

# Digital Storage Strategy: A Crash Course

11/19/2018

Ethan Gates

November Webinar for the Memory Lab Network

# Agenda

- Introduction (~5 min.)
- Risk Factors and Causes of Digital Loss (~15 min)
- Strategy and Planning (~20 min)
- Practical Options for Storage and Recovery (~30 min)
  - Physical Media/Formats
  - Cloud and Online Storage
  - Software
  - Data Recovery
- Q&A (~15 min.)

# Risk Factors and Causes of Digital Loss

**What are we scared of?**

- Hardware failure
  - Software failure/threats
  - Degradation and bit rot
  - Natural disasters
  - Human error
-

# Hardware Failure

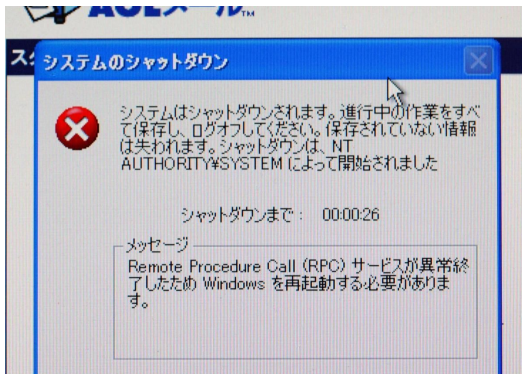
- Mechanical failure
- Electrical failure
- Obsolescence (physical interfaces, peripherals, etc.)



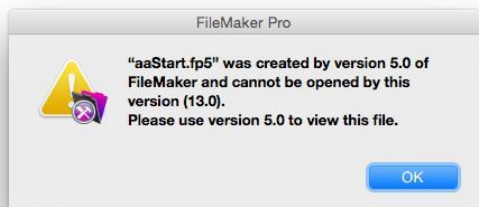
shutterstock

IMAGE ID: 46257946  
www.shutterstock.com

# Software Failure/Threats

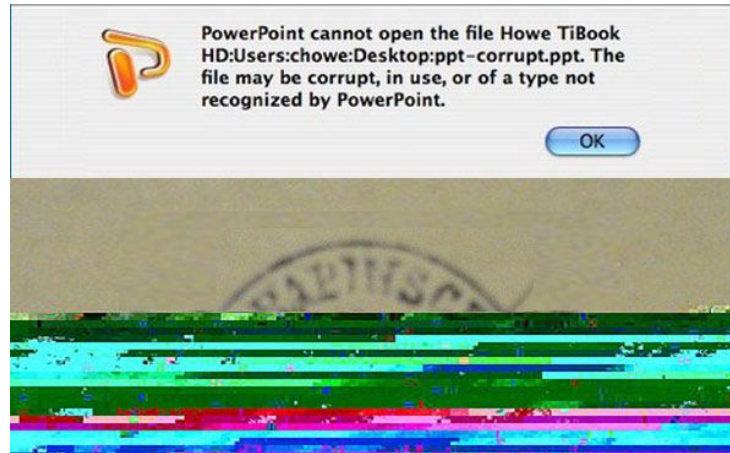


- Obsolescence (“software rot”)
- Incompatibility (between operating systems, file systems, more)
- Lack of transparency
- Viruses/malware



# Degradation and Bit Rot

- Physical deterioration of media (binder, dyes, etc.)
- Loss of magnetic orientation (hard disk, magnetic tape)
- “Bit rot” - gradual corruption of data due to accumulation of non-critical failures  
- which is to say, ￣\\_(\ツ)\\_/



# Natural Disasters

- Fire, earthquake, flooding
- Threat to both local and networked storage
- Affect manufacturing as well

## Computing

### Thailand's devastating floods are hitting PC hard drive supplies, warn analysts

Country is centre of world hard drive manufacture but factories have been seriously affected by floods, which could affect supplies early in 2012



Thai residents make their way through a flooded street in Pathum Thani, Thailand. Photograph: Daniel Berehulak/Getty Images



Home > Data Center

NEWS ANALYSIS

## Hurricane Sandy leaves wounded servers in its wake

As disaster recovery firms struggle to restore damaged data centers, experts warn of further storm-related breakdowns in the months ahead.



By Patrick Thibodeau | Follow

Senior Editor, Computerworld | NOV 19, 2012 6:00 AM PT



Data recovery experts have been busy in the wake of Hurricane Sandy, which left a slew of data centers underwater, damaging equipment and posing a significant threat to business-critical data.

Apparently disregarding weather forecasters' widespread warnings and underestimating the power of the storm that hit the East Coast late last month, many businesses didn't begin moving computer and IT communications equipment out of harm's way until it was too late, say officials at companies that specialize in data recovery.

Many data centers were casualties of the massive storm, and the damage threatened to shut down major New York-based businesses and interrupt Internet service across the country, according to experts.

For instance, the storm forced two so-called carrier hotels -- monolithic buildings that serve as major U.S. network hubs -- in lower Manhattan to operate on generator power for a significant period of time.

### MORE LIKE THIS

Sandy wounded servers, some grievously, say services firms

Storm forces Internet hubs to run on generator power

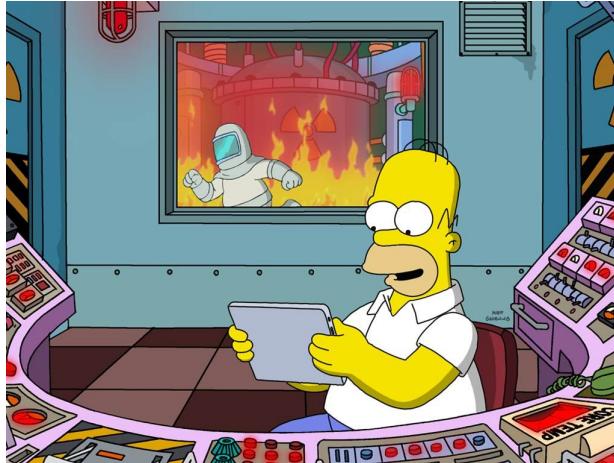
Drama in NYC as data center temp passes 100 degrees



VIDEO  
Need a data center location? Iceland wants you

# Human Error

- Poor handling of physical media
- Unintentional deletion/overwrite/modification





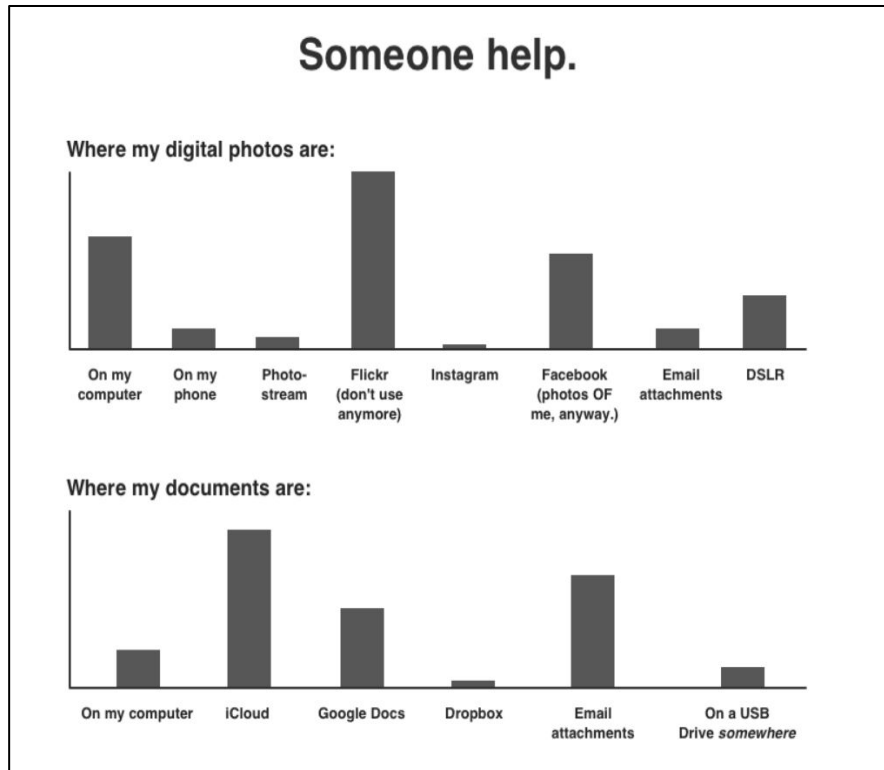
# Strategy and Planning

**What actions can we take to avoid loss?**

- Intellectual Control
  - Redundancy
  - File Fixity
  - Migration
  - Tech Support
-

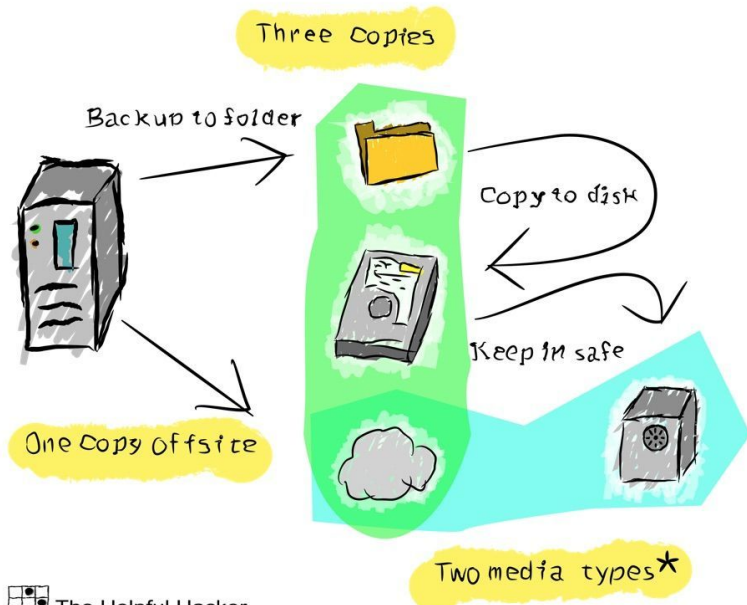
# Intellectual Control

- Inventory
- File Naming



# Redundancy

## The "321" Rule



- Lots of Copies Keeps Stuff Safe (LOCKSS)
- 3-2-1 strategy:
  - 3 total copies
  - 2 local, on different devices
  - 1 "offsite" (geographically separated)



The Helpful Hacker

<http://thehelpfulhacker.net>

\*Yes, technically "the cloud" is probably using a harddisk too

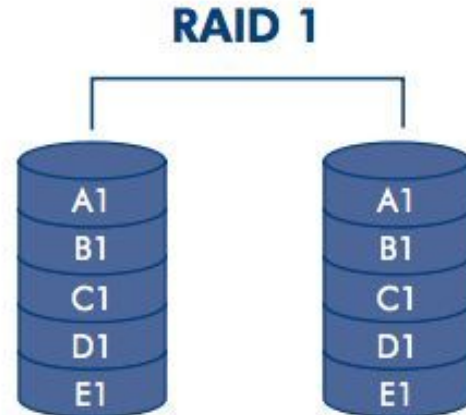
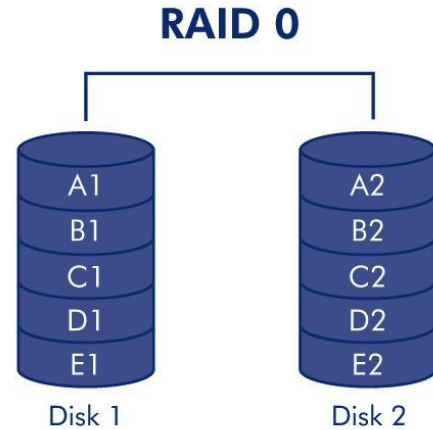
# Fault Tolerance

- RAID (Redundant Array of Independent Disks) - make multiple disks into one logical volume



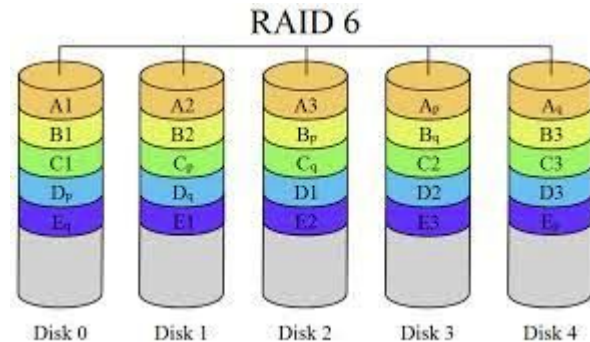
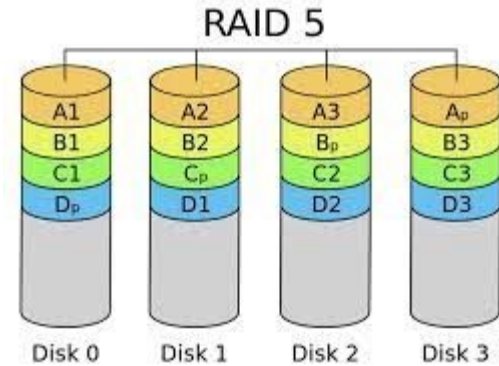
# RAID Levels

- JBOD (Just a Bunch Of Disks)
  - Not really a RAID
  - Boosts capacity, not redundancy
- RAID 0
  - Data striping
  - Improves read/write speed
  - Maintains total physical capacity
  - No redundancy - lose one disk, lose all data
- RAID 1
  - Data mirroring
  - Slower write speed
  - Half total physical capacity
  - Ensures data protection



# RAID Levels

- RAID 5
  - At least 3 drives
  - If one drive fails, RAID can rebuild all data
  - Hot swappable
- RAID 6
  - At least 5 drives
  - If two drives fail, RAID can rebuild all data
  - Hot swappable



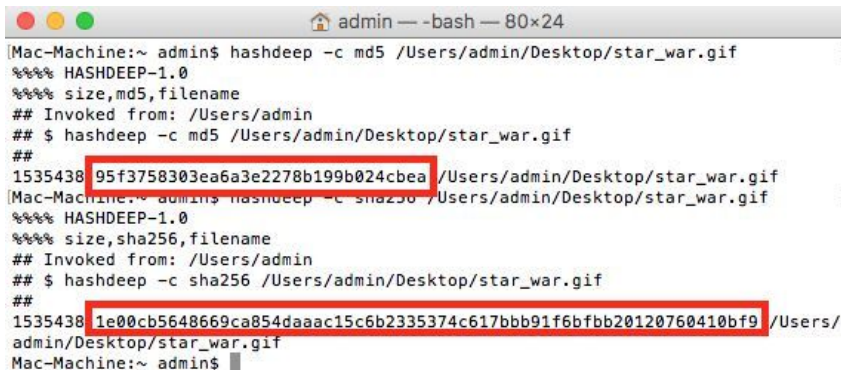
# Migration

- Regular, scheduled movement to new drives/media (~5 years)
- Maintain fixity during transfer



# File Fixity

- Make sure files don't change (corrupt) over time
- Accomplished via computing and validation of checksums
- Multiple checksum hash generators are common, but often aimed at encryption
  - MD5
  - Security Hash Algorithm
    - SHA-1
    - SHA-2 (SHA-256, SHA-512)



A terminal window titled 'admin -- bash -- 80x24' showing the output of the 'hashdeep' command. The first command is 'hashdeep -c md5 /Users/admin/Desktop/star\_war.gif', which outputs the MD5 checksum '95f3758303ea6a3e2278b199b024cbea'. The second command is 'hashdeep -c sha256 /Users/admin/Desktop/star\_war.gif', which outputs the SHA256 checksum '1e00cb5648669ca854daaac15c6b2335374c617bbb91f6bfbb20120760410bf9'. Both checksums are highlighted with red boxes.

```
Mac-Machine:~ admin$ hashdeep -c md5 /Users/admin/Desktop/star_war.gif
#### HASHDEEP-1.0
#### size,md5,filename
## Invoked from: /Users/admin
## $ hashdeep -c md5 /Users/admin/Desktop/star_war.gif
##
1535438 95f3758303ea6a3e2278b199b024cbea /Users/admin/Desktop/star_war.gif
Mac-Machine:~ admin$ hashdeep -c sha256 /Users/admin/Desktop/star_war.gif
#### HASHDEEP-1.0
#### size,sha256,filename
## Invoked from: /Users/admin
## $ hashdeep -c sha256 /Users/admin/Desktop/star_war.gif
##
1535438 1e00cb5648669ca854daaac15c6b2335374c617bbb91f6bfbb20120760410bf9 /Users/
admin/Desktop/star_war.gif
Mac-Machine:~ admin$
```



# Tech Support

- Commercial
- Community



# Practical Options for Storage

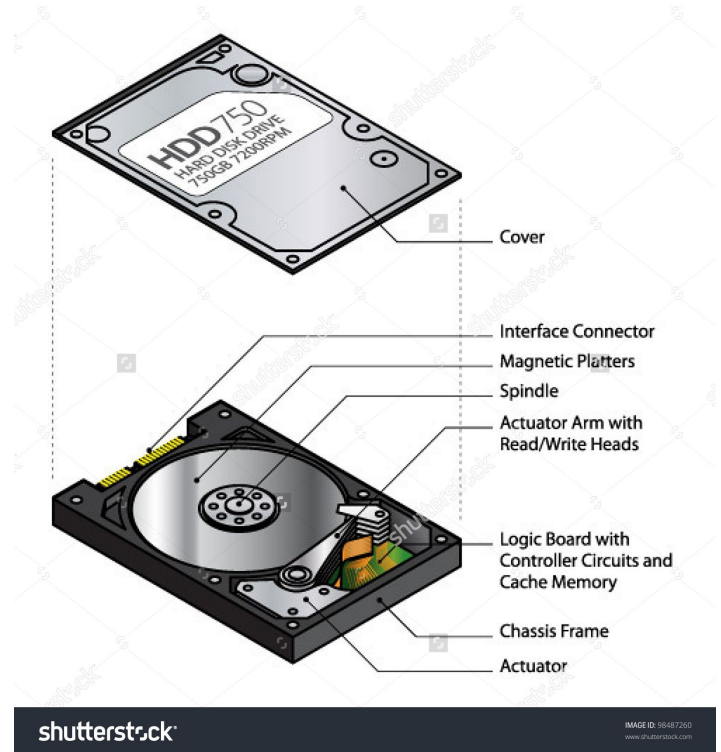
**What are our options?**

- Physical Media/Formats
- Cloud and Online Storage
- Software

---

# Hard Disk Drives (HDDs)

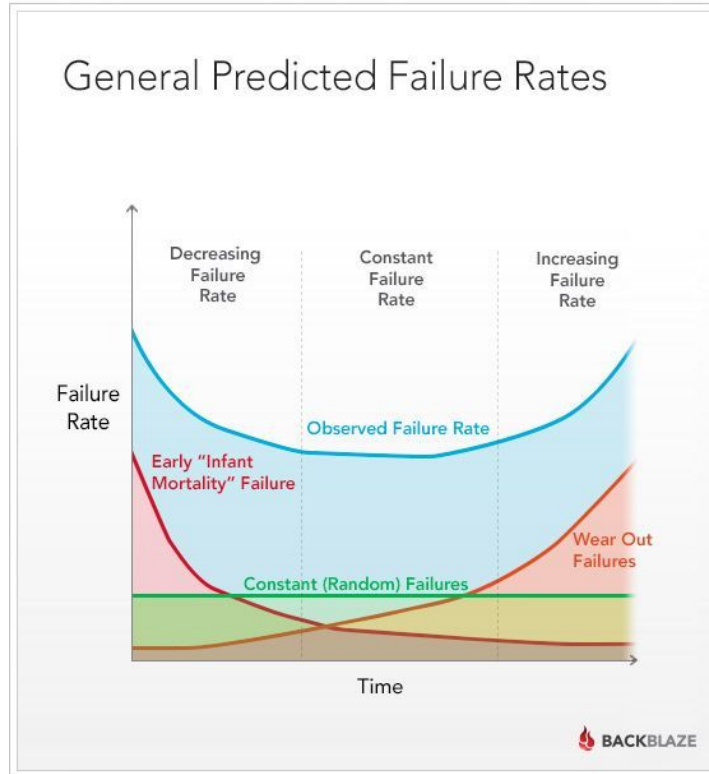
- Mechanical-magnetic (platter, spindle, head)
- Capacity still steadily increasing vs. cost
- “Bathtub curve” failure



shutterstock

IMAGE ID: 98487269  
www.shutterstock.com

# HDD Failure Rates



## Hard Drive Failure Rates by Manufacturer

For 2016, as of 12/31/2016

MFG	Drive Count	Drive Days	Drive Failures	Failure Rate
HGST	24,545	8,028,068	132	0.60%
Seagate	45,531	14,194,485	1,029	2.65%
Toshiba	237	86,308	3	1.27%
WDC	1,626	573,321	61	3.88%
<b>Totals</b>	<b>71,939</b>	<b>22,882,182</b>	<b>1,225</b>	<b>1.95%</b>



## Hard Drive Failure Rates by Drive Size

For 2016, as of 12/31/2016

Drive Size	Drive Count	Drive Days	Drive Failures	Failure Rate
3 TB	6,605	2,416,353	93	1.40%
4 TB	54,189	18,504,977	1,042	2.06%
5 TB	45	16,425	1	2.22%
6 TB	2,335	850,992	41	1.76%
8 TB	8,765	1,093,435	48	1.60%
<b>Totals</b>	<b>71,939</b>	<b>22,882,182</b>	<b>1,225</b>	<b>1.95%</b>



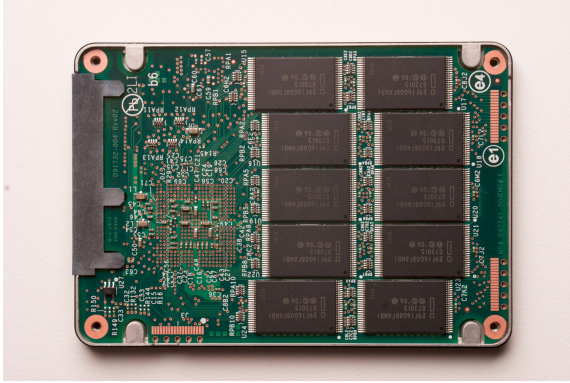
## Hard Drive Annualized Failure Rates for 2016

Reporting period 1/1/2016 - 12/31/2016 inclusive

MFG	Model	Drive Size	Drive Count	Avg Age (months)	Drive Days	Drive Failures	Failure Rate
HGST	HUH728080ALE600	8 TB	45	22.99	16,155	-	0.00%
Seagate	ST8000DM002	8 TB	8,660	4.72	1,075,720	48	1.63%
Seagate	ST8000NM0055	8 TB	60	1.44	1,560	-	0.00%
Seagate	ST6000DX000	6 TB	1,889	21.48	684,840	16	0.85%
WDC	WD60EFRX	6 TB	446	24.14	166,152	25	5.49%
Toshiba	MD04ABA500V	5 TB	45	22.15	16,425	1	2.22%
HGST	HDS5C4040ALE630	4 TB	2,625	45.35	987,011	14	0.52%
HGST	HMS5C4040ALE640	4 TB	7,014	29.48	2,579,698	28	0.40%
HGST	HMS5C4040BLE640	4 TB	9,407	15.51	2,436,130	34	0.51%
Seagate	ST4000DM000	4 TB	34,738	21.73	12,359,750	938	2.77%
Seagate	ST4000DX000	4 TB	184	38.54	72,615	27	13.57%
Toshiba	MD04ABA400V	4 TB	146	20.61	52,983	-	0.00%
WDC	WD40EFRX	4 TB	75	17.16	16,790	1	2.17%
HGST	HDS5C3030ALA	3 TB	4,476	55.87	1,647,137	34	0.75%
HGST	HDS723030ALA	3 TB	978	61.21	361,937	22	2.22%
Toshiba	DT01ACA300	3 TB	46	44.12	16,900	2	4.32%
WDC	WD30EFRX	3 TB	1,105	30.39	390,379	35	3.27%
<b>Totals</b>			<b>71,939</b>		<b>22,882,182</b>	<b>1,225</b>	<b>1.95%</b>



# Solid-State Drives (SSDs)/Flash



- Electrical (non-mechanical)
- Can lose charge due to poor electrical insulation or frequent use over time (relatively poor “write-endurance”)
- Limited capacity/more expensive to manufacture





# Optical Discs

- Hardware becoming less common in consumer environments
- Prone to mishandling
- Lifespan depends on materials/dyes used, storage environment
- Limited capacity compared to other media



# Magnetic/Data Tape (LTO)

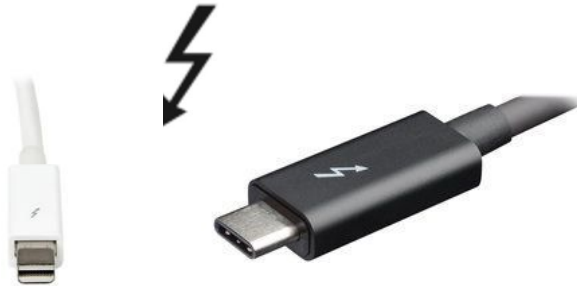
- Shares physical concerns of all magnetic tape formats
- Large capacity, quick write speed
- Sequential access = slow read speeds
- Requires specialized equipment, far more enterprise than consumer support





# Physical Interfaces/Data Transfer Protocols

- USB
- FireWire/Thunderbolt
- Ethernet/network



# Cloud and Online Storage

- Cheap
- Handled by third-party vendors
- Ability to access/manage/download files will vary depending on service
- No control over physical management



# Social Media

- Sharing, not saving
- Terms of Service
- Ad targeting

**Let's be candid.** Flickr at its best is a place to connect, to discover, and to evolve as photographers and lovers of photography. This is the world's largest photographer-focused community. Here, together, this newly-independent community can shape the future of photography itself.



Today we are announcing updates to Flickr's  
Pro and Free account offerings.

# Software

- Closed source versus open source
- Commercial versus free
- Look for wide adoption



# Data Failure and Recovery

What do I do if it's “too late”?

- Signs of Data Failure/Loss
- Data Recovery

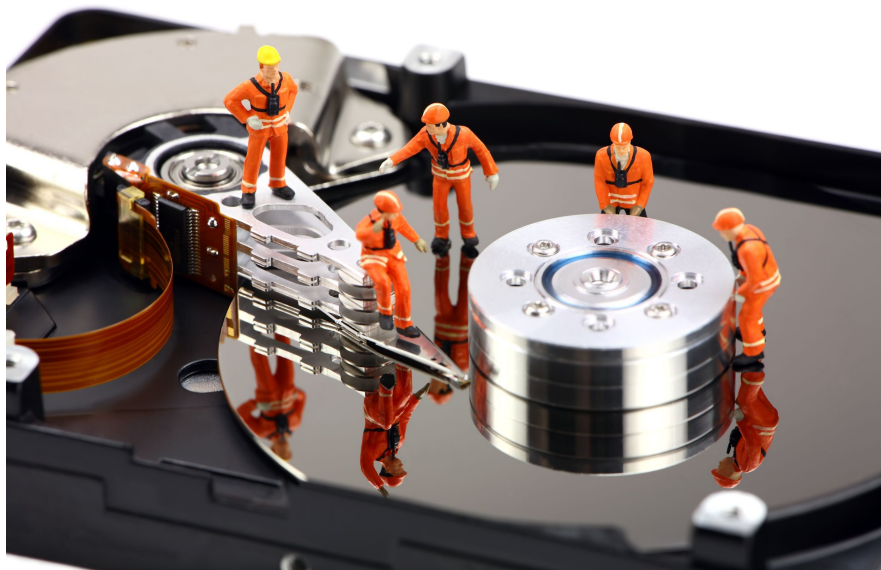
---

# Signs of Data Failure or Loss

- Trust your eyes
  - Error messages
  - Missing files
  - Folders “hanging” when you try to open them
- Trust your ears
  - Hard drive not spinning at all
  - Hard drive makes a beeping sound when you try to power it on
  - Drive makes unusual scratching, clicking or other noise

# Data Recovery

- Plan ahead, plan ahead, plan ahead
- DIY software options
  - Cost
  - User-friendliness
  - OS compatibility
- Commercial services
  - Cost
  - Transparency
  - Proximity



# Choosing Your Storage Strategy

How do I break it down?

- Total size of your stuff (GB? TB?)
- Growth
- Budget
- Users
- Vendor/developer support
- Community adoption/support

---



# Resources

## Data recovery software:

- [ddrescue](#) (all systems)
- [Recuva](#) (Windows)
- [Stellar Phoenix Data Recovery](#) (MacOS + Windows)
- [EaseUS Data Recovery](#) (MacOS + Windows)
- [PhotoRescue](#) (MacOS + Windows)

## File fixity tools:

- [Fixity](#) (MacOS + Windows)
- [File Verifier++](#) (Windows)

## File transfer:

- [Exactly](#) (MacOS + Windows)
- [Bagger](#) (all systems)
- [Grsync](#) (all systems)

# Resources

Video storage calculators:

- <https://www.digitalrebellion.com/webapps/videocalc>
- [AJA DataCalc app](#) (iOS and Android)

# Contact

Ethan Gates

Software Preservation Analyst, Yale University Library

Scaling Emulation as a Service Infrastructure (EaaS)

[ethan.t.gates@gmail.com](mailto:ethan.t.gates@gmail.com)

Twitter: @The\_BFOOL

Mastodon: @The\_BFOOL@digipres.club

[ethan-gates.com](http://ethan-gates.com)

[patchbay.tech](http://patchbay.tech)