Research Data Management
For Researchers

Dr Joanna Goodger
Information Hertfordshire

With Bill Worthington and Mohamed Hansraj
Research Data Management

SAFEGUARDING DATA
In this module, we’ll discuss how best to set up your research:

- Storage solutions – where should you keep your data
- Keep your work safe; keep it backed up
- Remote access – carry on working off campus and on other devices
- Share your data safely with collaborators
- Keep you sensitive data secure
Research Data Management

STORAGE SOLUTIONS
Keeping your data just on your working machine, be it a laptop or a desktop, is the perfect way to lose your data easily and permanently – this is not what you want!

UH offers a range of facilities for securing your data helping it live a long and useful life.

- UH personal storage space (U: drive)
- UH shared drives (X: drives)
- UH research drives
- UH Document Management System (DMS)

All these help to keep your data safe and accessible to you and your collaborators within and without UH.
# Research Data Management

## UH Shared Drives

<table>
<thead>
<tr>
<th>Drive</th>
<th>Capacity</th>
<th>Security</th>
<th>Back Up</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>My Docs on local machine</td>
<td>machine disc capacity</td>
<td>password protected</td>
<td>Up to you!</td>
<td>contact your local IT staff</td>
</tr>
<tr>
<td>U: personal network drive</td>
<td>Staff, 5GB; students, 2GB</td>
<td>password protected</td>
<td></td>
<td>Helpdesk: ext. 4678 or email</td>
</tr>
<tr>
<td>Research networked drive</td>
<td>&gt; 5GB</td>
<td>password protected and accessed by authorised members only</td>
<td>Daily UH backups held at College Lane and de Havilland data centres.</td>
<td>RDM Website</td>
</tr>
<tr>
<td>School networked drive</td>
<td>Unlimited (within reason)</td>
<td>password protected but accessed by all of the school members.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document Management System (DMS)</td>
<td>&gt; 5GB</td>
<td>password protected and accessed limited to group members only</td>
<td>Replicated the Disaster Recovery System on the other campus. Nightly backups to tape.</td>
<td></td>
</tr>
</tbody>
</table>
## Research Data Management
### UH Shared Drives

<table>
<thead>
<tr>
<th>Document Management System</th>
<th>Research Drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>External access</td>
<td>Free of file structure</td>
</tr>
<tr>
<td>Device independent</td>
<td>No automatic version control</td>
</tr>
<tr>
<td>Consistent file structure in place</td>
<td></td>
</tr>
<tr>
<td>Granular security</td>
<td>Ideal for large files which are not</td>
</tr>
<tr>
<td>Automatic retention</td>
<td>documents, and would be costly to duplicate after every alteration.</td>
</tr>
<tr>
<td>Automatic file reporting and auditing</td>
<td></td>
</tr>
<tr>
<td>Full text search</td>
<td></td>
</tr>
<tr>
<td>Scanning straight to file</td>
<td></td>
</tr>
</tbody>
</table>
Research Data Management
UH Storage

The shared storage is accessible using Novell and the UH intranet.

**On Windows** machines, from off campus, or when connected to the student network, you need to activate Network Connect before logging into Novell. This allocates your connection a UH IP address.

**On Linux** machines, you need to use a Windows Virtual Machine to run Novell, but you can configure the network settings directly and do not need to run Network Connect.

**Mac** users should be able to connect direct to the drives without using Novell.

- Novell
- Network Connect
Sign in with your staff user name and password. In order to use the “drag and drop” facility you need to download additional add-ons.

Your project will be equipped with a project file structure, secure to your group with additional security for sensitive files in personnel and consent folders.
Research Data Management
Document Management System (DMS)

Safeguarding data with Research Data Management
Research Data Management

KEEP IT SAFE - BACK UP!
Research Data Management

Back Up

Backing up should be an automatic part of your everyday research activities.

In 2005, an electrical fault in the electronics and laser research building at the University of Southampton cost £50-100M including temporary building hire and transfer of work to Holland.

Imagine if a fire or similar disaster happened at UH

How much would it cost you?

Storing your data on the UH network means that it is stored at de Havilland and at College Lane in the data centres.

Mountbatten Building, Southampton University.

Safeguarding data with Research Data Management
Research Data Management

Back Up

Theft and Loss are real risks:

In 2012, a Dell survey demonstrated that **12000 laptops per week** are left at US airport security points.

They cannot be turned on or analysed so unless your name is on the outside of your mobile device, it will not be returned to you.

You have 30 days before auctioning them as government property.

Next time you travel, make sure that

- your device is **labelled** and **locked**
- your data is **backed up**.
There are UH facilities, but these should be considered secondary back ups. You should have your own back up.

Safeguarding data with Research Data Management
Research Data Management

Back Up

**Windows**
Backup and restore

- Set an automated backup through control panel

**Mac**
Time machine

- Back up your entire content to another disk or to the net.

Safeguarding data with Research Data Management
rsync
- Updates the changes to files between two directories and servers

/usr/bin/rsync -avu /data/someuser/ /local/data/
/usr/bin/rsync -avu /home/someuser/ /local/data/home/

cron
- Timed schedule to perform tasks – your rsync for example

SHELL=/bin/tcsh
PATH=/sbin:/bin:/usr/sbin:/usr/bin
MAILTO=a.user@herts.ac.uk
17 3 * * * /usr/bin/rsync -avu /data/someuser/ /local/data/
Research Data Management

SHARING
Research Data Management
Sharing

Many methods for sharing data and documents during the analysis and the writing up of projects have been developed by researchers, but most are unsecure and violate the UH data policy.

- Emails have a limit to the attachment size and can be intercepted or simply miss-sent
- Web space is generally unsecured and openly accessible
- Cloud storage is unreliable and gives third parties access to your data
- Storage media, such as DVDs and flash pens, can be lost, intercepted, or broken in transit

Safeguarding data with Research Data Management
Research Data Management
Sharing

- Access only to UH members
- Versioning
- OS independent
- Set file structure

DMS

- Access only to UH members
- Undefined file structure
- No versioning

Research Drives

- Send large files using the UH server
- Web based only
- Open and Free

UH FTP

Safeguarding data with Research Data Management
Research Data Management
Sharing

UH Sharing solutions are

• more secure and reliable
• accessible on multiple operating systems and on the web,
• accessible on and off campus.
• accessed by listed members only

Collaborators can be granted visiting member status; on a separate hierarchy so your collaborators will only be able to access folders that you assign to them.
Cloud solutions are easy to use, but open your data up to third parties;

- some providers **take ownership** of your data but most declare that you are alone responsible for your data and that you retain your intellectual property rights,
- **the rules change** depending on where in the world your data is being held,
- back-up policies and versioning vary,
- providers reserve the right to **close your account** if they decide that you have misused it – there’s no appeal process,
- providers **scan your files** for illicit files including political and commercial threats,
- providers will **report files** that are deemed illegal,
- you may not be able to use encryption or password protection on your files.
Research Data Management
Sharing across OS

Sharing between operating systems can also be achieved using the shared drives and DMS; however, WinSCP is a fast, effective way of moving data between operating systems.

Free SFTP, SCP and FTP client for Windows

Log into a server from a windows machine.

Move files across using drag and drop.
Research Data Management

REMOTE ACCESS
On Campus

There are two networks – the staff network and the student network.

Staff network:
• Access to the shared drives using Novell on Windows
• Access to StaffNet and core services

Student network:
• Access to student system personal drive only.

To access StaffNet, core services, and the shared drives, you need to activate network connect and log into Novell.
Network Connect
– allocate your computer a UH IP address adding it to the staff network.

First time: log in to the UH VPN at  https://uhvpn.herts.ac.uk/

Using your UH member username@staff and the password.

Safeguarding data with Research Data Management
Select **Start** to activate Network Connect.

Activating network connect will initiate a download of the Java application to your machine.

Next time, you will not need to web interface, you can activate it directly from the Start menu.

Enter your username@staff and password.

You can then browse the online services; core, StaffNet, and engage, as you would at your desk on campus.
Next log into **Novell**;
- This enables connections to your staff personal drive and shared drives.

You will need to install the Novell package onto your machine including the UH settings.
**Emails**

The University of Hertfordshire uses a Windows Microsoft Exchange server. This requires manual setup on most devices, and a number of different domain, server and URL details.

Step by step instructions are provided in the UH Remote Access guide.
Research Data Management

SECURITY
Laptops go missing very regularly; Intel’s study in 2012 surveying 329 private and public organizations demonstrated that

- On average, 2.3% of laptops assigned to employees are lost each year
- 7.1% of employee laptops were lost or stolen before the end of their usefulness lifespan

In education & research that rises to

- 3.7% per year
- with 10.8% of laptops being lost before the end of their useful life

75% are lost outside the workplace, such as in cars, on public transport, in hotels.

Have you lost one yet?
If you lost your laptop or it was stolen, how easily could your data be stolen?

Source: BBC.co.uk/news
Research Data Management

Security

Password protect your devices:

• Do not write your password down and leave it lying around in full view
• Do not use the same password for personal and work related activities
• Do not reuse an old password when asked to update your password
• Do not share your password with others for any reason
• Do not enable the save password option on your computer
• Lock your machine when you step away from your desk (⊞+L on windows)
At UH, the UPR12 Data Management Policy refers to how staff should handle their PCI:

http://sitem.herts.ac.uk/secreg/upr/IM12.html

The Managing Personal & Confidential Information (PCI) Guide:


Unacceptable, but common practice:

- Saving PCI on a non-University computer;
- Use of portable media devices to store or backup PCI;
- Regular transfer or unencrypted transfer of PCI via portable media
Research Data Management
Encryption

**Windows 7+ only**

- Bitlocker
  - password protected, 128-bit or 256-bit AES encryption.
  - Encrypt the entire device, folder or file

**Mac Secure disk image**

- password protected, 128-bit or 256-bit AES encryption.
- automatically expands

Safeguarding data with Research Data Management
If you’re sharing with collaborators and partners with unknown operating system, or without root access – use TrueCrypt.

- Creates a virtual encrypted disk within a file and mounts it as a real disk.
- No installation required for sharing; can be packaged with mobile software
- Encryption is automatic, real-time (on-the-fly) and transparent.
- Encryption can be hardware-accelerated on modern processors.
- Provides plausible deniability, in case an adversary forces you to reveal the password.
If you have information that is not vital to the study, then don’t share it, anonymise it.

- Remove direct identifiers (e.g., personal information such as addresses)
- Aggregate or reduce the precision of variables that might be identifiable (such as postcode).
- Generalise text variables to reduce identifiability
- Restrict continuous variables to reduce outliers
- Pay particular attention to anonymising relational data - some anonymised variables may become identifiable when considered in combination.

Whenever editing is done, researchers need to be aware of the potential for distorting the data. For example, deleting all possible identifiers from text or sound recordings is a simple but blunt tool that creates data that are confidential but may be unusable.