

## WP2 – Digital Preservation Strategies

Richard Wright, PrestoPRIME WP2  
BBC Research & Development  
UK

### PrestoPRIME Summary

KEEPING  
AUDIOVISUAL  
CONTENTS  
ALIVE

- digital technology – the solution to analogue preservation and the new problem
- digital library technology – what it can and can't do (without PrestoPRIME)
- digital preservation technology – what it means for audiovisual content
- what PrestoPRIME should do

## How we got here: Analogue Digitisation Strategy

Audio: IASA TC-04; Film: EDCine; Video:

1. Save the original
2. Digitise @ SDI = 4:2:2 = 200 megabits/sec
3. Save exactly as digitised = uncompressed
4. Use an open source file format (MXF or ?)

3b – or lossless = lossless JPEG2000

3c – or ‘mezzanine’ = high-quality lossy, on the way to lossless (but only for low quality originals)

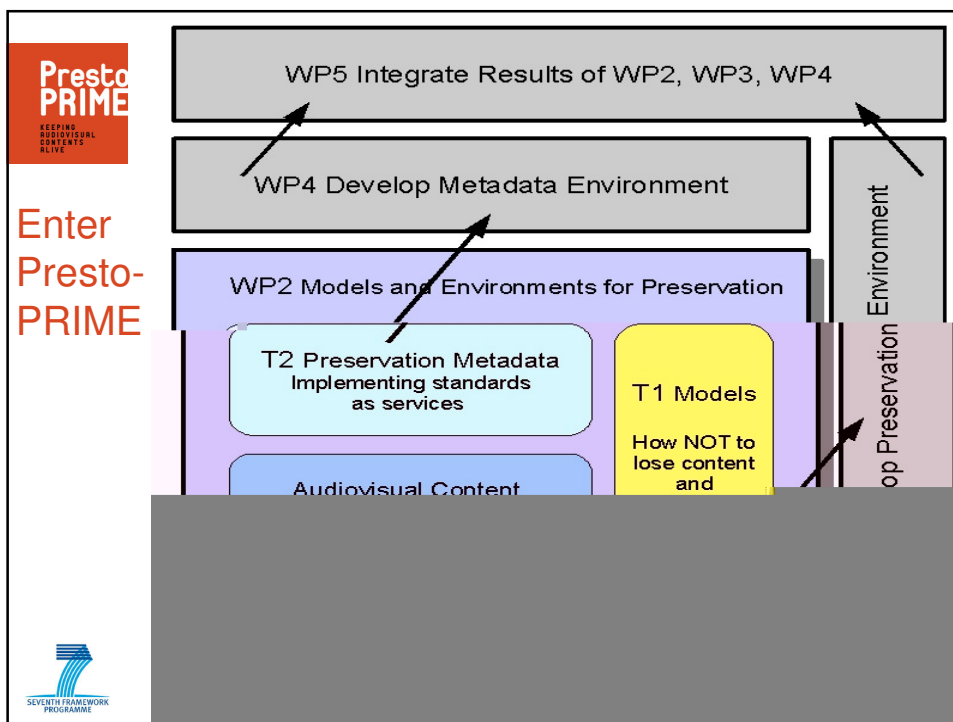
(digibeta is a ‘mezzanine’ format)

## What's needed now

- A *digital preservation* strategy
- Recognition of the problems inherent in digital technology
- Recognition of the work already done in the digital library / digital preservation communities
  - which mainly applies to text and still images
- Applying what's already known to audio and video files
- and filling in the gaps

## Problems with digital library and preservation technology in broadcasting

- Because MXF is the primary professional broadcast wrapper format
  - In Europe, at LOC, in US Public Broadcasting and for digital cinema
- Because MXF isn't supported by any digital library tools: JHOVE, PRONOM, metadata extractors
- Because OAIS is little-known in broadcasting
- Because broadcasters use MAM, not digital archives much less digital repositories
- Because AIP etc has few broadcast exemplars
- Hence: a shotgun marriage of OAIS and MXF

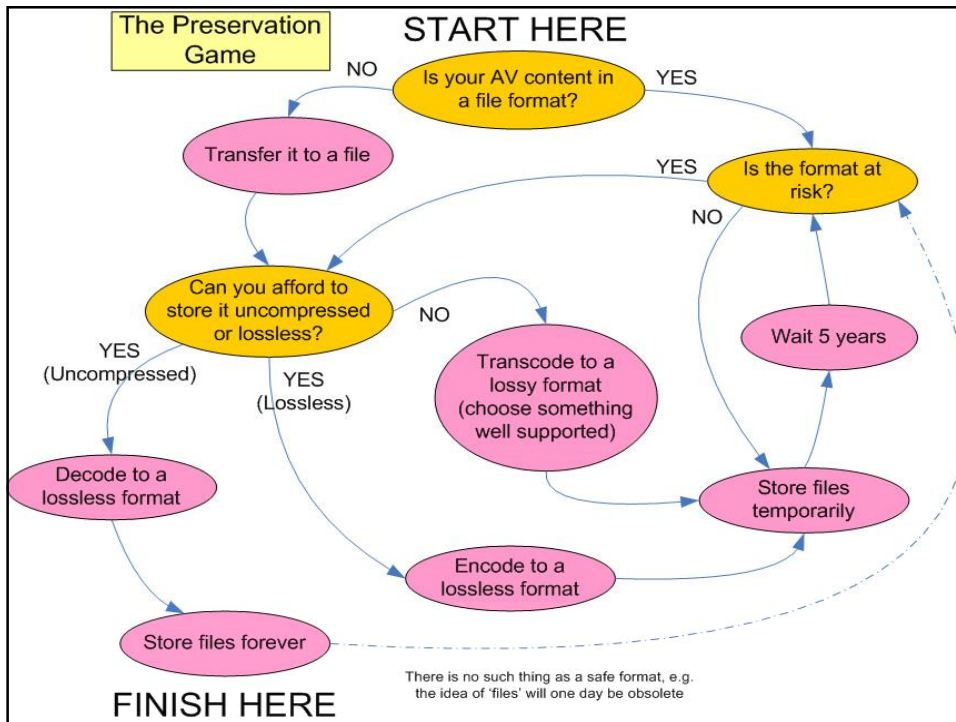


## A Digital Preservation Strategy

- The cornerstone is uncompressed (unreduced) data
- Why not go from one high-quality compressed format to the next, forever?
  - Because there will not, in general, be a transcoder
  - So the material c1 has to first be decompressed to u1, then encoded to c2
  - At the next cycle, c2 is decoded to u2
  - $u1 \neq u2 (\neq u3 \neq u4 \dots)$
  - So the “real thing” drifts away

## A format roadmap

- Avoid: moving from c1 to c2 to ...
- Requires going from c1 to u1, and keeping u1
- Does NOT require going from a poor or middle quality analogue format direct to u1
- Hence ‘temporary archiving’, and the roadmap on the PrestoSpace wiki



**Presto PRIME** KEEPING SUBVERSIVE CONTENTS ALIVE

## Format Roadmap: low quality media

Input	Output	Notes
W/S ape	W/S ape	Access Adeq a e f o q a y. M n e d a a e s (M b/s): 500 b/s. T e e a e M A t p o e n a a c c e s s i o a s, a n d t e y c o e a n d g o.
W/S ape	M b/s r e s	Archi e (e p o a y) 25 M b/s, 2 B. M g a e q o s s e s f o p e s e a o n.
W/S ape	W r e s	Archi e (e p o a y) B q o e f o a o W r o a t b e c o e s o b s o e e, g a e q o s s e s f o p e s e a o n.
'o' end' d g a f e s	S a e a s, A M s a e a s W o o s s e s	Archi e (e p o a y) B q o e f o a o W r o a t b e c o e s o b s o e e, g a e q o s s e s f o p e s e a o n.
W	W r e s	Archi e (e p o a y) 25 M b/s, 2 B. M g a e q o s s e s f o p e s e a o n.

SEVENTH FRAMEWORK PROGRAMME

Richard Wright / BBC FP7-ICT-2007-3 231161 PrestoPRIME 10

**Presto PRIME** Format Roadmap: medium quality

KEEPING SUPERVISOR CONTENTS ALIVE

Ma_c	W	Accessed ces q a y; s ab e on yf'o o ng
Ma_c	W r es	Archi e (e po a y) 25 M b/s, 2 W B. M g a e o oss ess'ro p ese a_on.
W , W AM	W r es ( ean ng .a r es' na e cod ng)	Archi e (e po a y) ans'e o co p e a 25 M b/s, es' ng n an .a r e'c one' q' e'o g na W ape. M g a e o oss ess'ro p ese a_on.



Richard Wright / BBC

FP7-ICT-2007-3 231161 PrestoPRIME

11

**Presto PRIME** Format Roadmap: high quality

KEEPING SUPERVISOR CONTENTS ALIVE

Be a S' g be a, o' e po r'o a s'	o' p essed	Archi e o' p essed s' anda d d' n on deo: 200 Mb/s. Abo 00 W B' o' n' e' o' e' ean ng 25 W Ms (o pa q' one da a ape o' a d d' e).
Be a S' g be a, o' e po r'o a s'	Mo on W 2000 (oss ess e s on)	Archi e oss ess co p ess on, a es an da a a e q' a o nd 90M b/s. Abo 0 W B' o' one o' e.
W 50	.a r es, W cod ng	Archi e (e po a y) As'ro W , b a' c' e' da a a e. Less s' cep b e o oss on' e' g' a' ons. M g a e o oss ess'ro p ese a_on.
W end' d g a r es	Sa e as s	Archi e (e po a y) B' o' e' r' o a' beco es obso e' e, g a e o oss ess'ro p ese a_on.



Richard Wright / BBC

FP7-ICT-2007-3 231161 PrestoPRIME

12

## PrestoPRIME Modelling the future

KEEPING SUPERVISORIAL SUBJECTS ALIVE

- Total cost of ownership / Lifecycle costs
- Requires overall process modelling
  - with decisions about what factors are significant
  - and decisions about parameters of the model
    - How many, how inter-related
  - and decisions about the values to assign to the parameters
  - plus how parameters (or the processes themselves) evolve over time (disruptive change)
  - eventually accumulated uncertainty becomes the dominant factor (the more you do, the worse it gets)



Richard Wright / BBC

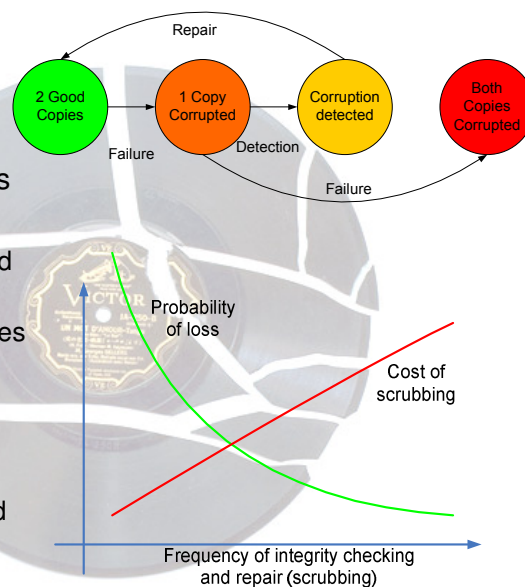
FP7-ICT-2007-3 231161 PrestoPRIME

13

## PrestoPRIME

KEEPING SUPERVISORIAL SUBJECTS ALIVE

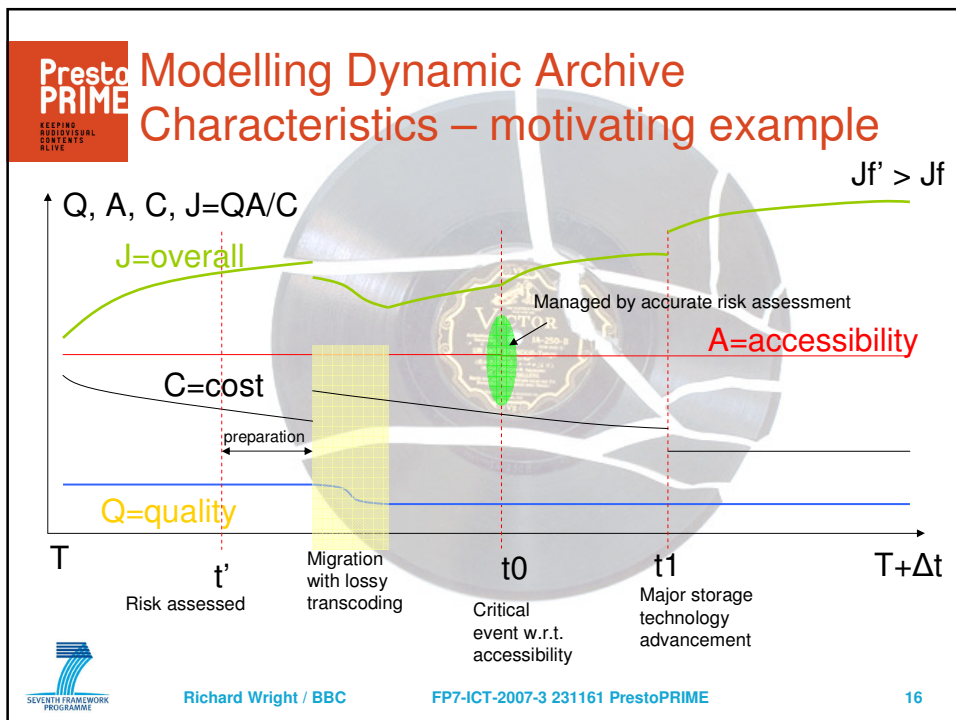
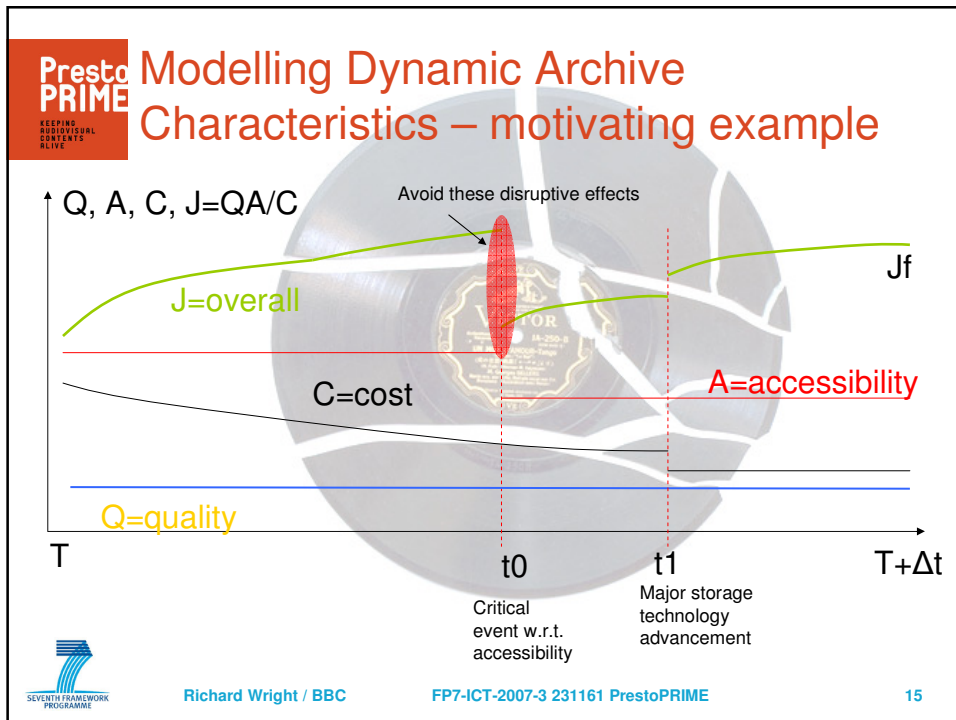
- Predict cost of risk of loss
- Factors:
  - Storage types, costs and failure rates
  - Corruption detection rates and costs
  - Repair rates and costs
  - Storage lifetime and migration
  - Use of compression and sensitivity to corruption
  - Number of copies



Richard Wright / BBC

FP7-ICT-2007-3 231161 PrestoPRIME

14





## Conclusions of WP2 on Encoding / Migration

- It's too bad that the best advice we can come up with is *continuous migration*
- -- except for emulation; do we trust it? Is it cost and time effective? Where is it effective? (see WP3 presentation)
- Continuous migration has the benefit of allowing all forms of obsolescence to be dealt with, with no long gaps, so it ensures *maintenance*
- Continuous migration gets cheaper and cheaper – and we can anticipate 'built-in maintenance'
- There should be very few uncompressed formats to support (but lots of compressed ones)

## WP2 Deliverables

<b>D2.1.1</b> Audiovisual preservation strategies, data models and value-chains	Review Compare Business models	BBC, RAI (2) B&G, ULiv (3) INA, ORF (1) ITInnov (6)
<b>D2.2.1</b> Review of semantic process modelling and workflow language	Evaluate Compare	ITInnov (9) ExLib (2)
<b>D2.3.1</b> Service-oriented models for audiovisual content storage	Define use of storage models	INA (1) Eurix, ORF (2) B&G, ULiv (3) BBC, RAI (4) ITI (12)

## D2.1.1 Audiovisual preservation strategies ...

1. **Current and emerging value-chains and business-models for audiovisual preservation**
  - Maintaining assets
  - New means of access
  - Public value
  - Commercial value
2. **A comparison of preservation strategies ...**
  - Dimensions of preservation
  - Total cost of ownership    Basic aspirations
  - Migration    Emulation    Multivalent
  - Recommendations

## D2.1.1 Audiovisual preservation strategies (cont.)

3. **preservation metadata models**
  - PREMIS
  - ExLibris (NZDL, Rosetta)
  - PrestoSpace data model
  - Planets project data model
  - What audiovisual preservation projects mean by “preservation metadata”
  - Information security in preservation environments
  - Audiovisual requirements (cf significant properties)
  - Current support for these requirements
  - Needed extensions

## D2.2.1 Preservation Process Modelling

- Processes relevant to digital preservation
  - digitisation
  - migration and emulation
  - life-cycle
- Process models relevant to digital preservation
  - OAIS
  - Project SHAMAN
  - common sense? traditional library processes?
- Process modelling formalisms
  - BPEL, XPD, UML, WSML ...

## D2.3.1 Storage Strategies

- Buy technology
- Rent technology
- Hire a service – *storage as a service*
- Requires knowledge of current costs for various options
  - plus some idea of how technology and costs will evolve
  - plus risk assessment

## What we want to know – about your files

- what file formats are in use or planned
- what standards are followed
- what systems are in use for producing files and exchanging files
- file road maps
- file, system and storage migrations
- general file problems

## What we want to know – details (1)

- what file formats they use
  - with a pause to distinguish between wrappers and encodings
  - and then get information from them on both wrappers and encodings
- what standards they follow
  - for production of files
  - for quality checking of files
  - for putting files into 'a safe place' (eg trusted digital repository)
  - probably with another pause to say more about repositories
- what systems they use for producing files and exchanging files
  - asset management systems
  - edit systems
  - storage systems
  - overall architecture eg web service or what

## What we want to know – details (2)

- file road maps
  - have they migrated from one file type to another?
  - have they made a map for how they would migrate?
  - do they anticipate problems with any of their existing file types? (example BBC problem with MPEG-2 browse)
- file, system and storage migrations
  - anyone moved from one robot to another?
  - anyone changed encoder eg IMX to ??
  - anyone changed wrapper eg RealMedia or WindowsMedia to a non-proprietary format
- general file problems
  - have they lost anything?
  - have they had quality issues (file contents not agreeing with metadata; file quality not as expected; problems in opening/reading/rendering files)?
  - do they anticipate problems