Archaeology Data Service

ADS Case study. Data Storage

Dr Tim Evans

07/11/2024

ads Archaeology Data Service

Planning for Disaster

Always try and plan for the worst!

Natural disasters AND Human factors.

ads Archaeology Data Service

OAIS

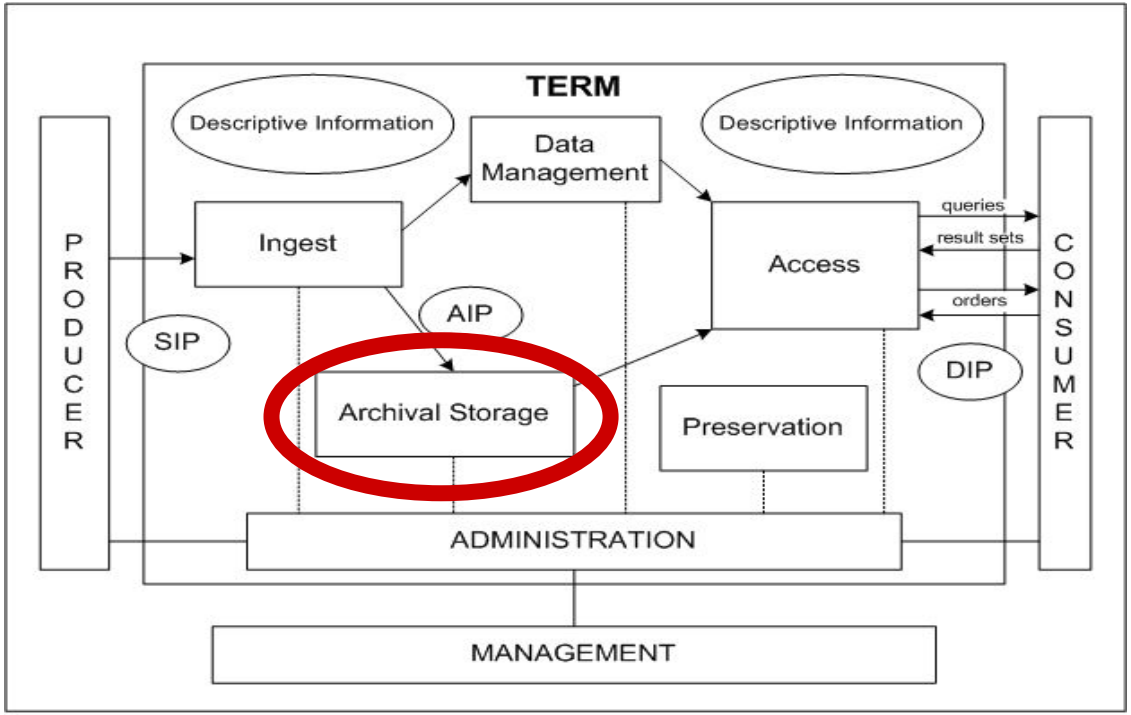
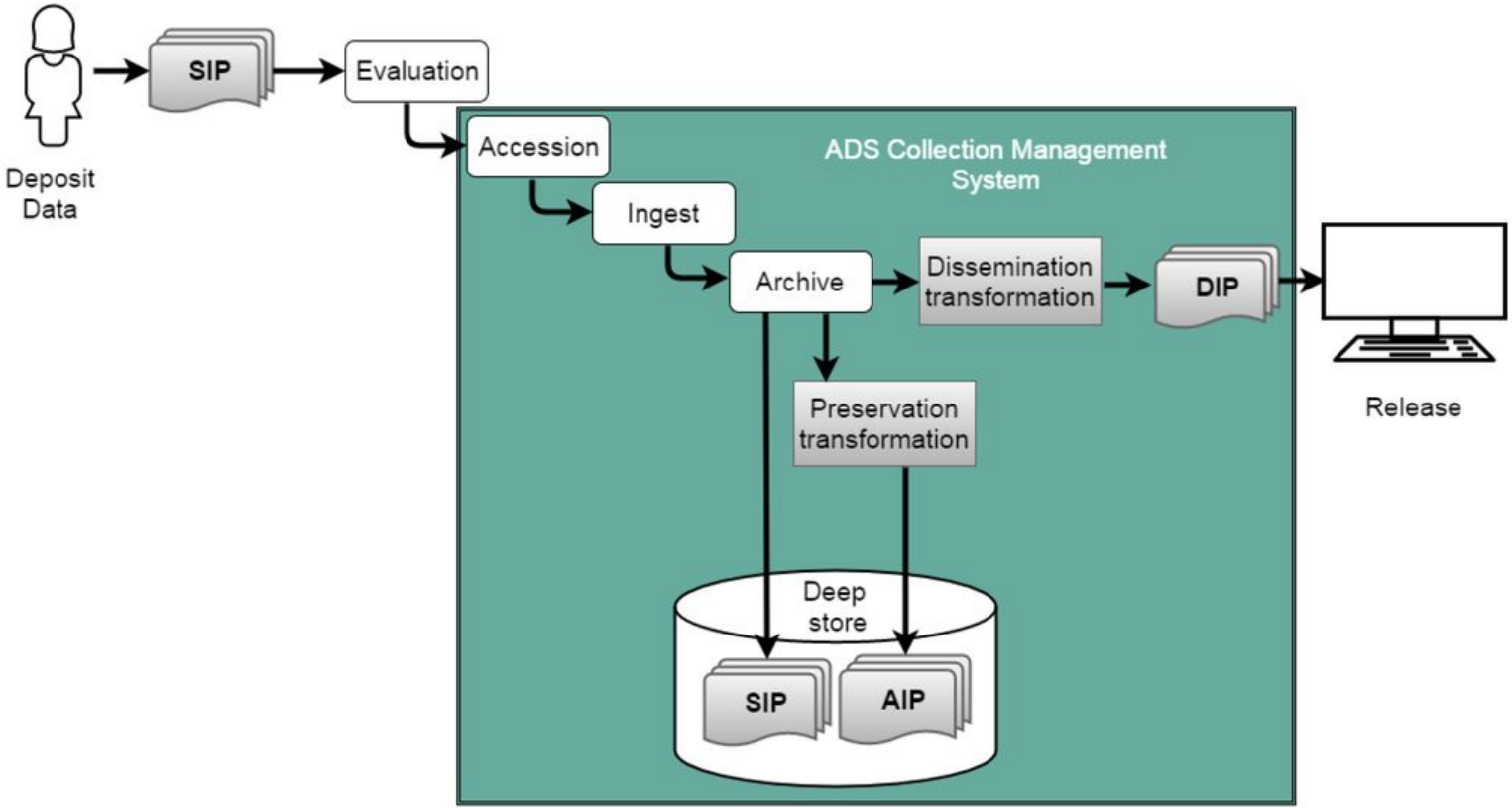
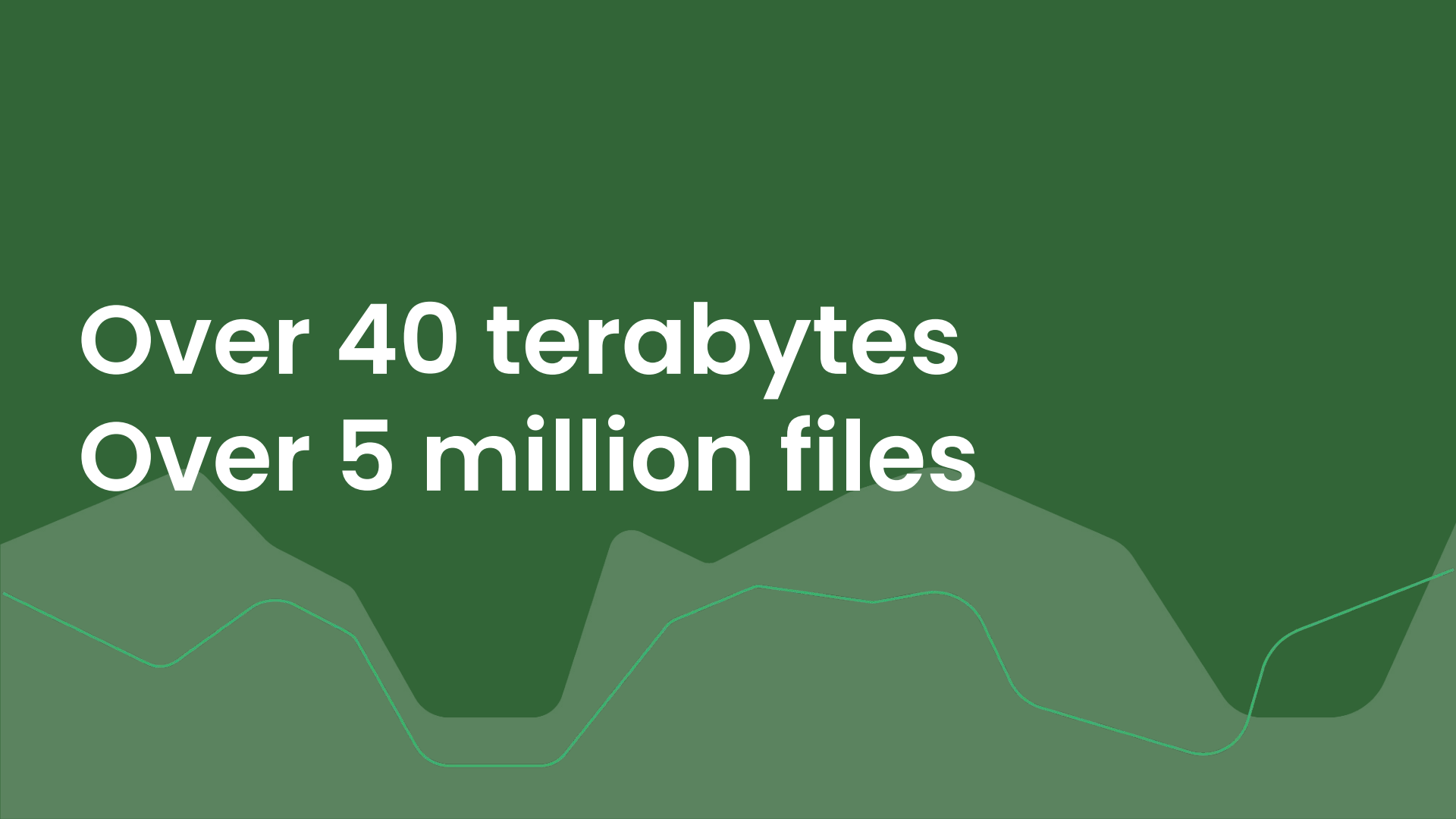


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

ads Archaeology Data Service




Over 40 terabytes
Over 5 million files

ads Archaeology Data Service

3-2-1

- Have at least three copies of your data
- Store the copies on two different media types
- Keep one backup copy offsite

Archaeology Data Service

Storage Array (A group of hard drives or solid state drives (SSDs) built into a single unit.

Archaeology Data Service

LTO (tape backup)

ads Archaeology Data Service

Amazon Glacier

Tape backup (assumed....)

ads Archaeology Data Service

Primary storage

i) Who is responsible in your organisation for monitoring your storage? Define the Role and Responsibility

ii)Decide and document where your storage is and what media is being used:

- Where is it? Is the space secure and monitored?
- What safeguards are in place to guard against data loss?

iii)Ensure your staff know what your workflows and policies are for data storage. Don't assume it's not important.

ads Archaeology Data Service

ADS example

All data and metadata (AIP, DIP and SIPs) are stored and backed up by ADS staff on the University of York (UoY) network.

On this network ADS data is stored on a pair of Dell Compellent enterprise storage arrays (current capacity ~1Pb), located in two different data centres, 2 km apart.

Each data centre is dedicated and purpose built, and has full UPS, fire suppression, generators and is "lights out" and alarmed. Within each site, data is protected by being spread redundantly across multiple disks ("RAID").

Between data centres it is replicated asynchronously, with a maximum data loss of 2 hours.

ads Archaeology Data Service

Stage 2: on site backup

i) Who is responsible for the backup?

ii) How often does the backup take place?

iii) Where is the backup?

iv) What is the backup?

v) How long is the backup kept?

vi) How do we get the backup back?

vii) Mitigate Risk

ads Archaeology Data Service

Stage 2: on site backup

i) Who is responsible for the backup?

Named member of staff (Paul Young) and a named lead in ITS

ii) How often does the backup take place?

Every night

iii) Where is the backup?

In a secure UoY ITS store - fire suppression, alarmed and secured room.

ads Archaeology Data Service

Stage 2: on site backup

iv) What is the backup?

Linear Tape-Open (LTO). Check which version e.g. The latest generation as of 2021, LTO-9, can hold 18 TB in one cartridge.

v) How long is the backup kept?

90 days

vi) How do we get the backup back?

Email request to ITS

ads Archaeology Data Service

Stage 2: on site backup

vii) Mitigate Risk

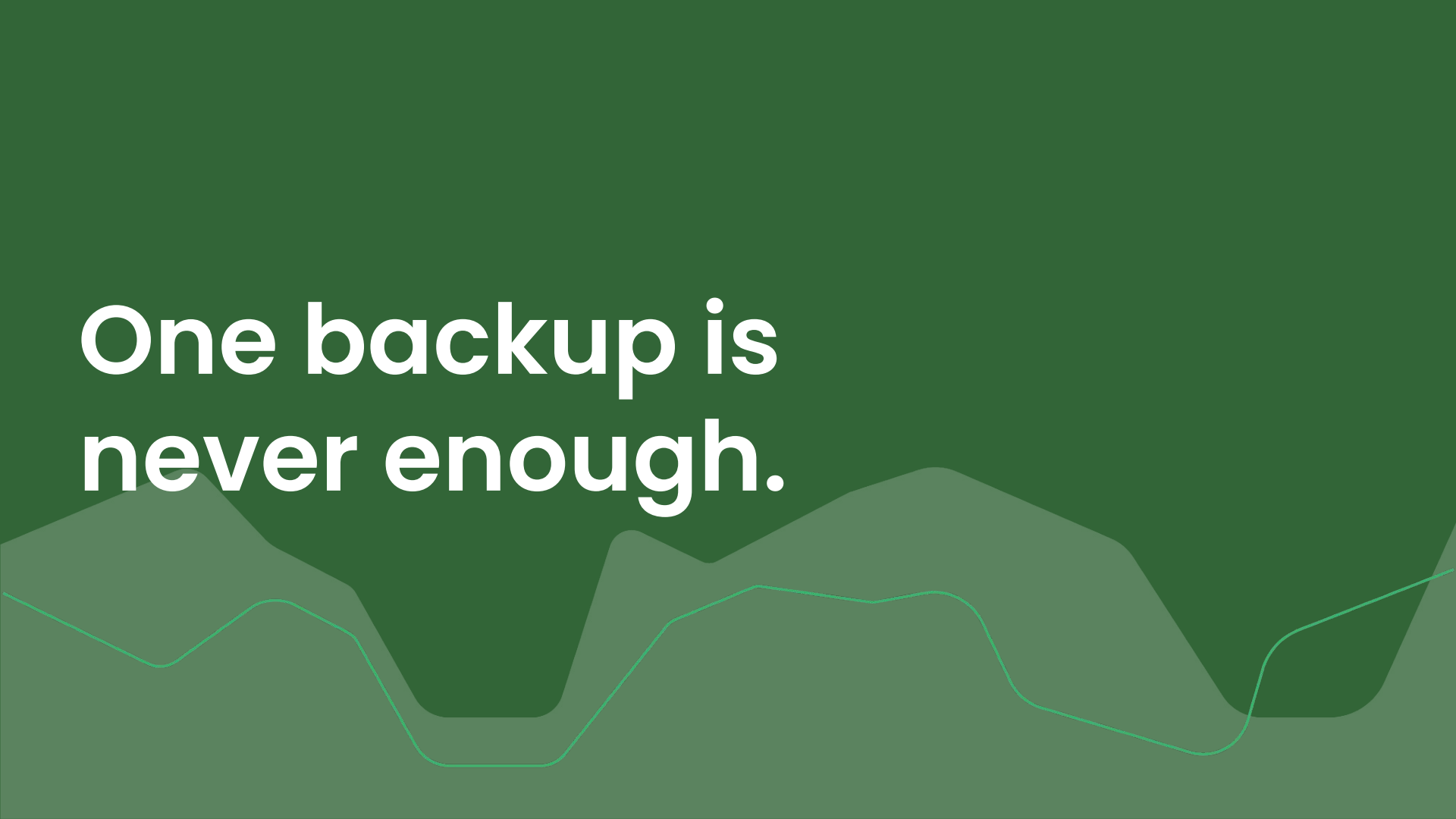
Basic: Information Security Risk Assessment

Review all the ways in which we can lose data - threats but also internal mistakes and oversights. Useful to list hypothetical examples that can inform future tests or exercises e.g. recovery of data.

Advanced: Policy for data integrity

Named responsibility for checksums and fixity values.

One backup is never enough.



ads Archaeology Data Service

ads Archaeology Data Service

Stage 3: Off-site storage

This is, effectively, a third copy of the AIP

Has to be independent of our main File Store and Backup.

Has to be reliable.

Has to be affordable - costs have to be clearly understood.

Have to be able to get data back - no hidden barriers.

Preferably “air gapped”: data cannot be deleted or altered in any way, by anyone, throughout its storage lifetime.

ads Archaeology Data Service

Cloud storage - Amazon Web Services

ALWAYS A RISK!

Advantages

Quick

Cheap

Scaleable / different options for use of APIs

“Too big to fail”

ads Archaeology Data Service

Cloud storage - Amazon Web Services

ALWAYS A RISK!

Disadvantages

Lack of clarity over exactly *where* data is

Lack of clarity over what data is stored on

Verifying local (AWS) integrity of files is in their hands…

“Too big to fail”

Hidden Charges…

ads Archaeology Data Service

Cloud storage - Amazon Web Services

ADS:

All data stored in Republic of Ireland

- Data subject to EU legislation (inc. GDPR)
- Not in the UK (disaster…)

Use of ‘deep glacier’ tier of storage: most economic for infrequent access

Use of non-public buckets (restrict access)

Cost modeller: allows me to understand what we’re spending and what we’re likely to spend

ads Archaeology Data Service

Cloud Storage - other considerations

Big commercial operations offer different types of storage (ultimately all based on the media on which it is stored) which can suit your needs.

For example, we use AWS Glacier (presumably tape) as we only need to access it in an emergency = relatively cheap. However it takes more effort and cost to retrieve files. NOT LIKE A FILESTORE WHERE YOU CAN IMMEDIATELY ACCESS THE FILE

If you use S3 which is more flexible = greater cost.

ads Archaeology Data Service

Cloud Storage - other considerations

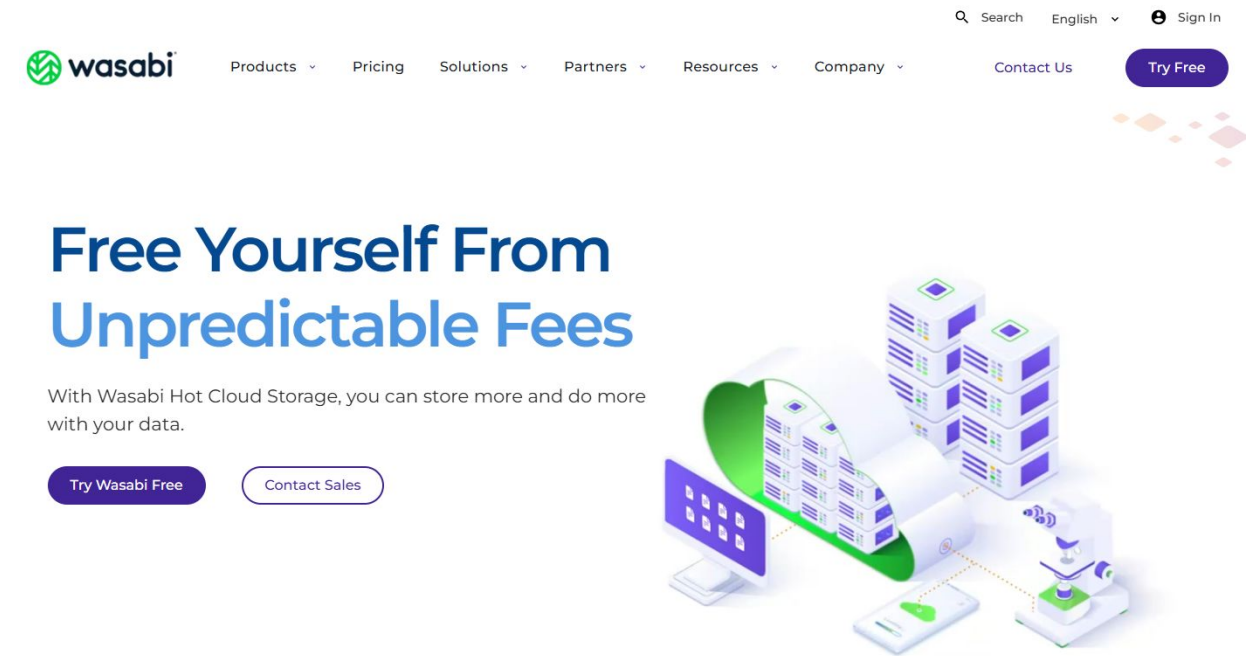
A growing number of small vendors that offer ‘simpler’ services at a competitive price.

Focus on security of your backup (“immutable”).

Focus on carbon emissions.

ads Archaeology Data Service

Cloud Storage – other considerations



ads Archaeology Data Service

Summary

There is no one-size-fits-all solution!

Key requirements are:

- Try and follow the 3-2-1.
- You have a named responsibility!
- This person (or people) have time to do this job.
- There's clear documentation on your storage.
- There are always risks with using a cloud-based service, but it does work. Just check the details!

ads Archaeology Data Service

Thankyou!

tim.evans@york.ac.uk

Archaeology Data Service
Department of Archaeology
University of York
The King's Manor
Exhibition Square
York, YO1 7EP

www.archaeologydataservice.ac.uk

help@archaeologydataservice.ac.uk