Applying Open Storage to Institutional Repositories
It helps to be open...

IRs are largely characterised by “Openness”

Open Source Software
Open Access (Initiative)
...
Open Storage
How can we facilitate effective preservation survives for digital repositories?

Preserv2 is a Project, not a service provider.
Independence is also good...

No single vested interest in...

Repository Software
EPrints and Fedora are represented in the project

Storage
Working with Sun among others via the CRIG

Preservation Tools
TNAs Pronom Droid, Some connections with Planets

Preservation Service Providers
2nd European Workshop on the Use of Digital Object Repository Systems in Digital Libraries
Taking the Bull by its Horns

• We can focus on building & testing an infrastructure that connects up components.

• The aim is to serve repository managers not by answering their preservation questions one at a time or per object.

• All approach, manual curation does not scale!
Interoperability

- OAI is one enabler to Interoperability
What OAI did for Repositories

Provided a standardised way to share information pertaining to Objects in a Repository.
Interoperability

- OAI is one enabler to Interoperability

Diagram:

- Messaging
- Web Service
- OAI-PMH
  - oai_dc
  - mets
  - rem_atom
  - ...

Logos:

2nd European Workshop on the Use of Digital Object Repository Systems in Digital Libraries

Preserv.org.uk
Repository Preservation and Interoperability
ORE

OAI-ORE Serialisation
REM_ATOM / REM_RDF

REM
Description + Aggregation

DC  FOXML  METS

PDF  EPXML

2nd European Workshop on the Use of Digital Object Repository Systems in Digital Libraries
What ORE does for Repositories

Reveals the first class objects/resources in the repository.

More than just the metadata is guaranteed to be described in a Resource Map (REM)
Mining with ORE

OAI-ORE

EPrints & Fedora

Which is which?
Open Storage

• Reliable
  • Self Checking and Self Healing file System

• Resilient
  • Must have the capability to be robust in the case of part failure

• Simple & Expandable
  • Must be made of parts which are easy to expand / upgrade way into the future.

• Open
  • Any software developed to enable all of the above must be open, same with any hardware specifications.
Open Storage (2)

• Objects are addressable via HTTP
  • http://url/uuid

• Each object should be heavily tied to its UUID; you should be able to find the UUID from the object.
Repositories and Open Storage

Repository Software

Storage Plugin

Open Storage
Services over Storage

- Ingestor
- Repository Software
- Index & Search Service

Dynamic Open Storage

Preservation Storage
Summary

• Objects, including REMs are stored in Open Storage.
• Web Services like OAI-PMH will still be used.
• The repository software can be removed, replaced and more easily upgraded.
• Enables preservation as you can let a 3rd party have some say in the management of your objects.
Open Storage & The Cloud

• Open Storage provides a controlled web based environment in which objects can be stored.

• Still in the Cloud

• Reliable, Resilient, Expandable.

2nd European Workshop on the Use of Digital Object Repository Systems in Digital Libraries
Unlocking the Repository

Repository as a service rather than software running on a local ‘box’

Facilitates a greater number of services to operate directly with your resources

LOCKSS
Format Classification
Format Migration

2nd European Workshop on the Use of Digital Object Repository Systems in Digital Libraries
“The coolest thing to do with your data will be thought of by someone else!”
One more thing...

• Smart storage.

• Utilising the masses of processor power in the storage cloud to benefit your repository.

• Storelets on Honeycomb
Thanks

JISC

David Tarrant
University of Southampton
dct05r@ecs.soton.ac.uk

http://www.preserv.org.uk
Questions

• But how do I control access to my objects?
  – Use $n$ storage plugins and $n$ indexers and then perform federated search.
  – Authentication
    • Use Open Auth
    • Local Login Servers
    • Eduroam
• Complete control costs $$$
Questions

• Q: What happens if you inherit a single hard disk with a series of files on it but don’t know any of the file identifiers. You have ORE maps for the resources. How do you identify the files which go with the URIs?

• A: When droid or an identification tool classifies the file It outputs XML which contains a binary stream of the start n bytes of the file it relates to, or a hash of the file. Smart storage!