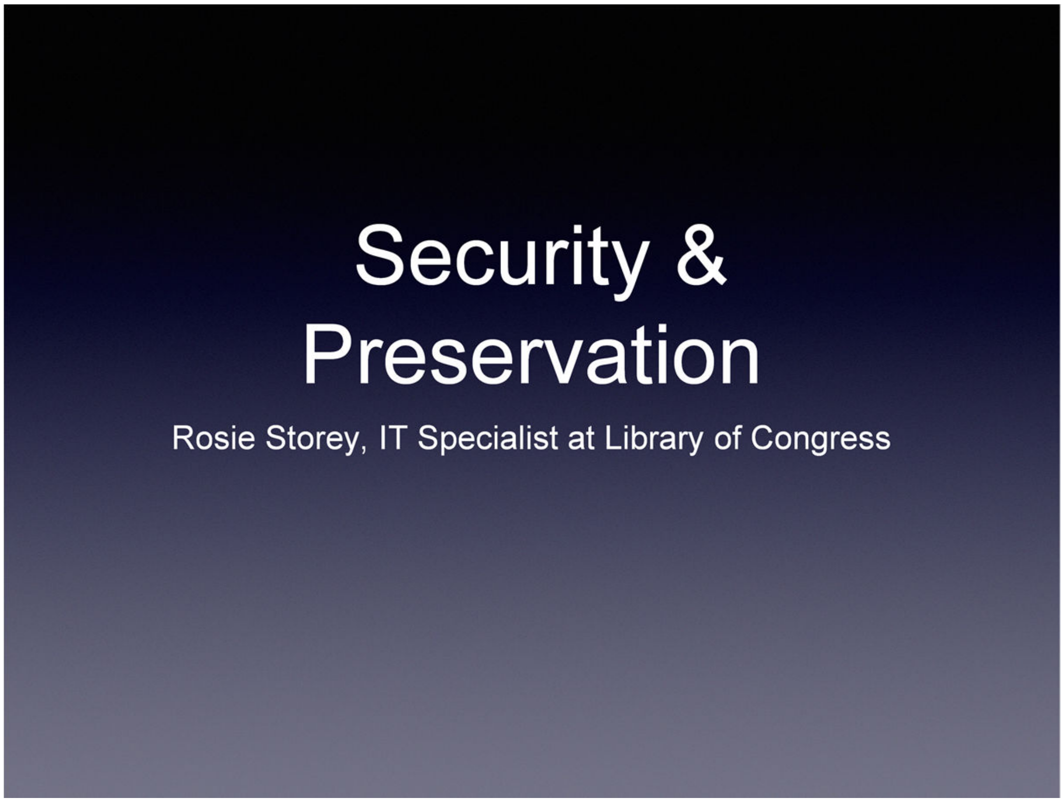
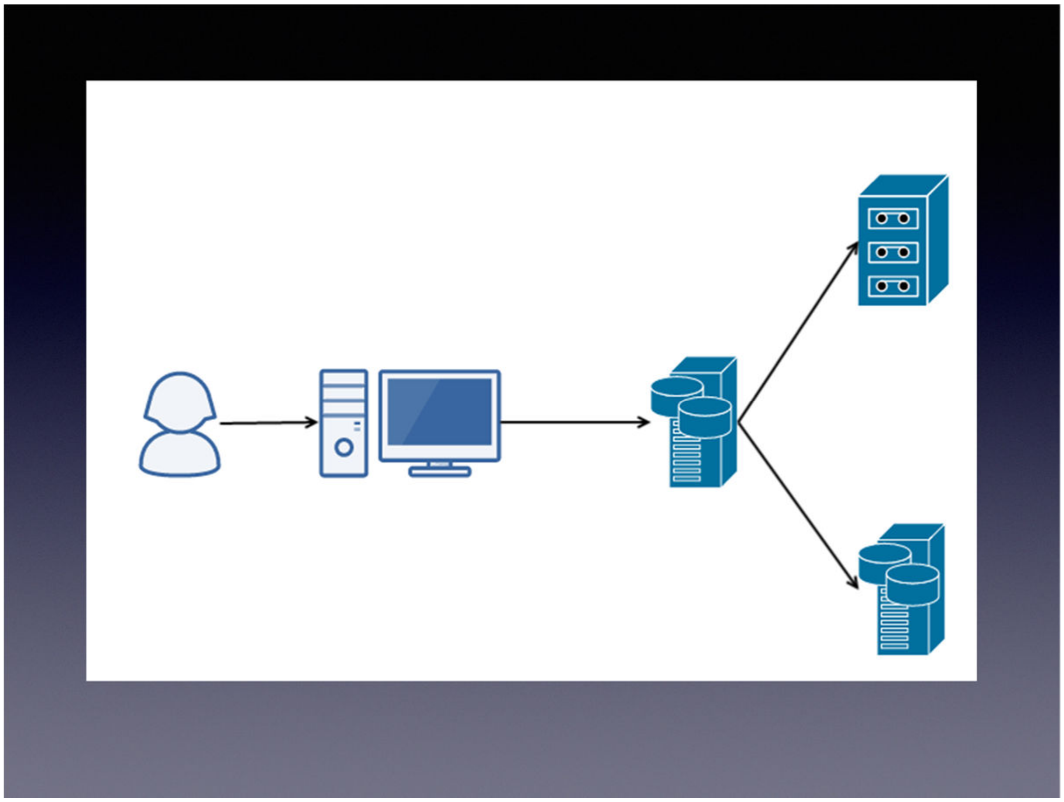
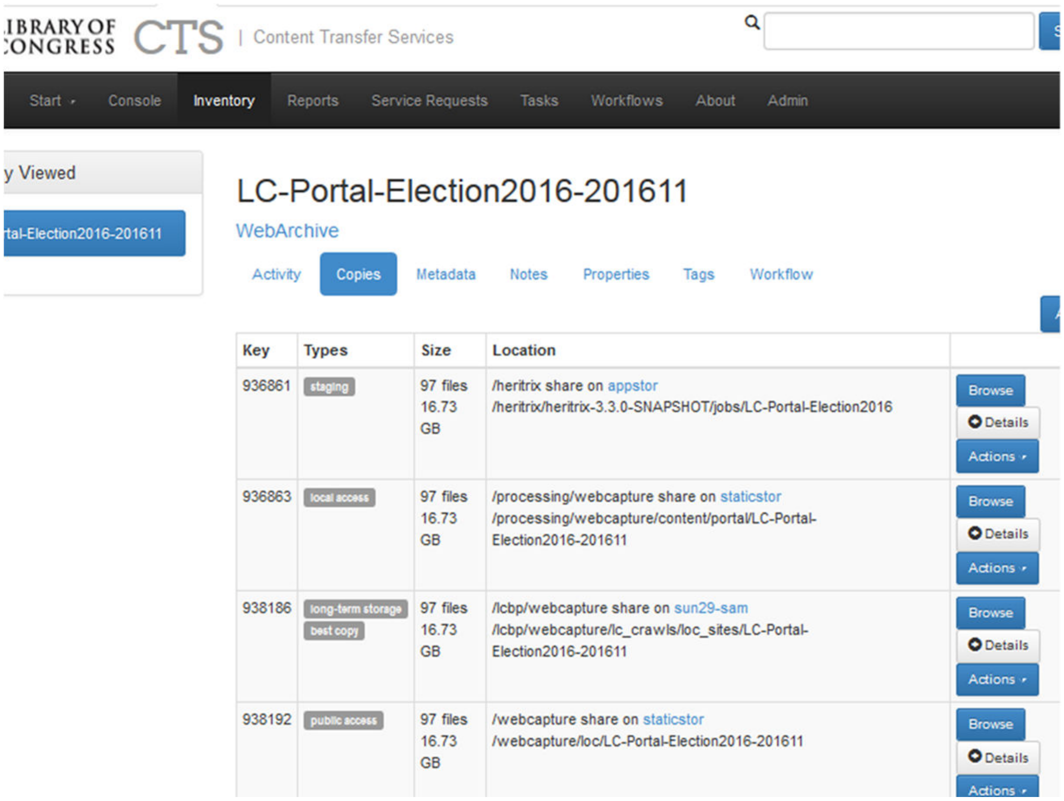


Security & Preservation

Rosie Storey, IT Specialist at Library of Congress

My background is in software development. I maintain and develop the software that picks up source material, runs it through a defined process with both automated and human tasks, and delivers it to its final destination including long-term tape archive storage and other storage if applicable. We refer to it as CTS or Content Transfer Services. I like to think of it as digital library repository services, or as the interface to our data center. Our data center at LC is a heterogeneous environment with several different operating systems, storage types, backup strategies, availability requirements, and capacities depending on the needs of the collection being stored. The inventory database has over 3.7 petabytes of unique inventoried content, we have quality-reviewed and accepted millions of images through the image quality review tool, and we're currently adding 1-2 TB per day of new content.

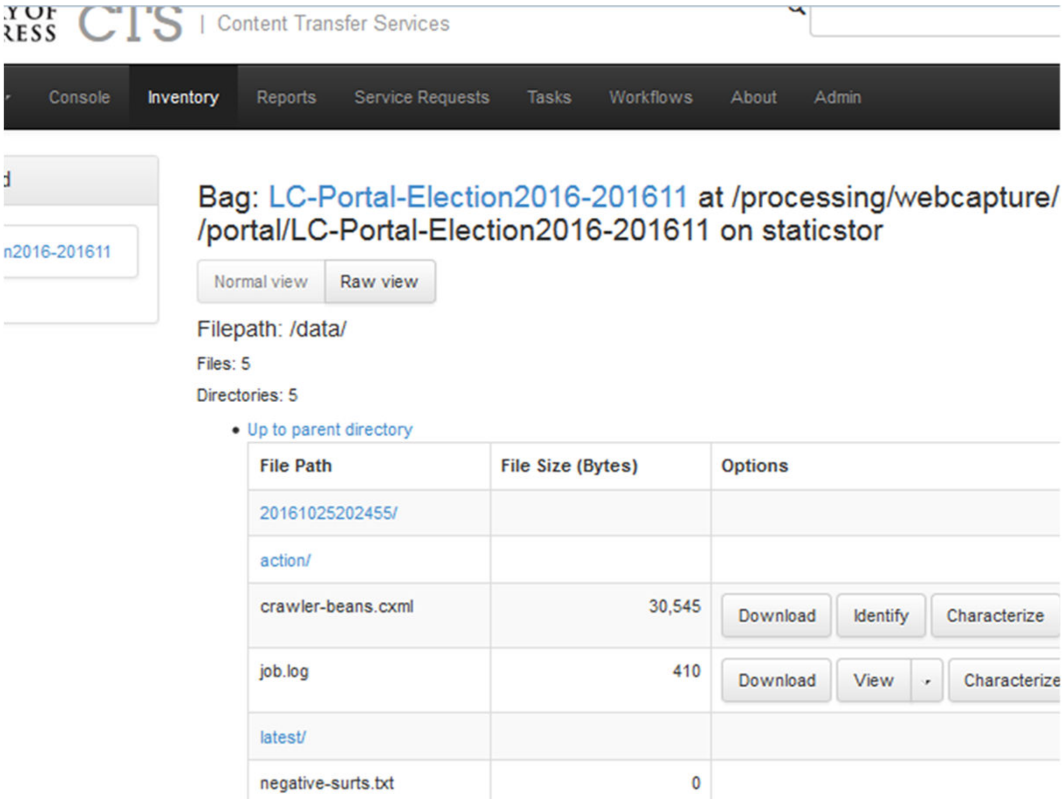




LC-Portal-Election2016-201611

WebArchive

Key	Types	Size	Location	
936861	staging	97 files 16.73 GB	/heritrix share on appstor /heritrix/heritrix-3.3.0-SNAPSHOT/jobs/LC-Portal-Election2016	Browse Details Actions
936863	local access	97 files 16.73 GB	/processing/webcapture share on staticstor /processing/webcapture/content/portal/LC-Portal-Election2016-201611	Browse Details Actions
938186	long-term storage best copy	97 files 16.73 GB	/lcbp/webcapture share on sun29-sam /lcbp/webcapture/lc_crawls/loc_sites/LC-Portal-Election2016-201611	Browse Details Actions
938192	public access	97 files 16.73 GB	/webcapture share on staticstor /webcapture/loc/LC-Portal-Election2016-201611	Browse Details Actions



Bag: LC-Portal-Election2016-201611 at /processing/webcapture/ /portal/LC-Portal-Election2016-201611 on staticstor

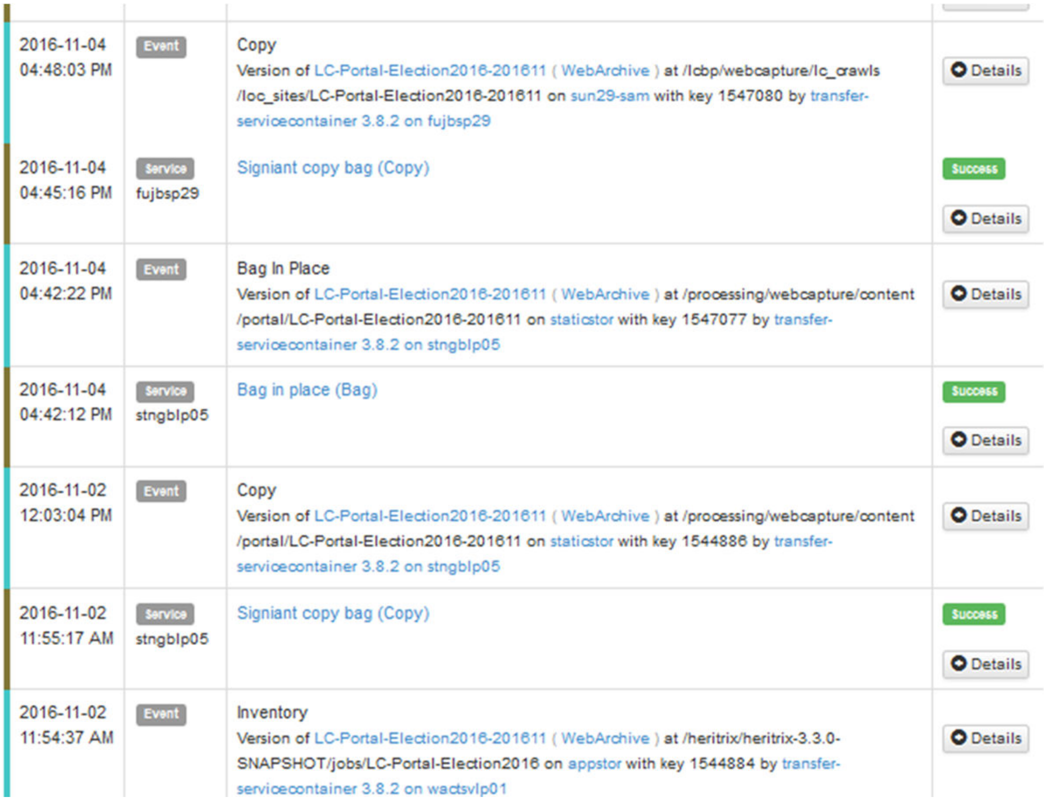
Filepath: /data/

Files: 5

Directories: 5

- Up to parent directory

File Path	File Size (Bytes)	Options
20161025202455/		
action/		
crawler-beans.cxml	30,545	Download Identify Characterize
job.log	410	Download View Characterize
latest/		
negative-surts.txt	0	



2016-11-04 04:48:03 PM	Event	Copy Version of LC-Portal-Election2016-201611 (WebArchive) at /lcbp/webcapture/lc_crawls /loc_sites/LC-Portal-Election2016-201611 on sun29-sam with key 1547080 by transfer-servicecontainer 3.8.2 on fujbsp29	Details
2016-11-04 04:45:16 PM	Service fujbsp29	Signiant copy bag (Copy)	Success Details
2016-11-04 04:42:22 PM	Event	Bag In Place Version of LC-Portal-Election2016-201611 (WebArchive) at /processing/webcapture/content /portal/LC-Portal-Election2016-201611 on staticstor with key 1547077 by transfer-servicecontainer 3.8.2 on stngblp05	Details
2016-11-04 04:42:12 PM	Service stngblp05	Bag in place (Bag)	Success Details
2016-11-02 12:03:04 PM	Event	Copy Version of LC-Portal-Election2016-201611 (WebArchive) at /processing/webcapture/content /portal/LC-Portal-Election2016-201611 on staticstor with key 1544886 by transfer-servicecontainer 3.8.2 on stngblp05	Details
2016-11-02 11:55:17 AM	Service stngblp05	Signiant copy bag (Copy)	Success Details
2016-11-02 11:54:37 AM	Event	Inventory Version of LC-Portal-Election2016-201611 (WebArchive) at /heritrix/heritrix-3.3.0-SNAPSHOT/jobs/LC-Portal-Election2016 on appstor with key 1544884 by transfer-servicecontainer 3.8.2 on wactsvlp01	Details

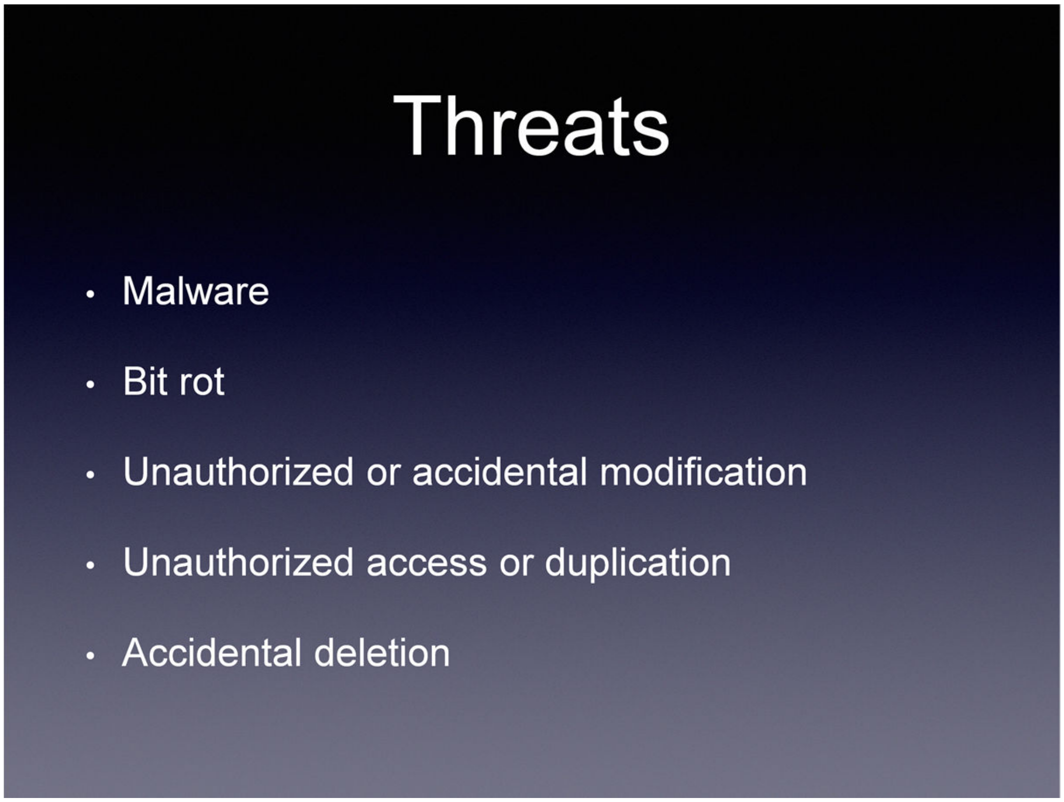
Security

- What’s the threat model?

Both preservation and security have to do with keeping collections safe. But unlike preservation, the study and practice of security assumes that your data or your business operations are under some kind of threat or attack. In these cases, we must ask ourselves: What’s the threat model? From what type of attack are you keeping the content safe?

You don’t have to be an expert in technology to think about the information security of your practices and systems. If you can ask yourself “What’s the threat model?” and have a reasonable conversation following, then you are making yourself more secure.

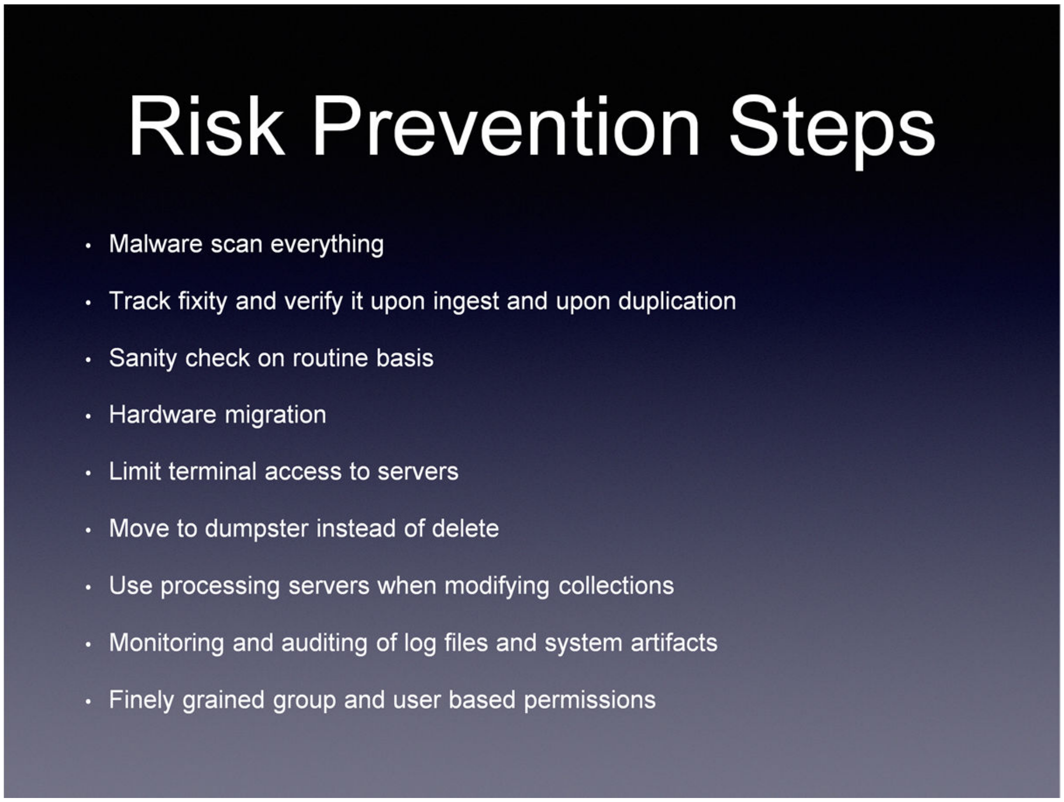
Awareness of potential risks and visibility into the status of the storage, systems, network and software that are responsible for handling collections are our best tools for ensuring security of digital collections.



Threats

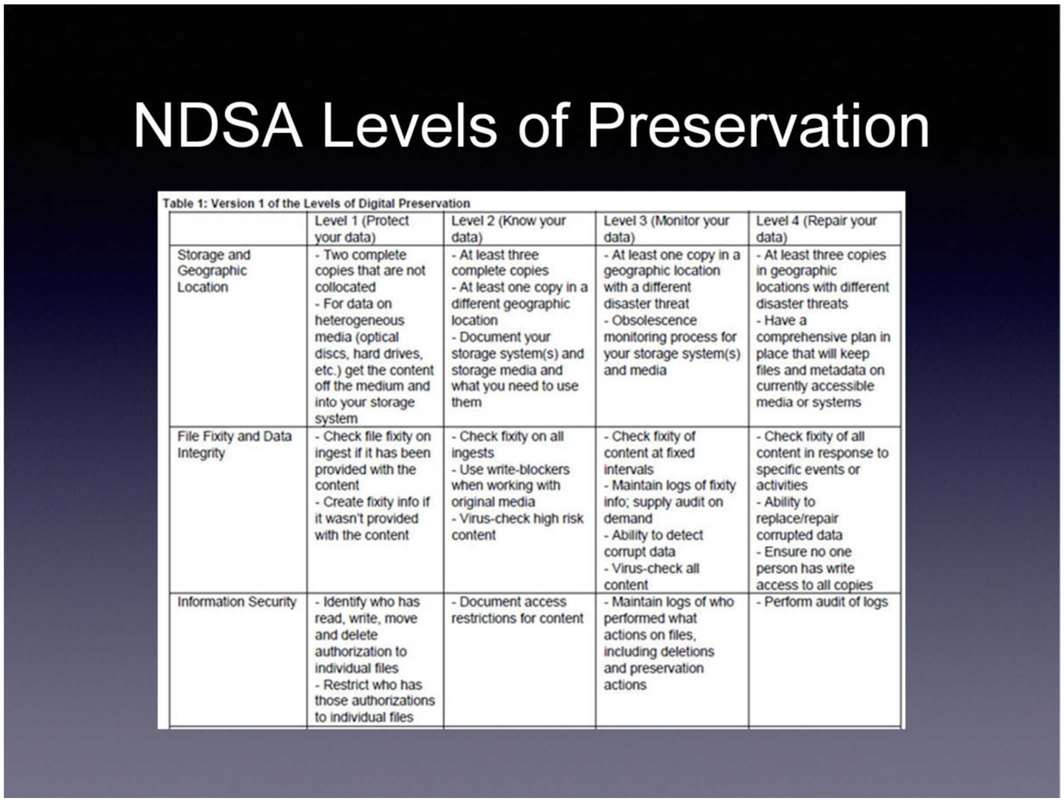
- Malware
- Bit rot
- Unauthorized or accidental modification
- Unauthorized access or duplication
- Accidental deletion

When it comes to technology, security is one of those attributes that is sometimes, mistakenly, checked against a checklist one time and then forgotten. Rather, we should practice information security continuously. Indeed, security is one aspect of technology quality, just like reliability, availability, and usability and it should be woven into any good solution. From there, as we develop and use technology solutions for our digital collections, our awareness of potential risks and visibility into status of everything should be always increasing.



Risk Prevention Steps

- Malware scan everything
- Track fixity and verify it upon ingest and upon duplication
- Sanity check on routine basis
- Hardware migration
- Limit terminal access to servers
- Move to dumpster instead of delete
- Use processing servers when modifying collections
- Monitoring and auditing of log files and system artifacts
- Finely grained group and user based permissions



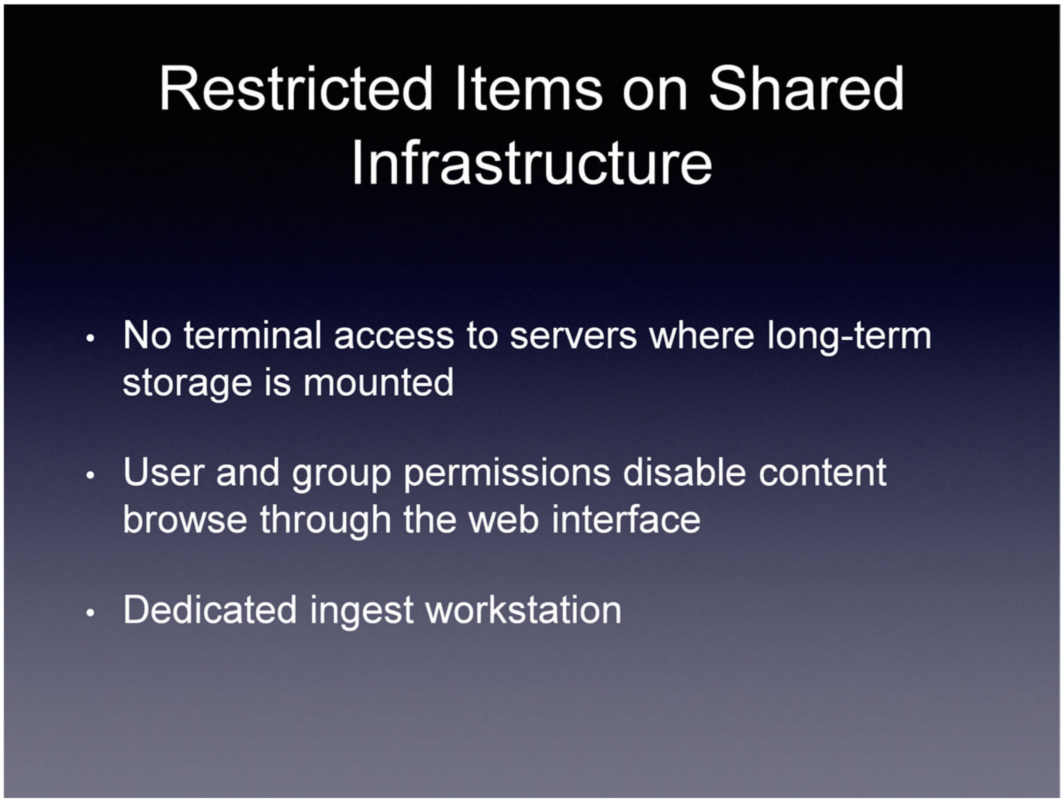
NDSA Levels of Preservation

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs

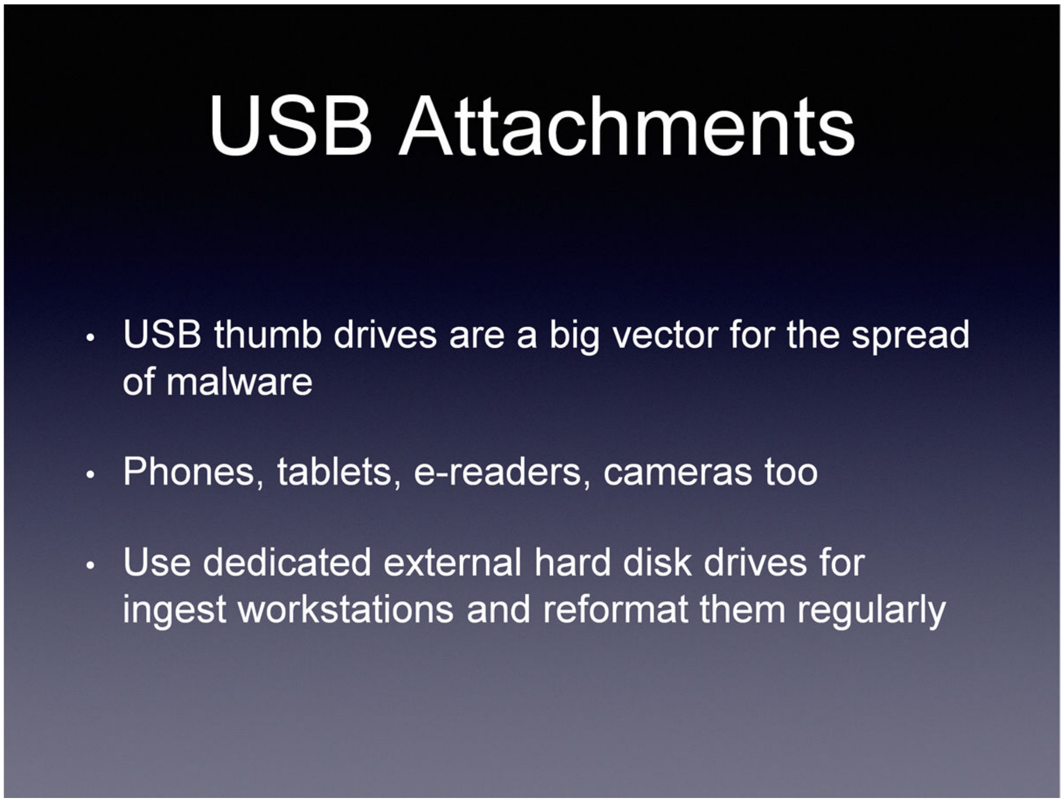
To me, all three of these rows represent security & preservation. The bottom row may be better off named something like “Information Secrecy” or something like that. Privacy of data, or keeping stuff secret, and the ability to do detective work to find out what happened to it, is just one aspect of security from the software developer / system administrator perspective. Redundant backups, failover strategy, intrusion detection, data integrity, system availability are all considerations in information security.

The other two rows not shown have to do with file formats and metadata.



Restricted Items on Shared Infrastructure

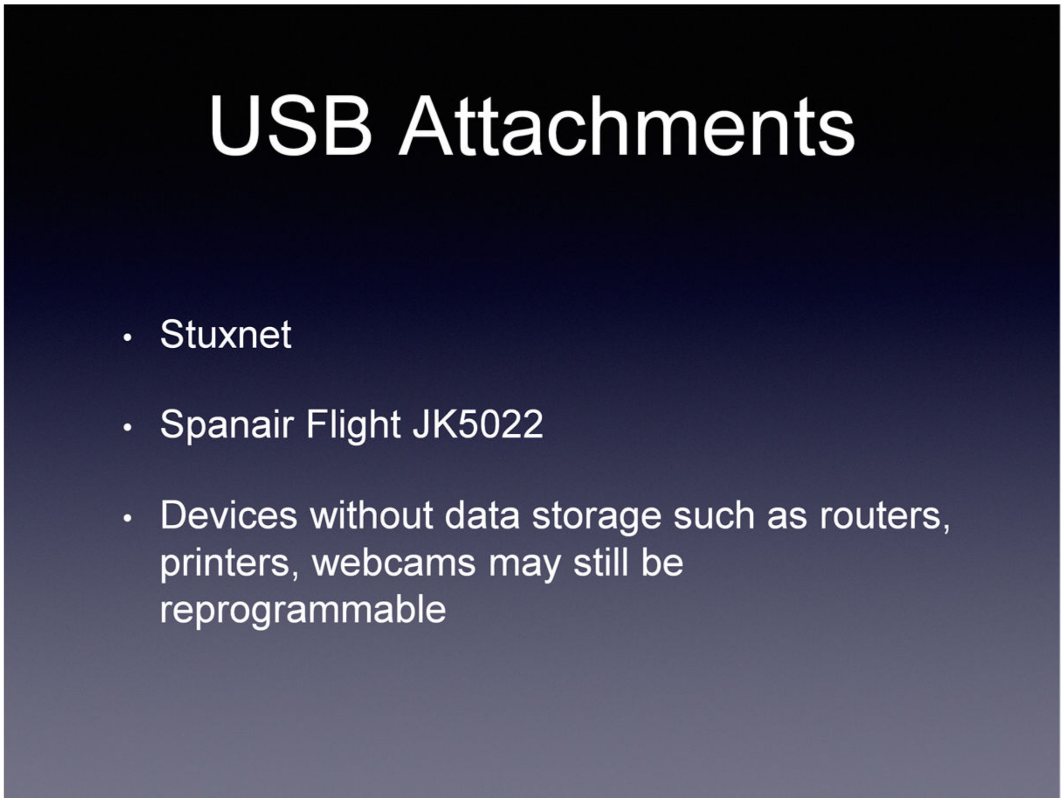
- No terminal access to servers where long-term storage is mounted
- User and group permissions disable content browse through the web interface
- Dedicated ingest workstation



USB Attachments

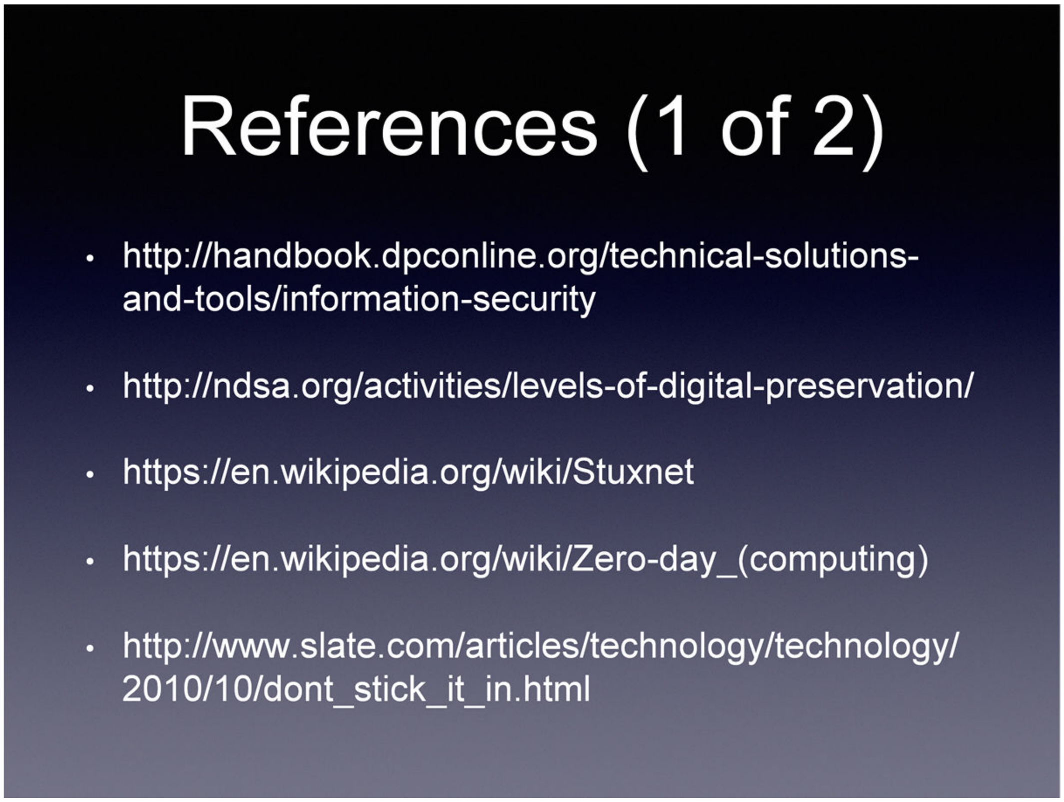
- USB thumb drives are a big vector for the spread of malware
- Phones, tablets, e-readers, cameras too
- Use dedicated external hard disk drives for ingest workstations and reformat them regularly

We restrict attachment of USB thumb drives by policy, since they have a tendency to be plugged into public computers and computers where risky activities such as email and web browsing take place. Use dedicated external hard disk drives to attach to workstations networked with the digital repository.



USB Attachments

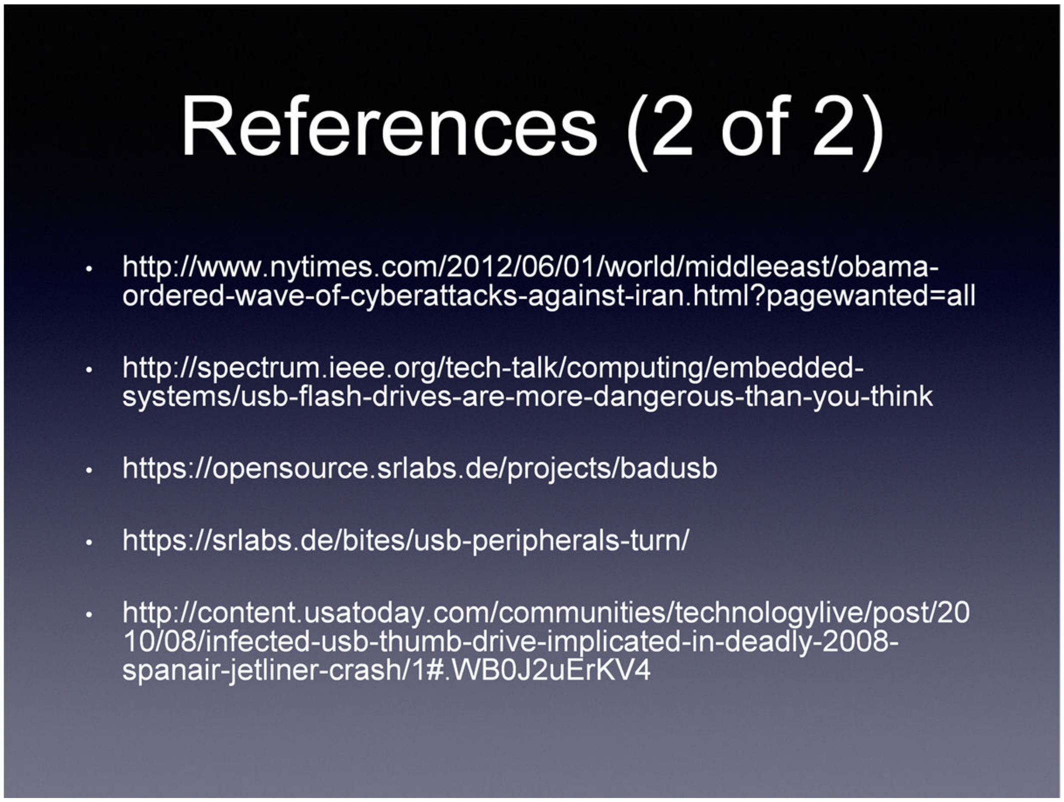
- Stuxnet
- Spanair Flight JK5022
- Devices without data storage such as routers, printers, webcams may still be reprogrammable



References (1 of 2)

- http://handbook.dpconline.org/technical-solutions-and-tools/information-security
- http://ndsa.org/activities/levels-of-digital-preservation/
- https://en.wikipedia.org/wiki/Stuxnet
- https://en.wikipedia.org/wiki/Zero-day_(computing)
- http://www.slate.com/articles/technology/technology/2010/10/dont_stick_it_in.html

http://handbook.dpconline.org/technical-solutions-and-tools/information-security
http://ndsa.org/activities/levels-of-digital-preservation/
https://en.wikipedia.org/wiki/Stuxnet
https://en.wikipedia.org/wiki/Zero-day_(computing)
http://www.slate.com/articles/technology/technology/2010/10/dont_stick_it_in.html



References (2 of 2)

- http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html?pagewanted=all
- http://spectrum.ieee.org/tech-talk/computing/embedded-systems/usb-flash-drives-are-more-dangerous-than-you-think
- https://opensource.srlabs.de/projects/badusb
- https://srlabs.de/bites/usb-peripherals-turn/
- http://content.usatoday.com/communities/technologylive/post/2010/08/infected-usb-thumb-drive-implicated-in-deadly-2008-spanair-jetliner-crash/1#.WB0J2uErKV4

http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html?pagewanted=all
http://spectrum.ieee.org/tech-talk/computing/embedded-systems/usb-flash-drives-are-more-dangerous-than-you-think
https://opensource.srlabs.de/projects/badusb
https://srlabs.de/bites/usb-peripherals-turn/
http://content.usatoday.com/communities/technologylive/post/2010/08/infected-usb-thumb-drive-implicated-in-deadly-2008-spanair-jetliner-crash/1#.WB0J2uErKV4