UNIVERSITY OF Hull

and

KING'S
College
LONDON

CLIF Project

_____

Use case summary

January 2010

JISC

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 CLIF Project is being undertaken by the Information Systems Group 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.

1. The use cases

This 'Use case summary' is based upon a number of interviews carried out at the University of Hull and at King's College, London. The interviewees were from a range of backgrounds, learning & teaching and administrative, and were chosen to cover a wide range of possibilities. The interviews sought to discover how people dealt with digital content and what kinds of software were used to manage it. Accepting that digital content may move within a range of software environments during its lifecycle, the interviewers sought to discover where repository software might usefully be employed as part of the long-term management process.

The use cases were:

- Hull
    - Records manager
    - Head of department
    - Senior lecturer

- King's College, London
    - Crystallography
    - Environmental Research Group (Modelling)
    - Environmental Research Group (Monitoring)

2. Sources and use of content

The following table takes information from the use case interviews which have been conducted and attempts to map it onto the Curation Lifecycle Model

developed by the Digital Curation Centre.[1]   Notes in blue attempt to identify points where text or data files might usefully be transferred to a Fedora-based storage environment.

Note: eBridge, mentioned in this table, is the University of Hull's implementation of Sakai as its Virtual Learning Environment.

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
Hull: Records Manager	**Policies and procedures**	Records manager gets completed P&P documents, generally via committees, to keep safe and accessible.	n/a	Repository general ingest tool	Some policies are reviewed annually, some 'as-and-when', some have specific review dates. Documents are regarded as permanent unless superseded	Local hard drive; Institutional repository	Institutional repository	
				Copies of record to Fedora – to be retained even when superseded				
	General departmental documents	Scanned into Amistore, sometimes all (space saving exercise), sometimes a more careful approach.	Records manager offers advice on selection if asked. Cost/benefit factors may influence selection.			Stored locally. Currently no central, authoritative source.	Local. Can also be accessed from financial and AIS systems.	
						May be a case for key materials to be held in an authoritative (Fedora) store?		

---

[1] See: http://www.dcc.ac.uk/lifecycle-model/

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Customer Relations Management	Documents can also go 'into' CRM either directly or linked, export is easy.			A lot of data is generally archived – 'deactivated' in CRM terms – to hide it and stop cluttered displays. It is easily reactivated. This is effectively the CRM preservation approach.		"Someone is looking at the idea of integrating CRM and SharePoint."	
	Training documents	Some are updated quite a lot and version control is tricky.			Consumers really need to know they have the up-to-date version where systems change.	There is no agreed central store where they are generally accessible. Some in Portal, eBridge, perhaps should have just used the repository.	Portal; eBridge	
						Case for creating a central, authoritative (Fedora) store accessible to other systems?		
Hull: Head of Department	**Standard policy documents**	Usually Word document managed in private area of SharePoint with its versioning controls turned on.	Final version transferred to public area of SharePoint.	Final version transferred to public area of SharePoint.	The visible library is refreshed every session. Old documents are archived into a folder structure. If a document is completely superseded it may be deleted but it will not be removed from backup storage.	Departmental SharePoint installation.	Public area of SharePoint	

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
						If these (visible) documents are intended to be accessible outside the department (FOA?) there is a case for a central, authoritative (Fedora) store?		
	Lecture content	Now generated largely in PowerPoint on a lecturer's 'own' machine.	Final version transferred to SharePoint.	Final version transferred to SharePoint.	The area is wiped every year.	Departmental SharePoint installation.	Migrated to a read-only area on SharePoint for students to access.	
	Examination papers	The documents use 'track changes' in Word and are accompanied by a comments form that may be electronic or manual. Computer Science is moving to a SharePoint-managed workflow to assist the process: it is not just a single document but the paper, answers, approval form(s) etc.			Old exam papers are archived. Only the previous year's papers are available to students at any given time.		n/a	
						A more formal archiving policy, perhaps as part of a wider university initiative, might prove useful in the long term.		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	[Course] Programme specifications	Handling programme specifications is about starting with a form coming in and then repeatedly adding to centrally held documents. Each stage in the approval process means adding more detail. The three stages of form are progressive and would usefully be managed by a SharePoint workflow.					At the moment the department's definitive final copy is held within SharePoint though Quality Office should have the 'real one'.	
						Arguably a central, authoritative store of these documents should exist for easy access by staff, students and others.		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Student software code	For big coding Team Foundation Server is available and may be mandatory for some. More generally, documents may be set up in a collaborative space in SharePoint but the lack of a decent drop-box facility is an issue. eBridge is used for the drop-box and the drops are pulled back in a big zip file. Some staff put responses back on eBridge. The size of submissions can be an issue with some assessments being greater than 50MB					[SharePoint] [eBridge]	
	Undergraduate dissertations	The undergraduate dissertation lifecycle involves now a CD. These dissertations are kept, never thrown out. The size of files can be an issue – the Games and Graphics MSc's frequently generate 50MB submissions and it may be necessary to submit these on CD rather than through a limited drop box.	Digital copies of all are transferred to repository.	Repository general ingest tool		Institutional repository. [Hard copies retained in departmental library.] Useful lifespan is probably only five years.	Institutional repository (restricted access).	

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Research	Internally most people will put up a non-publisher-format copy of the work as a non-locked PDF.			[Department does not use the metadata formally associated with documents in SharePoint.]	Local or SharePoint. SharePoint will now search across its own resources and local drives.	Departmental access only.	
						It may be useful in the context of a Research Assessment Exercise to have these accessible from a central university store. This would also facilitate open access?		
Hull: Senior Lecturer	**Student lab instructions**	The process starts with something like a lab sheet, a Word document, describing pictures of data; it is presented as a set of web pages in eBridge's weird templates, or it has been converted to web pages by some external tool and presented in eBridge as links.					eBridge	

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Student lab work	At different stages in the process involves large Word files, PowerPoints. Converted to PDF on submission.			Engineering does not provide for students in future years to see past attempts at similar work (there is a black market in past, marked reports).	All the main documents are converted to PDF, stored on TW's machine and backed up in various places. Accreditation panels may wish to review previous year's work. Material from the various discussion forums that students will have contributed to is wiped. In another module this may be captured to validate against WebPA for peer assessment.	n/a	
	Past undergraduate examination papers			Normal repository ingest tool.		Institutional repository	Institutional repository (and linked from eBridge)	
Kings: Crystallography	**Data (general)**	Data is captured on a server local to the capture device, is transferred (via DVD or external drive) to a researcher's desktop for processing and finally publishing.	'Reflection file' and final 'coordinate file' deposited in the Protein Data Bank.		Local archive copy (offline).	Protein Data Bank (PDB)(community database)	Protein Data Bank (PDB)(community database)	

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Data (instrumental)	In the laboratory, there is a system called the Laboratory Information Management System (LIMS), which can automatically capture data from certain instruments. Lab users can then retrieve the data in various formats such as PDF and XML.					Repository?	
		Automatic capture is achieved via 'Laboratory Information Management System' (LIMS).			LIMS can create very big files which prevent the system managing them long-term to provide an archive. Case for capturing them to a repository?	Repository?	Repository?	
	Diffraction images, reflection files, coordinate files and associated scripts.	Once captured, images are compressed and saved to a DVD or external drive, then uncompressed on the researcher's desktop prior to processing.				Diffraction images are vital during research period (1-5 yrs) until processed; reflection and coordinate files are vital until they appear in PDB; scripts are important during the research period. All are important for up to 10yrs after the project.	Some go to PDB.	

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
					Currently there is no formal backup procedure to ensure longevity of diffraction images etc on which the published output is based.	Is there an argument for duplicating these in a local repository – especially if the diffraction images are 'at risk'? Storage of the diffraction images would allow analysis to be re-examined or repeated.		
Kings: Environmental Research Group (ERG) **1. Modelling**	**Storage generally**	Data processing is performed on local machines and transferred to a network drive when completed.	ERG operate a semi-formal policy of retaining all data since their creation in 1993.		The modelling team standardise data files into a normalised format for import into the London Atmospheric Emissions Inventory (LAEI) and Emissions Toolkit.	LAEI.		
	Third-party raw data	Data may be provided in one of several formats depending on content and/or source: .xls, csv, ASCII, CMAC ASCII, SQL database. Normalised to an Access database.				No minimum or maximum time for retention. Vital for the operation of the LAEI.		
		It is not clear whether the raw (un-normalised) files have an ongoing value and should therefore be captured.						

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Emissions estimates	Emission estimates are generated by the ERG Modelling team using the Emissions Toolkit.				The ERG has a contractual obligation to store data on different modelling scenarios for a five year period, in order to answer questions regarding its validity. The Lancet journal indicates that "*authors may be asked to provide the raw data used for research papers when they are under review and up to 10 years after publication in The Lancet*".		
						Although not specified, it is likely these files are held on network storage. A Fedora solution, with appropriate metadata, could be envisaged?		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Pollution modelling	Pollution models are generated by the ERG Modelling team using the Air Pollution Toolkit when considering different pollution scenarios. Uses Access plus other commercial software (eg Golden Software's 'Surfer' (GIS)) to generate emissions data . Access plus other commercial software (eg Surfer) to generate pollution predictions. Three types of file produced: Emissions data (.emi), emissions model (.mod), Surfer grid (.grd).				The ERG has a contractual obligation to store data on different modelling scenarios for a five year period, in order to answer questions regarding its validity. The Lancet journal indicates that "*authors may be asked to provide the raw data used for research papers when they are under review and up to 10 years after publication in The Lancet*".		
						Although not specified, it is likely these files are held on network storage. A Fedora solution, with appropriate metadata, could be envisaged?		
	Surfer grids							

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Research papers	Most of the departments work is published via *The Lancet* who expect Word (.doc) or PDF files (PowerPoint and Excel are accepted for specific forms of data.)						
						It is not clear how these are stored long-term or if they are made accessible on-line. A Fedora solution could be envisaged.		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Emails					E-mails that contain raw data as attachments are retained as evidence of receipt. However, criteria for their retention are not explicitly stated in a formal or informal policy. An email with data attachment may be used to establish the first stage in a data audit trail and therefore may potentially be subject to the five year retention period established in modelling contracts. Some e-mails are retained as evidence of a negotiation process or subsequent discussion. However, criteria for their retention are not explicitly stated in a formal or informal policy. An e-mail that fits into the category is likely to have value during the lifetime of the contract and during the five year review period.		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	LTS model / road network map	ASCII-based format (.sel) for mapping sections of a road network onto a geographic map.				Local servers.		
Kings: Environmental Research Group (ERG) 2. Monitoring	General		ERG operate a semi-formal policy of retaining all data since their creation in 1993.			Retention period for individual items may be mandated by funding bodies. Monitoring team is required to provide full traceability of all data for five years.		
	Raw measurement data	Created by bespoke 'MONNET' software. Exact format depends on capture device.			Converted to normalised format for import to the London Air Quality Network database.	Structured network storage. Held for five years in case of query but in practice it is de-scaled and/or ratified data that is requested.		
	Calibration reports	Created every two weeks for each site. May reach ERG by email or fax – if the latter scanned to .jpg. Emailed may be .doc or .xls.				Stored in recipient mailbox and may additionally be stored as a text file in an appropriate project/device directory on shared drive.		
	Email	May contain text, data or both.						

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Normalised data	Raw data is converted to normalised .csv files.				Raw data is normalised by MONNET, inspected and, if accepted, stored in a SQL database. Held for five years in case of query but in practice it is de-scaled and/or ratified data that is requested.	London Air Quality Network database.	
	De-scaled data	Normalised data that has been modified using calibration data.					London Air Quality Network database.	
	Ratified data	Normalised and/or de-scaled data that is confirmed as providing an accurate measurement. Stored in a SQL database.					London Air Quality Network database.	
						It seems clear that ratified data should be stored long-term; it is not clear how much of the contributory data it would be useful to retain beyond the mandatory five year period.		
	Email					Limited long-term value? Important during the five year mandated retention period.		

		Create/receive	Appraise and select	Ingest	Preservation action	Store	Access, use and reuse	Transform
	Administrative metadata	Created by ERG at each stage in the process to provide an audit trail.						
						Clearly this should be retained for as long as data to which it refers is retained.		

3. Generic use cases

Using the information above, it is possible to propose a set of generic use cases which the CLIF Project might usefully address. Not all 'real-life' use cases will have all the stages identified here, nor will they necessarily map exactly onto what is proposed.

3.1 Generic data use case

The first generic use case deals with experimental data and accompanying documentation.

Process	CLIF functionality
Raw data is collected from automated systems	CLIF should investigate functionality to allow copies of the unprocessed data to be stored in a repository against future re-use
Raw data is normalised and stored in a 'common' file format	CLIF should investigate functionality to allow copies of the normalised data to be stored in a repository against future re-use
Data is processed through key intermediate stages	CLIF should allow repository storage of key stages in data to allow for partial re-use without the necessity to start again from 'square one'
Results of experiment are produced	Experimental results (data and accompanying diagrams, charts etc) should be captured
Accompanying materials	Any written materials accompanying the above stages should likewise be captured (Lab books? Audit metadata?)
Written documentation is produced	Author's reports and papers should be captured
Preservation	At the points of ingest to the repository, appropriate consideration should be given to preservation issues.

3.2 Generic text use case

The second generic use case deals with essentially textual material.

Process	CLIF functionality
Document is cyclically drafted and revised	Any precursor document (previous edition?) should be considered for archiving. Unless the drafting is part of a 'creative process' (eg, a literary work) it is probably not necessary to capture versions although this option should be offered where possible. In the case of a literary work or similar the development of the document may form an important part of the historical record.
Final version of document is produced	CLIF should allow a version of record to be stored in a local repository. Where the contributing system provides additional information (metadata, permissions etc) these should be retained in the repository object if possible and be available to inform creation of the equivalent repository information.
New version of document required	CLIF should allow the download from the repository of a document, if possible in its

	original unchanged format, so that it can be opened in its original authoring environment as the starting point for a new version or edition.