# Organising, storing and securely handling research data

**Scott Summers**

UK Data Service

University of Essex

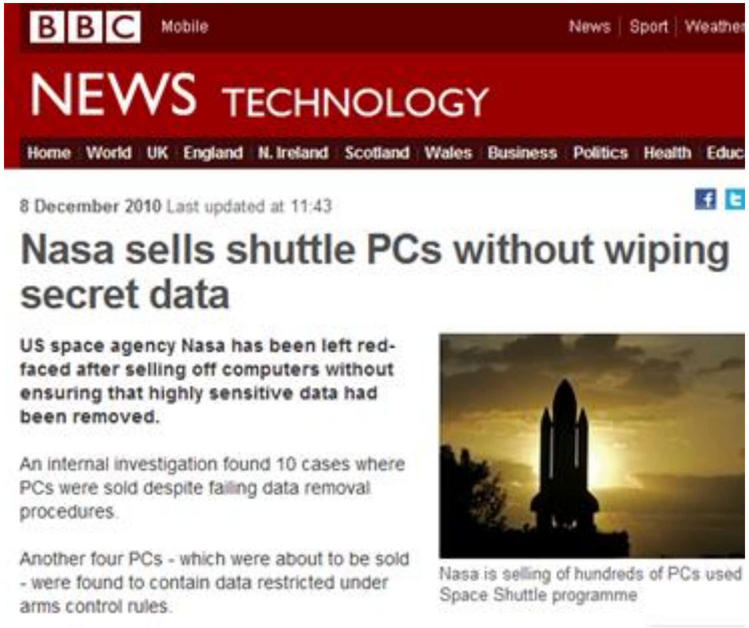
Managing, sharing and archiving social science research data

15th June 2016

UK Data Service

# Overview

- Looking after research data for the longer-term and protecting them from unwanted loss requires having good strategies in place for:
  - securely storing
  - backing-up
  - transmitting/encrypting
  - and disposing of data



8 December 2010 Last updated at 11:43

## Nasa sells shuttle PCs without wiping secret data

**US space agency Nasa has been left red-faced after selling off computers without ensuring that highly sensitive data had been removed.**

An internal investigation found 10 cases where PCs were sold despite failing data removal procedures.

Another four PCs - which were about to be sold - were found to contain data restricted under arms control rules.

Nasa is selling of hundreds of PCs used Space Shuttle programme

- Collaborative research brings additional challenges for the shared storage of, and access to, data

Stuff happens!

# Stuff happens: data inferno

- A fire destroyed a University of Southampton research centre resulting in significant damage to data storage facilities

- What if this was your university, your office or your data?
- Source: BBC

UK Data Service

# Stuff happens: fieldwork nightmares

- “I’m sorry but we had to blow up your laptop.”

- “What….all my client case notes and testimony, writing, pictures, music and applications. Years of work. NO!!!!”
- Source: https://lilyasussman.com

# Stuff happens: data theft

- What would happen if you lost your data?
- Imagine if you lost four years worth of research data - this nightmare situation happened to Billy Hinchen

https://www.youtube.com/watch?v=3xlax_Iin0Y

- Source: https://projects.ac/blog/the-stuff-of-nightmares-imagine-losing-all-your-research-data/

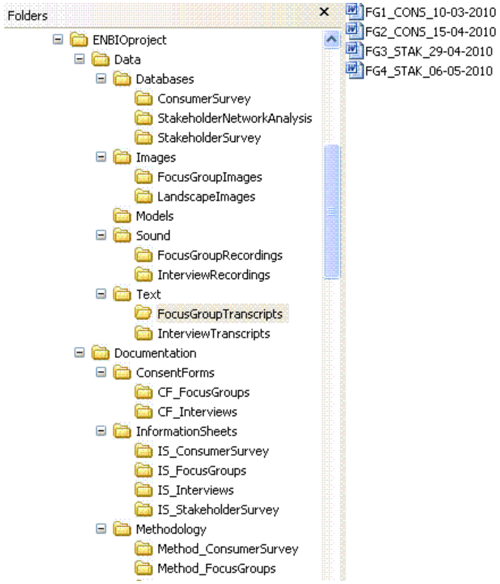
# Organising and storing data

# Organising data

- Plan in advance how best to organise data
- Use a logical structure and ensure collaborators understand

Examples

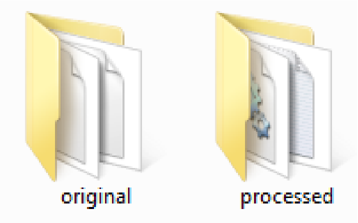
- hierarchical structure of files, grouped in folders, e.g. audio, transcripts and annotated transcripts
- survey data: spreadsheet, SPSS, relational database
- interview transcripts: individual well-named files

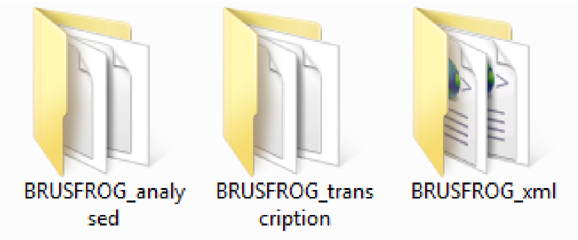


Folders

- ENBIOproject
  - Data
    - Databases
      - ConsumerSurvey
      - StakeholderNetworkAnalysis
      - StakeholderSurvey
    - Images
      - FocusGroupImages
      - LandscapeImages
    - Models
    - Sound
      - FocusGroupRecordings
      - InterviewRecordings
    - Text
      - FocusGroupTranscripts
      - InterviewTranscripts
  - Documentation
    - ConsentForms
      - CF_FocusGroups
      - CF_Interviews
    - InformationSheets
      - IS_ConsumerSurvey
      - IS_FocusGroups
      - IS_Interviews
      - IS_StakeholderSurvey
    - Methodology
      - Method_ConsumerSurvey
      - Method_FocusGroups

FG1_CONS_10-03-2010
FG2_CONS_15-04-2010
FG3_STAK_29-04-2010
FG4_STAK_06-05-2010





# Data storage

- Local storage
- University and collaborative storage
- Cloud storage
- Data archives or repositories

UK • DATA ARCHIVE

zipcloud
box
justcloud.com
bitcasa
INFINITE STORAGE
OneDrive
Dropbox
IDrive
Google Drive
MediaFire
iCloud

# Local data storage

- Internal hard drive/flash drive
- Note that all digital media are fallible
- Optical (CD, DVD & Blu-ray) and magnetic media (hard drives, tape) degrade over time
- Physical storage media become obsolete e.g. floppy disks

- Data files should be copied to new media every two-to-five years after they are first created

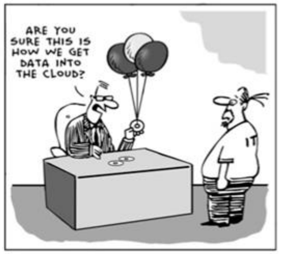
# University and collaborative storage

- Your university or department may have options available. For example:
  - Secure backed up storage space
  - VPN giving access to external researchers
  - Locally managed Dropbox-like services such as OneDrive and Essex ZendTo
  - Secure file transfer protocol (FTP) server

Sharing data between researchers

- Too often sent as insecure email attachments
- Physical media?
- Virtual Research Environments
  - MS SharePoint
  - Clinked
  - Huddle
  - Basecamp

# Cloud storage services



By David Fletcher
http://www.cloudtweaks.com/2011/05/the-lighter-side-of-the-cloud-data-transfer/

- Online or ‘cloud’ services are becoming increasingly popular
- Google Drive, DropBox, Microsoft OneDrive and iCloud

Google Drive Dropbox OneDrive

- Benefits:
  - Very convenient
  - Accessible anywhere
  - Good protection if working in the field?
  - Background file syncing
  - Mirrors files
  - Mobile apps available

**But**,

  - These are not necessarily secure
  - Potential DPA issues
  - Not necessarily permanent
  - Intellectual property right concerns?
  - Limited storage?

# Cloud storage services

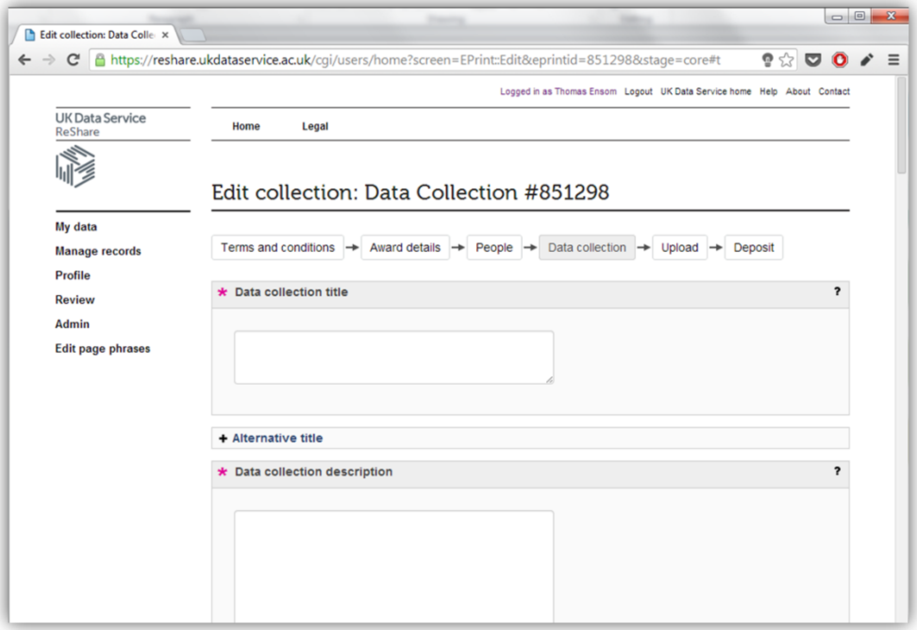
- Perhaps more secure options?

[Mega.nz](Mega.nz) [SpiderOak](SpiderOak) [Tresorit](Tresorit)

- Cloud data storage should be avoided for high-risk information such as files that contain personal or sensitive information, information or that has a very high intellectual property value.

# File sharing – data archive or repository

- A repository acts as more of a ‘final destination’ for data
- Many universities have data repositories now catering to its researchers, e.g. Research Data Essex
- UK Data Service has it’s own service called ‘ReShare’, for social science data of any kind
- http://reshare.ukdataservice.ac.uk/



https://reshare.ukdataservice.ac.uk/cgi/users/home?screen=EPrint::Edit&eprintid=851298&stage=core#t

Logged in as Thomas Ensom

Edit collection: Data Collection #851298

Terms and conditions → Award details → People → Data collection → Upload → Deposit

* Data collection title

+ Alternative title

* Data collection description

# Backing-up data

# Backing-up data

- It is not a case of *if* you will lose data, but *when* you will lose data!
- Keep additional backup copies and protect against: software failure, hardware failure, malicious attacks and natural disasters
- **Would your data survive a disaster?**

# Digital back-up strategy

Consider

- **What’s backed-up?** - all, some or just the bits you change?
- **Where?** - original copy, external local and remote copies
- **What media?** - DVD, external hard drive, USB, Cloud?
- **How often?** - hourly, daily, weekly? Automate the process?
- **What method/software?** - duplicating, syncing or mirroring?
- **For how long is it kept?** - data retention policies that might apply?
- **Verify and recover** - never assume, regularly test and restore

Backing-up need not be expensive

- 1Tb external drives are around £50, with back-up software

Also consider non-digital storage too!

“We back up our data on sticky notes because sticky notes never crash.”

# Verification and integrity checks

- Ensure that your backup method is working as intended
- Automated services - check
- Be wary when using sync tools in particular
  - Mirror in the wrong direction or using the wrong method, and you could lose new files completely

- You can use checksums to verify the integrity of a backup
- Also useful when transferring files
- Checksum somewhat like a files’ fingerprint
- …but changes when the file changes

# Checksums

- Each time you run a checksum a number string is created for each file
- Even if one byte of data has been altered or corrupted that string will change
- Therefore, if the checksums before and after backing up a data file match, then you can be sure that the data have not altered during this process
- A free software tool for computing MD5 checksums is MD5summer for windows
- We will run through a demonstration of this later

# Data security

UK Data Service

# Data security

Protect data from unauthorised:

- Access
- Use
- Change
- Disclosure
- Destruction

Who knows who is watching, listening or attempting to access data…

GCHQ IS NOW FOLLOWING YOU ON TWITTER!... AND FACEBOOK ... AND EMAIL... AND..
dink

# Data security strategy

- Control access to computers:
  - use passwords and lock your machine when away from it
  - run up-to-date anti-virus and firewall protection
  - power surge protection
  - utilise encryption
  - on all devices: desktops, laptops, memory sticks, mobile devices
  - at all locations: work, home, travel
  - restrict access to sensitive materials e.g. consent forms and patient records
  - personal data need more protection – always keep them separate and secure

- Control physical access to buildings, rooms and filing cabinets

- Properly dispose of data and equipment once project is finished

# Encryption

- Encryption is the process of encoding digital information in such a way that only authorised parties can view it.

- **Always** encrypt personal or sensitive data
  - = anything you would not send on a postcard
  - e.g. moving files, such as interview transcripts
  - e.g. storing files to shared areas or insecure devices

- Basic principles
  - Applies an algorithm that makes a file unreadable
  - Need a ‘key’ of some kind (passphrase or/and file) to decrypt

- The UK Data Service recommends Pretty Good Privacy (PGP)
  - More complicated than just a password, but much more secure
  - Involves use of multiple public and private keys

# Encryption software

Encryption software can be easy to use and enables users to:

- encrypt hard drives, partitions, files and folders
- encrypt portable storage devices such as USB flash drives

VeraCrypt

BitLocker

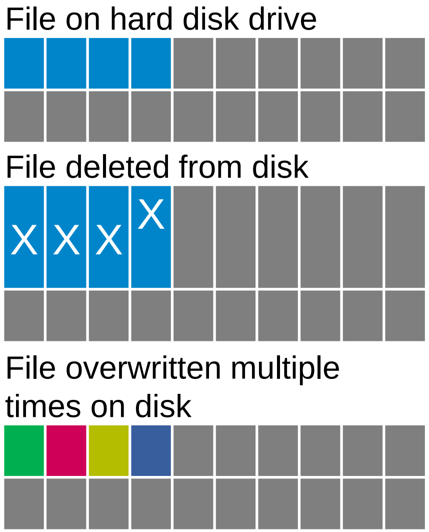
Axcrypt

FileVault2

We will run through a demonstration of VeraCrypt later

# Data disposal

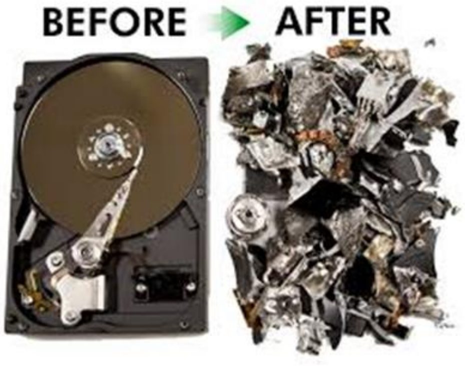
- When you delete a file from a hard drive, it is likely to still be retrievable (even after emptying the recycle bin)
- Even reformatting a hard drive is ***not*** sufficient
- Files need to be overwritten multiple times with random data for best chances of removal
- The **only** sure way to ensure data is irretrievable is to physically destroy the drive (using an approved secure destruction facility)



File on hard disk drive

File deleted from disk

File overwritten multiple times on disk



# Data disposal software

- **BCWipe** - uses ‘military-grade procedures to surgically remove all traces of any file’
  - Can be applied to entire disk drives

- **AxCrypt -** free open source file and folder shredding
  - Integrates into Windows well, useful for single files

- Physically destroy portable media, as you would shred paper

# Summary of best practices in data storage and security

- Have a personal backup and storage strategy: (a) store an original local copy; (b) external local copy and (c) external remote copy
- Copy data files to new media every two-to-five years after first created
- Know your institutional back-up strategy
- Check data integrity of stored data files regularly (using checksums)
- Create new versions of files using a consistent and transparent system structure
- Encrypt data – especially when sensitive or transmitting and sharing
- Know data retention policies that apply: funder, publisher, home institution
- Archive data and securely destroy data where necessary

# Resources

**UK Data Service Website resources**

- Organise data - https://www.ukdataservice.ac.uk/manage-data/format/organising
- Data storage - https://www.ukdataservice.ac.uk/manage-data/store/storage
- Data security - https://www.ukdataservice.ac.uk/manage-data/store/security
- Data encryption - https://www.ukdataservice.ac.uk/manage-data/store/encryption
- Data backup - https://www.ukdataservice.ac.uk/manage-data/store/backup
- Checksums - https://www.ukdataservice.ac.uk/manage-data/store/checksums
- File sharing - https://www.ukdataservice.ac.uk/manage-data/store/file-sharing
- Data disposal - https://www.ukdataservice.ac.uk/manage-data/store/disposal
- Further resources - https://www.ukdataservice.ac.uk/manage-data/store/disposal

**Video Tutorials**

- VeraCrypt - https://www.youtube.com/watch?v=Ogm9QHQpFqU
- AxCrypt - https://www.youtube.com/watch?v=ACcRlnsoYZg
- FileVault 2 - https://www.youtube.com/watch?v=JIZ9EFMS0ic
- Time Machine - https://www.youtube.com/watch?v=hlsQaVj7WtA
- MD5 Summer - https://www.youtube.com/watch?v=VcBfkB6N7-k

# Questions?