

Reliable Audiovisual Archiving Using Unreliable Storage Technology and Services

Matthew Addis

IT Innovation Centre

12 September 2009

IBC session "Archives: Just how reliable are they"



IBC2009

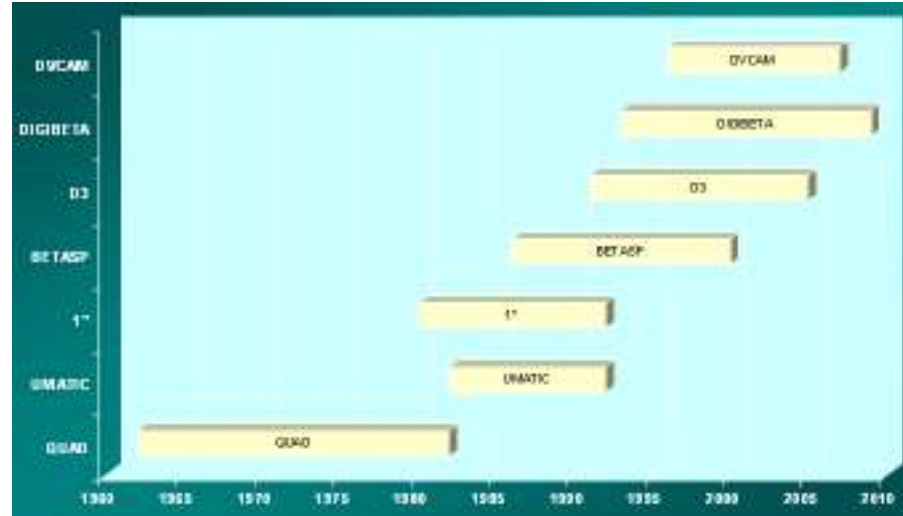
Traditional AV archiving



PrestoSpace estimate: 6M hrs across 20 major European archives
UNESCO estimate: 200M hrs of film, video and audio in total

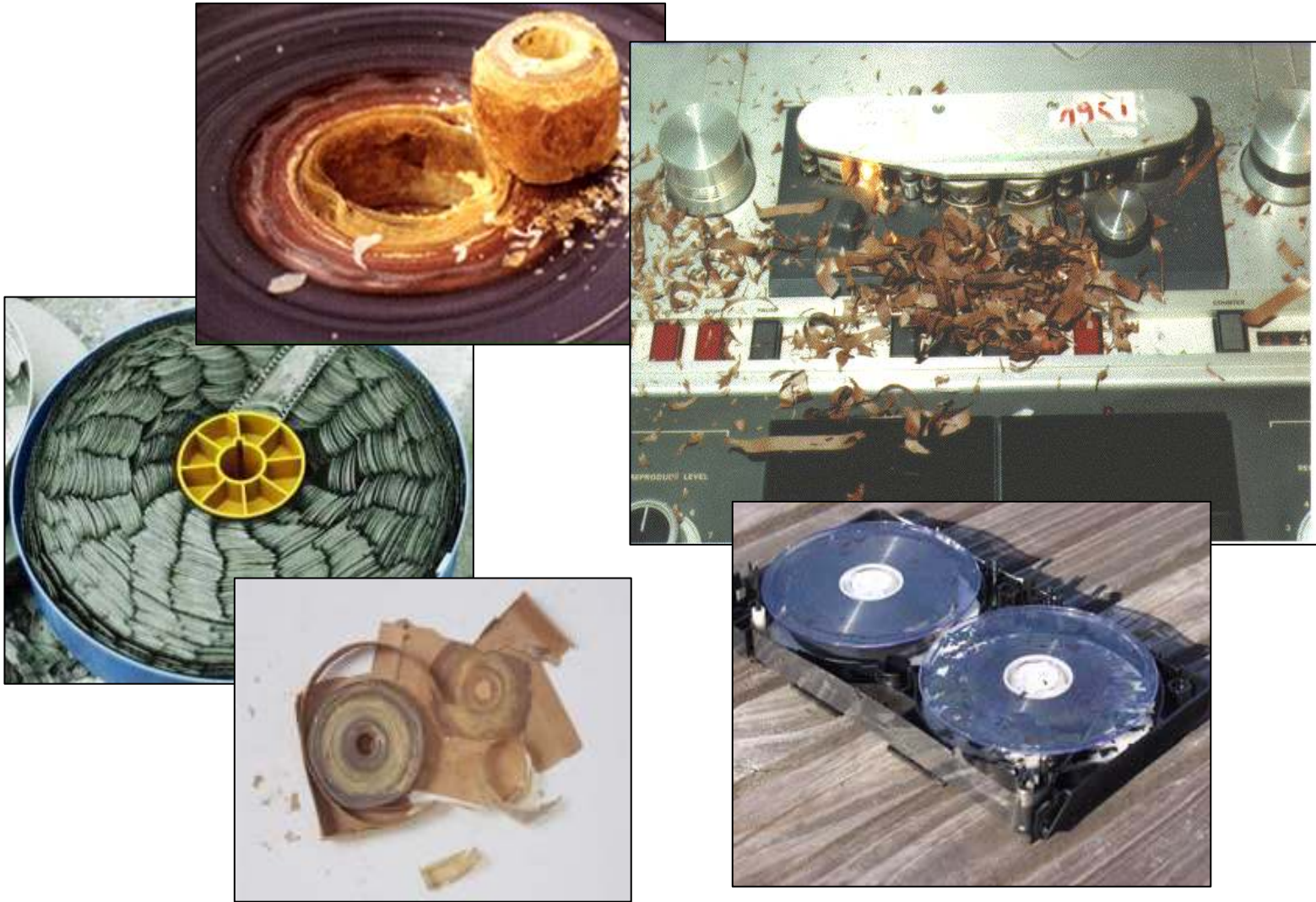


IBC2009





IBC2009



1/3 of material has deterioration

1/4 of material cannot be released as it is too easily damaged



IBC2009



FOR COLOR TAPES... This... reveals high level video recorder. Does everything. Not recording (and) that this very level perfectly available records video head of it addition, it offers a whole lot of new features. It "master" equipment from 302.



v/ounds



At least 2/3 of the material cannot be easily used



IBC2009

Expectations/Assumptions

- Archivist
 - Over 50 years of experience with media on shelves
 - Well developed practices
 - Uncertainty on how to transfer this to a digital world
- Management
 - Archiving of files is an IT problem
 - IT technology is safe
 - IT service providers know how to deal with AV
- Assumption: mass storage technology and digital files will make many problems a thing of the past...



IBC2009

Media degradation

- Is digital better?
- No, it's worse!

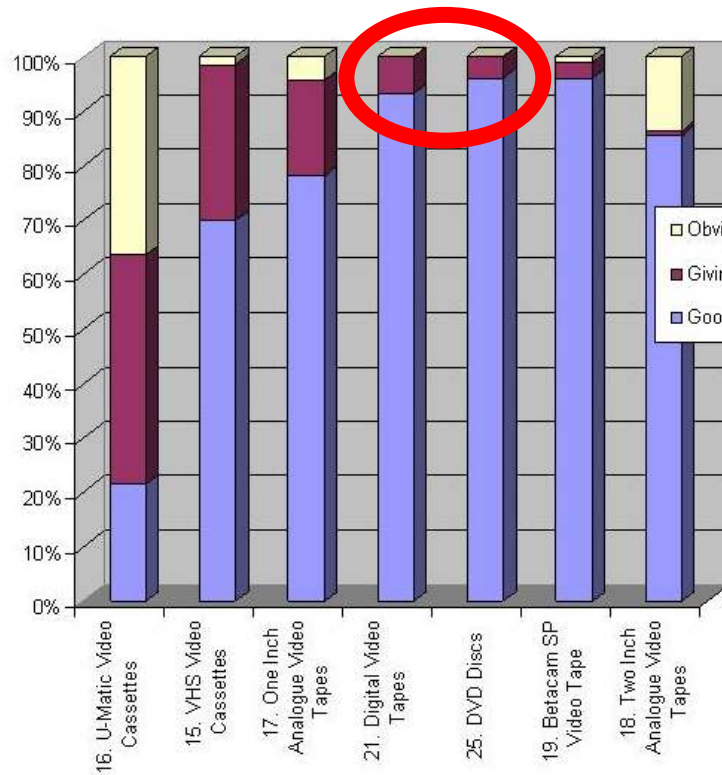
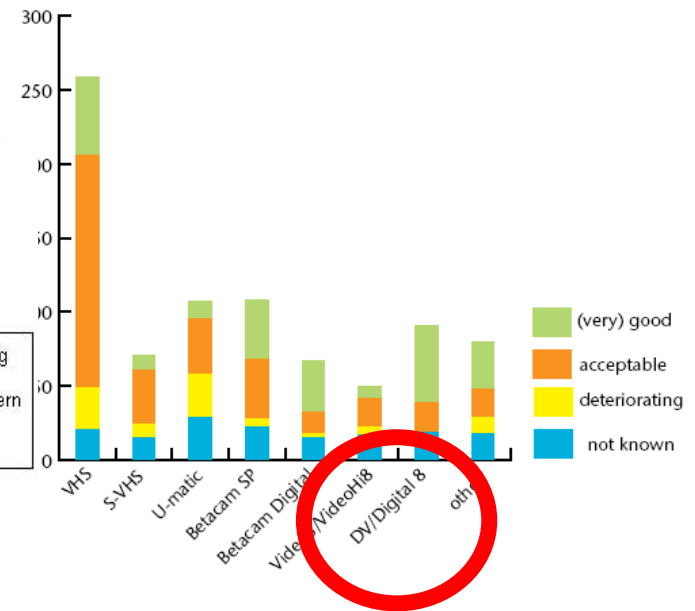


Figure 4-11 Condition of video collections

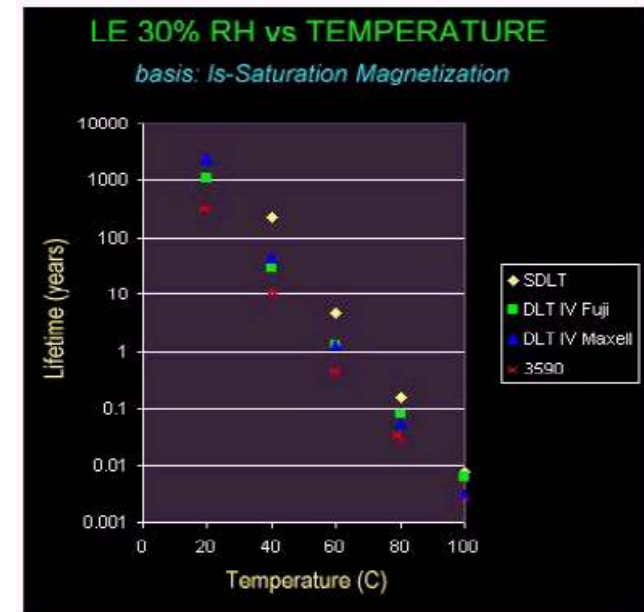




IBC2009

Data tape

- Claims
 - LTO tapes last for 30 years
- Reality
 - Tapes are not 100% reliable
 - You get your money back, not your data

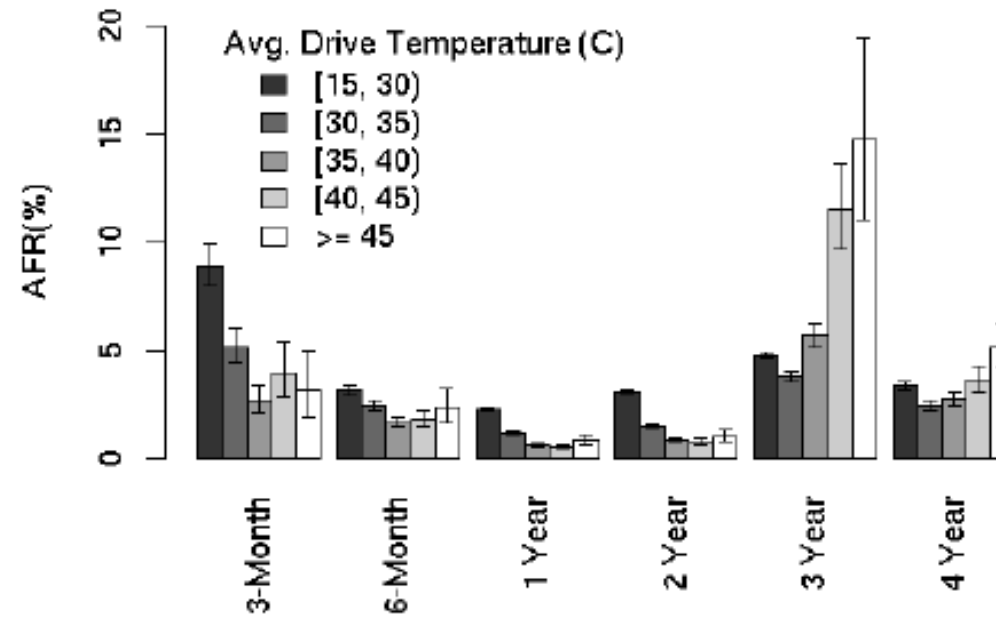


	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
3M	17	14	12	9	11	14	9	13	13	10	6	6	6	5	10	14
AGFA					43		51	67			42		60	40	14	0
AMPEX						34		14	17	20	12	9	10	8	8	0
BASF				13	23	3	21	41		9	11	19	14			
FUJI		15		5	24	49	13	3	8	8	11	12	17	9	8	
KODAK										20	23	17	12	3	3	
SONY	28	21	19	18	21	23	22	26	19	13	14	14	9	10	8	4



IBC2009

Hard drives



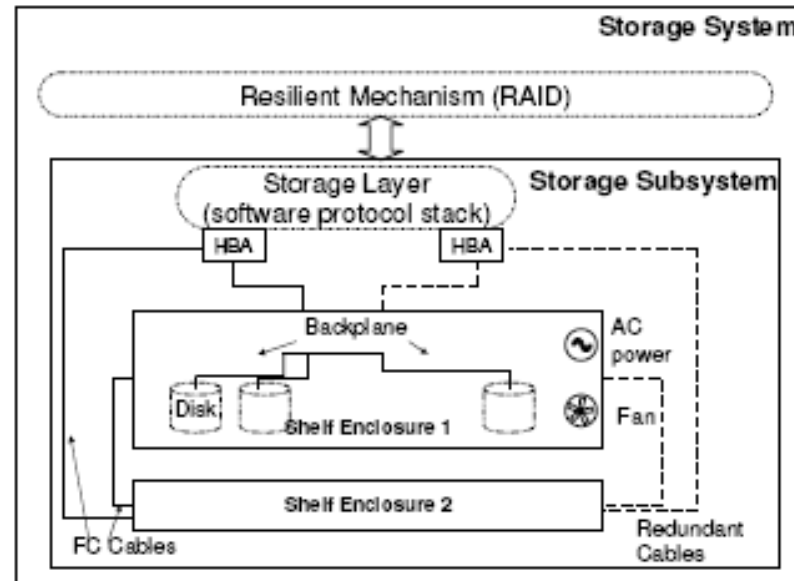
(Schroeder and Gibson 2007)

- Manufacturers: MTTF is 1,000,000 hrs, i.e. 100 years.
- Google study : real-world MTTF is up to 10 times higher



IBC2009

Storage is a system



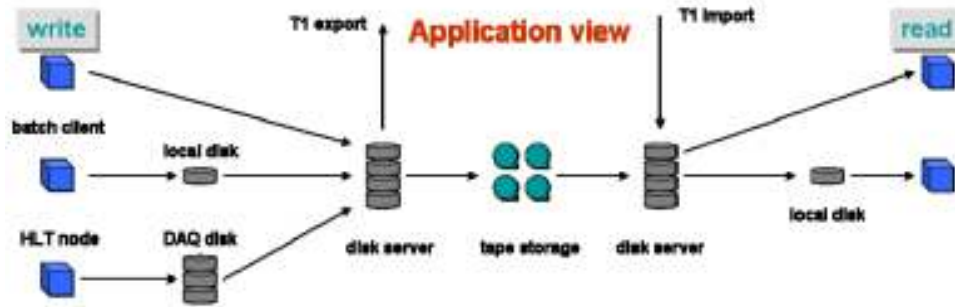
(Jiang 2008)

- Finding (1): In addition to disk failures (20-55%), physical interconnect failures make up a significant part (27-68%) of storage subsystem failures. Protocol failures and performance failures both make up noticeable fractions.
- Implications: Disk failures are not always a dominant factor of storage subsystem failures, and a reliability study for storage subsystems cannot only focus on disk failures. **Resilient mechanisms should target all failure types.**



IBC2009

Systems are never perfect



CERN found 1 bit in 10^9 permanently corrupted within 6 months



You can't stop bit rot!

Manage it

- Quantify
- Minimise
- Manage

RAID	Scrub	Sector Checksum	Block Checksum	Parent Checksum	Write-Verify	Physical ID	Logical ID	Version Mirror	Chance of Data Loss
✓									0.602%
✓	✓								0.602%
✓	✓	✓							0.322%
✓	✓		✓						0.041%
✓	✓			✓					*0.486%
✓				✓					*0.153%
✓	✓		✓	✓					0.002%
✓	✓		✓			✓			0.038%
✓	✓		✓				✓		*0.033%
✓			✓				✓		*0.010%
✓	✓		✓			✓	✓		*0.031%
✓			✓			✓	✓		*0.010%
✓	✓		✓				✓	✓	*0.004%
✓			✓				✓	✓	*0.002%
✓	✓		✓			✓	✓	✓	0.000%

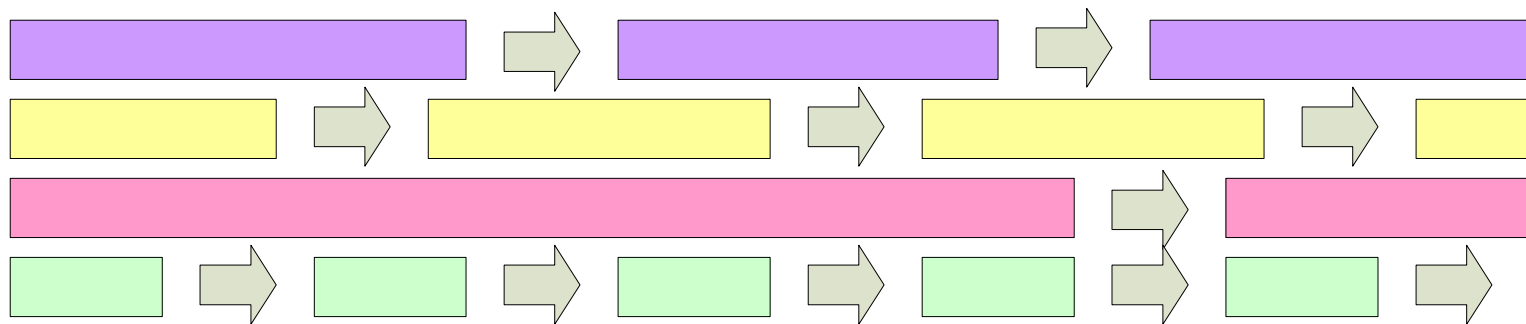
Table 3: **Probability of Loss or Corruption.** The table provides an approximate probability of at least 1 data loss event and of corrupt data being returned to the user at least once, when each of the protection schemes is used for storing data. It is assumed that the storage system uses 4 data disks, and 1 parity disk. A (*) indicates that the data loss is detectable given the particular scheme (and hence can be turned into unavailability, depending on system implementation).



IBC2009

Systems become obsolete

- Encoding formats
- Media formats
- Storage hardware
- Operating systems
- Management software
- Networking





IBC2009

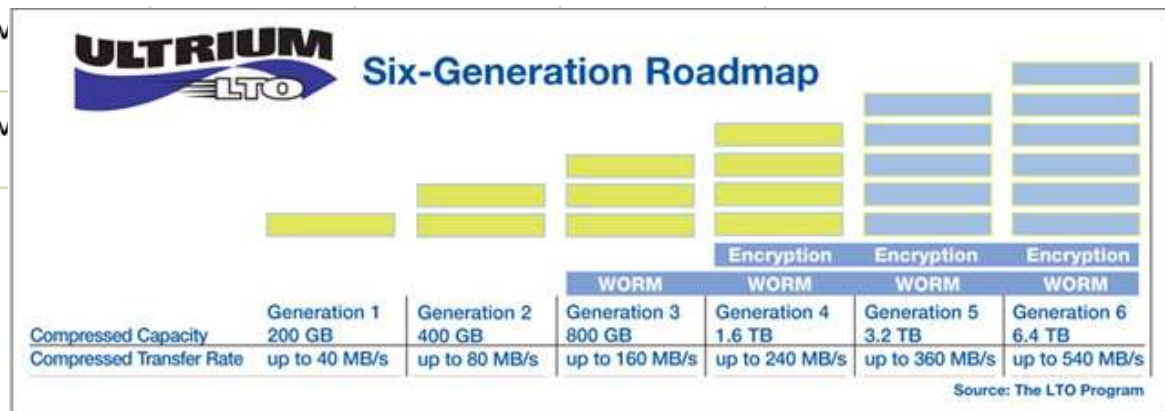
Obsolescence is getting faster



	LTO 4	LTO 3	LTO 2	LTO 1
Ultrium 4 Drive	Read/Write at 800 GB LTO 4 WORM Cartridge: Read/Write once	Read/Write at 400 GB LTO 3 WORM Cartridge: Read/Write once	Read only	Not compatible
Ultrium 3 Drive	Not compatible	Read/Write at 400 GB LTO 3 WORM Cartridge: Read/Write once	Read/Write at 200 GB	Read only

Ultrium 2 Drive

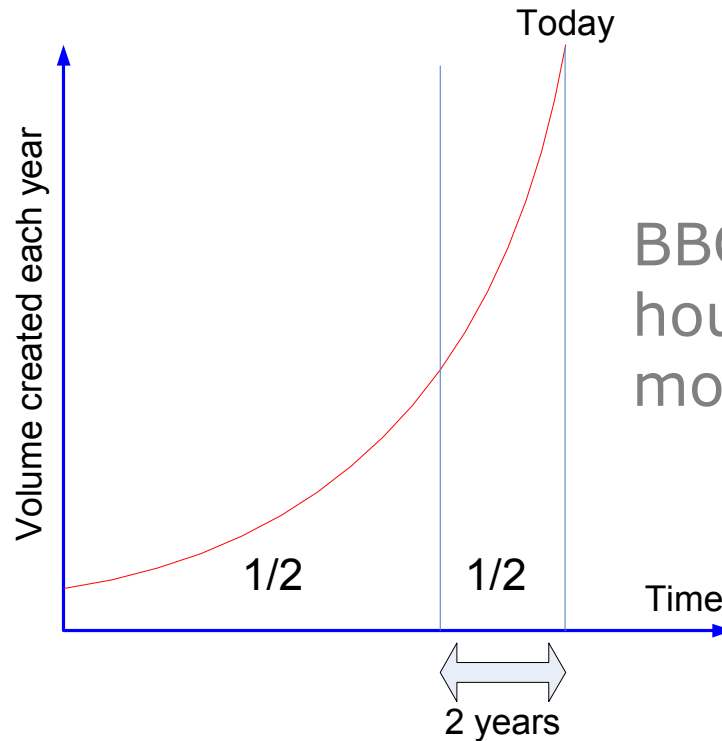
Ultrium 1 Drive





IBC2009

Growth



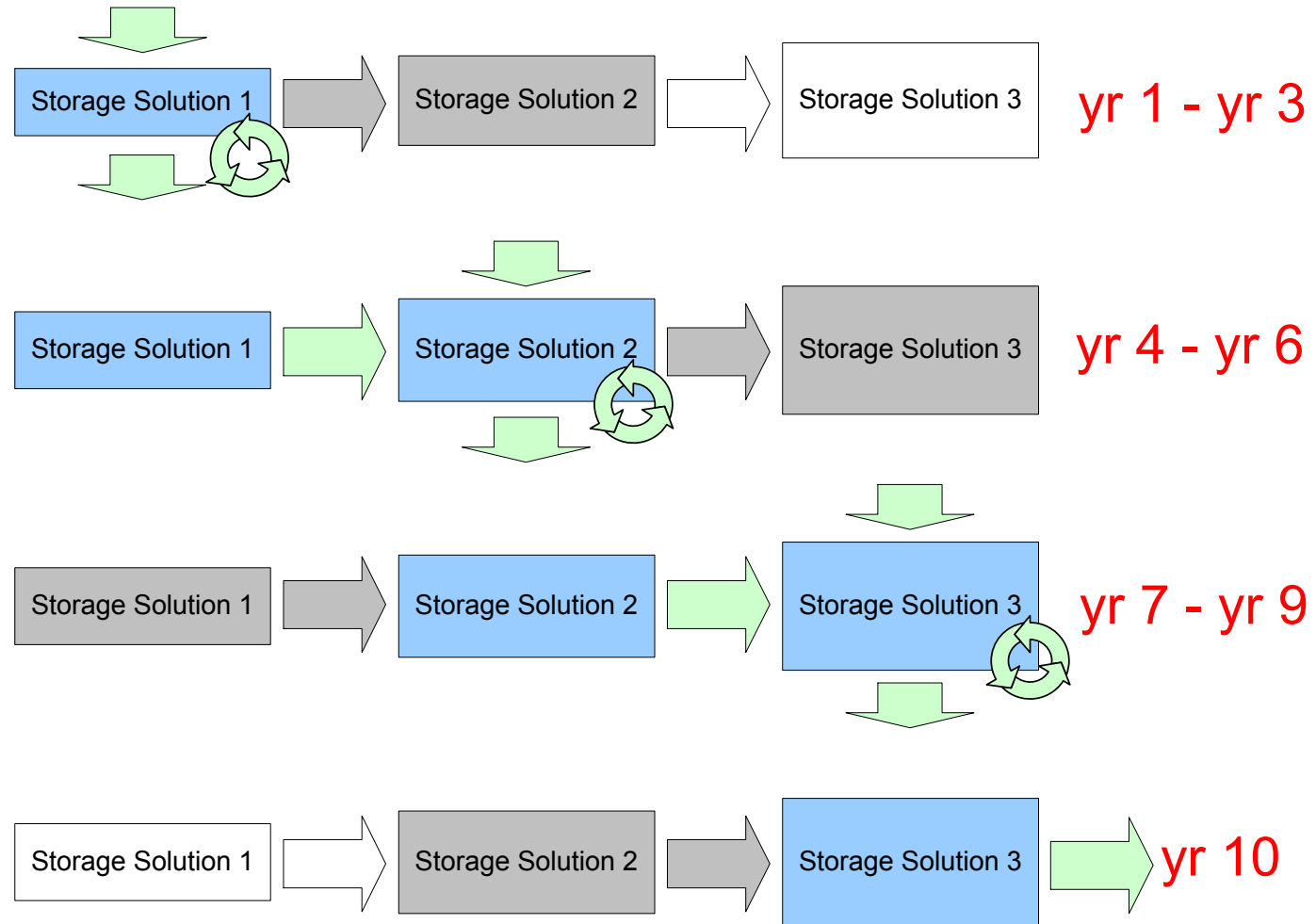
BBC archive intake is about 300 hours/week = 30 TB/week, moving to 120 TB/week for HD

Broadcast archives grow at 5M new hours per year
Data volumes can double in 2 years



IBC2009

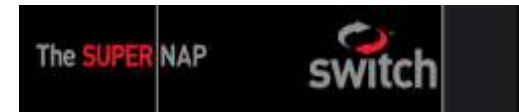
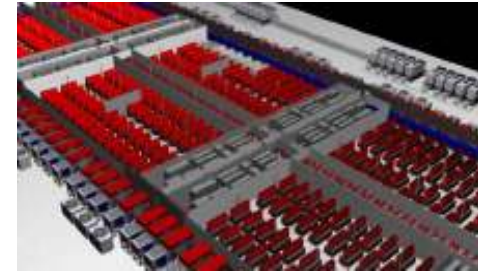
Automation and migration





IBC2009

...and it all costs **money**



- Amazon: \$0.12 per GB per month for more than 500TB
- SDSC: \$1000 for 1TB online for 1year (half this for tape)
- BBC: traditional tapes on shelves costs 20 times less (but neglects migration)
- 1PB online for 1 year has a TCO of \$1M

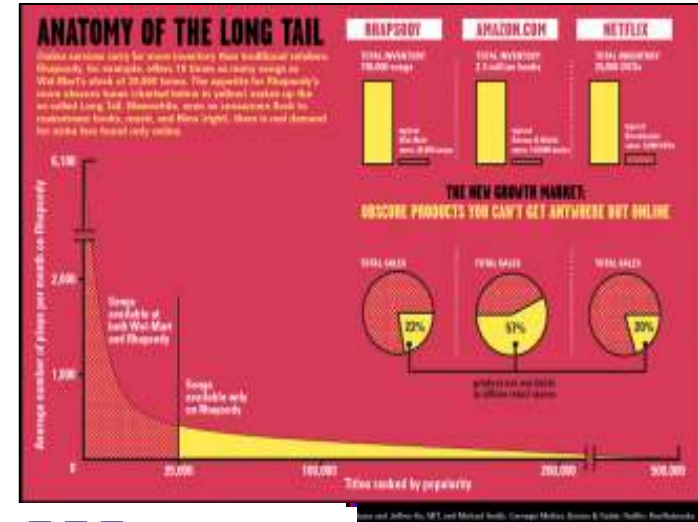


IBC2009

...and everyone wants to access it



100 000
émissions
radio télé



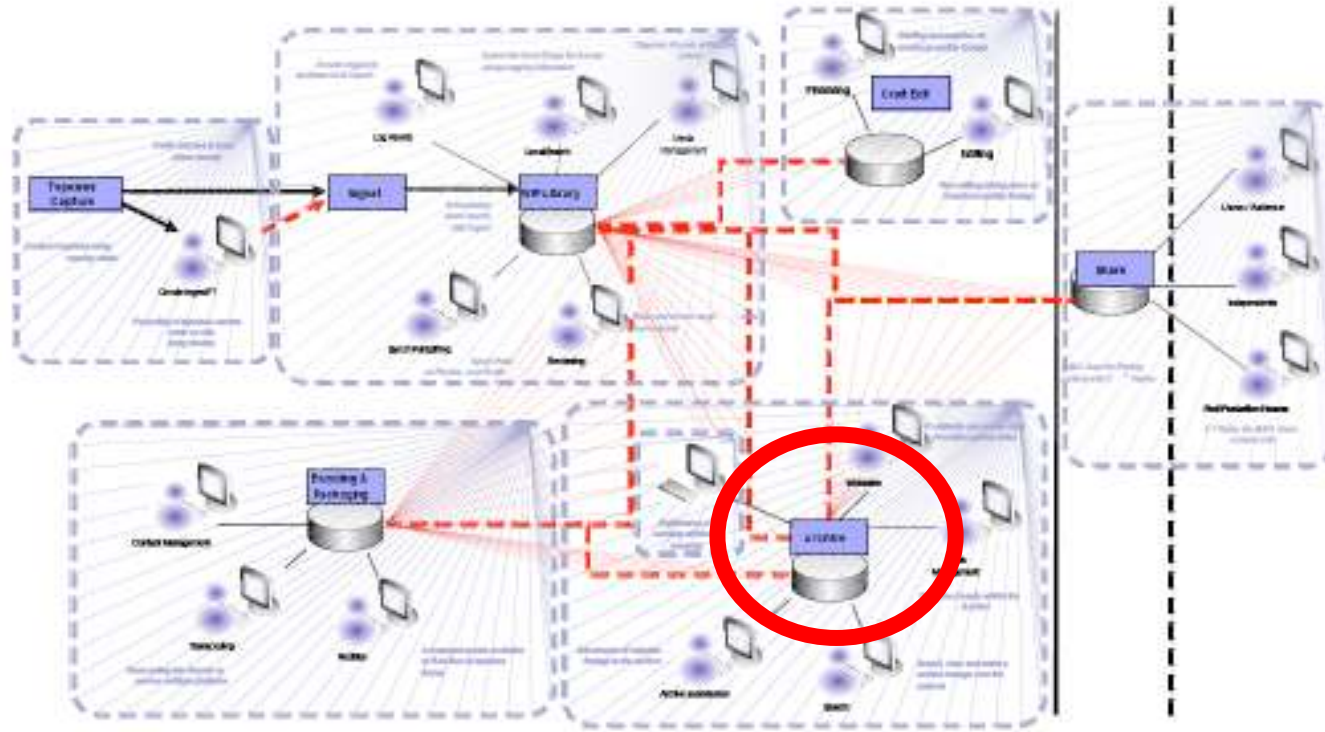
This is 4oD
The best of Channel 4, on demand





IBC2009

Business need for integration



(BBC DMI 2007)

- 95% of BBC archive use is internal
- The broadcast world is going tapeless
- Networked sharing of file-based content



IBC2009

So what do archives want?

- Long term data integrity
- From large scale storage
- At low cost
- Based on open and portable solutions
- That provide easy access to files for a wide range of applications

- Not much to ask for then!

(essentially a risk management problem)

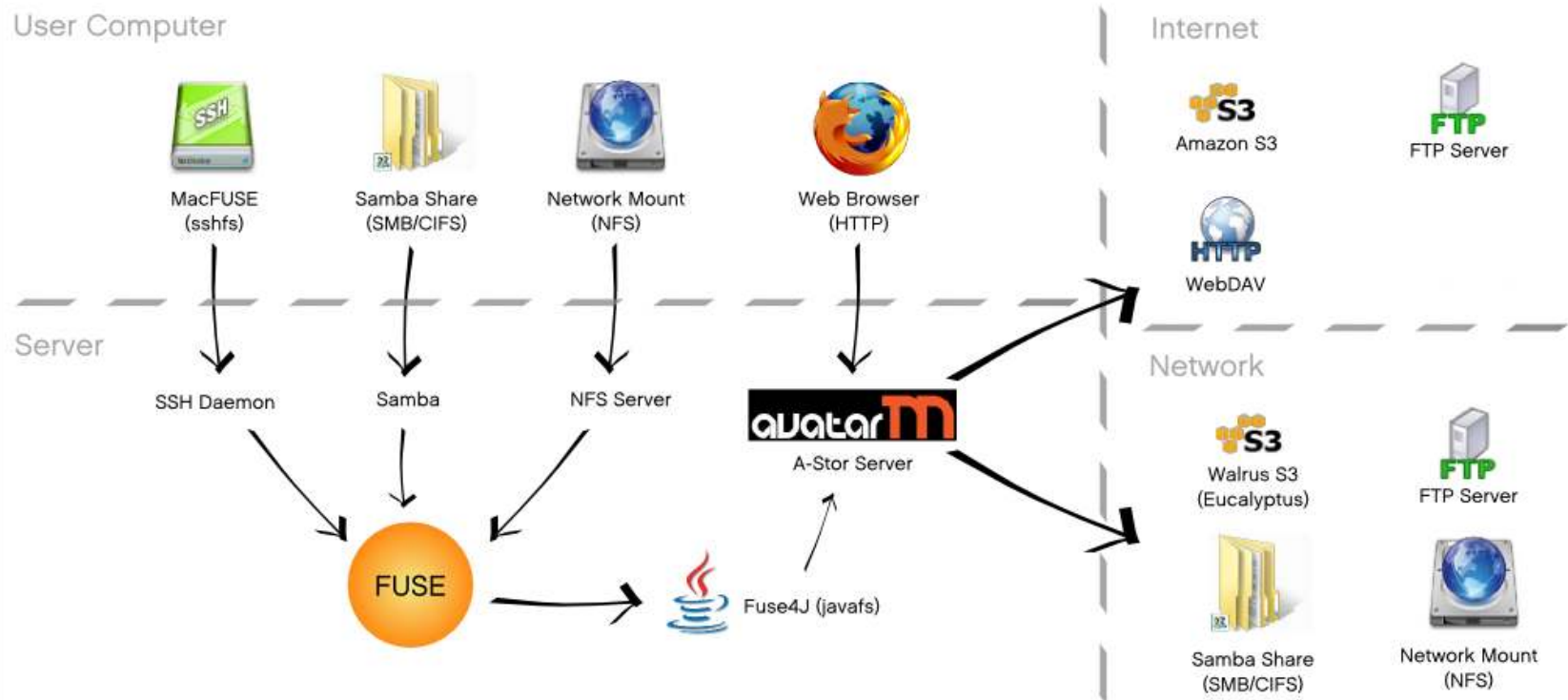


IBC2009

Combining storage



Connecting with A-Stor





IBC2009

Easy to add storage

The screenshot shows a web browser window displaying the AVATAR-m storage management dashboard. The dashboard has a navigation bar with 'Dashboard | Browse' and the AVATAR-m logo. Below the navigation bar, there are tabs for 'Summary', 'Manage Storage', 'Default Rules', and 'Configuration'. The 'Manage Storage' tab is active, showing a 'Storage Locations' panel on the left and a 'Managing Storage' panel on the right. The 'Storage Locations' panel lists two storage locations: 'Disk 2' (Access: 37.5 MB/s, Ingest: 8.6 MB/s, Reliability: +2, 8 KB used, 1.0 GB free) and 'Temp Storage' (Access: 6.5 MB/s, Ingest: 12.1 MB/s, Reliability: N/A, 5.4 GB used, 4.0 GB free). The 'Managing Storage' panel contains the text: 'The AVATAR-m storage to view their properties. Click on a storage location to manage the storage locations defined within the system and'. A modal dialog box titled 'Add New Storage' is open in the foreground, showing a list of storage options: 'Local', 'Windows Network Share (CIFS)', 'FTP', 'WebDAV Server', and 'S3 Bucket'. The 'Local' option is selected, and its configuration is shown: Name: RAID Array, Limit: 1TB, Path: /mnt/data. The dialog box has 'Cancel' and 'Add' buttons at the bottom.



IBC2009

East to access content

The screenshot shows a Mozilla Firefox browser window displaying a file index page. The address bar shows the URL `http://localhost:8182/browse/`. The page title is "Index of /". The page content includes a table of files and folders, and a "Create New Folder" dialog box is open in the foreground.

Index of /

Select a file to view its details or click its link to preview or open the file.

Name	Size	Type	Last Modified	Actions
BBC		Folder	last week	/ x
Rockstar Games		Folder	last week	/ x
20081130-Orda_Curta.m4v	36.72 MB	MPEG-4 Video	4 days ago	/ x
20080215_2331_Europa.ts	486.2 MB	TS File	3 days ago	/ x
lrmn.m4v	508.44 MB	MPEG-4 Video	5 hours ago	/ x
A4-lea442.pdf			last week	/ x
CFTA_3_Scotland_200p_mpeg2.mpg			8 hours ago	/ x
ed_300-mpeg4_rps.mp4			4 hours ago	/ x
install.txt			3 hours ago	/ x
OpenSUSE 11 (DVD).vd			last week	/ x
paprika-movie-trailer.avi			3 hours ago	/ x
protocolus-effects-shrinkvcs.js	102 KB	File	last week	/ x
test1.mkv	1.17 GB	MKV File	last week	/ x
test2.mkv	2.36 GB	MKV File	last week	/ x
test_2.txt	674 KB	Text Document	last week	/ x
TRANSIT.MIX	4.11 MB	MIX File	last week	/ x
VirtualBox.xml	2 KB	XML Document	last week	/ x

19 file(s), 5.02 GB total

© 2007-2008 University of Southampton IT Innovation Centre

Powered by A-Star (0.2.0408-4beta)

Done



IBC2009

Expandable and manageable

The screenshot displays the AvatarM dashboard interface. At the top, there's a navigation bar with 'Dashboard' and 'Browse' links, and the AvatarM logo on the right. Below this, there are tabs for 'Summary', 'Manage Storage', 'Define Rules', and 'Configuration'. The 'Summary' tab is active, showing system status and data transfer graphs.

Status reporting for daily

Server:	wakefield (Windows)		
State:	Online	Uptime:	4h 43m
Locations:	2	Files:	214
Folders:	12	Chunks:	5329

Storage usage: 6.4 GB / 6.8 GB free

Activity

There are currently no queued or active tasks.

Data Transfer watching the data go by

Ingest

Access

Cold transient files

- test_2.txt (accessed last week)
- A4 leaflet2.pdf (accessed last week)
- vmware-1.log (accessed last week)

Hot cache of attention

- 8mm.m4v (accessed 8 just now)
- ed_320-mpeg4_psp.mp4 (accessed just now)
- test2.mkv (accessed 2h 9 1m ago)

Old variable and wild

- test_2.txt (created last week)
- A4 leaflet2.pdf (created last week)
- vmware-1.log (created last week)

New well behind the wall

- install.txt (created 3h 44m ago)
- CFTA_2_Scotland_720p_mpeg2.m4v (created 4h 57m ago)
- paprika-movie-trailer.avi (created 4h 59m ago)

Big topping the scales

- test2.mkv (file size is 2.3 GB)
- test1.mkv (file size is 1.2 GB)
- 8mm.m4v (file size is 900.4 MB)

Small quality in appearance

- replay1.rsp (file size is 24 bytes)
- Rockstar Games Website.url (file size is 84 bytes)
- GTA Website.url (file size is 85 bytes)

© 2007-2009 University of Southampton IT Innovation Centre. Powered by 4-Star 0.2 (6c05-606a) using three engines and huge used under license.



IBC2009

Data management rules

Dashboard - Mozilla Firefox

http://localhost:8182/dashboard/

Dashboard | Browse **AVATAR-m**

Summary Manage Storage **Define Rules** Configuration

Rule Types

- Ingest Rules**
When should new data go?
- Move Rules**
To data in the right place?
- Replication Rules**
How many copies should be kept?
- File Splitting Rules**
How should files be split?

Defining Rules

The AVATAR-m storage system tries to store data in the best place possible. Normally this is done by optimising access rate and reliability, so that you'll always be able to get your data as quickly as possible.

However this isn't always the best approach and so trade-offs have to be made between ease of access, safety and cost. By defining rules you can control how data is stored.

Click a rule type on the left to manage rules of that type.

© 2007-2009 University of Southampton IT Innovation Centre. Powered by A-Star 0.2 (beta-alpha) Icons from oxygen and fugue used under license.

Done



IBC2009

Approaches to data integrity

- Multiple copies of whole files
 - Complete copy of each file in 3 locations
 - Expensive!
 - Storage costs, repair costs, access costs
 - But, if all else fails, you still have the files
- Erasure-correcting codes
 - RAID, RAIN etc.
 - Re-encode and distribute the data
 - Availability + safety by using many nodes
 - But, if you loose the encoder/decoder, you can lose everything



IBC2009

Factors to consider

- Assets are not all equal
 - Cost, safety, accessibility, longevity
 - Collections, genres, individual items
 - Rushes, edits, broadcast masters
- Parts of assets are not all equal
 - Metadata, video, audio, captions
 - Shots, scenes, frames
- Storage systems are not all equal
 - Spinning disk, tape robots, shelves



IBC2009

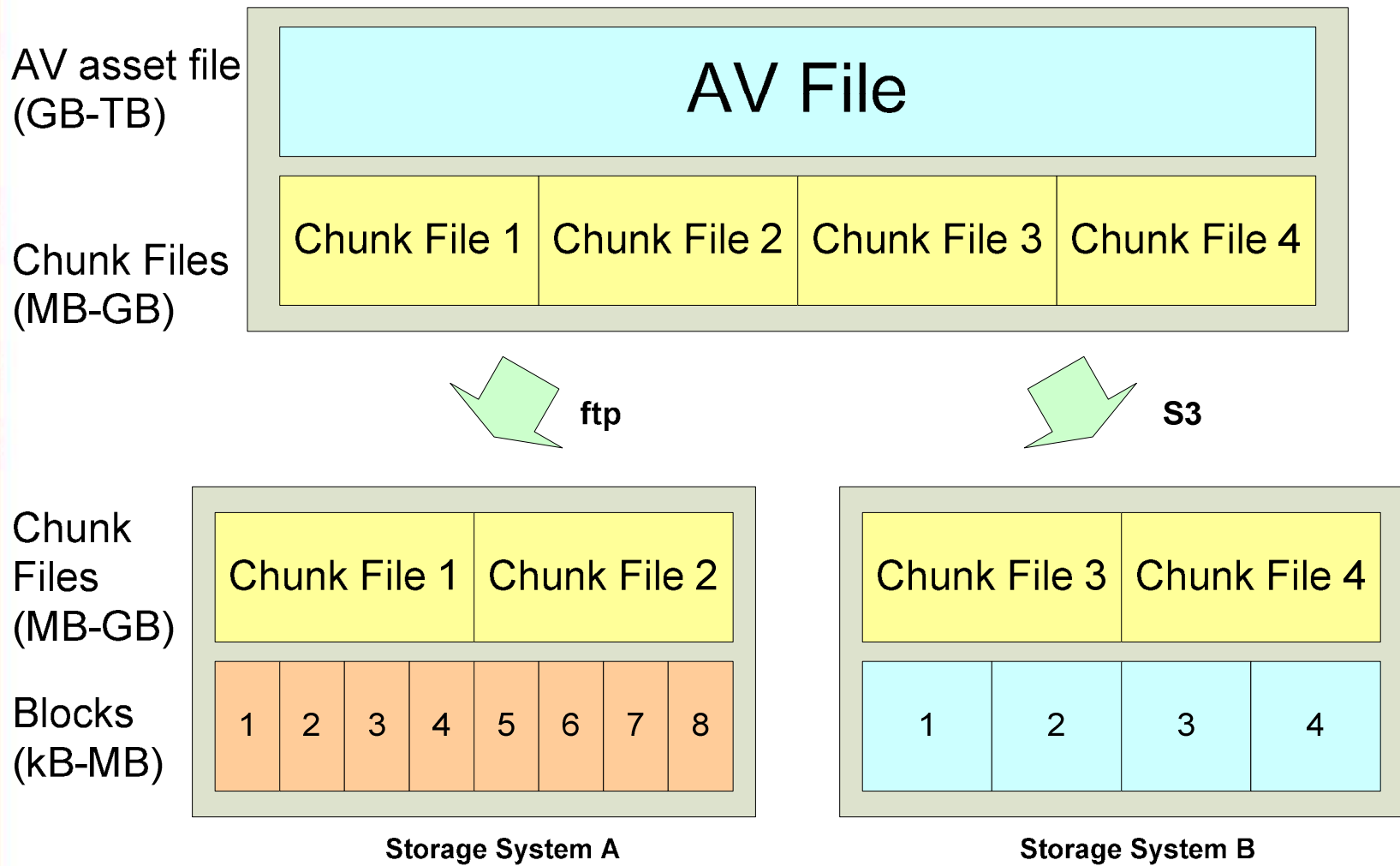
AV aware integrity management

- Split up files into pieces
 - Deconstruct MXF
 - Split video
 - Chop up large files
- Manage pieces according to policies
 - Number of copies
 - Type of storage
 - Frequency of integrity checking and repair



IBC2009

File chunking example





IBC2009

Chunking strategy

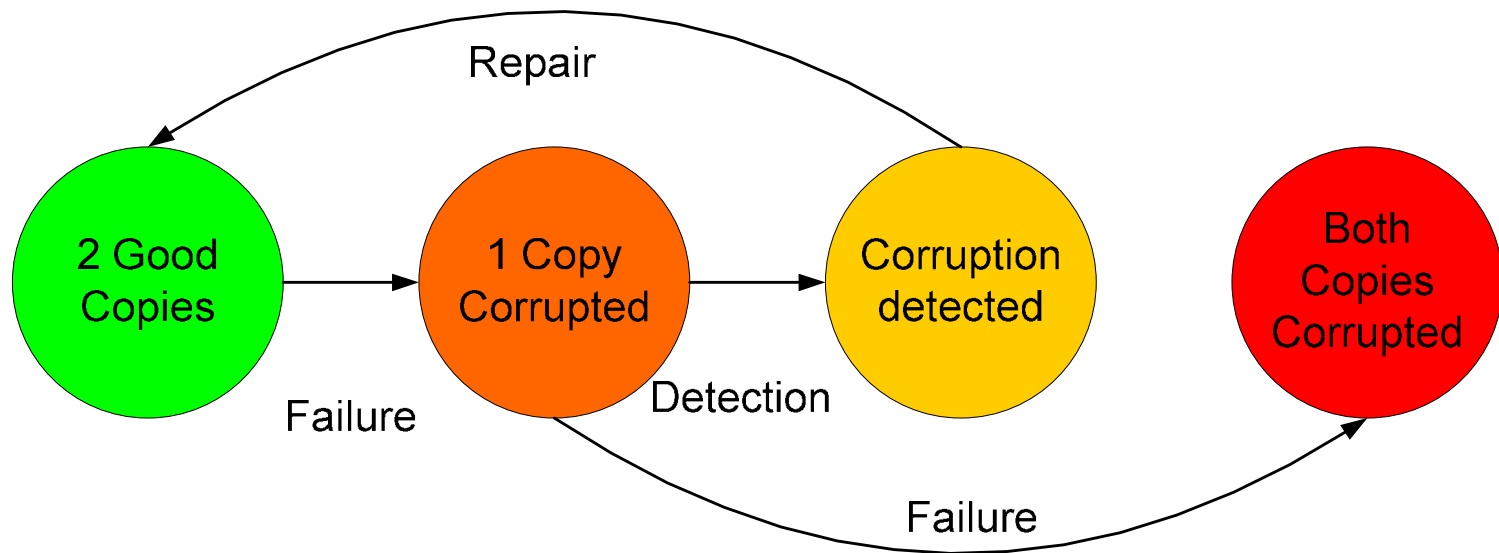
- Replicate chunks for safety
 - Match chunk sizes to AV content formats and storage corruption modes
 - Locations chosen based on need for safety, security, access, longevity
- Actively monitor for integrity
 - Nothing goes in, out or is left alone without the relevant chunks being checked for integrity



IBC2009

Optimisation

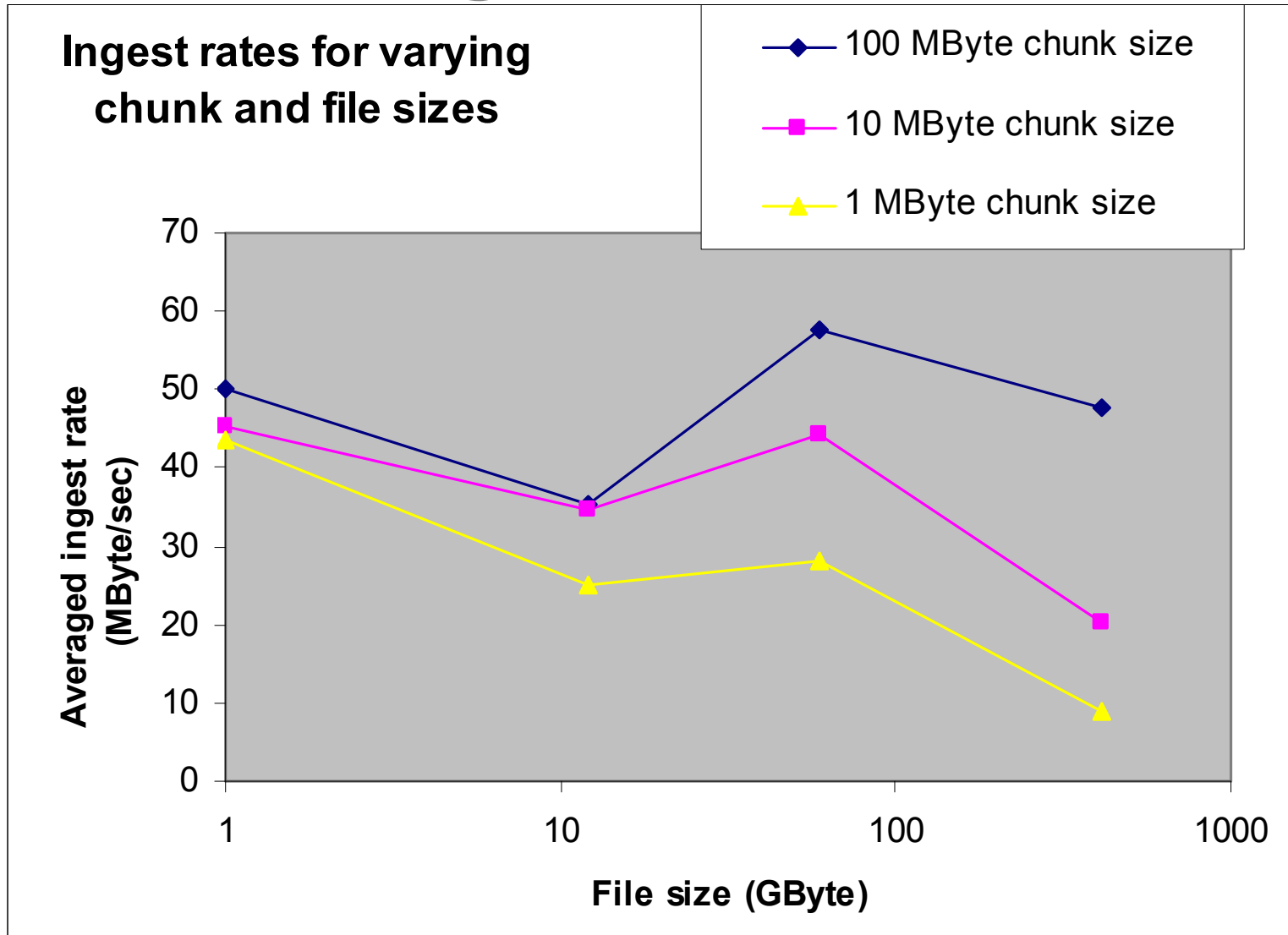
- Markov model for chunk states
- Components of the model:
 - Failure rates and costs
 - Detection rates and costs
 - Repair rates and costs
 - Storage lifetime and migration
- Predict cost v.s. probability of loss





IBC2009

Chunking adds little overhead to ingest or access

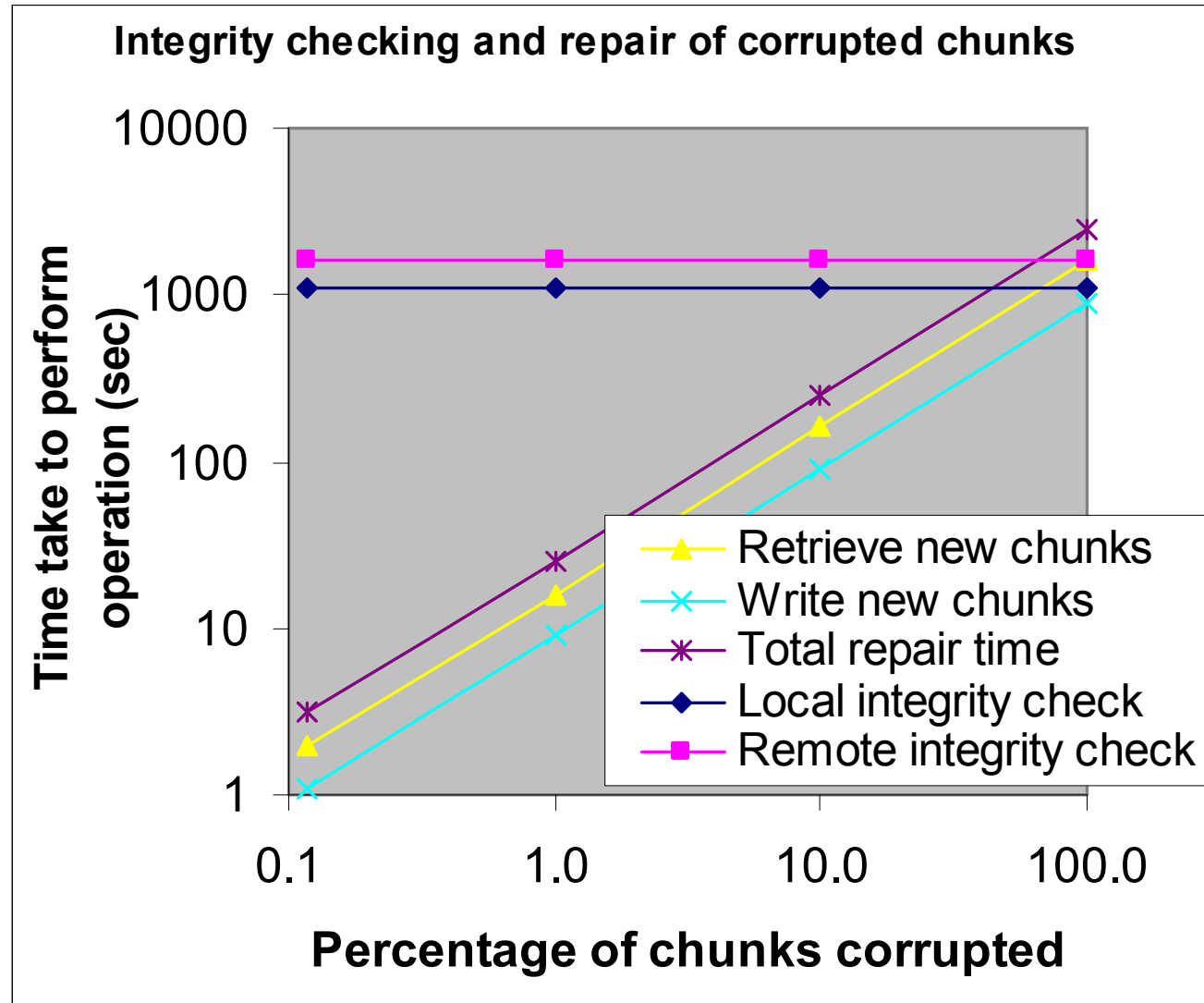


File ingest rate of AVATAR for a range of file and chunk sizes.



IBC2009

Chunking reduces repair costs



Repair of a 59GB file in 10MB chunks with varying levels of corruption



IBC2009

How does this fit with AV encoding?

- Uncompressed:
 - Good as a long-term preservation format
 - Storage can be hugely expensive
 - Robust to storage errors
- Compression:
 - Content often born this way
 - Risk of format obsolescence
 - Avoid transcoding to lossy formats
 - Lossless is possible, e.g. JPEG2000
 - More copies for the same money
 - But...



IBC2009

Corruption amplification

- Take out redundancy and you take out robustness
- Can only partially mitigate the problem
 - Intraframe compression
- New approaches needed
 - Digital file formats that behave like analogue media

(Heydegger 2008)





IBC2009

Graceful degradation

- Dirac based encoding
 - Designed to be robust to storage failures
- Retrieval is no longer an 'all or nothing' problem
- Compatible with chunking
 - Check and repair less often





IBC2009



Conclusions

- Archives: how reliable are they?
- Ongoing struggle to balance cost, safety and accessibility
- New technical approaches needed to equip archives for the world of digital files



IBC2009

Future work

The Presto PRIME logo, consisting of the word 'Presto' in a smaller font above 'PRIME' in a larger, bold font, both in white on a red square background.

Presto
PRIME

- PrestoPrime: keeping audiovisual content alive
- European collaboration supported by the EC
- 12M€, 42 months, started Jan 2009

www.prestoprime.org

- Models and environments for long-term audiovisual preservation
- Data management and processing for media preservation
- Long-term audiovisual access and use in changing contexts



IBC2009

Thank you

- mja@it-innovation.soton.ac.uk
- www.avatar-m.org.uk
- Exhibition Stand 7.A08e



AVATAR-m Overview