I’m Stacey Wolf and this is Jeremy Berg. We are both catalogers at the University of North Texas.

In 2011, UNT began a project to digitize our collection of print theses and dissertations, called the Retrospectively Digitized Theses and Dissertations project, which we all shorten to Retro ETDs so we don’t have to burn 20 syllables every time we talk about it.

Theses and Dissertations were broken down for mass scanning. Preservation cut off the spine and bindings so they were now just a stack of paper.

The digital projects unit handled all of the scanning and loading each Thesis or Dissertation into our Digital Library.

Then the cataloging department updated the records in two places: our ILS and in the Digital Library platform

- In our ILS, we already had MARC records for all of these actually, because they had been cataloged. However, theses and dissertations before 1974 only had minimal records.
- We corrected the fixed fields, added the online location, added the link to the digital library record, added a local collection note, and created an item record for the digital version.
- Most consequentially for our purposes here today, we added a 700 field for the committee chair/major professor.

The Digital Library record was not in MARC format, but local version of Dublin Core. These records contain similar information found in the MARC record, but sometimes in a different format. Or the record may contain more elements than what is found in our ILS record.

- For example,
  - The summary is new; very few of our MARC record contain a summary in the 520 field; or at least none that are substantive.
  - There are keywords in addition to the LCSH terms.
  - And we recorded the remaining committee members as well as the major professor/committee chair in the DL record.

As with any project, there are good and not so good parts

The good parts
They are important for preserving institutional history. As with most institutions, the only copies of these Theses and Dissertations were in our print collection at our campus, so they’re pretty inaccessible outside of our institution. Putting them online enables us to serve the scholarly community to a much greater extent.

The project was split between Preservation, Cataloging and Metadata Services and the Digital Projects Unit.

- The not so good parts
  - All this work had to be done for THOUSANDS of ETDs (we’ve completed about 11 thousand so far)
  - The signatures were frequently indecipherable,
  - and the individual’s name was not printed under them.
  - We had no idea about what we were going to run into with the signatures.
  - We started on the earliest theses and dissertations which had fewer variety in professors, only two committee members, and better legibility of signatures. The number of signatures we couldn’t recognize were minimal.
  - However, as we continued on the project, the signatures became more indecipherable, there were more faculty members to pick from, and committee grew to 4 or more members. And there were no longer as many resources that listed all of the faculty in one place.

Why is it important?

- Why important? It was a ton of work to figure out the signatures, easily the most time consuming and complex part of cataloging any given Retro ETD. Why go to all the effort?
  - We’re catalogers and it’s information, there is no why.
  - Can see what kind of research people were involved with, especially as major professors. Search that name in either the DL or the ILS and everything they were involved with pops up.
  - Can see who they mentored.
  - Are those things important? We don’t know, but we can see why they might be, and provide people the ability to find out.
  - The answer mirrors the process we used to decipher them: because you never know when a shred of information will become important, a connection between
people, or between people and subjects. We found those same small threads invaluable in investigating who these people were, trying to match up possible names with corroborating facts.

- Especially when we know the connections go beyond just the world of UNT—lead outwards, connecting people in the wider world.
  - There are professors from other universities.
  - There are executives from various companies.
  - Former U.S. Rep Dick Armey is in there—used to be a professor of economics at UNT.
- And this is all super rare information. It can feel mundane and only relevant to our institution—just some internal institutional stuff—but think of what we have here. There are a lot of ETDs out there, especially for retro projects like this, where your institution is the only place to find this information. Or the only place to find it for free, or it’s in the catalog and we’re the only ones who can push it out into the web.
  - That’s something else the keywords did for us in addition to providing alternative phrasing to the concepts in the LCSHs, or listing a concept not in LCSH. They make it easier to find the ETD on the web, which is where most people come to the DL records from, not the catalog.

**CLICK — SLIDE 6**

- Today we will talk about
  - **CLICK — SLIDE 6.1**
    - Chair and committee signatures
  - **CLICK — SLIDE 6.2**
    - How to decipher those signatures
  - **CLICK — SLIDE 6.3**
    - How to organize those signatures
  - **CLICK — SLIDE 6.4**
    - The Lessons we learned
  - **CLICK — SLIDE 6.5**
    - And finally what you can learn from us

**CLICK — SLIDE 7**

- Informal Poll
  - How many of you include the committee chair and/or the other committee members as metadata on ETDs?

**CLICK — SLIDE 8**

- Informal Poll
  - How many of you think it would be relatively easy to decipher a name using only a written signature?
- We have a test for you... ;)}
These are signatures that have been indecipherable in the past or we still don’t know who they are.

Also, just to show you how weird signatures are...

This signature (that we lovingly referred to forever as just “Loops”) is for a person named, Alexander Pettit.

How we deciphered the signatures

- Institutional knowledge through co-workers
  - Our coworker, Glenda, helped a TON
    - She had done acquisitions work in the 1980s and 1990s and knew some of the signatures and names.
    - Note that institutional knowledge in the form of people doesn’t last forever. A few more years and she might’ve retired. There are a lot of names we’d never get without her.
  - What information do your people have that may need to be recorded while they’re still around?
  - What other jobs have they done? We didn’t know Glenda had been involved in materials ordering—there was a length of employment gap in the department; no one involved with cataloging retro ETDs had been around long enough to know she’d done that. I forget how, but she happened to find out about our signatures trouble and offered her help. Similarly, Kevin Yanowski had worked in the English department, and was quite familiar with the infamous “Loops” signature we showed. Had we not managed to decipher that one by other means, his previous employment would have given us that invaluable piece of information, and he was able to help with other English signatures as well.

- Others in the library
  - Liaisons, subject librarians, public services librarians, people in acquisitions—anyone who regularly crosses paths with the faculty may be able to help.

- Admins in departments
  - We emailed departmental secretaries on several occasions, and they were usually able to help us—if you can’t reach faculty directly, or they’re no longer there, think about who was likely to see their signatures.

Official Records

- Budgets
Many of the University’s budgets have been digitized and reside in our digital library. These were a huge help.

There were often separated by department, so we could browse a department list and see if a name seemed to fit the scrawl in front of us.

- Not always guaranteed to help—sometimes there are similar departments, sometimes the sig’s from what’s really another department entirely, or even from outside the university. And sometimes the name’s just not there, for no reason we can figure out.
- Department sometimes seems to come up in multiple places.
- These have OCR, so you can search within the document by a fragment of the last name, if you have a guess as to what it might be, or by department, to try to find all mentions of it in the budget.

- No matter what shortcomings there are though, a hugely valuable resource.

### Graduate Catalogs

- Similar issues as budgets—last name first initial—but also similarly valuable.
  - Anytime you can end up with a list of names to try to match the signature to instead of just a totally acontextual scrawl to interpret, it’s helpful.
- Also arranged by department.
- A lot of the ones we have are print, not digitized.

### Yearbooks

- Some of the older yearbooks included a list of faculty for that school year, so you had a list of names to compare to the signature.

**[Jeremy chimes in here:]** I feel it’s important here to discuss just how crucial having these resources available to us has turned out to be. It’s a difficult truth that a lot of the faculty who pass through a university aren’t going to leave a huge mark—at least, not of the kind that’s helpful in matching their name to a largely illegible scrawl. The institution itself, however, was all kinds of concerned with them. They had to be. They hired these people, paid them, insured them, gave them statuses, assigned or allowed them to teach certain classes—their fates were intertwined. So it’s in the university’s papers, and there alone, that we are guaranteed to find at least some information on the professors associated with these ETDs. All of these sources are available to us because they were retained, and the budgets are only as usable as they are because they were digitized. Having digitized budgets in the Digital Library not only ensured they were available to us, it made them searchable, which was invaluable. It’s easy to say “why hang on to a 50 year old budget?” let alone go through the effort of scanning it and applying OCR. Even if it was just retained, this is by definition archival material—content generated by an organization’s standard operations, not designed to be published, distributed, and generally consumed. Well, this is why. Every piece of information that we describe and make accessible contributes to our ability to do so for other
pieces of information. There’s a building effect in preserving resources that gives us the knowledge to assign metadata so others can acquire still more knowledge. All of which can come in handy when your econ professor becomes a big name in Congress and you maybe want to find out who he worked with and what the capstone papers of his students looked like.

○ CLICK – SLIDE 11.3 [Stacey again]

○ Outside knowledge
  ▪ Newspapers
    • Both School and City/Town
    • Your local newspapers especially, which might carry items about the local profs that didn’t make the national news.
  ▪ Google Searching
    • Using quotes around the name
    • Searching city of school plus the name.
  ▪ Similar to doing authority work
    • Looking at subjects
    • Similar fields of study

○ CLICK – SLIDE 11.4

○ Citations in the ETD
  ▪ People will cite their professors, they usually have research in common.
○ Acknowledgements/Thanks
  ▪ If such a section exists, they sometimes thank their major professor/committee chair, which gives us some names for sure.

○ CLICK – SLIDE 11.5

○ For harder to read signatures
  ▪ Multiple eyes on the signature
    • Screenshots of a decent quality and pass it around via email.
  ▪ Look for similar letter forms
  ▪ Know/understand cursive writing
    • If you can find just a few letters in a row, sometimes that is all you need to stumble upon the rest of the name.

○ CLICK – SLIDE 11.6

○ Look at already known signatures (if you have them)
  ▪ This is really where this whole presentation came from. We kept screenshots of the hard-to-decipher signatures in a folder and list to compare and make it easier to spot them.
  ▪ Writing changes over time and with age
○ Rare, someone signing for a person
  ▪ Much clearer or different writing, but certainly the same name
  ▪ How we collected, recorded, and made usable the data.
At first, there was no thought to a sustained knowledge base. Just recorded what were deemed problem signatures.

- Consequently, we approached the problem piecemeal.
  - Pointed to a failure to see the full scope again. We didn’t step back and say “OK, if we’ve had this much trouble in the first couple months, what does that imply for the future?”
  - What’s not a problem to read for some may be for others, so it’s better to just record everything.
  - Even if they’re perfectly legible, it still takes time to read the sig and look up the authorized form (if applicable), faster to have a list you can just consult.

The signature situation reached a point where we had to create a centralized place to put them—there were too many to keep as individual images.

Eventually we started archiving all signatures, not just those deemed problems.

Began using MS Excel to list the authorized name forms in both the digital library and in our ILS (because of course they weren’t always the same).

- Excel was good for sorting and organizing, but...
  - Couldn’t put pictures with the names—had to cross reference. Wastes time, and Excel not well suited to images. Those being crucial to the entire process, we had to try something else.

So we moved to a table in MS Word.

- Good for putting everything together, handles images fine, none of Excel’s temperamental formatting issues, but it’s terrible for sorting and organizing.
  - The best we could do was to do a search in the document for department, then go through the results and see if we could find the person.
    - That’s dreadfully inefficient even for a reasonably-sized document, and it became rather glitch as the list topped 100 pages. We had to do something else.

After Excel and Word proved insufficient, we went with MS Access on the advice of Karen Harker, who is in another department entirely and has nothing to do with either ETDs or IT, but who we’d been told knows Access. That will be important for future considerations.

- Listed first name, last name, department, years known active, and a signature image.
  - You could search and order by any of those categories, so if you had a guess at the last name, you could confirm it. Or you could browse by department or any other confirmed characteristic.
Problem seemed to be solved and we were happy with this iteration for a few years. But eventually, it too became glitchy.

- CLICK – SLIDE 16 [Jeremy’s turn]
  - Access would display the wrong signature for names, but only every now and then.
    - Images did not match up
  - You had to view it in a certain mode, and it wasn’t intuitive. We had to mess around to figure it out.
  - Then it got really weird.
    - Copies of the Signatures DB started just appearing. It would just randomly copy itself and back itself up.
    - The database started slowing down, taking forever to open.
- Consulted Karen again.
  - Karen told us that the unending automatic backups and other weirdness were side effects of the database being way larger than an Access file should be. Access, apparently, was never really intended to handle images, or at least, not to be the platform for a full on image database. Though there’s something to be said for the innovative use of a tool, sometimes you hit its limits trying to make it perform in a different way than it was designed for. Especially when you don’t know that that’s what you’re doing.
  - Finally reached the point where the file straight up wouldn’t open.
    - We now had many hours of work and an essential resource rendered unusable and possibly lost.

- CLICK – SLIDE 17
  - Our Library IT department created custom online database to our specs: fields, image supportive, could handle the info load.
    - Also, they were able to put this database behind a secure log in, so that only people with permissions could log in and see/edit the data.
      - This was important because the legal implications of having a database of names and signatures could be quite problematic
- CLICK – SLIDE 18
  - They built a nice interface to enter new data
- CLICK – SLIDE 19
  - The table that is populated as data is entered is also editable so that we can correct or add information as needed
  - Note though:
    - Didn’t have it til the Access DB cacked for good. We lost valuable time trying to diagnose and fix the problem and figuring out how to operate around it, plus having to go back and do the signatures later.
• We got lucky, IT was able to salvage data from Access and populate the text portion of the online database.
  o But they weren’t able to autoload the signature images and the custom database couldn’t handle copy and paste, so we had to upload images, saving each one individually and loading them one at a time to the database.
  o Would’ve been faster to copy and paste.
    ▪ That was on me (JB), didn’t specify that when building the db.

• CLICK - SLIDE 20 [Stacey’s part]
• Lessons learned:
  o While some of these might seem like common sense now, if we were to start over again, these would be important things to address
    o CLICK – SLIDE 21
    o CLICK – SLIDE 21.1
    o Figure out the scope of your project early.
    o CLICK – SLIDE 21.2
    o Learn tools and what they do.
      ▪ Know if you’re using the right ones.
    o CLICK – SLIDE 21.3
    o Consult with the right people. This was something we didn’t do until it was way too late.
      ▪ Karen’s not IT. She does know Access well and is very computer savvy, but we still should’ve gone to IT, who are there to support us on technical issues, rather than choosing someone we happened to know who worked nearby.
        • I don’t mean to imply that no one knows anything outside of their job title, and certainly it’s faster to go to the people in your office who you can consult right away. But it’s a sign of the way that we continued to be kind of insular with the project even as aspects of it drifted out of our discipline.

• CLICK – SLIDE 22 [Jeremy’s turn]
• Takeaway that you can use
  o CLICK – SLIDE 23
  o CLICK – SLIDE 23.1
    ▪ Understand the full scope of the project
  o CLICK – SLIDE 23.2
    ▪ Recording information is important
      • You never know who might need the info
      • You never know what the info might be used for
  o CLICK – SLIDE 23.3
    ▪ Use multiple resources
      • If you can’t find information or can read a word, ask around
Search for other institutional knowledge to help clarify what you’re seeing.

- All the data was created as a side effect of the institution’s functionality—there will be pieces of knowledge in different areas that overlap and inform each other.
- And since you never know which will be helpful, preserve it and make it as available and searchable as possible.

- There will be institutional knowledge, but it might not be apparent who knows what
  - People know things, but they won’t work there forever—collect their knowledge so you have it when they’re gone.
    - Not just Glenda, but departmental secretaries we talked to, the faculty members themselves, etc.
    - It can happen sooner than you think. Recently, Stacey and I realized that after Glenda and Cathy, our principle cataloger, we were the next most senior members of our department. This was when we’re receiving our five years of service awards.

- Don’t forget to follow the unusual path

  **CLICK – SLIDE 23.4**
  - Ask for help
    - There might be an expert working right next to you or at least in the same building
    - It has been said, but...there are no stupid questions

  **CLICK – SLIDE 23.5**

- Work closely with IT.
  - They won’t know what you need, think the project through.
  - While your IT Department might not be able to custom build an online database, they will be able to help
  - It will save a lot of headaches down the road
  - They will know software (and possibly hardware) that can make your life easier
  - There might be an out of the box solution or possibly an open source solution that is exactly right for what you need, but you won’t know if you don’t ask.

**CLICK – SLIDE 24**

Questions?

**CLICK – SLIDE 25**

Thanks!