



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. “

<http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/april07/tenassumptions.cfm>



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

	High Value Very important	Low Value Not as important
Deadline High urgency	1. Do it now! <ul style="list-style-type: none">• Active reference or faculty request• Unique materials with high research value• Grant or donor expectation• It's a hidden gem!	3. Gotta minute? <ul style="list-style-type: none">• Fragile unique items needing reformatting, but having only medium or low research value• Donor or administrative request
No Deadline Low urgency	2. I really should... <ul style="list-style-type: none">• Unique materials with semi-high research value• No grant or donor expectations	4. I really shouldn't... <ul style="list-style-type: none">• Not related to strategic planning• Not unique materials• No donor expectation• No active reference request



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.



Examples cont.

- 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.



Examples cont.

- 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.



Examples cont.

- 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...”)

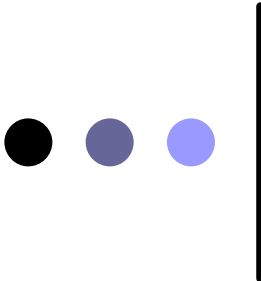
- Don’t digitize them at all.



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)

The screenshot shows the Rice Digital Scholarship Archive interface. The main heading is "Moody family genealogical correspondence, 1946-1967". Below this, there is a "Show full item record" link and a metadata table.

Title:	Moody family genealogical correspondence, 1946-1967
Author:	Snow, Helen Foster
Abstract:	Correspondence between Helen Foster Snow, a relative of the Moody family, and Jean Halder Walker two genealogists that were researching the Moody family ancestry and property during the 1960's.
Description:	
URI:	http://hdl.handle.net/1911/21674
Date:	1947

Below the metadata, there is a section titled "Files in this item" which contains a table of 9 PDF files:

Files	Size	Format	View
ms181_1.pdf	4.681Mb	application/pdf	View/Open
ms181_10.pdf	3.092Mb	application/pdf	View/Open
ms181_2.pdf	20.64Mb	application/pdf	View/Open
ms181_3.pdf	8.389Mb	application/pdf	View/Open
ms181_4.pdf	8.451Mb	application/pdf	View/Open
ms181_5.pdf	2.107Mb	application/pdf	View/Open
ms181_6.pdf	2.487Mb	application/pdf	View/Open
ms181_7.pdf	2.921Mb	application/pdf	View/Open
ms181_8.pdf	8.528Mb	application/pdf	View/Open
ms181_9.pdf	1.428Mb	application/pdf	View/Open

Below the file list, there is a section titled "The following license files are associated with this item:" with a link to "Original License".

A blue callout box on the right side of the screenshot contains the text: "Notice multiple pdfs associated with one folder level record, little folder level metadata." An arrow points from this box to the "View/Open" links in the file list table.

Finding aid link to IR for *text* items at *folder* level (a minimal digitization approach)

Moody, John Wyatt, 1946-1967. Papers. (MS 181) — Fondren Library - Mozilla Firefox

http://library.rice.edu/collections/WRC/finding-aids/manuscripts/0181?portal_status_message=Changes%20saved.

Restrictions on Use

Permission to publish from the John Wyatt Moody Biographical Papers must be obtained from the Woodson Research Center, Fondren Library, Rice University.

[Return to the Table of Contents](#)

Index Terms

Subjects (Persons)
Houston, Sam, 1793-1863
Moody, John Wyatt
Snow, Helen Foster

Formats
Correspondence
Notes

[Return to the Table of Contents](#)

Administrative Information

Preferred Citation
John Wyatt Moody Biographical Papers, 1946-1967, MS 181, Woodson Research Center, Fondren Library, Rice University

Acquisition Information
The collection was a gift of Helen Foster Snow, 1968.

[Return to the Table of Contents](#)

Detailed Description of the Collection

Documents

Folder

1

Group A: Transcribed and annotated versions of correspondence between John Wyatt Moody and Sam Houston, along with various other revolutionary era figures.
Group B, 1946-1967: Correspondence between Helen Foster Snow, a relative of the Moody family, and Jean Halden Walker two genealogists that were researching the Moody family ancestry and property during the 1960's. [View this folder on-line.](#)

[Return to the Table of Contents](#)

Fondren Library Rice University
6100 Main Street
Houston, Texas 77005

Website Problems, Questions, Comments?
Copyright © 2008

Done

start

Moody, John Wyatt, ...

Microsoft PowerPoint ...

4:27 PM



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.

Rice Scholarship Home > Woodson Research Center > Texas City disaster records (MS 529) > View Item

Texas City disaster area, pre-explosion photographs

[Show full item record](#)

Title:	Texas City disaster area, pre-explosion photographs
Author:	Unknown author
Abstract:	Photographs of the Texas City, Texas waterfront which was later devastated by the April 16, 1947 disaster which occurred when a French ship carrying fertilizer on Galveston Bay caught fire and exploded. The nearby Monsanto styrene plant and other petroleum refineries, homes, warehouses, ships and buildings were largely destroyed. At least 576 were killed and 4,000 injured, with the town being almost completely destroyed.
Description:	Photographs taken from the air. Photographer unknown.
URI:	http://hdl.handle.net/1911/21675
Date:	1947

Files in this item

Files	Size	Format	View
pre_1.jpg	2.209Mb	image/jpeg	View/Open
pre_2.jpg	2.187Mb	image/jpeg	View/Open
pre_3.jpg	2.216Mb	image/jpeg	View/Open
pre_4.jpg	1.978Mb	image/jpeg	View/Open
pre_5.jpg	2.245Mb	image/jpeg	View/Open
pre_6.jpg	2.009Mb	image/jpeg	View/Open
			Need help?

The following license files are associated with this item:

- [Original License](#)

This items appears in the following Collection(s)

- [Texas City disaster records \(MS 529\)](#)

[Show full item record](#)

<http://scholarship.rice.edu/handle/1911/21675/show-full>

Finding aid link to IR for *photo* items at *folder* level

Original material or photo by William Lane, and 1947 Monsanto newsletters; 1997 correspondence relates to anniversary events.

Series VI: Anniversary events, 1997

1	13	Correspondence, flyers, brochures and ephemera regarding 1997 anniversary events in Texas City
2	-	Newspaper coverage of 1997 anniversary

Series VII: Maps and photographs

1	14	Site maps, 1947 Includes a site map annotated with numbers relating to photographs and an index to those photographs, but it is not clear if the largely unassociated photographs in this collection relate to this index.
	15	Images, pre-explosion (2 black and white 8x10) See folder on-line.
	16	Images, post-explosion (6 black and white 8x10s, 6 black and white postcards) See selection of items from this folder on-line.
2	1	Images, oversize (5 black and white 11x14s)

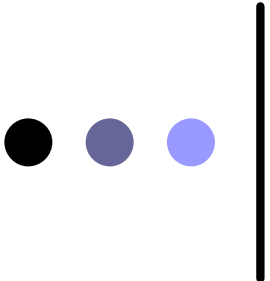
Series VIII: Video, 1957 & 1997 notes

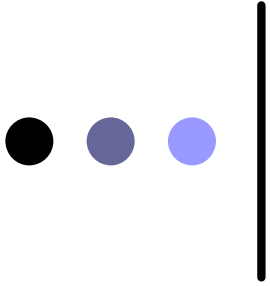
1	17	Manuscript and typescript notes regarding "Texas City Film" and "Introduction to Safety Movie", 1952 with 1997 notes
2	2	VHS videotape "Texas City - 1997 Reviews" 20 minutes. Includes "Eyes of Texas" (Houston KHOU-TV Channel 2) segment and two Houston KTRC-TV News Channel 13 segments. Not for reproduction.
	3	VHS videotape "Texas City - History Channel" 45 minutes. Not for reproduction.
	4	DVD access copy of videotapes: "Texas City - 1997 Reviews" 20 minutes & "Texas City - History Channel" 45 minutes. Not for reproduction.



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)

The screenshot displays the Rice Digital Scholarship Archive interface. At the top, the Rice Digital Scholarship Archive logo is on the left, and the Fondren Library logo is on the right. The breadcrumb trail reads: Rice Scholarship Home > Woodson Research Center > Lamar travel journal > View Item. A search bar is located in the top right corner.

The main content area features the title "Mirabeau B. Lamar travel journal, 1835" with a "Show full item record" link below it. A metadata table follows:

Title:	Mirabeau B. Lamar travel journal, 1835
Author:	Lamar, Mirabeau Buonaparte, 1798-1859
Abstract:	On his 1835 trip from Georgia to Texas, Mirabeau Buonaparte Lamar kept a manuscript diary. This journal is written in continuous narrative form, with frequent historical or descriptive passages inserted, covering the months June-October, 1835, the period during which Lamar apparently made his decision to settle in Texas permanently and join in the Texian battle for independence from Mexico. Lamar went on to serve the Republic of Texas' second President.
Description:	64 handwritten pages in a ledger
URI:	http://hdl.handle.net/1911/21658
Date:	1835

Below the table, there are links for "Files in this item": "Full book in HTML, with images (preferred reading version)" and "XML version of TEI-encoded text (for download and analysis)".

A section titled "This items appears in the following Collection(s)" lists "Lamar travel journal".

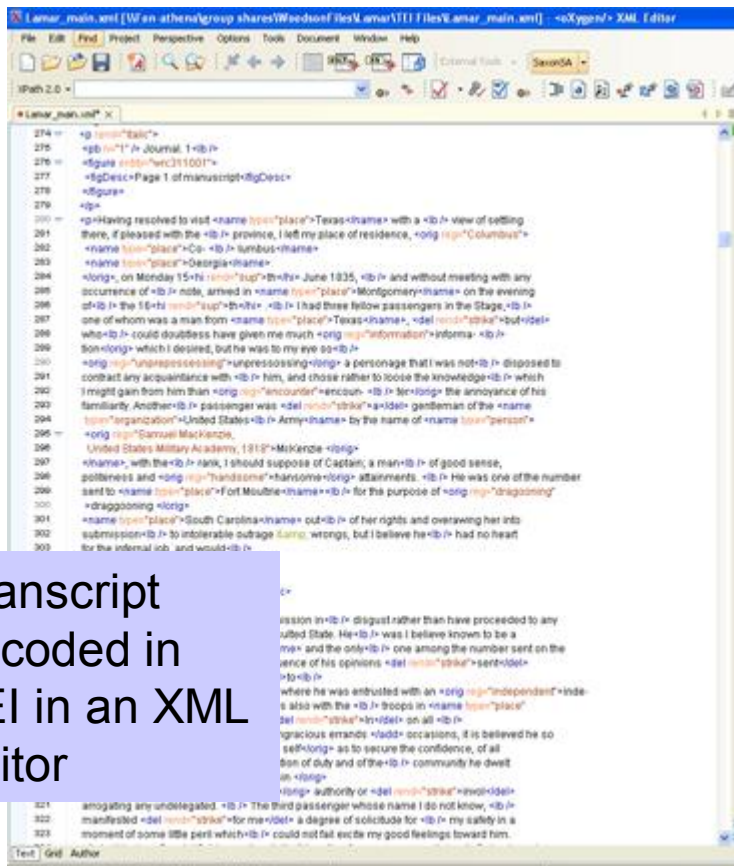
On the right side, there are navigation sections: "Browse" with links for "Entire Archive" (Communities & Collections, Titles, Authors, Subjects, By Dates) and "This Collection" (Titles, Authors, Subjects, By Dates); and "My Account" with links for "Login" and "Register".

The footer contains the breadcrumb trail, management information: "Managed by the Digital Library Initiative of Fondren Library, Rice University", and a "contact us" link.

- o Full item record shows all 17 qualified Dublin Core metadata fields

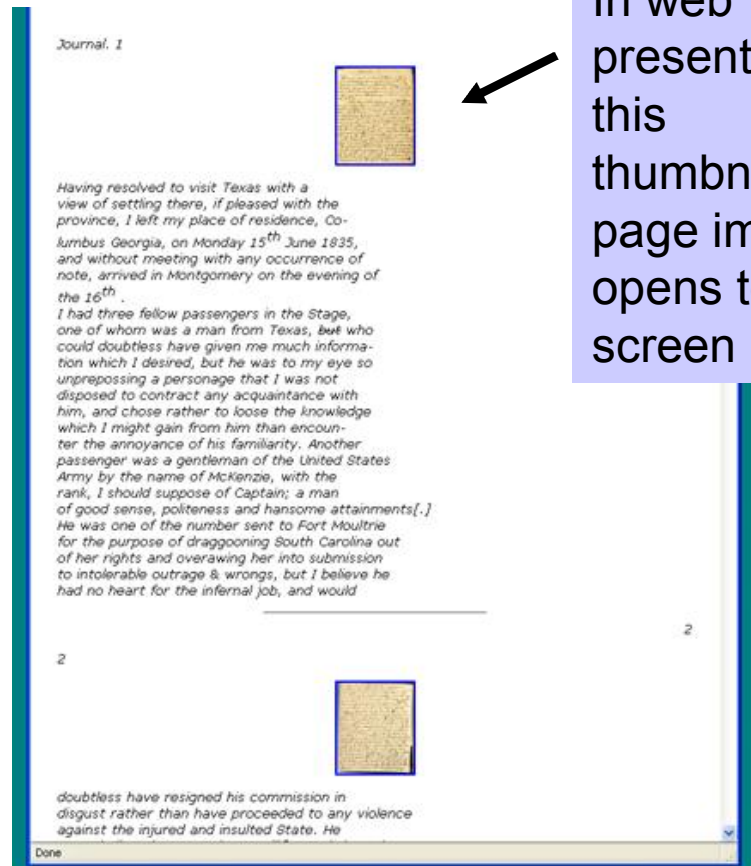
Full process for *handwritten text at item level cont.:*

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



```
<?xml version="1.0" encoding="UTF-8" type="text" />
<TEI xmlns="http://www.tei-c.org/ns/1.0" type="text">
  <fileDesc>
    <title>Journal. 1</title>
    <pageDesc>Page 1 of manuscript</pageDesc>
  </fileDesc>
  <text>
    <p>Having resolved to visit <name type="place">Texas</name> with a <ref type="view" /> view of settling there, if pleased with the <ref type="place">Co. </ref> province, I left my place of residence, <ref type="place">Columbus</ref> <ref type="place">Ga. </ref> on Monday 15th June 1835, <ref type="place">Montgomery</ref> on the evening of the 15th <ref type="place">Montgomery</ref> I had three fellow passengers in the Stage, <ref type="place">Texas</ref> one of whom was a man from <ref type="place">Texas</ref> who <ref type="place">Texas</ref> could doubtless have given me much information <ref type="place">Texas</ref> which I desired, but he was to my eye so <ref type="place">Texas</ref> unprepossessing <ref type="place">Texas</ref> a personage that I was not <ref type="place">Texas</ref> disposed to contract any acquaintance with <ref type="place">Texas</ref> him, and chose rather to loose the knowledge <ref type="place">Texas</ref> which I might gain from him than <ref type="place">Texas</ref> the annoyance of his familiarity. Another <ref type="place">Texas</ref> gentleman of the <ref type="place">Texas</ref> United States <ref type="place">Texas</ref> Army by the name of <ref type="place">Texas</ref> Samuel McKenzie, United States Military Academy, 1812 <ref type="place">Texas</ref> rank, I should suppose of Captain, a man <ref type="place">Texas</ref> of good sense, politeness and <ref type="place">Texas</ref> attainments. <ref type="place">Texas</ref> Fort Moultrie <ref type="place">Texas</ref> South Carolina <ref type="place">Texas</ref> submission <ref type="place">Texas</ref> wrongs, but I believe he <ref type="place">Texas</ref> had no heart for the infernal job, and would <ref type="place">Texas</ref> <ref type="place">Texas</ref> State. He <ref type="place">Texas</ref> was I believe known to be a <ref type="place">Texas</ref> and the only <ref type="place">Texas</ref> one among the number sent on the <ref type="place">Texas</ref> service of his opinions <ref type="place">Texas</ref> where he was embued with an <ref type="place">Texas</ref> independent <ref type="place">Texas</ref> also with the <ref type="place">Texas</ref> troops in <ref type="place">Texas</ref> on all <ref type="place">Texas</ref> gracious errands <ref type="place">Texas</ref> occasions, it is believed he so <ref type="place">Texas</ref> self <ref type="place">Texas</ref> as to secure the confidence, of all <ref type="place">Texas</ref> of duty and of the <ref type="place">Texas</ref> community he dwelt <ref type="place">Texas</ref> <ref type="place">Texas</ref> authority or <ref type="place">Texas</ref> <ref type="place">Texas</ref> The third passenger whose name I do not know, <ref type="place">Texas</ref> manifested <ref type="place">Texas</ref> a degree of solicitude for <ref type="place">Texas</ref> my safety in a moment of some little peril which <ref type="place">Texas</ref> could not fail excite my good feelings toward him.
```

Transcript
encoded in
TEI in an XML
editor



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



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)

[Show full item record](#)

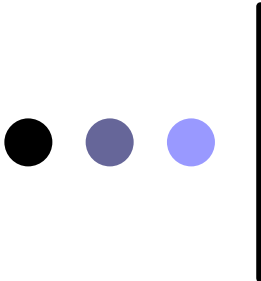
Title:	Texas currency: Diez pesos
Author:	Mexico. Tesoreria de la Nacion.
Abstract:	<p>This piece of Texan currency, printed for Stephen F. Austin's colony in 1823 when Texas was still a territory of Mexico, represents part of the transition that took place after Mexico declared its independence from Spain in 1821. The first side of the currency is actually an outdated Papal bull that was printed in 1784 while Mexico was still firmly under Spanish control. It makes reference to a recent decree from Madrid that Catholics could now eat egg and milk products except during Holy Week and that, in exchange for this "Indulgence," members of the church should contribute to a tax that was to be used to fund a war against "infidels." The document states that the amount of this tax was dependent upon the position of the church member. Due to a severe paper shortage the currency for the Austin colony was printed on the back of this outdated bull after Mexico had declared its independence from Spain, though this currency became quickly outdated as well. Following the Mexican War of Independence, Agustin de Iturbide proclaimed the Plan de Iguala, which declared Mexico to be independent of Spain and established the Mexican Empire in 1821. However, when the second side of this currency was printed just two years later, the Mexican Empire had already dissolved and the liberal-controlled Mexican Congress had disavowed the conservative Plan de Iguala and was working to establish the 1824 Constitution of Mexico. The printing on the second side of this document, then, represents an interim moment in the government of both Texas and Mexico.</p>
Description:	Printed document, 1pp. Very early currency for Austin's colony and Mexico. Due to a severe paper shortage, the currency was printed on the reverse of an out-dated papal bull.
URI:	http://hdl.handle.net/1911/9236
Date:	1823

Files in this item

Files	Size	Format	View
aa00015_0001.jpg	140.9Kb	image/jpeg	
aa00015_0002.jpg	174.8Kb	image/jpeg	

[Need help?](#)

- 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