CLIF: Moving repositories upstream in the content lifecycle

Simon Waddington, Richard Green, Chris Awre[1]

Abstract

The UK JISC-funded Content Lifecycle Integration Framework (CLIF) project has explored the management of digital content throughout its lifecycle from creation through to preservation or disposal. Whilst many individual systems offer the capability of carrying out lifecycle stages to varying degrees, CLIF recognised that only by facilitating the movement of content between systems could the full lifecycle take advantage of systems specifically geared towards different stages of the digital lifecycle. The project has also placed the digital repository at the heart of this movement and has explored this through carrying out integrations between Fedora and Sakai, and Fedora and SharePoint. This article will describe these integrations in the context of lifecycle management and highlight the issues discovered in enabling the smooth movement of content as required.

Keywords: repositories; integration; digital preservation; content lifecycle

Introduction

At the heart of meeting institutional needs for managing digital content is the need to understand the different activities that the content goes through, from planning and creation through to disposal or preservation. Digital content is created using a variety of authoring tools. Once created the content is often stored somewhere different, made accessible in possibly more than one way, altered as required, and then moved for deletion or preservation at an appropriate point. Different systems can be involved at different stages: one of these may be a repository. To embed repositories in the content lifecycle, and prevent them becoming yet another content silo within the institution, they thus need to be integrated with other systems that support other parts of this lifecycle. In this way the content can be moved between systems as required, minimising the constraints of any one system.

The CLIF (Content Lifecycle Integration Framework) Project was a two-year joint venture between Library and Learning Innovation (LLI)[2] at the University of Hull and the Centre for e-Research (CeRch)[3] at King's College, London. Funded by a grant from the Joint Information Systems Committee (JISC)[4] in the UK, the work was completed in the spring of 2011. The project set out to explore the digital content lifecycle in the context of users' work with two content management systems used predominantly in different parts of a University, Microsoft SharePoint[5] and the

---

[1] Simon Waddington is Software Development Manager at the Centre for e-Research, King's College, London; Richard Green is a consultant working with the Library and Learning Innovation (LLI) repositories team at the University of Hull; Chris Awre is Head of Information Management within LLI at the University of Hull

[2] Library and Learning Innovation, University of Hull, http://www2.hull.ac.uk/lli/

[3] Centre for e-Research, King's College London, http://www.kcl.ac.uk/innovation/groups/cerch/index.aspx

[4] JISC Repository enhancement and start-up programme strand, http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2.aspx

[5] Microsoft SharePoint, http://sharepoint.microsoft.com/en-us/Pages/default.aspx

learning management system/virtual learning environment Sakai[6], and to investigate how these could be made to interoperate with the Fedora Commons repository application[7]. Each of these systems provides management of digital content yet they approach this from different perspectives and fulfil different purposes at different parts in the lifecycle; how might some form of integration allow users to manage their content more effectively? Sakai, Microsoft SharePoint and Fedora Commons were chosen as being of particular interest on the campuses of the two partner institutions.

The project commenced with a literature review (Awre 2010) focusing on the digital content lifecycle and how this might be applied across different systems. In parallel, the project team worked with creators of digital content at the two partner universities to develop an understanding of how they might like to deal with authoring, collaboration, delivery and potential preservation of their work in a world where transfer between SharePoint and Fedora, one the one hand, and Sakai and Fedora on the other was possible. At the University of Hull several of the staff interviewed were already familiar with the idea of a repository being part of a workflow having previously contributed to the work of the RepoMMan[8] and REMAP[9] projects which explored this area. (Green, Dolphin, Awre and Sherratt 2007; Green and Awre 2008; Green and Awre 2009)

Once this initial phase of the project was complete, the project team set out to design and produce open source, and where possible standards-based, software that would allow this transfer of content between systems in order to meet the lifecycle requirements expressed by the potential users. Following an initial technical review, the software was designed in such a way that it appears to be a natural extension of Sakai and SharePoint and thus to allow maximum flexibility for users to transfer content between the systems at what they consider to be their point of need.

Managing the Lifecycle

Pennock (2007), in research carried out at UKOLN, identified the following reasons why management of the lifecycle of digital content is necessary.

- 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."

Long (2003), in a presentation to the Common Solutions Group, describes the benefit of using a lifecycle approach to help manage the digital landscape: different content types have different lifespans, and thus are best managed through understanding these lifespans. Rusbridge (2005) highlights that there are usually multiple lifecycles acting in tandem, and these need to be managed in relation to each other. Digital preservation investigations (notably the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008) have suggested that effective preservation needs

[6] Sakai, http://sakaiproject.org/
[7] Fedora Commons, http://www.fedora-commons.org/
[8] RepoMMan project, http://www.hull.ac.uk/esig/repomman/
[9] REMAP project, http://www2.hull.ac.uk/discover/remap.aspx

to take account of the lifecycle of the content: preservation actions can be carried out at all stages of a lifecycle. Ross and Hedstrom (2005) flip this round and emphasise the benefit of digital preservation taking place "in the context of a lifecycle."

Wu and Liu (2001) had proposed an early model of how systems can support the management of the digital content lifecycle, enabling stages within this to be automated using a calendaring solution. Deegen (2001) highlights that digital content needs more careful attention in order to get the most out of it, and that using a lifecycle approach and implementing this within appropriate systems can allow the content to be used and managed as best it can. 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. Tzitzikas (2007) describes the ubiquity of dependency in systems in managing the lifecycle, whilst Ioannidis et al. (2005), identify the stages of user interaction with systems that can be applied at each part of the lifecycle. The limitation of using only a single system was highlighted by the Rights and Rewards project at Loughborough University (Bates et al., 2006).

Away from the specifically technical aspects of managing the digital lifecycle, it is important to be clear about the rights associated with content at each stage of the lifecycle (Pauli 2009) and having effective policy for the management of content throughout its lifecycle is also advised (Erpanet 2003). The use of standards throughout the lifecycle management of content is very much also supported, particularly by the Digital Curation Centre[10].

Case studies

The CLIF project studied the content lifecycle for different types of content as they are used for the purposes of research, teaching and administration in higher education. In particular, we were interested in determining requirements for moving content across different systems within the institution and the extent to which the repository could provide a staging post for content at intermediate stages of the lifecycle.

Interviews were conducted with staff at the University of Hull and King's College London to investigate their current and future information management requirements. The staff interviewed included department heads, records managers, archivists, academic researchers in environmental science, biophysics and war studies, and lecturers in engineering.

These were distilled into a set of requirements and exemplary scenarios that were used as the basis for the technical integration work. Sample scenarios reflecting uses cases for CLIF covering the areas of teaching, research and administration are described in the following subsections.

Examinations and tests

The management of content relating to examinations and online tests is a ubiquitous scenario for higher education institutions. Examination papers are typically stored and prepared in systems such

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

as SharePoint. Exam papers undergo a review process, which is currently carried out via email with the moderators, but could in the future be implemented as a SharePoint approval workflow. The exam materials themselves represent an output, for which the institutional repository is used as the final store for access and preservation. The deposit can be automated by linking it to the approval workflow. The exam materials are initially deposited in a folder with restricted access. Following the exam, the materials are made public as a resource for students studying for exams in subsequent years.

Exams and tests that are performed or made available online are moved to the learning management system (Sakai) for completion by the students. Student marks are again retained for several years in order to answer requests for references by employers and potential appeals by the candidates. Hence the marks and student solutions are retained in a private area in the repository. Finally, administrators and teaching staff need to have access to the exams to perform analysis, which is typically performed in SharePoint. SharePoint provides convenient tools for sharing documents and controlling access, as well as Excel spreadsheet integration. The spreadsheets and associated materials are similarly deposited in the repository for future reference.

Research papers and datasets

Many researchers use collaboration systems such as SharePoint or Sakai for preparation of academic papers, as they enable editing by multiple authors, access management and version control. Once a publication or preprint is completed, it can be deposited in the public institutional repository. SharePoint in particular is also suitable for storing materials such as raw experimental data sets and image collections that need to be shared across research teams and are too large to be sent by email. Many funding councils and journals are demanding that research data associated to publications should be retained for 10 years or more, and increasingly also demanding that the datasets be shareable. Datasets required only for internal use or associated to unpublished work can be retained in repository folders with restricted access. When journal papers appear, the associated data can be deposited in public repository folders. Approval by heads of department or research groups is sometimes required before a paper can be moved into the public domain.

Providing the capability for researchers to search and retrieve documents from repositories is essential to support their research. Researchers require the ability to browse specific collections, search for documents by fields such as title and author as well as carrying out keyword searches on the full contents of textual documents. For users of SharePoint and Sakai, providing repository integration enables direct import of the documents into their workspace.

Administrative documents

Preparation of policy documents is typically a task performed by several staff, for which a collaboration system such as SharePoint or Sakai is suitable. In particular, version control is required to manage the editing process, and approval workflows to obtain sign-off of the resulting outputs. Many administrators currently use email for exchanging live documents, resulting in a large number of versions in circulation that are on occasions being worked on simultaneously. Approved policy documents are deposited as a matter of procedure to a public area of the institutional repository, from where they can be referenced by other systems and end users. Policy documents are periodically updated, requiring retrieval to another system where they can be edited before re-submission to the repository.

Technical review and design

Development of the software for CLIF started with a technical review (Awre, Green, Thompson and Waddington 2010). The purpose of this work was to understand the rich and often complex functionality available to us in the various systems under investigation, how and where content transfer between them might best be achieved to support users' needs, and to determine what standards might be used effectively in the process.

Selection of systems

The systems selected for investigation with the CLIF project were chosen because of their role in the management of digital content from different perspectives within a University.

- Sakai needs to hold content to inform learning activities, and enable remote teaching. It can also be applied in research space and provide access to research materials.
- Microsoft SharePoint is a generic platform that can be applied to the development and sharing of content. It is often associated with the management of administrative functions and the workflow associated with these, though there is use of the system in teaching and research contexts
- Fedora provides a repository platform that can be applied to any digital content management use case required.

Each of the three applications is of mature design and thus offered a number of versions as starting points: some proven and stable, others somewhat newer. In the event, and notwithstanding the fact that a new release of SharePoint emerged during the course of the project (SharePoint 2010), the team focused its efforts around the following, and developed our integration software for these accordingly:

- Sakai 2.6.1
- SharePoint 2007
- Fedora 3.4

Fedora and Sakai have well-defined web service APIs that were used in support of the CLIF Project's code. In the case of SharePoint the interaction was achieved by using a C# middleware layer, called Hydranet, to communicate with the Fedora APIs. Whereas the APIs and interfaces for all three systems were available to allow for systems integration, it became swiftly apparent that applying these was not as straightforward as might have been hoped. There have been a number of initiatives over the years to identify a common API for integrating content management systems. In the library world Z39.50[11] had this as one of its original foci by providing an abstracted interface through which different systems could talk to one another. Open Service Interface Definitions (OSIDs)[12] also aimed to deliver this abstraction for learning systems, and Java community processes have also tackled the issue through development of the JSR 170 specification[13] (and subsequent versions, JSR-283 and JSR-333). The main issue for all of these has been the relative low level of adoption, certainly by the systems under investigation in CLIF, but also within the University sector.

---

[11] Z39.50, http://www.loc.gov/z3950/agency/

[12] OSIDs, http://en.wikipedia.org/wiki/Open_Service_Interface_Definitions

[13] JSR content repository specifications, http://www.jcp.org/en/jsr/detail?id=170 (change the number at the end to 283 or 333 for subsequent versions)

A current initiative attracting a high level of interest and buy-in is CMIS (Content Management Interoperability Services)[14]. Whilst not available to the CLIF project, the advent of this latest specification to facilitate effective API links between systems will be of interest for the future.

Shaping the content within the systems

In attempting to produce tools that might be of broader applicability than just the partner institutions, the team were faced with the problem of structuring the digital content being used across the systems. This is of particular relevance when using Fedora where it is important to have a clear understanding of this structure so as to properly build the digital objects within it. The Fedora repository software provides an architecture with which myriad content structures can be created and within them digital objects conforming to any number of patterns. How, then, can digital objects be produced in a way that might have broad acceptance? The team was fortunate to be operating in parallel with developments in the Hydra project[15] , and chose to adopt the data modelling recommendations from this[16]. The work of the Hydra project has attracted international interest and many universities (and others) worldwide have been supportive of the guidelines that it has produced for building digital objects within Fedora. Through adopting this approach we thus set out to enable Sakai and SharePoint to create Hydra-compliant objects.

Development process

This review and design stage valuably informed work to develop code that would enable the two integrations: Sakai with Fedora and SharePoint with Fedora. Each site (Hull and King's) concentrated primarily on one of the integrations and code was then shared through a common SVN for review and comment where this was appropriate. Each site also then used the other's developer to test what was written. Thus it was that the majority of the work on Sakai was done in Hull (where Sakai is the institutional virtual learning environment) and King's took the lead on SharePoint development, building on local institutional SharePoint developments. The detail of the development work, the outcomes and the installation procedures are documented at length in the technical appendix to the CLIF Project's Final Report (Awre, Green, Thampi, Thompson and Waddington 2011). All code produced by the project has been made available via a github site[17] for others to work with and, hopefully, develop further.

Fedora-Sakai integration

Sakai describes itself as a "collaboration and learning environment"[18]. In the UK we would describe it as a 'virtual learning environment' whereas colleagues in the US would more likely use the term 'learning management system'. Sakai is organised around the provision of sites: each module may have its own site, but organisational units such as departments, or teams or individuals within these may also have their own site, all with appropriate membership. Part of the functionality available to Sakai users, within the sites of which they are a member, is a resources area where they can store digital materials of their own and potentially share them across sites with other users of their Sakai

[14] CMIS, http://www.oasis-open.org/committees/cmis/
[15] Hydra project, http://projecthydra.org
[16] Hydra objects, content models and disseminators, https://wiki.duraspace.org/display/hydra/Hydra+objects%2C+content+models+%28cModels%29+and+disseminators
[17] github site for University of Hull, https://github.com/uohull
[18] Sakai Collaboration and Learning Environment (CLE), http://sakaiproject.org/node/2260

installation. Users are able to organise their materials in a conventional tree structure and the system provides them with the expected management functions: upload, copy, edit, move, delete and so on. These functions can be applied to individual files, to whole folders or to sections of the tree. Shared resources from other sites can similarly be represented in the user's resources area. Building on work started in the JISC CTREP project[19] at University of Highlands and Islands, the CLIF integration work has allowed us to portray a Fedora repository as one such 'other site' through a specific application of the ContentHostHandler, and to allow appropriate Sakai management functionality around its content.

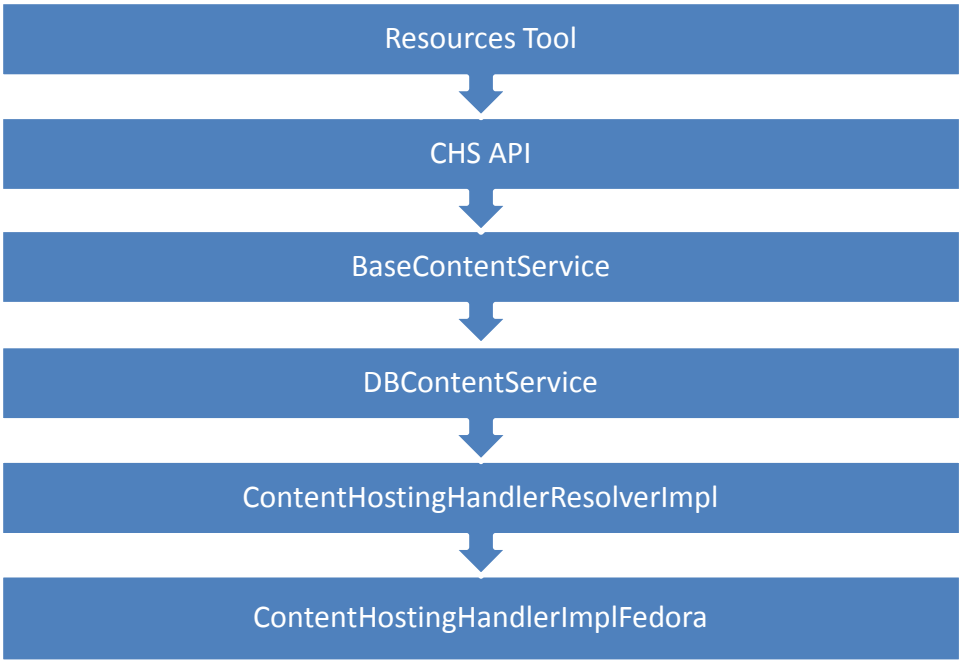

Figure 1: Diagram of Sakai Content Hosting Handler model

From a user's perspective the repository looks and behaves like any other Sakai site and is represented to them using the familiar tree paradigm. Crucially, the standard Sakai management functionality allows them to copy and move digital content in either direction between their workspace and an 'other site' and thus between their workspace and a Fedora repository so represented.

The flexibility of this approach means that the CLIF Sakai integration software allows the user to choose if and when they should transfer materials between the two systems according to lifecycle requirements. Sakai users are not necessarily able to interact with all the content of the repository: a configuration file specifies an access point in the repository hierarchical structure below which they have access: thus, interaction may be focused around a collection of open educational resources within the repository. By transferring it into repository folders digital content is potentially shared outside the Sakai environment and may be in a better location for medium- or long-term preservation.

[19] CTREP project, http://www.jisc.ac.uk/whatwedo/programmes/reppres/sue/tetracam.aspx

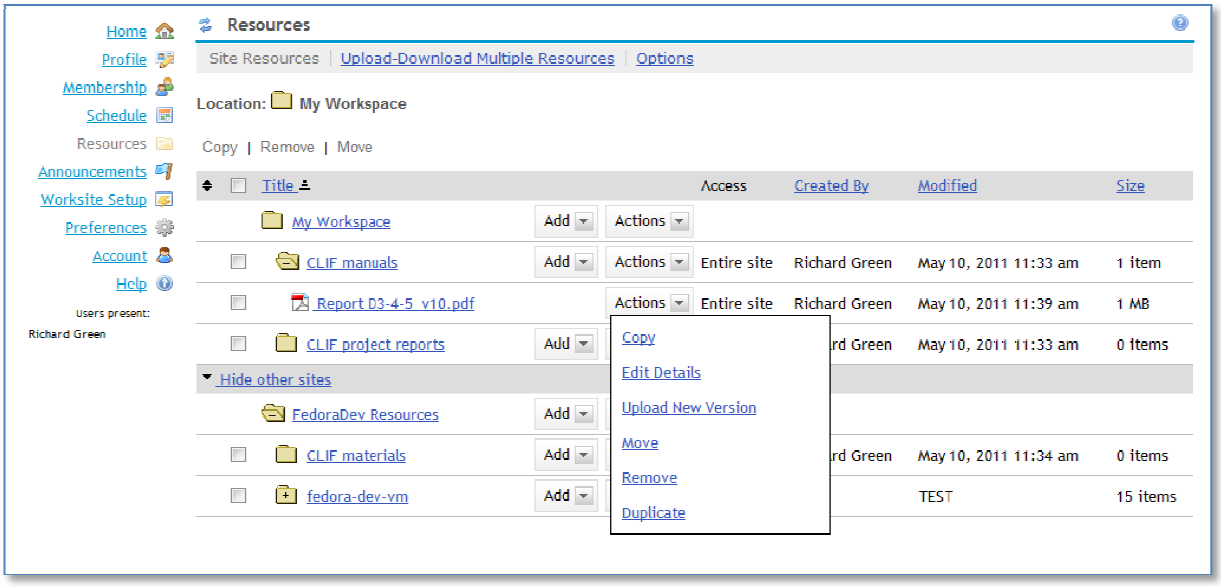

Figure 2: Sakai resources screen showing local and repository resources ready for copy/paste

When digital content is transferred to the repository from Sakai the CLIF software creates a Fedora repository object around it. In creating these objects we attempt to capture any descriptive metadata that Sakai may contain about the content and store it in the repository as MODS. It is regrettable that Sakai does not currently provide support to bring such metadata back into its environment when transfer occurs in the reverse direction.

The code produced during the CLIF project assumes that a Sakai user has full read/write permissions in the area of the Fedora repository with which they can interact. This may not be an appropriate approach in some institutions; indeed it is not suited to the way that the University of Hull operates its institutional repository. Post project, additional work has been undertaken to produce a version of the Sakai integration code which allows users to browse (read-only) all the materials in the area of the repository under Sakai's access point but write only into one part of that tree. In Hull, all materials deposited into our repository pass through a quality assurance stage and this writeable area will correspond to the quality assurance queue for learning and teaching materials. The team managing this queue will be able to supplement the limited descriptive metadata coming from Sakai before exposing the new materials to appropriate repository users.

Lifecycle management

Those testing this integration software had no difficulty understanding how it should be used, as it follows, deliberately, the recognised Sakai paradigm for accessing resources. The simplicity of operation perhaps obscures the complexity of the underlying processes and the work that was undertaken to achieve them, though the seamless integration resulting from this was much appreciated. The ability to browse the Fedora repository from within Sakai was considered helpful as a way of identifying and using materials not already within the Sakai environment. It was emphasised that the most valuable way of accessing content was in context, with links to wherever it was, so taking users to the repository may be useful on occasion rather than seeing it through Sakai. In discussing what movement of content there would be between the two, a separation between throwaway or repeatable content and reference or irregular use content was made, with the repository playing a greater role for the latter. In such cases, copying or moving content needed

to be undertaken with a clear understanding about storage, to avoid duplication. A key piece of Sakai functionality, to upload a new version, was considered to be helpful in allowing management of different versions. Where content was being referenced in Sakai, there seemed little point moving it into Sakai when it could be referenced where it was. Moving content into the repository was most obviously useful when archiving it, and it was considered to be particularly valuable if many objects could be moved in one go (resulting in bulk ingest to the repository). In general, the ability to move around lots of content quickly was considered essential to save time.

Fedora-SharePoint integration

The SharePoint MySite template was chosen as the base site template for the implementation of the CLIF integration work. As a dedicated personal site, MySite provides users with a single location to manage all of their documents (through a document library), other content, and tasks, as well as their calendar and contacts. It enables users to create their own workspaces and share selected personal information, content and documents with other users. Privacy groups also allow users to specify permissions to access information on shared pages. MySite thus offered a logical place to facilitate interaction with a repository for personal content lifecycle management. However, much of the CLIF functionality is independent of the site template chosen and could equally well be deployed on other SharePoint site templates.

Overall, the SharePoint Fedora integration enables deposit to Fedora and retrieval of documents from Fedora via the SharePoint user interface. When moving documents from Fedora, metadata is appended to the Fedora object to enable subsequent identification, search and retrieval of the content. When documents are retrieved from Fedora to SharePoint, only the content payload is returned. The CLIF SharePoint integration software enables documents to be transferred to private repository locations that have restricted access as well as public repository folders that are open access.

Deposit of individual documents to a private repository folder can be performed in one of two ways. *Copy to Repository* adds a Fedora object containing the document and associated metadata to the repository, leaving the SharePoint document in situ. *Move to Repository* performs a similar operation, except that the document is removed from the SharePoint document library. In order to enable rapid retrieval of the document to SharePoint, an Archive list is provided in the SharePoint document library that contains a hyperlink to the location of the document in Fedora.

Publish to Repository enables deposit of a document and metadata to a public repository folder. The list of publishable repository locations is configured by an administrator and is made visible to a MySite user via a pull-down menu. Publish to Repository also provides the option to initiate an approval workflow, taking advantage of SharePoint's capabilities in this area. Once submitted by the user, an approval task is created for the approver and the task appears on the MySite homepage of the approver. Additionally, if connection to an email server has been configured, the approver can receive a notification by email of the pending task. The approver has the option to approve or reject the document as well as providing comments in a text box for the submitter. Once the review task has been completed, the submitter can review the task status via their MySite homepage, possibly also after email notification.

Deposit of individual documents can be selected from the pull-down menu associated to a document in a document library as illustrated in Figure *. Custom columns have been added to the document library. Project Title and Content Subject are entered by the user in the document upload form. Publishable Status indicates whether documents have been published or copied to a private repository folder. Document Approval Workflow indicates the current status of the publication approval workflow.

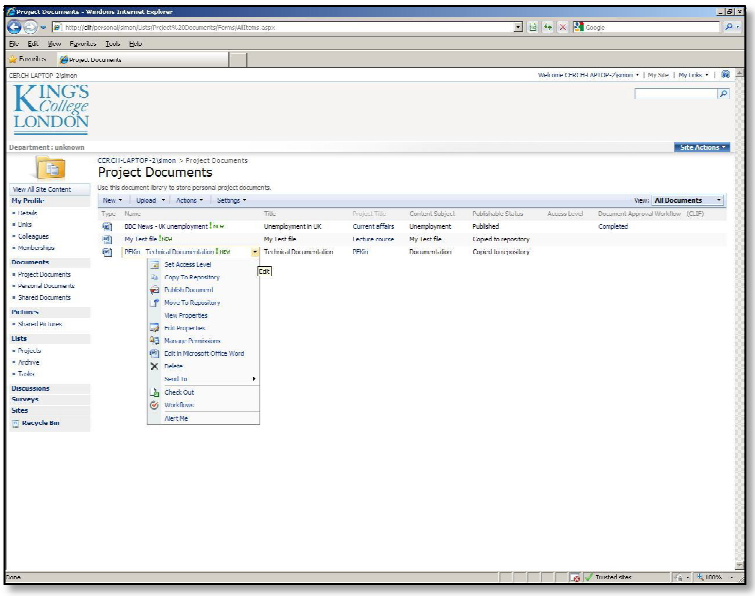

Figure 3: Pop-up menu with options for depositing item to repository.

The features Copy to Repository and Move to Repository enable the user to browse to specific locations in the repository. The repository browse form is illustrated in Figure *. A set of default folders can be configured by the administrator, which is named _private and _archive in this example. In addition, the user can create new folders.

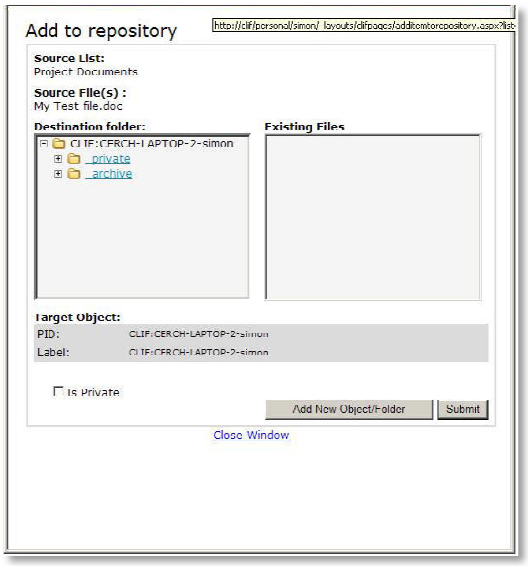

Figure 4: Browse to private repository location

In order to support deposit of multiple documents to the repository, bulk move and copy operations are supported. These are named *Copy Multiple Files* and *Move Multiple Files* (the copy and move options retaining the file or a link to the repository as for single files). These bulk operations can be selected from the Site Actions tab of the document library as illustrated in Figure *.

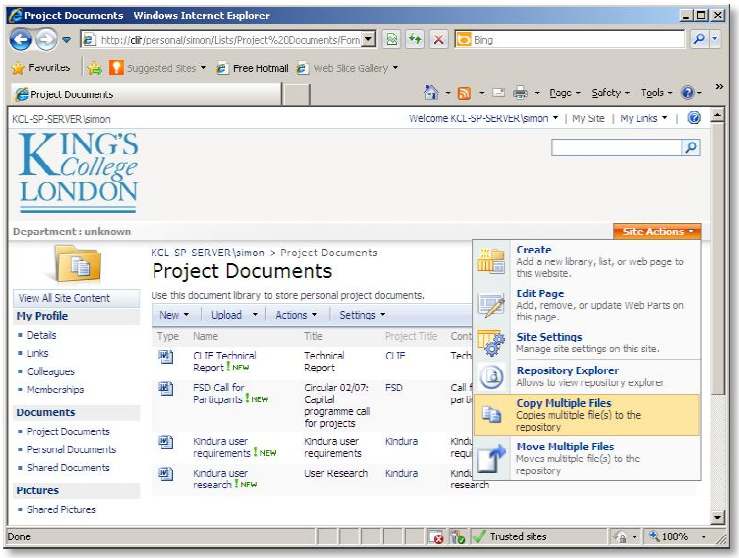

Figure 5: Deposit of multiple documents to the repository

As noted earlier, when using the Move option for single or multiple files the files themselves are not retained in SharePoint. However, a link is. Such documents can be retrieved by navigating to the Archive list from the left side bar of the document library, selecting the pop-up menu from the required document and clicking on Retrieve from Repository as illustrated in Figure *.

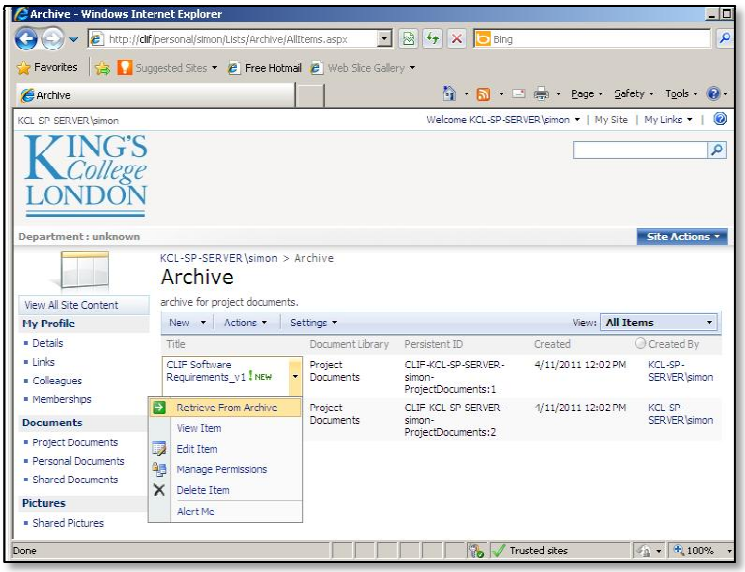

Figure 6: Retrieval of moved documents from archive

Alongside this route for retrieval of documents, a repository browser enables more general browsing and retrieval of documents in Fedora. The browser can be used to retrieve documents from both public and private repository folders. The Repository Explorer option is selectable from the Site

Actions tab of the MySite document library as illustrated in Figure *. The user has the option to select the root of the repository folder tree they wish to browse. The options to browse from the MySite root folder and the Publishable Locations are standard for every user. If the user has administrative rights, they can also navigate the CLIF root folder.

In summary, the SharePoint-Fedora integration is based on a set of 34 features that can be deployed as a complete solution. Individual features can be activated or deactivated by a MySite administrator in order to customise the functionality available to MySite users. The CLIF.Solutions project contains the feature definitions. All the C# code is contained in separate projects. The creation of Fedora XACML policies to manage Fedora access permissions is contained in the PolicyManagement project. The Hydranet project contains methods that call the Fedora API-M SOAP services, and perform Fedora object creation.

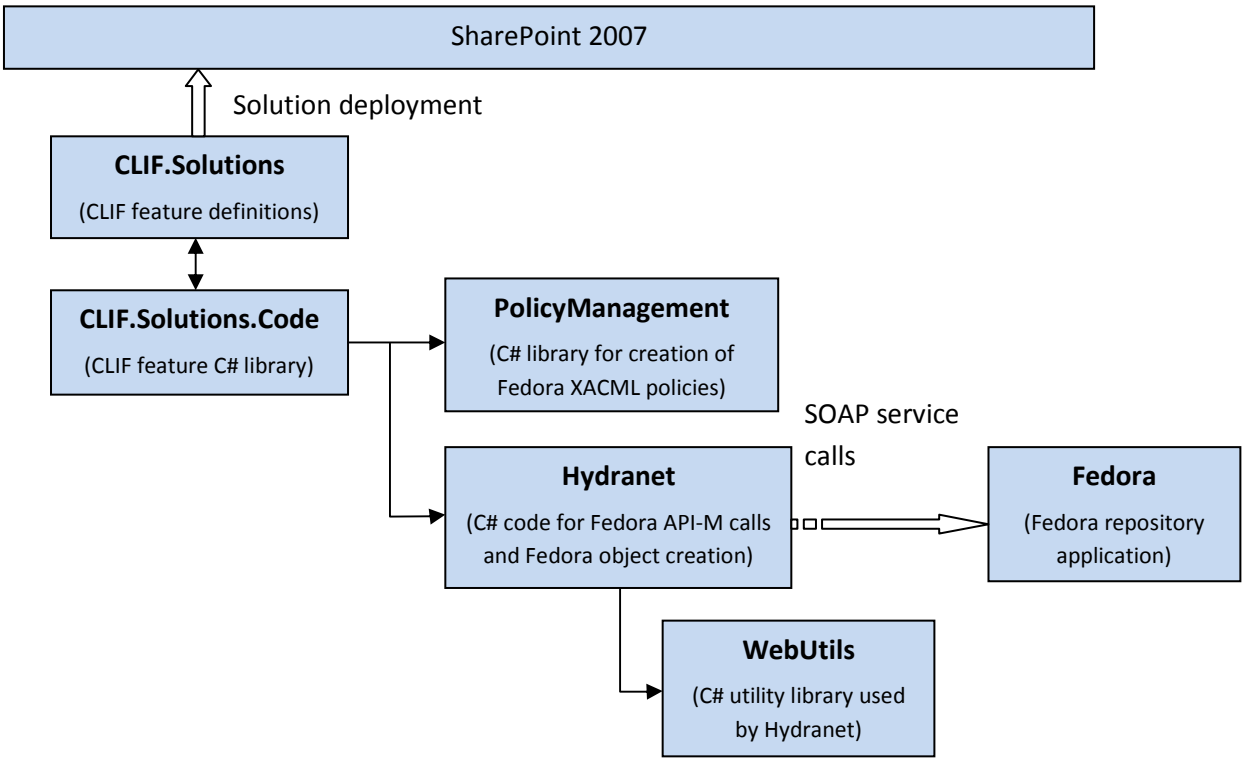

Figure 7: Structure of the SharePoint-Fedora integration solution

The current solution has been developed and tested on SharePoint 2007. The CLIF team performed selective porting of the solution to SharePoint 2010 following its release. Modifications are required to the CLIF.Solutions project, reflecting the fact that feature definitions in SharePoint 2007 and 2010 are incompatible, but the C# libraries can themselves be reused without change.

Lifecycle management

In discussing the ability to simply copy materials to the repository from SharePoint, it was stressed y those testing it that there needed be clear motivation for doing so. This is no different to previous experience of getting people to deposit in a repository, but is a motivation that needs to address the context of the user: if the content is sitting well and being managed well within SharePoint, why do I need to move it? The archival case was considered one of the strongest cases, but needs to be made clear (or mandatory) as it is also recognised that not everything needs archiving. There was some concern raised in allowing copying that this would result in duplication, which may have an

auditing implication. The ability to move content also needed such auditing so as to clarify where the version of record is being kept, though as a process it could be useful in formal archiving tasks. The ability to publish to the repository via a workflow was felt to be very useful and capitalising on a strength of SharePoint. This was a way to incorporate relevant checks on the content before deposit. If browsing the repository from SharePoint, it was necessary to make it clear that the results are externally hosted, but also helpful that it would be possible to access external sources from within the working environment.

The common view expressed was that content would primarily move from SharePoint to the repository, but this was mainly as a result of a lack of immediate use cases apparent in moving content the other way (even if you can search it usefully). This in part related to needing to be sure about access arrangements (what can I use?) and in part the need to keep track of what was where. Where preservation was noted as a requirement there was a clearly recognised need to also make sure it was clear what was where, and to minimise system use and duplication where possible.

Conclusions

The work of the CLIF project has been a successful exploration of how integration between different types of content management systems can be enabled. These integrations help to demonstrate how digital content lifecycle management can take place across systems. More specifically, the project reached the following conclusions:

- The management of digital content lifecycles has been extensively explored in the literature, from many different perspectives and in many different subject and content domains. The majority of these explorations focus on the processes involved in managing the different steps of the lifecycle, and whilst there is variation there is also a great deal of consensus in the descriptions of digital content lifecycles. This project has not sought to replicate this work or add to the variations in existence, rather to focus on implementation of digital content lifecycle principles across multiple systems. This practical aspect of how a digital content lifecycle can be put into practice is far less explored in the literature. This may be because technologies change and consistency in process is more important that focusing on specific systems; it may be that different domains put their findings into practice using technology designed for that domain, and do not have an identified need to move out of that domain. The literature suggests both. CLIF challenges in particular this latter position by recognising that different systems used to manage digital content within a University do not have to work in isolation, but can be used together.

- The technical integration work carried out has successfully demonstrated that diverse content management systems can be brought together to allow the seamless movement of content between them. Having identified a set of use cases from interviews with local users, we were nevertheless keen to ensure that implementing these use cases did not preclude other uses for the movement of content between the systems, and implemented them in as generic a way as possible. This has resulted in a flexible set of outputs that can be further developed and applied. Our evaluations revealed additional functionality and use cases that could be implemented, and we anticipate further use cases emerging as we implement the project's outputs more widely and more users become familiar with what is feasible.

- The work required to carry out the integration has been extensive and detailed, and it can also be concluded that the lack of the most up-to-date standards in the interfaces for content management presented by all three systems under investigation does not make the task of getting such systems to work together any easier. It is concluded from this experience that content management systems should be encouraged to make it as easy to get content out as it is to get content into them in order to facilitate seamless flow and enable the digital content lifecycle across systems.

- An assumption at the start of the project was that we would be agnostic about the direction in which content might flow between the systems once integrated. Evaluation feedback clearly suggests that the repository's archival capability is regarded as one of its strongest assets, and the area that the other systems could not offer comparable functionality on. Hence, the primary flow of content is into the repository. Having said this, the capability of moving content directly from the repository to the content management system also encouraged more active use of the repository. The role of the repository within a University would seem be regarded in terms of what it can offer that the other systems cannot, rather than try and compete on all levels. Whilst there is clear benefit in playing to one's strengths there is a challenge to clarify better at an institutional level what functionality is offered by different content management systems, so as to more fully understand how different stages of the digital content lifecycle can be best enabled.

Acknowledgements

We gratefully acknowledge funding for the CLIF Project from the 'Repositories Enhancement' strand of the Information Environment initiative of the Joint Information Systems Committee (JISC).

We are also grateful to colleagues, academic and administrative, at both the University of Hull and King's College London. Their time in helping us to formulate case studies and, subsequently, to evaluate the project's work against those starting points was much appreciated.

Finally, we acknowledge assistance from the University of the Highlands and the Islands (UHI). Ours was not the first attempt to integrate Sakai and Fedora, colleagues at UHI tried it some years ago as part of the JISC CTREP project. That they were only partially successful was entirely due to restrictions of the (then) Sakai code. Their work provided a very helpful starting point for the CLIF team and their time in helping us understand it was invaluable

References

Awre, C. (2010) "CLIF Literature Review". The University of Hull. Last updated January 31 2010. http://hydra.hull.ac.uk/resources/hull:2430

Awre, C., Green, R., Thampi, S., Thompson, A. and Waddington, S. (2011) "CLIF Project Final Report and appendices". The University of Hull. Last updated May 20, 2011. http://hydra.hull.ac.uk/resources/hull:4194

Awre, C., Green, R., Thompson, A. and Waddington, S. (2010) “CLIF technical design”. The University of Hull. Last updated August 13, 2010. http://hydra.hull.ac.uk/resources/hull:2697

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

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

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

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

Green, R. and Awre, C. (2009) "The REMAP Project: Steps towards a Repository-enabled Information Environment". Ariadne, University of Bath, Issue 59. Last updated June 09, 2009. http://www.ariadne.ac.uk/issue59/green-awre/

Green, R. and Awre, C. (2008) "RepoMMan: Delivering Private Repository Space for Day-to-day Use". Ariadne, University of Bath, Issue 54. Last updated February 26, 2008. http://www.ariadne.ac.uk/issue54/green-awre/

Green, R., Dolphin, I., Awre, C., and Sherratt, R. (2007) "The RepoMMan Project: Automating workflow and metadata for an institutional repository". *OCLC Systems & Services: International digital library perspectives*, Vol. 23, No. 2, 210-215

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

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/

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 (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

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

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

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