Follow-Up Questions

ASERL Webinar: "Intro to Digital Preservation #3 -- Management of Incoming Born-Digital Special Collections"

Speaker = Gretchen Gueguen, University of Virginia. April 16, 2013
Session Recording = **https://vimeo.com/64166341**
Speaker's PPT = **http://bit.ly/17CoNxN**

1) What does a virus look like? How do you check for it?

Virus files or virus-infected files may look just like any other file. Run a virus-scan software which will both identify them and take steps to remove them. The big names in virus software are McAfee and Norton, but there are open-source options like ClamAV.

2) What are the preservation needs for disks, in terms of temperature, humidity control, etc.?

Pretty similar for paper collections actually, dry and cool. CDs and DVDs in particular will delaminate in high temperatures. Floppy disks can become de-magnetized as well. Here are some sources for more info: http://www.clir.org/pubs/reports/pub121/sec5.html, http://www.atarimagazines.com/creative/v9n12/205_Floppy_disk_handling_and_.php

3) Can you provide an image of a 3.25 disk to show the step you talked about in the "read only" [write protection] process?

I'll just point you to the section on this page under "Write Enable Notch" for 5.25" http://electronicstechnician.tpub.com/14091/css/14091_261.htm and this one for 3.5" http://electronicstechnician.tpub.com/14091/css/14091_262.htm

4) Where can I find a glossary of terms?

The AIMS project whitepaper has a glossary, as do the Interpares project (http://www.interpares.org/book/interpares_book_q_gloss.pdf) and the Paradigm Workbook (http://www.paradigm.ac.uk/workbook/glossary/)

5) How do you get extracted metadata to a finding aid or repository record?

This completely depends on your system for creating finding aids or your repository. Usually if you can produce metadata in a tab-delimited, or comma-separated value (CSV) form, you can then either batch it into another structured data system or script it into some other form of XML. If you are creating your finding aids natively in XML this could be a useful way to generate portions of container lists. The instructions in this tutorial can show you how to export XML from an Excel spreadsheet: http://www.youtube.com/watch?v=9bat12gH3Qs

6) How do you take single metadata records and batch compile them?

I believe a scripting language like Perl or Python could do this, but this is not something that I do with any regularity.

7) What do you advise for managing file names for born-digital materials from donors, that cause migration issues? Especially when they're bundled?

Archivematica has a nice utility for dealing with this problem. It will record the original file name in the PREMIS record it creates for the file, then change the filename and record the change. If you don't want to adopt Archivematica to do this automatically, I'd advise creating a similar process.

8) Do the techniques you showed us apply to audio and A/V materials as well?

Yes, you can image A/V disks (CD and DVD) and then treat them as you would other image files.

9) How do you tell if a file with a .pdf extension is pdf/A, pdf/X, simple pdf, or something else?

PDF/A is identified through natively embedded metadata in the file (http://en.wikipedia.org/wiki/PDF/A#Identification). PDF.org recommends the use of these tools to verify whether pdfs are pdf/a: http://www.pdfa.org/?s=product

10) What recommendations can you make for preserving genealogy databases (e.g., files in GEDCOM) Is there a free/low-cost version to open these files?

I apologize, I have no experience with this format. A quick Wikipedia search leads me to believe these are records in a standard data format, in which case I would recommend backups and checking/refreshing them periodically.

11) Advice on setting priorities for archives with small staffs – what to preserve first?

This is exactly what the inventory helps you do. Once you finish it you could prioritize different ways: oldest materials (therefore high-risk of loss) first? High priority collections first? Materials for which you have the capability? Institutional commitment? I can't answer for you which things are the highest priority, but I do think doing the inventory will help you make a plan.

12) What are the pros/cons of different types of write-blockers / forensic bridges?

I think the basic breakdown is between cost and performance. Performance can be affected by what you need (do you need IDE connections for hard drives, or USB connections for modern external drives). I'd suggest asking this question of the Digital Curation Google Group, you may get a variety of good advice there.

13) Any advice for organizations that create and archive their own content (e.g., audio/video content?)

Since you are creating it, you can try to create it in standard formats with good file names and organization from the beginning. This will only help you in the long run. You can also create some useful documentation as things are being created, even just descriptive metadata that no one will have the time to create later. Look at best practice guides from institutions that are doing audio/video reformatting for ideas of what their acceptable quality standards are. The Federal Agencies Digitization

Guidelines Initiatives created some resources that might help you sort this out:
http://www.digitizationguidelines.gov/guidelines/

14) How can you determine which file formats are best for user access? For preservation?

In terms of access, I would say common sense rules. Go with the majority…jpeg, pdf, mp3…but know your audience, if they are going to be constantly requesting higher quality copies or original digital files, you may consider making those accessible as well. For preservation, I would say the original is the gold standard. You would have to have a really compelling reason to migrate to another format and throw out the original entirely. There is a process called "normalization" but this isn't the same as migrating…it's more like file repair. It doesn't fundamental change the file (making a pdf into a pdf/a for example is just a process of adding metadata).

15) Do you have recommendations for evaluating archival material that is in obsolete formats when it is offered for purchase?

Be realistic about your institutional capabilities. Be prepared to invest time and money into research and testing. Don't make promises that are unsustainable. I apologize that I can't be more specific, but a lot of the resources in the bibliography should help with figuring this out for your institution.

16) What type of computer(s) do you use for accessioning (make/model)? Do you use the same workstation(s) for all accessions/processing?

We have a Forensic Recover of Evidence Device (FRED) at UVa. This is an all-in-one product with a write-block bay and ports for many disk and connector types, a raid-array for storage, and hot-swap bays for additional hard drives. I am not endorsing the purchase of one of these devices. It was purchased with grant funding in a testing capacity, but it does work well for what we are doing. It is powerful enough to run FTK software and has a lot of the peripherals needed (as long as you have external drives for older disks like floppies and zip disks). We have also purchased a second PC with a large monitor and enough memory and storage to run the FTK application. It cost us about $3,000. The FRED was more expensive, but I am not able to say exactly the price.