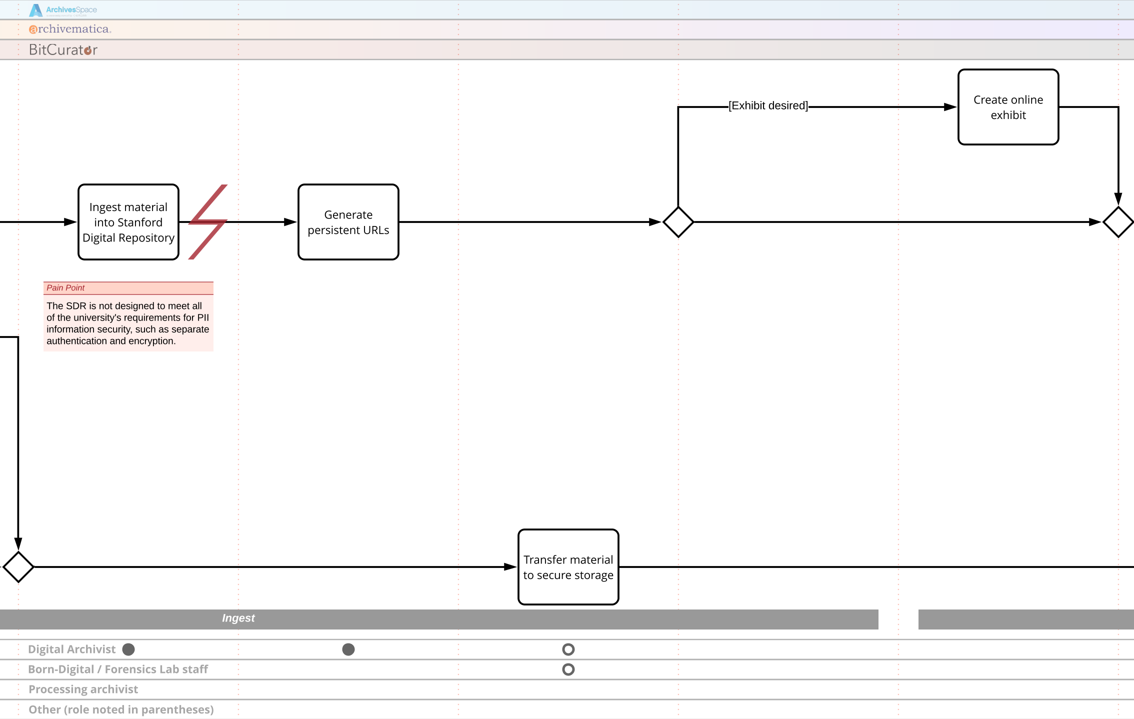
PDF reports generated by FTK are not machine-actionable and cannot be used to further automate the accessioning process.


Examine accessioned material
Identify clusters of content
Create additional high-level description
Analyze content for PII
[No PII]
[PII or other medium- or high-risk data present]
Create manifest spreadsheet
Generate Digital Resource Unique IDs (DRUIDs)
project]
Processing

Last updated 2018-05-02 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017 Stanford University ***OssArcFlow Project***

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Ingest material into Stanford Digital Repository

Generate persistent URLs

[Exhibit desired]

Create online exhibit

Pain Point

The SDR is not designed to meet all of the university's requirements for PII information security, such as separate authentication and encryption.

Transfer material to secure storage

Ingest

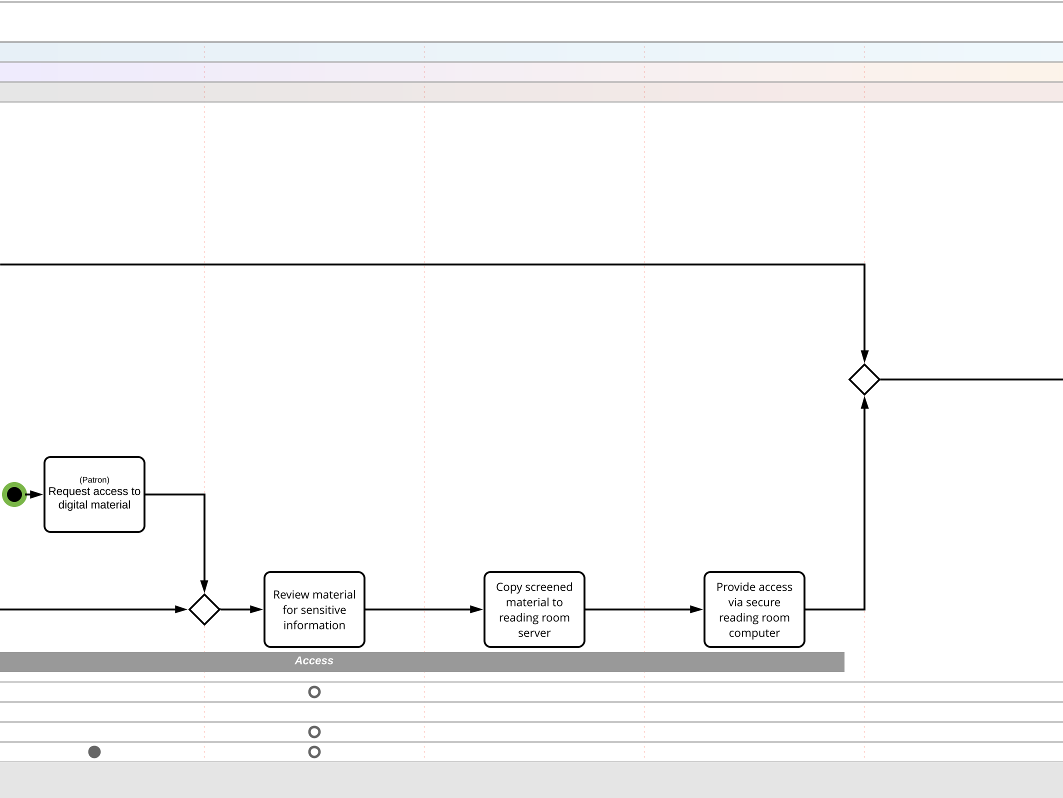
Digital Archivist ● ● ○

Born-Digital / Forensics Lab staff ○

Processing archivist

Other (role noted in parentheses)

- ● *Primary / preferred / required*
- ○ *Secondary / optional*
- *Concurrency*
- ● ◉ *Start / end point*
- ◇ *Split / Merge*
- ▢ *Action*
- ○ *Connection*
- ▭ *Object / Data*
- ◗ *Delay*
- ϟ *Pain point*


(Patron)
Request access to digital material
Review material for sensitive information
Copy screened material to reading room server
Provide access via secure reading room computer
Access

Last updated 2018-05-02 by Andrew Rabkin

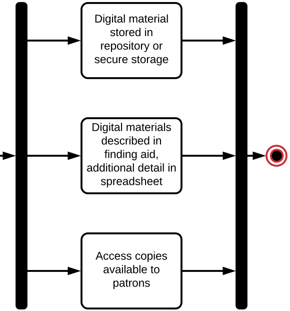
As-Is Digital Curation Workflow, August 2017 Stanford University ***OssArcFlow Project***

https://educopia.org/research/ossarcflow

ArchivesSpace

archivematica.

BitCurator



Digital material stored in repository or secure storage

Digital materials described in finding aid, additional detail in spreadsheet

Access copies available to patrons

Digital Archivist

Born-Digital / Forensics Lab staff

Processing archivist

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

Connection

Object / Data

Delay

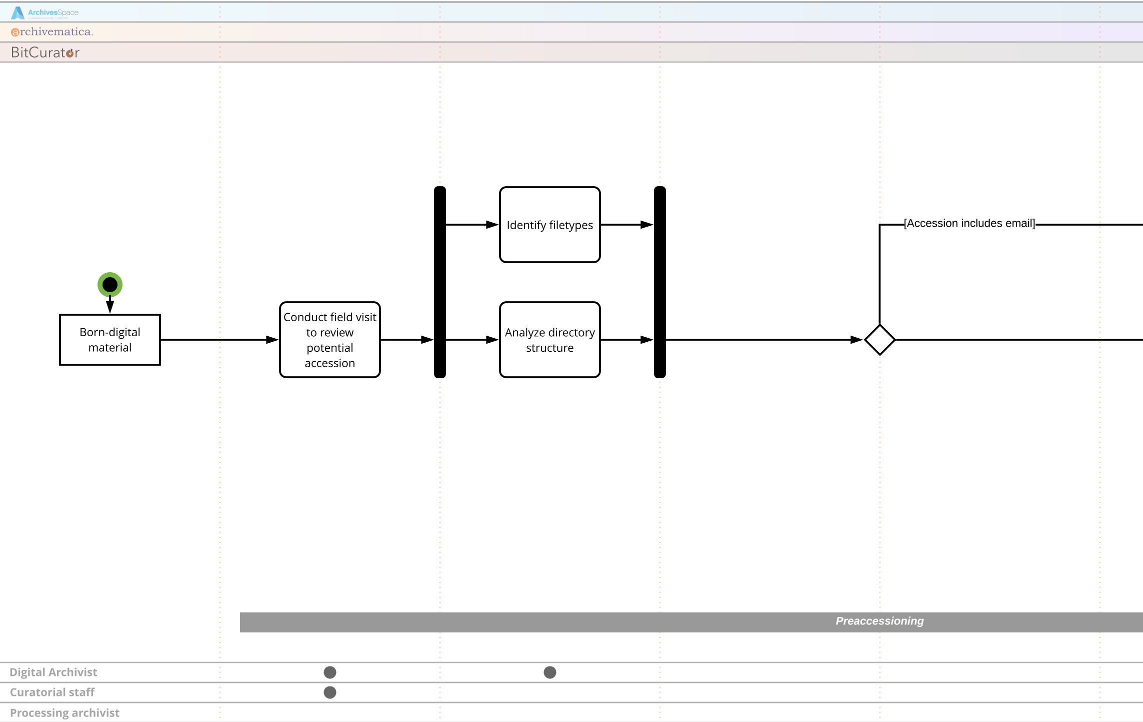
Pain point

As-Is Digital Curation Workflow, August 2017

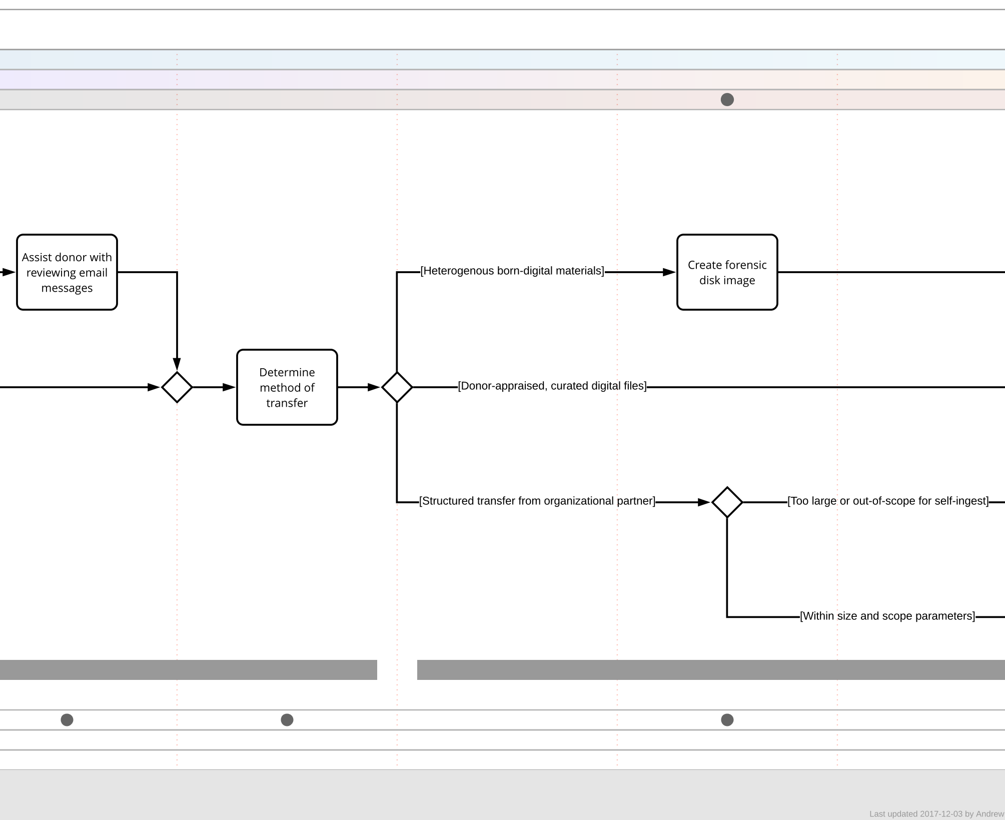
New York University

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica.
BitCurator
Born-digital material
Conduct field visit to review potential accession
Identify filetypes
Analyze directory structure
[Accession includes email]
Preaccessioning
Digital Archivist
Curatorial staff
Processing archivist

- Primary / preferred / required
- Secondary / optional
- Concurrency
- Start / end point
- Split / Merge
- Action
- Connection
- Object / Data
- Delay
- Pain point


Assist donor with reviewing email messages
Determine method of transfer
[Heterogenous born-digital materials]
Create forensic disk image
[Donor-appraised, curated digital files]
[Structured transfer from organizational partner]
[Too large or out-of-scope for self-ingest]
[Within size and scope parameters]

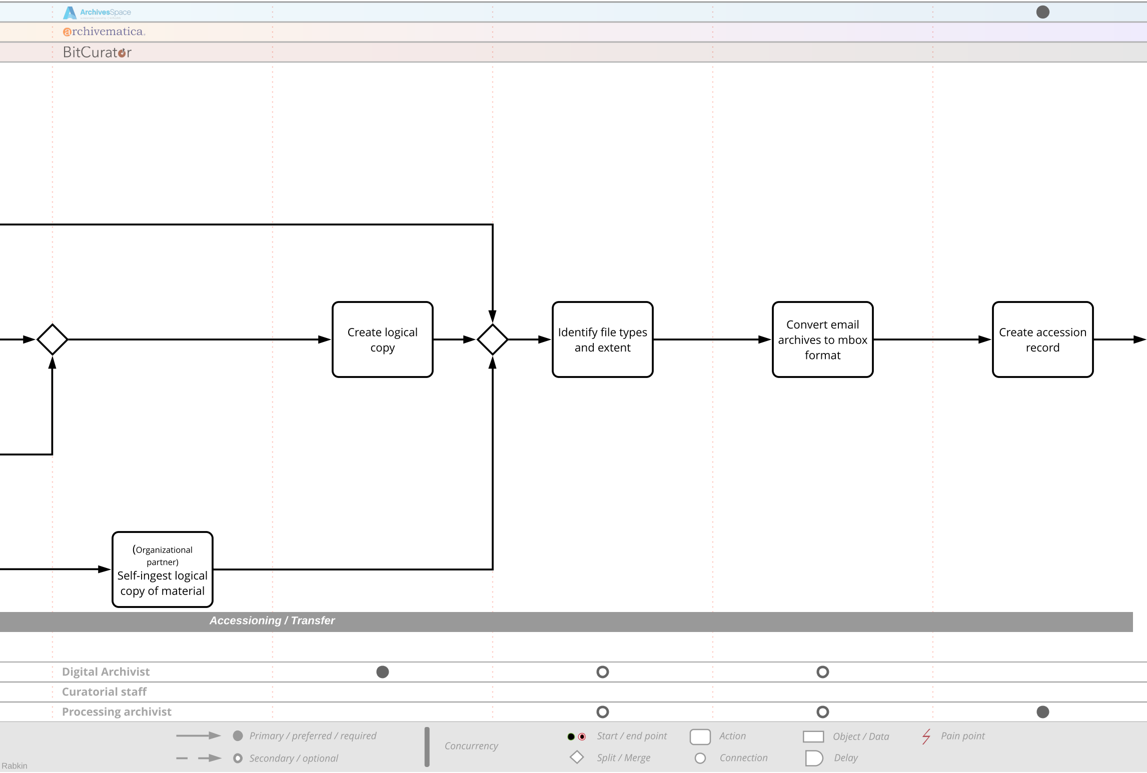
Last updated 2017-12-03 by Andrew

As-Is Digital Curation Workflow, August 2017

New York University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica.

BitCurator

Create logical copy

Identify file types and extent

Convert email archives to mbox format

Create accession record

(Organizational partner)
Self-ingest logical copy of material

Accessioning / Transfer

Digital Archivist

Curatorial staff

Processing archivist

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

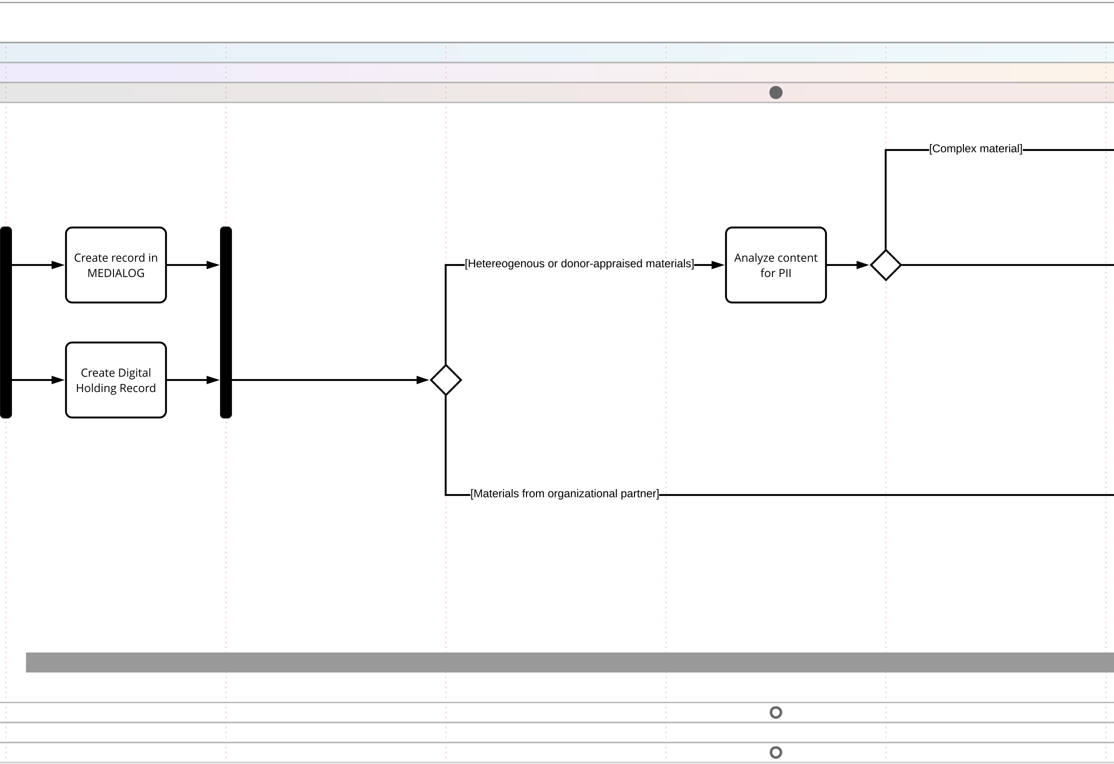
Connection

Object / Data

Delay

Pain point

Rabkin


Create record in MEDIALOG
Create Digital Holding Record
[Hetereogenous or donor-appraised materials]
Analyze content for PII
[Complex material]
[Materials from organizational partner]

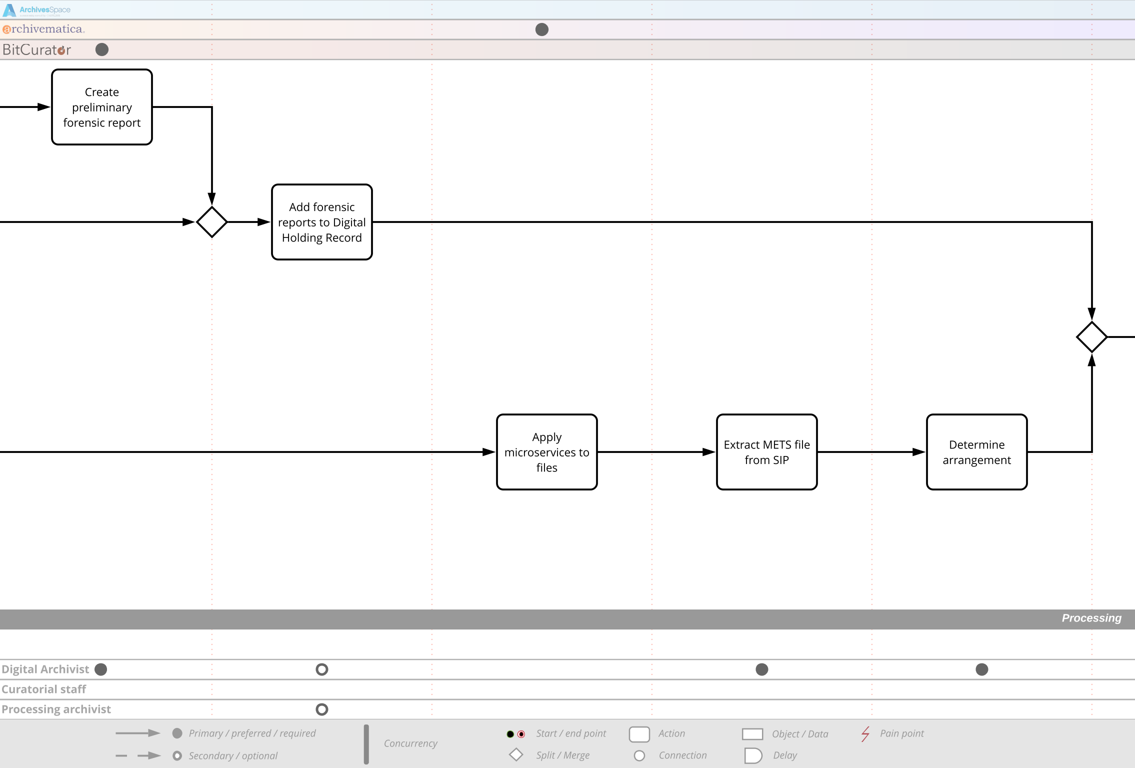
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

New York University

OssArcFlow Project

https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

Create preliminary forensic report

Add forensic reports to Digital Holding Record

Apply microservices to files

Extract METS file from SIP

Determine arrangement

Processing

Digital Archivist

Curatorial staff

Processing archivist

● *Primary / preferred / required*

○ *Secondary / optional*

Concurrency

Start / end point

Split / Merge

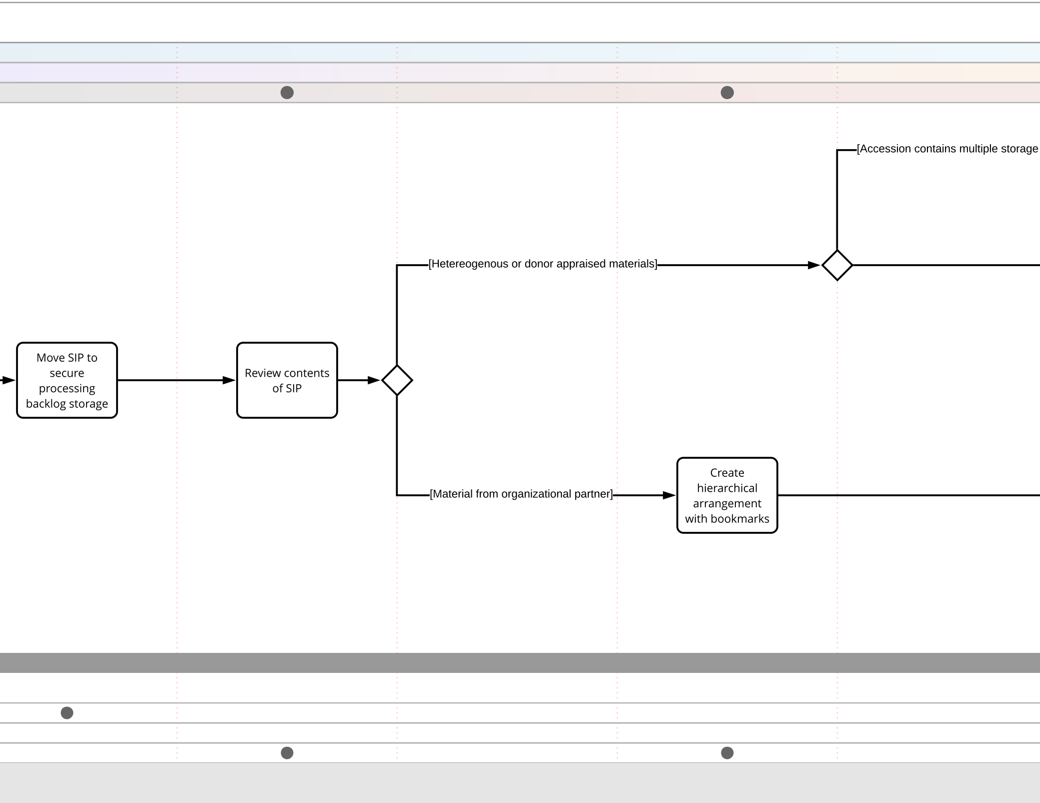
Action

Connection

Object / Data

Delay

Pain point


Move SIP to secure processing backlog storage
Review contents of SIP
[Hetereogenous or donor appraised materials]
[Accession contains multiple storage
[Material from organizational partner]
Create hierarchical arrangement with bookmarks

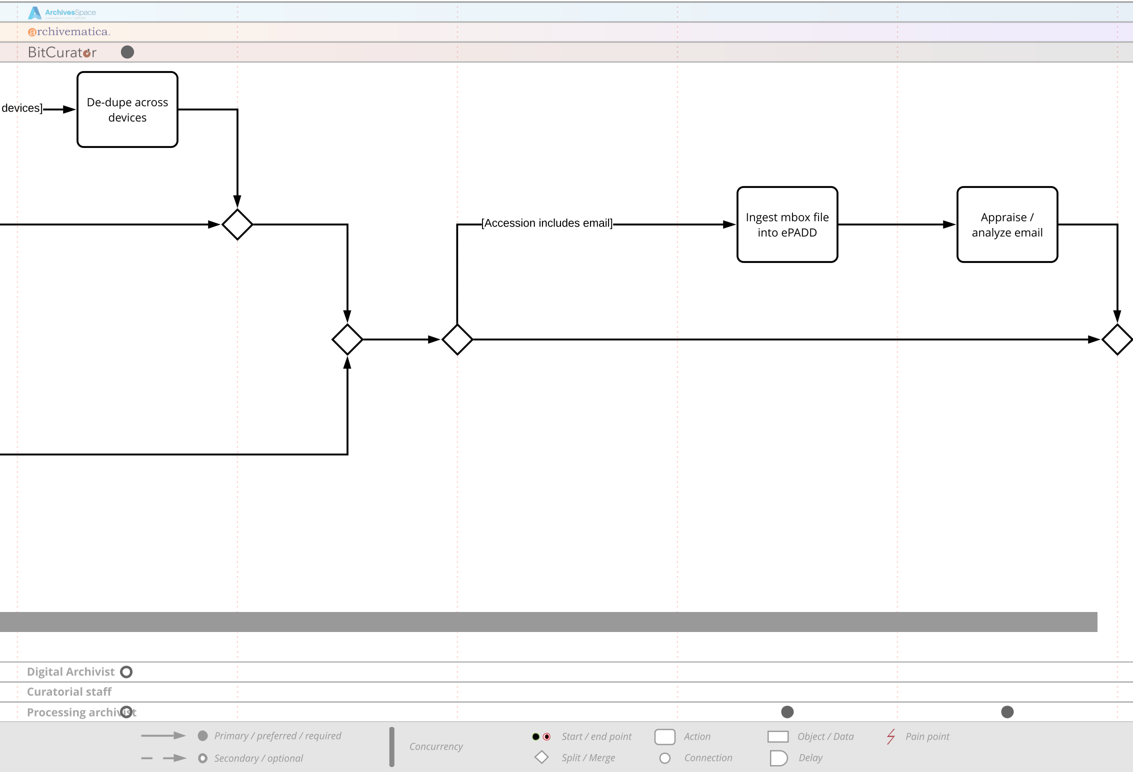
Last updated 2017-12-03 by Andrew Rabkin

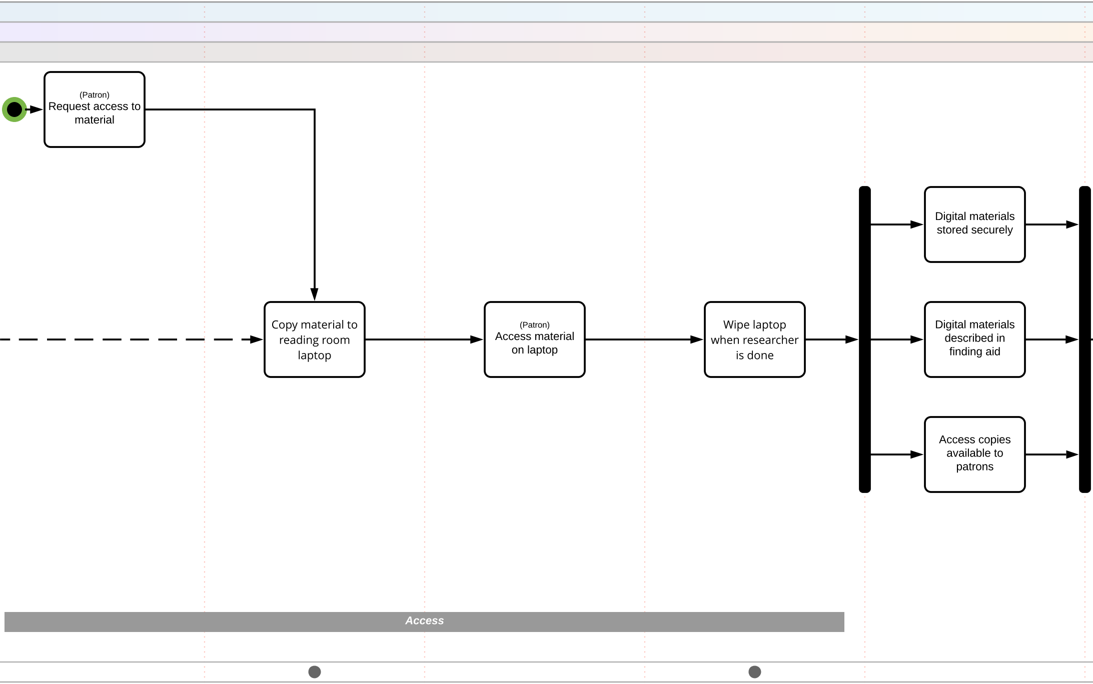
As-Is Digital Curation Workflow, August 2017

New York University

OssArcFlow Project

https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica
BitCurator
devices]
De-dupe across devices
[Accession includes email]
Ingest mbox file into ePADD
Appraise / analyze email
Digital Archivist
Curatorial staff
Processing archivist
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


(Patron)
Request access to material
Copy material to reading room laptop
(Patron)
Access material on laptop
Wipe laptop when researcher is done
Digital materials stored securely
Digital materials described in finding aid
Access copies available to patrons
Access

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

New York University

OssArcFlow Project

https://educopia.org/research/ossarcflow

ArchivesSpace
archivematica
BitCurator
Digital Archivist
Curatorial staff
Processing archivist

→ ● Primary / preferred / required

– → ○ Secondary / optional

| Concurrency

● ◉ Start / end point

◇ Split / Merge

▢ Action

○ Connection

▭ Object / Data

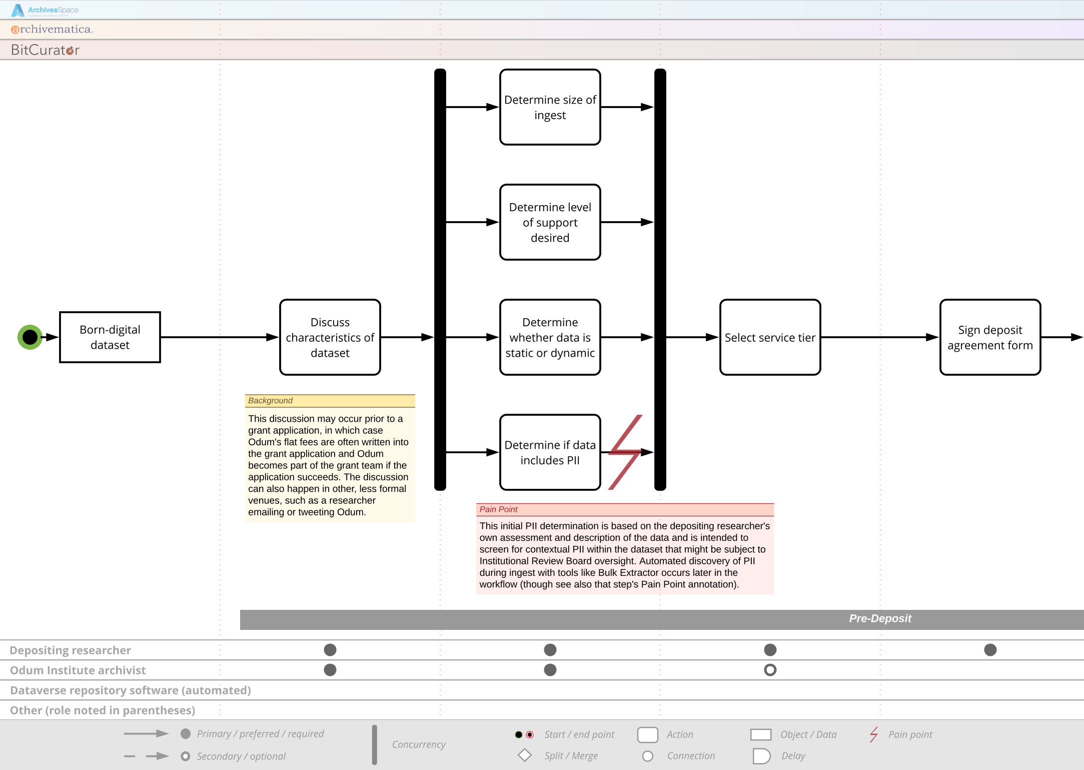
◗ Delay

ϟ Pain point

As-Is Digital Curation Workflow, August 2017

The Odum Institute
University of North Carolina at Chapel Hill

OssArcFlow Project
https://educopia.org/research/ossarcflow



ArchivesSpace
archivematica
BitCurator
Born-digital dataset
Discuss characteristics of dataset
Determine size of ingest
Determine level of support desired
Determine whether data is static or dynamic
Determine if data includes PII
Select service tier
Sign deposit agreement form
Pre-Deposit

Background

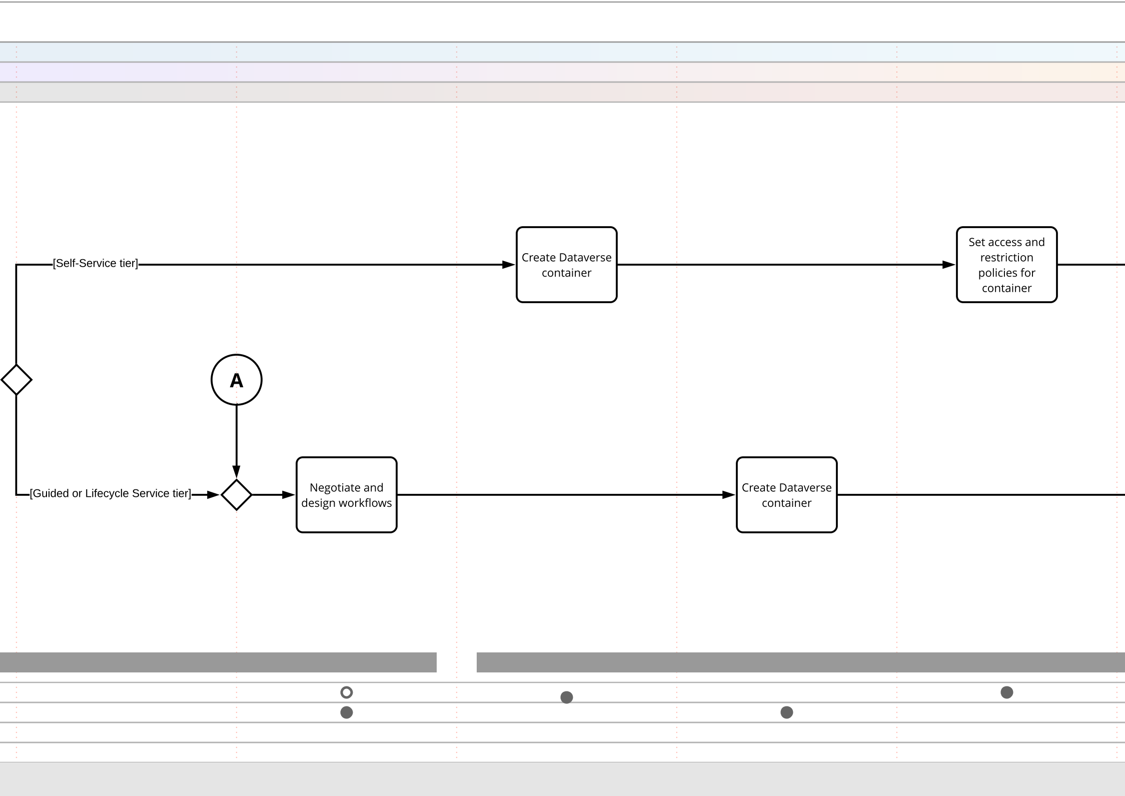
This discussion may occur prior to a grant application, in which case Odum's flat fees are often written into the grant application and Odum becomes part of the grant team if the application succeeds. The discussion can also happen in other, less formal venues, such as a researcher emailing or tweeting Odum.

Pain Point

This initial PII determination is based on the depositing researcher's own assessment and description of the data and is intended to screen for contextual PII within the dataset that might be subject to Institutional Review Board oversight. Automated discovery of PII during ingest with tools like Bulk Extractor occurs later in the workflow (though see also that step's Pain Point annotation).

Depositing researcher		●	●	●	●
Odum Institute archivist		●	●	○	
Dataverse repository software (automated)					
Other (role noted in parentheses)					

● *Primary / preferred / required*
○ *Secondary / optional*
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point

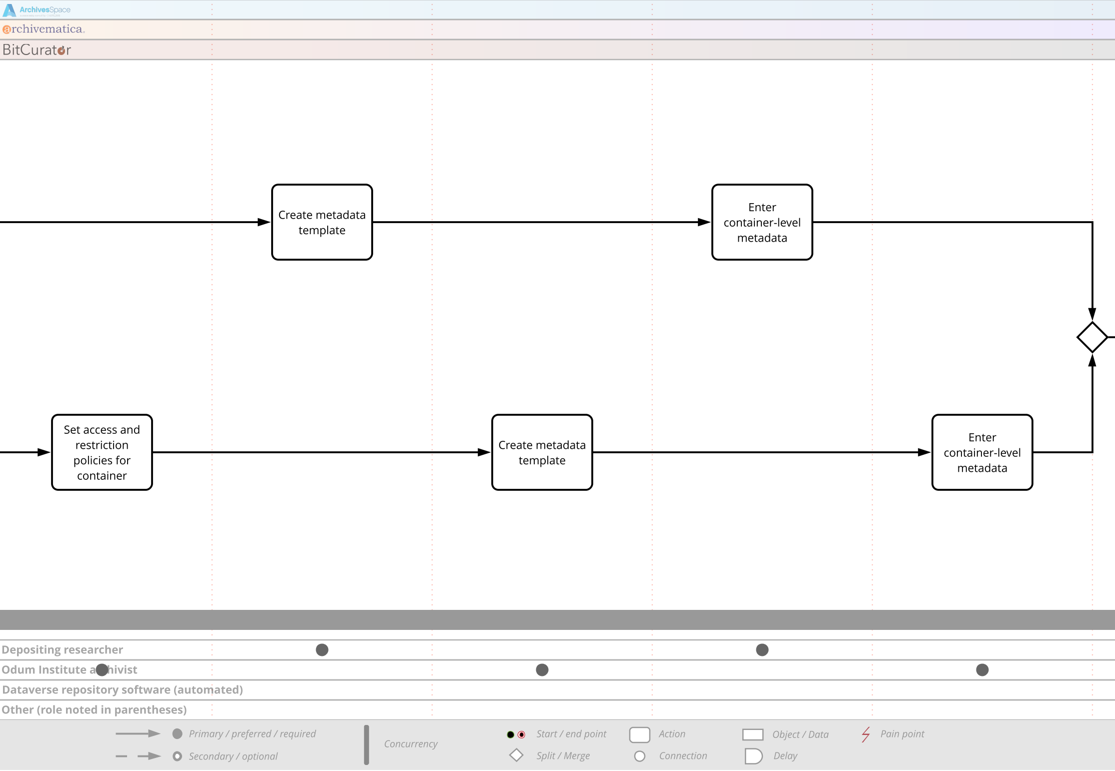

[Self-Service tier]
Create Dataverse container
Set access and restriction policies for container
A
[Guided or Lifecycle Service tier]
Negotiate and design workflows
Create Dataverse container

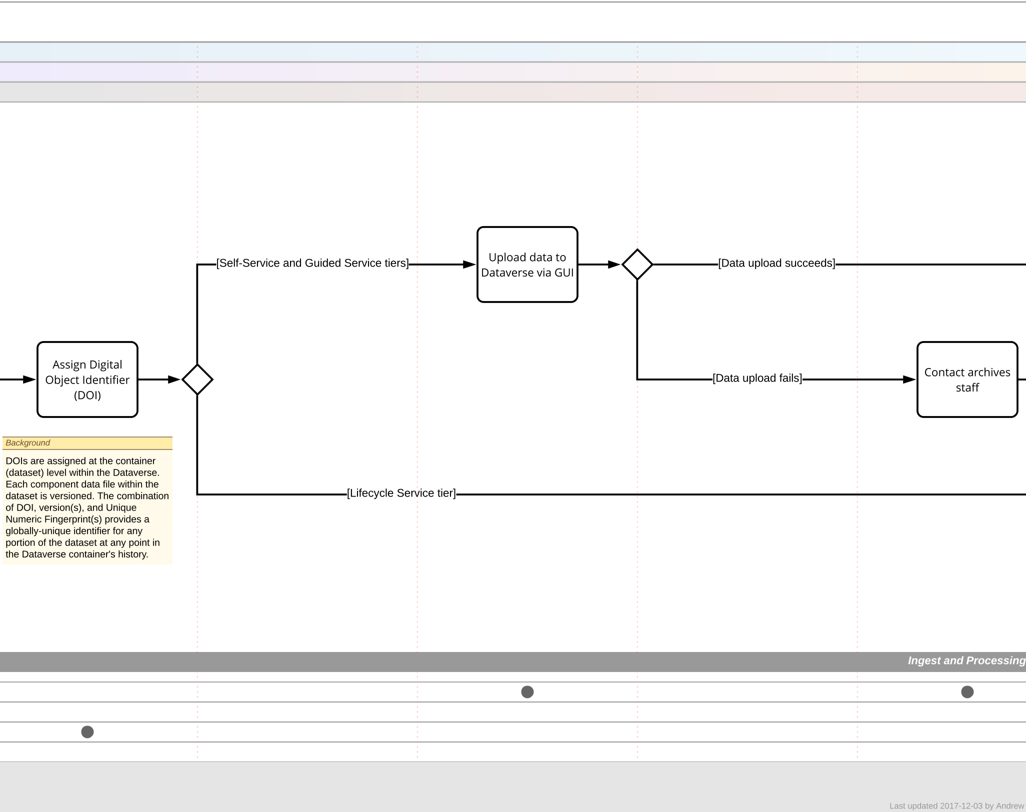
Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

The Odum Institute
University of North Carolina at Chapel Hill

OssArcFlow Project
https://educopia.org/research/ossarcflow


ArchivesSpace
archivematica.
BitCurator
Create metadata template
Enter container-level metadata
Set access and restriction policies for container
Create metadata template
Enter container-level metadata
Depositing researcher
Odum Institute archivist
Dataverse repository software (automated)
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point


Assign Digital Object Identifier (DOI)
[Self-Service and Guided Service tiers]
Upload data to Dataverse via GUI
[Data upload succeeds]
[Data upload fails]
Contact archives staff
[Lifecycle Service tier]
Ingest and Processing

Background

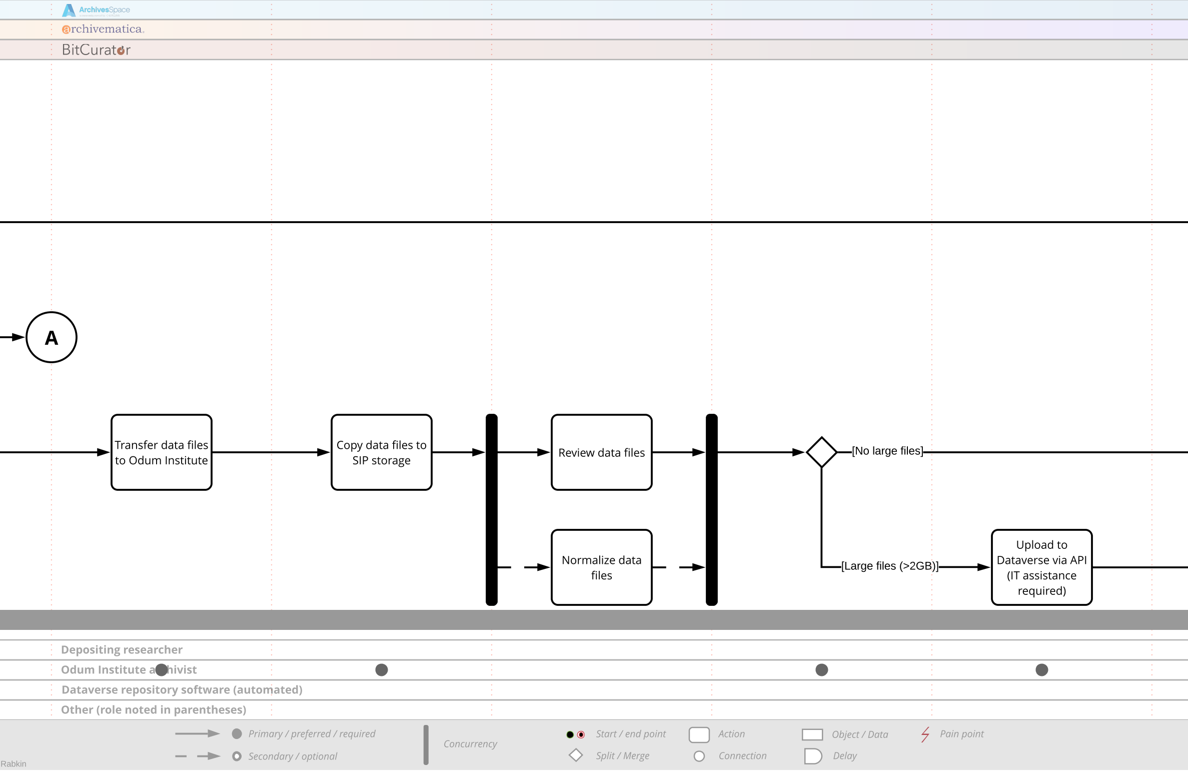
DOIs are assigned at the container (dataset) level within the Dataverse. Each component data file within the dataset is versioned. The combination of DOI, version(s), and Unique Numeric Fingerprint(s) provides a globally-unique identifier for any portion of the dataset at any point in the Dataverse container's history.

Last updated 2017-12-03 by Andrew

As-Is Digital Curation Workflow, August 2017

The Odum Institute
University of North Carolina at Chapel Hill

OssArcFlow Project
https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

A

Transfer data files to Odum Institute

Copy data files to SIP storage

Review data files

Normalize data files

[No large files]

[Large files (>2GB)]

Upload to Dataverse via API (IT assistance required)

Depositing researcher

Odum Institute archivist

Dataverse repository software (automated)

Other (role noted in parentheses)

Primary / preferred / required

Secondary / optional

Concurrency

Start / end point

Split / Merge

Action

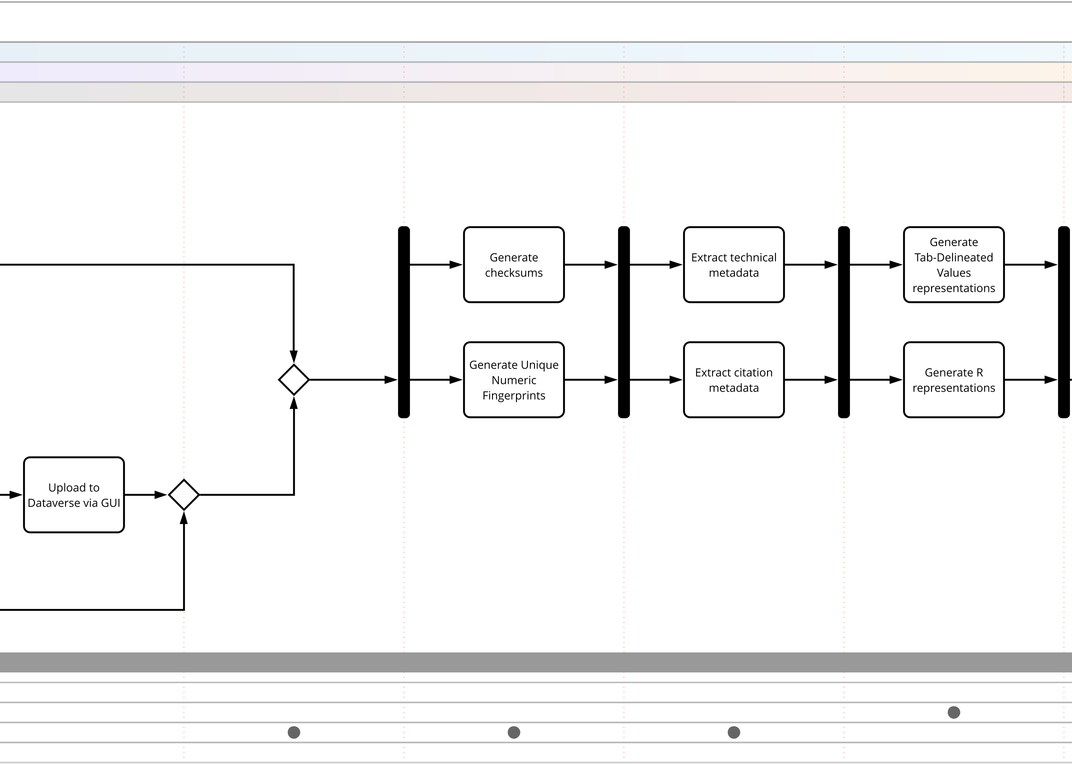
Connection

Object / Data

Delay

Pain point

Rabkin

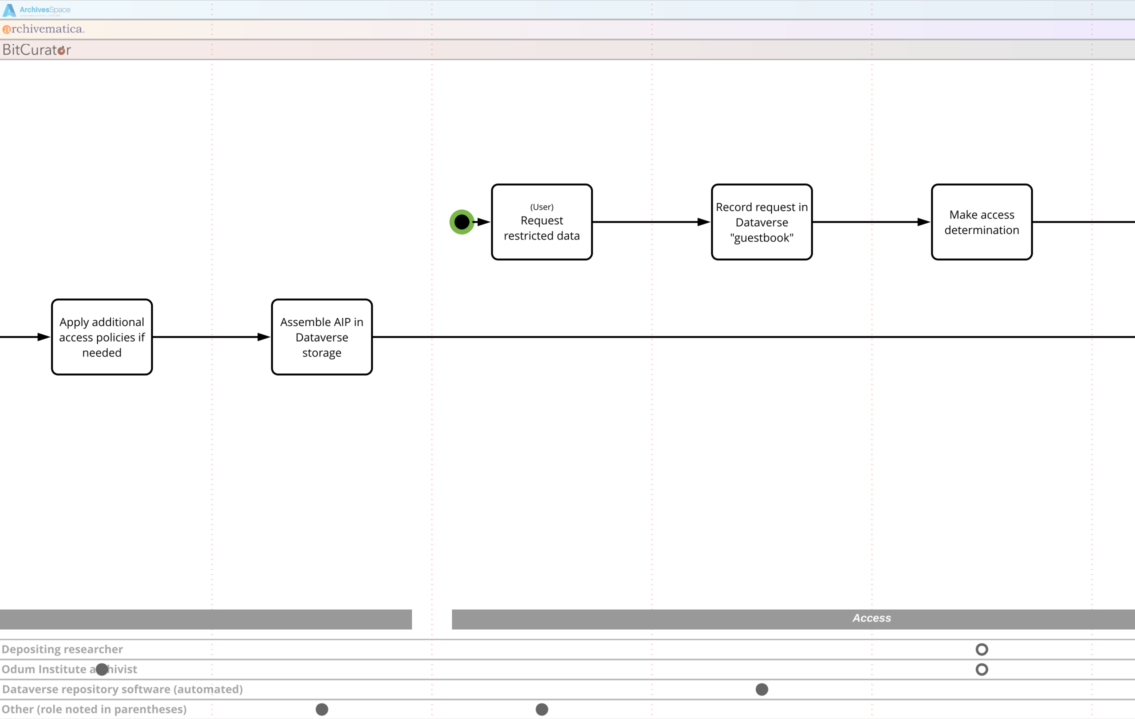

Generate checksums
Extract technical metadata
Generate Tab-Delineated Values representations
Generate Unique Numeric Fingerprints
Extract citation metadata
Generate R representations
Upload to Dataverse via GUI

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

The Odum Institute
University of North Carolina at Chapel Hill

OssArcFlow Project
https://educopia.org/research/ossarcflow



ArchivesSpace

archivematica

BitCurator

(User) Request restricted data

Record request in Dataverse "guestbook"

Make access determination

Apply additional access policies if needed

Assemble AIP in Dataverse storage

Access

Depositing researcher

Odum Institute archivist

Dataverse repository software (automated)

Other (role noted in parentheses)

→ ● *Primary / preferred / required*

⇢ ○ *Secondary / optional*

Concurrency

● ◉ *Start / end point*

◇ *Split / Merge*

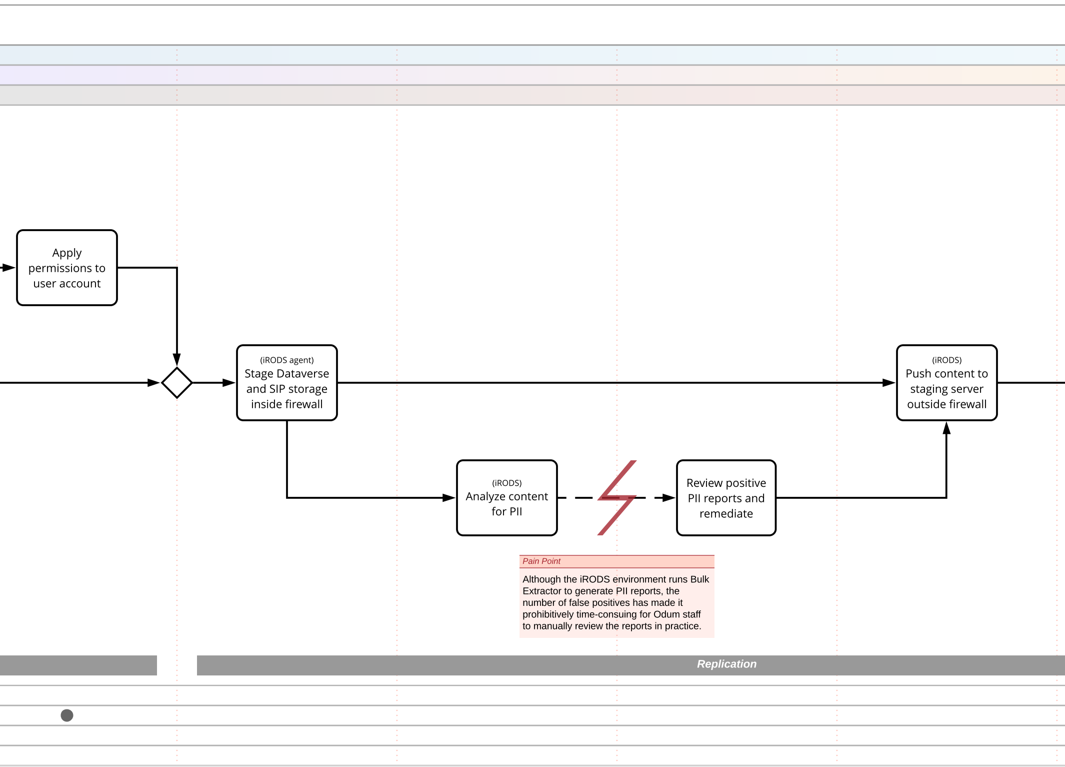
▢ *Action*

○ *Connection*

▭ *Object / Data*

◗ *Delay*

ϟ *Pain point*


Apply permissions to user account
(iRODS agent)
Stage Dataverse and SIP storage inside firewall
(iRODS)
Push content to staging server outside firewall
(iRODS)
Analyze content for PII
Review positive PII reports and remediate
Replication

Pain Point

Although the iRODS environment runs Bulk Extractor to generate PII reports, the number of false positives has made it prohibitively time-consuing for Odum staff to manually review the reports in practice.

Last updated 2017-12-03 by Andrew Rabkin

As-Is Digital Curation Workflow, August 2017

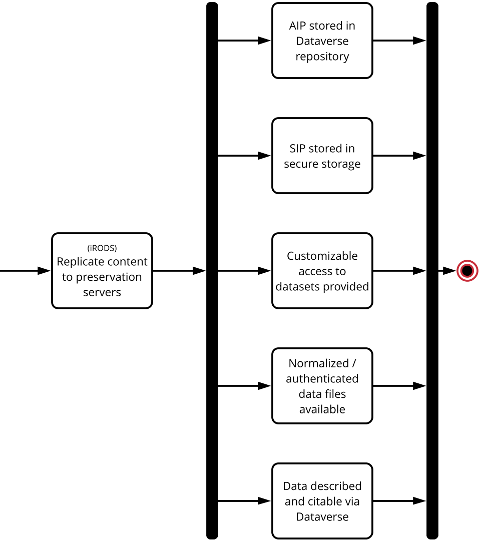
The Odum Institute
University of North Carolina at Chapel Hill

OssArcFlow Project
https://educopia.org/research/ossarcflow

ArchivesSpace

archivematica

BitCurator



(iRODS)
Replicate content to preservation servers

AIP stored in Dataverse repository

SIP stored in secure storage

Customizable access to datasets provided

Normalized / authenticated data files available

Data described and citable via Dataverse

Depositing researcher

Odum Institute archivist

Dataverse repository software (automated)

Other (role noted in parentheses)

- ⟶ ● *Primary / preferred / required*
- – ⟶ ○ *Secondary / optional*
- ▍ *Concurrency*
- ● ◉ *Start / end point*
- ◇ *Split / Merge*
- ▢ *Action*
- ○ *Connection*
- ▭ *Object / Data*
- ◗ *Delay*
- ϟ *Pain point*

Last updated 2017-12-03 by Andrew Rabkin