

# Digital Scavenging

A Repository Rant at Open  
Repositories 2014, Helsinki,  
Finland

Linda Newman

University of Cincinnati

[newmanld@ucmail.uc.edu](mailto:newmanld@ucmail.uc.edu)

# Digital Scavenging

We've accomplished both a lot and a little. We're building new platforms using agile software development, separating out the CRUD successfully with an eye to persistence and long term preservation, supporting sophisticated computing environments with multiple VMs and load balancing, and moving from simple formats with a 1:1 ratio of file to intellectual object, to complex multi-part formats. We are exploring digital forensics.

# Digital Scavenging

But our users may still not know or care who we are. We have a 'build it and they will come' attitude. We need to do more than build the platforms and acquire content piece by piece.

- Should we become anthropologists and archaeologists of the future – ravenously, even stealthily, scavenging our digital environments for what we can squirrel away, perhaps hidden, for our future colleagues to find?
- Could we capture entire systems as objects in our repositories, hoping that our future colleagues will be able to unfold them like origami, able to explore whole past worlds?

# Digital Scavenging

We are competing for overstressed time of faculty and researchers and asking them to make time to store what should be persisted. Government mandates for publicly funded research help, requirements to preserve born-digital legal documents help, yet the producers of the born-digital content we covet may continue to view our systems as an afterthought – a nice to have.

Instead of collecting our cultural heritage piece by piece, with major gaps, how could we truly preserve entire environments, not just pieces of them?

# Digital Scavenging

## Ideas:

- ❖ Actively acquire backups of content systems on our campuses – course delivery systems, campus research information systems and researcher directories, classroom video-delivery systems, iTunesU and store them (perhaps embargoed) for a future digital archivist/archaeologist to parse apart, long after IP considerations have expired.
- ❖ Run an Internet Archive-like Way-Back Machine against our campus systems – websites, library catalogs, Twitter and Facebook feeds of our campus administrators, and going as deep down the web tree as we can?

# Digital Scavenging

Ideas (continued):

❖ Create Virtual Machine Disks of entire systems (as if we were going to migrate those systems), and which our future colleagues, we hope, can later unfold to explore whole past worlds.

- In early 2014 we moved a repository system from Columbus Ohio to our Cincinnati campus.

- 10 TBs of data, in 16 different VMDKs (virtual machine disk images) was transferred over the internet pipeline

- Checksums were created for each VMDK and verified upon receipt.

- We have saved the VMDKs themselves on tape – perhaps we should have saved them in our repository itself?

# Digital Scavenging

Fair criticisms:

We would be acquiring sensitive data, with access rights buried in that data. Much of this content could need to be embargoed for years if not decades.

We do not know the best standards for such digital objects and scavenged images of entire systems to increase the likelihood of their use in a few hundred years! (Perhaps research in digital forensics and specialized disk images suggests a direction.)

# Digital Scavenging

Some of us are using systems which have been developed in law enforcement for forensic archiving with our born digital archives.

FRED – Forensic Recovery of Evidence Device is a common hardware/software option that can read old media (zips, floppy, etc. and archive in various formats considered reliable by forensic archivists.



A photo of FRED from  
<http://www.digitalintelligence.com/products/fred/>



# Digital Scavenging

Another approach to digital forensics is provided by

# BitCurator

<http://www.bitcurator.net/> -- a joint effort led by the School of Information and Library Science at the University of North Carolina, Chapel Hill & the Maryland Institute for Technology in the Humanities.

BitCurator provides:

- Pre-imaging data triage
- Forensic disk imaging - Expert Witness Format (.exx), the linux raw image (.dd), 'advanced forensic format' (.aff)
- Filesystem analysis and reporting
- Identification of private and individually identifying information
- Export of technical and other metadata

# Digital Scavenging

In their article “**Extending Digital Repository Architectures to Support Disk Image Preservation and Access**” (<http://www.ils.unc.edu/callee/p57-woods.pdf>) Woods, Lee and Garfinkle not only offer a good overview of the Disk Image as Digital Object, but call for these forensic disk images to be themselves collected and stored in our digital repositories.

Could we use forensic disk imaging techniques to not just image the content of old floppies, hard drives, etc. but to harvest entire platforms/delivery systems at our institutions?

# Digital Scavenging

Once I placed a political poster on several area bulletin boards and I observed an archivist carefully carry one away. I suspected I had been ‘marked for accession’ – that my poster was viewed as potentially valuable ephemera to store in an archival folder where it might be found by some future researcher. That’s what archivists have, quite laudably, often done with what they see around them – I’m suggesting we should re-tool and re-think to do this programmatically – maybe just as sneakily – with the born-digital content that surrounds us.



# Digital Scavenging

But we have more work to do to develop best practices, or our cultural record will become an unworkable scrap heap for future digital archaeologists trying to mine the past:



Photograph credited to Robert & Mihaela Vicol at [http://publicphoto.org/people/scrap-metal-scavengers-poor-people-hard-work/attachment/scrap-metal-scavengers-poor-people\\_img\\_1292\\_cr/#](http://publicphoto.org/people/scrap-metal-scavengers-poor-people-hard-work/attachment/scrap-metal-scavengers-poor-people_img_1292_cr/#)