cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Storing and Securely Handling Research Data

Dr Scott Summers
UK Data Service
University of Essex

Pravni in etični vidiki ravnanja z raziskovalnimi podatki - 11th April 2018

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

BBC Mobile News | Sport | Weather

NEWS TECHNOLOGY

Home World UK England N. Ireland Scotland Wales Business Politics Health Educ

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

cessda eric

Stuff Happens!

cessda eric

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

cessda eric

Stuff Happens: Data Loss

- What would happen if you lost your data?
- Imagine if you left your bag on a train, containing your laptop (with all your digital research notes on) and your paper based notes too – this situation happened to Andrew Penson
- Source:
- https://twitter.com/ADPenson/status/883637257323896832

cessda eric

Stuff Happens: Data Theft

- What would happen if you data was stolen?
- Imagine if you lost four years worth of research data – this situation happened to Billy Hinchen
- https://www.youtube.com/watch?v=3xlax_Iin0Y
- Source:
- https://figshare.com/blog/The_stuff_of_nightmares_imagine_losing_all_your_research_data/121

cessda eric

Stuff Happens: Data Theft

- What would happen if you data was stolen?
- Imagine if seven years worth of your Ebola research was stolen – this situation happened to Dr Fitzgerald
- Source:
- https://www.standard.co.uk/news/crime/burglar-stole-laptop-with-seven-years-of-ebola-research-from-doctor-s-house-a3689406.html

cessda eric

Storing Data

cessda eric

Data Storage

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

cessda eric

Data Storage: How to Decide

- How much storage space do I need?
- Who needs access?
- What precautions should I take to protect my data against loss?
- Which storage solutions are suitable for personal data?

cessda eric

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

cessda eric

University and Collaborative Storage

- Your university or department may have options available. For example:
  - Network attached drives
  - 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

cessda eric

Cloud Storage Services

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

- 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
  - Limited control over where data is stored
  - Not necessarily permanent
  - Intellectual property right concerns?
  - Limited storage?

cessda eric

Cloud Storage Services

- Perhaps more secure options?

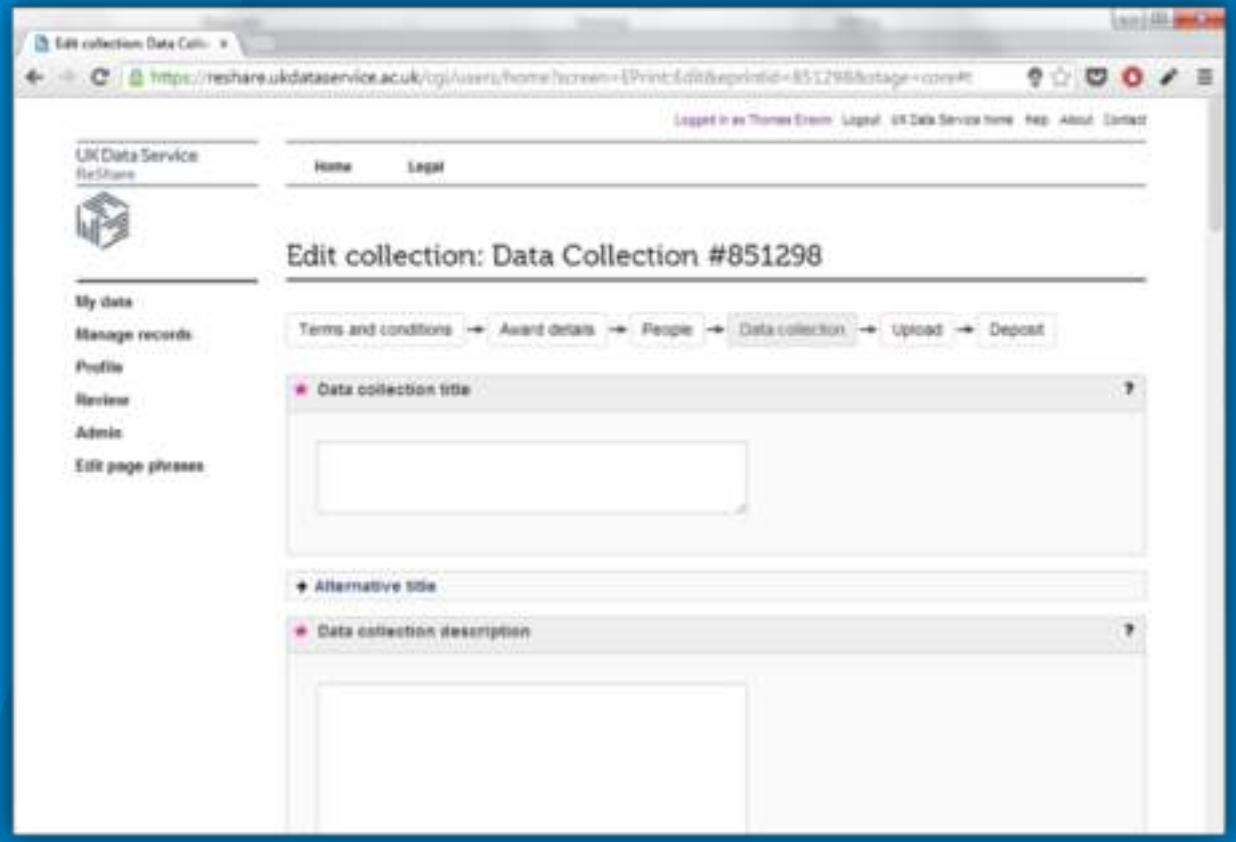
- Mega.nz
- SpiderOak
- Tresorit

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

cessda eric

File Sharing – Data Archive / 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 its own service called ‘ReShare’, for social science data of any kind
- http://reshare.ukdataservice.ac.uk/



cessda eric

Data Storage: Comparison

	Portable	Cloud	Local	Networked drive
Advantage	Easy transport No internet needed Low cost	Easy access and sharing Automatic backup Automatic version control	Full control Easy to protect from unauthorised access	Central storage Shared and remote access Central backup
Disadvantage	Easily lost and damaged Not for long-term storage	Not always secure No control over storage location (breach data protection) Free service may claim right to use content	No sharing	Higher security needed Higher cost
Sensitive data	Encrypt files Password protect	Encrypt files	Password protect PC Encrypt hard drive	Protect from unauthorised access

cessda eric

Backing-up Data

cessda eric

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?**

cessda eric

Common Causes of Data Loss / Damage

- Hardware failure
- Software malfunction
- Malware or hacking
- Human error (research data accidentally gets deleted or overwritten or is lost in transport)
- Theft, natural disaster or fire
- Degradation of storage media

Backups will permit you to restore data in the case of loss or damage

cessda eric

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?
  - How many copies? - minimum of three copies!
  - 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
  - 2Tb external drives are around£70, with back-up software
- Also consider non-digital storage options too!

cessda eric

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

cessda eric

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
- OS X has this functionality built into Terminal
- We will run through a demonstration of this later

cessda eric

Data Storage Strategy

1. Use two types of storage media
    - » At least two different types of storage media should be used, e.g. Solid State Disk (SSD) and CD-ROM or Hard Disk Drive (HDD) and SDD

2. Replace storage media
    - » Replace storage media after 2-5 years

3. Carry out integrity checks
    - » Frequently carry out integrity checks to ensure that the stored data has not been corrupted. This can be done with checksum tools. These allow you to detect if a file was changed in any way, intentionally or unintentionally

cessda eric

Data Security

cessda eric

Data Security

- Protect data from unauthorised:

- Access
- Use
- Change
- Disclosure
- Destruction

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

cessda eric

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
  - UPS power supplies
  - 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 your project is finished

cessda eric

Passwords

- Strong passwords are crucial
- Avoid using weak or easy to guess passwords and reusing passwords
- Consider password managers, complex passwords or stringing words together to create stronger passwords
- But, remember that you need to be able to remember the passwords!
- **Why does this matter?**
- No matter how good the encryption is that you use if you use a weak password the encryption will offer little protection
  - https://howsecureismypassword.net (*Never use real passwords*)

cessda eric

Password Security

» “Dog”

» “Password”

cessda eric

Password Security

» “I like to use secure passwords”

cessda eric

Password Security

- Edward Snowden on passwords
- https://www.youtube.com/watch?v=yzGzB-yYKcc

cessda eric

Encryption

- Encryption is the process of encoding digital information in such a way that only authorised parties can view it
- Basic principles
  - Applies an algorithm that makes a file unreadable
  - Needs a ‘key’ of some kind (passphrase or / and file) to decrypt
- Some types of encryption provide greater protection than others, the type and level of encryption used should correspond to the sensitivity of the data being protected
- As a general rule, more bits equals stronger encryption, therefore, 256-bit encryption is stronger than 128-bit encryption

cessda eric

Encryption

- When using encryption 128-bit encryption should be the minimum level used

- 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

- 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

cessda eric

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

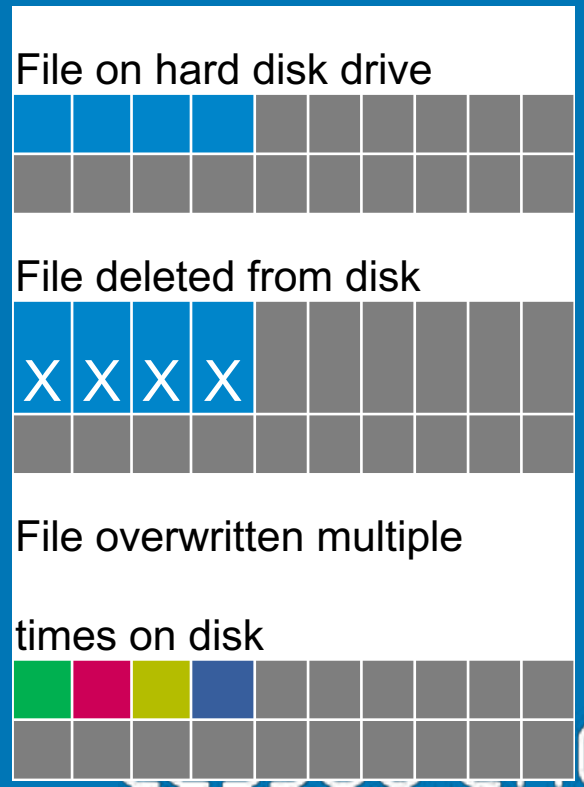
- FileVault2

- We will run through a demonstration of VeraCrypt later

cessda eric

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)



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
- Securely destroy data at the end of the project

cessda eric

Video Tutorials

- BitLocker tutorial - https://www.youtube.com/watch?v=y4Iosu-Yfsw
- FileVault 2 tutorial - https://www.youtube.com/watch?v=JIZ9EFMS0ic
- AxCrypt tutorial - https://www.youtube.com/watch?v=ACcRlnsoYZg
- Time Machine back up tutorial - https://www.youtube.com/watch?v=hlsQaVj7WtA
- MD5summer tutorial - https://www.youtube.com/watch?v=VcBfkB6N7-k&index=6&list=PLG87Imnep1Slj6Dxq1QQ4_WeQYU3vNBX

cessda eric

Questions

- Dr Scott Summers
- Collections Development and Producer Relations team
- UK Data Service
- University of Essex
- ukdataservice.ac.uk/help/get-in-touch

cessda eric