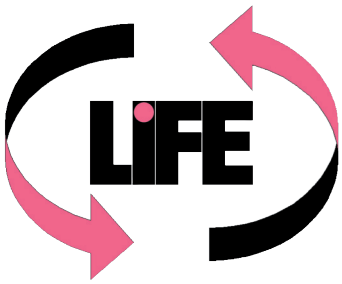



# The LIFE³ Project

Bringing digital preservation to LIFE

## Lifecycle Information for E-literature

An Introduction to the third phase of the LIFE project

A JISC and RIN funded joint venture project






Humanities Advanced Technology & Information Institute

## THE LIFE PROJECT TEAM

**Brian Aitken**
Software Developer (HATII)
**Paul Ayris**
Director of UCL Library Services (UCL)
**Brian Hole**
LIFE3 Project Manager (British Library)
**Li Lin**
LIFE3 Financial Modeler (British Library)
**Patrick McCann**
Software Developer (HATII)
**Caroline Peach**
Head of Preservation Advisory Centre (British Library)
**Paul Wheatley**
Digital Preservation Manager (British Library)

## GOVERNANCE

The Project is being governed by an international Project board. The full membership of the project board is as follows:

**Sheila Anderson**
Director, Centre for e-Research (CeRch)
**Kevin Ashley**
Director, Digital Curation Centre
**Dr Paul Ayris (Chair)**
Director, UCL Library Services
**Neil Beagrie**
Charles Beagrie Ltd
**Neil Grindley**
Programme Manager, JISC
**Brian Hole**
LIFE³ Project Manager (British Library)
**Michael Jubb**
Director, RIN
**Martin Moyle**
Digital Curation Manager, UCL Library Services
**Caroline Peach**
Head of Preservation Advisory Centre (British Library)
**Chris Rusbridge**
Chris Rusbridge Consulting Ltd
**Wouter Schallier**
Executive Director, LIBER
**Paul Wheatley,**
Digital Preservation Manager (British Library)

## ACKNOWLEDGEMENTS

The LIFE Project would like to extend its thanks and appreciation to everyone who has helped this Project attain its aims.

In particular LIFE would like to thank the following people:

Christine Adams (British Library)
William Anderson (University of Texas at Austin)
Simon Bains (University of Edinburgh)
Neil Beagrie (Charles Beagrie Ltd.)
Aly Conteh (British Library)
Helen Cooper (University of Central Lancashire)
Richard Davies (British Library)
Richard Davis (ULCC)
David Dinham (National Library of Scotland)
Isabelle Dussert-Carbone (BnF)
Marie-Elise Freon (BnF)
Louise Faudet (BnF)
Ed Fay (LSE)
Neil Fitzgerald (British Library)
Juan Garces (British Library)
Cody Green (Texas A&M University)
John Harrington (Cranfield University)
Lee Hibbard (National Library of Scotland)
Patrick Hinni (National Library of Switzerland)
Steve Hitchcock (University of Southampton)
Helen Hockx-Yu (British Library)
Ioan Isaac-Richards (National Library of Wales)
Wendy Johnston (Oracle)
Ulla Bøgvad Kejser (Danish national library)
Klaus Kjærgaard (State and University Library of Denmark)
Stephanie Meece (University of the Arts, London)
Erica McLaren (UCL)
Martin Moyle (UCL)
Anders Bo Nielsen (Danish state archive)
William Nixon (University of Glasgow)
Claire Parry (Queen Mary University, London)
Maggie Pickton (University of Northampton)
Will Prentice (British Library)
Keith Rajecki (Oracle)
Richard Ranft (British Library)
Christine Rees (University of Edinburgh)
John Rhatigan (British Library)
Robin Rice (University of Edinburgh)
Owain Roberts (National Library of Wales)
Olivier Rouchon (CINES)
Chris Rusbridge (Chris Rusbridge Consulting Ltd.)
Helen Shenton (British Library/Harvard University)
Katherine Skinner (Meta-Archive Cooperative)
Mary-Anne Spalding (University for the Creative Arts)
Neil Stewart (LSE)
Victoria Swift (International Dunhuang Project)
Alex Thirifays (Danish state archive)
Philip Thornborow (University of Northampton)

**JISC**
Thanks to Neil Grindley, Digital Preservation Programme Manager.

**RIN**
Thanks to Michael Jubb, Director.

# Executive summary

## Introduction

The first phase of LIFE (Lifecycle Information For E-Literature) made a major contribution to understanding the long-term costs of digital preservation; an essential step in helping institutions plan for the future. The LIFE work models the digital lifecycle and calculates the costs of preserving digital information for future years. Organisations can apply this process in order to understand costs and plan effectively for the preservation of their digital collections. The second phase then refined the LIFE Model adding three new exemplar Case Studies to further build upon LIFE[1].

The third phase of the LIFE Project, LIFE[3], is a collaboration between University College London (UCL) and the British Library, funded by JISC and RIN, and running from August 2009 to August 2010. LIFE[3] has produced a predictive costing tool, based on a refined version of the LIFE Model (v3), and several additional sources of case study information, and incorporating a further revision of the digital preservation costing model (GPM v1.2).

This summary aims to give an overview of the LIFE Project, outlining some of the key outputs.

There are three main areas discussed:

1 **From LIFE[1] to LIFE[3]** outlines some of the key findings from the first phase of the project as well as summarising the motivation behind this second phase.
2 **The LIFE Model** describes the current version of the model (version 3) which has been revised since the second phase.
3 **The LIFE Tool** presents the LIFE Digital Preservation Predictive Costing Tool.

## Further Information

On the inside of the back cover of this summary, there is a full listing of the project outcomes from both phases of the project. All project documentation, including Case Study results and spreadsheets with exact costings, are available from the LIFE website.

After each section in this document, there is a selection of links to further information. For example the box below contains links to the main project partners and project funder.

There is also a project blog (with RSS feed) which highlights any new project findings or documentation being made available.

| USEFUL LINKS | |
|---|---|
| Digital Preservation at The British Library | www.bl.uk/dp |
| JISC | www.jisc.ac.uk |
| RIN | www.rin.ac.uk |
| LIFE Project Website | www.life.ac.uk |
| LIFE Project Blog | www.life.ac.uk/blog |
| UCL Library Services | www.ucl.ac.uk/library |

# From LIFE[1] to LIFE[3]

What follows is a brief summary of the first two phases of the LIFE Project (LIFE[1] and LIFE[2]) and the motivation for the third phase of the project (LIFE[3]). All documentation referred to is available from the LIFE website (www.life.ac.uk).

## LIFE[1] Summary

Run from 2005 to 2006, the LIFE[1] Project made a major contribution to understanding the long-term costs of digital preservation. The project team felt that this was an essential first step in helping institutions plan for the future of digital collections.

Based on a comprehensive review of existing lifecycle models and digital preservation, the LIFE[1] Project developed a lifecycle-based methodology to calculate the costs of preserving digital information for the next 5, 10 or 100 years.

The LIFE Model broke down a digital object's lifecycle into six main lifecycle stages, identifying the costs of these elements over a specific time period, and thus providing a complete lifecycle cost.

→ A full breakdown of the lifecycle categories and elements, as well as analysis of each element is provided in the LIFE[1] Project final Report (Section 4, p.9-16)

## Generic Preservation Model

Due to the lack of work undertaken in the area of digital preservation costing before 2005, LIFE[1] also produced the Generic Preservation Model to further develop the Preservation stage of the model. This work allowed institutions to start to identify and reduce the spikes of cost, as well as the frequency of their preservation actions.

In the Generic Preservation Model, key elements of preservation activities were identified and the factors which contributed to their costs were modeled. A spreadsheet tool for calculating the costs for digital objects of varying file formats was also developed as part of the model.

→ A detailed introduction on the Generic Preservation Model (GPM) can be found in the LIFE[1] final Report (Section 8, p.90-107). The GPM spreadsheet is also available from the LIFE website.

## Case Studies and Findings

To test and evaluate the LIFE methodology, three Case Studies were chosen: Web Archiving, Voluntarily Deposited Electronic Publications (VDEP) at the British Library, and E-Journals at UCL. By using these Case Studies, which were vastly different in both content and workflow, key costs were identified for each element in the lifecycle, enabling the project to estimate the costs for a single title, item or instance over a given time period.

**Web Archiving**

This Case Study considered the costs of the British Library's web archiving activities, which selected and archives around 1000 web site instances each year.

→ The full Web Archiving Case Study can be found in the LIFE[1] final Report (Section 6, p. 52-63).

**E-Journals**

The e-journals Case Study was based at UCL Library Services. At the time of the Case Study, 8668 e-journal titles were logged in a UCL Access database.

→ The full e-Journal Case Study can be found from the LIFE[1] Project final Report (Section 7, p.64-89).

**VDEP**
Voluntarily Deposited Electronic Publications (VDEP) housed at the BL provided the final Case Study and involved the analysis of over 230,000 files.

→ The full Report of VDEP Case Study Report can be found from the LIFE[1] final Report (Section 5, p.17-51)

The three Case Studies proved to be highly effective in highlighting both the types of issues that can be encountered in a digital collection, and the ways in which a lifecycle methodology can be utilised to capture and apply a cost to solving these problems.

→ More detailed practical and strategic findings for each of the Case Studies can be found from the LIFE[1] Project final Report (Section 9, p.108-113)

## LIFE[2]

After completion of the LIFE[1] deliverables (i.e. developing and testing the initial LIFE Model), it became clear that the model and LIFE approach needed to be further tested, and expanded through a wider range of Case Studies.

One of the key deliverables for LIFE[2] was to make the LIFE model and findings more accessible to those institutions wishing to either adopt the model, or make use of the findings. Essentially, to answer the question – how is the LIFE work useful for our own collections?

The LIFE[1] Case Studies comprised born-digital collections, so a key area of expansion for LIFE[2] was the examination of non-born digital material (The British Library Newspaper Collection Case Study). This Case Study allowed for the comparison of analogue and digital lifecycles and costs.

Institutional Repositories were also addressed in two Case Studies (SHERPA-LEAP and SHERPA-DP). The costs of three Institutional Repositories were modelled to the LIFE work (SHERPA-LEAP Case Study), and the digital preservation services were examined through the SHERPA-DP Case Study.

USEFUL LINKS

| | |
|---|---|
| LIFE[1] Project Documentation | www.life.ac.uk/1/documentation.shtml |
| UK Web Archiving Consortium | www.webarchive.org.uk |
| VDEP at The British Library | www.bl.uk/aboutus/stratpolprog/legaldep/index.html |

# LIFE$^3$

The third phase of LIFE has delivered a predictive costing tool that significantly improves the ability of organizations to plan and manage the preservation of digital content. This tool is based upon a refined version of the LIFE model produced in phase two, following collection of additional case study and survey data. This has enabled the model to cover a wider range of preservation scenarios, including sound, web and e-journal archiving, in addition to print. In addition to producing the tool as an Excel spreadsheet, the LIFE team has also partnered with the Humanities Advances Technology and Information Institute at the University of Glasgow, to produce a web-based version of the tool.

A survey of digital preservation repositories was carried out in order to better understand their storage requirements and costs, with these being correlated to the size and purpose of each system. Aside from the number of mirror sites employed, the survey looked at the combination of storage technologies used for access as well as backup, the cost and expected lifetime of the hardware, and also as other factors such as support, infrastructure and electricity costs.

USEFUL LINKS

| | |
|---|---|
| LIFE$^3$ Project Documentation | www.life.ac.uk/3/documentation.shtml |
| LIFE$^3$ Project Blog | www.life.ac.uk/blog |
| Humanities Advances Technology and Information Institute | www.hatii.arts.gla.ac.uk |

# LIFE[3] MODEL

The LIFE Model provides a view into the typical processes applied to digital objects throughout their lifecycle, by an organisation acting as the custodian of those objects. The processes are loosely organised in a chronological order, from their creation through to eventual access. It should be noted however that processes can, in practice, overlap with each other or be executed in a different order. The model aims to capture common processes found in most digital lifecycles. While some processes may not be applicable to all lifecycles, the intention is to provide meaningful placeholders for the majority of typical lifecycle processes.

$$L_T = C + Aq_T + I_T + M_T + BP_T + CP_T + Ac_T$$

**L** = Complete lifecycle cost over time 0 to **T**.

Other catergories are:

C = Creation
Aq = Acquisition
I = Ingest
M = Metadata Creation
BP = Bit-stream Preservation
CP = Content Preservation
Ac = Access

| Lifecycle Stage | Creation or Purchase | Acquisition | Ingest | Metadata Creation[2] | Bit-stream Preservation | Content Preservation | Access |
|---|---|---|---|---|---|---|---|
| Lifecycle Elements | Conceive Activity* | Selection | Quality Assurance | Re-use Existing Metadata | Repository Administration | Preservation Watch | Access Provision |
| | Selection and Preparation* | Submission Agreement | Deposit | Metadata Creation | Storage Provision | Preservation Planning | Access Control |
| | Transport* | IPR & Licensing | Holdings Update | Metadata Extraction | Refreshment | Preservation Action | User Support |
| | Digitisation* | Ordering & Invoicing | Reference Linking | | Backup | Re-ingest | |
| | Digitisation QA* | Obtaining | | | Inspection | | |
| | IPR* | Check-in | | | | | |

* Optional items for projects involving digitisation.

Stages represent high level processes within the lifecycle which group related lifecycle processes together. Elements represent the next level down in the analysis of lifecycle processes. They are still relatively high level and but are focused on a distinct process within the lifecycle. The LIFE Model attempts to describe a standard set of elements to which most digital lifecycles can easily be mapped. Sub-elements represent the specific components of a lifecycle element. At this level of detail, lifecycles are expected to vary considerably from one to another and so the detailed sub-elements that are provided in the full Model documentation are for guidance only.

The breakdown of components within the LIFE Model:

| Lifecycle Level | Explanation |
|---|---|
| **Lifecycle** | The process from creation to access to preservation for a particular digital object, which can be broken down further into a number of distinct processes. |
| **Lifecycle Stage** | A high level process within a lifecycle. Provides a way of grouping related lifecycle elements. Processes within a Lifecycle Stage typically occur or recur at the same point in time. |
| **Lifecycle Element** | A distinct and significant lifecycle process that will provide useful costing information for organisations to support planning, evaluative or comparative exercises. |
| **Lifecycle Sub-element** | A suggested key component of a Lifecycle Element. Not significant enough to warrant inclusion as a distinct Lifecycle Element. |

A full explanation and analysis of the model is available in a separate document from the LIFE Website.

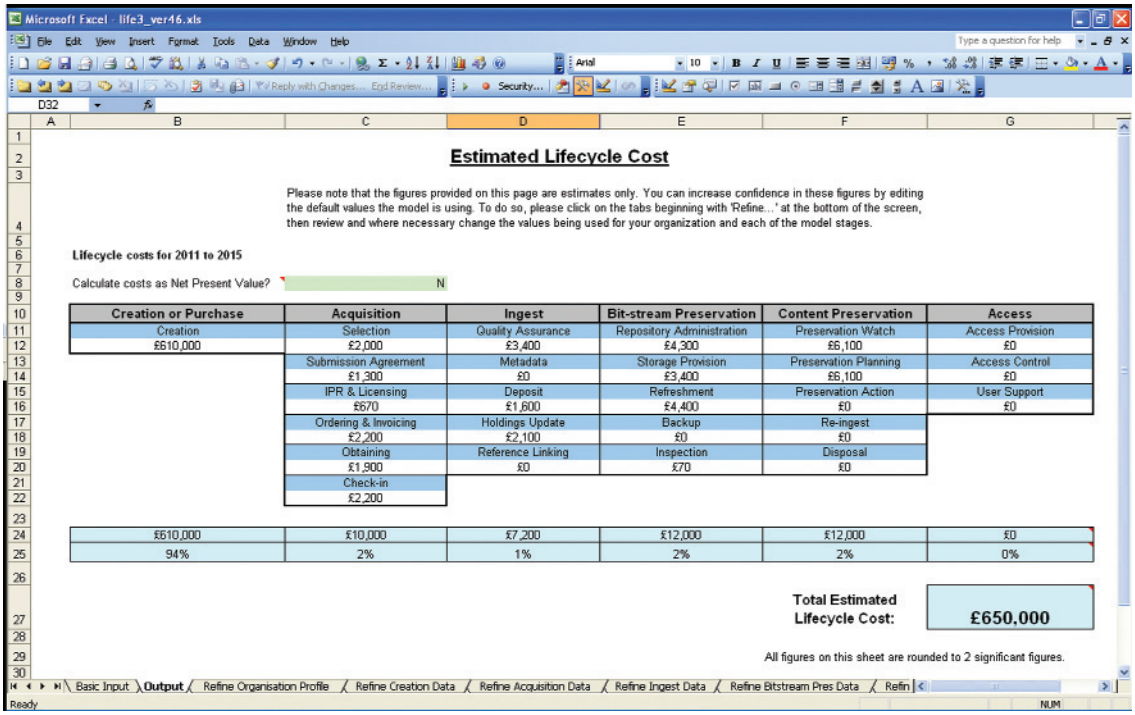
USEFUL LINKS

| | |
|---|---|
| LIFE[1] Model Explanation (in full report) | www.life.ac.uk/1/documentation.shtml |
| Economic Evaluation of LIFE[1] and LIFE[1] | www.life.ac.uk/2/documentation.shtml |
| LIFE[2] Model Update v2 | www.life.ac.uk/2/documentation.shtml |
| LIFE[3] Model Update v3 | www.life.ac.uk/2/documentation.shtml |

# LIFE³ Tool

The model is designed to produce estimated lifecycle costs with a minimum of effort from the user. A template approach was followed to allow the user to select from content and organisation categories into which their particular project falls. The model is then populated with default data calculated from the mean values of case studies that also fall into those categories. The user then has the opportunity to refine the results by reviewing the default data and assumptions made.

A user thus has to enter data into only five fields on the basic input sheet in order to receive an in initial cost estimate.



Microsoft Excel - life3_ver46.xls

**Estimated Lifecycle Cost**

Please note that the figures provided on this page are estimates only. You can increase confidence in these figures by editing the default values the model is using. To do so, please click on the tabs beginning with 'Refine...' at the bottom of the screen, then review and where necessary change the values being used for your organization and each of the model stages.

**Lifecycle costs for 2011 to 2015**

Calculate costs as Net Present Value? N

| Creation or Purchase | Acquisition | Ingest | Bit-stream Preservation | Content Preservation | Access |
|---|---|---|---|---|---|
| Creation | Selection | Quality Assurance | Repository Administration | Preservation Watch | Access Provision |
| £610,000 | £2,000 | £3,400 | £4,300 | £6,100 | £0 |
| | Submission Agreement | Metadata | Storage Provision | Preservation Planning | Access Control |
| | £1,300 | £0 | £3,400 | £6,100 | £0 |
| | IPR & Licensing | Deposit | Refreshment | Preservation Action | User Support |
| | £670 | £1,600 | £4,400 | £0 | £0 |
| | Ordering & Invoicing | Holdings Update | Backup | Re-ingest | |
| | £2,200 | £2,100 | £0 | £0 | |
| | Obtaining | Reference Linking | Inspection | Disposal | |
| | £1,900 | £0 | £70 | £0 | |
| | Check-in | | | | |
| | £2,200 | | | | |
| £610,000 | £10,000 | £7,200 | £12,000 | £12,000 | £0 |
| 94% | 2% | 1% | 2% | 2% | 0% |

**Total Estimated Lifecycle Cost:** **£650,000**

All figures on this sheet are rounded to 2 significant figures.

Basic Input | Output | Refine Organisation Profile | Refine Creation Data | Refine Acquisition Data | Refine Ingest Data | Refine Bitstream Pres Data | Refin

Ready NUM

This information is used to pre-populate the model with data averaged from relevant case studies where it is available, and the user is immediately presented with a cost estimate on the output page. They are able to drill down and change the default values at each stage of the life cycle in order to achieve a more precise result using the refinement sheets, or they can simply reset the model and try a different configuration.

In conjunction with HATII, a web-based tool incorporating the financial model has also been produced. The aim of the tool is to make the LIFE model both easily accessible and easy to operate for all levels and backgrounds of users. As an example of this, when using the tool in comparison to the spreadsheet, only the data that is directly relevant to the user at any point in time is displayed.

USEFUL LINKS

| | |
|---|---|
| LIFE³ Access to Excel and web-based tools | www.life.ac.uk |
| LIFE³ Tool documentation | www.life.ac.uk/3/documentation.shtml |

# Project Documentation

All project documentation and deliverables from all phases of LIFE are available on the LIFE website: www.life.ac.uk/

## LIFE[1] DOCUMENTATION

**LIFE[1] Project Summary**
A short Report providing an overview of the Project's results and findings.

**Research Review**
A detailed literature review that describes the background to the Project, and the selection and development of the methodology and lifecycle approach.

**LIFE[1] Project Final Report and Spreadsheets**
The Report describes the Project's approach, methodology and findings in developing lifecycle techniques to identify and cost the preservation of digital materials. Cost estimations for preservation activity for both the VDEP and Web Archiving Case Studies are also available.

## LIFE[2] DOCUMENTATION

**Economic Evaluation of LIFE[1] and LIFE[2]**
An independent Report evaluating the approach used in LIFE[1] as well as the intended approach for LIFE[2].

**LIFE[2] Model Update – version 1.1**
The working model update used during LIFE[2]. This version of the model will be updated to produce the final LIFE[2] Model v2 which will be included in the final Report.

**LIFE[2] Project Summary**
A short Report providing an overview of activities in the project's second phase.

**LIFE[2] Project Final Report**
The Report will describe the Project's approach, methodology and findings, including:

- A full description of the LIFE[2] methodology and Lifecycle Model
- Detailed Case Studies which apply the cost model to the areas of Institutional Repositories using SHERPA-LEAP and SHERPA-DP developments, and Digitised Newspapers at the British Library
- Findings and conclusions from the Project

**Case Study Spreadsheets**
Spreadsheets providing detailed lifecycle costing activity for each of the Case Studies.

**Project Papers and Presentations**
All journal and conference papers produced for the Project, as well as any other Project presentations.

# LIFE³ DOCUMENTATION

**LIFE³ Model Update – version 3**
The working model update used during LIFE³. This version of the model was updated to produce the final LIFE³ Model v3 which is included in the final Report.

**LIFE² Project Summary**
A short Report providing an overview of activities in the project's second phase.

**LIFE³ Project Final Report**
The Report describes the Project's approach, methodology and findings, including:

- A description of the LIFE3 methodology and Lifecycle Model
- A full description of the LIFE3 Predictive Costing Tool
- Details of the LIFE3 storage survey

**The LIFE³ Predictive Costing Tool (Excel)**
The Excel version of the tool, for advanced users.

**The LIFE³ Predictive Costing Tool (Web-based)**
The web-based version of the tool, for greater accessibility.

**Project Papers and Presentations**
All journal and conference papers produced for the Project, as well as any other Project presentations.

www.life.ac.uk

ISBN 978 0 7123 5839 2