

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

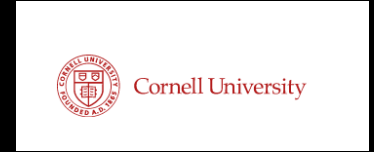
Research Project

- *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

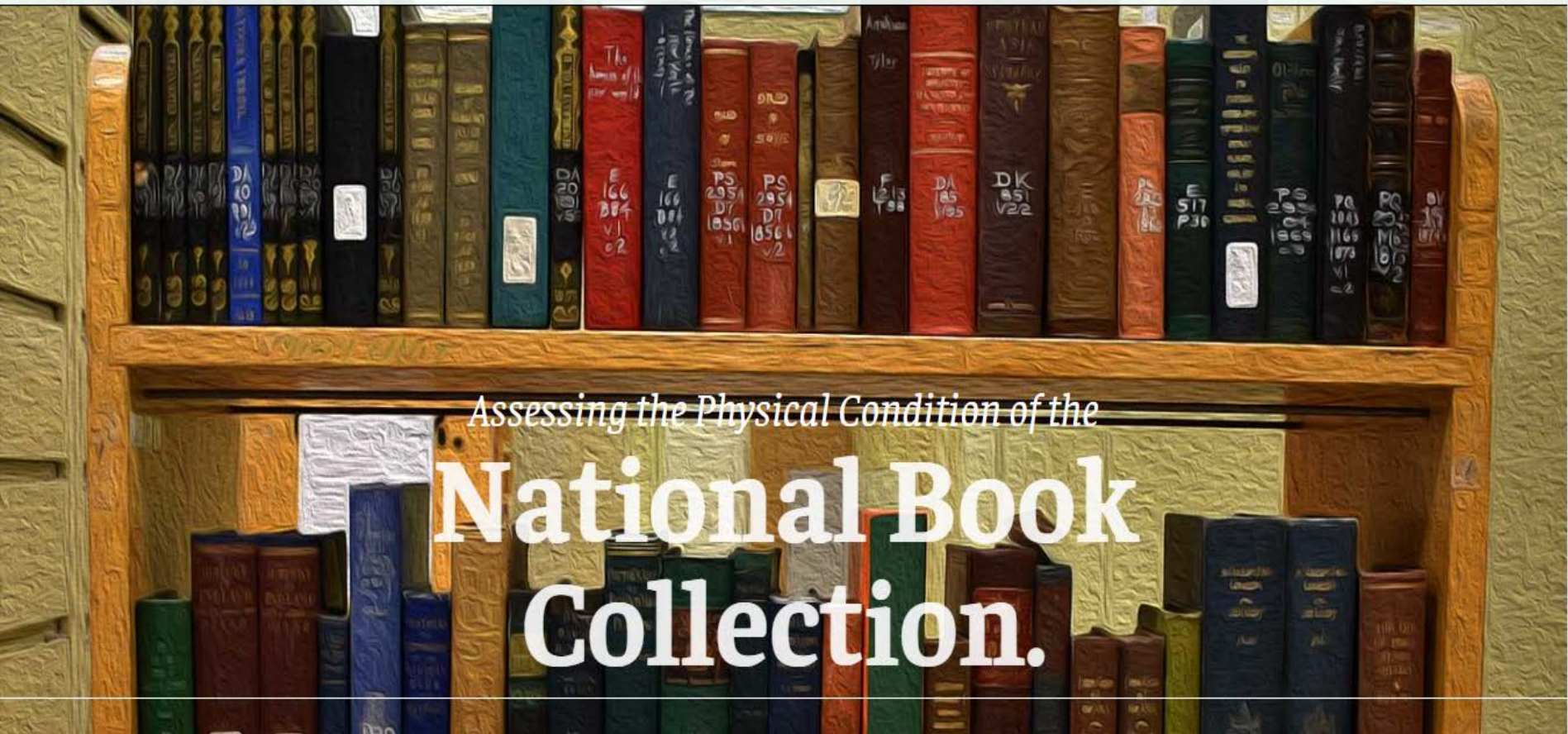
<https://nationalbookcollection.org/overview>

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

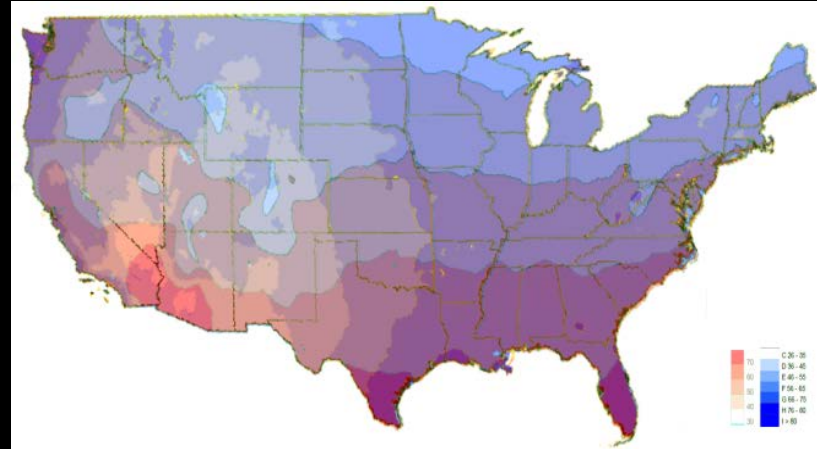
National Book Collection.

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)*

ANC Books

COMPARE

0

ADD

Institution Catalog Batch Batch as Received From To Search Reset

CONTROL LC ASUL CUL UCBL UML UWL 0 1 2 3 4 5 6 N/A 1 2 3 4 5 6 1800 2020 X

Page: 1 Rows per page: 25 1-25 of 165

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

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?

View an ANC Book



EDIT



SAMPLES



ANALYSES



GO BACK

Assessing the National Collection Project Details

Book ID
book:10012

(Read only)

Catalog ID
catalog:10016

(Read only)

Institution ID
institution:00002

(Read only)

Batch as Received *
1

From Partner (Required)

Catalog Book *
10016: Trelawny papers (1853)

Syntax: `Catalog ID: Title (abbrev.) (Date)` (Required)

Institution *
ASUL: Arizona State

(Required)

Received On
2019-06-20



Returned On
2019-12-05



Created: Bernardo_D, Jun 20, 2019.
Last Modified: Forsberg_A, Mar 10, 2020.
Assessed: Hodgkins_R, Jun 20, 2019.

Catalog Details

Title *
Trelawny papers

Author

Catalog Details

Title *

(Required)

Volume

Height (mm)

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.

Use the select field below to indicate which ANC catalog books have been bound together with this one:

Collection Includes

catalog:10015, catalog:10017, cat... ▾

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

e.g., box or tape. Leave blank for none.

Visual Assessment: Description

Description > Binding

Leaf Attachment

- Sewn
- Adhesive
- Metal Fittings

Material

- Cloth
- Leather
- Paper

Format

- Hardcover
- Softcover
- None
- Missing

Overall Color

- Black
- White
- Gray
- Brown
- Multi
- Red
- Orange
- Yellow
- Green
- Blue
- Purple

Other

- Obvious Rebinding
- Edge Decoration
- Collection of Volumes

Description

Binding Notes

bound with 10011, 10013

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

Sticker, bottom p iii:

```

---
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.

Binding Notes

bound with 10011, 10013

Description > Textblock

- Printing Ink: Black Color
- Textblock Paper: Normal Calendered Glossy
- Obvious Multiple Papers: Normal Calendered Glossy
- Other: Illustrations Letterpress Tip-Ins

Description

Textblock Notes

Description > Overall

Overall Description Notes

Visual Assessment: Condition

Condition > Lendable? Deteriorating?

- Yes, the library would lend this book.
- This book's condition is stable.



Condition

Visual Assessment: Condition

Condition > Lendable? Deteriorating?

Yes, the library would lend this book.

This book's condition is stable.

Condition > Cover

Damage

Worn Edges

Damaged Corners

Torn

Delaminating

Discoloring

Staining

Damage (cont.)

Damaged Spine

Broken Spine

Warping

Loss

Pastedown Lifting

Powdering

Partially Detached

Front

Back

Spine

Other

Obvious Repair

Label

Notations

Cover Condition Notes

Repair: rebound, see cover description.

Condition

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

Paper Edge Discoloring

Paper Color

White

Cream

Apparent Overall Yellowing

Event Traces

Obvious Repair

Water Damage

Insertions

Marginalia, Inscription, Ex Libris...

Written Annotations

None

Minor

Major

Missing Pages

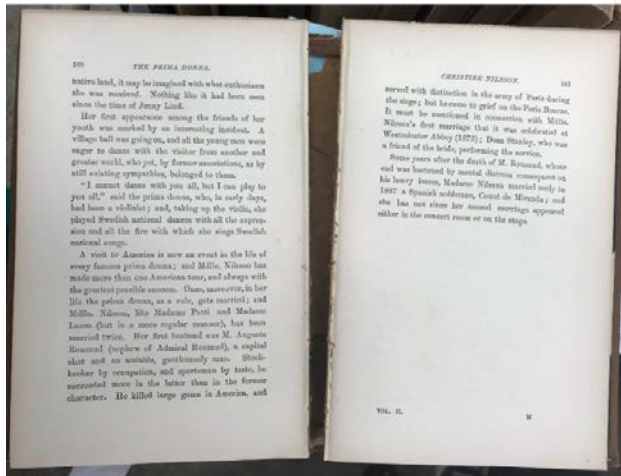
Enter as a space delimited list, e.g.: 12 38 98

Missing Partial Pages

Enter as a space delimited list, e.g.: 6 47 129

Textblock Condition Notes

Repair: textblock backing replaced with perfect binding, re-trimmed.



What does “identical” look like?



A stack of three books is shown from a side-on perspective, resting on a light-colored wooden surface. The top book has a light brown, textured cover and its pages are a pale, off-white color. The middle book has a dark brown, textured cover and its pages are a slightly darker, cream color. The bottom book has a black, textured cover and its pages are a light tan color. The text "Multiple Paper Types in One Volume" is overlaid in a black serif font across the top book's pages.

Multiple Paper Types in One Volume

Visual/Subjective Tests

- Double fold
- Spot tests to determine sizing (*starch, rosin, protein, alum*)



- *Correlate with infrared (non-invasive) data to assess accuracy of subjective tests*
- “Volatiles profile” (*book sniffing*)
 - Can we determine collection health from the “*breath*”?

10001

ASUL

The chronicle of William de Rishanger, of the barons' war, the miracles of Simon de Montfort.

William Rishanger

Linking Visual (subjective) Assessment with Lab Testing

Visual Assessment

ANC Book ID	10001
Published	London: The Camden Society, 1840. (180 years old.)
Double Fold (<=10)	10
General Book Notes	matches catalog record rebound collection of books
Cover Description	hardcover cloth sewn leaf attachment green overall color
Textblock Description	black ink normal paper letterpress
Overall Condition	lendable
Cover Condition	label
Cover Damage	damaged corners
Textblock Structural	non-uniform edges
Textblock Visual	paper color: yellowing edge discoloring
Textblock Event Traces	marginalia

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

SPECIMENS & STATS AVERAGES & S.D. AVERAGES ONLY

Barcode ▲	Page	Specimen	Probe	Mass (g)	Cal. Slope	pH
ANC001095	47	1	1	0.0102	97.6	5.02
ANC001095	47	2	1	0.0097	97.6	4.64
ANC001095 Avg.	47					4.830
ANC001095 S.D.	47					0.2687

pH Deltas

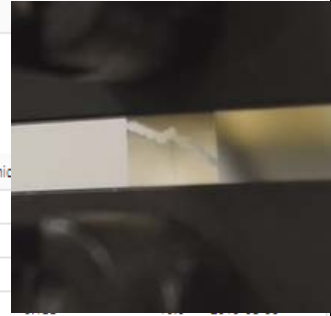
Sample ▲	ANC001095:1	ANC001095:2	ANC001095 Avg.
ANC001095:1	-	0.3800	0.1900
ANC001095:2	0.3800	-	0.1900
ANC001095 Avg.	0.1900	0.1900	-



Tensile

SPECIMENS & STATS AVERAGES & S.D. AVERAGES ONLY

Sample ▲	Pg.	Spec.	Orient.	Max Load (N)	Stress @ Max Load (MPa)	Strain @ Max Load (mm/mm)	Energy Absorpt. (N/mm)	Young's Mod. (MPa)	Thic
ANC001095:h:1	47	1	horiz.	12.90343	9.35031	0.01810	0.01310	1154.21267	
ANC001095:h:2	47	2	horiz.	14.37038	11.49631	0.03014	0.02711	1796.50136	
ANC001095:h:3	47	3	horiz.	10.76705	8.54527	0.01877	0.01113	922.47181	
ANC001095:h:4	47	4	horiz.	12.55204	9.43762	0.02326	0.01805	1461.74099	
ANC001095:h:5	47	5	horiz.	12.95558	9.81483	0.02015	0.01583	1746.24315	0.132 10.0 2019-08-06



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)



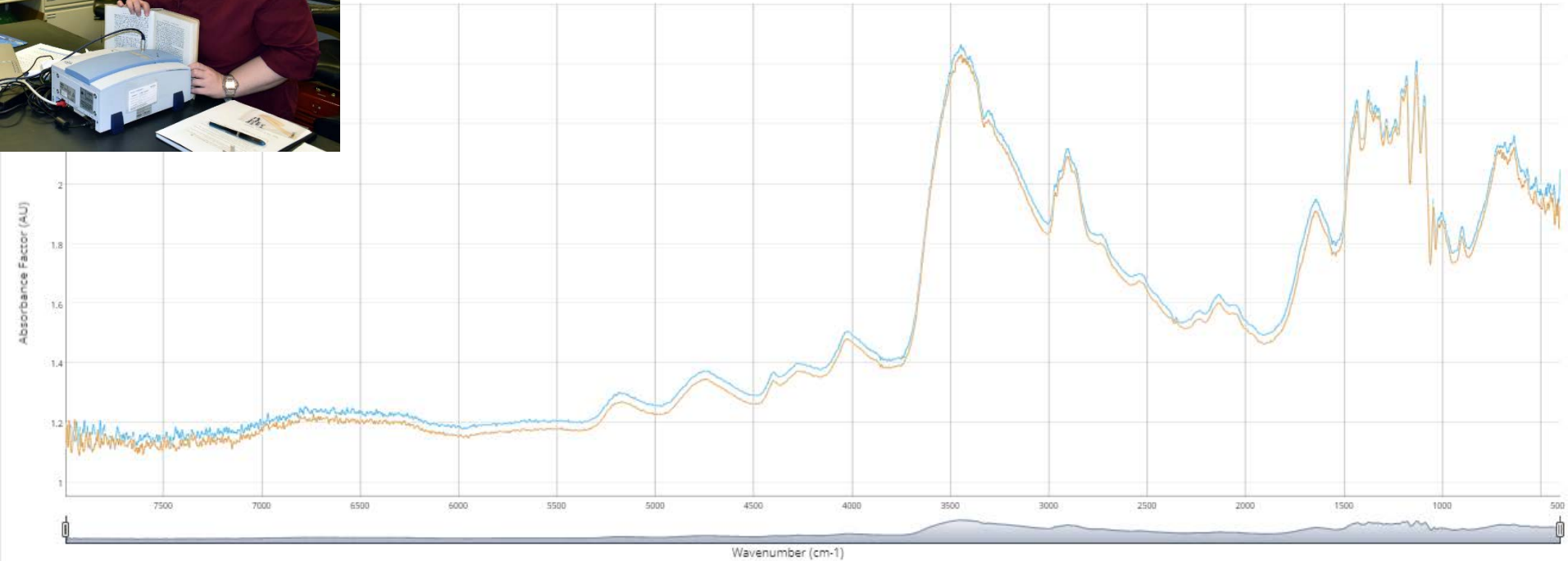
Fourier Transform Infrared Spectroscopy(FTIR)



/ses/book/10001/report



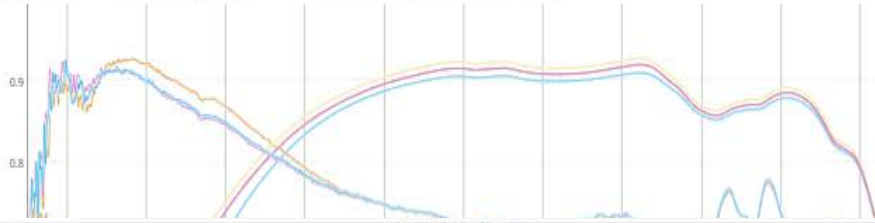
FTIR ☐



Fiber Optic Reflectance Spectroscopy (FORS)

admin.prtid.app/anc/analyses/book/10001/report

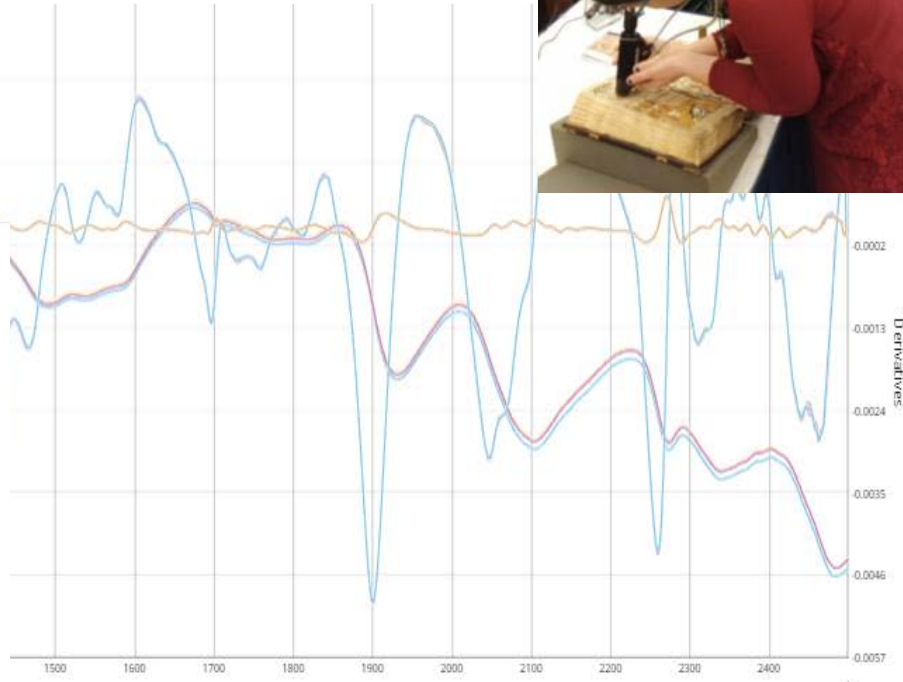
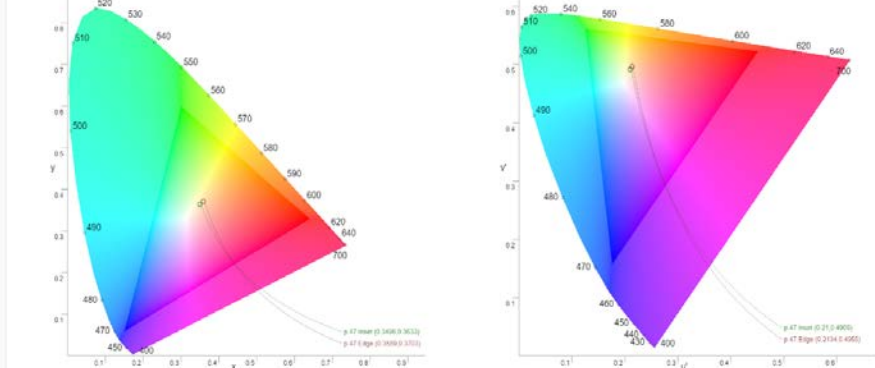
P.47 EDGE INSET GUTTER RAW 1ST.DERIV 2ND.DERIV Y NORM(Y) KM(Y)



Chromaticity Diagrams:

Reflectance Factor (AU)

Chromaticity Diagrams (CIE color space)



Wavelength (nm)



How do we know if the papers are the same or not?

Compare Books

 GO BACK

VIS. ASSESS. PHOTO DOC. PH TENSILE SEC SPOT XRF FORS FTIR CHROMATICITY COLORIMETRY **SAMPLE NOTES**

Reference Paper Notes

Paper:10000

Rag, c.1994, PAP00892c
Composition: Cotton Rag
Aging: Room, natural aging
Source: Unknown

Paper:10001

Rag, c.2007, PAP00891c
Composition: Cotton Rag
Sizing: Starch
Buffer: Buffered
Aging: Room, natural aging
Source: Stonehenge

Paper:10002

Newsprint, c.1979, PAP00880c
Composition: Wood Pulp
Aging: Room, natural aging
Source: Unknown

Paper:10003

Wood Pulp, c.1994, ISR5C0103b

ISR: 1

Composition: 100% Bleached Northern Softwood Kraft
Sizing: Alum/Rosin
Aging: 5°C, natural aging
Source: ASTM

Paper:10004

Wood Pulp, c.1994, ISR5C0302b
ISR: 3
Composition: 100% Softwood Bleached Chemi-thermomechanical Pulp
Sizing: Alum
Aging: 5°C, natural aging
Source: ASTM

Paper:10005

Rag, c.1994, ISR5C0503b
ISR: 5
Composition: 100% Cotton Rag
Sizing: Alum/Rosin
Aging: 5°C, natural aging
Source: ASTM

Paper:10006

Wood Pulp, c.1994, ISR5C0702b
ISR: 7
Composition: 20% Bleached Northern Softwood Kraft / 80% Slush-soft Ground Wood
Sizing: Alum
Aging: 5°C, natural aging
Source: ASTM

Paper:10007

Wood Pulp, c.1994, ISR5C1102b
ISR: 11
Composition: 50% Bleached Northern Softwood Kraft / 50% Bleached Northern Hardwood Kraft
Buffer: Na2CO3
Aging: 5°C, natural aging
Source: ASTM

Paper:10008

Wood Pulp, c.1994, ISR5c1502b

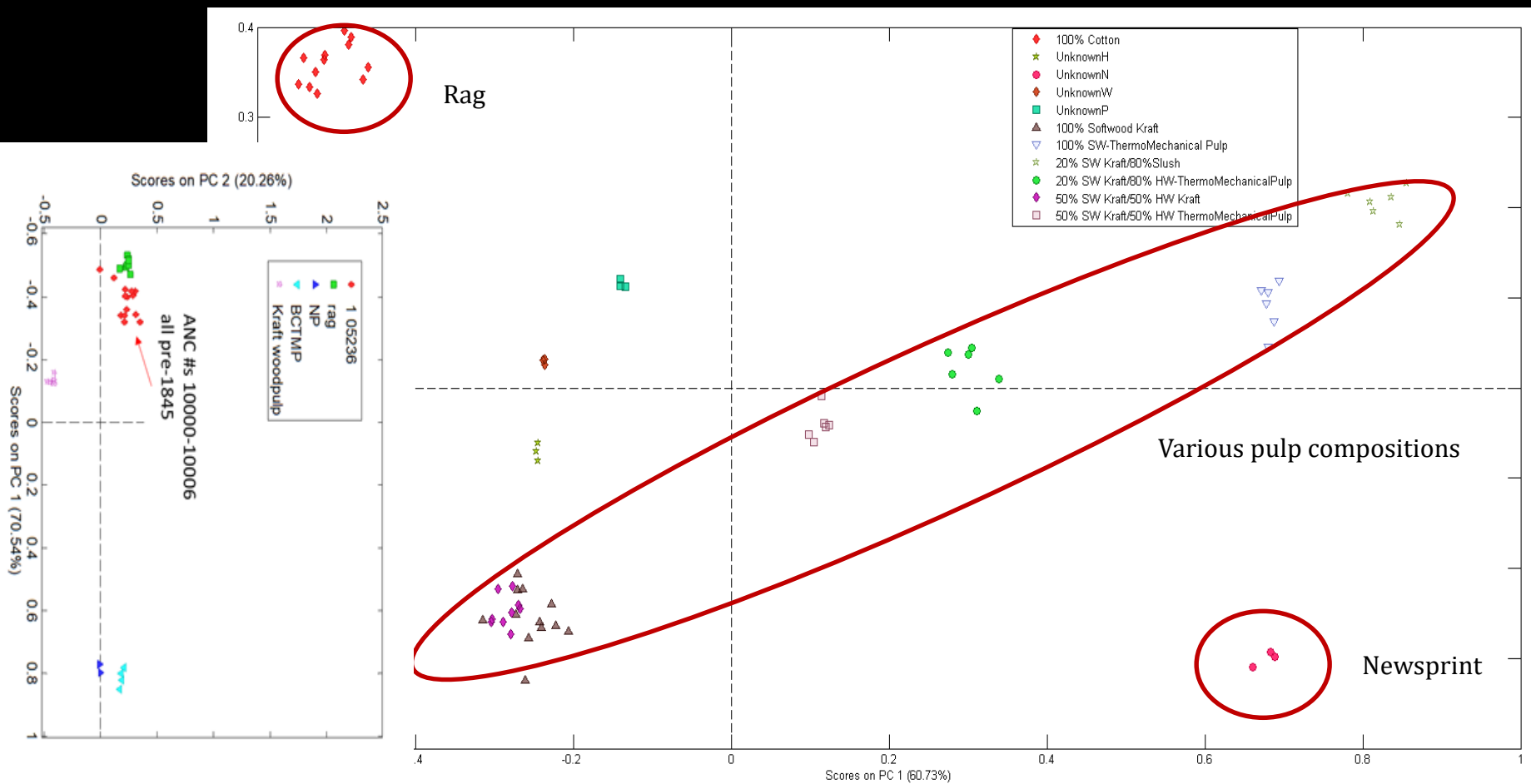
ISR: 15

Composition: 50% Bleached Northern Softwood Kraft / 50% Bleached Northern Hardwood Kraft
Sizing: Internal and external starch
Buffer: Na2CO3/CaCO3
Aging: 5°C, natural aging
Source: ASTM

Paper:10009

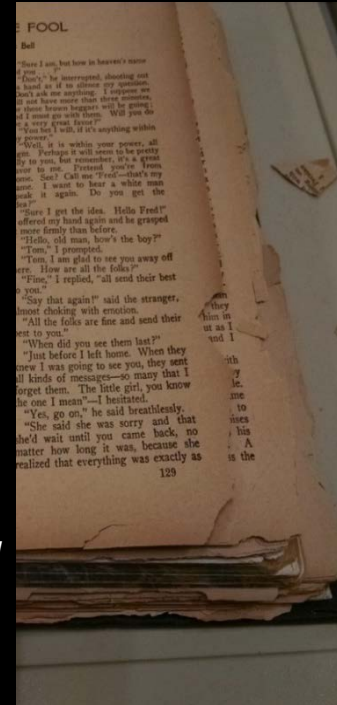
Rag, 2019, PAP00931a
Composition: Cotton Rag
Sizing: Gelatin
Buffer: Unbuffered (Ca might have been used during production)
Aging: Room, natural aging
Source: University of Iowa, Center for the Book

Paper Classification from Pulp Concentration (Chemometrics and FTIR)



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



Query Analyses

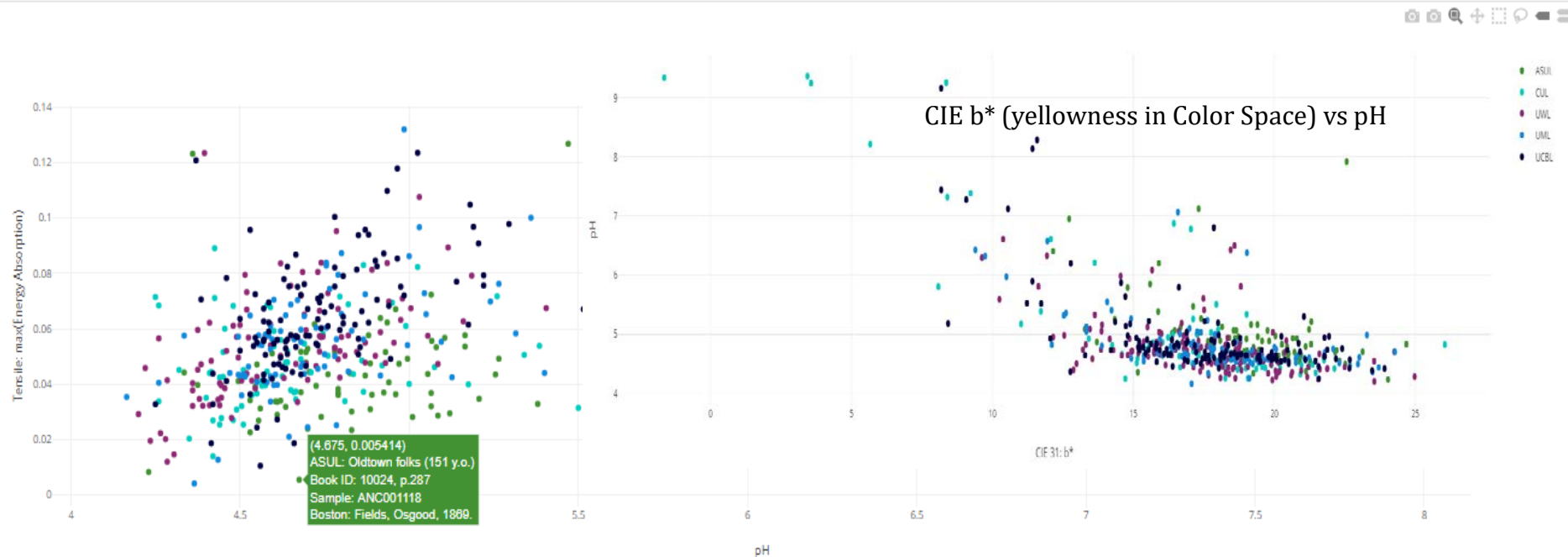
GO BACK

From Decade: 1840 To Decade: 1890

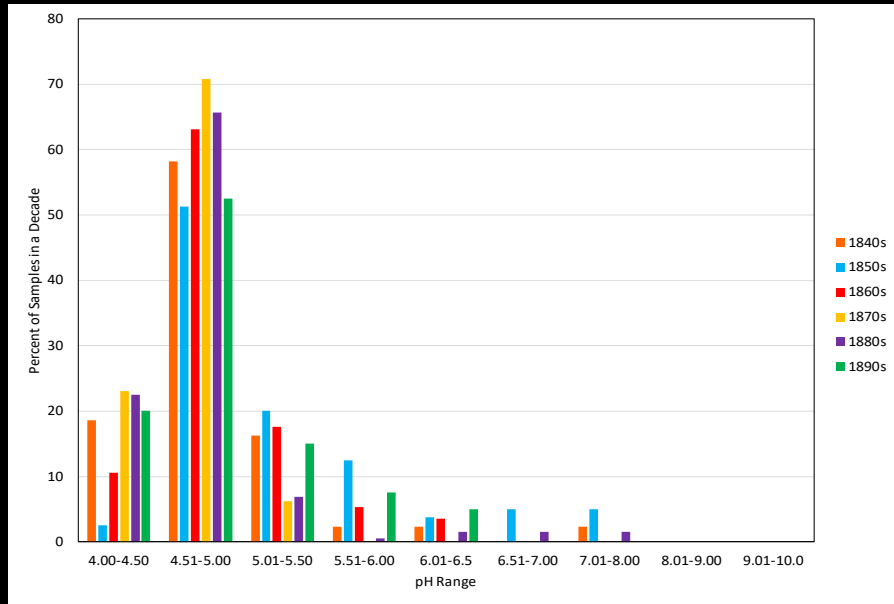
NORMAL
 CALENDERED
 ALUMINUM
 LIGNIN
 PROTEIN
 ROSIN
 STARCH
 Include Controls

X: pH CIE Observer: CIE 1931 2° Colorimetry Loc.: Edge ΔE* Formula: CIE 2000 Tensile Orient.: Max()

Y: Tensile: Energy Abs. CIE Observer: CIE 1931 2° Colorimetry Loc.: Edge ΔE* Formula: CIE 2000 Tensile Orient.: Max()

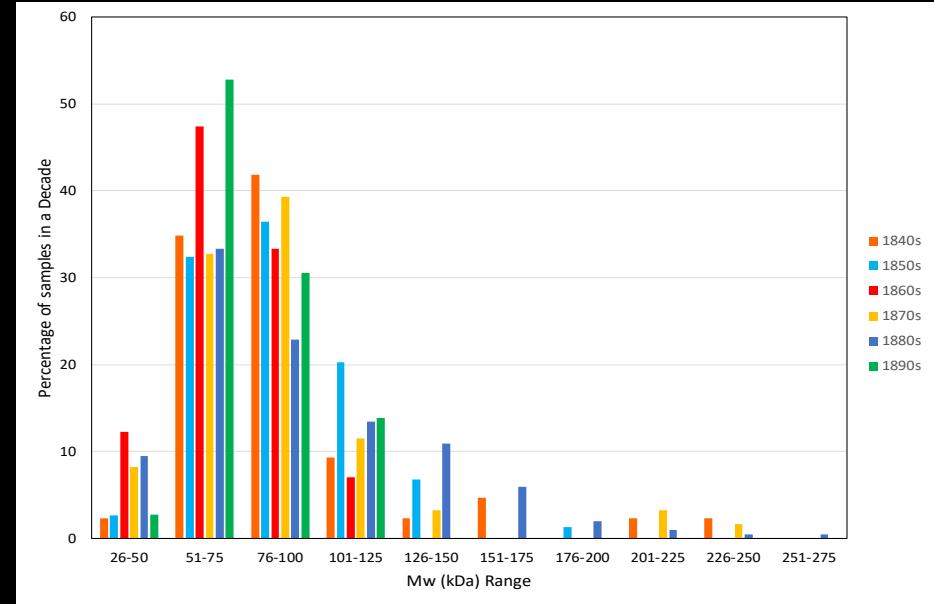


pH by Decade



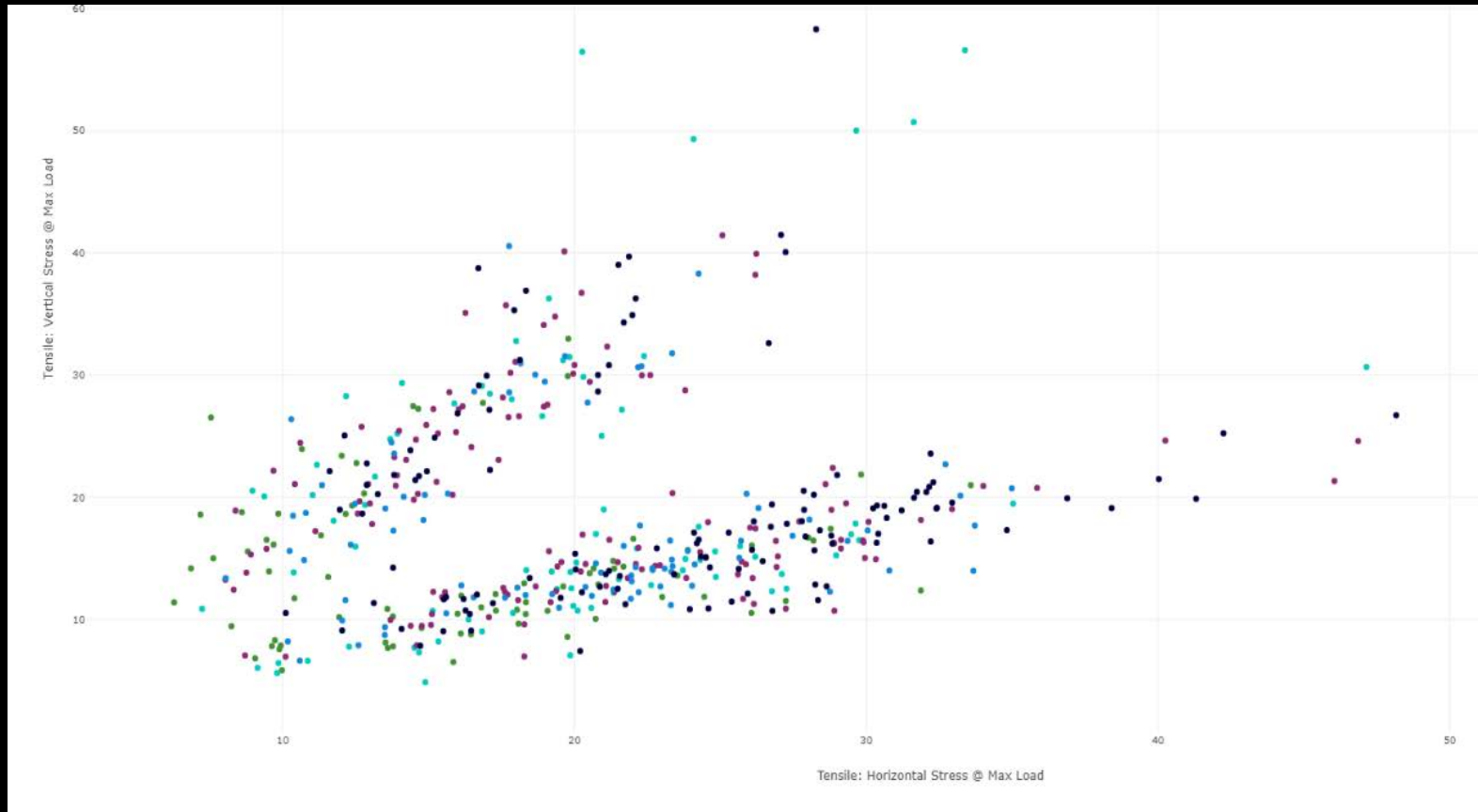
Distribution of pH results for books published **1840-1890** normalized by number of samples in specific decade

Size Exclusion Chromatography (SEC)



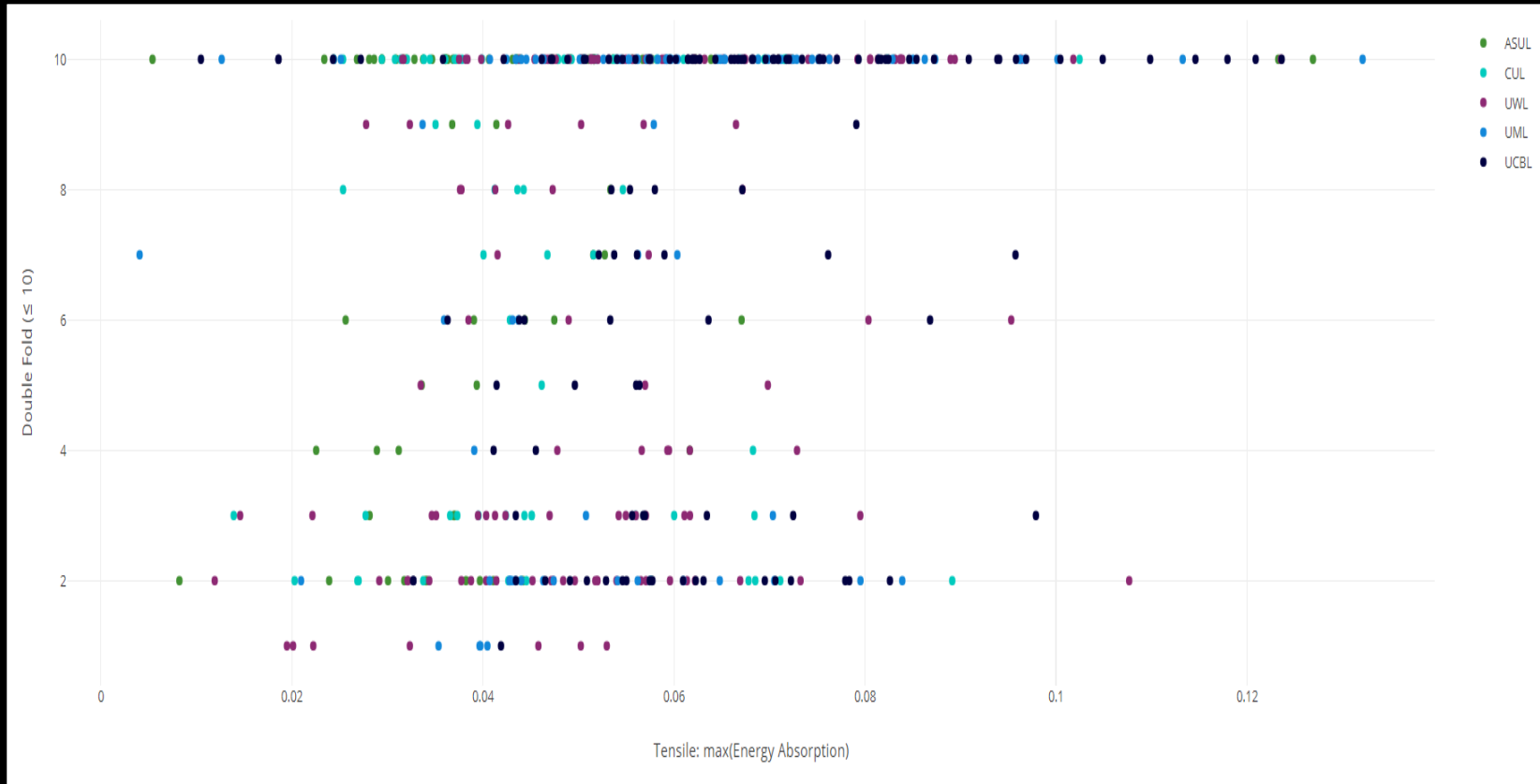
SEC Mw results by decade for samples published 1840-1900 normalized by number of samples in a decade

Horizontal vs Vertical Maximum Stress



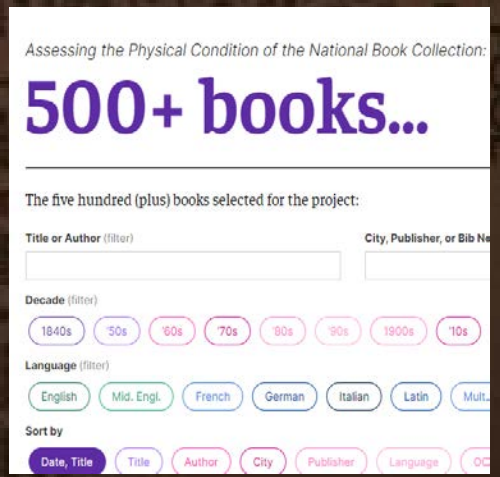
For the books received to date, it is almost as likely that the paper will be strongest in the vertical direction as it is in the horizontal. i.e., that books in this period are not necessarily bound to make best use of the papers' machine direction

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>