

The Book Traces Project: Implications for Shared Print Collections

Kara M. McClurken
Print Archive Network
June 23, 2017

UNIVERSITY
of VIRGINIA
LIBRARY

Thanks for having me here today—how many of you have heard about the Book Traces project?

Book Traces @ UVA

Looking at 180,000 pre-1923 volumes...

- Identify items with unique artifactual value
- Describe the unique features
- Route materials to preservation, special collections, or limited access
- Develop model protocol for identifying and describing
- Share protocol with other institutions

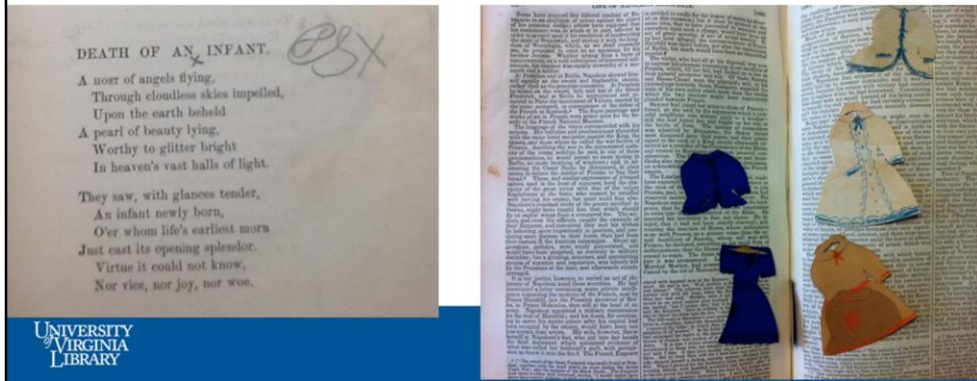


In 2014, UVA was awarded a CLIR Hidden Collections grant to examine the Library's pre-1923 items in our circulating collections. The goal of the project was to develop a protocol for identifying, describing and preserving evidence of past readership in our books located in the circulating collections. We hoped to develop both a model and a conversation to have with other institutions about what makes a book unique, how to identify and describe those unique features, and why those books might be worthy of preservation.

We particularly hoped to reach out to those institutions considering large catalog driven weeding and deduping projects, off-site storage or shared print repository projects—since I am here today, I guess we were successful in broadcasting our message.

What Makes a Book Unique?

- How do we define uniqueness?
- How do you discover in a world of copied catalog records?



When we started the project we had to ask the question—what makes a book unique enough to be worthy of this extra effort. There are two parts to this question. One, what is the threshold by which you define a book as unique? If I spill hot chocolate on a book, I have changed the colors of some of the pages? Is that worthy of preservation and description? Probably not...

On the other hand, if I am a 19th Century grieving mother who finds solace and comfort in the words of poet, and I write the names of my lost children in the margins, is that interesting enough? Especially if other mothers have done similar things in other copies of that volume of poetry?

Or if I am a soldier in the Civil War, and I use the flyleaves to sketch out the battlefield, presenting a view that maybe was unknown to historians before, would you want to preserve this volume?

The second question we wanted to explore is, if the volume is in the circulating stacks, how would you know about the battlefield sketch? Especially if my institution merely copied the record from some other institution when the book was acquired, so for all intents and purposes, the catalog reveals none of this volume's secrets?

Develop Criteria

- Examine samples
 - What is interesting?
 - What contributes to the larger body of scholarship?
 - Is enough information there to determine context?
 - Name
 - Date AND
 - Location
 - OR In combination with other, notable interventions

UNIVERSITY
of VIRGINIA
LIBRARY


The first question required that we develop a set of criteria to see if a book met our thresholds

To determine a threshold for uniqueness for the project, several of us with different scholarly interests sat around the table and physically examined samples of interventions. Should we describe every little underline or notation, no matter how tiny? Should we rate on a scale of interestingness?

Ultimately, we decided to look primarily for interventions that looked like they were from the original owner. Although sometimes that was impossible to determine, some things were obvious—we did not include ballpoint ink or highlighters, since those modifications occurred with technologies developed long after our 1923 cutoff publication date.


Common marks, like underlining alone, told us little information, so unless it was combined with other interesting features, we ignored underscoring. Also, many donated books had inscriptions or the names of the original owners, but without additional details to provide context, that information didn't tell us very much—We would only apply the term "inscription" if we had the owner's name, a date, and a place name (or if it had some of those features with additional information that would make it more likely to place the book in the context of the owner's life.)

Efficient Workflows



1) Identify candidates in the stacks

2) Record specific modifications at a desk



UNIVERSITY OF VIRGINIA LIBRARY

The second part of the question—how to identify these books that were sitting in the circulating stacks, usually without any local notes, was to design a survey and workflow that would allow us to work quickly and accurately.

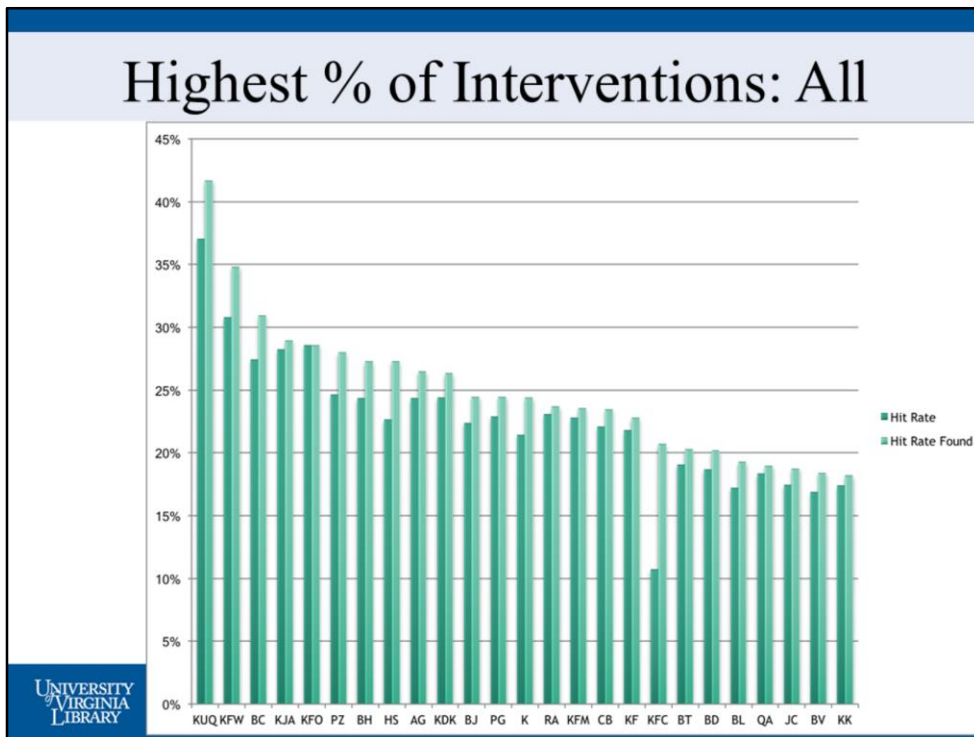
Having an efficient process for carrying out the survey has been essential to the cost-effectiveness of the work. Our process is designed to optimize a balance of speed and accuracy in recording data about the books in our open stacks. Our project assistants spend about 80% of their work time in the stacks, using a book truck as a mobile work station with a laptop and barcode scanner. Working from a list of pre-1923 monographs, the assistants pull each book from the shelf and give it a preliminary examination. Books that have unique modifications meeting our criteria are set aside on the book trucks for further analysis; others are returned to the shelf. Near the end of each shift, the assistants do a second inspection of the books they pulled and record the specific modifications, such as marginalia or insertions, using a Google form based on the controlled vocabulary of the Provenance Evidence Thesaurus. This two-step process promotes accuracy of description by allowing project assistants to concentrate on description separately from the task of finding and pulling books. Throughout the project, we have constantly gathered data about our student workers' efficiency; knowing how many books they can survey per hour allows us to plan our future work accordingly. Although each student had different rates of efficiency, in the end it took about a minute a book to do the survey.

[This workflow, including the statistical sampling approach, could easily be modified to survey collections for any other features of the physical books, such as preservation needs, notable bindings, or bookplate names. In fact, we are doing a bookplate census of all the books we touch: one of the steps in the workflow is to record the bookplate name for each book. Piggybacking this survey onto the Book Traces project has added a marginal amount of time to the handling of each book but has allowed us to gather lots more data about the library's holdings and reunite certain historical collections that were previously dispersed and essentially hidden in the vastness of the stacks. Similarly, you could piggyback a Book Traces step onto another type of collections survey, or any library process where large numbers of older books are being handled.]

Next Steps for Books

- Return to stacks as is
- Go to Metadata Analysis and Design (MAD) for bibliographic correction
- Go to Preservation
- Transfer insertions to Medium Rare
- Transfer identified bookplated collections to Special Collections

Where did books go after being surveyed? It depends on what they needed and how they had been identified. Most returned to stacks without any further work. The ones that had cataloging errors were sent to MAD for correction. Those with preservation needs sent to my unit, for repairs and boxing. Items with insertions were transferred to medium rare, where they received a better housing environment and are accessible via the Special Collections reading room. We used the opportunity of the survey to attempt to reunite some bookplated collections that had been distributed between circulating and Special Collections over the years. So the items designated for reunification were transferred to Special Collections.



Overall, we touched almost 116,000 volumes; We surveyed all of Alderman Library, our main social science and humanities collection, as well as our science and engineering library. We pulled samples from off-site storage, music and the Law Library. This chart includes the highest hit rates by call numbers. You can see that the Law collection skewed the results a bit.

Average hit rate was 12.08% but the top scorers were well above that

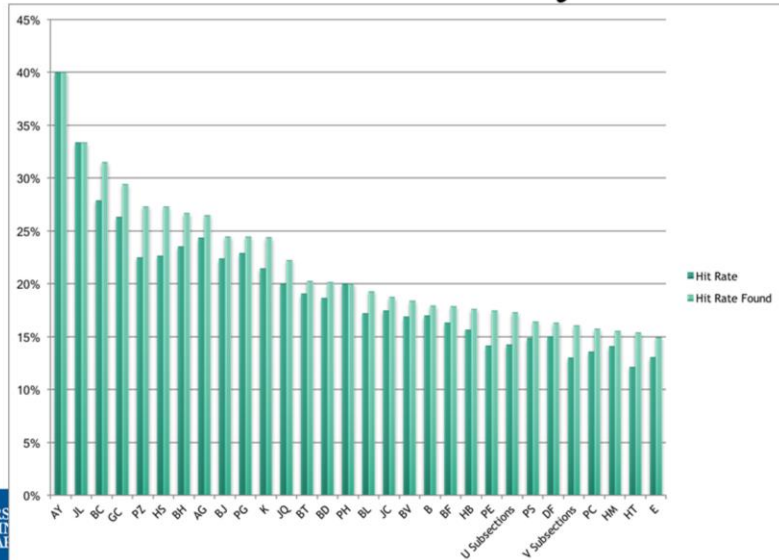
[If you are looking at this chart and trying to understand why there are two bars—hit rate found is the percentage of items that met our threshold of the books we found—some were checked out or missing, so the general hit rate number compared the number found to the total number of books the Library should have from that call number, as opposed to generating the percentage based on the number of books found. Hit rate found, will therefore be a higher percentage.

The K call numbers are for law--

Law is mostly US law/state law, but also Roman, Ireland and New Zealand

I always suspected from my own area of research that the Bs would prove to have many interventions, and we see that playing out in area of theology, ethics, philosophy and logic. The other call numbers are a random smattering of everything from secret societies to Slavic languages—and their high hit rates may be anomalies specific to our collection rather than general trends—we'd have to do more studies.

Highest % of Interventions: Alderman Only

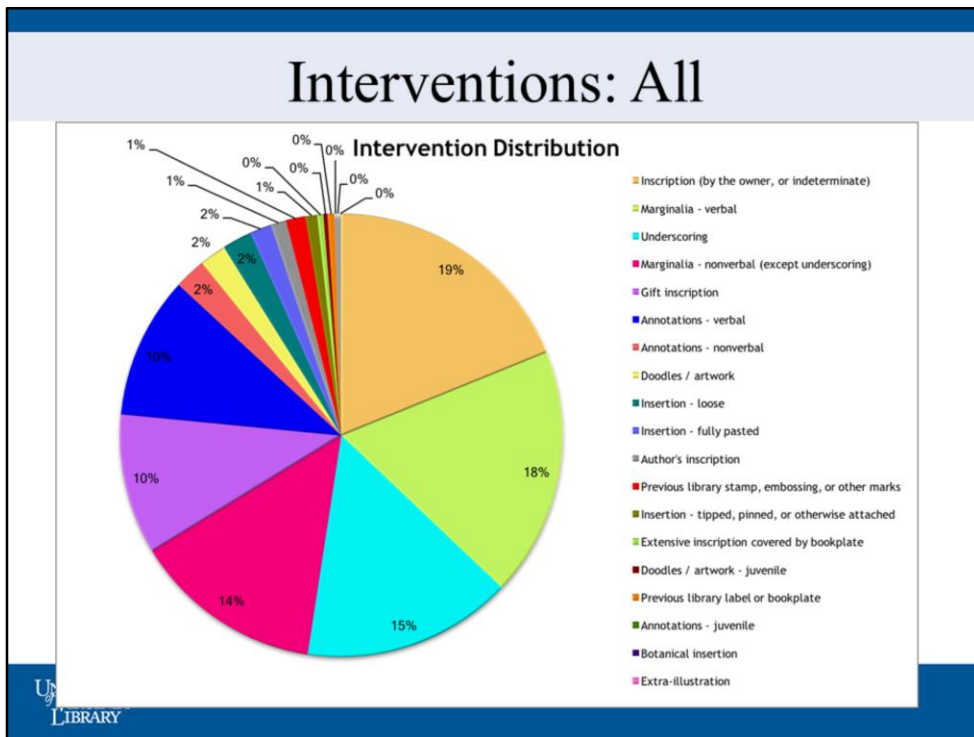


Because law skewed our numbers so much, I asked our statistician to pull data just from our main social science and humanities collection—

This reveals high rates in several areas of literature

Law and many of the same philosophy and religious subsections found in the other chart

And several areas of social sciences, like sociology and economic theory



What interventions did we find the most of?

Inscriptions, marginalia (which relate to the text) and annotations (which do not) were the most common interventions. Items deliberately or accidentally inserted into the volumes, doodles, and artworks were far less common

Some volumes had more than one intervention so this chart is a little confusing—this chart says, of the interventions found, 19% of them related were inscriptions, rather than 19% of the volumes HAD inscriptions

Note: our categories were a little broader than the Provenance Evidence Thesaurus—we will collapse them back down when we add them to the catalog, but we may end up making suggestions about expanding their definitions...

Lessons Learned



- Workflows much more efficient in open stacks
- Assess books *before transfer to off-site storage*

An important and relevant lesson we have learned is that it is considerably easier to survey books in bulk while they sit on open shelves in our main library. Calling books from U.Va.'s off-site storage facility was much more labor intensive for the facility staff, and we were unable to pull as many as we desired, even using a sampling method. The lesson learned here is that the ideal time to do this work is **BEFORE the items move off site. Trying to accomplish it after it a much more expensive endeavor.**

Implications

- Preliminary studies suggest UVA is not unique in its percentage of unique volumes in circulating stacks
- Once item is withdrawn, it is gone forever
- Do not simply rely on catalog-guided deduplication methods
- Consider keeping items that meet thresholds for uniqueness in addition to “clean copies”

UNIVERSITY
of VIRGINIA
LIBRARY

So, why am I telling you all this? Well, we have done some sampling at other institutions and our data suggests that UVA is not unique in its significant numbers of volumes that contain unique features.

As items move off site and the potential for serendipitous discovery is lost, it is even more important that items be well described. As my co-PI, English Professor Andrew Stauffer points out, the books he depends upon for his research of 19th Century reading practices— are out of copyright, freely available as page images online, and often in poor condition due to acidic paper – They are also the ones most at risk of being deaccessioned en masse by libraries. Once an item is withdrawn, it is gone forever.

I am here today to ask those of you who have shared print collections, especially those just starting out, to consider seeking out and keeping items that meet whatever your consortial definition of uniqueness is in addition to keeping “clean copies” for researchers. That rather than relying on the catalog record to tell you that two items are alike, that you figure out a way to take the time to eyeball items before deaccessioning them.

How to Get Started

- Develop own/similar criteria
- Survey all/sample
- Determine method for creating “easy” metadata
 - Check box
 - Provenance Evidence Thesaurus
 - Develop own controlled vocabulary



How might you get started?

First—you need to decide what your consortial group considers unique enough to retain. You could use our methods or develop your own.

Next, consider whether you will survey every single book or Target specific areas

You could do statistically significant sampling

Use results from UVA and focus on call number ranges with high hit rates

When we did sample of a few British authors, we found high numbers of interventions at several institutions—it is reasonable to believe the areas like British and American literature would result in high hits in most academic research collections. While our numbers may have had some anomalies, it would be a good jumping off point if surveying the entire collection was not feasible

Focus on your consortial/institutional areas of strength

Finally, determine how you will indicate that the item has “unique features.”

A simple but not very detailed option would be to just add a check box to your validation process—check if there are unique features. You need not describe what they are; just indicate that this object should be retained. While not as useful to the researcher, it might be a way to quickly and efficiently protect items with unique content. It also would eliminate the need to gain buy in for controlled vocabulary should we get to the point where we can share this data across institutions—a shared portal is a future desire of the Book Traces project.

A slightly more time intensive process would be to use the PEV, which would allow your efforts to be combined with ours in away that provided a bit more information to the researchers.

And finally, you could develop your own process, with the knowledge that merging your metadata with ours might prove more difficult later on.

Questions?

Kara M. McClurken
kara.mcclurken@gmail.com

<https://booktraces.library.virginia.edu/>

Twitter: @booktracesuva

Special thanks to:

Kristin Jensen

Jackie Morrogh

UNIVERSITY
of VIRGINIA
LIBRARY

I would love to talk with folks afterwords if you are interested in incorporating these ideas into your workflows.

Thanks.