Print to Petabyte:
Creating Knowledge from Collection Information

Fenella G. France PhD MBA FAIC
Chief, Preservation Research and Testing Division,
Library of Congress
Overview

• How do we effectively address the physicality issues of print collections?
• Understanding and interpreting collection data is a critical concern
• To advance presence in the digital realm we need to understand collections;
  – what can be digitized? - and depending on condition,
  – how effective decisions are made for prioritizing digitization
• Many institutions are currently making withdrawal and retention decisions based upon subjective and incomplete information
• The ultimate goal is to fill gaps in knowledge about materials at risk, and allow institutions to accurately predict good quality and poor quality copies of books
Research Project

- Andrew W. Mellon Foundation Grant (Scholarly Communications)
- $540,000 for 40 months Jan 2019 – Apr 2022
- Goal: to provide data to objectively assess the condition of the books held in the United States by performing an in-depth scientific analysis on a stratified random representative sample from the time period 1840-1940 (total of 3000 volumes)
- The physical, chemical and optical data will help ensure that large-scale withdrawal of materials does not compromise the overall robustness of cultural heritage collections, informing the shared print, preservation and digitization communities

https://nationalbookcollection.org/overview
Research Library Partner Institutions

- Cornell University
- University of Miami
- University of Washington
- Arizona State University
- University of Colorado at Boulder

*Analyze the same 500 volumes from all partners*
Assessing the Physical Condition of the National Book Collection.
Data Analytics and Possible Trends

What can we find out?
What factor(s) contribute most or have a greater influence on the current state of the text block?

- The inherent properties of the paper and paper composition?
- Impact of Environment
- Usage

Partners are providing environment and circulation data that is available

Data suggests a 5% loss in condition from collection use (twenty additional checkouts)*

*EAST (Boston Consortium)
<table>
<thead>
<tr>
<th>ID</th>
<th>Cat. ID</th>
<th>Inst.</th>
<th>Cat. Batch</th>
<th>Batch Recv.</th>
<th>Date</th>
<th>Title</th>
<th>Vol.</th>
<th>Double Fold</th>
<th>Catalog</th>
<th>Vis. Assess</th>
<th>Records Match</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>10000</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1840</td>
<td>Chronica Jocellini de Brakelond, de rebus gestis Semsonis abbatis ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10001</td>
<td>10001</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1840</td>
<td>The chronicle of William de Rishanger, of the barons' war, the ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10002</td>
<td>10002</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1842</td>
<td>Three early English metrical romances : with an introduction ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10003</td>
<td>10003</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1843</td>
<td>The diary of Dr. Thomas Cartwright, bishop of Chester; commencing ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10004</td>
<td>10004</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1844</td>
<td>Mexico as it was and as it is</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10005</td>
<td>10005</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1847</td>
<td>A commentary of the services and charges of William Lord Grey ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10006</td>
<td>10006</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1847</td>
<td>A relation, or rather A true account, of the island of England ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10007</td>
<td>10007</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1848</td>
<td>Diary of Walter Yonge, esq.; justice of the peace, and M.P. for ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10008</td>
<td>10011</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1849</td>
<td>The visitation of the county of Huntington, under the authority ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10009</td>
<td>10012</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1851</td>
<td>Parts, or, Scenes and adventures on the banks of the Amazon</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10010</td>
<td>10014</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1852</td>
<td>A journey to the tea countries of China : including Sung-lo and ...</td>
<td>1</td>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10011</td>
<td>10015</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1853</td>
<td>Account of the expenses of John of Brebet and Thomas and Henry ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10012</td>
<td>10016</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1853</td>
<td>Treawny papers</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10013</td>
<td>10017</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1853</td>
<td>Household expenses of the Princess Elizabeth during her residence ...</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10014</td>
<td>10018</td>
<td>ASUL</td>
<td>1</td>
<td>1</td>
<td>1853</td>
<td>The homes of the New World : impressions of America</td>
<td>1</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Visual Assessment

• How are we determining the condition of our collections in a standard / consistent way that allows us to know how our collection compares to another?
• Are we effectively using our knowledge to make retention decisions?
• Do we know what is the “same” volume?
• Can we trust our catalog information?
Assessing the National Collection Project Details

**Book ID**
- book:10012

**Catalog ID**
- catalog:10016

**Institution ID**
- institution:00002

**Batch as Received**
- 1

**Catalog Book**
- 10016: Trelawny papers (1853)

**Syntax**
- Catalog ID; Title (abbrev.) (Date)

**Received On**
- 2019-06-20

**Returned On**
- 2019-12-05

Created: Bernado_D, Jun 20, 2019.

**Catalog Details**

**Title**
- Trelawny papers
Catalog Details

Title: Trelawny papers

Volume: 1
Edition: 1853
Published Date (year): 1853
Publisher: Printed for the Camden
Location: London

Height (mm): 219
Width (mm): 165
Thickness (mm): 21
Number of Pages: 23

☑ Rebound collection of catalog books in one library volume?
Create separate book records for each ANC catalog entry in a (re)bound collection. Check this box for each record, assess, sample, and analyze each section as if they were distinct books. If an ANC book is bound with non-ANC books then describe the volume in the catalog notes field below.

Collection Includes:
catalog:10015, catalog:10017, c...
1853: Account of the expenses of John of Brabant and Thomas and Henry of Lancaster, A.D. 1292-3
1853: Household expenses of the Princess Elizabeth during her residence at Hatfield, October 1, 1551, to September 30, 1552
1853: Trelawny papers

Enclosures (describe)
e.g., box or tape. Leave blank for none.
**Description**

**Binding Notes**

- bound with 10011, 10013

Library rebinding with new blank endpapers. Perfect binding replaces stitching. Retrimmed.

Sticker, bottom p lli:

- 37
- 01/00
- 36 ASU
- Xl3
- 30-010-00
- ...
- 2240
- GBC
- ...
- ICI
- INFORMATION
- CONSERVATION, INC.
- ...

*Note especially attempted repairs that might have compromised the book's integrity e.g., the use of adhesive plastic to 'protect' the cover.*
Visual Assessment: Condition

Condition > Lendable? Deteriorating?
- Yes, the library would lend this book.
- This book's condition is stable.

Condition > Cover

<table>
<thead>
<tr>
<th>Damage</th>
<th>Damage (cont.)</th>
<th>Partially Detached</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worn Edges</td>
<td>Damaged Spine</td>
<td>Front</td>
<td>Obvious Repair</td>
</tr>
<tr>
<td>Damaged Corners</td>
<td>Broken Spine</td>
<td>Back</td>
<td>Label</td>
</tr>
<tr>
<td>Torn</td>
<td>Warping</td>
<td>Spine</td>
<td></td>
</tr>
<tr>
<td>Delaminating</td>
<td>Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discoloring</td>
<td>Pastedown Lifting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staining</td>
<td>Powdering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cover Condition Notes

Repair: rebound, see cover description.
**Condition**

- **Textblock**
  - **Physical**
    - [ ] Brittle or Crumbling
    - [ ] Corners
    - [ ] Damage
    - [ ] Loss
    - [ ] Pages
    - [ ] Loose (Free)
    - [ ] Torn
    - [ ] Partial Loss
    - [ ] Complete Loss
  - **Structural**
    - [ ] Textblock Separation
    - [ ] Tight Binding
    - [ ] Uncut
    - [ ] Untrimmed Pages
    - [ ] Non-Uniform Textblock Edges
    - [ ] Ink Changes
    - [ ] Fading
    - [ ] Strike-Throughs
    - [ ] Transfer
  - **Visual**
    - [ ] Staining
    - [ ] Foxing
    - [X] Paper Edge Discoloring
    - [ ] Paper Color
      - [ ] White
      - [ ] Cream
      - [X] Apparent Overall Yellowing
    - **Event Traces**
      - [X] Obvious Repair
      - [ ] Water Damage
      - [ ] Insertions
      - [ ] Marginalia, Inscription, Ex Libris...
    - **Written Annotations**
      - [ ] None
      - [ ] Minor
      - [ ] Major

- **Missing Pages**
  - Enter as a space delimited list, e.g.: 12 38 98

- **Missing Partial Pages**
  - Enter as a space delimited list, e.g.: 64 712 99

- **Textblock Condition Notes**
  - Repair: textblock backing replaced with perfect binding, re-trimmed.
What does “identical” look like?
Multiple Paper Types in One Volume
Visual/Subjective Tests

• Double fold

• Spot tests to determine sizing \(\text{\textit{(starch, rosin, protein, alum)}}\)

  – Correlate with infrared (non-invasive) data to assess accuracy of subjective tests

• “Volatile profile” \(\text{\textit{(book sniffing)}}\)

  – Can we determine collection health from the “breath”?
Linking Visual (subjective) Assessment with Lab Testing

<table>
<thead>
<tr>
<th>ANC Book ID</th>
<th>10001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td>London: The Camden Society, 1840. (180 years old.)</td>
</tr>
<tr>
<td>Double Fold (&lt;=10)</td>
<td>10</td>
</tr>
</tbody>
</table>

**Visual Assessment**

<table>
<thead>
<tr>
<th>General Book Notes</th>
<th>matches catalog record, rebound collection of books</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover Description</td>
<td>hardcover, cloth, sewn leaf attachment, green overall color</td>
</tr>
<tr>
<td>Textblock Description</td>
<td>black ink, normal paper, letterpress</td>
</tr>
<tr>
<td>Overall Condition</td>
<td>lendable</td>
</tr>
<tr>
<td>Cover Condition</td>
<td>label</td>
</tr>
<tr>
<td>Cover Damage</td>
<td>damaged corners</td>
</tr>
<tr>
<td>Textblock Structural</td>
<td>non-uniform edges</td>
</tr>
<tr>
<td>Textblock Visual</td>
<td>paper color: yellowing, edge discoloring</td>
</tr>
<tr>
<td>Textblock Event Traces</td>
<td>marginalia</td>
</tr>
</tbody>
</table>
Analytical Tests (*miniaturized*)

- Smallest sample required – a strip (10mm* x 100mm) taken from the replicate volumes for all the destructive (invasive) and non-invasive tests.

- **Mini tensile test**
  - The strength of the paper

- **Acidity - pH**
  - Mini-pH test requires 3 hole punches of sample compared to usual 2.5 g (1-2pgs)

- **Size exclusion chromatography (SEC)**
  - The chain lengths of cellulose in paper can indicate loss of physical integrity - molecular weight (1mg sample size)

*10mm ~ 3/8”*
### pH

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Page</th>
<th>Specimen</th>
<th>Probe</th>
<th>Mass (g)</th>
<th>Cal. Slope</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANC001095</td>
<td>47</td>
<td>1</td>
<td>1</td>
<td>0.0102</td>
<td>97.6</td>
<td>6.02</td>
</tr>
<tr>
<td>ANC001095</td>
<td>47</td>
<td>2</td>
<td>1</td>
<td>0.0097</td>
<td>97.6</td>
<td>4.64</td>
</tr>
<tr>
<td>ANC001095 Avg.</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.836</td>
</tr>
<tr>
<td>ANC001095 S.D.</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2689</td>
</tr>
</tbody>
</table>

### Tensile

<table>
<thead>
<tr>
<th>Sample</th>
<th>Drg</th>
<th>Spec.</th>
<th>Orient.</th>
<th>Max Load (N)</th>
<th>Stress @ Max Load (MPa)</th>
<th>Strain @ Max Load (mm/mm)</th>
<th>Energy Absorb. (N/mm)</th>
<th>Young's Mod. (MPa)</th>
<th>Thickness</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANC001095.t</td>
<td>47</td>
<td>1</td>
<td>horiz.</td>
<td>12.953</td>
<td>3</td>
<td>9.88031</td>
<td>0.01810</td>
<td>0.01310</td>
<td>1154.21267</td>
<td>10</td>
</tr>
<tr>
<td>ANC001095.t</td>
<td>47</td>
<td>2</td>
<td>horiz.</td>
<td>14.9708</td>
<td>11.09631</td>
<td>0.03214</td>
<td>0.02711</td>
<td>1766.60196</td>
<td>10</td>
<td>2019-06-06</td>
</tr>
<tr>
<td>ANC001095.t</td>
<td>47</td>
<td>3</td>
<td>horiz.</td>
<td>10.76705</td>
<td>8.54327</td>
<td>0.01677</td>
<td>0.01113</td>
<td>827.47181</td>
<td>10</td>
<td>2019-06-06</td>
</tr>
<tr>
<td>ANC001095.t</td>
<td>47</td>
<td>4</td>
<td>horiz.</td>
<td>12.55204</td>
<td>9.4762</td>
<td>0.02362</td>
<td>0.01805</td>
<td>1461.74999</td>
<td>10</td>
<td>2019-06-06</td>
</tr>
<tr>
<td>ANC001095.t</td>
<td>47</td>
<td>5</td>
<td>horiz.</td>
<td>12.95588</td>
<td>9.81488</td>
<td>0.02015</td>
<td>0.01598</td>
<td>1746.24915</td>
<td>10</td>
<td>2019-06-06</td>
</tr>
</tbody>
</table>

### Size Exclusion Chromatography – *Size of Molecule*
Analytical Tests (*miniaturized*)

Non-invasive

- **Infrared Spectroscopy (IR)**
  - Looking at IR data to see how it can be linked, integrated and correlated with the collective physical, chemical and optical measurements

- **Ultraviolet / Visible spectroscopy (UV/VIS)**
  - Optical measure of the difference between volumes held in different locations, (potential for moving towards completely non-invasive)
Fourier Transform Infrared Spectroscopy (FTIR)
Fiber Optic Reflectance Spectroscopy (FORS)

Chromaticity Diagrams (CIE color space)
How do we know if the papers are the same or not?
Paper Classification from Pulp Concentration (Chemometrics and FTIR)

Samples include known pulp type standard ISR papers, with known and unknown lab reference. Various pulp compositions. Rag, Newsprint.
Using results for On-site Evaluations/ “Stack Tests”

- This project is developing field evaluations or “stack tests” designed to assess the relative condition of monographs in the stacks, without requiring them to be brought to the laboratory.

- Stack tests have three constraining factors:
  - inexpensive, so can be readily available at multiple sites
  - quick, so that results can be useful in decision making, and
  - Simple, so they can be performed by available staff
CIE b* (yellowness in Color Space) vs pH
**pH by Decade**

- Distribution of pH results for books published **1840-1890** normalized by number of samples in specific decade

**Size Exclusion Chromatography (SEC)**

- SEC Mw results by decade for samples published 1840-1900 normalized by number of samples in a decade
For the books received to date, it is almost as likely that the paper will be strongest in the vertical direction as it is in the horizontal. i.e., that books in this period are not necessarily bound to make best use of the papers’ machine direction.
Correlating Double Fold to Other Measurable Parameters
Conclusions

• To assess physicality issues of print volumes, interpreting collection data is critical.

• Collecting physical, chemical, and optical data from 3000 volumes for books from 1840-1940 allows us to identify materials at risk, and allow institutions to accurately predict good and poor quality copies.

• Creating knowledge from this data has required information reuse, and constant re-interpretation from an active data infrastructure.
Acknowledgements

- Mellon Project Researchers: Drs. Andrew Forsberg, Robyn Hodgkins
- Ian Bogus (CoPI), Partners and partner institutions
- Jake Nadal
- Advisory Board
- Institution Partners
- PRTD Staff (Andrew Davis, Chris Bolser, Amanda Satorius, Jamie Shetzline, Kelli Stoneburner, Meghan Wilson, Emin Voloder)
- Multiple amazing interns...

https://nationalbookcollection.org/overview