

**BRITISH  
LIBRARY**

# **MODELLING THE DIGITAL PRESERVATION COSTS**

**Paul Wheatley**

**Digital Preservation Manager**

**British Library**

- Overview of the model:
  - Aims
  - Development process
  - Model
- Results
- Evaluation
- Conclusions

## Scope

- Acquisition
- Ingest
- Metadata
- Storage
- Access
- **Preservation**

## Background and aims

- Previous work (see Final Report):
  - National Archief, Digital Bewaring – full costing/audit approach
  - Oltmans, Kol – lifecycle and strategies
- Key aims:
  - Make the first major step in defining and estimating the lifecycle cost of digital preservation activities.
  - Propose a model for comment by the wider preservation community
  - Enable the LIFE Case Studies to be compared and contrasted by providing some cost estimates for “P” in the Lifecycle Model.
  - Attempt to identify the scale of preservation costs. Are they dramatically high as suggested previously by many in the preservation community or are they more achievable as suggested recently (see Rusbridge, C, “Excuse Me... Some Digital Preservation Fallacies?”)?

## Development process

- Key cost factors, experimentation, iterative development and refinement
- Based on evidence or indications of trends where possible
- Editable inputs where key estimation or assumptions made
- Cost component review
- Application of draft model, refinement of inputs
- Team review, refinement of model weaknesses

## The Generic LIFE Preservation Model

$$\text{Preservation} = t * \text{TEW} + (t / \text{ULE} + \text{PON}) * (\text{CRS} + \text{UME} + \text{PPA} + \text{QAA})$$

### Expansion of calculated components:

- ULE – Unaided Life Expectancy of a Format =  $\text{BLE} + 0.1 * t$
- CRS – Cost of new rendering solution =  $(1 - \text{PTA}) * \text{TDC} * \text{FCX} + \text{PTA} * \text{COA}$
- PPA – Performing preservation action =  $\text{PON} * (\text{SCM} + n * \text{HVM})$
- QAA – Quality Assurance =  $n * \text{BCT} * \text{FCX}$
- PTA – Proportion of Tool Availability =  $\text{STA}(1-t/20) + \text{ETA}(t/20)$

### Expansion of scaling components:

- PON – Proportion of normalisation = 0.4
- FCX - Format complexity (e.g. JPEG = 0.2, WMF = 0.4, PDF = 0.6, Word = 0.8)

### Expansion of cost component inputs:

- HVM – High volume migration cost per object = £0.05
- BCT – Base cost of testing a preservation action per object = £0.17
- UME – Update Metadata = 2 metadata officer weeks @ £30k annual salary = £1250
- TDC – Tool development cost = 24 programmer months @ £30k annual salary - £60000
- COA – Cost of available tool = £1500
- TEW - Technology Watch = 1 metadata officer week @ £30k annual salary = £625
- BLE - Base life expectancy = 8 (years)
- STA – Starting tool availability = 0.5
- ETA – Ending tool availability = 0.9
- SCM – Setup cost of migration = £340

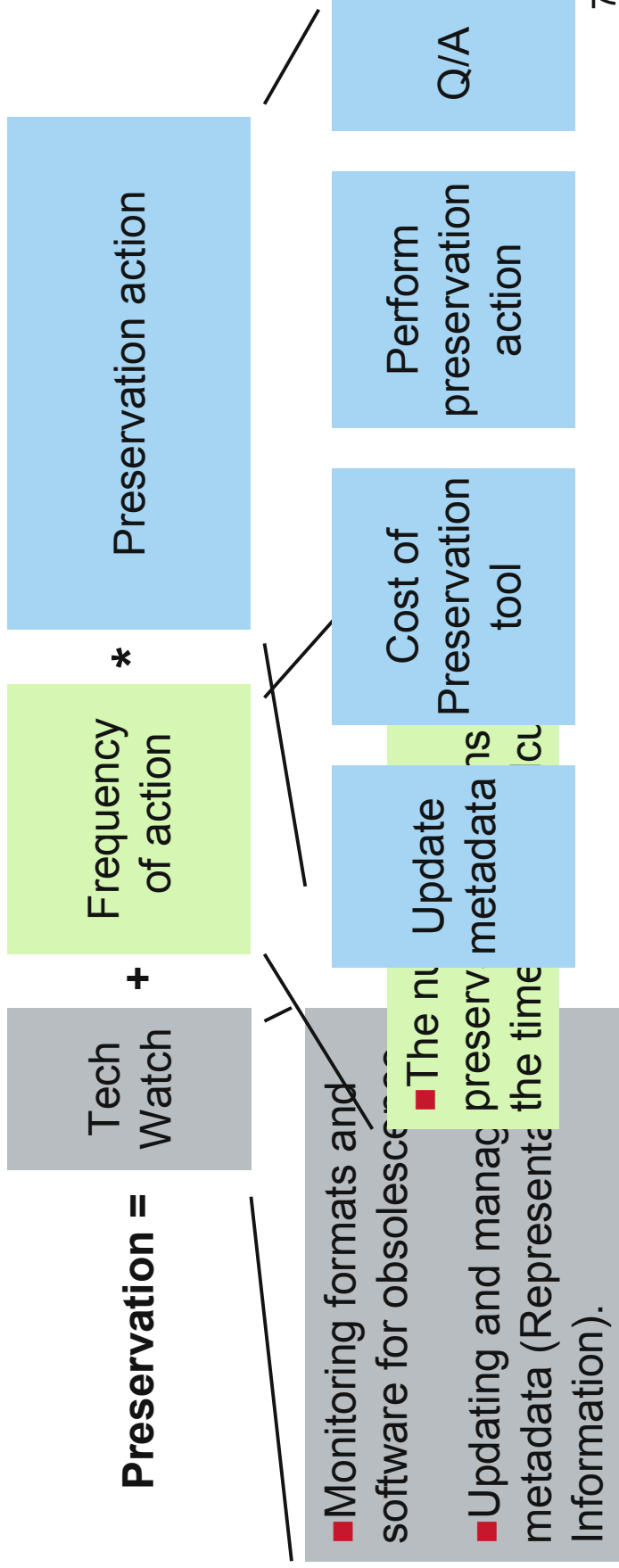
# The Generic LIFE Preservation Model : key elements explained

Preservation cost of  $n$  objects of a particular format for the period 0 to  $t$ .

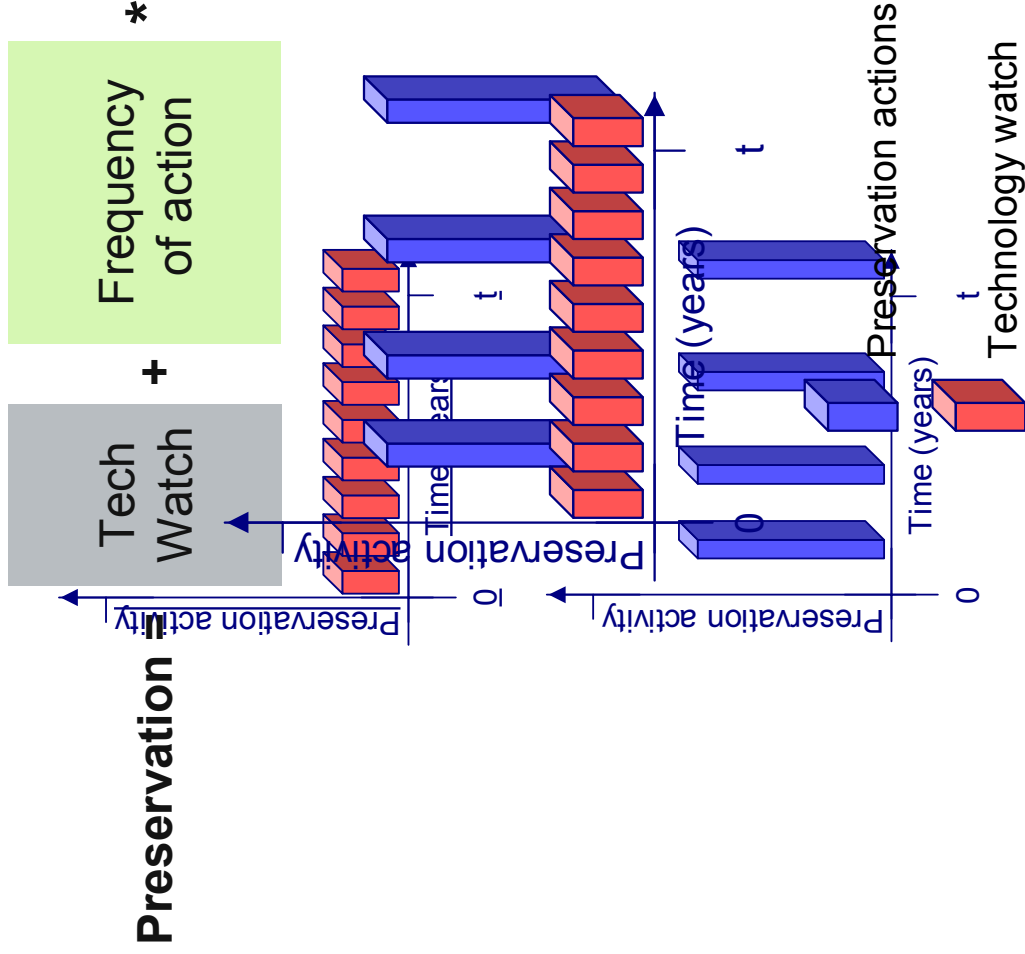
Eg. 20000 objects of the GIF format for a period of 10 years.

$$\text{Preservation} = t * \text{TEW} + (t / \text{ULE} + \text{PON}) * (\text{CRS} + \text{UME} + \text{PPA} + \text{QAA})$$

Preservation =



# The occurrence of costs (1<sup>st</sup> detailed sample of the model)



Preservation action

Example : FCLA Action Plans

<http://www.fcla.edu/digitalArchive/>

Series of small technology watch events and spikes of preservation activity at increasing intervals

Base life expectancy = 8 years  
Increases by a year every decade



Complexity of file formats  
(2<sup>nd</sup> detailed sample of the model)

Preservation =

Tech Watch + Frequency of action \*

Preservation action

Format Complexity

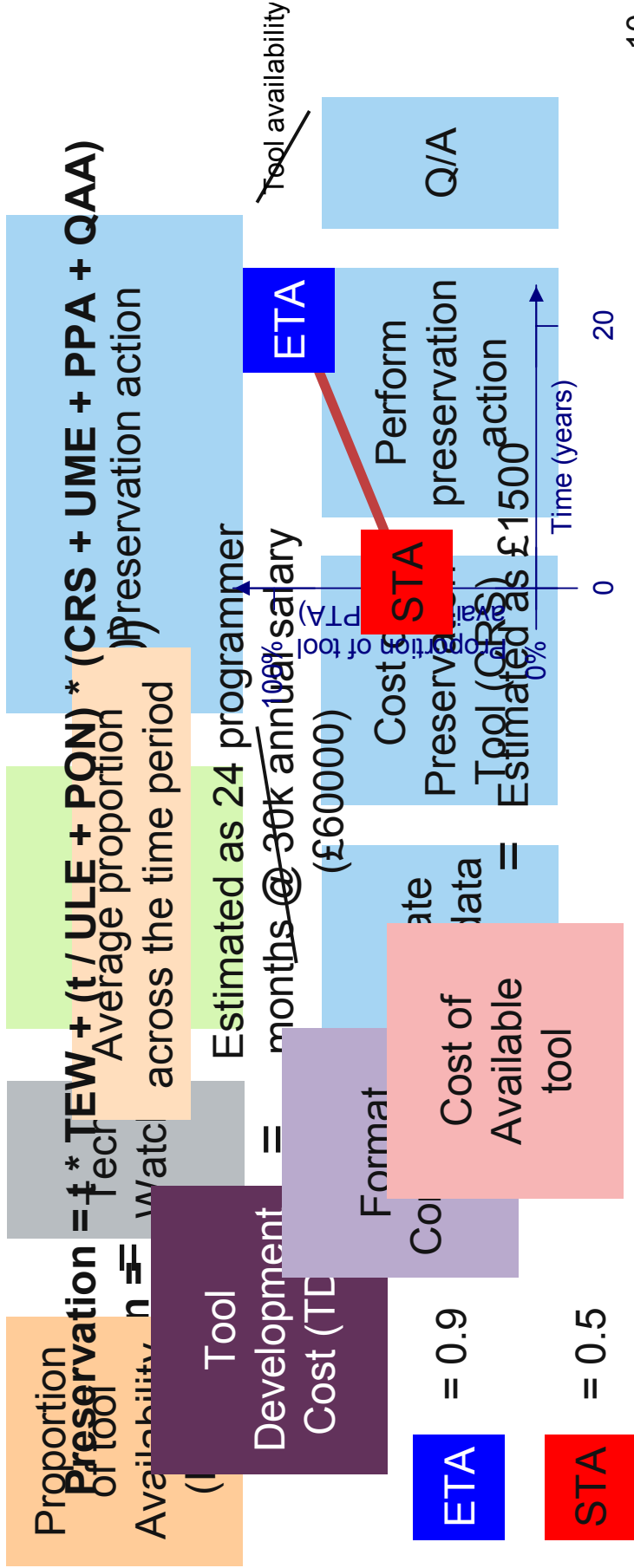
=

Category	Complexity	Examples
Simple	0.1	ASCII, Unicode
Bitmap	0.2	JPEG, GIF
Media	0.3	MP3, XML, HTML, Perform
Proprietary metadata	0.4	Preservation, DRA, Preservation
Open multimedia	0.6	MPEG3, WAV
Standardised	0.8	Word, PDF
Complex	1	Oracle database dump

Q/A

# Preservation tool cost (3rd detailed sample of the model)

$$= (1 - \frac{PTA}{A}) \text{ Cost of developing a new tool} + \text{ Cost of acquiring an existing tool}$$



## Estimated costs using the model

File Format	Format Complexity	Number of objects	Frequency of pres action
GIF	0.2	225079	1.51

Estimated preservation costs for GIF files in the Web Archiving Case Study

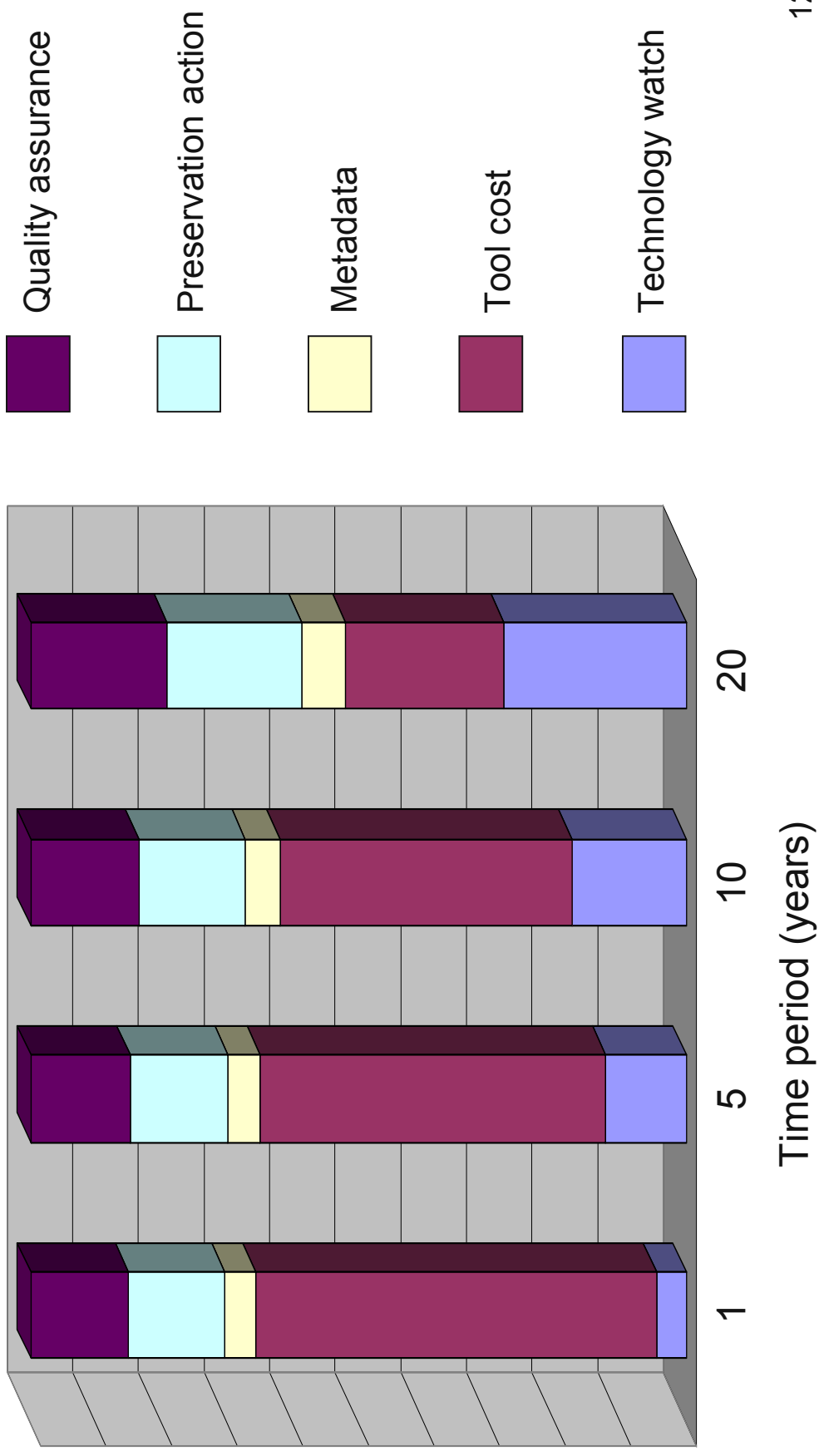
File Format	Technology watch	Preservation tool cost	Metadata	Preservation action	Quality assurance	Total cost (over 10 years)
GIF	£6,250	£7,027	£1,889	£7,008	£11,564	<b>£33,738</b>

Case study name	Sub category	Year1	Year 10	Percentage of total lifecycle cost
VDEP	e-monographs	£0.89	£1.45	4%
VDEP	e-serials	£10	£27	2%
Web archiving		£425	£8509	62%

Comparison of average object preservation costs across the Case Studies

# Model outputs: WA Case Study, percentage breakdown

Breakdown of complete preservation costs over time in the WA Case Study



## Self evaluation of the model

Evaluation against key aims:

- Make the first major step in defining and estimating the lifecycle cost of digital preservation activities.
- Propose a model for comment by the wider preservation community
- Enable the LIFE Case Studies to be compared and contrasted by providing some cost estimates for “P” in the Lifecycle Model.
- Attempt to identify the scale of preservation costs. Are they dramatically high as suggested previously by many in the preservation community or are they more achievable as suggested recently (see Rusbridge, C, “Excuse Me... Some Digital Preservation Fallacies?”)?

## **Further work and refinement**

- Refinement based on real cost data, removal of assumptions
- Level of detail
- Format complexity
- Re-ingest
- More detailed discussion in the Final Report...

## Summary and conclusions

- Estimating the cost is not easy but appears to be possible!
- Provides a useful perspective on performing preservation
- Focuses on achieving cost effective preservation

Two appeals to the audience:

- Please cost, record and publish your preservation work
- Provide comment on the preservation model:

Questions, comments, evaluation:

[paul.wheatley@bl.uk](mailto:paul.wheatley@bl.uk)