and

CLIF Project

Literature Review

January 2010

The CLIF Project

Project Director:	Chris Awre	(c.awre@hull.ac.uk)
Project Manager:	Richard Green	(r.green@hull.ac.uk)
Project Site Manager for King's College:	Mark Hedges	(mark.hedges@kcl.ac.uk)

The Content Lifecycle Integration Framework (CLIF) Project is being undertaken by the Information Management and IT Systems Groups at the University of Hull and the Centre for e-Research (CeRch) at King's College London. It is funded by the JISC Information Environment Programme 'Repositories Enhancement' strand.

Introduction

As indicated in the proposal for the CLIF project, the concept of a content lifecycle is not a new one. Records managers have long recognised its importance to their work. Projects within the JISC Supporting Institutional Records Management programme in 2003[1] covered this and looked to build on the previous JISC Study of the Records Lifecycle project (Parker, 2000). The lifecycle being described in these studies is not limited to digital content, but all content that might be considered a 'record'. The lifecycle concept has also, though, been specifically adopted where records management has moved into the digital arena. The MoReq2 specification[2] for electronic records management systems refers to the document lifecycle, though it is not an extensive part of the document. In the commercial world of enterprise content management (ECM) there is much consideration of how the content lifecycle can be improved to maximise the benefit the content offers a business, and there is no shortage of commercial offerings to enable this[3].

As digital content has become more and more prevalent in business and other walks of life there have been equivalent developments in the provision of systems and associated businesses to assist in managing it. The ECM systems referred to in the previous paragraph are one example of a system being designed solely for that purpose. In a presentation at the Digital Preservation Coalition (DPC) workshop 'Preservation of e-Learning Materials and Cost Models for Digital Preservation' in 2002, Meg Bellinger from OCLC described the approach OCLC was taking to developing digital life cycle management services. This revolved around how a third party could support preservation. This contrasts with the ECM commercial approach of selling a system, but reiterates the idea that the digital lifecycle requires active management (Granger, 2002).

Many other systems designed to provide a specific functional purpose also incorporate content management capability, as quite often the system has to hold the content in order to carry out its other functions, e.g., the ability of virtual learning environment (VLE) systems to hold content that can then be used for teaching & learning. There are thus many approaches to managing digital content.

The approach of using the content lifecycle to guide the management of content, by applying the content lifecycle concept to a specifically digital environment, is an evolving one. At the same DPC event referenced in the paragraph above Helen Shenton from the British Library (BL) described the BL's work in defining a content management lifecycle for the increasing amount of digital content being acquired by the Library. The investigation of such a digital lifecycle sought to see how traditional models for managing the lifecycle of print resources could be adapted. The BL also sought to identify how to allocate practical resources to dealing with different stages of the lifecycle.

The evolution of a digital content lifecycle can thus be considered from two perspectives:

[1] JISC Supporting Institutional Records Management programme, http://www.jisc.ac.uk/whatwedo/programmes/supportingirm
[2] MoReq2 specification, http://www.moreq2.eu/index.htm
[3] For example, see http://www.ecmconnection.com/article.mvc/Learn-From-Content-Lifecycle-Transformation-0002 or http://www.opentext.com/2/global/sol-products/sol-pro-extensions-microsoft/pro-clm-sp-edocs.htm

- Retaining emphasis on the process that is being performed, irrespective of the system, manual or digital, which is being used to carry out the management. This takes the records management view as a starting point, where the focus is the record itself, not the system.
- Giving attention to the system or systems involved in facilitating the management of the content lifecycle, to understand how these systems can be best utilised to support the lifecycle management of the digital content. The term 'system' could here refer to non-technical organisational systems as well as technical ones, though the emphasis here is on the latter.

This literature review takes as its starting point the need to understand better the second of these perspectives, though recognising that they are often blurred. It does not aim to produce or summarise the many different examples of lifecycle in existence, but instead addresses the issues that have emerged through developing or examining such lifecycles. In doing so, there is a related need to better understand the first perspective above so we can appreciate what systems are being asked to do as well. Examples of the literature addressing both perspectives have been drawn on, with the aim of informing technical development within the CLIF project, and potentially elsewhere, on how the digital content lifecycle can be managed across different systems at different stages of the lifecycle.

The description of the literature has been divided up into a number of themes, as listed below. These are somewhat arbitrary in nature, but represent those areas that emerged from the literature as the most useful way of understanding the digital content lifecycle for the needs of the CLIF project.

- Content lifecycles
- Lifecycles and digital preservation
- Lifecycles and content management
- Technology and the lifecycle
- Living the lifecycle
- Standards
- The knowledge lifecycle
- Developing the lifecycle
- Conclusions

The literature review starts, however, with a brief consideration of the terminology used to help provide a degree of consistency in reporting this multi-faceted field.

Terminology

In researching the literature, it is apparent that the terms digital, lifecycle (or life cycle), and content have been used and referred to in a variety of ways and in different combinations. Some papers refer to the digital lifecycle, some to content lifecycle, some to the digital content lifecycle. The latter has been used as the preferred term for this review, and differences in definition highlighted where these are considered valuable. The differences possibly highlight a lack of definition or meaning about what is being described, and some of these differences are discussed further in the text.

There is also a discrepancy in whether what is being described is a life cycle or a lifecycle. The latter is used for consistency except where it is referenced as a direct quote, although there does not appear to be any particular preference for one or the other in wider usage.

Content lifecycles

The Digital Curation Centre (DCC) Curation Lifecycle Model is one of the major outputs from the Digital Curation Centre, and reflects the coming together of a body of views and models for managing digital preservation. The initial draft publication of this (Higgins, 2007) highlights that the Lifecycle Model is a generic overview, and that more granular functionality can be mapped onto it according to the specific content domain being addressed for curation. It also highlights that there is existing work in a number of those specific domains that can be applied: for example, examinations of the role of a lifecycle within e-science research (e.g., Humphrey, 2006), or the work of the Paradigm project on personal archives[4]. The final publication of the Lifecycle Model (Higgins, 2008) emphasises that a lifecycle management approach to digital materials "…is necessary for their continuity."

This builds on an earlier piece of research at UKOLN (Pennock, 2007). This stated:

"The life cycle approach is necessary because:

- Digital materials are fragile and susceptible to change from technological advances throughout their life cycle, i.e. from creation onwards;
- Activities (or lack of) at each stage in the life cycle directly influence our ability to manage and preserve digital materials in subsequent stages;
- Reliable re-use of digital materials is only possible if materials are curated in such a way that their authenticity and integrity are retained."

The DCC Curation Lifecycle Model was also informed by the presentations at the DCC event 'Digital Curation and Preservation: defining the research agenda' held in Warwick in November 2006. One of the strands of this event was about Data Life Cycle Management[5]. In her introduction, Anne Trefethen emphasised that when looking to preserve research it is valuable not just to capture the core of what the research did or produced, but the whole context or lifecycle of the ideas and processes involved. This would involve different stakeholders at different points in the cycle. Jeremy Frey described the scholarly life cycle in his presentation, complementing ideas proposed by Liz Lyon (Lyon, 2003) that demonstrated how many different parts of scholarly activity interconnect, whilst Helen Shenton highlighted aspects of lifecycle cost that have been investigated through the LIFE projects at UCL and the British Library. The formula developed by the LIFE project, $L_T=Aq+I_T+M+Ac_T+S_T+P_T$, identified criteria that affected the long-term cost management of content (McLeod et al., 2006), the focus being on the costs of acquisition and the tasks required to manage the content once acquired. The systems that this

[4] Paradigm project, digital archives and the records cycle diagram, http://www.paradigm.ac.uk/workbook/introduction/paradigm-lifecycle.html

[5] Data Life Cycle Management strand, Digital Curation and Preservation: Defining the Research Agenda for the Next Decade, http://www.dcc.ac.uk/events/warwick_2005/sessions/data_lifecycle_management/

will be carried out through, whilst recognised as integral to the different lifecycle stages, were not a cost factor that could be effectively incorporated into a lifecycle approach (Ayris et al, 2008).

A range of examples of digital content lifecycles building on the direction the DCC set down is evident.

1. In a case study for the SCARP project (Irshad and Ure, 2009) on TeleHealth the concept of the TeleHealth lifecycle is described, highlighting the generation of data and its subsequent processing for effective use. The case study highlighted the system designed to support this, but the subsequent discussion around data curation did not address specific system issues.
2. OCLC (Anon, 2005) describes the four stages of its digital content lifecycle as planning, processing, presentation and preservation, and is building and delivering services around each of these in an integrated fashion (as previously reported in 2002 – see Introduction).
3. The UK Research Data Service Feasibility Study (2009) is also building its model around the concept of a research data lifecycle, taking its lead from related work at the University of Exeter and within the DCC.

As is evident from the previously described references, from the beginning of the DCC there has been an appreciation of the value of a lifecycle approach to curation. Rusbridge (2005) presents a generic overview of the role it can play. He also demonstrates that different lifecycles can operate in parallel or tandem with each other. For example, an experimental lifecycle and the lifecycle of managing the information that feeds into the experiments and comes out of them.

In a presentation to the Common Solutions Group in the US Long (2003) describes how the management of the digital landscape relies on understanding and managing the digital content lifecycle. The digital landscape is made up from many different types of content. These have different lifespans, and thus move through their lifecycle over different times. Understanding the lifecycle can assist in managing these different types of content regardless of the length of lifespan. In the same presentation, Long's co-presenter Ann Green proposes that the lifecycle view can be applied to projects as well as digital content resources, and, indeed, many lifecycles are initiated by faculty members requesting a piece of work. In both cases there are four main identified stages to the digital lifecycle: produce, publish, repurpose, preserve. Common requirements across these stages include: the use of standards, storage capability, and persistence. Not all content or projects will go through every stage, but putting in place infrastructure and tools that support these stages can then enable a variety of initiatives over time. The authors concede that it remained to be identified what the costs of each individual stage were, both initially and ongoing, but consider that the approach can facilitate the management of new demands more flexibly.

In an examination of the management of personal digital archives, Williams et al. (2009) reviewed the literature on this topic to identify a coherent intellectual framework to aid understanding of how individuals create, organise, manage, use and dispose of digital content within their personal archives. They discovered that whilst there is much mention of lifecycles in the literature in various guises (as identified within this literature review), its application to personal collections has been limited (though

note the work of the Paradigm project[6]). They highlight the work of Bruce et al. (2004) who identified the concept of a short-term management stage within the lifecycle of personal digital archives, where content created or obtained is kept, left or ignored depending on its perceived value at different times. This matches the proposal by Sellen and Harper (2002) that content can be hot, warm or cold in terms of current activity and/or interaction. This view highlights the dynamic nature of personal archives, and how they fluctuate, and the need to identify more closely how they can be managed to facilitate this across systems. Williams et al. also report on the issue of enabling personal digital archives when the content within them is so varied. Some specific systems have emerged to help address this issue, and these appear to work through a combination of gathering materials together in one place, and integrating with systems holding the content to permit access from a central point.

Lifecycles and digital preservation

As indicated in the Introduction, one of the key stimuli to investigating the idea of a digital content lifecycle has been the growth of interest and activity around digital preservation. The driving issue has been recognition that digital preservation cannot start just when a piece of digital content is handed over to archivists or equivalent for it to be 'preserved'. It should instead start from the point at which the digital content is created, or whenever it can prior to being formally considered ripe for preservation: this emphasises that the content passes through different stages that can be considered its lifecycle. A number of studies have examined this space.

The Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access in December 2008 (Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008) highlighted that economically sustainable digital preservation required:

- Recognition of benefits
- Incentives for decision-makers to act
- Selection
- Mechanisms to support ongoing, efficient allocation of resources
- Appropriate organisation and governance

Key to the incentives was the need to orchestrate incentives at all stages of the digital content lifecycle, recognising that preservation within the lifecycle cannot be enabled unless it is considered at earlier stages.

Pennock (2006) in providing guidance on the curation of emails makes use of the lifecycle approach to be subsequently seen in the DCC Lifecycle Model. Email curators are advised to consider curation activities at all stages of the email lifecycle, not just at the point at which someone feels it needs to be preserved. A general agreement with this point is espoused in the DPC's Preservation of Digital Materials: a Handbook (DPC, 2008), which also recognises that different stakeholders may be involved at different stages of the digital content lifecycle, and that working with these stakeholders will be

[6] Paradigm project, http://www.paradigm.ac.uk/

necessary to effect this whole lifecycle approach as well as being beneficial simply in engaging with them. Hockx-yu (2006) emphasises JISC's commitment to the integration of digital preservation within the lifecycle of information management and not as a separate activity.

Taking account of the many views that digital preservation needs to take place at all stages of the digital content lifecycle, Ross and Hedstrom (2005) flip this round and emphasise the benefit of digital preservation taking place "in the context of a lifecycle."

Lifecycles and content management

Digital content lifecycle management doesn't have to have digital preservation as its originating focus, though. Reviewing content management in the context of information architecture, Batley (2006) summarises previous work (incl. Boiko, 2001 and Tredinnick, 2005) on the information lifecycle as a way of guiding content management strategy, a similar approach to that taken above. She emphasises the importance of addressing the needs of information in 'information architecture' to ensure that a theoretical architecture can be effectively implemented. The five stages described are:

- Information creation and collection
- Information approval
- Information deployment (or publication)
- Information review
- Information archiving and/or deletion

Wu and Liu (2001) in an early review of how content management systems might impact academic libraries, consider that placing much of the digital content libraries have to manage in a content management system would bring benefits. They reference the lifecycle of the content as being that time that the content has to be made available to users, and suggest that calendaring within the content management system could control the switch on and off of availability, ensuring that the digital content lifecycle, and the stages listed above, are managed in an automated way.

Deegen (2001), in a general review of digital library developments, highlights that the most important reason to adopt a lifecycle approach is to create a sustainable resource. Whilst this is related to digital preservation she also emphasises that the nature of digital materials requires a more proactive management approach generally anyway: when a library receives or generates a digital resource it cannot just put in on the shelf, so to speak, as it does for books and assume it will still be accessible and available in the future. As proposed in digital preservation discussions, there is benefit in taking account of the whole lifecycle of a digital resource from the start so it can be actively managed.

Part of this management involves describing the digital content in an effective way so it can be organised and accessed. In an article describing the Metadata Encoding Transmission Scheme (METS), McDonough (2006) describes the structure of the scheme and its constituent parts. One of these parts is designed to record provenance information, or as he puts it, information related to the lifecycle of the digital object being described. This information can also be described as information about a particular

state or instance of the object, for example, when an image was created or when it was migrated between formats. This information does not manage the lifecycle itself, but it can be used to inform curatorial actions at later stages in the digital content lifecycle.

Technology and the lifecycle

The European Task Force on Permanent Access (now the Alliance for Permanent Access) reiterates the first of the points highlighted by Pennock in 2007 in its Strategic Action Programme 2006-10, reporting that the "life of a digital text is determined by the information carrier and by the hardware and software that make it accessible to users. These means of storage and intermediaries are very vulnerable and have short lives." Inevitably taking a preservation viewpoint of managing this problem, the proposed solution, which the Task Force went on to explore in detail, was the transfer of the bits and bytes to a new carrier. The Information Technology and Information Storage industries (SNIA association) concurs having addressed this issue, referring to the "… most appropriate and cost effective IT infrastructure from the time information is conceived through its final disposition…"[7].

So technology is clearly important, or at least a necessity. But what type? Although heavily promoted by the ECM industry, monolithic systems for the management of the digital content lifecycle can have disadvantages. They are not dynamic and can be difficult to scale and interoperate with other enterprise systems. Chieu et al. (2008) describe an approach using a service component architecture, modelled along SOA lines. This approach allowed the authors to successfully demonstrate the system's extensibility and scalability.

Tzitzikas (2007) describes the ubiquity of dependency in systems, using the development of preservation information systems, as undertaken by the European Union (EU) CASPAR project[8], as an exemplar for the lifecycle of information is affected and influenced by dependencies that need to be taken into account when designing relevant systems. This sense of dependency influenced Ioannidis et al. (2005), when considering the necessary systems infrastructure for digital libraries. They identified five stages of their own lifecycle in interaction with a digital library, and subsequently describe the research issues related to each of these that require further investigation. The stages identified are:

- The user interacts with the system
- The system processes the request
- The information requested is accessed and retrieved
- The information is presented to the user (following any processing required)
- The user organises the information presented

[7] See http://en.wikipedia.org/wiki/Information_lifecycle_management
[8] EU CASPAR project, http://www.casparpreserves.eu/

Whilst not focusing on the content itself, these stages highlight how interaction with digital content through systems can have its own set of stages that may impact on the lifecycle of the content concerned: a case of different types of lifecycles overlapping with and influencing each other.

The Rights and Rewards project at Loughborough University addressed the issue of digital content lifecycles as part of its assessment of which teaching materials to include in its repository (Bates et al., 2006). Other systems were not investigated, but the project identified a set of criteria to assess materials so as to determine whether the repository was the most appropriate place to hold them and, if so, in what format and state of organisation (e.g., packaged or granular) and when. The criteria used were:

- Persistence – the management of content will vary depending on whether it is static (i.e., won't change) or dynamic (i.e., may change) whilst in the repository. Different strategies are required for each.

- Versions – Where content is dynamic, there may be versions to manage. How does a lifecycle approach deal with reiterations of the same material? What constitutes a new version? These matters were queried by the project, and have also been addressed extensively elsewhere (Brace, 2008; Razum et al., 2007).

- Creation (method and workflow) – A third criterion was the way in which materials were created, or how they were adapted for local use from external sources. Different mechanisms had different lifecycles and the best time for interaction with the repository varied slightly between them.

Another factor that was highlighted for further investigation was the nature of collaborative working on the lifecycle of content, and the degree to which a repository could be involved in stages of a digital content lifecycle involving minor editing or reviews. The impression given, albeit one concluded in 2006 prior to the growth of Web 2.0, was a doubtful one, preferring the repository to be used for stable materials.

Living the lifecycle

Many of the conclusions from the Rights and Rewards project related to the practical implementation of the digital lifecycle, and understanding how this can be done most effectively. Appreciating the value of a lifecycle approach to managing digital content, and seeking to put this into regular practice on a day-to-day basis, the University of Illinois Library has proposed the restructuring of its Library staff to align them with different stages of the digital content lifecycle (Digital Content Life Cycle Management Team, 2008). A specific coordinator is recommended to assist different teams in working more closely together to facilitate lifecycle management and there are suggestions based around different stages: creation, selection, management (including access and stewardship), metadata, and preservation.

Also examining how to put the management of the digital content lifecycle into practice, the LIFE-SHARE project[9], currently running at the Universities of Leeds, Sheffield and York, is aiming to build on the outcomes of the LIFE projects and identify the skills, strategies and knowledge required at each stage of the digital content lifecycle. No mention is made of systems specifically, though its outputs are keenly awaited.

One of the emphases of the LIFE-SHARE project is the digitisation of content and the lifecycle of this process: another example of a specific lifecycle that can have an impact on the wider management of digital content. The 'Make It Digital' website in New Zealand, a service developed by Digital New Zealand, presents its own digital content lifecycle in respect of the digitisation process[10]. This incorporates the following stages:

- Selection of content
- Creation of the digitised content
- Describing the content
- Managing the content (including transfer between systems as hardware/software becomes obsolete)
- Discovering the content
- Using and reusing the content
- Preserving the content

There are clearly some similarities here with more generic digital content lifecycle stages, with this being another specific instance. It raises the possibility that process lifecycles can run in parallel and intersect as and when useful.

The Interuniversity Consortium for Political and Social Research (ICPSR) (McGovern and Stuchell, 2009) has published a template to aid the description of the lifecycle of digital content that could be used to identify how different lifecycles relate to each other. The template moves from an overview of the content to the different lifecycle stages and how they are managed, and prompts consideration of implications for management of the whole lifecycle, including the legal position and the technical tools required. It is intended to act as a guide to those looking to manage digital content and an aid in considering the different needs through the content's lifecycle.

On the legal side, Korn has prepared three briefing papers for the Strategic Content Alliance that address this important topic in managing the digital content lifecycle, examining the IPR aspects from creation through to curation (Pauli, 2009). Regardless of the system or systems being used, it is vital that the rights associated with the content at different stages of its lifecycle are clearly understood so that the content can pass from stage to stage.

[9] LIFE-SHARE project, http://www.leeds.ac.uk/library/projects/lifeshare/
[10] The Make It Digital guides, http://makeit.digitalnz.org/guidelines

The development of appropriate policy to guide digital content lifecycle management is also advised. The Erpanet Digital Preservation Policy Tool (Erpanet, 2003) highlights the advantages of having a policy to guide how often transient digital materials should be dealt with. Coming from a digital preservation perspective it also clarifies that digital preservation does not sit alone and that policy in this area should sit comfortably with other organisational policies and practices to ensure it can be adhered to. A good example of such a policy-driven approach is presented by Columbia University Libraries, whose policy for preservation includes a statement of commitment to lifecycle management[11].

Producing a policy can provide guidance when dealing with new types of digital content that emerge over time. Digital datasets have grown in scope and number in the past few years, an issue raised by Lyon (2007), and have their own specific lifecycle aspects to contend with. Wright et al., (2007), in discussing the connection between digital libraries and eScholarship, identifies the importance and value of data to these communities, and the need to make a connection through the scholarly communication lifecycle, which encompasses both. There is a need to tie in the management of data and its lifecycle with this communication lifecycle in order to get the best value from it. In providing guidance on the management of research datasets, and discussing this management with researchers, Witt and Carlson (2007) at Purdue University highlight lifecycle factors that may affect the management. How does the dataset evolve as it is generated or processed? Who has custody of the data and how does this affect its provenance? What funder demands are there on describing and archiving the dataset for future accessibility and integrity? Not all the answers to these questions are yet clear and the JISC is currently funding a range of projects to explore this area further[12].

As part of understanding what we do with digital content it is valuable to evaluate activity undertaken. Khoo (2006) highlights the benefits of using a digital content lifecycle approach to facilitate the evaluation of digital libraries, allowing the library, or system, to be analysed according to each stage of the lifecycle. In performing such an evaluation Khoo (2007) also builds in the fact that lifecycles can be iterative, and that the lifecycle of one piece of digital content may initiate the lifecycle of another.

Standards

Durga (2007), in a blog posting, describes a potential ECM environment that incorporates a number of different systems used to create and/or capture information for different purposes. So, a web front-end may present a series of forms that data is entered into. The submission of this data may lead to other forms, or it may result in some content being generated and/or delivered (the example given is of an insurance company website providing online quotes). The lifecycle of the content related to the insurance quote request relies on different systems talking to each other, and the author concludes that implementation of technical standards to facilitate this communication is essential. One of the aims of the DCC is to provide information on technical standards that can be used to foster the digital content

[11] Columbia University Libraries Policy for Preservation of Digital Resources, http://www.columbia.edu/cu/lweb/services/preservation/dlpolicy.html
[12] JISC Research Data Management Programme, http://researchdata.jiscinvolve.org/

management Durga describes. The Curation Lifecycle Model has a number of different stages, and the DCC highlights the different standards that might apply or which could be used at each of these stages[13].

Further guidance in the DPC's Handbook includes the recognition that migration of files (between formats and/or systems) is a potential strategy for the preservation part of the lifecycle, though it has its critics (e.g., Rothenburg, 2000). Adherence to standards is highlighted as having the potential to save effort on behalf of those working on later stages of the lifecycle, whilst providing continual access can cause difficulties and increase the effort required if this lifecycle stage is to be periodically re-visited following preservation.

Recording information about data and other types of digital content in a standard way can also facilitate management decisions about sharing to foster ongoing research (Wallis et al., 2007). Working in the ecological field, they observe that many researchers do not share their data. This appears to be because of a lack of standards in way data is organised and made available. If there are no standards, there isn't the ability to easily use other data, therefore no need to request data from others, leading to the conclusion that there is no need to share data, and no need for data standards. This vicious circle can be broken by managing the lifecycle of the data from its creation, so that it can be used at different stages of its lifecycle. In a separate article (Wallis et al., 2008), the authors describe a nine-stage lifecycle for the data that they have identified to help establish standards and facilitate sharing.

Adhering to standards and being open in how digital content is managed can also lead to increased levels of trust. In adopting its policies around digital content lifecycles, OCLC was aiming to be regarded as a trusted digital repository provider (Research Libraries Group, 2002). A key to this was compliance with the Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems, 2002). This itself describes a lifecycle approach, addressing how content can be made ready for ingest to a system, managed within the system, and then made ready for export to another system. Whilst frequently associated with supporting ongoing preservation, the OAIS reference model can also be applied to day-to-day lifecycle management of digital content, prior to active preservation activity taking place.

The knowledge lifecycle

Content lifecycles relate to the management of actual content files throughout their lifecycle. A related concept looks at the management of the lifecycle of the knowledge that is contained or generated by content. A knowledge lifecycle consists of (Sunassee and Sewry, 2002):

1) Create new knowledge
    i) Identify new knowledge
    ii) Identify old and existing knowledge
2) Identify knowledge relevant to organisation
3) Verify selected knowledge

[13] Browse all standards by lifecycle action, http://www.dcc.ac.uk/diffuse/lifecycle/

4) Capture and organise knowledge
5) Disseminate and use knowledge
6) Combine new knowledge and re-evaluate assumptions to create knowledge

Systems to deal with knowledge lifecycles have been investigated (e.g., Schlorlemmer et al., 2002), and it is clear that this is another direction through which digital content lifecycle management has and could evolve. Another view of the knowledge lifecycle is proposed by Millard et al. (2006), which encompasses knowledge acquisition, knowledge modelling, knowledge annotation, knowledge reuse and knowledge maintenance. By using tools to semantically enhance the knowledge the authors propose how the knowledge lifecycle for e-learning might be improved.

Developing the lifecycle

With evidence aplenty that digital content lifecycles exist and are being described, it is of interest to see what may happen next in the evolution of the digital content lifecycle. The InterPARES 2 project (Duranti and Preston, 2008) had built on the previously established view (in InterPARES 1) that a lifecycle approach to managing and preserving records was useful in outlining the different stages of the lifecycle. InterPARES 2 described an extensive number of tasks that are involved throughout the lifecycle, concluding that these different stages could not easily be treated as separate entities. A continuum approach, treating each stage in the context of those around it, was proposed as more valuable in considering how to manage each stage through the lifecycle.

Also addressing the context within which lifecycles sit, the Rutgers University Community Repository (RUCore)[14] has at its heart a data model that tracks the lifecycle and ecology of the data held within it. This recognises that "Digital data lives in relationship to other data and may be repurposed many times over its lifecycle." As such, the repository is a representation of 'living data', and holds information on events and relationships related to the data as part of its overall life.

Lehrmann et al. (2007) tackled the issue of multiple people being involved in the management of learning resources, and also the adaptation or re-authoring of the learning resource as part of its re-use in a different context. In many cases such re-authoring is considered to be the creation of a new object and the start of a new lifecycle for a new and different object. The authors make the point, though, that the original and re-authored resource are connected, and that there is information generated as part of the re-authoring that can be captured ("lifecycle information") that can enrich a user's understanding of both original and re-authored versions. There is an acknowledgement that many systems are not good at capturing this lifecycle information and recognition that where it is captured it often gets stuck at system borders and thus sits in isolation and is not captured in its entirety with the resource.

Two particular types of lifecycle information are proposed:

- Relation information. This encompasses all types of relationship that the content may have, including aggregation relations (links to any aggregation), sequence relations (links to other

[14] RUCore description, http://rucore.libraries.rutgers.edu/how/

resources and the order the links are in) and version or variant relations (links to alternative instances of the resource). Considering the relationship between the original and re-authored resource there might also be reduction or extension relations to indicate if whether the resource has been simplified or expanded.

- Context information. Whereas relation information describes information about two or more resources, context information pertains to a resource in its own right. It can be anything about the resource, but is most likely to be about the resource's use or what is happening to it within the overall lifecycle.

Both these might be considered as metadata about the resource. The authors thus propose the use of an extension to the Learning Object Metadata (LOM) standard under 'category 7: relation' to store this lifecycle information. The metadata associated with the resource can then act as a route for capturing and transferring the lifecycle information as a whole (akin to the role of METS referenced earlier). An architecture is, however, also proposed for a centralised Lifecycle Information System (LIS), which tools involved in the learning resource lifecycle can interact with via capture and access components. The former assist in capturing the lifecycle information, and the latter makes it available in an analysed form for use where appropriate. The authors have implemented an initial version of such a system and indicate a plan to develop more plugins to facilitate interaction between LIS and other tools.

Other aspects of digital content can be considered to have their own lifecycle characteristics. Barton and Robertson (2005) organised a Conceptions of Library and Information Science (CoLIS) workshop on the development of a metadata lifecycle model. Their starting point was a recognition that a conceptual framework to understand how digital repositories and related repository services interacted was absent. The development of such a framework, they proposed, needed to include the object lifecycle and the lifecycle of the metadata associated with the digital objects. At the workshop, Chen and Chen (2005) presented findings from their own development of a metadata lifecycle model, first presented at IFLA in 2003. Their reasoning for the development of such a model was to provide cost effectiveness, quality assurance, consistency and interoperability in their projects.

In a related paper given at the DELOS Digital Repositories: Interoperability and Common Services Workshop (2005), Barton and Robertson expanded on this approach, highlighting that the lifecycle models need to be associated with an ecology of repositories, a model in its own right that lays out the relationships between repositories and surrounding systems or services (the repository's environment). They also propose a methodology for developing a metadata lifecycle model to suit specific needs and highlight how it can help with optimising workflows, e.g., by facilitating the interaction between repositories or between a repository and other systems that adhere to the common model. This ecology of repositories was developed further by the authors and is now available as a resource[15]. Unfortunately, the associated object and metadata lifecycle model approaches proposed by Barton and Robertson have not been developed in the same way.

[15] Ecology of repositories, http://www.ukoln.ac.uk/repositories/digirep/index/Ecology

Conclusions

This literature review has, in many ways, raised a range of different views, opinions and approaches to dealing with digital content lifecycles. These come from different perspectives and starting points, which seems to emphasise that wherever you are coming from consideration and management of digital content lifecycles is important, and even sometimes viewed as necessary, in the variety of environments described. It is perhaps surprising that literature on specific system aspects of this management approach was not found, but this may be due to the flux in technology adoption and the rapid pace of change. Nevertheless, the technology involved must be taken into account when implementing a digital content lifecycle management approach as it is core to the day-to-day running of this. Consideration of the appropriateness of different systems, is useful to inform this, and this can be extended to consideration of appropriate systems at different stages of the lifecycle, the starting point of CLIF.

Specific points emerging from the literature that will be taken into the CLIF project are:

- The importance of consistency in terminology and clear definition of what this terminology refers to.
- There are different types of lifecycle involved in managing digital content, including content lifecycles, user interaction lifecycles, knowledge lifecycles, digitisation lifecycles, metadata lifecycles, etc. These may overlap, they may run in parallel, or one may lead to or initiate another. Since lifecycle terminology is not fixed it is difficult to consider any sort of classification, but it is useful to be aware that any lifecycle approach will not sit in isolation.
- It is valuable to break down the lifecycle into its constituent stages, but that these too should not be treated in isolation but dealt with acknowledging external influences. It is potentially useful to ask the question 'What can be done at this stage to facilitate other stages?', as is the case presented for dealing with digital preservation.
- Lifecycle stages can be long or short depending on the nature of the material, and that this needs to be accounted for when considering system aspects
- Lifecycle information could be potentially valuable, but the implications and complexity of capturing it need to be thought through. Such lifecycle information has to be initially generated, as it is not always immediately available, and then captured and stored in an appropriate fashion to allow it to be used further along the lifecycle. Understanding the purpose of capturing particular lifecycle information should guide practice.
- There appears to be increasing interest in how people fit into the digital content lifecycle, both as actors in it (those who facilitate or carry out different stages) and as stakeholders (those who are affected or influenced by the different stages). This may well move digital lifecycle management to a different footing as roles align themselves with lifecycle stages and potentially offers greater clarity in adopting this approach.

- Understanding digital content lifecycle stages can ease the management of new forms of content, such as data, though acknowledging that each will have its own specific aspects.
- Policy is as important as technology and other factors in embedding a lifecycle approach.
- Moving between lifecycle stages in a technical environment is greatly facilitated by the use of standards, both technical and organisational.
- Understanding of digital content lifecycles is evolving and maturing, though not yet at a stage where they can be easily implemented. There is, though, much available to guide additional work and implementations in new environments.
- Digital preservation needs to be considered at all stages of the digital content lifecycle, rather than only as a concluding step. In particular, a lifecycle approach needs to deal with reiterations of existing material, and determine effective policies and procedures to deal with dynamic and unstable content.

Finally, this literature review has deliberately not addressed where responsibility lies for overseeing or managing digital content lifecycles, as this lies beyond the scope of the CLIF project. Any implied assumptions that can be read into the literature reviewed, for example about the role of a library, are coincidental. It is acknowledged, though, that as digital content lifecycles become better understood in terms of their design and technical implementation there will be a need to place this in the context of how the lifecycles are managed organisationally and see the lifecycles put into practice.

References

All links included were available at the end of February 2010

Anon (2005). *Digitization lifecycle solutions: an integrated approach*. OCLC presentation, ALA Annual Conference 2005, available at http://www.oclc.org/news/events/presentations/2005/digitzn_solutions_0506.ppt

Ayris, P, Davies, R, McLeod, R, Miao, R, Shenton, H and Wheatley, P (2008). *The LIFE2 final project report*, available at http://eprints.ucl.ac.uk/11758/

Barton, J and Robertson, RJ (2005). Developing a metadata lifecycle model. In: *Lecture Notes in Computer Science*, eds. Crestani, F and Ruthven, I, 3507: 249-250

Bates, M, Manuel, S, Loddington, S and Oppenheim, C (2006). *Rights and Rewards Project – Digital lifecycles and file types: final report*, University of Loughborough, available at http://rightsandrewards.lboro.ac.uk/index.php?section=21

Batley, S (2007). The I in information architecture: the challenge of content management. *Aslib Proceedings: New Information Perspectives*, 59 (2): 139-151

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008). *Interim Report*, available at http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf

Boiko, B (2001). Understanding content management. *Bulletin of the American Society for Information Science and Technology*, 28 (1): 8-13

Brace, J (2008). Versioning in repositories: implementing best practice. *Ariadne*, Issue 56, available at http://www.ariadne.ac.uk/issue56/brace/

Bruce, H, Jones, W and Dumais, S. (2004). Information behaviour that keeps found things found, *Information Research*, 10 (1), available at: http://InformationR.net/ir/10-1/paper207.html

Consultative Committee for Space Data Systems (2002). *Reference model for an Open Archival Information System (OAIS),* CCSDS 650.0-B-1 Blue Book, available at http://public.ccsds.org/publications/archive/650x0b1.pdf

Chen, Y-N and Chen, S-J (2005). *Metadata lifecycle model and metadata interoperability*. Presentation at the CoLIS 5 workshop, Glasgow, UK

Chieu, TC, Zeng, L and Mohindra, A (2008). An extensible enterprise content management system with Service Component Architecture. In: *IEEE International Conference on Service Operations and Logistics and Informatics*, 1131-1137, doi:10.1109/SOLI.2008.4686569

Deegen, M (2001). Management of the life cycle of digital library materials. *Liber Quarterly*, 11: 400-409

Digital Content Life Cycle Management Team (2008). *Final report*. University of Illinois Libraries, available at http://www.library.illinois.edu/nsm/digcon/digitalcontentcharge.html

Digital Preservation Coalition (2008). *Preservation management of digital materials: the handbook*, available at http://www.dpconline.org/graphics/handbook (in particular note chapter 4. Organisational Activities)

Duranti, L and Preston, R, Eds. (2008). *International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records*. Associazione Nazionale Archivistica Italiana, Padova, Italy, available at http://www.interpares.org/ip2/book.cfm

Durga, A (2007). Standards and the content lifecycle. In *Portals & Content Management: News, Views & Analysis Blog*, available at http://it.toolbox.com/blogs/pcm/standards-and-content-lifecycle-18174

Erpanet (2003). *Digital preservation policy tool*. Erpa guidance document, available at http://www.erpanet.org/guidance/index.php#policy

European Taskforce on Permanent Access (2006). *Permanent access to the records of science: strategic action programme 2006-2010*, available at http://www.alliancepermanentaccess.eu/Strategic%20Action%20Programme.pdf

Granger, S (2002). *Preservation of e-learning materials and cost models for digital preservation*. Meeting report, DPC website, available at http://www.dpconline.org/events/e-learning-materials-and-cost-models.html

Higgins, S (2007). Draft DCC curation lifecycle model. *The International Journal of Digital Curation*, 2 (2): 82-87, available at http://www.ijdc.net/index.php/ijdc/issue/view/3

Higgins, S (2008). The DCC curation lifecycle model. *The International Journal of Digital Curation*, 3 (1): 134-140, available at http://www.ijdc.net/index.php/ijdc/issue/view/4

Hockx-Yu, H (2006). Digital preservation in the context of institutional repositories. *Program: electronic library and information systems*, 40 (3): 232-243, doi:10.1108/00330330610681312

Humphrey, C (2006). *E-Science and the life cycle of research*, paper available at http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc

Ioannidis, Y, Maier, D, Abiteboul, S, Buneman, P, Davidson, S, Fox, E, Halevy, A, Knoblock, C, Rabitti, F, Schek, H and Weikum, G (2005). Digital library information-technology infrastructures. *International Journal on Digital Libraries,* 5: 266-274, doi:10.1007/s00799-004-0094-8

Irshad, T and Ure, J (2009). *Clinical data from home to health centre: the Telehealth curation lifecycle.* SCARP project case study no. 3, available at http://www.dcc.ac.uk/scarp/

Khoo, M (2006). A sociotechnical framework for evaluating a large-scale distributed educational digital library. In: *Lecture Notes in Computer Science*, ed. Gonzalo, J, 4172: 449-452

Khoo, M (2007). Evaluating digital libraries: a sociotechnical approach. *Proceedings of the American Society for Information Science and Technology*, 43 (1): 1-9, doi:10.1002/meet.14504301236

Lehrmann, L, Hildebrandt, T, Rensing, C and Steinmetz, R (2007). Capturing, management and utilization of lifecycle information for learning resources. In: *Lecture Notes in Computer Science*, eds. Duval, E, Klammer, R and Wolpers, M, 4753: 187-201

Long, P, Green, A and Nitecki, D (2003). *The campus digital landscape through the digital lifecycle*. Presentation from Yale University to Common Solutions Group meeting, University of Texas, Austin, available at http://www.stonesoup.org/meetings/0301/library.pres/

Lyon, L (2003). eBank UK: building the links between research data, scholarly communication and learning. *Ariadne*, Issue 36, available at http://www.ariadne.ac.uk/issue36/lyon/

Lyon, L (2007). *Dealing with data: roles, rights, responsibilities and relationships: consultancy report*. UKOLN report, University of Bath, Bath, available at http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

McDonough, JP (2006). METS: standardized encoding for digital library objects. *International Journal on Digital Libraries*, 6 (2): 148-158, doi:10.1007/s00799-005-0132-1

McGovern, NY and Stuchell, L (2009). *Digital content review – lifecycle report outline*. ICPSR paper, available at http://www.icpsr.umich.edu/files/ICPSR/curation/DigitalContentReviewTemplate.pdf

McLeod, R, Wheatley, P and Ayris, P (2006). *Lifecycle information for e-literature: full report from the LIFE project*, available at http://eprints.ucl.ac.uk/1854/

Millard, DE, Tao, F, Doody, K, Woukeu, A and Davis, HC (2006). The knowledge life cycle for e-learning. *International Journal of Continuing Engineering Education and Lifelong Learning*, 16 (1/2): 110-121

Parker, E (2000). Study of the records lifecycle. *New Review of Academic Librarianship*, 6 (1): 73-123, doi:10.1080/13614530009516803

Pauli, M (2009). Intellectual property rights at a glance. In *JISC SCA Blog*, available at http://sca.jiscinvolve.org/2009/10/13/ipr-at-a-glance/

Pennock, M (2006). *Digital Curation Manual – instalment on "Curating e-mails: a life-cycle approach to the management and preservation of e-mail messages",* Digital Curation Centre, available at http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-e-mails/

Pennock, M (2007). Digital curation: a life-cycle approach to managing and preserving usable digital information. *Library & Archives Journal*, Issue 1, e-print, available at http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf

Razum, M, Schwichtenburg, F and Fridman, R (2007). Versioning of digital objects in a Fedora-based repository, paper given at German e-Science Conference, available at http://www.ges2007.de/papers/

Research Libraries Group (2002). *Trusted digital repositories: attributes and responsibilities*. An RLG-OCLC report, RLG, Mountain View, California, USA, available at http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf

Robertson, RJ and Barton, J (2005). *Optimising metadata workflows in a distributed environment*. Paper given at DELOS Digital Repositories: interoperability and common services workshop, Heraklion, Crete, available at http://delos-wp5.ukoln.ac.uk/forums/dig-rep-workshop/intro.html

Ross, S and Hedstrom, M (2005). Preservation research and sustainable digital libraries. *International Journal on Digital Libraries*, 5: 317-324, doi:10.1007/s00799-004-0099-3

Rothenberg, J (2000). An experiment in using emulation to preserve digital publications: a report commissioned by the Koninklijke Bibliotheek (KB), available at http://nedlib.kb.nl/results/emulationpreservationreport.pdf

Rusbridge, C (2005). *Information life cycle and curation*. DCC presentation, available at www.dcc.ac.uk/docs/dcc-life-cycle.ppt

Schorlemmer, M, Potter, S, Robertson, D and Sleeman, D (2002). Knowledge life cycle management over a distributed architecture. *Expert Update*, 5 (3): 2-19, eprint available at http://eprints.aktors.org/135/

Sellen, AJ and Harper, R (2002). *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA

Shenton, H (2005). *Life-cycle costing*. Presentation at Digital Curation and Preservation: Defining the Research Agenda for the Next Decade, available at http://www.dcc.ac.uk/events/warwick_2005/DCPWarwick_DLM_HelenShenton_71105.ppt

Sunassee, S and Sewry, DA (2002). A Theoretical Framework for Knowledge Management Implementation. *ACM International Conference Proceeding Series*: Vol. 30. Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology, Port Elizabeth, South Africa, 235 – 245

Tredinnick, L (2005). *Why Intranets Fail (and How to Fix Them).* Chandos, Oxford, UK

Trefethen, A (2005). *Data Life Cycle Management Strand welcome*. Presentation at Digital Curation and Preservation: Defining the Research Agenda for the Next Decade, available at http://www.dcc.ac.uk/events/warwick_2005/DCP_DLM_Breakout_session.ppt

Tzitzikas, Y (2007). Dependency management for the preservation of digital information. In: *Lecture Notes in Computer Science*, eds. Wagner, R, Revell, N and Pernul, G, 4653: 582-592

UK Research Data Service Feasibility Study (2009). *Newsletter for stakeholders November 2009*, available at http://www.ukrds.ac.uk/resources/

Wallis, JC, Borgman, CL, Mayernik, MS and Pepe, A (2007). *Moving archival practices upstream: exploration of the life cycle of ecological sensing data in collaborative field research*. Presentation at 3rd International Digital Curation Conference, Washington, USA, available at http://www.dcc.ac.uk/events/dcc-2007/programme/

Wallis, JC, Borgman, CL, Mayernik, MS and Pepe, A (2008). Moving archival practices upstream: exploration of the life cycle of ecological sensing data in collaborative field research. *The International Journal of Digital Curation*, 3 (1): 114-126, available at http://www.ijdc.net/index.php/ijdc/issue/view/4

Williams, P, John, JL and Rowland, I (2009). The personal curation of digital objects: a lifecycle approach. *Aslib Proceedings: New Information Perspectives*, 61 (4): 340-363, doi:10.1108/00012530910973767

Witt, M and Carlson, JR (2007). *Conducting a data interview*. Purdue Libraries Research Publication, Purdue University, available at http://docs.lib.purdue.edu/lib_research/81/

Wright, M, Sumner, T, Moore, R and Koch, T (2007). Connecting digital libraries to eScience: the future of scientific scholarship. *International Journal on Digital Libraries*, 7: 1-4, doi:10.1007/s00799-007-0030-9

Wu, YD and Liu, M (2001). Content management and the future of academic libraries. *The Electronic Library*, 19 (6): 432-439