



Digital**Preservation**Coalition

# Digital Preservation Tools

## 3 (4) things that might help

Ingest: Pronom + Droid

Costs: LIFE

Certification: Data Seal of Approval

William Kilbride

[william@dpconline.org](mailto:william@dpconline.org)



Digital **Preservation** Coalition

## **Pronom + Droid: eh?**

Pronom: a register of file formats and their behaviours

DROID: a tool that uses Pronom to analyse the files on a system and report any issues that arise

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>



# file formats and their characteristics



### The technical registry PRONOM

- Welcome
- About
- Add an entry
- Search
- Help
- Information resources

#### Welcome to PRONOM

[PRONOM changes and DROID signature file release notes.](#)

Find out more about our plans to make PRONOM's data available

The online registry of technical information. PRONOM is a repository of software products and other technical components required to preserve digital information of historical or business value. Find out about the future of PRONOM

Search PRONOM ▶

- Simple search
- File format
- PRONOM Unique Identifier
- Software
- Vendor
- Lifecycles
- Migration Pathways

#### Details for: Tagged Image File Format 3

Save as... XML | CSV | Print

Go to: [Summary](#) | [Documentation](#) > | [Signatures](#) > | [Compression](#) > | [Character encoding](#) > | [Rights](#) > | [Reference files](#) > | [Properties](#) >

#### Summary

<b>Name</b>	Tagged Image File Format
<b>Version</b>	3
<b>Other names</b>	TIFF (3)
<b>Identifiers</b>	PUID: fmt/7 MIME: image/tiff Apple Uniform Type Identifier: public.tiff

#### Family

<b>Classification</b>	Image (Raster)
<b>Disclosure</b>	Full

#### Description

The Tagged Image File Format (TIFF) is a raster image format originally developed by the Aldus Corporation, primarily for use in scanning and desk-top publishing. When Adobe Systems Incorporated purchased Aldus in 1994, they acquired the rights to the TIFF format and have maintained it since then. TIFF files comprise three sections: an Image File Header (IFH), an Image File Directory (IFD), and the image data. TIFF files can contain multiple images (multi-page TIFF), and each image has a separate IFD. The IFH always appears at the beginning of the file, and is immediately followed by a pointer to the first IFD. The IFD contains metadata which describes the associated image, stored as a series of tags. The IFD also contains a pointer to the actual image data. TIFF 3.0 supports colour depths from 1 bit to 24 bit (e.g. monochrome to true colour), and a range of compression types (RLE and CCITT Group 3 and Group 4).



Digital **Preservation** Coalition



*DROID: Search and report on files from an entire network*

*Identify files by extension*

*Identify files by contents*

*Report errors and concerns*

*Automated processes and Scalability ...*

## Pronom + Droid: Ingest processes

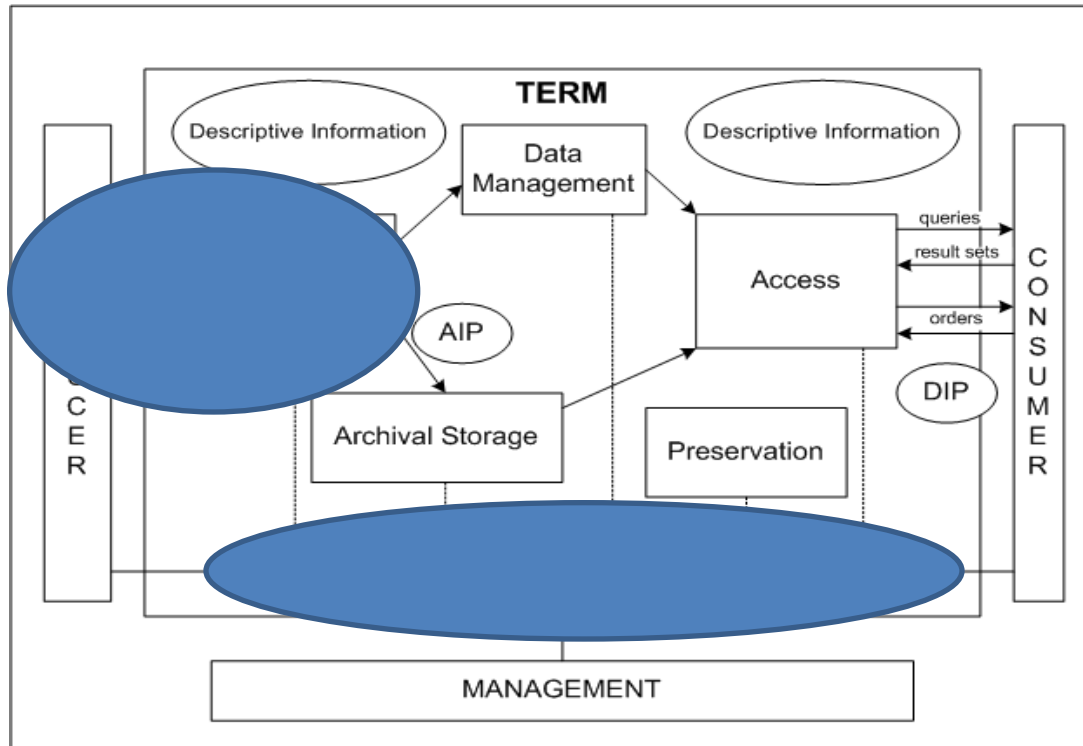


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, *Producer-Archive Interface Methodology Abstract Standard*. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

Need to know  
what you've got.

Is it what it says  
it is?

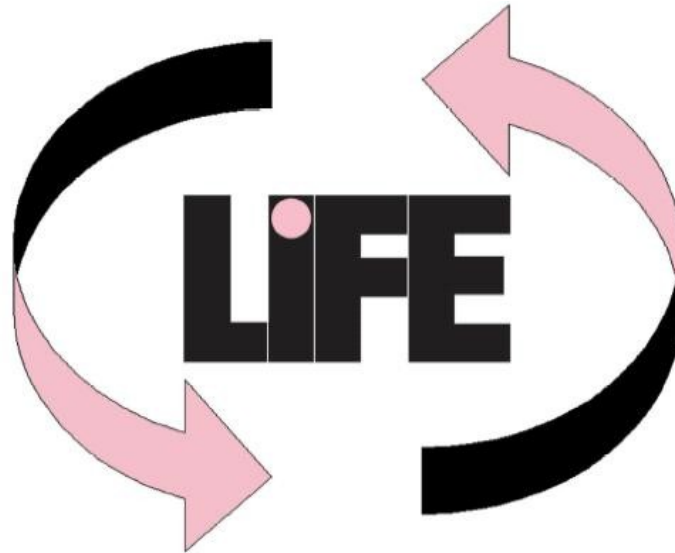
Planning and  
management

# LIFE Project

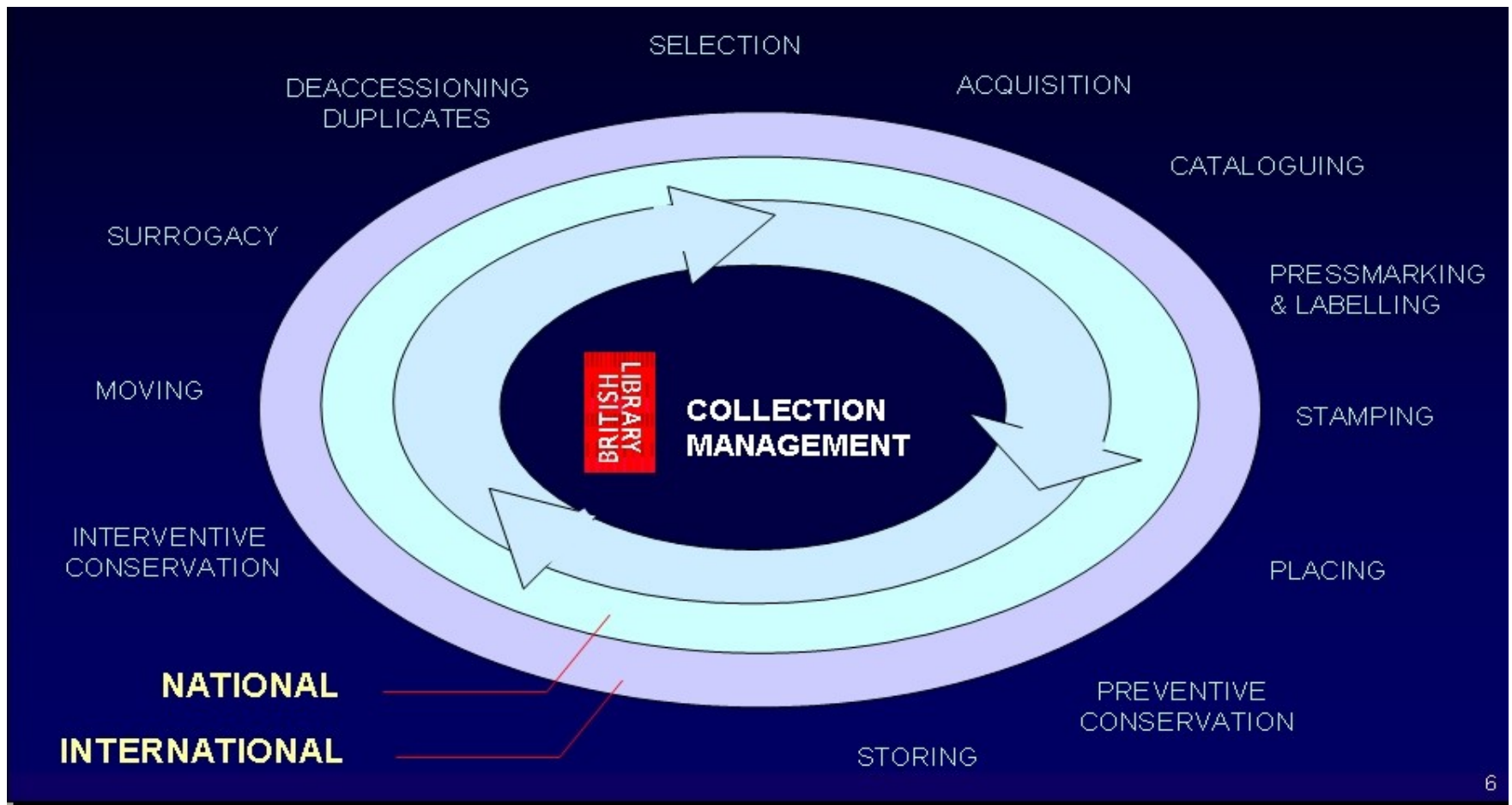
Partners, UCL/BL

Funded by JISC

<http://www.life.ac.uk/>



# LIFE: BL's lifecycle collection management cost model



# LIFE: lifecycle Costing formula

## A regular monograph

$$K(t) = s+a+c+cs+hl+p(t)+cs(t)+r(t)$$

s = selection cost;

a = acquisition cost (excluding the purchase price);

c = cataloguing cost;

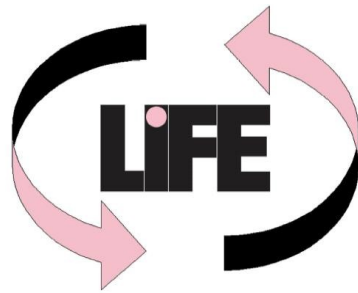
pl = initial preservation cost (such as an archival enclosure);

hl = initial handling cost (including pressmarking, labelling, and placing);

p(t) = the likely preservation cost over time (including interventive conservation); cs(t) = the collection storage cost over time;

r(t) = the likely retrieval and replacement cost over time.





<http://www.life.ac.uk/>

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

L is the complete lifecycle cost over time 0 to T. Other categories are

Aq - Acquisition,

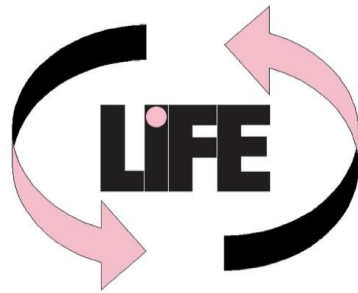
I - Ingest,

M - Metadata,

Ac - Access,

S - Storage,

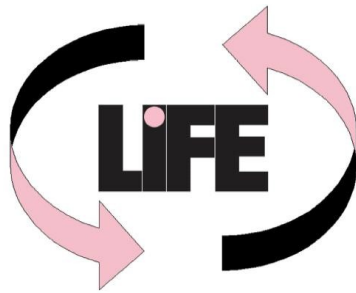
P - Preservation



<http://www.life.ac.uk/>

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

Lifecycle element	Acquisition	Ingest	Metadata	Access	Storage	Preservation
Element 1	Selection (Aq1)	QA (I1)	Characterisation (M1)	Reference linking (Ac1)	Bit-stream storage costs (S1)	Technology watch (P1)
Element 2	IPR (Aq2)	Deposit (I2)	Descriptive (M2)	User support (Ac2)		Preservation tool cost (P2)
Element 3	Licensing (Aq3)	Holdings update (I3)	Administrative (M3)	Access Mechanism (Ac3)		Preservation metadata (P3)
Element 4	Ordering and invoicing (Aq4)					Preservation action (P4)
Element 5	Obtaining					Quality assurance (P5)
Element 6	Check-in (Aq6)					



<http://www.life.ac.uk/>

Wednesday 21 October 2009



```
010101001001011001011001 00100110101
10100100101100101100100 00101011100
000101001001011001011001 00100110100
10100100101100101100100 10101011100
010101001001011001011001 00100110101
10100100101100101100100 10101011100
000101001001011001011001 00100110101
10100100101100101100101 10101011100
010101001001011001011001 00100110101
10100100101100101100100 10101011100
010101001001011001011001 00100110101
10100100101100101100100 10101011101
```

...helping to bring digital preservation to LIFE

Location: LIFE homepage » LIFE<sup>3</sup>

Menu
Home
What is LIFE?
LIFE Team
LIFE <sup>1</sup>
Documentation
Conference
LIFE <sup>2</sup>
Documentation
Conference
► LIFE <sup>3</sup>
LIFE Blog
LIFE Cycles
Glossary

## LIFE<sup>3</sup>

The third phase of LIFE commenced in August 2009, and will run for one year with funding from JISC and RIN. By producing a predictive costing tool, LIFE<sup>3</sup> will significantly improve the ability of organisations to plan and manage the preservation of digital content. The project will expand its existing Generic Preservation Model to create a comprehensive suite of models covering all life cycle stages, providing greater accuracy and assurance in estimation. The predictive costing tool will be made available towards the end of 2010, as both a web application and an Excel-based model. The project team would be delighted to hear from organisations interested in assisting with trials of the tool.

**WP1:** Development of the Predictive Models. The objective of the WP is to further refine the GPM (Generic Preservation Model) and develop new estimative models for the other stages of the lifecycle. The GPM was first developed in LIFE<sup>1</sup> in order to estimate the cost for the preservation stage of a digital object's lifecycle. The GPM is truly innovative, because very little costing information for this activity exists. Few details are available of either the breakdown of what the process might involve or of the costs of each of these elements for the large scale preservation of digital collections.

Because of lack of historical figures, a strategy of estimation has been





Digital **Preservation** Coalition

What does success look like? Certification

*Trusted Digital Repository*

*TRAC*

*DRAMBORA*

*BS 10008*

*Data Seal of Approval*





Digital **Preservation** Coalition

## *Data Seal of Approval*

*<http://datasealofapproval.org/>*

16 criteria

Self assessment against criteria

Aimed at academic research data

UK Data Archive & Archaeology Data Service

'Entry level', 'Bronze' certification

Brings certification within reach ...



Digital **Preservation** Coalition

# Digital Preservation Tools

## 3 (4) things that might help

Ingest: Pronom + Droid

Costs: LIFE

Certification: Data Seal of Approval

William Kilbride

[william@dpconline.org](mailto:william@dpconline.org)