Planning and Managing Automated Services for Ingest, Storage and Access

Martin Hall-May
IT Innovation Centre
18th October 2012
PASIG Dublin
Introduction

• We have developed software
  – for archives and service providers
  – who want to be able to plan for the future and have control through service level agreements, monitoring and automatic management systems

• Three tools integrate to implement the whole of the plan-do-check-act control cycle:
  – a preservation planning tool
  – a service management framework
  – a data storage and processing service
Storage

- Is not 100% safe
- Becomes obsolete quickly
- Total cost is high, but falls quickly
- Fast access and safety don’t always go together

<table>
<thead>
<tr>
<th>Medium</th>
<th>Storage Density bits/cm²</th>
<th>Life, years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone</td>
<td>10</td>
<td>10000</td>
</tr>
<tr>
<td>Paper</td>
<td>$10^4$</td>
<td>1000</td>
</tr>
<tr>
<td>Film</td>
<td>$10^7$</td>
<td>100</td>
</tr>
<tr>
<td>Disc</td>
<td>$10^{10}$</td>
<td>10</td>
</tr>
</tbody>
</table>
An archive provides **services**

- Ingest, access
- Safe storage
- Formats
- Metadata
- Rights

- They all cost money
- They all take time
- Never enough of either!
Plan-do-check-act

- Storage planning & Optimisation
- Storage & Processing
- Monitoring, Management & SLAs
- Policies & Automation
Services need planning and managing

- Service Level Agreements (SLA)
  - What the service does
- Quality of Service (QoS)
  - How well it does it

- If you can’t **measure** it then you can’t **manage** it
  - Throughput
  - Quality
  - Cost
  - Risk

© University of Southampton IT Innovation Centre 2012
Suppliers

Content Producers and Consumers

Archive Operator

QoS, QoE & SLAs

Management Channel

Set policies

Modelling

Plan

Management actions

Monitoring data

Act

Do

© University of Southampton IT Innovation Centre 2012
Example

SLA Monitoring

Too many invalid SIP uploads?  Suspend ingest for customer

Verification results

Suspend ingest

Ingest

Verify MXF wrapper  Transcode for archive  Checksum  Replicate

SIP upload

Too many invalid SIP uploads?

Suspend ingest for customer

Verification results

Suspend ingest

Ingest

Verify MXF wrapper  Transcode for archive  Checksum  Replicate

SIP upload
Planning and managing include compromises

- Volume
- Quality
- Deadline
- Budget

- Digitisation workflows
  - Cost, throughput, quality
  - E.g. QC v.s. automation

- Storage strategies
  - Cost, risk of loss
  - E.g. copies v.s. cost

- Online access services
  - Cost, QoS, Users
  - E.g. KPI v.s. customers
Storage SLA Terms

- **Availability**
- **Integrity / Safety**
  - How to measure?
- **Ingestion time**
  - Indexing, generating access copies
- **Search time**
- **Delivery time**
  - From request to start of delivery
- **Bandwidth**

- **Subscription fee**
- **Charge for data**
  - On disc
  - Ingest
  - Access
- **Charge for CPU**
- **Charge per user**

- **Maximum storage size**
- **Maximum number of users**
Availability

I need the service to be available almost all the time

Can you be more specific?

I need the access service available 99.9% of the time

Is that measured over a day, month or year?

The access service must be available 99.9% of the time each month

That’s 43 minutes of downtime each month – what if that’s all in one go one afternoon?

When can maintenance be done on the service?

Do you want different uptimes for day and night?

When is “daytime” for an international operation?
MServe – Ingest
MServe – Workflows
Ting – SLAs

Please choose component that matches your role and task below:

Customer-facing
- Define service offer
- Monitor usage
- Aggregate across customers
  - Understand load
  - Inform resourcing and pricing decisions
- Automate mgt. action
  - Reduce costs

Supplier-facing
- Monitor use of resources
- Aggregate across suppliers
  - Compare service offering with experience

Plan & Optimise
For Service Providers
Predict future trends
Optimise your configuration
Improve your policies

Open Planner
Policy-based Mgmt.

Monitors:
- Availability
- Data ingest and access
- Disc space
- Errors in files
- Delivery time

Manages:
- Access (e.g. bandwidth)
- Ingest (e.g. suspend for invalid SIPs)
Predict Future Trends

If we change nothing:
- Will I lose any data next year?
- How many assets will be at risk?
- What will the running costs be?

If we store another copy:
- How much will storage costs increase?
- How much safer will it be?
Given the current state:

- How often should I be scrubbing the data?
- How many copies should I keep?
- How much resource should I dedicate to access?

... whilst keeping the data **safe** and the cost within **budget**.
Probability of losing 2% of archived assets over 10 years

Number of additional copies

Scrubbing every X months

Cost

© University of Southampton IT Innovation Centre 2012
Summary

• Archives provide an increasingly important set of services
• These services are now used throughout the media lifecycle
• Industry simultaneously moving to file-based workflows
• Planning and active management becomes essential
• Our tools help show that’s possible
Software and Publications

http://prestoprime.it-innovation.soton.ac.uk/