

Storage 101: “The Basics”

Storage for digital content

Storage for your digital objects

Storage for your archival digital objects

Storage for your Web content

Storage for working on projects

Storage for your day job

Many storage needs

Many storage options

RAID, NAS, SAN, CD, DVD, Blu-ray, JBOD,
MAID, Tape, S3, “The Cloud”, SRB, IRODS

Many storage options

Which one is going to be right for you?

As with all projects... Start with some planning

What is your goal?

Who are your users?

How do they use the storage?

How does the storage scale?

What are the startup costs?

What are the ongoing costs?

Quite a few things to think about...

What kinds of things are we trying to archive?

Do they change over time?

Are they stable?

How do you want to access your storage?

Instant access?

Have to pull it from an archive?

Have to pull it from a tape or disk manually?

How much storage are you dealing with?

100GB?

1TB?

100TB?

1PB?

How is your data growing?

Daily
Weekly
Monthly
Yearly

Do you want any special services on top of your
“archive”?

Search?
Fixity checks?
Data migration?
Data replication?

Understand that you have to maintain your content that sits on your storage.

The biggest, best, most environmentally stable physical archive becomes pretty useless if you just throw stuff in and don't organize things.

Storage on optical disc

(CD, DVD)

(cdr, cdrw, dvd+r, dvd-r, BlueRay)

Storage on a local harddrive

Storage on a removable harddrive

Storage on a NAS

Storage on a SAN

Now how are NAS and SAN different?

Storage in “the cloud”

Each have benifits

Each have challenges

Honestly when was the last time something didn't
come down to cost?

There is the “right way” and then there is the “right way”*

*based on available budgets and priority

Preservation Networks

Geographically Separated Storage

Lots of copies keeps stuff safe.

Even in your own institution

Replicate your data

Have an onsite and offsite version

Replicate, you mean back up right?

Not really...

But what about RAID?

RAID should be thought of as fault tolerance.

But what about mirroring (RAID 1 or RAID 0+1)

What happens if the physical system goes away?

But we are small, we don't have support, we have only a city it person we share, they don't want to store this much stuff, they don't see the value, we don't have the people, we don't have the time, we don't have the energy...

Then collaborate...

There are people in the state that are setup to help.

Not every institution needs a massive digital preservation system.

Can't we pool resources and share?

Are there models out there?

Can we work out the challenges?

Can we just try it for 5 years?

(5 years buys you a huge amount of time in the technology world)

Conclusions

Multiple copies are good

Mind your data

Keep up with things

Plan

Thank You.