Thanks Cathy, for that great introduction and for providing the foundations for our discussion today.
“Ask any Canadian archivist about electronic records today and no matter what the specific question, you will likely get in response a great sense of frustration. For years, we have agreed that electronic records present archives and archivists with thorny problems ranging from appraisal and acquisition through conservation and public service.”


About a year and a half ago, I came across Archivaria 36 - the Special Issue on Electronic Records from the Fall of 1993 (I kicked myself that it had somehow escaped my radar previously!) As I read through the issue, I was struck by the similarities between then and now. Though there have been a lot of developments in the field of digital preservation and archiving of born-digital records since 1993 - I wondered why we were still having similar discussions about capacity gaps and other issues 26 years later? Or so it seemed, anecdotally.

As Cathy mentioned, Greg Bak has done a lot of the legwork in summarizing the writings on digital archiving in Canada from the 1960s to 2011. So I tip my hat to him and suggest you all read his two articles if you haven’t already.

Greg aptly argues, “the decision to turn away from media specialization resulted in two decades of wandering in the wilderness as Canadian archivists grappled with the tsunami of digital records that followed the mainstream adoption of desktop computing.” [2]

“...Canadian memory institutions still need to stop building digital memory infrastructure for the future and start building it for the present and past.” [3]

(I’ll talk a bit later on about some of the ways UTARMS is making progress in the present)

As a newer archivist, I’m left to wonder - has our community’s quest for perfection (and fear of mistakes) slowed us down? It is only in the last ten years or so that there has been a reversal of this position of shying away from media specialization, where we’ve seen a proliferation of new digital archivist positions. But yet - I can count on my hand the number of archivists in Canada with the word “digital” in their title. And as a result of all this “wandering in the wilderness”, we are left to play catch-up and deal with large backlogs of legacy media carriers within our collections. Now I know this is a bit of a generalization, and that there are some who don’t have the word “digital” in their title who are making in-roads in accessioning born-digital records, but
anecdotally, it feels like most archivists are still “confronting the digital”. To quote my friend Anna St. Onge from an email conversation we had on the topic, “if we aren’t comfortable with getting into the uncomfortable/risky spaces of digital preservation, we shouldn’t be archivists”. The current reality is that we all need to have some skills to manage “the digital.”

There are two further points I want to make about this before moving on to reflect on my work as a digital archivist over the last three years.

[3] Ibid., 286
[4] See Greg’s comment and notes about “confronting the computer”, 286
I can’t help but think that part of the reason why there has been such a steep learning curve and slow adoption to working with digital records is because our profession skews largely female - but there are gender and diversity issues at play when it comes to working in tech-related fields. It can be an intimidating and unwelcoming place for many, with few role models to look up to. A 2017 survey of more than 900 Canadian tech firms found women account for just 5 per cent of CEO roles and 13 per cent of executive team positions, while more than half – 53 per cent – of tech companies have no female executives. On average, women account for 8 per cent of director roles, while 73 per cent of firms have no women on their boards. Similarly, women represent only a quarter of the information and communications technology (ICT) workforce, even though they represent half of the overall Canadian workforce.

Ciaran Trace
What It Means for Records to Be “Born Digital”

- Born-digital records are forged in the alliance of the user, the computer hardware, the operating system software, and the application software.
- Understanding the levels of abstraction that form a computer system simplifies the inherent complexity of these machines and shows how records exist as physical, logical, and conceptual objects.

When I think back upon my formal training as an archivist - I graduated from the U of T iSchool in 2013, 6 years ago now - I don’t think there was enough emphasis on the shifting nature of recorded information and how that is changing with new technologies, or as Ciaran Trace so aptly put it, “What it means for Records to Be “Born Digital”. And even when I talk to current students about the present curriculum - in some of the archives courses, a lot of the discussion around “digital records” is still being framed as an afterthought for the end of the semester, as a “new challenge”, something to overcome, instead of addressing the work head-on with practical solutions about how to manage these types of records in the present (at UofT, in any case). This is discounting the course in Digital Preservation, which though important, as it is related to digital archiving - it encompasses a much larger field of practice.

With all that in mind - let me reflect on a few observations as a digital archivist for the last three years.

My role at UTARMS is a new one - I’m the first person to hold this position. As a result, over the last three years, a lot of my work has been figuring out how to insert myself into existing systems and processes, while also disrupting some of those existing processes. This has been challenging at times - for all parties involved - as we figure out who does what - especially as we are still operating in a hybrid realm - most of the collections we are acquiring, include both analog and digital files. So it helps that I have some experience working with textual records processing analog collections because lately I’ve ended up processing most collections that include a large digital component - in addition to the textual records. Time permitting, we are moving away from siloing the work - it is easier for me to make an informed appraisal decision about the digital files after comparing them to their analog counterparts, when possible. I believe digital archivists need to have some footing in the earlier archival traditions, because it informs so much about what we do, how we are doing it, and why it may or may not need to change.
Tools and utilities I like to use

- DROID and Brunnhilde
  ○ file format identification and reporting
- Teracopy / rsync / robocopy
  ○ file copying
- LibreOffice & QuickViewPlus
  ○ Viewing and rendering *some* older file formats
- Bagger / bagit.py
  ○ To create bags
- DataAccessioner
  ○ http://dataaccessioner.org/
- Community Resource for Archivists and Librarians Scripting (CRALS)
  ○ https://dd388.github.io/crals/

At UTARMS, we acquire records under two large umbrellas - university records and private records (i.e. personal papers) - and they each present with them their own set of different challenges. The goal is that over time, all the archivists will upskill to work with born-digital acquisitions as they increase, especially once these policies and workflows have been hammered out further. But as we all know - developing workflows is an iterative process and requires a certain degree of flexibility - especially in this realm - because it is sometimes difficult to pin down patterns - though they do start to emerge over time. So in the meantime, we are figuring it out as we go along. I’m experimenting with tools and seeing what I like best - and what works best for certain types of situations versus others. Some examples of these tools are listed on the slide here. There’s no one size fits all for this kind of work - it can vary from collection to collection based on a variety of factors including size, file format, and content. And I think that can be the most challenging sometimes - but personally, it’s the type of challenge I like!

Part of the current reality is that we are operating in a hybrid realm - most of the donors I’m working with are still operating in-between the paper and the digital spheres. Computing habits are varied, and there can sometimes be a lot of duplication and overlap with their textual files. Just like people can be messy with their paper files - how people work in the digital realm can be equally disorganized. But that’s ok!

Here’s an example: We acquired the digital files of a deceased political science professor - as I looked over the files I realized he would paste some of his emails into
Word docs - so I decided not to capture his unwieldy email inbox - just the Word docs because he was deeming these conversations important by saving them in Word. In this specific case, I was fortunate to be able to have a conversation with his widow, who happened to collaborate closely with him, and she was able to provide insight into his work habits with his laptop.
With regards to university units we are increasingly seeing digital-only transfers and acquisitions. This is often after building up a relationship with an office or donor and explaining to them our capacity to maintain their digital files - and the value in keeping them in their native digital formats.

In order to raise awareness around these issues, and make it as easy as possible to get small accessions of born-digital files into the Archives, we created the UTransfer service. We wanted something to be as user-friendly as possible. In discussion with staff from the library's ITS department - we found out there was existing cloud infrastructure in place for research data management, that we could piggy-back onto, for our own purposes.

We ask staff to fill out a questionnaire - the idea for this was based on what Richard Dancy and his team at Simon Fraser University had created with their MoveIt project, so we weren’t re-inventing the wheel, but rather building upon their work.

UTransfer has come with its share of successes and hiccups - it doesn’t work well with large accessions over several GB for example. As the project has matured, the procedures have evolved and been refined. For example, we no longer need to give offices a login, we can just share a link with them to a folder for them to upload their files to.

In any case, it’s a success I’m proud of, because it highlights the collaborative nature of this work and the fact that we are doing SOMETHING and it serves as an outreach
tool and conversation starter when training staff about the records management program at the U of T.

Some examples of recent born-digital acquisitions include...

Last year, receiving a transfer of digital files after my colleague worked with them to schedule their records, a few months prior. We could see the results of his work in helping then schedule their records.

One of the things we're grappling with however, is that we are only receiving instances of records – a snapshot in time. It is difficult for us to then tell them to delete these files – because they may want to hold onto them and reuse them – because it is so easy to do so nowadays. We are telling offices that they can delete them because we are safely preserving them now – but we have no means of making sure they do delete the files. Nor are we sure it is our responsibility to do so. Though we encounter duplicates in the paper world – it is much easier to hold on to and make duplicate digital files. It's something we're grappling with and still figuring out what the best solution is.

Another example: Student group who doesn't use email but instead communicates via Facebook Messenger Chat. Work with them to self-download/archive the conversation and then transfer to us. This was the first time I had worked to acquire something from a social media platform like this - it was a learning experience for all parties!
As you all know, we’re still very much in a paper world. I just wanted to include this photo of the UTL’s offsite storage facility in the Downsview neighbourhood of Toronto, partially because its just so impressive and an important reminder of our responsibilities as record custodians.
One of the things I struggle with the most - partially because it is so time consuming, is going back and addressing the legacy media carriers within our existing collections as well as file transfers that were put on our network drive and never fully properly described. The act of figuring out what's on the media, rendering the files, and then undertaking appraisal and description is a time consuming task, especially when you were not the archivist who took in the content in the first place. This is something I try and chip away at one day a week - but it is a slow one.

And sometimes it's just a matter of opening the boxes and removing the floppy disks - I think they can be distracting to keep there if we aren't going to be preserving their contents. It's important to remove them.
• Have a conversation with donors and get a sense of their computing habits
  ○ A bit exhaustive, but good template in Appendix F from AIMS Project (2012) whitepaper
  ○ https://dcs.library.virginia.edu/files/2013/02/AIMS_final_appF.pdf
    ■ Do you separate your personal files from your work files?
    ■ What types of digital files are created? (e.g. word processing files, images, spreadsheets, databases, etc.)
    ■ If you create files in both digital and paper formats, do certain files exist in both formats? (e.g. drafts of writings, email, etc.)
    ■ How are your digital files currently organized? (e.g. filed in named folders? by projects? by topics? some other scheme?)
    ■ Do you use more than one computer (e.g. office desktop, office portable computer, home desktop, etc.)? If yes, how do you synchronize files between different computers?
    ■ What media are used for backup files? (e.g. optical disk, hard disk, file server, web based backup service, etc.)
    ■ Do any digital files require passwords?

• Explain to them all that’s involved in preserving their digital files

**Donor handouts**

• Examples from the Beinecke Rare Book & Manuscript Library
  https://beinecke.library.yale.edu/article/born-digital-archival-acquisition

With regards to new accessions, I am much more comfortable - partially its a control thing, I think - because I am better equipped to respond and react to the types of content being acquired.

Thanks to the plethora of resources available online I am also better prepared now to know what kinds of questions to ask when speaking with donors or offices. It is important when possible to have a conversation with them about their computing habits and what operating system they use, what kinds of file formats they are working with, amongst other things - as this will inform and contextualize how to preserve the files but also better helps in the description.

[8] https://dcs.library.virginia.edu/aims/white-paper/
[10] https://beinecke.library.yale.edu/article/born-digital-archival-acquisition
I think part of the biggest hurdle can sometimes be not knowing what questions to ask or adequately communicating with IT to make the case for your needs.

“Just as archivists have a language specific to digital preservation; IT has its own for backups and storage. These languages can sound awfully similar as the two share common roots. But the use of similar key terms can often lead to confusion.” [11]

I recommend Scott Prater’s good article on this topic.

In my experience, the best way to break down barriers regarding lingo is to ask questions, read up on the topic and talk to fellow colleagues. In my experience, most people in the digital archiving / digipres community are a nice and generous bunch of folk. It is important to read up and keep abreast of new developments. This can be seem overwhelming at times, so try and carve out an hour a week to do so. I keep bookmarks and lists of things I see on Twitter and on listservs.

These are just some of the issues I’ve encountered on the job so far - if anyone has any specific questions about anything discussed here today, please send me an email.

Thanks for listening to Cathy and I today – and allowing us to reflect back on things in the 90s and look at where we are now - I hope it will spur a lot of discussion. The themes mentioned in 1993 are still ongoing, partially because the digital work is never ending - it is an ongoing task that needs constant monitoring and frequent adjustments. This is a shift in mindset from the past, wherein once something has been arranged and described and put on the shelf - you don’t need to think about it too much.

I know that we are slowly building up a community of practice here in Canada - and that is something of which I am proud to be a part of. Don’t be overwhelmed by the task at hand - we are all in this together!
Thank you!

e.sommers@utoronto.ca

https://xkcd.com/2143/