

Archival Storage vs. Data Backup



WHAT'S THE DIFFERENCE?

**JENNIFER RICKER
DIGITAL COLLECTIONS MANAGER
STATE LIBRARY OF NORTH CAROLINA**

Overview



- **Why store at all?**
- **What IS the difference?**
 - General differences
 - Do we need to do both?
- **Do they have different IT requirements?**
 - Can one system do both?
- **What else do I need from my backup storage?**
- **What else do I need from my archival storage?**
- **Summary**

Why Store at All?



Hardware errors

Human error

Storage media degradation

Hardware/software obsolescence

Viruses

PREVENT DATA LOSS

Lost or stolen computers

Accidental deletions

Natural disasters (floods, fires, quakes, or storms)

Power failures and surges

Malicious activity

What IS the Difference?



- **Data Backups**

- Short term (kept for days or weeks then overwritten)
- Disaster recovery for ALL data
- Restoring lost/corrupted data

Image: CompRearSmall11 by pyroclastichawk on Flickr.



- **Archival Storage**

- Long term (kept for years or forever)
- Discovery/Access to SELECT data
- Compliance/e-discovery/file history

Image Archive on the American Eugenics Movement

The screenshot shows the website's search and browse interface. At the top right is a magnifying glass icon with the word "SEARCH" inside. Below it is a search input field with the placeholder text "Enter keyword(s) or ID# (e.g., id735)" and a "Search" button. To the left of the search field is a yellow sticky note with the following text: "SEARCH To search for an image, enter a keyword or image ID number in the text box. or BROWSE The Archive image database by clicking a category below. Mouse over the icon for a description of the category." Below the sticky note are four categories: "Topics" with a book icon titled "ETHIC, GENETIC AND RACIAL THEORY" and a family pedigree chart; "Object Type" with a stack of photographs; "BROWSE a category: Time Period" with a calendar icon showing dates from 1915 to 1933; and "Archive" with an open box containing various documents and photos.

Image: <http://www.eugenicsarchive.org/eugenics/list2.pl>

What IS the Difference?



Image: Five Day's Backup by daryl_mitchell on Flickr.

VS.



Image: Internet Archive's server farm from Archive.org.

What IS the Difference?



- Why not just keep backups longer & use them as the archival storage?
 - Requires restoring to a particular date and extracting files that meet criteria
 - Rinse
 - Repeat



Image: washing my dog by strelitzia on Flickr.

What IS the Difference?



- What data they store

- Active vs. inactive



Image: Sleeping again by Miikka Skaffari on Flickr.

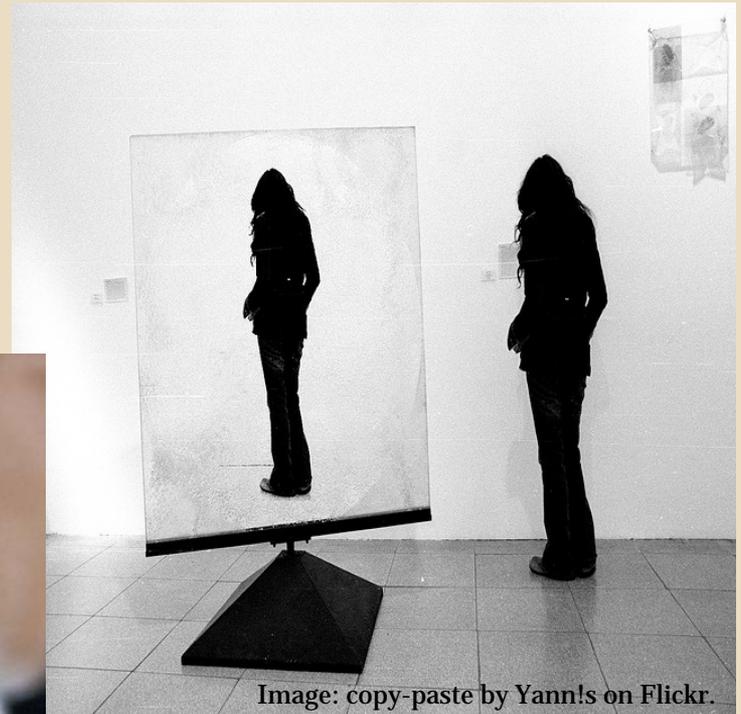


Image: copy-paste by Yann!s on Flickr.

- Copies vs. originals

What IS the Difference?



- **When they store the data**
 - Scheduled batch process vs. triggering event



Image: found umbrellas by Paul Keller on Flickr.

- **How they handle metadata**
 - Not extracted from file nor supplemented vs. extracted, supplemented, and stored in database (i.e., searchable)

What IS the Difference?



Do we need to do both?

Absolutely! They serve completely different purposes and both are absolutely necessary.

Do they have different IT requirements?



Data Backup	Archival Storage
	Data authenticity
	Indexing capabilities
High media capacity	Extended media longevity
High performance read/write streaming	High performance random read access
Low storage cost	Low total cost of ownership



Image: "Technology" -- A gift that keeps on giving by Curtis Palmer on Flickr.

Do they have different IT requirements?



Can one system do both?

Absolutely! Just make sure it does actually do what you need it to do for both.

What else do I need from my backup storage?



- **Full or differential**
- **Audits**
- **Location**
 - Local or remote
 - ✦ transferring large volumes to remote servers can take time though
 - Multiple and geographically disparate copies may be appropriate
 - Self-managed or outsourced
- **Security**
 - Encryption for confidential information

What else do I need from my archival storage?



- **OAIS compliance**
 - **File integrity/authenticity**
 - ✦ Virus checking/quarantine, hashing, and versioning
 - ✦ File migration capability
 - ✦ Audits
 - **Security**
 - ✦ Limited write capabilities
 - ✦ Protection for confidential information
 - **Self-managed or outsourced**
 - **Geographically disparate backups (yep...backups)**
 - ✦ Ideally with synching capabilities
 - **Scalability**
 - ✦ Persistent identifiers
 - ✦ Storage media refreshing
 - **Information Lifecycle Management**
 - ✦ Integration into content creation systems
 - ✦ Implement storage management policies

Summary



Data Backup	Archival Storage
Operational backup and disaster recovery	Regulatory compliance and legal records retention
Cost efficient	Cost efficient
No individual file	Retrieval
No additional metadata	Metadata
Short retention	Long retention
Scheduled	Event
High capacity, fast I/O w/ streaming performance storage media	Authenticated, long lived, high random access performance storage media
IT responsibility	Executive and legal responsibility

Remember...

Archival Storage

must be backed up!

Thank you!



Questions?

jennifer.ricker@ncdcr.gov