
THE UNIVERSITY *of* EDINBURGH

igmm

INSTITUTE OF GENETICS
& MOLECULAR MEDICINE

MRC Medical Research Council


CANCER RESEARCH UK

# Data Storage & Security

Dr Alastair F. Brown

*Head of Computing*

MRC Human Genetics Unit

MRC Institute of Genetics and Molecular Medicine

The University of Edinburgh

DIY Research Data Management Training Kit for Librarians

# Purpose of This Presentation

- Topic Overview
- Fill some gaps
- Real life examples

# Topic Overview

- Where to store data
  - Local Drive, Network Drive, Cloud
  - Consider: Capacity & Access by co-workers
- Data backup
  - Disaster Recovery (Business Continuity)
  - Long Term Backup  (Archiving)
- Data security
  - Corruption or Loss (hardware failure or data deletion)
  - Confidentiality (personal or intellectual property)

# Digression on Two Issues

- Two issues which are often overlooked but are worth highlighting are
  - Usernames and passwords: they are so common users often forget they are still a key part of security on most systems
  - Public WiFi hotspots: safe or not?

# Usernames and Passwords

- If possible NEVER use your username as your e-mail address e.g. fbloggs27@staffmail.ed.ac.uk…
  - …always use an alias e.g.: Fred.Bloggs@ed.ac.uk
  - with a valid username, the bad guys only have to guess your password
- Do not write passwords on Post-Its/say them out loud
- Do not use untrusted computers (e.g. internet café)
- Do not use the obvious (car reg., phone no., pet's name)
- Do not use **any** dictionary words (including foreign)

# Public WiFi

- Do not be afraid of WiFi hot spots
  - Just be careful
  - Treat them as untrusted computers…
  - …unless you use a VPN* connection.

  * A Virtual Private Network link provides end-to end encryption between your laptop and the system you are connecting to.

# Example 1 – Who Needs Passwords?

- "I don't need a good (or any) password because…":
  - "I have no important/private information in my data area"
  - "I don't care if someone else can read my files"
- Any authorised access is a first step for the bad guys
  - And they may just delete all your work
  - Or worse, change your data which you may not notice
- Though you have no sensitive files, you may have access to parts of the system which DO
- A security hole within the system may be exploited once the bad guys have gained access by legitimate means

# Example 2 – Backup for How Long?

- Researcher has accumulated several years of data and software on a departmental computer backed up remotely every day
- Researcher leaves for another job
- Replacement not found for 6 months
- Replacement tries to log on to computer to find the hard disk had failed 5 months previously
- Asks for a backup to be restored to a new disk, but discovers that backup tapes are recycled after 4 months
- Result – misery!

# Example 3 – Sharing Personal Data

- Share a database between 3 sites
  - Data are clinical in nature, mostly images
  - User uses a database program specially written
  - User assures Sysadmin that all data in database are encrypted
- Solution:
  - Place database in DMZ (Demilitarised Zone) with very tight firewall restrictions
  - Only specific workstations at the 3 sites can connect to the database server
  - Connection to server requires username/password
  - As does access to database itself and to decrypt the data

# Example 3 – [continued]

- Problem – user had not checked how the database worked
  - Sysadmin asked the right questions…
  - …but trusted the user's answers
  - The database contents **were** encrypted but…
  - The database contained only pointers to the images
  - The images were stored as plain files, unencrypted, in a folder/directory outside the database!
  - And to make matters worse, the user decided to keep all their clinic appointment and follow up letters in the same directory – these were Word documents (not even password protected!!)
- Result – a close shave!
  - Good example of defensive, multi-level security

# Example 4 – Where's the Metadata?

- PI needs data generated by a post-doc 3 years previously – data are on backup/archive tapes
  - PI knows the directory/filenames and dates
  - Data files are restored from tape
  - Data files are DNA sequences with no annotations and no metadata files
  - PI cannot find lab notebook of post-doc
  - Post-doc's memory does not persist for 3 years
- Result – misery!

# Thank You

Questions?

Alastair.Brown@igmm.ed.ac.uk


THE UNIVERSITY *of* EDINBURGH

igmm

INSTITUTE OF GENETICS
& MOLECULAR MEDICINE

MRC Medical Research Council


CANCER RESEARCH UK