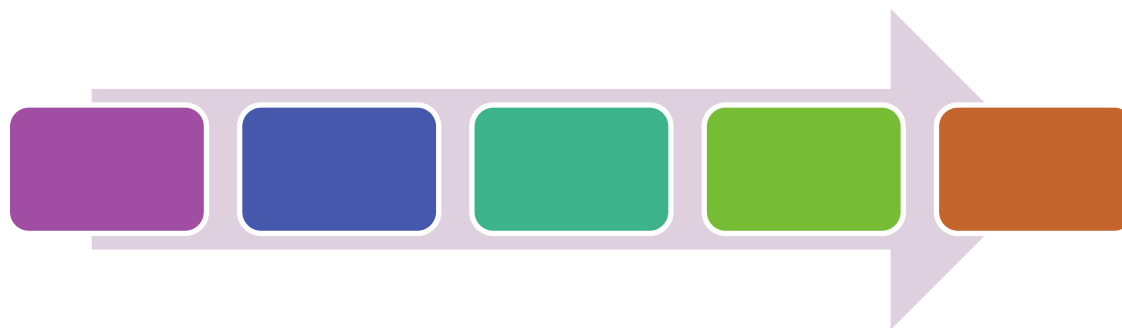


# Developing a Framework for File Format Migrations



iPRES 2015 Chapel Hill, NC 3 November 2015

Joey Heinen and Andrea Goethals



HARVARD  
LIBRARY

# Outline

1. Project Drivers
2. Migration Framework
3. Key Lessons Learned
4. Q&A

**Researcher** to **Librarian A**: I somehow can't open the file extension .smi (see attachment) working on a mac. Any ideas? I'm so sorry to bother you with this.

**Librarian A** to **Audio Engineer**: This is now a recurring problem. Is there another audio option for the website? Right now, the files on my computer seem to try to download into VLC perpetually, but never play. They used to open in Itunes. Any advice?

**IT Support**: Unfortunately it is not possible to play RealPlayer SMI playlist files on a Mac. The user should try playing the files on a Windows PC instead. The latest version of RealPlayer does not work with Real Audio files. The files will not play, and the application itself may crash on some Windows 7 machines. The user should download and install an older version of Real Audio Player

**Librarian A**: This is why we need a new audio platform... I can't even play the files anymore... ??

# Obsolete Formats in the DRS

- Audio
  - RealAudio files & SMIL Playlists (use copies)
- Images
  - Kodak PhotoCD images (archival and production masters)
- Obsolescence revealed through use

Adobe Photoshop CS3 Extended



Could not complete your request because it is not the right kind of document.

OK

# Ongoing Need

- Self-assessments revealed our inexperience with format migrations as a gap area in our repository practices
- Need a format migration *framework*
  - Cover whole workflow
  - Flexible structure
    - Format-agnostic
    - Scalable up or down
    - Plug-n-play roles and responsibilities

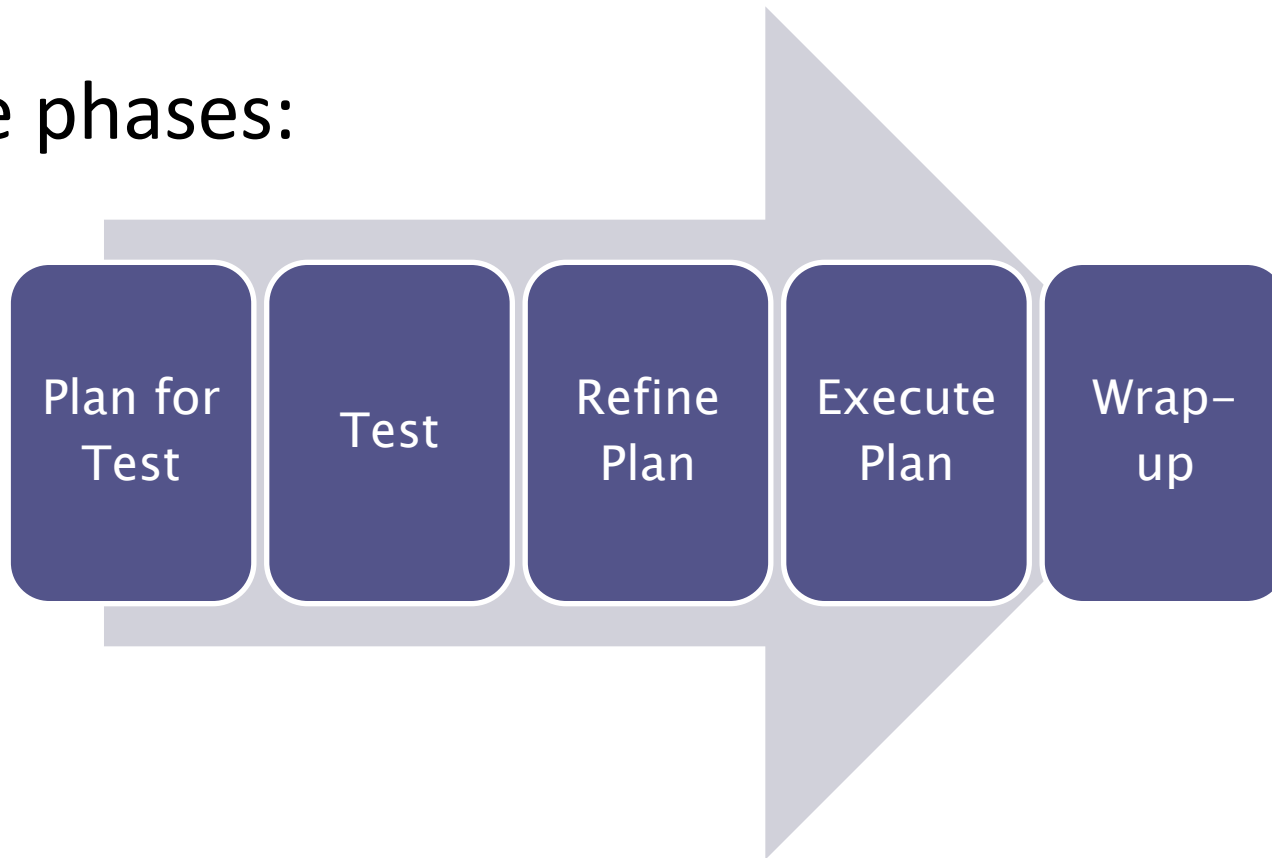
# Opportunity

- IMLS-funded NDSR Boston project
- 9-month resident to work on strategic projects designed by the host institution
- Uninterrupted time to focus on developing a format migration framework by working through 2 real use cases



# Format Migration Framework

- Five phases:

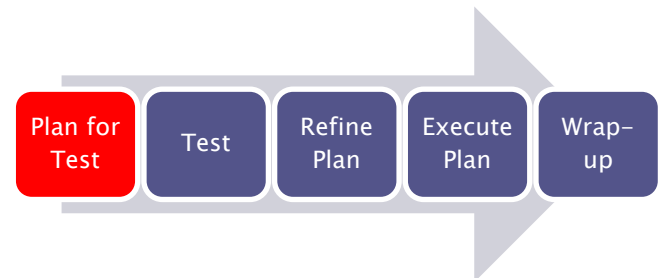


Each with activities and deliverables

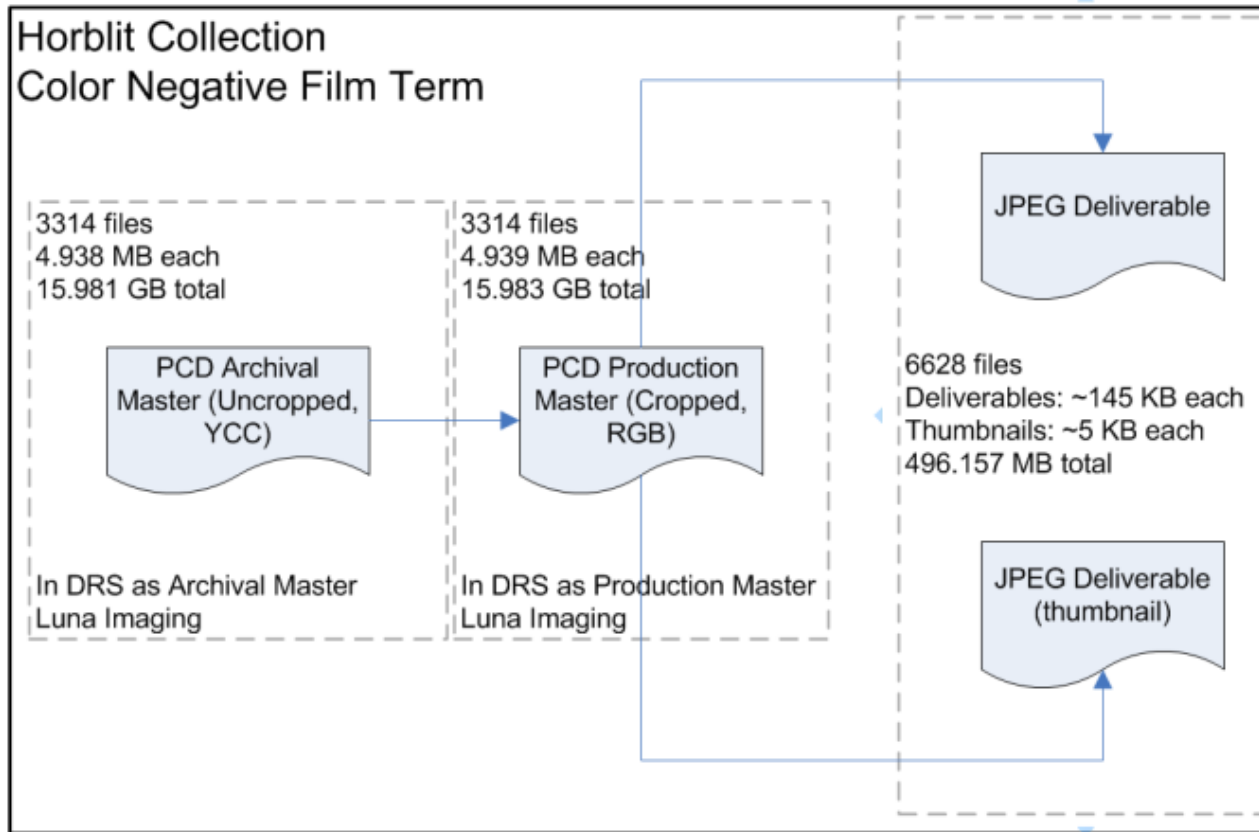


# Phase One: Plan for Test

- Define stakeholders and their migration roles
- Identify concurrent projects that can affect migration
- Acquire specifications
- Research formats and identify key properties
- Identify the content “buckets” that can be migrated together
- Analyze content, metadata, migration tools



# Deliverable Ex. for Kodak PCD: Content Groupings



Plan for  
Test

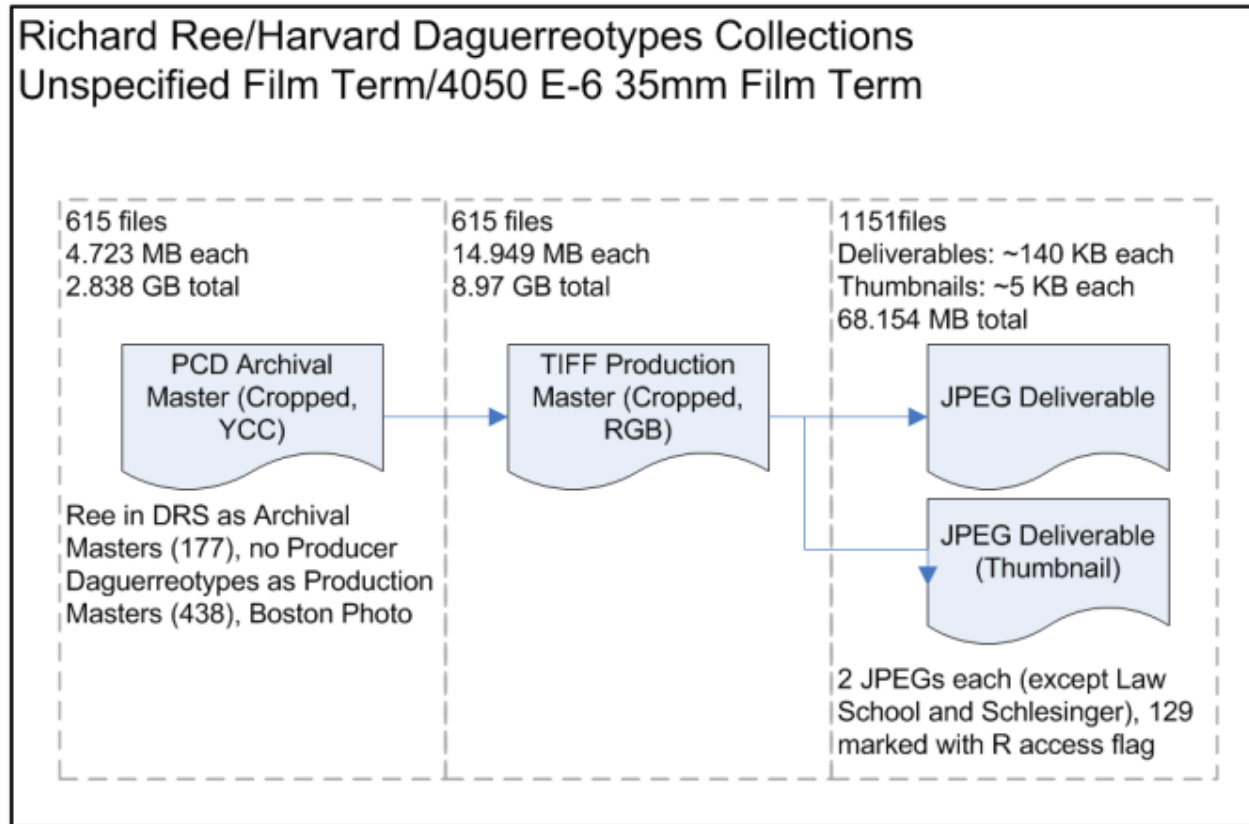
Test

Refine  
Plan

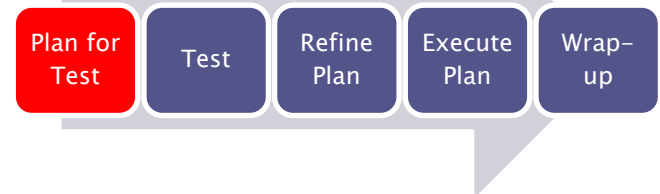
Execute  
Plan

Wrap-  
up

# Deliverable Ex. for Kodak PCD: Content Groupings

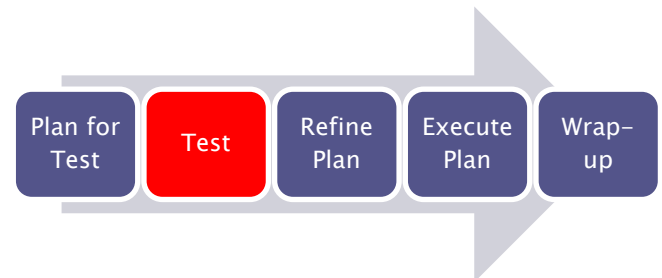


Different cropping, roles, and  
format of Production Master



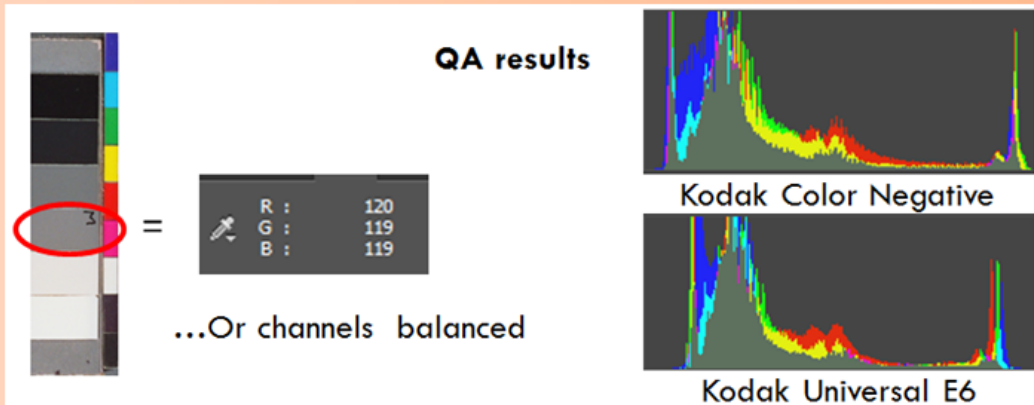
# Phase Two: Test

- Select sample content to test with
- Design migration tests to exercise tools and different processes
- Perform migration tests
- QA results using authoritative metrics
- Decide on the migration tools and process based on the test



# Ex. for Kodak PCD: Testing Metrics

Use Photoshop to check color balance (if image contains color bars).  
Compare gamut of histograms to find best range.



Noting subtle differences  
— though seemingly similar, different film terms produce slightly different results.



Plan for  
Test

Test

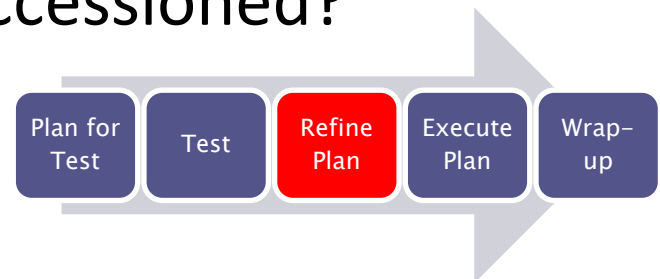
Refine  
Plan

Execute  
Plan

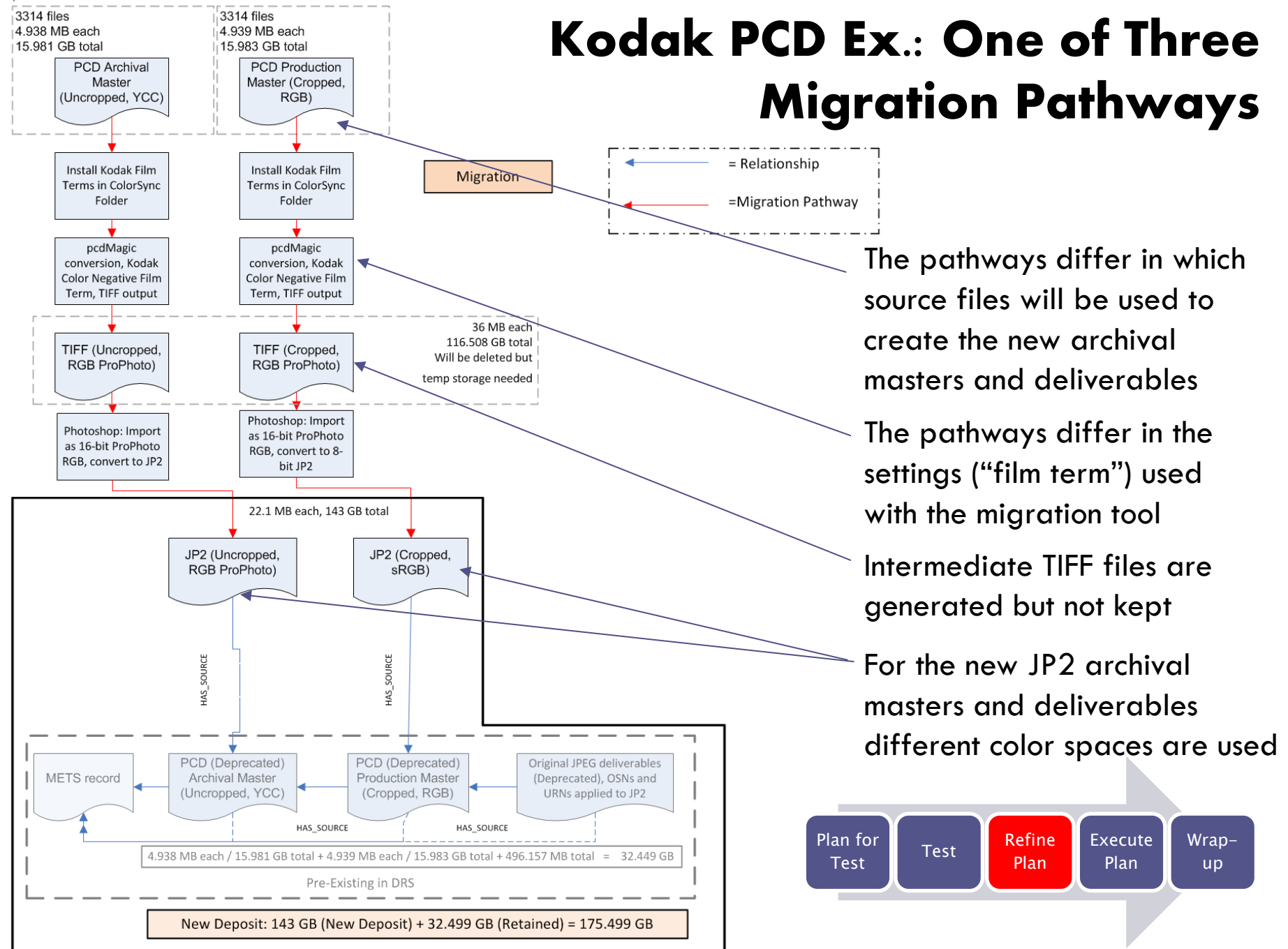
Wrap-  
up

# Phase Three: Refine Plan

- Design how content created by the migration will interact with existing content and the repository
  - How will it be ingested into the repository?
  - What metadata will be added about the content and migration?
  - What existing metadata will change?
  - What will be retained / de-accessioned?

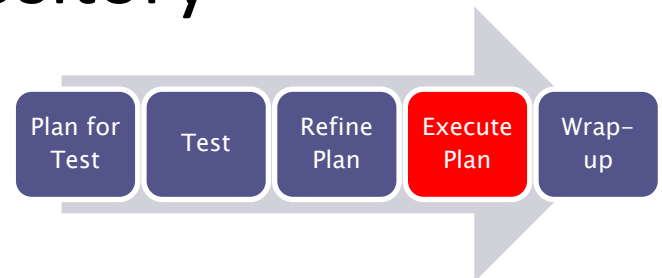


# Kodak PCD Ex.: One of Three Migration Pathways



# Phase Four: Execute Plan

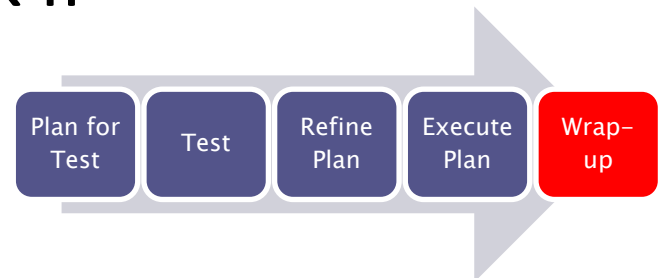
- Set up migration environment
- If necessary, custom development
  - create scripts to automate migration process
  - modify existing ingest tools
- Schedule migration
- Perform migration
- Ingest content into the repository





# Phase Five: Wrap-Up

- Verify and document results
- Schedule any clean-up needed, e.g.:
  - de-accessioning of replaced content
  - changes to identifiers for new use copies
- Decide on final disposition of migration documentation and “file” it (e.g. delete, save, preserve, expose)
- Adjust migration framework if needed



# Key Lessons Learned

- Users are the experts on obsolescence
  - Repository managers need to solicit use problems from end users, support staff, etc.
- Migrations are more complex/nuanced than commonly depicted in the literature
- This generic framework is useful in practice & can help us institutionalize format migrations

# Thank You

**Joey Heinen**

Digital Production Coordinator  
Northeastern University Libraries

j.heinen@neu.edu

@joey\_heinen

**Andrea Goethals**

Manager of Digital Preservation and Repository Services

Harvard Library

andrea\_goethals@harvard.edu