More Pixels, Less Process: Decision making for minimal processing digitization

Amanda Focke, Rice University
afocke@rice.edu
Trends

- Increasing numbers of archives are putting the Meissner-Greene Minimal Processing ("MP-LP") approach into practice for processing collections.
- Expectations that materials be made available on-line only grow.

What we think of as “proper digitization” processes are at odds with the MP-LP approach.

…But what is “proper”?
ACRL Top 10 Assumptions for the future of academic libraries

Top ranked assumption:

“There will be an increased emphasis on digitizing collections, preserving digital archives, and improving methods of data storage and retrieval. “

Assumption goes on to say:

- “Academic libraries have an opportunity to make their unique collections available to the world in unprecedented ways.

- In fact, the digitization of unique print collections may emerge as one of the primary missions of academic libraries in the 21st century.

- Librarians should collaborate with disciplinary colleagues in the curation of data as part of the research process.”
Part I:
Decision factors for selecting specific materials for digitization

Part II:
Choosing a digitization approach

Our main focus for today is really Part II, but your thought process for Part I can help you determine your repository’s digitization approach.
Part I:  
Decision factors for selecting specific materials for digitization

- Reference requests for materials, including faculty & general researchers
- Perceived research value
- Fragile originals not suited for access
- Content supports strategic initiatives (university goals for courses, outreach, etc.)
- Donor expectations
- Staff time / resources
## Prioritizing digitization projects

<table>
<thead>
<tr>
<th><strong>High Value</strong></th>
<th><strong>Low Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Very important</strong></td>
<td><strong>Not as important</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Deadline</strong></th>
<th><strong>High urgency</strong></th>
<th>1. Do it now!</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Deadline</strong></td>
<td><strong>Low urgency</strong></td>
<td>2. I really should…</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Gotta minute?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. I really shouldn't…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>High Value</strong></th>
<th><strong>Low Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active reference or faculty request</strong></td>
<td><strong>Fragile unique items needing reformatting, but having only medium or low research value</strong></td>
</tr>
<tr>
<td><strong>Unique materials with high research value</strong></td>
<td><strong>Donor or administrative request</strong></td>
</tr>
<tr>
<td><strong>Grant or donor expectation</strong></td>
<td><strong>It's a hidden gem!</strong></td>
</tr>
<tr>
<td><strong>No grant or donor expectations</strong></td>
<td><strong>Deadline</strong></td>
</tr>
<tr>
<td><strong>High urgency</strong></td>
<td><strong>Low urgency</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>High Value</strong></th>
<th><strong>Low Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Not as important</strong></td>
<td><strong>Very important</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>High Value</strong></th>
<th><strong>Low Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Not related to strategic planning</strong></td>
<td><strong>Not related to strategic planning</strong></td>
</tr>
<tr>
<td><strong>Not unique materials</strong></td>
<td><strong>Not unique materials</strong></td>
</tr>
<tr>
<td><strong>No donor expectation</strong></td>
<td><strong>No donor expectation</strong></td>
</tr>
<tr>
<td><strong>No active reference request</strong></td>
<td><strong>No active reference request</strong></td>
</tr>
</tbody>
</table>
You’ve prioritized digi-projects – what’s next?

- The decision factors by which you prioritized your digital projects should suggest the digitization approach you will use.
- Consider starting with minimal level digitization as your default plan, doing item level full process digitization only when the resources to do it can sincerely be justified.
For example

- Do you have a specific request for an item from a faculty member, needed quickly? (matrix category 1: “Do it now!”)

  Consider minimal processing digitization to get it out quickly.
Do you have high-research value materials with no specific urgency for access? Have the resources and administrative support to spend more time on it? (matrix category 2: “I really should…”)

Consider full process or minimal process.
Do you have what seem to be medium to low research value materials but a donor or admin request to digitize them?
(matrix category 3: “Gotta minute?”)

Find a way not to work on this material, but if you must digitize it, use minimal processing.
Do you have materials which seem to have little research value and there are typically no requests for them? (matrix category 4: “I really shouldn’t…”)

Don’t digitize them at all.
Part II: Choose digitization approach

- **Minimal processing digitization:** At a grouping level, such as folder level (or larger), with lower resolution digital files and less metadata
- **Item level full process digitization:** At item level with rich digital files and full metadata
- **Somewhere in between?**
Introducing these examples of approaches...

- There is no standard for the following examples of approaches.
- These approaches include many parts which could be altered to suit any situation.
- This “minimal process” approach example could be made even more minimal, this “full process” approach example could be developed even more.
- Mix and match parts to meet your needs, same as you would when formulating a collection processing plan.
Main parts to consider for any approach

- Scan quality
  - Quickly-scanned photocopies? High-res color scan?

- Amount of metadata
  - Just title/date/source – or full qualified Dublin Core, or MODS?

- Mode of delivery
  - Bundles of pdfs? High res tiffs?
Minimal digitization for *text* items at *folder* level

- Original pages photocopied
- Photocopies scanned with sheet feed scanner
- Page images made into PDF files reflecting natural groupings within folder
- PDF files posted on-line as links from finding aid and in Rice's institutional repository (IR)
- Basic folder level metadata includes dc.Title, dc.Date, dc.Source (dc = Dublin Core)

- example URL: http://scholarship.rice.edu/handle/1911/21674
Rice’s IR showing text items at folder level (a minimal digitization approach)

Notice multiple pdfs associated with one folder level record, little folder level metadata.
Finding aid link to IR for text items at folder level
(a minimal digitization approach)
Minimal digitization for *photo* items at *folder* level

- Photos scanned at 300 dpi greyscale
- Image files posted on-line in Rice's IR under basic folder level metadata and linked from finding aid
- Basic folder level metadata includes dc.Title, dc.Date, dc.Source

- example URL: http://scholarship.rice.edu/handle/1911/21675
IR showing *photos at folder level* (a minimal digitization approach)

- The full item record view also shows:
  - “Source” field with citation info
  - URL for the finding aid.
Finding aid link to IR for photo items at folder level
What can minimal processing digitization do for you?

- Higher digitization statistics?
- More hits on your website?
- Stars in your archives crown?
- Gets more material out faster

- Can be incorporated into reference work – anything you get a photocopy request for, you can then also digitize in this manner and post on-line (pending any © issues)

- Involves less cherry-picking of materials by archivists
And now to look at the other end of the digitization spectrum…
Full process digitization for *handwritten text at item level*

- Item level Dublin Core full metadata
- Transcription of handwritten text
- TEI XML mark-up of transcription
- Full color 600 dpi TIFF master on high-end scan-back machine at lab
- Posted in our IR and linked from EAD finding aid

example URL:
http://hdl.handle.net/1911/21658
IR showing *handwritten text at item level* (a full process digitization approach)

- Full item record shows all 17 qualified Dublin Core metadata fields
Full process for *handwritten text at item level* cont.:

Transcribed text, marked up in TEI/XML, shown in HTML

In web presentation, this thumbnail page image opens to full screen

Transcript encoded in TEI in an XML editor
Full process digitization for *photographs* at *item* level

- Item level Dublin Core full metadata
- Full color 600 dpi TIFF master on high-end scanner at lab, showing front and back of image
- Posted in our IR and linked from EAD finding aid

- example URL: http://scholarship.rice.edu/handle/1911/9236
IR showing a *photograph* record at *item* level
(a full process digitization approach)

- The full item record shows 17 qualified Dublin Core metadata fields
- Notice front and back of image are presented
Pros and cons of item level full process digitization

- **Pros**
  - Makes item very discoverable
  - Benefits research community, particularly in example of handwritten text becoming full-text searchable

- **Cons**
  - Extremely resource intensive
  - Lots of time spent on one object means no time for many, many others – so the selection of individual items must be well done
Conclusions

- The user communities expect that archives will digitize and share as much of their collections as they can.

- Looking realistically at resources needed for item level full process digitization it becomes clear that this cannot be our only approach.
Conclusions

- Just as there is a wide spectrum of processing approaches, there is a wide spectrum of digitization approaches.
- You develop your digitization approach by choosing
  - scan quality
  - amount of metadata
  - mode of delivery.
Decide what your repository’s variety of digitization approaches may be, but consider making minimal processing digitization your default approach.
Thank you!

Amanda Focke
Rice University
afocke@rice.edu