

Lifecycle Information for E-literature

An Introduction to the second phase of the LIFE project

Produced for the LIFE² Conference, 23 June 2008

A JISC-funded joint venture project under 03/06, Repositories and Preservation Capital Programme, and supported by the LIBER Access and Preservation Divisions









THE LIFE PROJECT TEAM

Paul Ayris

Director of UCL Library Services (UCL)

Richard Davies

LIFE² Project Manager, British Library

Rory McLeod

Digital Preservation Manager, British Library

Rui Miao

LIFE² Project Assistant, British Library

Helen Shenton

Head of Collection Care, 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)

Dr Paul Ayris (Chair)

Director, Library Services (UCL)

Sayeed Choudhury

Associate Director for Library Digital Programs (The Johns Hopkins University)

Richard Davies

LIFE² Project Manager, British Library

Rory Mcleod

Digital Preservation Manager, British Library

Helen Shenton

Head of Collection Care, British Library

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 reach its aims.

In particular LIFE would like to thank people in the following project areas:

The LIFE Model

Thanks to the following for feedback and comments on the updated model – Neil Beagrie, Bo-Christer Björk (Hanken, Swedish School of Economics and Business Administration), Ulla Bøgvad Kejser (Royal Library, Denmark), Peter Bright (British Library), Anders Bo Christensen (State Archives, Denmark), Birte Christensen-Dalsgaard (State and University Library, Denmark), Adam Farquhar (British Library), Birgit Nordsmark Henriksen (Royal Library, Denmark), Jan Nebber-Christensen (State Archives, Denmark)

SHERPA-LEAP Case Study

Thanks to Rebecca Stockley (SHERPA-LEAP), Martin Moyle (UCL), Jacqueline Cooke (Goldsmiths), and Adrian Machiraju (Royal Holloway).

SHERPA-DP Case Study

Thanks to Sheila Anderson, Steve Grace, and Gareth Knight at the Centre for e-Research (CeRch).

Medical Research Council Case Study

Unfortunately the MRC case study was not able to be completed due to staffing difficulties. However thanks go to Paul Watters and Diane Kurr for their support of the LIFE work.

British Library Newspapers Case Study

Thanks to Jane Shaw, Lucy Evans, Patrick Fleming, Richard Gibby, Ed King, Stephen Morgan, Deborah Novotny, Dawn Olney, Bhavna Tailor from the British Library.

JISC

Thanks to Neil Grindley, Digital Preservation Programme Manager (JISC).

Project Support

Thanks to Chris Carrington at UCL for the website development and support.

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 of the LIFE Project, LIFE², has refined the LIFE Model and added three new exemplar Case Studies to further build upon LIFE¹. LIFE² is an 18-month JISC-funded project between UCL (University College London) and The British Library (BL), supported by the LIBER Access and Preservation Divisions. LIFE¹ was completed in April 2006. LIFE² started in March 2007, and will complete in August 2008.

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

There are three main areas discussed:

- 1 **From LIFE**¹ **to LIFE**² 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 1.1) which has been thoroughly updated from the first phase.
- 3 **LIFE**² **Case Studies** describes the three new Case Studies for LIFE². It does not include the results from the Case Studies (these will be available in the final LIFE² Report), rather some background on each of the studies as well as discussion of why they were chosen.

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

JISC

LIFE Project Website LIFE Project Blog

UCL Library Services

www.bl.uk/dp

www.jisc.ac.uk www.life.ac.uk

www.life.ac.uk/blog www.ucl.ac.uk/library

From LIFE¹ to LIFE²

What follows is a brief summary of the first phase of the LIFE Project (LIFE¹) and the motivation for the second phase of the project (LIFE²). All documentation referred to is available from the LIFE website (www.life.ac.uk).

LIFE¹ Summary

Run from 2005 to 2006, the LIFE¹ 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 LIFE¹ 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 LIFE¹ Project final Report (Section 4, p.9-16)

Generic Preservation Model

The development of the Generic Preservation Model (GPM) helped to establish the cost to preserve digital assets within the Lifecycle Model, but in isolation from other areas such as ingest and metadata. Further development of the model, integration with the broader lifecycle approach and refinement of its inputs using real data will be crucial in taking this forward.

→ A detailed introduction to the Generic Preservation Model can be found in the LIFE¹ 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 LIFE¹ 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 LIFE¹ 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 LIFE¹ 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 LIFE¹ Project final Report (Section 9, p.108-113)

LIFE²

Feedback received at the LIFE¹ conference provided a useful steer on the design of the LIFE² Project. Key aspects include further validation of the model, economic assessment of the LIFE approach and further testing and evidence generation via further case studies.

One of the key deliverables for LIFE² is 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 final LIFE² End of Project Report will capture key elements of this discussion. The Report will be made available at the end of the Project in August 2008.

The LIFE¹ Case Studies comprised born-digital collections, so a key area of expansion for LIFE² 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 to begin a cost comparison.

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

Mindful of feedback from the LIFE¹ Conference one of the first steps taken for LIFE² was to undertake a full economic evaluation of the approach used for the LIFE work. This independent evaluation largely validated the LIFE approach, and provided valuable feedback for the LIFE² Model updates. As with all the project outcomes, this independent report is available from the LIFE website.

USEFUL LINKS

LIFE¹ 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

LIFE² 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 to long-term preservation 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.

$$L_{\tau} = C + Aq + I_{\tau} + M_{\tau} + BP_{\tau} + CP_{\tau} + Ac_{\tau}$$

L = Complete lifecycle cost over time 0 to T.

C = Creation Aq = Acquisition I = Ingest

M = Metadata Creation
BP = Bit-stream Preservation
CP = Content Preservation

Ac = Access

| Lifecycle Stage | Creation or Purchase ¹ | Acquisition | Ingest | Metadata Creation ² | Bit-stream Preservation | Content Preservation | Access |
|--------------------|---|-------------------------|----------------------|-----------------------------------|------------------------------|--------------------------|---------------------|
| Lifecycle Elements | | Selection | Quality Assurance | Re-use Existing Metadata | Repository Administration | Preservation Watch | Access Provision |
| | | Submission Agreement | Deposit | Metadata Creation | Storage Provision | Preservation Planning | Access Control |
| | | IPR & Licensing | Holdings Update | Metadata Extraction | Refreshment | Preservation Action | User Support |
| | | Ordering & Invoicing | Reference Linking | | Backup | Re-ingest | |
| | | Obtaining | | | Inspection | | |
| | | Check-in | | | | | |

¹ This stage may be beyond the scope of some costing activities. Creation may occur outside the view of the costing institution. It should therefore be considered to be optional. Where considered within scope, elements will need to be tailored to the specific lifecycle case in question.

² Metadata Creation is often considered to be part of the process of Ingest. Due to its significance within the lifecycle, it has been represented within the Model as a distinct lifecycle stage.

Stages represent high level processes within the lifecycle that group related lifecycle processes together. Elements represent the next level down in the analysis of lifecycle processes. They are still relatively high level 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

LIFE¹ Model Explanation (in full report) Economic Evaluation of LIFE¹ and LIFE² and LIFE² Model Update v1.1 www.life.ac.uk/1/documentation.shtml www.life.ac.uk/2/documentation.shtml

Case Studies in summary

As with LIFE¹, the Case Studies form the basis of the LIFE² Project. Three Case Studies were chosen to help refine and review the Lifecycle Model developed in LIFE¹, as well as to expand the testing of the Model to new areas.

The three Case Studies chosen for LIFE² were:

- SHERPA-LEAP Institutional repositories in the federal University of London
- SHERPA-DP Distributed repository environment for digital preservation of content
- British Library Newspapers Digitisation as surrogacy

SHERPA-LEAP Repository Case Studies

The Project wished to provide a tool that can be used throughout the UK, and globally, to cost the lifecycle and long-term digital curation of deposited research outputs. LIFE² developed a range of costing studies to complement the outputs of the Case Studies in LIFE¹. Based on repository development, these case studies used the SHERPA-LEAP and SHERPA-DP Projects as testbeds for identifying lifecycle costs and the costs of digital preservation in these areas.

SHERPA-LEAP (a partner in the UK SHERPA consortium) is a University of London consortium, led by UCL, which has helped to create open access institutional EPrints repositories at 13 University of London institutions. Established in 2004, the two overriding aims of the project were to create EPrints repositories for each of the partner institutions, and to populate these repositories through collaborative advocacy.

Three institutional repositories were chosen for the Case Study, to represent a range of material:

- Goldsmiths, University of London contains examples of research output from the visual and performing arts.
- Royal Holloway, University of London contains a range of text-based materials (mostly journal and conference papers).
- UCL contains mostly text-based objects (journal and conference papers, and book chapters); only a few audio-visual objects are currently held.

Goldsmiths Research Online

Goldsmiths Research Online, the institutional repository of Goldsmiths University of London, was set up in 2006 on the SHERPA-LEAP server hosted at UCL. It was intended to represent the diverse, creative qualities of the research environment at Goldsmiths. The Case Study has been included because it contains examples of research outputs from the visual practice and performing art perspectives, and few repositories hold this kind of material.

The LIFE Model was utilised in Goldsmiths to assist in the evaluation of repository development, identifying the problematic elements of including practice-based arts research in repositories, and assessing the different levels of attention and expertise required for different types of digital objects. The model was also used to assist in assessing each specific type of output, to identify whether collections, cataloguing and curation fell within the current scope of the repository.

Royal Holloway Research Online

During 2005, Royal Holloway Research Online was set up for Royal Holloway University of London on the SHERPA-LEAP server hosted at UCL. The purpose of the repository is to increase access to Royal Holloway's research output. Currently all the material is text based. The great majority of the collection comprises journal papers, working papers, or chapters from books.

While the repository was hosted on the SHERPA-LEAP server, UCL has been responsible for the digital preservation of the repository. With planning now underway to transfer the repository to a server on campus in 2008, Royal Holloway Research Online started facing the issue of long-term preservation. This Case Study is a good example for repositories with a lack of experience in digital preservation. The LIFE Model was used to address the lifecycle stages of its digital collection,

to estimate the long-term cost of digital preservation, and to make strategic plans for future preservation activities.

UCL EPrints

UCL EPrints was founded in March 2004. The initial decision was taken to build the repository 'from the ground up', with the aim of building a critical mass of expertise, content and support, backed by usage data, in order to be able to demonstrate the value of the repository to the institution and to secure permanent recurrent funding.

Most of the content in UCL EPrints is textual, with only a few audio-visual objects currently being held. Journal articles, conference papers and book chapters form a high proportion of the content. Working papers, theses, patents, reports and other outputs are also represented. It is noted that around half of the records in the repository are currently metadata-only records, due to copyright restrictions or the need to maintain complete and updated publication listings for academic departments.

With a digital preservation policy now prepared, and adequate arrangements for bit-level preservation in place, UCL Library Services has put a lot of effort into investigating preservation strategies and solutions for the content of its EPrints repository. The LIFE Model will be used as a tool in assisting the repository to decide its long-term preservation solutions (e.g. in-house, outsourcing, etc.), as well as its deposit workflows.

SHERPA-DP Case Study

The Centre for e-Research (CeRch) at King's College London was established following the demise on 31 March 2008 of the Arts and Humanities Data Service (AHDS). AHDS was funded to provide a national service of digital curation and preservation for arts and humanities collections. It is also possible that the new service could provide a shared preservation environment for several universities' institutional repositories, following the outcomes of the SHERPA-DP Project. SHERPA-DP posits a preservation layer on top of repository curation, sharing AHDS's expertise in preservation planning and activities across multiple institutional repositories.

The SHERPA-DP Case Study is used to test the implementation of the LIFE Model in an explicit context of digital preservation, and to consider the broader external factors affecting preservation costs. AHDS undertook digital preservation for more than a decade, with all collections ingested and managed in explicit conformance with the OAIS model, enabling the Case Study to show costs relating to preservation when conducted as a third-party, or outsourced, service for Institutional Repositories, and to compare them with the preservation costs for in-house preservation.

The LIFE Project developed the lifecycle costing methodology that focused specifically on the cost implications of the preservation workflow and the content that was being preserved. However, as explored in the Generic Preservation Model, there is a range of external influences that have the potential to change the Lifecycle Model significantly and to reduce costs. The SHERPA-DP Case Study also considered the implications of digital lifecycle elements being undertaken by different institutions, and the significant potential of these centralised and distributed processes for cost savings.

British Library Newspapers Case Study

In LIFE², a key expansion of the LIFE¹ work is to examine whether the LIFE Model could be used to capture the costs of both analogue and digital lifecycles, and if possible, to compare the costs of both lifecycles at the same stages of the LIFE Model. The comparisons between analogue and digital lifecycles are crucial to making future collection management decisions. For example, when faced with the decision to acquire an analogue or digital version of the same object, which one provides the better solution in terms of cost and sustainability?

To help identify solutions for this question, this Case Study used the LIFE Model to provide:

- 1. A direct cost comparison between paper and digital formats.
- 2. A possible method of supporting decision making to help libraries decide what to keep when space or cost is a concern.

Two British Library Newspaper Collections were used to track the applicability of the LIFE Model and to compare the lifecycles of digital and analogue collections.

Burney Digital Collection

The Burney collection is a collection of Newspapers purchased from the Reverend Dr Charles Burney in 1818 for £18,500. It comprises over 1,100 volumes of the earliest-known newspapers from the 17th and 18th centuries.

In order to preserve this rare collection, the Burney collection was microfilmed in the 1970's and then digitised, generating close to 1,000,000 page images. This content forms the Burney Digital Collection.

■ Legal Deposit Newspaper Collection

The British Library receives a copy of every national newspaper, and the majority of regional daily and weekly newspapers, under legal deposit legislation. 133,000 issues arrive every year, and the costs for one year's curation of the collection are used for the analogue part of this Case Study.

The Case Study applied the LIFE Model to both collections, producing a detailed explanation and comparison of each of the lifecycle stages in the analogue and digital materials. Additionally, a comparison of the processing workflows in the two collections was also conducted.

The detail in the Case Study will have direct relevance to all libraries which are digitising newspaper collections, or facing the debate of choosing to preserve either digital or analogue versions of the same material. The key output from this Case Study will also provide Higher and Further Education Libraries with guidance in the preparation and planning of a bid for the digitisation of material in their care, as well as in the preservation of outputs from digitisation projects.

USFFUL LINKS

Full LIFE² Case Study Reports www.life.ac.uk/2/documentation.shtml

AHDS www.ahds.ac.uk

Burney Newspaper Collection www.bl.uk/collections/burney.html

Newspapers at The British Library www.bl.uk/collections/newspapers.html

OAIS Model www.dcc.ac.uk/resource/curation-manual/

chapters/oais-model

SHERPA www.sherpa.ac.uk
SHERPA-DP www.sherpadp.org.uk
SHERPA-LEAP www.sherpa-leap.ac.uk/

LINKS TO INSTITUTIONAL REPOSITORIES

Goldsmiths Research Online http://eprints.goldsmiths.ac.uk
Royal Holloway Research Online http://eprints.rhul.ac.uk
UCL EPrints Repository http://eprints.ucl.ac.uk

PROJECT DOCUMENTATION

LIFE Project Summary

All project documentation and deliverables from both LIFE¹ and LIFE² are available on the LIFE website: www.life.ac.uk

LIFE¹ DOCUMENTATION

LIFE¹ 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.

LIFE¹ Project Final Report & 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.

LIFE² DOCUMENTATION

Economic Evaluation of LIFE1 and LIFE2

An independent Report evaluating the approach used in LIFE¹ as well as the intended approach for LIFE².

LIFE² Model Update - version 1.1

The working model update used during LIFE². This version of the model will be updated to produce the final LIFE² Model v2 which will be included in the final Report.

LIFE² Conference Project Summary

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

LIFE² Project Final Report

The Report will describe the Project's approach, methodology and findings, including:

- A full description of the LIFE² 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
- The Report will be made available at the end of the Project (August 2008)

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.

www.life.ac.uk ISBN 978 0 7123 5065 5