

# From the Desktop to the Cloud

---

## Leveraging Hybrid Storage Architectures In Your Repository

David Tarrant, Tim Brody & Les Carr

davetaz / tdb2 / lac @ecs.soton.ac.uk

School of Electronics & Computer Science

UNIVERSITY OF  
**Southampton**

# Summary

## 1. Storage Ecosystem

- Environmental study

## 2. Storage Controller

- Interacting with your environment

## 3. Managing Stored Assets

- Ensuring the future of your data

Where can we store data?

**STORAGE ECOSYSTEM**

# Local Disk Storage

- No local bandwidth costs
- Hard to expand
- Locally Managed
- High overheads cost
- Requires space and cooling
- Tied closely to the software



# Local Archival Storage

- Specialist
- Expensive to purchase
- Locally Managed
- Space and running costs
- Expandable



# Cloud Storage

- Scalable
- Externally controlled
- Known Costings
- Unclear retention policy
- Re-Useable (using simple APIs)
- Global Scale



# But Clouds Blow Away

In the last 10 months:

- Yahoo Briefcase
- XDrive
- AOL Pictures
- HP Upline
- Sony Image Station

# Why use Hybrid Storage

- Use the best features of each storage type
- Performance
  - Scaling-up bandwidth
- Optimisation
  - Large-file handling
  - Multimedia streaming
- Localised Delivery
  - Local delivery from the cloud

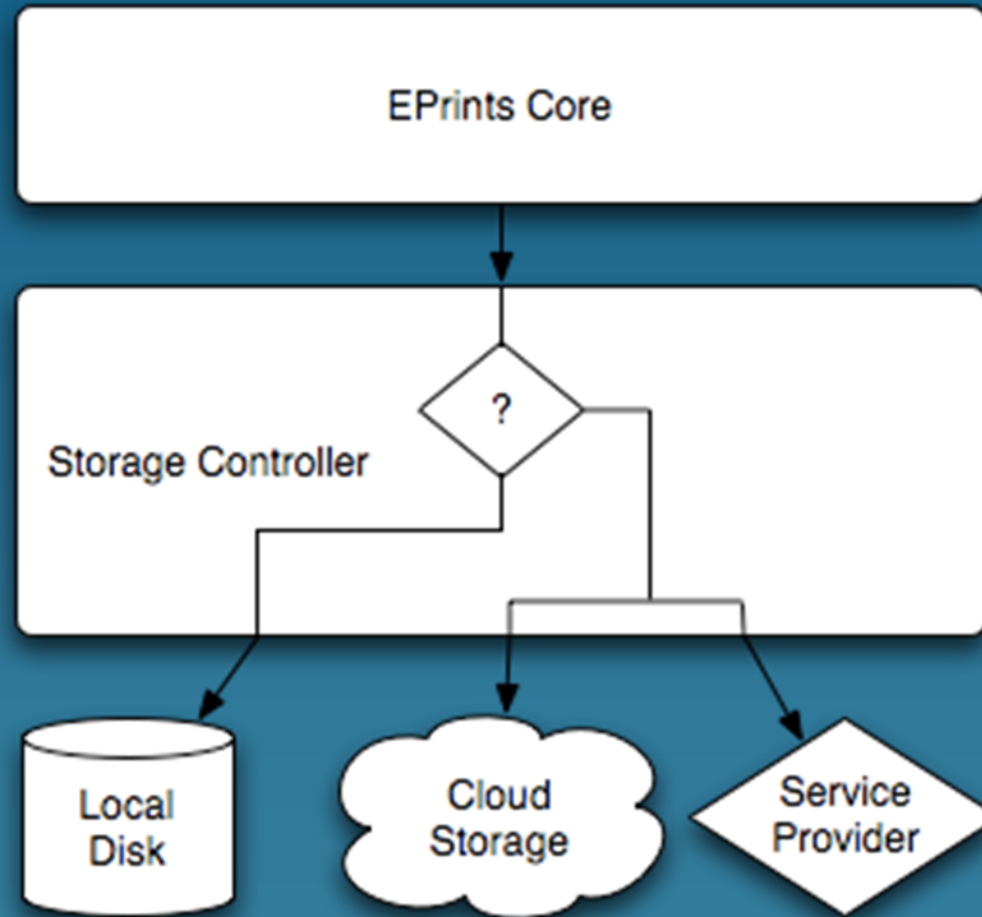
Which storage should we use?

**STORAGE CONTROLLER**

# EPrints Storage Controller

- The storage controller decides where to put a file.
- Uses rule based policy defined by simple configuration file (XML)
- Examples:
  - Large binary files of scientific data (raw machine result data) can be stored in a large disk (slower access) system and sent to a tape company for long term storage.
  - Processed results can be stored locally and in the cloud ready for rapid delivery to end points.

# Architecture Diagram



# Controller Ruleset (1)

```
<choose>
  <when test="datasetid = 'document'">
    <choose>
      <when test="$parent{relation_type} = 'isVolatileVersionOf'">
        <plugin name="Local"/>
      </when>
      <otherwise>
        <plugin name="AmazonS3"/>
      </otherwise>
    </choose>
  </when>
  <otherwise>
    <plugin name="Local"/>
  </otherwise>
</choose>
```

# Controller Ruleset (2)

```
<choose>
  <when test="datasetid = 'document'">
    <choose>
      <when test="$parent{relation_type} = 'isVolatileVersionOf'">
        <plugin name="Local"/>
      </when>
      <otherwise>
        <plugin name="AmazonS3"/>
      </otherwise>
    </choose>
  </when>
  <otherwise>
    <plugin name="Local"/>
  </otherwise>
</choose>
```

How do I move data around?


# MANAGING STORED ASSETS

# EPrints Storage Manager

## Storage Manager

Amazon S3 storage

There are 217 total files stored using this back-end, taking 3126Kb.

Documents:  217


Copy to

Delete Copies

---

Local disk storage

There are 289 total files stored using this back-end, taking 1649Kb.

History:  289

Copy to


Delete Co

- Compressed local disk storage
- Amazon S3 storage

---

Compressed local disk storage

There are 85 total files stored using this back-end, taking 293Kb.

History:  85


Copy to

Delete Copies

# Amazon S3 Localisation (1)

**African Elephants in Τάρταρος**

Léricolais, Y. (1998) *African Elephants in Τάρταρος*. In: 4th Conference on Animal Things, 1-9 June, Dallas, Texas.

 PDF  
[Download \(11Kb\)](#) | [Preview](#)

**Abstract**

This is where the abstract of this record would appear. This is only demonstration data.

<http://yomiko.ecs.soton.ac.uk:8080/1/1/paper.pdf>

[P Language and Literature > PN Literature \(General\) > PN2000 Dramatic representation: The Theater](#)

<b>ID Code:</b>	1
<b>Deposited By:</b>	Unnamed user with email <a href="mailto:admin@localhost">admin@localhost</a>
<b>Deposited On:</b>	12 May 2009 16:25
<b>Last Modified:</b>	12 May 2009 16:25

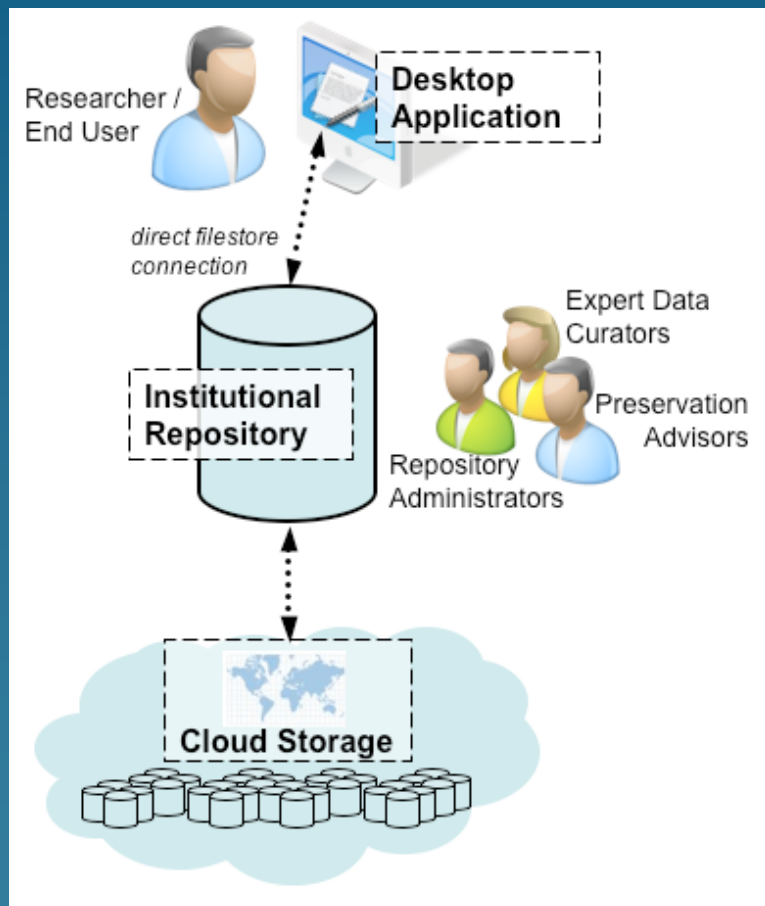
Repository Staff Only: [item control page](#)

http://yomiko.ecs.soton.ac.uk:8080/1/1/paper.pdf

# Amazon S3 Localisation (2)



# One More Thing



- Full end to end integration is ready to go in EPrints 3.2
- Come to our user group sessions to find out more on EPrints 3.2 and for demo's.

# Thank You

Storage Controller for EPrints with thanks to

The logo for JISC, consisting of the letters 'JISC' in a bold, orange, sans-serif font.

David Tarrant, Tim Brody & Les Carr

davetaz / tdb2 / lac @ecs.soton.ac.uk

School of Electronics & Computer Science

UNIVERSITY OF  
**Southampton**

# Hybrid Storage In EPrints

- Using a single storage platform or solution has drawbacks.

## Cost vs. Speed vs. Reliability

- If repositories are to provide good preservation then they need to utilize and be able to migrate to new platforms.
- In this presentation we look at backing your repository with a Hybrid storage solution.
- Gives you the power to utilize the benefits of each solution without losing control of your digital objects.

