



Archiving Digital Objects

Simon Kravis, Principal



Why Digital?

- Any kind of data can be stored digitally – text, images, video, audio
- Very economical - exploits miniaturisation and mass production
- Perfect copying is possible
- Perfect preservation if bits are preserved and delivery mechanism available
- Data storage may be remote from delivery mechanism

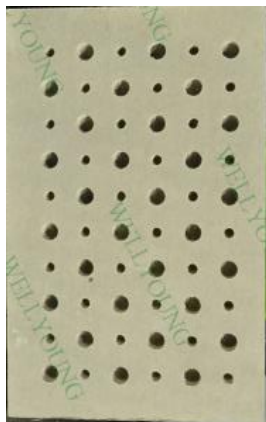


Digital Storage Miniaturisation

4000 Year old
Sumerian clay
tablet

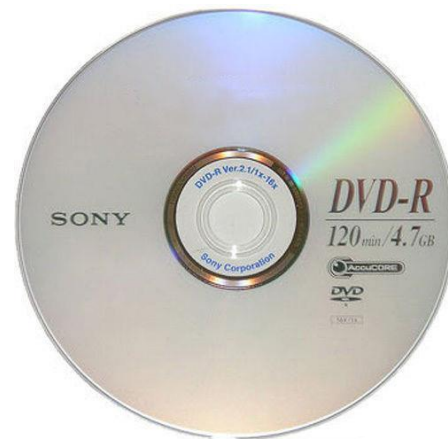


Big hole = 1
Small hole = 0
60 bits



x 626 million =

DVD
37 billion bits
1500 digital photos



If digital tablet weighed = 100 g, storing digital data that would fit on 60 DVDs would require a weight of tablets equivalent to the weight of the Great Pyramid of Giza



Digital Downsides

- Separation of storage from delivery
 - Delivery mechanisms (computers) are very complex and prone to obsolescence
- Archival storage requires both digital data and delivery mechanism
 - Future delivery mechanisms may not be able to decode digital data
 - Digital data storage media may have degraded or become inaccessible
- Huge data volumes – 2 trillion photos taken per year
- Archiving digital objects requires much more technical knowledge than physical artefacts



Personal Digital Archiving Resources

- Books
 - [“Managing the Digital You”](#) (Melody Condron)
 - [“The Complete Guide to Personal Digital Archiving”](#) (Brianna Marshall)
- Conferences
 - HASTAC (Humanities, Arts, Science and Technology Alliance and Collaboratory)
- Online Guides
 - Universities, Libraries, Cultural Organisations



Comments on Resources

- Strong on processes and organisational principles
- Common features
 - Assess importance of data items
 - Make a plan
 - Store data in more than 1 place – sharing helps!
 - Advice may be dated or unsound on technical issues



Storage Media

| Storage Medium | Melody Condron's Advice | My Advice |
|--------------------------------|--|--|
| External Magnetic Disk Drives | 20% of drives fail after 4 years. Replace every 3 years. | Analysis based on continuously spinning drives. Drives will last much longer if kept mostly unpowered. Replacement period much longer. |
| Optical Disks | Not recommended. Risk of scratching, coating flakes off (based on report), UV damage, limited rewrites | Recommended through lack of electronics and robustness. Scratches can be polished out with Brasso, flaking based on report exposing disks to extreme conditions not found in domestic storage. |
| Cloud Storage | Recommended in conjunction with other storage media | The same. Possibility of access failure, need to store credentials, and make periodic payments if required. |
| Flash (thumb) drives, SD cards | Transfer only. 5-10 year life. | The same |



Digital Preservation for Future Genealogists

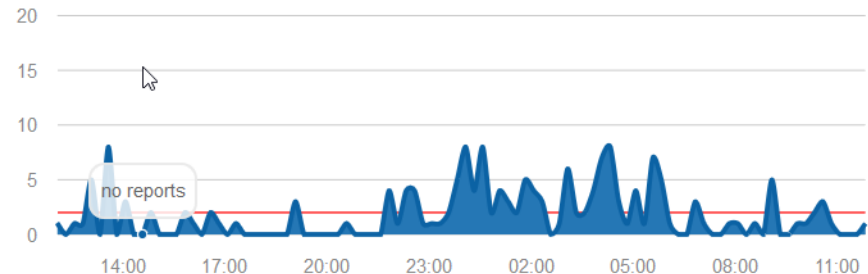
- [My article on this topic](#)
- Availability of data migration services a key factor for individuals as media become obsolete
 - Use common data formats – e.g. jpeg, mp3
 - Format obsolescence less of a problem than expected
- Store data on
 - DVD or optical disks (no electronics)
 - Removable magnetic USB drives kept unpowered, but occasionally plugged in and spun up
 - Cloud server in addition to local physical copies
- Fire/water proof storage useful for local copies



Storage Media - Cloud

- Someone else's computer anywhere in the world
- Access failures rare but not absent
- Free quota, then pay per Gbyte
- What happens if payments stop?
- Preserving username and password

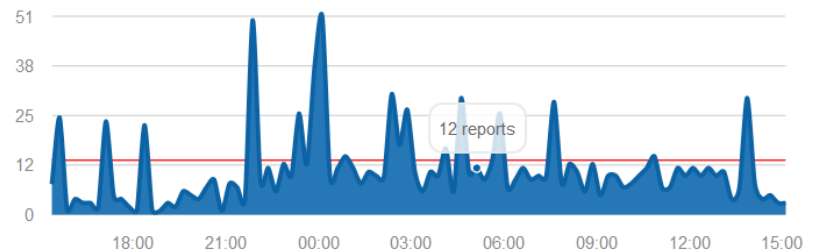
Google Drive problems last 24 hours



Most reported problems:

- Can't access files online (45%)
- App not loading (42%)
- File syncing (11%)

Facebook problems last 24 hours



[I have a problem with Facebook](#)

[Check past issues](#)

Most reported problems:

- Newsfeed (53%)
- Total blackout (25%)
- Log in (20%)

Resolved issues:

- **26 April:** [Problems at Fac](#)
- **25 April:** [Problems at Fac](#)
- **24 April:** [Problems at Fac](#)



Storage Media – Removable

Magnetic USB drives

- Electro-mechanical devices which store bits as areas of magnetisation in disk
- Technology rapidly changing to increase capacity – currently ~100 Mbytes/sq in
- USB sockets disappearing from modern laptops
- Be gentle on USB connectors
- Avoid condensation

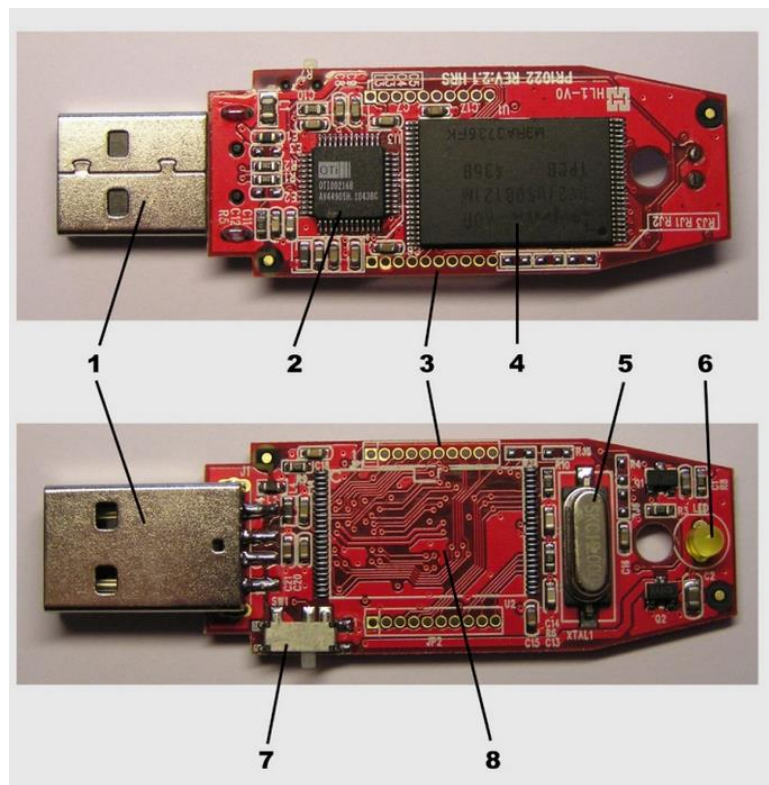


- Drivers may not be available for newer operating systems
- Lifetime if kept mostly unpowered not known
- Plugging in and spinning up disk occasionally will minimise mechanical problems
- Data recovery if drive fails possible, but expensive



Storage Media – Flash Devices

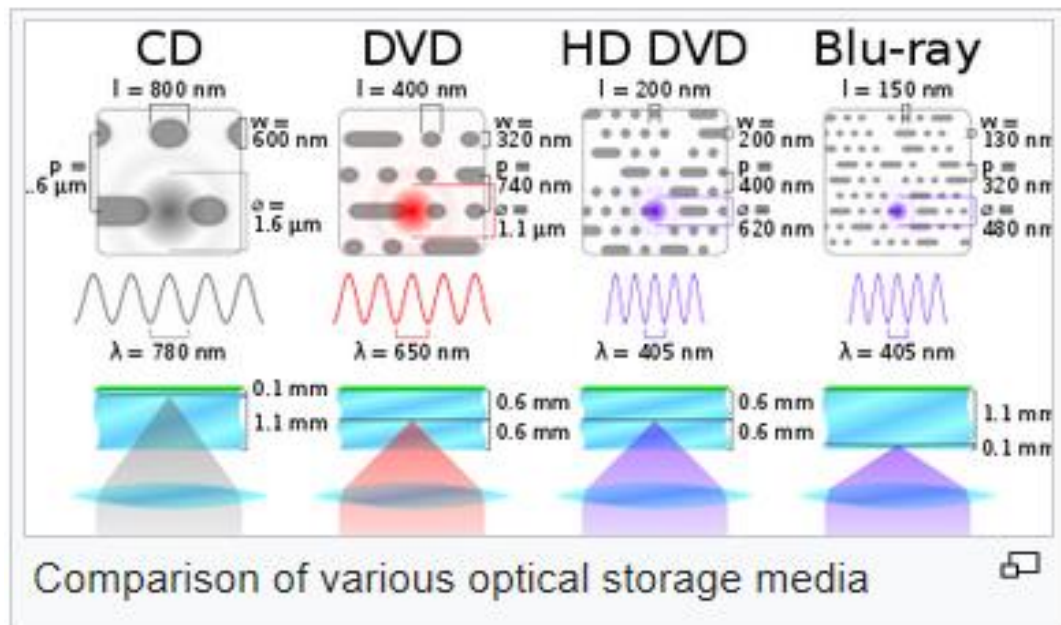
- Store bits as electrical charge in transistors
- Thumb drives
 - Cheap and convenient
 - No moving parts
 - Anecdotally prone to “Device not recognised”
 - Small, hard to label, easily lost
- Solid State Drives
 - Higher quality storage





Storage Media – Optical Disks

- DVDs, CDs, Blu-Ray discs – no electronics
- Store bits as reflective pits in spiral track going from the inside outwards
- Read failures common as CD/DVD reader lasers lose power but disks are fairly stable
- Will survive immersion in dirty water
- Disc scratches and scuffs can be polished out with Brasso and a soft cloth.





Digital Photos

- Most digital image formats use image compression – tradeoff between image quality and file size. JPEG format dominant.
- Descriptive file and folder names not visible on phones or tablets
- Photo Grouping
 - Most apps support albums but app may become unsupported
 - Windows Photo Gallery
 - Picasa
 - Folders may not be visible but always present – most robust grouping
- Weed multiple photos of the same scene with a near-matching app like [Duplicate Photo Cleaner](#)
- Embed information you'd like to write on the back in image with an app like [Caption Pro](#)



Visible Captioning of Photos



Children's Medical Research Foundation Button Day
1981

Jill Howell, left, of Yarralumla, Maureen Cassar of Deakin and Ellen McCuaig of Narrabundah preparing collecting trays for Children's Medical Research Foundation Button Day



Tidbinbilla, ACT, Australia, 1982

Simon Kravis and Dinah Kravis meet an Emu

captionpro.com.au

Caption is part of the image

Image Analysis may be able to extract text from images (Windows Photos)



Digital Video

- Files are big and getting bigger as resolution increases
- Compression methods more variable than for photos
- 1 min iPhone 5 video uses same space as 30 still photos (100 Mbytes)
- Editing digital video is difficult & slow
- No dominant file format
- Tape formats (eg VHS, Video 8)
 - No electronics
 - Very mechanically sophisticated players/recorders
 - Good archival properties if stored well



Scanning

- Makes photos and documents easily shareable and viewable via a computer
- Some colour adjustment, dust and scratch removal possible
- PDF format widely used but no searchable text from scanned documents unless Optical Character Recognition (OCR) used



Captioning Multiple Scanned Images

- Save scans as images, then either
 - Caption with Caption Pro or similar and combine into PDF with [PDF Shaper](#)or
 - Import images into Word, add captions to page and save as PDF or leave as Word document





Webmail (eg Gmail, Outlook.com)

- Messages stored on a remote server
- Loss of data not uncommon
- Advice to regularly back up your archive should be heeded
- Backup format may not be easily readable

Email

☰ Gmail Help

 **Original Poster**
Jan Vinhage 1/8/16

I lost 2 years of emails in my inbox  [2 expert replies](#)

Good morning,
for some reason I can't explain emails between 30.12.2015 and 19.07.2013 have disappeared from my inbox. I have checked in the bin but the stuff I have deleted is still there. I have tried to contact google but to no avail. Is there any way to contact tech support or similar because I have the distinct feeling that there is nothing that I can do at my end.
J.

Desktop Email

- Messages pass through remote servers and may remain there
- Limit to archive sizes on servers
- Download mailboxes and treat email archives like other files



SMS Messages

- Apple provide backup via iTunes for iPhones
- Need 3rd party app for Android
- No natural format for SMS conversations



Proprietary and Open-Source Software

- Good and bad software and support can be found in both categories
- Open-source software depends on continued enthusiasm of a group of developers
- Open-source does not offer much for individuals unless you are a programmer
- Free does not mean open-source.
- Free versions of commercial software limited compared to paid versions



Dealing with Obsolescence

- All data formats and media eventually become obsolete
- Most widely used formats and media tend to survive for longer
- Conversion services most readily available for common media and formats
- Use common formats and be prepared to use conversion services



Social Media

- All want to maintain or increase your use of their platform to deliver advertising
 - Minimal interest in downloading data
- What happens if you die or are incapacitated?
 - Usernames and passwords hard to pass on
 - Google and Facebook offer 'digital will' facilities
- Is Facebook forever?
 - The MySpace story
- Security
 - Google Plus shut down due to security breach
 - Facebook has had numerous breaches



Conclusions

- Archiving digital objects much more difficult than for physical objects due to miniaturisation
- No future-proof storage medium for digital objects
 - Organisations have in-house expertise
 - Individuals must rely on specialist services
 - Maintenance over decades is required
- [Articles](#) on captioning, digital preservation & other topics