This presentation was part of a three person panel, “Going Beyond the Click: A Moderated Conversation on the Future of Archival Description” at the 2016 Radcliffe Institute Workshop on Technology and Archival Processing, held at the Schlesinger Library at Harvard University, April 4-5 2016. This panel aims to consider this transformation and the following questions. How does archival description change when there are no paper records? In digital record environments, what will define the provision of meaningful access? How will workflows change in the future?

The title, Stop Looking for the Quarter Under the Light, refers to a comic from 1942 wherein a person is looking for a lost item under a lamp because there’s more light there ... even though they lost the item somewhere else. I used this a frame for thinking about how we shouldn’t continue to look at how we do processing for analog collections as a model for born-digital collections. We might need to step away from those known processes and look for others to best do our work with born-digital collections.
When we think about born-digital records and arrangement and description activities I believe it’s useful to keep in mind the competencies that our profession has agreed upon as important. I’m calling out these four to situate my presentation: Arrangement; Description; Discovery; and Ethics.
## Born-digital Archives Use Cases

- **Known records that are individual objects** (e.g. annual reports)
- **Accruals of records or new administrative collections** (semi-organized business records)
- **Personal Papers** (Evidence of Me - HT to Sue McKemmish)
- **Productivity accounts; databases; Web archives, etc.**
- **Reformatted collection material**

There are five use cases that I consider when thinking about digital archives and how to approach processing and sustainability of them. Reformatted collection material is often a/v although sometimes digitized text and images can benefit from some of the same techniques.
There are tools we can use, such as the script written and used at Princeton University Archives, to automatically extract the **arrangement** and basic information of files in folders and their dates and extents.
Using tools like ePADD, we can also automatically extract agent and location information, which can aid us with Description.
Our researchers of born-digital material, especially Web archives, need more than usual amount of information about provenance that we provide for physical collections. The mechanism and criteria used in collection is critical for **discovery** and understanding the content in our collections.
Productivity Accounts

- Email messages
  - Message attachments
- Calendars
- Task lists
- Contact lists

Local or hosted accounts?
Organizational server or cloud service provider?
Transformation tools used? What has been restricted?
How does this effect email “conversations”?
Do the contact names and email names resolve to XXXX@XXX.XXX?

How we process and understand productivity accounts, such as email, address books, and calendars, needs to follow our guidelines for archival ethics and safe keeping personally identifiable information.
Other tools, such as Tagul, that we can use on born-digital content are things like auto summation and word clouds, which can produce a visual discovery tool for words that are in a text resource.
Auto Transcription

Running auto transcription tools, like in the ResCarta Toolkit, against audio files (or audio tracks from videos) can aid us by producing a good-enough version of an audio resource that can be used for description.
There are many standalone tools and ones that work together, such as within the BitCurator environment, that help us to understand the information about what is in a digital object (it's an artwork, or a story, or a report, or a photograph of a butterfly) and also about the digital object itself (a complex object consisting of HTML text file, an image jpg file, and an audio mp3 file). This is important for the metadata that we need to collect and keep for the sustainability of the digital collection.
Being transparent about our preservation actions on the digital content that we hold and are responsible for sustaining is important not only for future archivists and content users to understand the life story of born digital content, but also increases trust between donors, archivists, and content users.
The second half of my presentation discusses lessons learned and offers some suggestions for how to approach and work with born-digital collections and creatively think about their arrangement, description, and discovery.
Move away from the streetlight

- Much of the descriptive information for born-digital collections will not comfortably fit into DACS and EAD structured finding aids.
- Presume researchers will be using software tools to render, sort, extract, and produce their own collection or partial descriptions.
- The burden on archival repositories to provide researchers with provenance, fixity, and transformation information will continue to rise with born-digital collections.
Expand the search area

- Provenance information and processing actions can be documented in a variety of records, including accession records.
- Implement technology solutions that include URLs that can be cited for provenance information as well as collection objects.
- Build robust Information Packages that can be delivered to researchers consistently, over time, with added metadata as data transforms.
- Adopt and adapt tools from other domains.
Look to additional Standards

- PAIMAS
  - Producer-Archives Interface Methodology Abstract Standard
- OAIS
  - Reference Model for an Open Archival Information System
- ISO 15489: Information and Documentation - Records Management Part 1 (General)

It’s really important to do your work in accordance with the Archival and preservation standards we have. They may seem a bit daunting at first but you will be surprised at how helpful they are with doing your work in preparing, describing, and making available born digital collections.
There is a lot more metadata that’s necessary to document and keep track of for digital files than there is for physical materials. In fact, most of the metadata about the digital collections you will be keeping is not about the description of the content in the collections at all! It’s about the digital files themselves and how they are made, how big they are, what software is needed to render the content, what happened to the files, when was it copied and transformed to a different file format, etc. Without this information in the future (and that can be in just 5 years from now!) you will only have digital files that you might know about but can’t use. Don’t let that happen!
Look at the life cycle models that are available for learning what happens or should have happened with managing digital material over time. The DCC model, the Records Continuum model, and the Digital Preservation Management Life Cycle models are three examples.
Think modularly when it comes to the software tools and people that you use for processing and delivering your digital collections and description. There is no one-system-fits-all.
Consider letting your users have much more information that you might for your physical collections. In the Dissemination Information Package, your access handbag of digital objects, you will not only have the digital files from the collection but many of the other kinds of automatically extracted data that you can share with researchers.
An example here is a MIT Institute Archives and Special Collections diagram of where tools fit in the digital archives processing, preservation, and delivery ecosystem.
Questions & Comments?

Kari Smith
smithkr@mit.edu
@karirene69

*Engineering the Future of the Past*
http://libraries.mit.edu/digital-archives/