CEPLAS
Cluster of Excellence on Plant Sciences

Friday FAIRday
Session II: Storage & Backup

CEPLAS FAIR Data Team
Dominik Brilhaus & Hajira Jabeen

Friday FAIRday

Data Management Workshop Series

Save the dat(e/a)

July 2nd | 2 - 2:30 pm

Data Storage & Backup

- Independent sessions
- Open to everyone
- Online via zoom
  (check ceplas.eu for details)

Manipulated from Zoë Roth, "Disaster Girl"

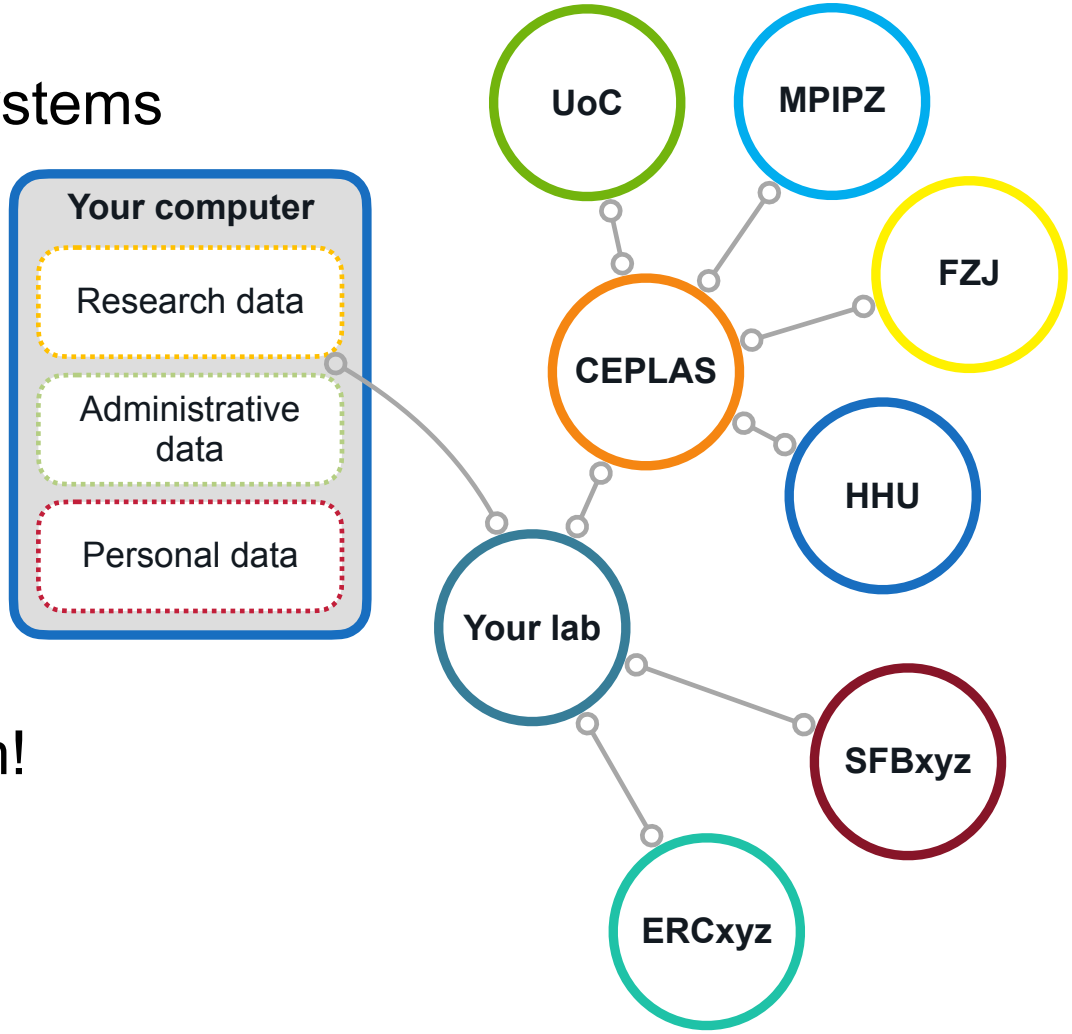
hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

In case of fire, which data will you save before running out?

Would you loose important data?

Research Data Flow

- Different environments use different systems
  - For file naming
  - Directory structure
- In the end - it's up to you:
- Define the scope
  - personal, lab, consortium
  - Don't try to change a running system!



hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Store : Planning

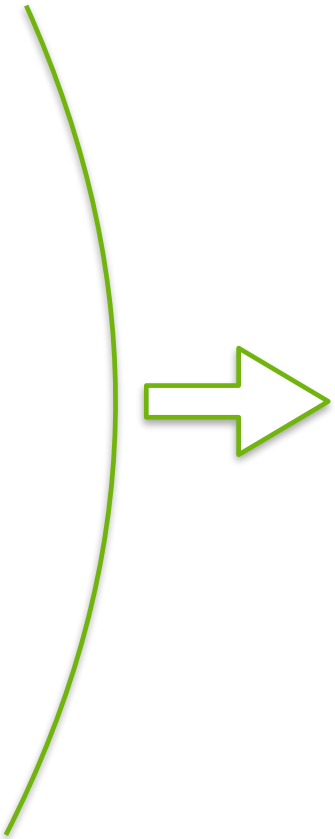
- Storing data while:
  - Data collection
  - Data analysis
  - Sharing
  - Backup
  - Archive
- Recovering data
  - Loss
  - Damage
  - Corruption

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Store

First Step: Planning

- Storing data while:
  - Data collection
  - Data analysis
  - Sharing
  - Backup
  - Archive
- Recovering data
  - Loss
  - Damage
  - Corruption



- Keeping in view
  - Volume
  - Format
  - Update frequency
  - Versioning
  - Security

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Store

Local hard disks

Local server

University server / HPC

Cloud services

sciebo
dieCampuscloud

Chat

Your data

eMail

Project management platforms / Wiki

Electronic lab notebooks

eLabFTW

labfolder

Git repositories

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Personal Computer Institute

Pros

- Easily accessible
- Fast access
- Personal preferences
- Freedom of use

Cons

- Failure prone
- Missing
  - Portability
  - Suitability for remote work
  - Backup
  - Version control
- Personal view only

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Laptop

Pros

- Same as PC
- Portable
- Remote work

Cons

- Same as PC
- Easily
  - Damaged
  - Lost
  - Stolen

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Removable media (Hard disk, USB, Tape)

Pros

- Portable
- Accessible
- Secure
- Low Cost
- High capacity
- Two layered security
  - Place where it is stored (locked environment)
  - Encryption

Cons

- Susceptible to
  - Failure
  - Shock
  - Damage (dropped, dampness etc), lost, stolen
- Encryption needs to be implemented
- Not robust for long term reliable storage
- Active backup and quality control is needed

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Institutional services (File servers)

Pros

- Professionally maintained
- Depends on policies
  - Storage
  - Easy sharing
  - Secure
  - Back-up
  - Version controlled
- Access Control
- Remote work possible

Cons

- Latency in access (Network dependent)
- Permissions and rights
- Unclear policies (e.g. security implementation)
- Lack of control/flexibility
  - e.g backups (e.g. once in 24 hours, or once a week)
- Quota for storage

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Cloud (Google, Dropbox, Sciebo)

Pros

- Usually easy to use
- Professionally maintained and backed up
- Possibly version controlled
- Easy sharing
- Remote work possible

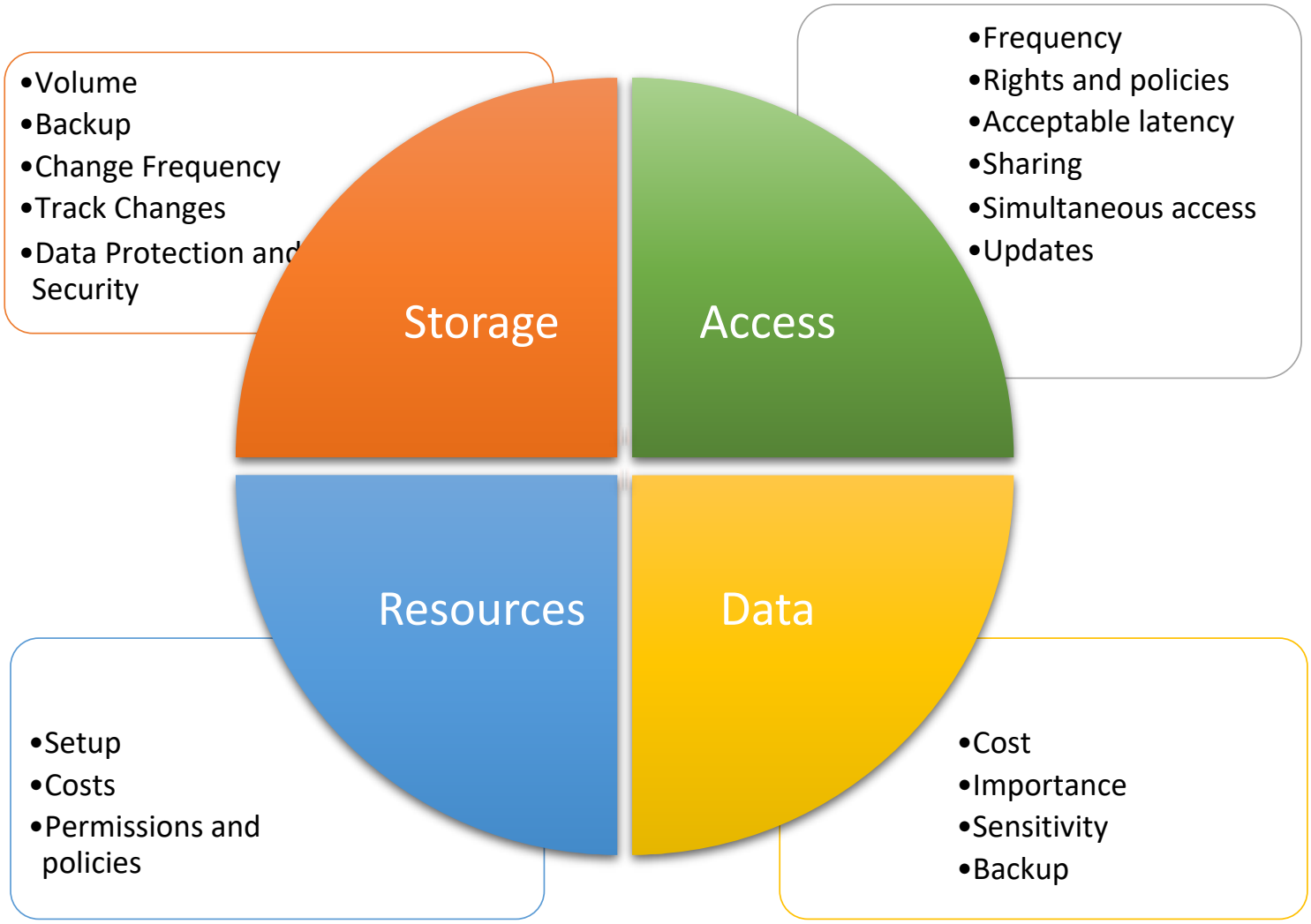
Cons

- Costly
- Security issues
  - Data Protection (unclear policies)
  - Permissions and rights
  - GDPR compliance
- Latency in access (network dependent)
- Lack of control/flexibility
  - e.g backups (e.g. once in 24 hours, or once a week)
- Quota for storage

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Storage

Key Considerations



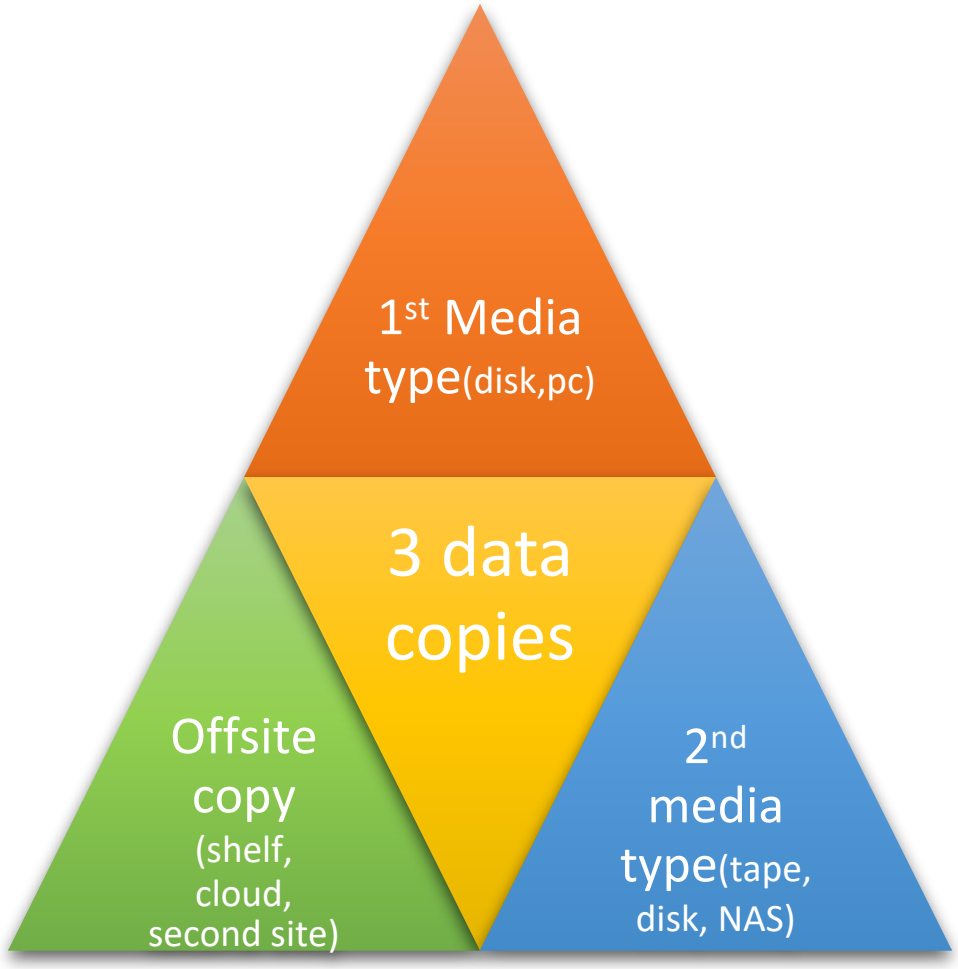
hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Backup

- Keeping multiple copies of data
- Backup process:
  - Identify important files
  - Copy them to a safe location
  - Repeat (and test) after scheduled intervals
- Caution!
  - Security
  - Cost-Benefit
  - Policies
  - Rules (funding)

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

3-2-1 Backup rule



hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Archiving

Key Considerations

- Longterm Store
  - Regularly maintained
  - Cost
  - Trustworthy
  - Longevity of the solution
- Reliability
- Sustainability
  - Reusable file formats, or also store the software
  - Well documented data (more later)

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Backup Vs Archive

	Backup	Archive
Reason	Duplication	Migration
Purpose	Disaster recovery, Data loss	Long term store, Compliance
Changes	Short term updates	No updates
Trend	Cyclic, Replacement	Growing
Store type	Short-mid-term data	Long term
Usage	Work in progress	Cold, Unused data
Retrieval	Fast/Costly	Slow/Cheaper

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Data Backup & Archiving

Key Considerations

Backup

- Volume
- Importance
- Backup frequency
- Maintenance frequency
- Sharing
- Security and policies
- Latency of access
- Cost
- Redundancy

Archive

- Volume
- Space
- File format
- Duration
- Access rights
- Documentation
- Maintenance
- Trust
- longevity

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Backup Toy Story

Toystory4 (2019)

https://i.redd.it/dwa2p7pc0wl61.jpg

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Resources

- Related Material
- Store, Backup and Archiving Checklist

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de

Thank you

Questions and Discussion

hajira.jabeen@uni-koeln.de, dominik.brilhaus@hhu.de