# Introduction to Research Data Management

Storage, backup, and security

http://hdl.handle.net/1969.1/164383

LIBRARIES
TEXAS A&M UNIVERSITY

# Introduction

Focus: Digital data storage.

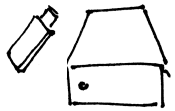
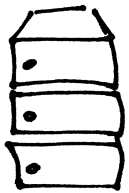
- Storage locations,
- data backup,
- and security.

Goal: Think "safety first" and reduce risks.

# What risks?

- Data loss.
- File corruption.
- Unauthorized data access.

# Discussion


- Accessibility
- Space
- Security
- Integrity

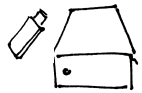
# Personal computers

1. Accessibility
2. Space
3. Security
4. Integrity

# Personal computers

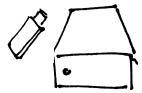
Recommended for storing working data that aren't sensitive or restricted.

Depend heavily on personal practices.



# External hard drives or USB sticks

1. Accessibility
2. Space
3. Security
4. Integrity



# External hard drives or USB sticks

Usually used as a location for backup or data sharing.

Depend heavily on personal practices.

# University networked drives

1. Accessibility
2. Space
3. Security
4. Integrity

# University networked drives

Generally recommended for storing master copies of data.

Managed by professional IT staff.

# Third-party services

TAMU-sponsored services:

- TAMU Google Drive
- Syncplicity

Other common services:

- Amazon S3 (cloud computing storage facilities)
- Dropbox
- SpiderOak

# Third-party services

1. Accessibility
2. Space
3. Security
4. Integrity

# Third-party services

Recommended for sharing with collaborators and for working data.

Read the fine print.

- Security depends on the service provider.
- May have size, cost, or other limitations.

# TAMU Google Drive

**Not** for:

- Electronic Protected Health Information (EPHI) subject to the Health Insurance Portability and Accountability Act (HIPAA).
- Data controlled for export under Export Control Laws (EAR, ITAR).
- Certain types of Personally Identifiable Information (PII), including Social Security Numbers, bank or financial account numbers.
- High Risk Activities in which loss or inappropriate disclosure would result in large consequences in terms of economic loss, loss of trust, or legal liability.

# TAMU Syncplicity

Secure cloud environment for users in the Texas A&M University System to store documents.

# Exercise

Where do you think you will you store your data?

How many copies of your data will you keep?

# Good practices for storing data

Keep at least 3 copies of data, in at least 2 geographic locations.

1. Original.

2. External copy, kept locally.

3. External copy, kept in a remote location.

Move data files to new media 2 to 5 years after first created.

# Create a backup system

1. How much space to store your original data?
2. How often should backups be made?
3. Full backup or an incremental backup of changes?
4. How long should each backup be stored before being over-written?
5. How much additional space is required to maintain these backups?
6. How will you keep track of versions when backing up to multiple devices?
7. Where will backups be stored?

# Software can help

Software that automatically backs up files can simplify the process considerably.

# Test backups

Check the integrity of backups.

- Restore from backups.
- Generate checksums for files.

# Break

# Security tips

- Follow policies for moving, copying, and sharing data.
- Avoid logging into secure space using untrusted computers or networks.
- Ensure physical storage media are locked up.
- Use strong passwords.
- Store and transmit sensitive data with encryption.
- Ensure complete destruction when deleting sensitive data.

# Passwords

- Unique.
- Max them out, 18 or more characters.
- Nonsense “passphrases,” fake words, digits, symbols.


# Discussion

Does having a password protect you data if your computer is stolen?

# Encryption

Encrypting data will help ensure they remain safe from disclosure in the event that a laptop, desktop, USB stick, or external hard drive are lost or stolen.

**Bitlocker**: Whole disk encryption tool included in Windows Vista and later.

**FileVault**: Built-in tool on Macs for whole disk encryption.

# Sending files

Files in transit should also be protected.

**Filex**: provides a secure and easy-to-use file distribution system.

https://filex.tamu.edu

GNU Privacy Guard (GPG) to encrypt files, folders, and emails.

**Gpg4win**: Made easy for Windows.

**GPG Suite**: Made easy for Mac OS.

# Virtual Private Networks

Virtual Private Networking (VPN) is a method of providing a more secure network connection from public or untrusted networks.

**AnyConnect VPN client:** Cisco tool allowing users to connect to the Texas A&M VPN.

# Sensitive data deletion

- Data erasure.
- Degaussing.
- Physical destruction.

# Conclusion

Reviewed risks associated with data storage.

Discussed data back up strategies.

Identified tools and methods for keeping data secure.

# References and resources


- DMPTool. "Data Management General Guidance" [Website](https://dmptool.org/dm_guidance)
- Google Apps at Texas A&M. "Terms Of Use and Privacy" [Website](http://google.tamu.edu/Resources/terms_of_use.html)
- Hicock, Robyn. 2016. "Microsoft Password Guidance" [PDF](https://www.microsoft.com/en-us/research/publication/password-guidance/)
- MANTRA. "Storage and security" [Module](http://mantra.edina.ac.uk/storageandsecurity/)
- New England Collaborative Data Management Curriculum. "Module 4: Data Storage, Backup, and Security" [Website] (http://library.umassmed.edu/necdmc/modules)
- TAMU. "Protecting Confidential Information" [Website](https://security.tamu.edu/protect_my_work/Protecting_Confidential_Information.php)
- TAMU. "SAP29.01.03.M1.16: Information Resources — Portable Devices: Information Security"[PDF](http://rules-saps.tamu.edu/PDFs/29.01.03.M1.16.pdf)