#s702
Cultivating your Archival Garden

Cultivating your Archival Garden

digging into data for assessment and programmatic improvement

Megan Mummey, University of Kentucky
Hannah Pryor, University of Louisville
Morgen MacIntosh Hodgetts, DePaul University

MAC 2024

Hello everyone! I'm Megan Mummey and I'm the Director of Manuscript Collections at the University of Kentucky Libraries. Today, myself, Hannah Pryor of UofL, and Morgen MacIntosh Hodgetts of DePaul University will be talking to you about everything data-related in our presentation: Cultivating your Archival Garden: digging into data for assessment and programmatic improvement! We'll be talking about different kinds of data, ways to gather it, and how to use it. We'll each discuss a project framed around using data to guide and improve our work.

Processing Priorities, Researcher Use, and Programmatic Improvement

MAC 2024

Megan Mummey, Director of Manuscript Collection, University of Kentucky Libraries

From the Louis Edward Nollau F Series Photographic Print Collection

Today I'm going to talk about a data analysis project I've been working on since 2018. In short, I've been using the data generated by our research requesting system to make decisions and set priorities for the archival processing of manuscript collections. And let me tell you I find this process to be one of the single most enjoyable things that I do. A. I love spreadsheets and structured data! B. ask any processing archivist and they will tell you that there is nothing like the thrill (the absolute high) of seeing someone use a collection you've poured your blood, sweat, and tears into (not literally).

How are processing priorities set?

- Subjectively
- "Problem in front of you"
- Processing on demand
- As acquired (MPLP)
- Donor needs
- Use based

One of my "problems": 36 ft collection found in storage - unprocessed with no accession record - acquired in 1978

So I first want to talk about how processing priorities are set. There are many different methodologies for this and in a nutshell most of the time it is incredibly subjective. Some of it is "the problem in front of you", like this photograph of a problem I found in storage. You know…just a small 36 cubic foot collection with no documentation. We've had it since 1978. You can do "processing on demand" (so when a researcher wants access to an unprocessed collection you minimally process it before they arrive). You can do it in a "as collections are acquired approach." And still sometimes you have to drop everything because of donor relations. Not all of these methodologies are wrong or necessarily bad, some are better than others. Oftentimes the way we determine our priorities at UK takes a little bit from all of these things.

Context

- UK SCRC fully implemented AEON in 2016
- In 2018 (with 18 months of data) began analyzing check-out data
- Have done this process once a year on use data from 2016-2022
- Never had a true backlog busting project or made accession information available
- Many requests for collections in our backlog

UKL SCRC Research Room

But in 2016 when my institution implemented AEON – I had a question –what would happen if we looked at the data AEON collects and have it inform our processing decision making. What would the data tell us? This was the first time that my institution had any reliable system for tracking collection usage in our research room. Before it – everything was paper-based, and nothing was done with those paper slips. AEON collects a lot of data! Which makes it at times challenging to use said data.

So, in 2018 with about 18 months of well-formed, consistent data – I began analyzing which boxes and collections were checked out in our research room. In Spring 2018, I gave a presentation as a part of a panel at the National AEON symposium about my process, the results, and future directions. This presentation is essentially an update of that presentation with data from 2016-2022. (I have not had time to really dig into 2023 yet).

Other bits of context you might need to know is that we've never had a true backlog busting project or made our accession information publicly available. Now – we do get a lot of requests for unprocessed collections in our backlog though. Because of past management decisions there are a lot of citations floating around in the scholarship for collections in our backlog.

Process

- AEON Transaction Report - Frequency
- Exclude staff check-outs
- Add in collection titles
- Clean up data
- Check online status
- Code collections - subject, collecting area, formats
- Analyze data and generate visualizations
- Help set priorities

	A Accession	B Title	C Number of uses	D Online	E Topic	F Topic 2
1	Accession	Title	Number of uses	Online	Topic	Topic 2
2	91M2	Ann and Harry Caudill papers	60	yes	Authors	Appalachia
3	72M2	Frederick Moore Vinson papers	51	yes	Judiciary	Public Policy
4	63M349	Wickliffe Preston family papers	50	yes	Kentucky -- History	Lexington
5	49w1	Samuel Wilson vertical file collection	48	no	Kentucky -- History	Lexington
6	1979ua003	Otis Singletary papers	42	no	UK President record	University of Kentu
7	2001ua062	Margaret Lantis papers	34	yes	Faculty papers	Public Health
8	2011ms055	John Arthur Dearinger papers	32	no	Theater	
9	0000ua007	John W. Oswald papers	31	no	UK President record	University of Kentu
10	1997MS271	Altrusa International of Lexington records	31	yes	Education	Non-Profit
11	2007ua023	Athletic Players Files	27	yes	UK Athletics	University of Kentu
12	76M2, 82M6, 84M	Earle Clements papers	26	yes	Public Policy	
13	1997MS213	Hillenmeyer family papers	25	yes	Kentucky -- History	Lexington
14	85m1	Frontier Nursing Service records	24	yes	Appalachia	Public Health
15	81M3	Stanley Forman Reed papers	23	yes	Judiciary	Public Policy
16	2009ms132.0081	Wade Hall: Jack and Norah Parker letters	23	no	WWI	Military history
17	63M202	Hunt-Morgan family papers	22	yes	Kentucky -- History	Lexington
18	2004ua046	UK photographic services negatives	21	no	University of Kentu	Photographs
19	1997MS282	John D. Whisman papers	21	yes	Appalachia	
20	2005ua015	Joseph Baber papers	20	no	Faculty papers	University of Kentu
21	2009MS172	Pettit, Duncan, Gibson family papers	20	yes	Kentucky -- History	Suffrage
22	2009ms132 0004	Wade Hall: George Canary letters	20	no	WWII	Military history
23	1997ms373	Episcopal Dioces of Lexington records	19	no	Lexington	Religion
24	ua	UK general reference files	18	no	University of Kentu	Reference files
25	2010MS041	James Clay family papers	18	yes	Clay family	Kentucky -- History
26	61M158	Linda Neville papers	18	yes	Appalachia	Public Health
27	80m1	John Sherman Cooper	18	yes	Public Policy	
28	90M1	Appalachian Regional Commission record	17	yes	Appalachia	
29	2004av001	John C. Wyatte Herald Leader photograph	17	no	Photographs	Lexington
30	82M9	John Jacob Niles papers	16	yes	Appalachia	Music
31	2014ms0041	Kentucky Hemp Growers Association and	15	yes	Hemp	Public Policy
32	46M53	John Winston Coleman, Jr. collection on s	14	yes	Slavery	African-Americans
33	xx	WWII posters	14	no	WWII	Military history
34	58M25	Cora Wilson Stewart papers	14	yes	Appalachia	Women
35	77M1	Happy Chandler papers	13	yes	Public Policy	Civil Rights
36	2009ms132.0084V	Wade Hall: Dessie Gum Sharp papers	12	no	WWII	WWI
37	2009ms043	E. I. "Buddy" Thompson papers	12	no	Women	University of Kentu
38	96M8	Frontier Nursing Service medical surveys	12	yes	Appalachia	Public Health
39	1997ms295	Laura Massie papers	12	yes	Civil Rights	African-Americans
40	0000UA107	Margaret Ingels papers	11	yes	Faculty papers	STEM
41	0000ua197	UK Training of the Fighting Mechanic reco	11	yes	WWI	Military history
42	2004ua015	UK Public Information Audio Collection	11	yes	University of Kentu	A/V

My data process is relatively simple, and only uses excel, but of course it involves cleanup and takes time. AEON spits out a frequency report that gives me a spreadsheet of collection identifiers and numbers of box checkouts. Staff check-outs are excluded from this report. Unfortunately, this report does not differentiate between books and collections, so those have to be weeded out. Also the identifiers are only as good as the data entry. Sometimes there is some noise, junk identifiers, identifiers that are essentially the same but entered differently (you know normal human error; capitals vs. lowercase; extra spaces). So there's data cleanup that I perform as well as adding collection titles to be more human readable than just the collection identifier. I also check whether these collections have an online guide including whether it has a full box listing vs. just a collection-level guide. Over the years, I've added some data points such as year the collection guide went online and what program they're a part of (these are totally artificial, but can be helpful – so manuscripts, university archives, and audiovisual archives). Once my data is well-formed, I start coding the collections by general topic. I've developed a controlled vocabulary for this process, that I usually add to every year. I try to keep the coding consistent from year to year, so I look at how I've coded the collections in past years. I use my knowledge of the collections to assign between 1 and 6 topics. These can be collecting areas (like Kentucky history, Appalachia, public policy), subjects like specific wars or subject areas like medicine, STEM, suffrage, etc. and sometimes formats microfilm, photographs etc. I compile the number of box checkouts and the number of collections used for each subject. Then I look at the data, analyze it, generate some visualizations, and draw some conclusions to help set future priorities.

2016-2017 Research Room box Check-outs by Subject
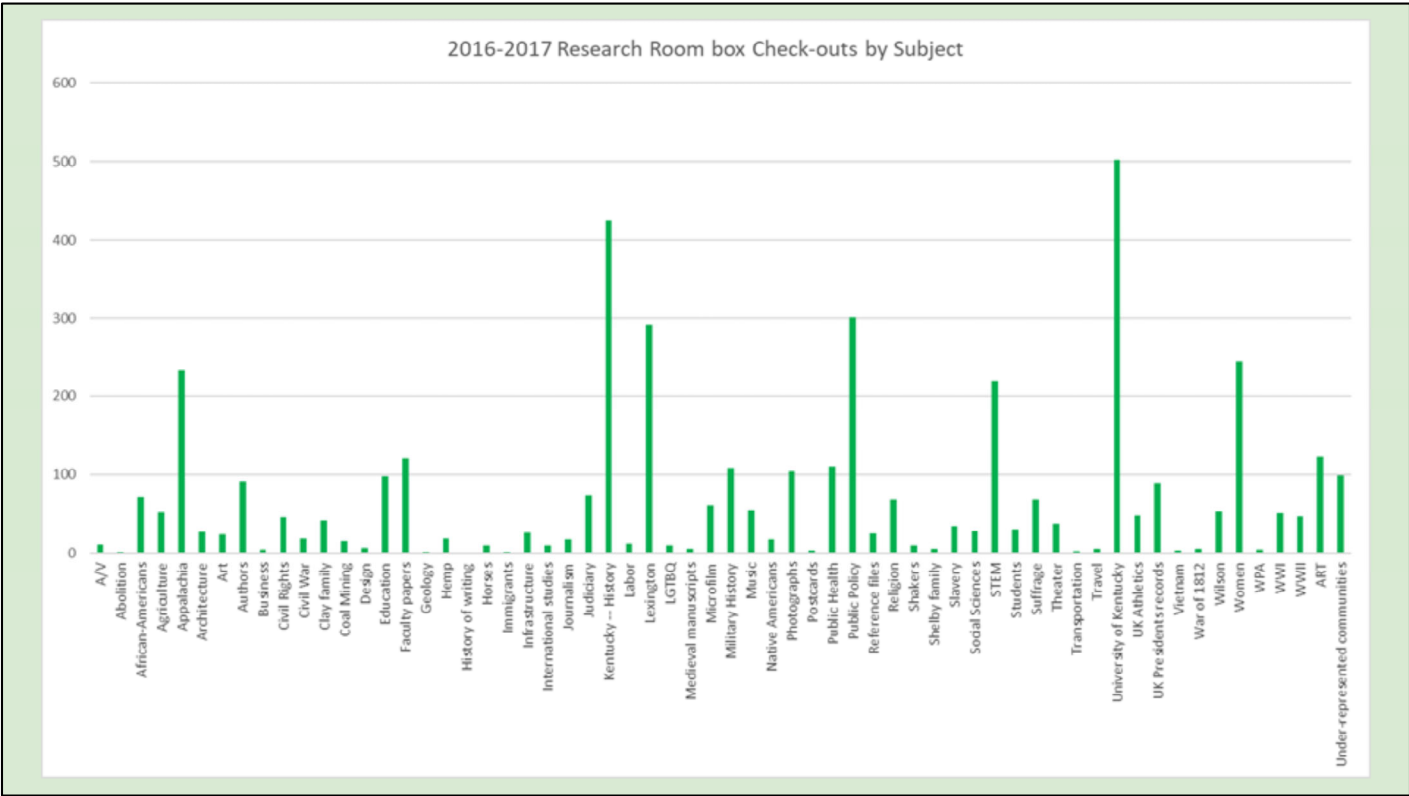
One of the visualizations I generate is a count of each coded term for both collections and box checkouts. This bar graph represents the subject-based topics by box checkouts. This is the first subject term graph I generated from the 2016/2017 data. These largest spikes here, not surprisingly, correspond with our largest collecting areas: Appalachia, University of Kentucky, Kentucky history, Politics and Public Policy.

2022 Research Room Box Check-outs by Subject
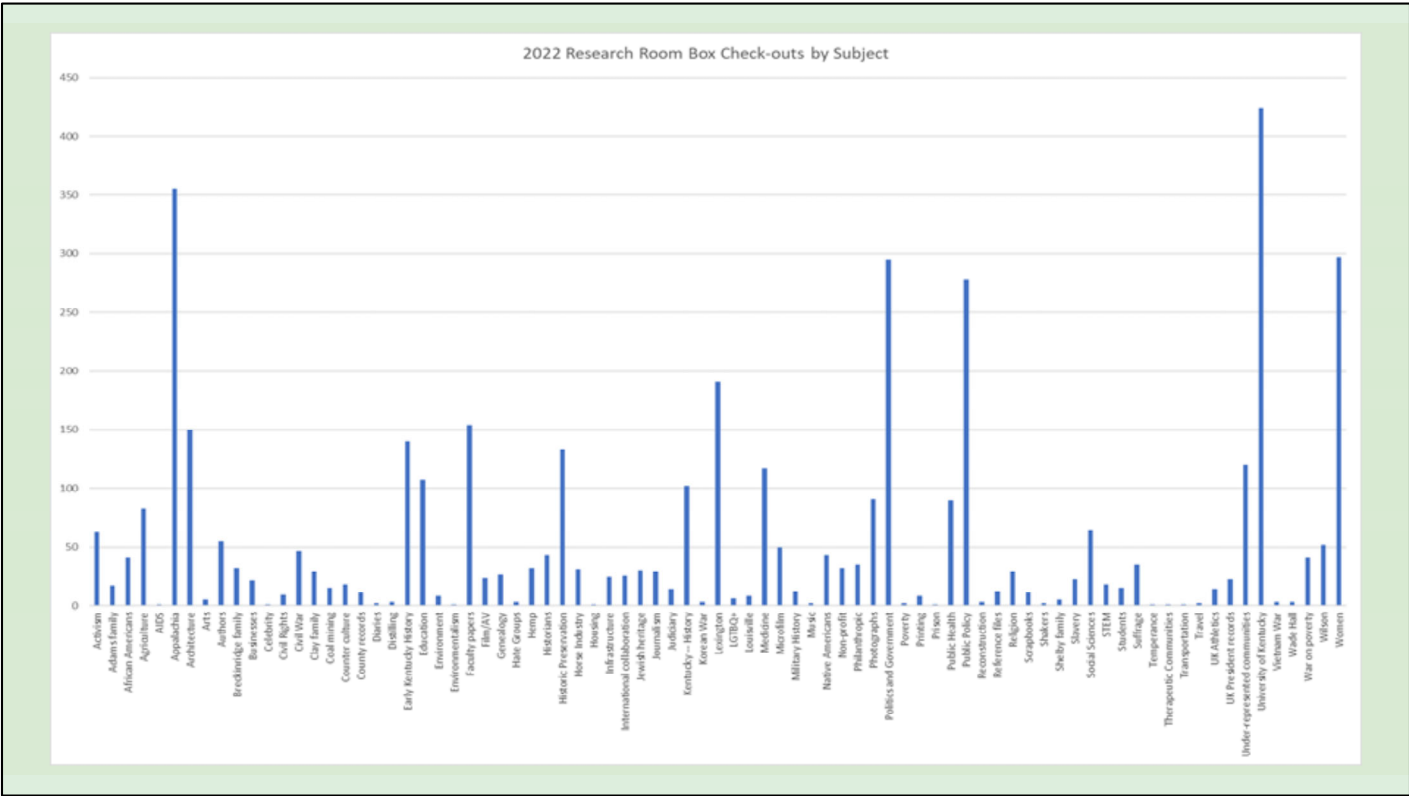
Here is the latest graph I have of data from 2022. As you can see – the large spikes still correspond and haven't changed much. We still do a lot with Kentucky history, Appalachia, politics and public policy, and UK. We're always going to prioritize our main collecting areas when we process. But what I'm really interested in with these graphs – is these smaller categories, which rise and fall in interesting ways from year to year. I find this is where the most meaningful analysis lives and where I can draw conclusions of what to process

Programmatic Improvement

- Set processing priorities for new collections and the backlog
- Helps identify problem areas:
  - Legacy collections in need of better access
  - High-use collections without an online finding aid
- Helps identify collections for digitization
- Has implications for collection development
- Can be used for advocate for high use areas in need of support

I use this process to make some decisions to set the direction for the manuscripts processing program. The subject analysis helps me set priorities for backlog processing as well as for new material coming in. (Granted I use this analysis along with some of the considerations I discussed on the first slide – donor promises, responding to patron requests, etc.) This process has also helped identify legacy collections that are highly used and in need of better access as well as high-use collections that for whatever reason do not have an online collection guide. It can also help us decide whether to digitize a high-use collection for better access and better preservation (you know fewer people touching delicate things is always good). Looking at micro-trends in research at our own institution can also inform collection development decisions about whether or not to take a collection on a certain topic or how to aim our purchasing funds for the year. The data can also be used for internal advocacy, so if there is an area, program, etc. that has high use and low access I can use this data to back up requests for support.
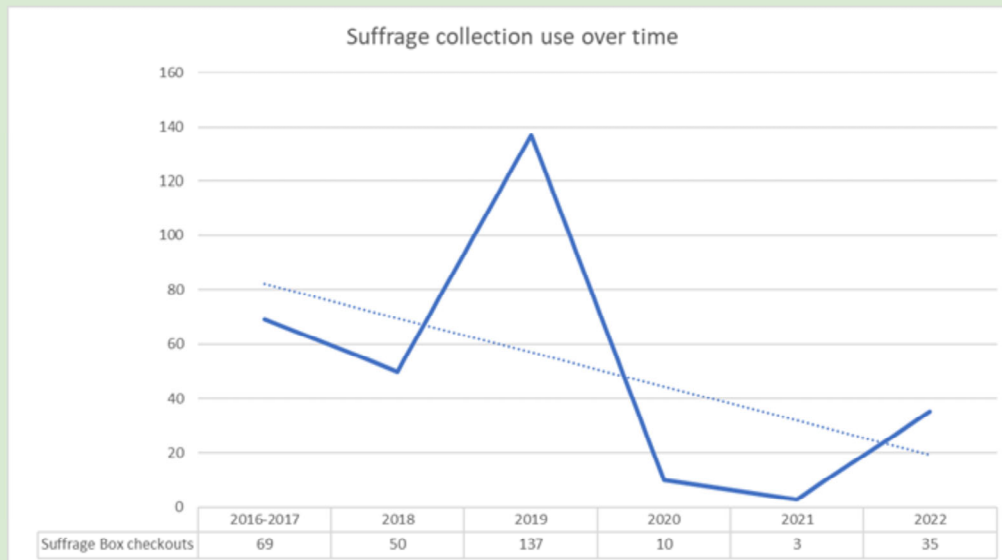
So now instead of doing a deep (and possible boring) dive into this body of data – I want to use my remaining time to talk about a few real-world examples of how we've used this data to make decisions. This first example is about identifying research trends to inform processing and digitization. This graph shows the rise and fall in use for suffrage related collections. Starting in 2016/2017 the in-person use of our collections about suffrage increased rapidly. This was of course due to an increase in scholarship about suffrage surrounding the various anniversaries of the 19th amendment in 1919-1920 (it was passed in 1919 and completely ratified by the states in 1920). The box checkouts shown in this graph peaked in 2019 and cratered out in 2021. I think if not for the pandemic the use in 2020 would still have been high. We worked hard in late 2018/early 2019 to process, improve description for collections dealing with women's suffrage. Additionally, we digitized some highly used collections that were starting to experience wear and tear from repeated use. These images came in handy when we were all sent home in 2020 during the pandemic.

Examples - Finding Aid Access Project

Collections checked out by online status 2016-2017

54%
15%
31%

finding aid   collection-level   none
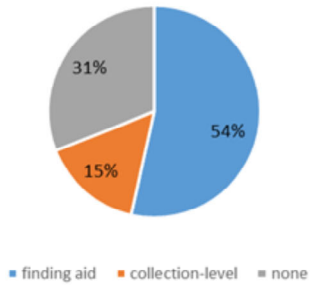
- From 2016-2018 - ⅓ of collections checked out were not online
- Represents a lot of work for research desk staff and subject specialist staff
- In 2019 ran a project to address this:
  - Identified 110 collections in this gap
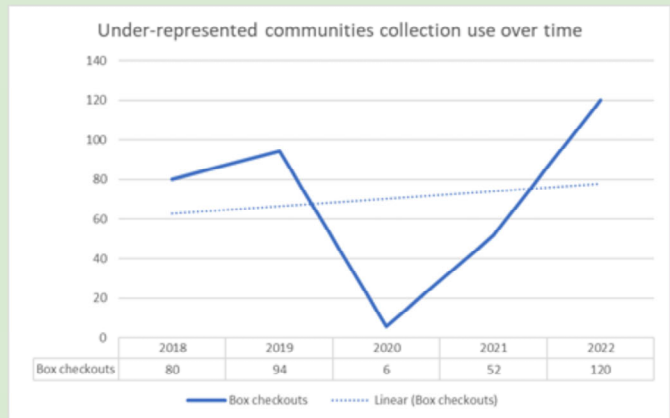 - 68 have gone online

So when I first did this process with the 2016-2017 data – I noted that a third of the collections checked out in our research room had no online access. Remember I mentioned how there are citations for backlog collections out there… We have to generate pdfs of collection guides in progress (if they exist) and send them to researchers, we have to assess these collections to see if they can even be served to patrons (are they stable? Are they in good enough boxes? Are they in any order at all?). Also, if they're not online – then patrons cannot help themselves by finding it. This represents a lot more work for our research room staff as well as our processing staff, and that work has to happen in the moment. In 2019 – I ran a project to identify some "abandoned" collection guides in our Manuscripts holdings, categorize them by difficulty to complete, resolve their problems, and put them online. I identified 110 collections in this gap. 68 have now gone online. (53 were categorized as easy; 9 medium; and 6 hard). We're still working on some of the ones identified as hard and medium, but we're definitely still making progress on them!

Examples - Find Aid Access Project

Collection checkout online status over time
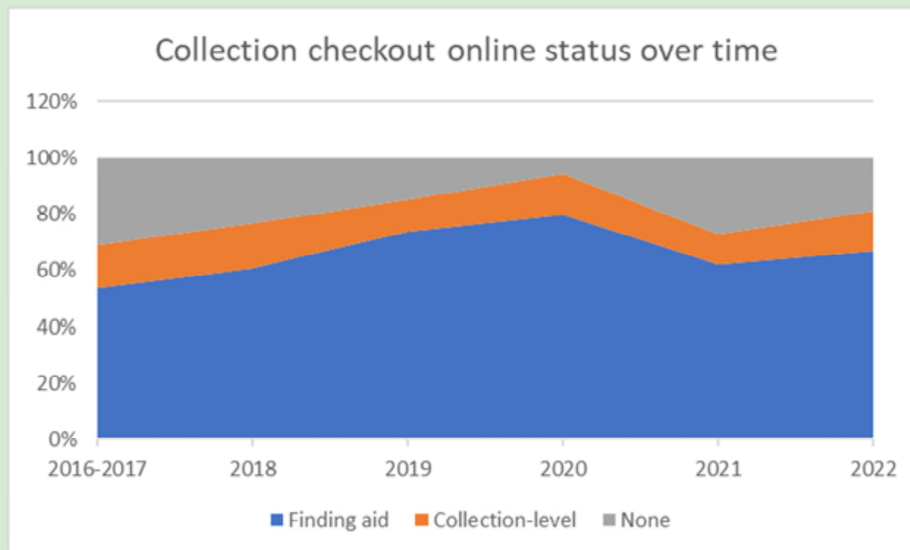
By 2020 you can see that these use numbers drastically improved. Use of collections in our research room that have no online access declined 15% from 2016-2019! Which is awesome. It's been climbing back up again – But that's just another opportunity to push some collections online.

Examples - Advocacy

Faculty papers collection use over time

	2018	2019	2020	2021	2022
Box checkouts	36	119	10	55	154

Under-represented communities collection use over time

	2018	2019	2020	2021	2022
Box checkouts	80	94	6	52	120

You can also use this data to advocate for certain types of collections! This graph shows the impact of faculty papers (something my institution is so-so on collecting). But they're actually used pretty heavily as long as the content is of high quality. I also have a graph here showing how often collections documenting under-represented communities are used. I'm happy to report that use is growing! This is a legitimate research area of interest. And we should continue to prioritize and process them because people want to use them! And…I had to put in these trend lines because of the pandemic…my numbers crater in 2020 and 2021! The pandemic ruined so many things, but it also ruined my dataset!

Examples - Bad Data Points

Checkouts by year EAD online 2022
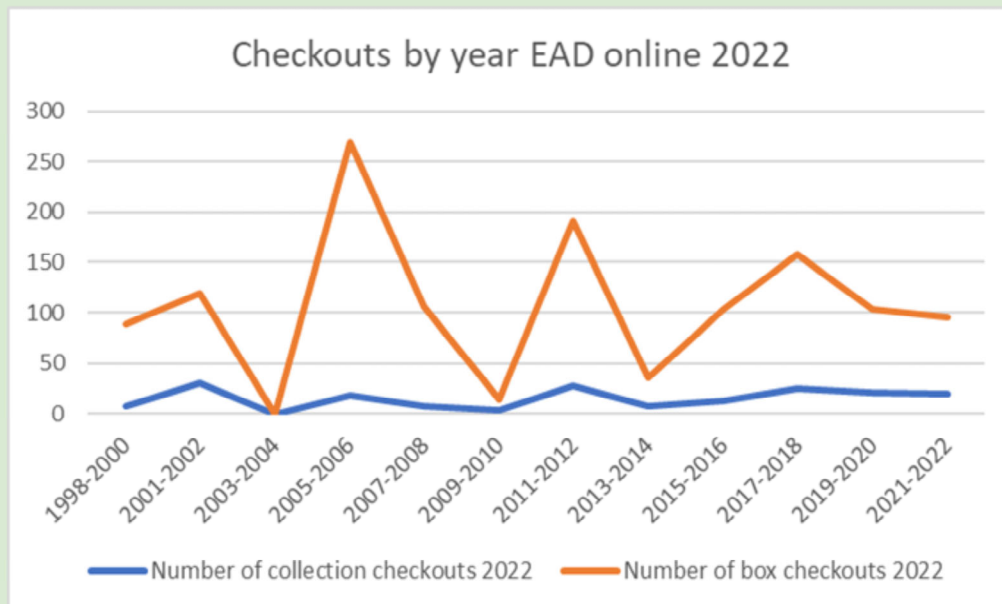
Now I want to take a minute to discuss failures. Not every data point I have investigated over the years….has panned out. I decided to look at the checkouts and numbers of collections used by when their finding aid went online. My results were not great. These spikes in 2005-2006 actually show when some of our biggest, most well-known, and highly used collections went online – like the frontier nursing service…this spike is when we had a project to put legacy finding aids online in 2011/2012. This downturn in 2009/2010 is when our new digital library was being built and we couldn't put anything online. I was hoping this data would show the longer something has been online the greater the use…but right now it just points out years we put a lot of finding aids or extra large finding aids online!

Examples - Bad Methods

Term	Count
kentucky	557
ky	192
states	170
united	163
lexington	159
history	156
county	97
john	94
war	88
family	85
clay	76
william	76
century	62
james	62
women	61
university	58
american	57
records	52
henry	46
1865	45
appalachian	43
1861	39
robert	39
correspondence	38
1945	37
african	37
world	37
20th	36
company	36
government	36
politics	36
civil	35
region	35
court	34

Another investigation that went nowhere…extracting and analyzing the subject headings from finding aids for collections used in our Research Room. This was a suggestion from someone who saw my 2018 presentation. I created an xslt to batch extract our library of congress subject terms from individual EAD. Then I ran the corpus through Voyant tools (which is a web based text analysis tool) and was extremely disappointed.

Examples - Bad Methods

Worst word cloud ever

Term	Count
kentucky	557
ky	192
states	170
united	163
lexington	159
history	156
county	97
john	94
war	88
family	85
clay	76
william	76
century	62
james	62
women	61
university	58
american	57
records	52
henry	46
1865	45
appalachian	43
1861	39
robert	39
correspondence	38
1945	37
african	37
world	37
20th	36
company	36
government	36
politics	36
civil	35
region	35
court	34

It created basically the worst wordcloud ever! 46 henrys…62 james…170 states.

Examples - Bad Methods

Term	Count	Length
shelby isaac 1750 1826 shelby james 1784 1848 sh	3	13
breckinridge robert j robert jefferson 1800 1871	2	12
alexander robert spreul crawford aitcheson 1819	2	10
church and state united states history 20th cent	2	8
marshall george s marshall silas norchutt joseph	2	8
aspects appalachian region coal mines and minir	3	7
united states history civil war 1861 1865	9	7
african american fraternal organizations kentuck	2	6
agricultural and mechanical college of kentucky	2	6
aspects kentucky coal mines and mining	2	6
coleman j winston john winston 1898	2	6
first african baptist church lexington ky	2	6
lincoln grant high school covington ky	3	6
african american baptists kentucky lexington	2	5
african american musicians kentucky lexington	2	5
baptist church fayette county ky	2	5
birney james gillespie 1792 185	2	5
black james dixon 1849 1938	2	5
bullock waller overton 1875 1953	2	5
church records and registers kentucky	2	5
fouse william henry 1868 1944	2	5
history civil war 1861 1865	6	5
jewell robert berry 1896 1986	2	5
kentucky lexington women college students	2	5
kinkead george blackburn 1849 1940	2	5
may andrew jackson 1875 1959	2	5
mcdowell henry c henry clay	2	5
morrow edwin porch 1877 1935	2	5
politics and government 1861 1865	2	5
siebert wilbur henry 1866 1961	3	5
u.s university of kentucky chapter	2	5
underwood warner lewis 1808 1872	2	5
united states history 20th century	4	5
united states politics and government	2	5

It's just as bad when it looked at the phrases rather than the individual words. The whole process was time intensive. And to draw any conclusions from it – I would need more programming skills and text mining knowledge than I currently have.

Conclusions

- Data is messy; you have to clean it up!
- Data cleanup can take a long time
- Look for data your institution may already be collecting
- Sometimes your data points mean nothing
- Use data can help point you in the right direction
- Results
  - More finding aids online!
  - More resources
  - Can give direction when faced with huge backlogs

A Suffragette Primer, 1911, from the Laura Clay papers

In conclusion I have some takeaways and reflections on my results. Data is messy – you gotta clean it up! Data cleanup can take a long time. Look for data your institution may already be collecting. Sometimes your data points mean nothing. But collection use data can help point you in the right direction! Because of this process we've put more finding aids online! Ones that are used! I can advocate for more resources in a high use area because I have numbers. Administrators love numbers. There's never a shortage of things to process or improve access to – but looking at the data can give you direction when faced with a huge backlog! And I'll repeat it again I find it so affirming to see that people are using our collections – it always makes me feel optimistic and positive.

UNIVERSITY RECORDS DATA ASSESSMENT

Hannah Pryor, Archivist for University Records and Records Manager
University of Louisville
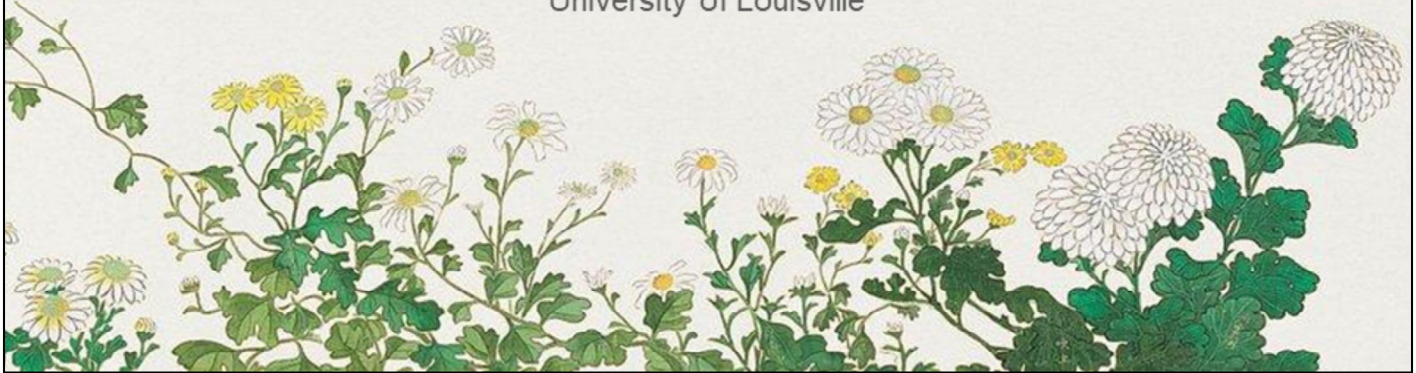
Hello and thank you. Introduction. Education, history, certifications.

OUTLINE

I. Institutional Context
II. Key Questions
III. The Process of Turning ArchivesSpace Data into Usable Information
IV. Main Challenges
V. Findings
VI. Conclusions

Hannah Pryor, University of Louisville | 2

Quick outline overview

UofL Archives and Special Collections

- Consists of University Archives and Records Center (UARC), Photographic Archives, Rare Books, Oral History Center, and Digital Initiatives
- UARC founded in 1973
- Records management program serves university community of 13 academic units, administration, 23,000 students, and 7,100 employees
- Around 11,000 linear feet of university records

https://library.louisville.edu/archives/home

Hannah Pryor, University of Louisville | 3

Team: archives director, myself, and archives assistant

KEY QUESTIONS

1. How can we visualize and assess over 100 years of university records?

2. How do we identify gaps?

3. How do I balance conducting this assessment with other job responsibilities?

4. How can we prioritize outreach in the future, based on our findings?

Hannah Pryor, University of Louisville | 4

Why I started this project: encountered gaps in records during a reference request one year into the job, wanted to get a better birds eye view of collections. What I wanted to accomplish: complete an assessment to drive future decision surrounding outreach and reappraisal. Key questions:

THE PROCESS

- Exported .csv report of university records information from ArchivesSpace
- Filtered out non-permanent records and oddities
- Cleanup, cleanup, cleanup!
- Analysis
- Planned next steps

Image credit: "Registration, UofL, 1980s," UofL Digital Collections, ULUA 001_0465
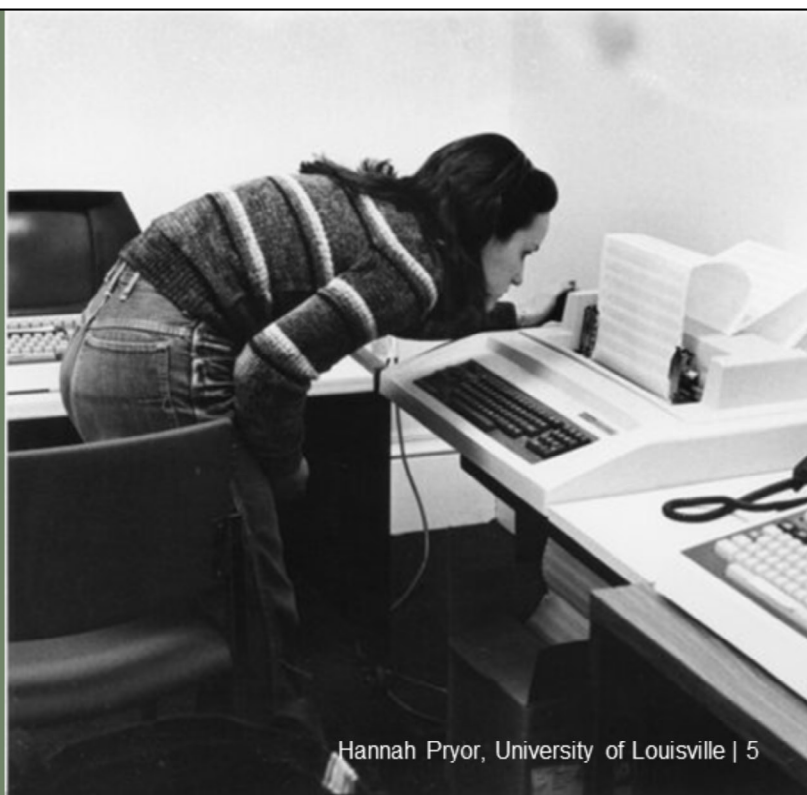
Hannah Pryor, University of Louisville | 5

Overview of my process. Little background: we use ASpace to accession university records.

MAIN CHALLENGES

- Data cleanup
- Data visualization
- Time constraints

Image credit: "Paul Blake cutting field with horses,"
UofL Digital Collections, ULPA SONJ_15325

Hannah Pryor, University of Louisville | 6

Overview of main challenges

This is the step that took the longest. Once I had my .csv: Coding qualitative data that I wanted to capture such as unit, subunit, and record type. Inconsistent unit names (e.g. School of Business vs. College of Business, President's Office vs. Office of the President). Unit names that merged and changed over time. To identify blank fields, improperly formatted dates, etc I explored using tools like OpenRefine, but most of it was manual clean up of around 2500 rows. Iterative, pick up and put back down, but important to know when to stop.

DATA VISUALIZATION

- Exploring options—Power BI? Pivot tables? Airtable? Libapps?
- Getting data do what you want it to do
- Visualizing gaps in date ranges
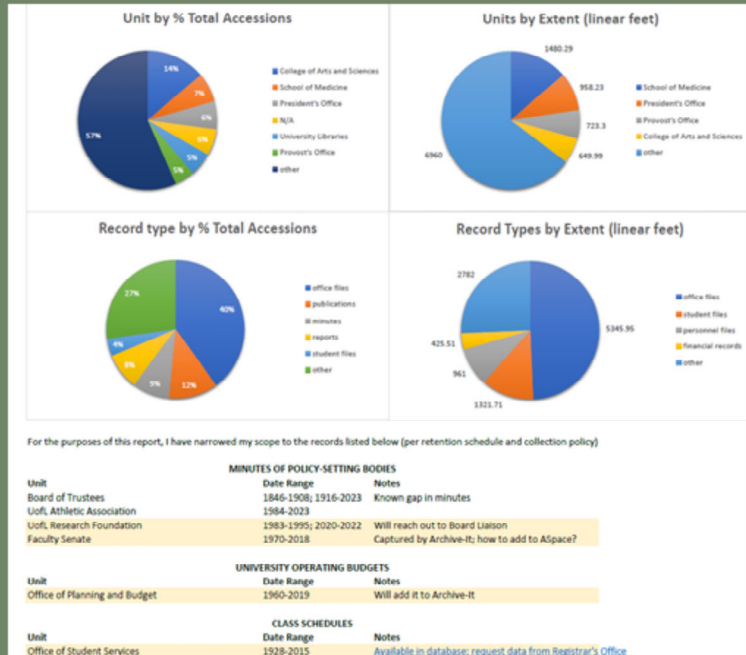
Hannah Pryor, University of Louisville | 8

Knew what I wanted, but it was difficult to make it happen. A million solutions, but I didn't have time to learn how to use them. For example: wanted to use a scatter plot graph to visualize accessions over time, but it didn't happen (and that's okay). Decided to visualize big points and use basic text fields to list record groups that I want to focus on for outreach.

MANAGING TIME CONSTRAINTS

- Multiple projects at once
- Detours into exploring different solutions
- Coping methods

I STARTED THE DAY WITH LOTS OF PROBLEMS.

BUT NOW, AFTER HOURS AND HOURS OF WORK,

I HAVE LOTS OF PROBLEMS IN A *SPREADSHEET.*

Image credit: https://xkcd.com/1906

Hannah Pryor, University of Louisville | 9

Same old story. Sometimes, finding new solutions just served as a distraction when I had decent tools already at hand. Utilized project management techniques to stay on track while juggling other responsibilities.

FINDINGS

1   The nearly unlimited potential of data vs. my personal limitations (time and ability)

2   Adjusting expectations for outcomes (shift gears from prioritizing outreach to balancing it with processing and reappraisal)

3   Created plan for outreach to governing bodies with gaps in records and to underrepresented student and employee groups

CONCLUSIONS

- Create a plan, but stay flexible
- Look at what the patterns of data are telling you, rather than try to fit your preconceived notions onto them
- Become comfortable with the imperfect—data is a tool to help you make decisions, not an end product

HELPFUL RESOURCES

- OpenRefine
- Project management software (e.g. Trello, Asana, Confluence)
- *Data-Driven Decisions : A Practical Toolkit for Library and Information Professionals* by Amy Stubbing (2022)

MY CONTACT INFO

hannah.pryor@louisville.edu

Hannah Pryor, University of Louisville | 11

Seeds of Change:

Designing and Implementing a Collection Survey

Morgen MacIntosh Hodgetts
Senior Archivist and Coordinator of Archival Operations
MAC, Des Moines, Iowa: Saturday, May 4, 2024
mmacinto@depaul.edu 773-325-5081

DEPAUL UNIVERSITY
UNIVERSITY LIBRARY
Special Collections and Archives

Seeds of Wisdom - Appraisal

"What were we thinking when we took this in?"

~ Mark Greene, 2002~

"All well managed archives need periodic cleanups."

~Elena S. Danielson, 2010~

"The reappraisal process is systematic, nuanced, and time-consuming."

~SAA Guidelines for Reappraisal and Deaccessioning, 2012~

MMH Slide 1 (presentation slide 31) - I am not a person who can be described as 'trendy' or 'fashionable' so when I heard Rachel Christine Woody, archives and museum consult and founder of Relicura, describe deaccessioning as a "hot-topic" and "in-vogue" in two of her recent webinar videos, I have to admit to feeling a little bit pleased. Due to a recent restructure of the positions in our department, last year the pieces fell into place for me to design and implement a survey of our collections. Part of the design process included revisiting materials from an SAA Reappraisal and Deaccessioning workshop that I attended in 2017, reviewing notes from my 1998 archives class taught by Northwestern University Archivist, Patrick Quinn, and rereading some classic and more recent articles and case studies about archives appraisal. In my 23 years at DePaul, I know that we have accepted and processed collections without performing what Patrick referred to as the "most significant and important" archival function – appraisal. Which naturally means we have entire collections or portions of collections that are candidates for deaccessioning and suddenly I find myself (perhaps the least cutting-edge person in the department) leading a project that is "in-vogue." On this slide are a few seeds of wisdom about the topic…

Seeds of an Idea - Collections Survey
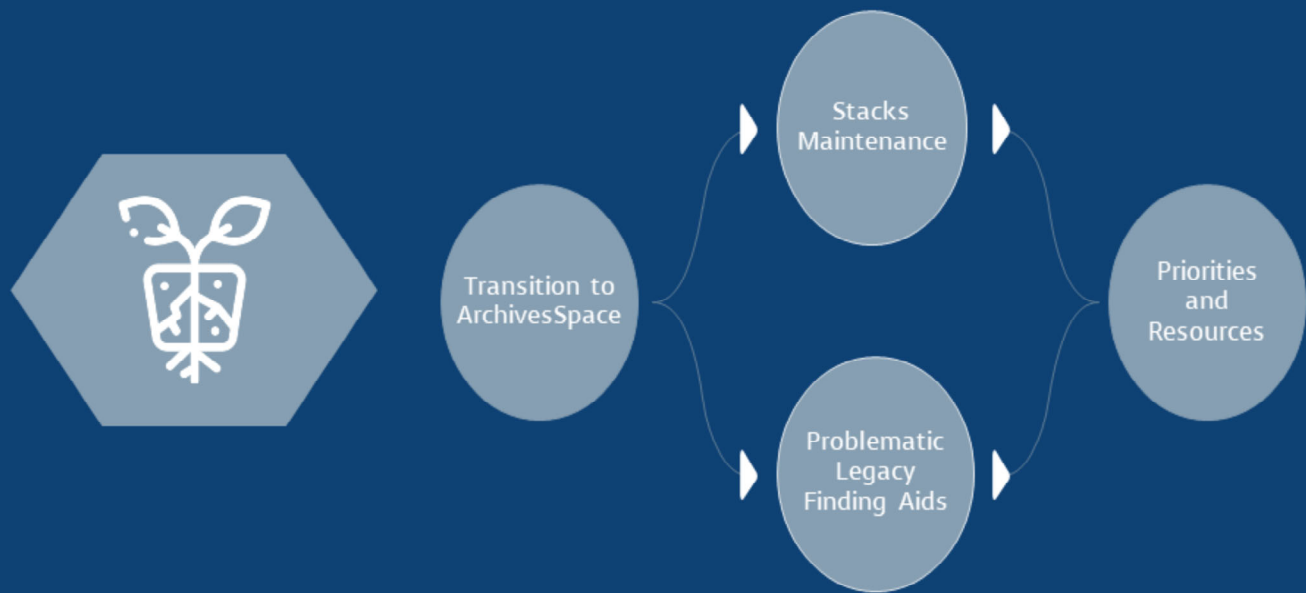
MMH Slide 2 (presentation slide 32) - First a disclaimer - we have not completed the survey of our collections that we started last year; the work I am discussing today is an ongoing project.

We were prompted to design and implement a collections survey for several reasons:

- Our transition to ArchivesSpace revealed a high number of inconsistencies and errors in our legacy finding aids.
- Those subpar collection descriptions were surfacing in Google search results which led to confusion for both researchers and staff.
- We recognized that areas of our stacks were becoming crowded, and we needed to embark on a major reshelving project.
- I wanted a systematic method to help me develop priorities for correcting issues and problems which would inform the allocation of staff resources.
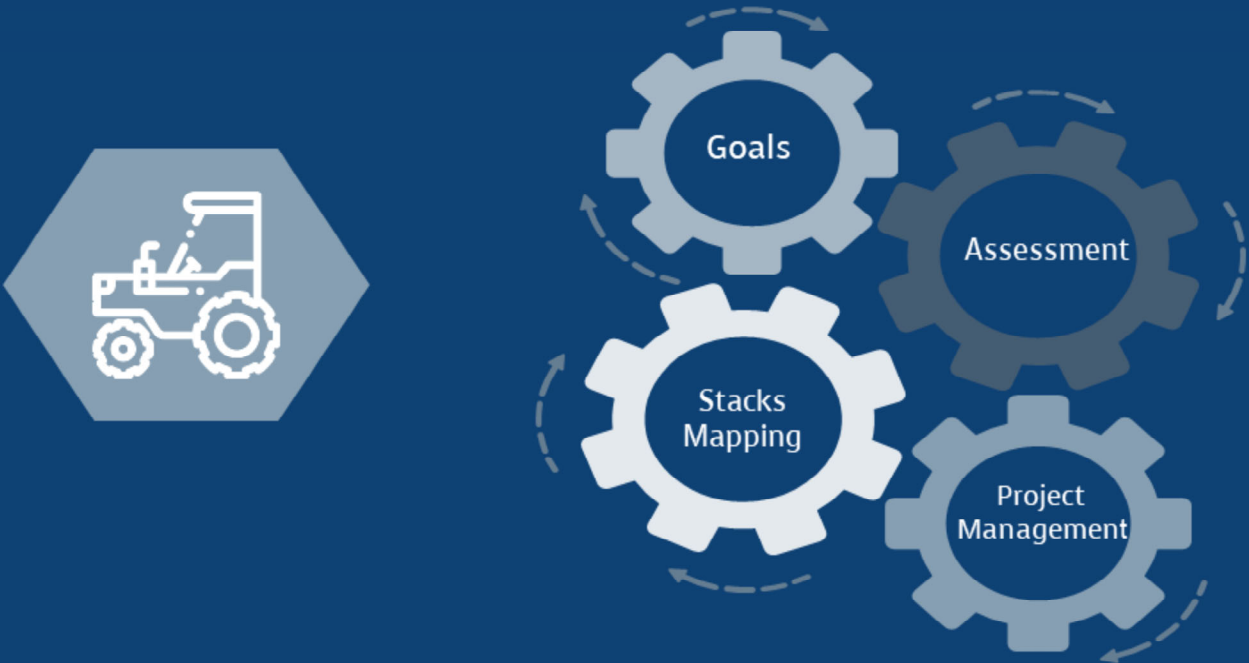
MMH Slide 4 (presentation slide 33) - I have three main goals with this project:
- Develop priorities for processing, reprocessing, and deaccessioning by applying a set of weighted criteria.
- Maximize storage capacity in our stacks.
- Revise our collecting policy.

I came to learn that Archives Space - the same system that shed light on our issues with legacy finding aids – could be used as a tool to help us systematically identify problems with our archives collections.

Following a conversation at MAC last year, Benn Joseph, Head of Collection Services at Northwestern University, graciously shared his repository's Backlog Assessment Tool Manual with me. Benn explained that the guide was developed to facilitate his team's evaluation of unprocessed collections. While the focus of our survey is our processed collections, this manual gave me an example of how I could customize the Assessment module in ArchivesSpace to meet our needs.

And, conversations with my colleague - Digital Archivist, Margaret G., convinced me that by utilizing the Locations and Container Profiles modules in ArchivesSpace we could gather information I needed to plan the long overdue reshelving project.

In addition to using the modules in ArchivesSpace, I am using Microsoft's Planner to assign clusters of related collections to each of the processing assistants as individual tasks. It is a very basic project management tool, but the features allow the processing team to attach

files, record notes, send notifications, and most importantly cross the assigned task off as completed which they've told me they really enjoy!

Harvesting the Data - Collections Survey
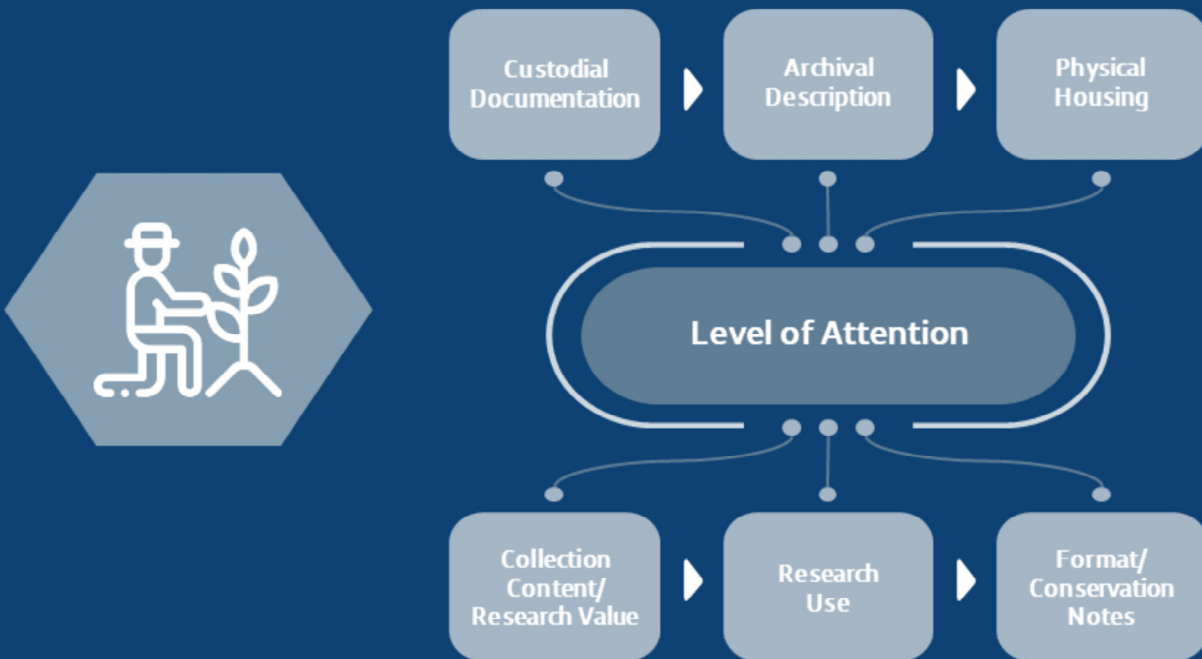
MMH Slide 5 (presentation slide 34) - How many folks are using ArchivesSpace as their content management system? How many of you are using the Assessment module?

Those familiar with this module will know that it includes 7 "out-of-the box" ratings which certainly reflected the type of information I wanted to collect but I found I disagreed with the alphabetical order in which these ratings are presented. After testing with a small handful of collections, I abandoned these 'out-of-the box' ratings and collaborated with Margaret to create our own in the 'additional ratings' section of the module which established a more logical workflow for the processing team and allowed for the collection of more data points. After customizing the assessment ratings, we utilized all the other standard sections of the module which include Basic Information, List of Material Types/Formats, and Conservation Issues. The manual that we created for this project instructs surveyors to skip the 7 'out-of-the' box ratings.

On this slide I've presented the categories of our additional assessment ratings and the general order in which the processing team is working through each collection. The first step is to review our donor files and record if there is a deed of gift or records transfer form. The surveyors also record the date of the deed of gift because prior to 2013 the form did not prompt donors to indicate their wishes for separations. We now ask donors to let us know if they would like separations returned, if we have their permission to discard, or if they have other plans for materials that are not retained for long term preservation. For collections donated prior to 2013, I will document efforts to contact the donors or their heirs

to inform them of separations prompted by our current related survey of the collections.

Before I talk about some of the initial findings from these assessments, I am going to share a bit about the stacks mapping part of the survey.

MMH Slide 6 (presentation slide 35) - The stacks mapping project was designed to gather information about the number and type of boxes across collections while assigning those boxes to a specific floor/range/unit location in our two stacks areas. Margaret built out the container profiles module in ArchivesSpace to provide a standard measurement for boxes used to house our physical collections and created location profiles which will allow us to see how much space is available on each range. As Katie K., a DePaul senior and a student assistant on our team, worked on this aspect of the project, she identified 11 mystery collections that were not described at any level in ArchivesSpace or had a Word or pdf finding aid on our network drive.
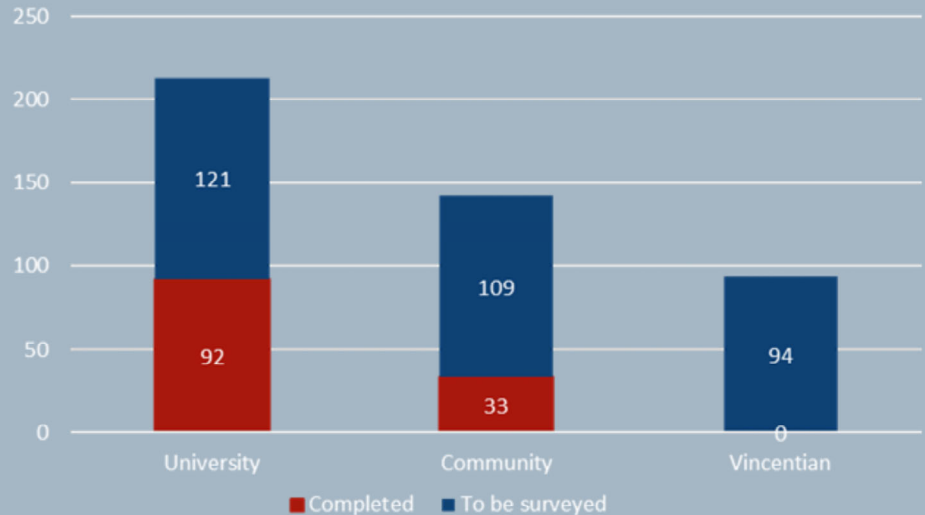
MMH Slide 7 (presentation slide 36) - As I mentioned the survey project is ongoing so the initial findings that I will be reviewing for the remainder of this presentation are taken from the analysis of assessments on 125 collections, 92 of those are part of our University Archives which is the largest of our three 'buckets' of archives collections and the current focus of the survey. The other two 'buckets' depicted on this slide's graph are our community archives and Vincentian archives. The time and extent metrics on this slide were captured in the Basic Information section of the Assessment module.

Digging in the Data - Archival Description

Quality of Administrative History/Biographical Note — Good, 36; Poor, 37; Fair, 52

Quality of Scope and Content Note — Good, 44; Poor, 29; Fair, 52

Quality of Arrangement — ?, 27; No, 33; Yes, 65
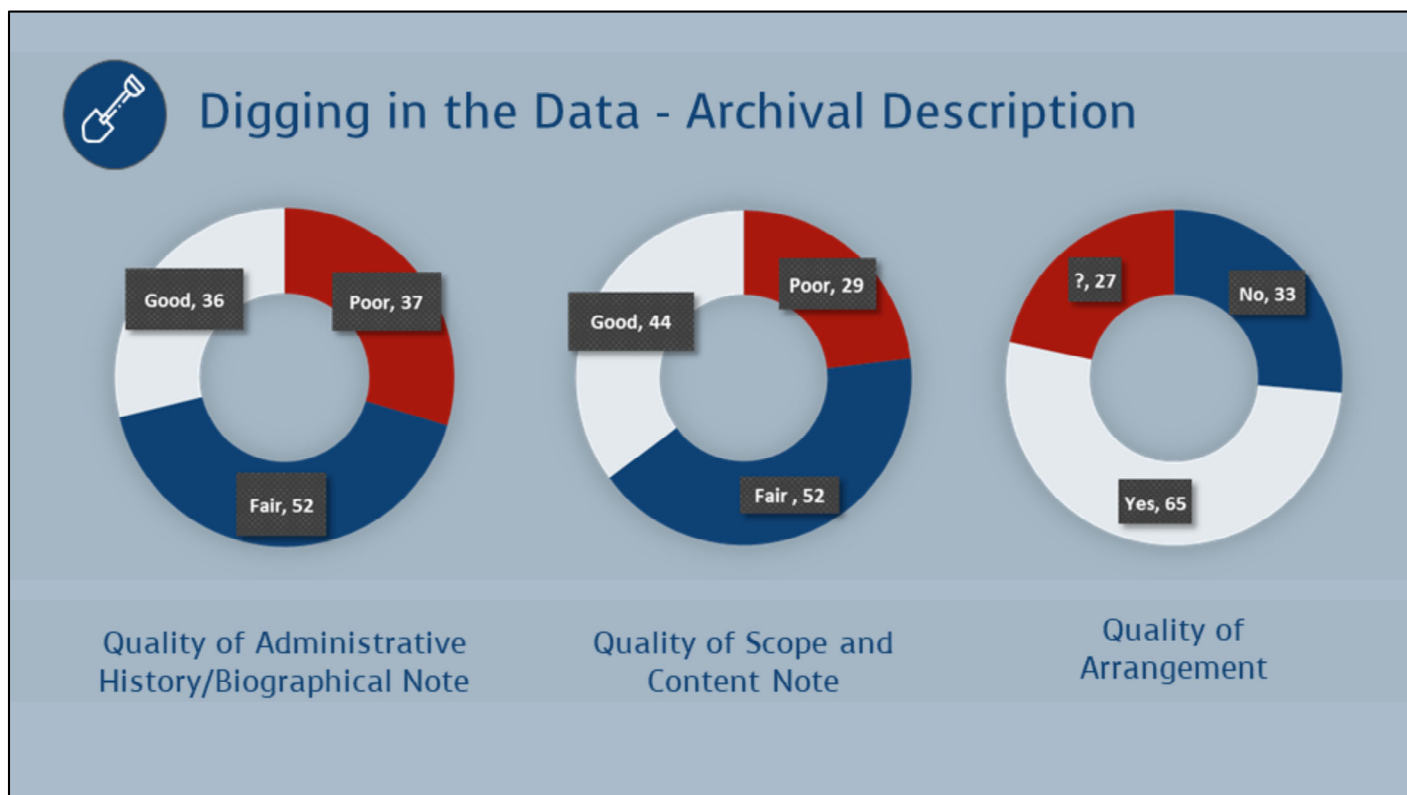
MMH Slide 8 (presentation slide 38) - A reminder – this survey project is designed to find problems so the data reflects our goal of identifying areas of weakness. To identify and track issues with the narrative descriptions of our collections, the processing team - Jeff E., Hannah K., and Ellie M. Mollman - had examples of well-written finding aids, prompt questions, and examples of common issues to guide them. We also have check-in meetings as a group to share information with each other.The portion of the survey that records the evaluation of a collection's narrative description includes the following prompts:

- Rate the quality of the Administrative History/Biographical Note – Does it provide contextual information about the university unit/division/department or organization/individual/family who created, collected, or maintained the papers? (Score 1 = Poor; Score 3 = Fair; Score 5 = Good)
- Rate the quality of the Scope and Content Note – Does it provide information about the materials and activities documented in the collection/records that helps a researcher evaluate its potential relevance to their research inquiry? (Score 1 = Poor; Score 3 = Fair; Score 5 = Good)
- Are you able to navigate the collection using the series and the file folder arrangement? (Score 1 = No; Score 5 = Yes)

You can see that more than half of the collections surveyed were rated as having poor or fair administrative histories and collection level scope and content notes.

And, the third chart demonstrates that survey tools are meant to be tested and adjusted. This is a great example of how a question we wrote has tripped up the processing team

and we will be revising it. To rate the quality of arrangement for a collection the response should have been either a yes or a no but 27 collections had a value of 3 which means that the surveyor was applying the poor/fair/good rating scale from the previous questions or it was left blank and when I reviewed the blank responses more closely the collection was rated poor and the accompanying notes recorded issues that overshadowed the question about a collection's arrangement.

Digging in the Data - Physical Housing

Issues/Problems…

box labels:
67
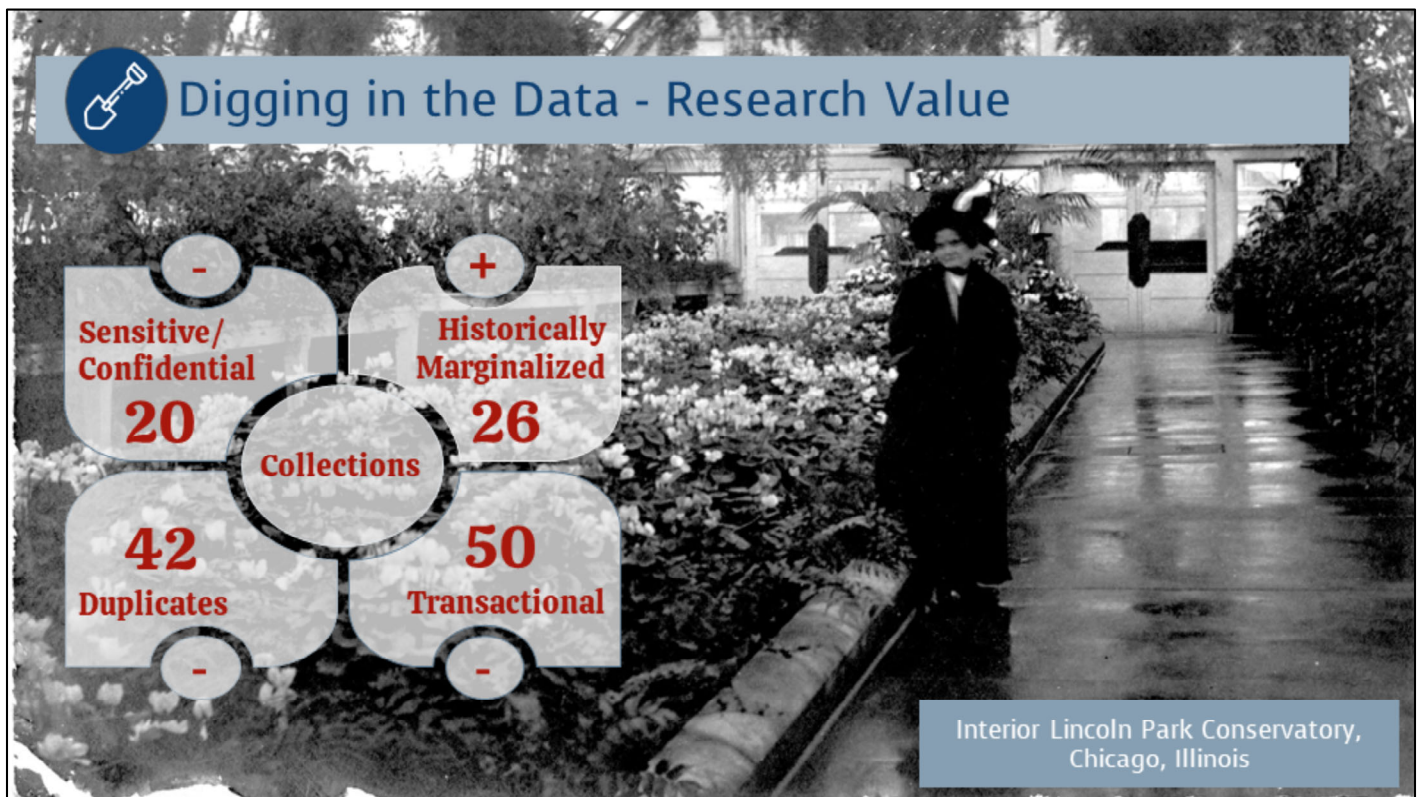collections

containers:
68
collections

file folder titles:
77
collections

"Fr. Smith's Farm" - Site of St. Vincent's Church and DePaul University, Chicago, Illinois

MMH Slide 9 (presentation slide 28) - The quote on my first slide – "All well managed archives need periodic cleanups" written by Elena Danielson in the 2010 book "The Ethical Archivist" applies to this part of the survey questions and the project goal of maximizing our storage capacity. I have pulled enough boxes for instruction or patron use to know that we have issues with many of the containers used to house a collection. For example box labels that rely on an acronym or an abbreviation or the legacy name of a university unit – these labels have caused confusion for new staff and novice researchers learning how to create citations and need the full name of the collection. I also wanted to collect information about the boxes used to house a collection and document if there are boxes that are nearly empty (we've pulled numerous boxes for instruction and patron use that have 1 or 2 file folders and a spacer) or overstuffed or if we need to find a better housing solution that will help with long term preservation.

Digging in the Data - Research Value

- Sensitive/Confidential 20
+ Historically Marginalized 26
Collections
42 Duplicates -
50 Transactional -

Interior Lincoln Park Conservatory, Chicago, Illinois

MMH Slide 10 (presentation slide 39) - Alongside issues with our archival descriptions and physical housing, there are known issues with the contents of collections.
For a variety of reasons, documents with little-to-no research value that could have or should have been separated during on-site appraisal, accessioning, or processing were retained, preserved, and described. We also have collections of high research value that document historically marginalized communities but the descriptions do not lift up these voices and perspectives or the finding aids contain outdated or hurtful terminology. In order to help us identify and make notes about the research value of collections, the survey includes questions/ratings about duplicates, transactional records, and sensitive/confidential documents. For example in University archives we are finding:

- Salaries tied to names and social security numbers
- Student disciplinary actions/suspensions
- Personnel disputes of a highly confidential nature
- Student grades and evaluations included in athletics reports
- Severance negotiations
- Tenure reviews
- Grade challenges
- Performance evaluations

And in our Community Archives we are finding:

- file folder titles that include the names of victims of crimes

- documents with personal identifying information of incarcerated individuals

Because of the sensitive nature of these materials we are considering the removal or redaction of these documents as a high priority and dedicating staff time to this activity rather than waiting until we complete the collections survey.
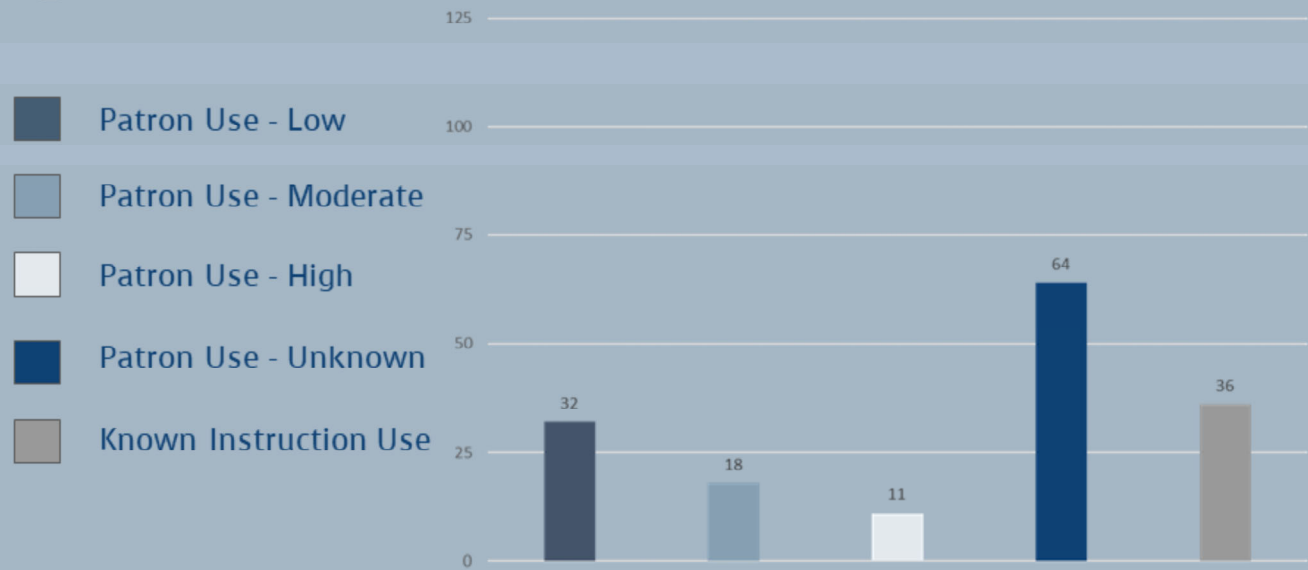
MMH Slide 11 (presentation slide 40) - In order to assign a value for a collection's research use, we are using data compiled from annual report patron statistics based on our paper registration forms (we do not use a system like Aeon) and instruction session bibliographies. It turns out that we have better information about patron use for our community archives because university archives are only tracked at the highest level. For example, if someone requested boxes from DePaul's Center for Latino Research or Division of Student Affairs those are simply recorded as "university archives." Thankfully, our practice of gathering usage information can be easily changed so that we have meaningful data going forward. And, of course collections are only requested by researchers if a finding aid is discoverable so the processing team is also noting the absence of a legacy pdf finding aid on our website, the lack of a record in our library catalog, or an unpublished resource in ArchivesSpace.

MMH Slide 12 (presentation slide 41) - I would like to end with an example of how we are using information gathered from the survey to correct issues.

This chart shows a summary of survey ratings for two university collections: University Ephemera which measures 5LF and University Memorabilia which measures 28 LF. These collections may not be used heavily by researchers in our reading room but they are used for instruction sessions and university events. Based on the issues noted during the survey the level of attention required to correct the noted issues was rated maximum because significant changes are needed at three levels: description, arrangement, and physical housing. We currently have a student assistant on our team, Ana O., who is a DePaul senior and a museum studies minor who enjoys working with objects and ephemera. We are not always in a position to match a staff member's interest to a collection that is next in the processing queue but this time it was possible. Ana is currently working to untangle these two collections along with similar ephemera and memorabilia scattered in our Student Affairs records and our Athletics records which were identified when those collections were surveyed.

Project Team: Ana, Katie, Margaret, Jeff, Ellie, and Hannah

MMH Slide 13 (presentation slide 42) - None of the work that I've described today would be possible without the curiosity, tenacity, and sense of humor contributed by my colleagues in this photograph: Student assistants Ana and Katie, Digital Archivist, Margaret, and Archives Processing Assistants: Hannah, Ellie, and Jeff. All of the team members gave their permission for me to share their full names (during the conference presentation; speaker notes only include first name and first initial of last name) and include this photograph.

Citations/Credit Information

Slide 1
- Mark Greene, "What Were We Thinking?: A Call to Embrace Reappraisal and Deaccessioning," *Provenance* 20, no. 1 (2002): 42.
- Elena Danielson, The Ethical Archivist (Chicago: Society of American Archivists, 2010)
- Guidelines for Reappraisal and Deaccessioning, May 2017, Society of American Archivists

Slide icons created by Flaticon; infographics by Freepik

MMH Slide 14 (presentation slide 43) - citations/credit information

Questions??

Megan Mummey, University of Kentucky, megan.mummey@uky.edu

Hannah Pryor, University of Louisville, hannah.pryor@louisville.edu

Morgen MacIntosh Hodgetts, DePaul University, mmacinto@depaul.edu