THE UNIVERSITY OF HULL

with

SPOKEN WORD SERVICES

REMAP Project

D1

Records management and preservation requirements

Richard Green, Chris Awre, Judy Burg, Vicky Mays and Iain Wallace

November 2007

The REMAP Project

Project Director:	Ian Dolphin, Head of e-Strategy, University of Hull	(i.dolphin@hull.ac.uk)
Project Manager:	Richard Green	(r.green@hull.ac.uk)
Repository Domain Specialist:	Chris Awre	(c.awre@hull.ac.uk)
Technical lead (Hull):	Robert Sherratt	(r.sherratt@hull.ac.uk)
Software developer (Hull):	Simon Lamb	(s.lamb@hull.ac.uk)
Archivist:	Judy Burg	(j.burg@hull.ac.uk)
Records Manager:	Vicky Mays	(v.mays@hull.ac.uk)
Project Director for GCU:	David Donald	(D.Donald@gcal.ac.uk)
Project Manager for GCU:	Iain Wallace	(iain.wallace@gcal.ac.uk)
Technical lead for GCU:	Graeme West	(graeme.west@gcal.ac.uk)

The REMAP Project is being undertaken by the e-Services Integration Group at the University of Hull and Spoken Word Services at Glasgow Caledonian University. It is funded by the JISC Repositories and Preservation Programme.

1 Introduction

The JISC has, over the past few years, funded a wide range of projects in the records management and digital preservation[1,2,3] (RMDP) arena, and established much good practice and knowledge. Many of the projects have worked on the basis of there being an acceptance that interaction with a repository for RMDP is a good thing. The work carried out has certainly demonstrated that this is the case. However, there remains a potentially damaging disconnect between the repository and the people expected to use it on a regular basis.

Web services development to support records management, a specific form of preservation, has so far tended to be through the provision of commercial systems, though there is potential applicability of preservation Web services to be focused around specific records management needs.

Web services, whilst offering great potential, do not in themselves offer integration between a digital repository and the desktop applications used to create digital materials. This gap can, though, be addressed through notifications, alerting material owners and/or creators to specific records management and/or digital preservation tasks that need to be carried out. The repository can thus enter into an engagement with the end-user in order to encourage RMDP processes over a period of time.

The JISC-funded RepoMMan project[4] based at the University of Hull has developed a tool to orchestrate Web services using WSBPEL (or just BPEL) over a Fedora repository[5]. The tool allows the user to interact with the repository in a variety of ways; the REMAP project will enhance and extend this tool to provide support for RMDP. The work will be based on user requirements for institutional RMDP and the development of related models describing RMDP workflow. REMAP will develop a notification layer as part of the overall orchestration to inform users of repository tasks that require their attention and encouraging their engagement with the system.

The first stage in the project was a user needs gathering exercise to determine records management and preservation requirements. This work, which forms the substance of this report, was conducted with a range of people from the University of Hull (UoH) and staff from Spoken Word Services (SWS) at Glasgow Caledonian University. The interviews at Hull concerned themselves primarily with material in text form whilst the work with SWS concentrated on their work with audio material.

The underlying intent of REMAP's work is to embed digital repository interaction within working practices for RMDP purposes and thereby demonstrate the potential benefits of so doing.

---

[1] Digital Preservation and Records Management Programme, http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation.aspx

[2] Supporting Institutional Records Management Programme, http://www.jisc.ac.uk/whatwedo/programmes/programme_supporting_irm.aspx

[3] Digital Preservation and Asset Management Strand, http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_404.aspx

[4] RepoMMan project, http://www.hull.ac.uk/esig/repomman/

[5] Fedora digital repository system, http://www.fedora.info/

2 Needs gathering

2.1 Overview

During early project meetings, considerable time was devoted to identifying a number of potential users for the eventual REMAP tool whose processes could be used as sources of RMDP requirements and whose day-to-day work gives rise to a set of use cases that can be used to ground the work of the project in the real world.

The first set of processes are external to the REMAP Project but based at the University of Hull:

- Preparation of committee papers by Committee Section
- Dealing with past undergraduate examination papers and providing them as learning resources
- Learning and Teaching Programme Approvals
- Maintaining the University's Register of Policies and Procedures
- Maintaining the University Quality Handbook

In addition there are a number of processes in which members of the REMAP Project team are directly involved:

At SWS, Glasgow Caledonian University:

- Development of RMDP for a large repository of audio materials

At the University of Hull:

- Development of the RepoMMan tool to 'publish' materials
- Development of repository functionality to support electronic theses and dissertations (ETD)
- Development of procedures specifically to support digital archives
- Development of an integrated web presence for the newly formed Centre for Spirituality Studies

For the first group of processes, interviews were carried out by members of the project team with staff involved and were documented. A meeting held over two days was devoted to considering the SWS requirements, summarised in a report. For the final group of processes a brief report on each was produced by the team member(s) involved. In some of these cases, though not all, the outline workflow was represented in diagrammatic form.

All these sources of information are considered later in this report and all are considered in order to identify possible user needs. A subset is then identified to guide and ground the technical development work of the project. More formal workflow diagrams will be developed for these identified use cases.

2.2 Preparation of committee papers by Committee Section

The management of documents in the University Committee Section is a complex process perhaps best explained with reference to the diagram that follows, and in particular to the second 'layer' of the diagram: 'Document process'.

There is a commonality of approach to the work with all the groups that the Section deal with from sub-committees through to Senate and Council. It is also the case that a document that starts with a relatively lowly sub-committee may cycle through the structure to emerge, probably in modified form, at the Senate or Council level.

A document arrives to be inserted into the committee system from an originator. Committee Section do not themselves originate documents other than procedural ones such as minutes and agendae. This received document may be in one of a number of common formats. The document is placed into the collection of papers for the particular committee.

In an ideal system, Committee Section would like to be able to expose these collected papers, together with the agenda, to members of the committee and other legitimate readers in electronic form so that paper versions do not have to be circulated. This would require the ability for secure sharing at an individual collection level. The discussion papers made available in this way need to be preserved, unaltered, for the record.

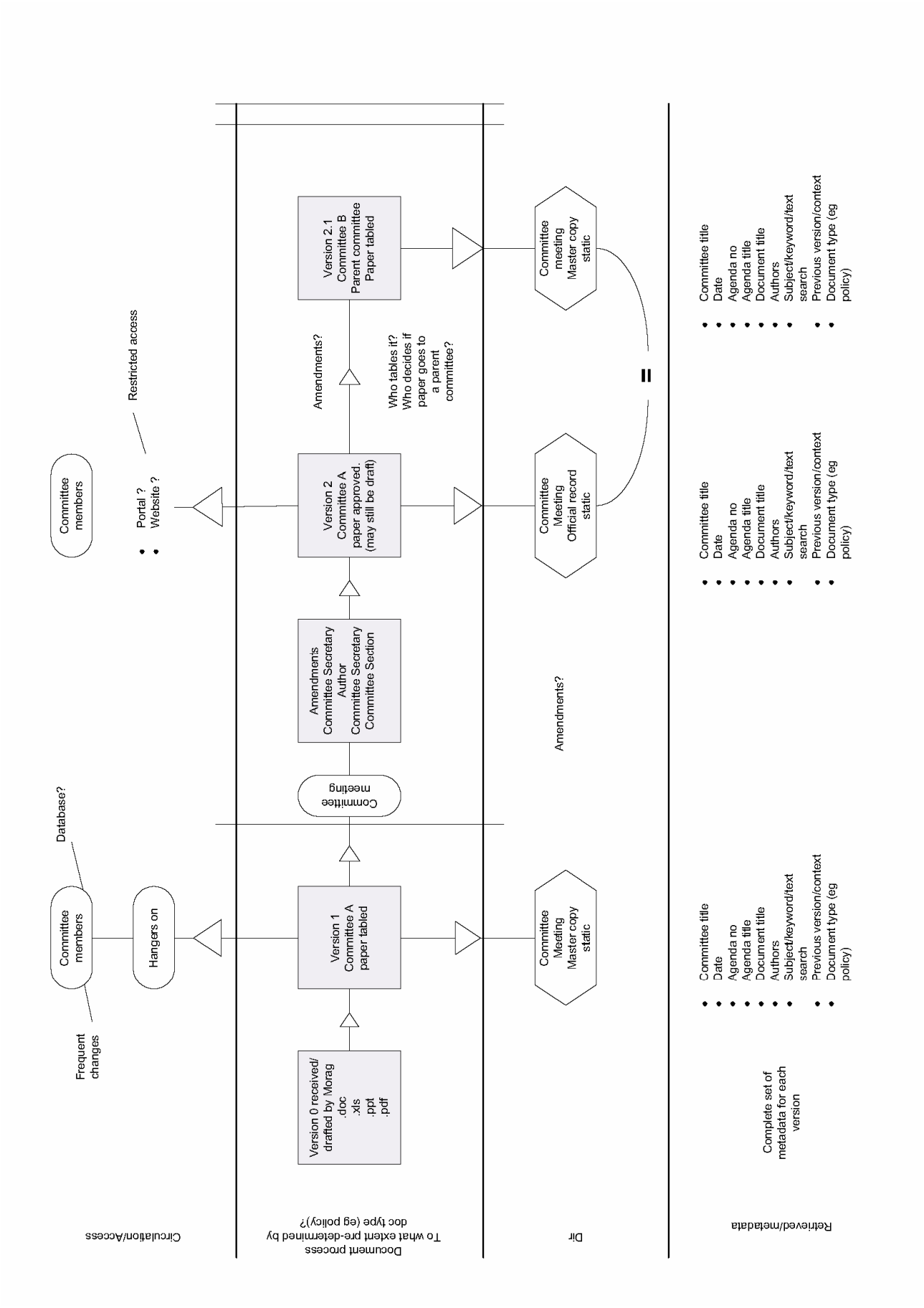
Following the meeting, the document may be deemed 'dealt with' and go no further within the committee structure. However, revisions may be required by the committee, in which case the document is amended by the author, and/or it may be passed up to a more senior committee either unaltered or in amended form. Thus a document comes into the system for a further committee meeting and the process starts again. (The double 'bar lines' at the end of the row indicate a potential repeat to the first bar line.) The committee section manage these processes

At each cycle a version of record needs to be kept as the paper submitted to the committee at a particular meeting. A version of record also needs to be kept of a paper approved by the committee. It may or may not have been amended according to the committee's recommendation, so there are potentially two identical versions of a paper which need to be kept, each with a separate context and therefore metadata. At any stage of the cyclical process it may become appropriate to make a record version of the document available more widely, perhaps with published committee minutes. These minutes may variously be available only to a restricted group of people or more widely - or even publicly. In whichever situation, access must be carefully controlled.

Committee agendas, papers and minutes, taken together, constitute a key part of the University's archive. Therefore permanent preservation of documents, metadata and interrelatedness of documents (within and between committees) is also a key factor.

Committee Section are not primarily interested in using REMAP processes to pass messages to the 'end-users' of their services (by which we mean committee members). Rather they are in interested in harnessing the repository to help with the workflow required to manage committee papers from day to day and in having the system alert them to coming cycles of events.

The ability to keep and control versions of documents is essential and it follows that there is a need to keep and control the metadata associated with those documents at each stage in their lifecycle. It is desirable that when a document of record is finalised at one committee level its metadata should go with it if the document is passed forward, perhaps to a higher committee. It is also likely to be important that relationships be established and maintained between these documents in the repository. The ideal would be that when a document of any sort is approved, the version of record in the repository has a complete, versioned metadata record going back to the point that the document first entered the committee system and that the version of record has declared relationships to all previous versions of the document as it passed through the system.


Circulation/Access
Document process
To what extent pre-determined by doc type (eg policy)?
Dir
Retrieved/metadata
Committee members
Frequent changes
Database?
Hangers on
Version 0 received/ drafted by Morag .doc .xls .ppt .pdf
Version 1 Committee A paper tabled
Committee meeting
Amendments Committee Secretary Author Committee Secretary Committee Section
Version 2 Committee A paper approved. (may still be draft)
Committee members
Portal ? Website ?
Restricted access
Amendments?
Who tables it? Who decides if paper goes to a parent committee?
Version 2.1 Committee B Parent committee Paper tabled
Committee Meeting Master copy static
Amendments?
Committee Meeting Official record static
Committee meeting Master copy static
Complete set of metadata for each version
Committee title
Date
Agenda no
Agenda title
Document title
Authors
Subject/keyword/text search
Previous version/context
Document type (eg policy)

2.3 Dealing with past undergraduate examination papers and providing them as learning resources

The provision of past undergraduate examination papers was a function carried out by the University library. In recent years, and for a number of reasons, this service has become fragmented and the mechanisms for such provision have become piecemeal. It is envisaged that provision of examination papers to the University repository might become the normal method of providing access to future students.

The process can be seen in three stages.

The first stage involves reminding departmental secretaries that digital copies of examination papers will be required for the repository; this reminder would be issued at the beginning of each examination cycle. At the end of the cycle an actual request to provide the papers would be sent to the departmental secretaries.

Secretaries would be given a month or so to respond to the request. Specific requests would then be sent out for examination papers that had not been received. The collected examination papers would then be checked, converted to locked pdf files and placed in the repository in time to inform students in the next examination cycle.

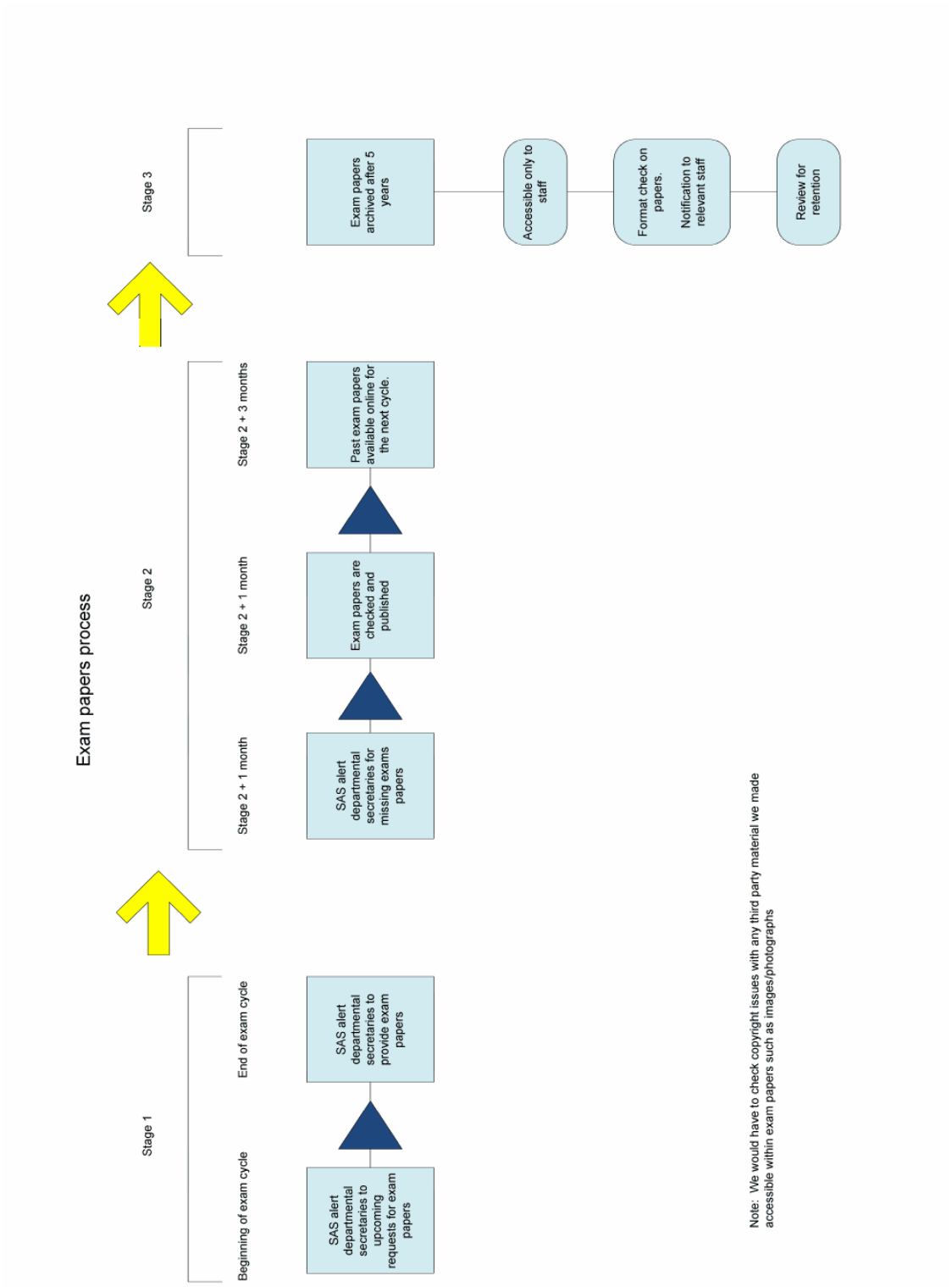
Checking the papers would potentially involve a number of things, amongst which a check that publication in the repository is not likely to infringe third-party copyright. The publication process would set a number of flags in the repository object. The primary flag would cause review of the object after five years. If the repository manager did not deem otherwise the paper would normally be made invisible to students, being potentially out of date; the paper would remain accessible to staff. Further flags would potentially trigger periodic format reviews and ultimately a retention review.

Messaging would be required:

- to send alerts (and reminders) to secretaries
- to notify the repository manager that a five-year review is due
- to notify the repository manager of any default action taken at this point
- to report periodic format checks
- to notify the repository manager that a retention review is due

It would be useful to be able to generate reports to the Examinations Office summarising the range of papers that has (or has not) been put up on the repository for each examination season. To facilitate this it might be appropriate to create a 'collection' object for each subject when alerts are sent to secretaries. This object would be able to carry alerts and flags even in the absence of papers and could serve as the trigger point for reminders and reports of non-deposit.

It was suggested, in the case of a five-year review message that the alert be triggered at collection level rather than send an alert for each individual paper in the collection. However, in responding to the alert it would be useful for the administrator to be able to remove (in the sense of hide from public view) some papers but extend the life of others. It might further be useful if the alert could be copied or forwarded in an e-mail to a third party: for instance, "I've just received this alert from the repository, have you a view on how we should respond?" There was some discussion of whether it should be possible to re-assign an alert to another person; this might be useful if that person also has appropriate rights to manipulate the repository content.


Exam papers process
Stage 1
Beginning of exam cycle
End of exam cycle
SAS alert departmental secretaries to upcoming requests for exam papers
SAS alert departmental secretaries to provide exam papers
Stage 2
Stage 2 + 1 month
SAS alert departmental secretaries for missing exams papers
Stage 2 + 1 month
Exam papers are checked and published
Stage 2 + 3 months
Past exam papers available online for the next cycle.
Stage 3
Exam papers archived after 5 years
Accessible only to staff
Format check on papers.
Notification to relevant staff
Review for retention
Note: We would have to check copyright issues with any third party material we made accessible within exam papers such as images/photographs

2.4 Learning and Teaching Programme Approvals

Learning and Teaching Programme approvals is the process of developing new courses at the University. The current process is not well documented, it is fragmented across the institution and currently under review.

There are three stages to the final approvals process which is currently delegated to faculties. The first stage is the development consent which is also linked to the planning round, the second stage is planning permission and the final stage is full approval. This cycle can take up to a year but can be compressed in to three months if necessary.

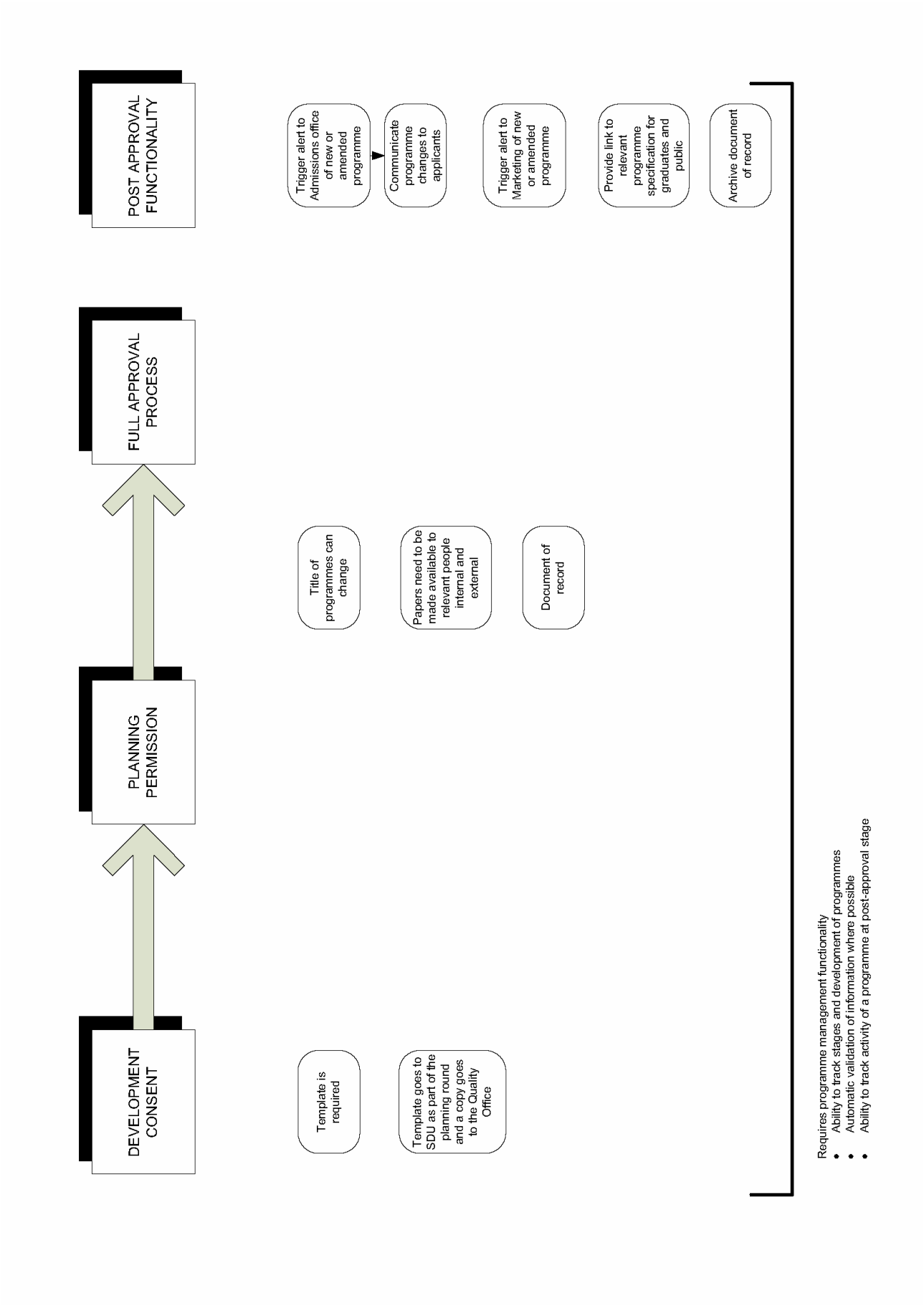
Possible requirements for a new programme approvals process to be managed through the repository may include the following:

The document needs to be monitored and tracked by relevant departments through the three stages of the approvals process

Version controls need to be applied to the development of the document.

The document will need have different access levels through the three stages of the approvals process.

It was felt that what was needed here was some form of 'workflow object'. This would be created when a new programme approval process was started and would contain alert and status flags for the various stages within it. It would effectively record progress through the process and perhaps act as the collection point for the various papers that need to be gathered as the process unfolds.


DEVELOPMENT CONSENT
PLANNING PERMISSION
FULL APPROVAL PROCESS
POST APPROVAL FUNCTIONALITY
Template is required
Template goes to SDU as part of the planning round and a copy goes to the Quality Office
Title of programmes can change
Papers need to be made available to relevant people internal and external
Document of record
Trigger alert to Admissions office of new or amended programme
Communicate programme changes to applicants
Trigger alert to Marketing of new or amended programme
Provide link to relevant programme specification for graduates and public
Archive document of record
Requires programme management functionality
Ability to track stages and development of programmes
Automatic validation of information where possible
Ability to track activity of a programme at post-approval stage

2.5 Maintaining the University's Register of Policies and Procedures

The University's policies and procedures are stored in a central register. The current process of capturing policies and procedures into the Register and ensuring an annual review is carried out is time consuming and problematic.

It is envisaged that provision of policies and procedures to the University repository might become the normal method of providing access to future students.

After a policy paper has been produced or amended through the committee cycle the paper is then sent through to be maintained on the Register of Policies and Procedures. Document metadata is captured in a database and the document is published on the staff portal.

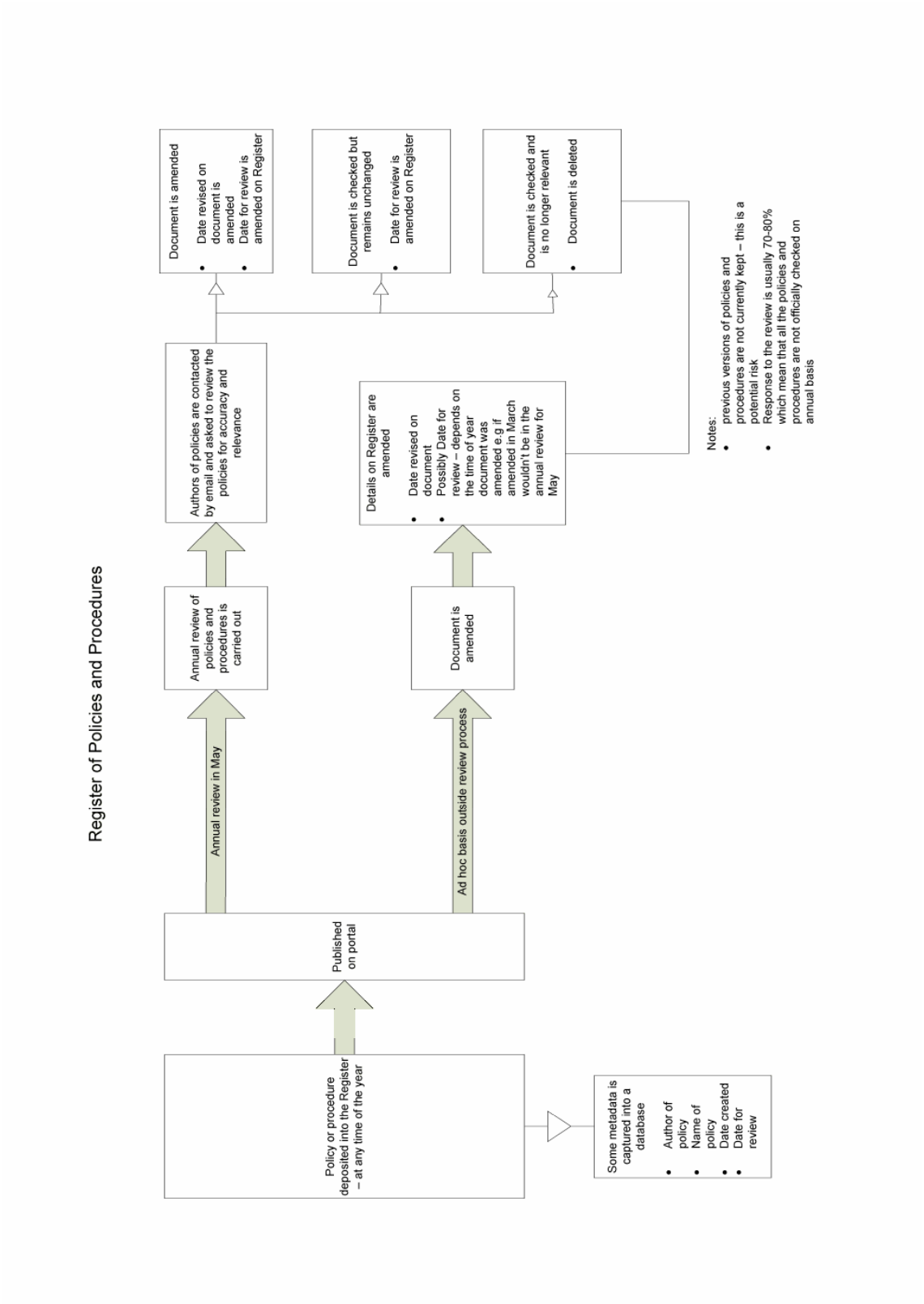
It is required that the policy document is reviewed on an annual basis. Staff responsible for those documents are contacted and asked to review the documents for accuracy and relevance. Documents can also be reviewed and amended outside the annual review period and therefore details can be sent to the Register of Policies and Procedures throughout the year.

Currently previous versions of policy and procedure documents are not held. The method of contacting staff to remind them to review documents is time consuming and does not have a good response rate.

It would be helpful if electronic documents entering the Register could arrive with all the metadata from their progress through committees attached. Once within the electronic register there will be a need for annual review and, following that, possible revision. It was agreed that a document in the Register might be in one of four states:

- current and up-to-date
- needing review
- in review
- archived

This last category reflects the fact that 'old' policies need to be retained for the record even though they have been superseded.


Register of Policies and Procedures
Policy or procedure deposited into the Register – at any time of the year
Some metadata is captured into a database
Author of policy
Name of policy
Date created
Date for review
Published on portal
Annual review in May
Annual review of policies and procedures is carried out
Authors of policies are contacted by email and asked to review the policies for accuracy and relevance
Document is amended
Date revised on document is amended
Date for review is amended on Register
Ad hoc basis outside review process
Document is amended
Details on Register are amended
Date revised on document
Possibly Date for review – depends on the time of year document was amended e.g if amended in March wouldn't be in the annual review for May
Document is checked but remains unchanged
Date for review is amended on Register
Document is checked and is no longer relevant
Document is deleted
Notes:
previous versions of policies and procedures are not currently kept – this is a potential risk
Response to the review is usually 70-80% which mean that all the policies and procedures are not officially checked on annual basis

2.6 Maintaining the University Quality Handbook

The Quality Handbook comprises of University regulations and codes of practice in accordance with Quality issues. It is a series of documents which are reviewed as an ongoing process and go through the committee cycle for approval.

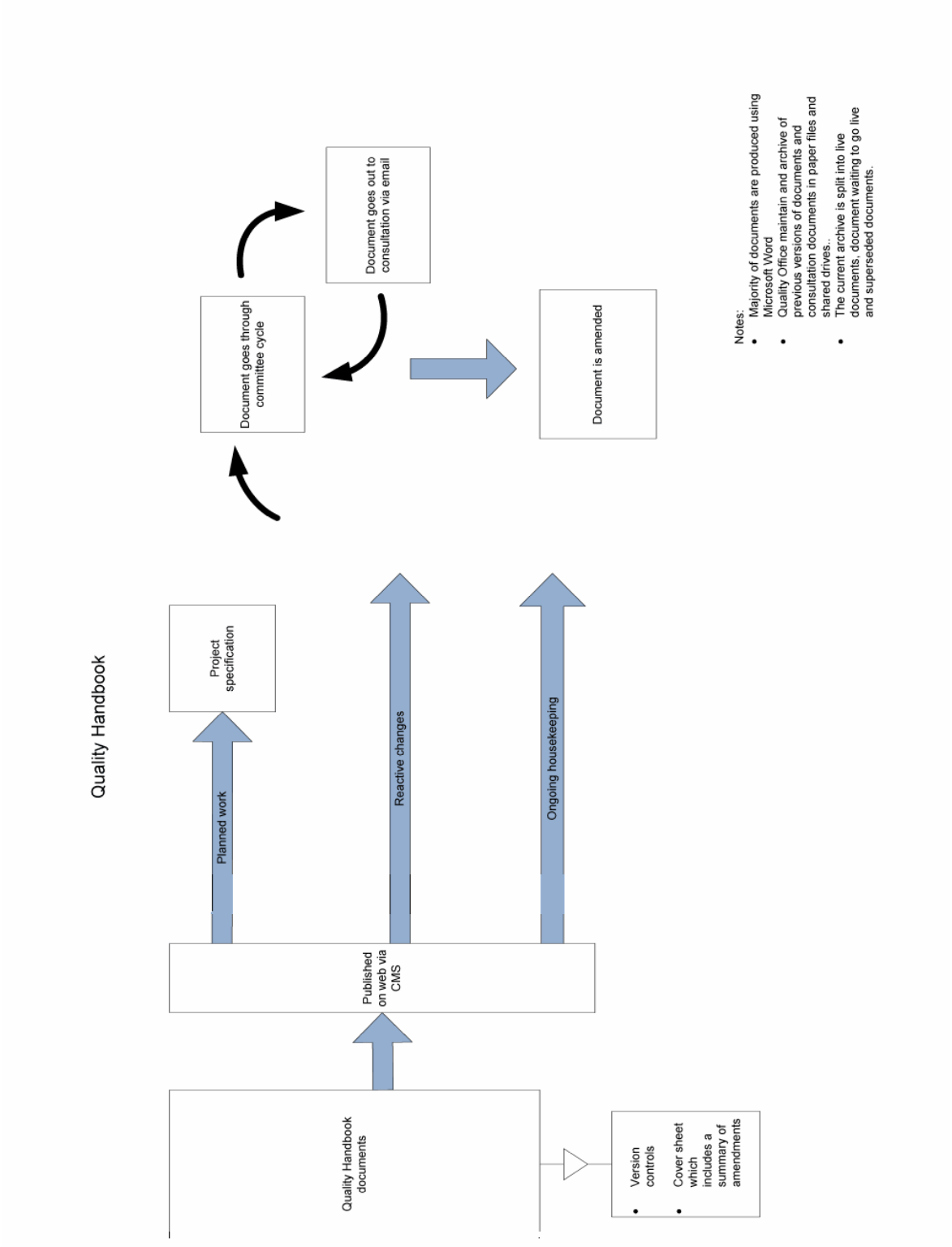
The Quality Office are responsible for publishing the documents which are currently made available on the University website. The Quality Office also maintain their own archive of quality handbook documents on the departmental shared network drive. Documents in this archive are split into live documents, pending live documents and archived previous versions of documents.

Consultation documents relating to an amendment of a Quality Handbook document are also stored in a combination of paper files and electronic documents stored in shared network drives. These documents need to be retained for QAA purposes and future redevelopment.

The majority of these documents are produced in Microsoft Word. Version control is applied to each document and draft document and currently each document has a cover sheet recording a summary of change

Documents are usually amended during one or more cycles of committee consultations. So far the issue of making papers available for consultation has not been resolved, currently consultation papers are circulated via email. The final approved Quality Handbook documents are published on the website.

The requirements for this process seem to be a combination of those required for previous examples: there is a need for elements of the committee process in terms of dealing with meeting papers and progression through committees, and there is an overlap with the Register of Policies and Procedures in that there is a need for a review and revision cycle operating to a known timetable.


Quality Handbook
Project specification
Planned work
Reactive changes
Ongoing housekeeping
Published on web via CMS
Quality Handbook documents
Version controls
Cover sheet which includes a summary of amendments
Document goes through committee cycle
Document goes out to consultation via email
Document is amended
Notes:
Majority of documents are produced using Microsoft Word
Quality Office maintain and archive of previous versions of documents and consultation documents in paper files and shared drives..
The current archive is split into live documents, document waiting to go live and superseded documents.

2.7 Development of RMDP for a large repository of audio materials at Spoken Word Services

2.7.1 General

Spoken Word Services, as the name implies, is a service based organisation based at the Saltire Centre in Glasgow Caledonian University (GCU). Its main function is to provide recordings of the spoken word, and related material, for the teaching and learning community - some at GCU but more likely elsewhere. By far the majority of the recordings currently provided are sourced from the BBC's archives, but the number sourced from elsewhere is increasing. Users of SWS include:

- external teachers/learners, even from overseas - sometimes quite unsolicited; the first contact is usually "do you have...?" GCU, as it is currently staffed, cannot meet the demand from this group and so tend to pick 'potentially interesting' people to work with.
- students (internal, external; under-/post-graduate)
- librarians
- SWS staff
- subject specialists
- ...

Note 1: teachers and post-graduate users are the most likely to be 'collectors' of sound recordings both from SWS and elsewhere.

Note 2: potential users who are not looking for BBC material are increasingly approaching SWS and saying "please will you record...?"

At present SWS have their materials and metadata in a home-grown system. They intend to move to a unified Fedora-based repository in the near future. At present, all management interaction with the repository materials (ie excluding download) is by SWS staff members and this is not deemed to be a sustainable model; SWS would like to move towards a situation where at least some users were able to deposit materials themselves.

The following two sections on workflow, whilst looking forward to some enhancements, essentially describe the situation 'as-is'. The document then moves on the consider how the ideas being developed by the REMAP Project might impact on SWS's records management and digital preservation (RMDP) processes.

2.7.2 Workflow (BBC materials)

The first stage in a request for BBC materials is to check whether SWS have the recording already. SWS is currently investigating integration of their materials into the GCU library catalogue to make this stage easier. The work is at proof-of-concept stage.

If SWS do not have the recording, a search is made by SWS of the BBC catalogue. Again, there is work at the proof-of-concept stage to 'scrape' the entire BBC catalogue on a regular basis (approaching one million objects) so that this can be held locally in a more user-friendly form. *[SWS recognise that there could be performance issues and that this may require 64-bit machines etc]*

If the recording is found, it is requested from the BBC and should arrive, typically, 3-6 weeks later. SWS's collectors are informed of this at the beginning of the process. SWS try to prioritise materials based on teaching and learning needs and deadlines - in practice it is difficult for the BBC to do the same: material can come from many different regions and then needs to be managed and distributed to SWS via Information & Archives in London. Reminders are sent and materials chased but this is done informally and with respect and awareness of the internal workings of the BBC. Requests are currently batched and submitted to the BBC

as a spreadsheet. (This is the BBC's preferred way of working.) At the stage a request is submitted, a 'proto-object' is generated in the repository as a placeholder; this prevents later bottlenecks in the workflow which could be caused by large batches from the BBC arriving at once. This is a low risk strategy because SWS typically have a 95% success rate with receiving materials that they request (occasionally the master tapes have been 'misplaced'!)

Once the recording is received, two parallel processes take place: the recording is transcoded into an appropriate format and its metadata is normally enhanced (over and above the BBC supplied version); and the process of obtaining appropriate rights clearance on the material is started. This is a formal process but is currently managed using a Filemaker database, separately from the main repository. SWS intend to bring this information into Fedora as they migrate. It may sometimes be that the transcoded material is made available to the requester for private analysis/use before rights clearance is obtained.

A number of things can then happen.

- Having had access to the material, the requester may find that it is not what they wanted after all and cease to have an interest in it. This would not stop SWS completing the accession process.
- Rights clearance may not be obtained. Again, this would not stop SWS accessioning the material but it would not be made public; the rights situation may change in the future.
- Rights clearance may be obtained and the material will be accessioned and made available.

The object that results in the repository should not be seen as a 'static' object. Once published, the metadata may well need to be corrected or enhanced. It may be that aspects of the metadata turn out to be wrong, hence correction. It may be that the metadata needs to be supplemented: perhaps the clip is rebroadcast by the BBC, perhaps Spanish television use it, perhaps the BBC discover that an end user has better metadata than they do and pull this back. This pull-back process raises issues of authority and provenance.

SWS are at the very early stages of looking at how they would cope with regular 'scrapes' of the BBC catalogue and dealing with revised metadata for recordings that they already list. Initial thoughts are to do this by comparing some sort of checksum; if the checksum from the BBC catalogue does not match that recorded in the SWS catalogue, a new version of the metadata stream would be created.

The SWS rights database is currently held separately from the metadata and recordings. (The details of the rights clearance process don't have to be in the object.) The detail in the database is held so that SWS could show 'due diligence' if ever asked. For BBC-derived materials the primary rights clearance is covered by a blanket agreement between the BBC and SWS and is taken as read; SWS are responsible for getting third-party rights clearance, for instance from the actors involved. Non-BBC materials are covered by other agreements such as Creative Commons licences. SWS does not have funding to buy rights; if an actor, for example, demands a fee for the use of the material in question it is likely that the recording will not be made available to the public but will, instead, be placed into the dark archive.

2.7.3 Workflow (non-BBC materials)

The workflow for non-BBC materials is essentially similar to that described above, however there will be no simple, 'routine' mechanism for obtaining a recording and primary rights clearance cannot be assumed. The process will be slightly different again where SWS are involved in making the recording.

2.7.4 SWS, RMDP and the RepoMMan repository model

Having established the essential elements of the existing and planned SWS workflows, discussion moved on to how RMDP could be integrated into them. The 'management' aspect was discussed mainly in terms of a messaging system that would aid the users in their interaction with the repository and, in so doing, improve the service that SWS provide. Discussion of 'preservation' was largely confined to matters of format support. Until now, SWS has dealt mainly with BBC materials for which the BBC retain responsibility for long-term preservation of the media and its core metadata; as SWS take on more non-BBC material, preservation becomes a more serious issue for them.

As this discussion progressed, it became increasingly clear that the system SWS hoped to establish had much in common with the repository model developed by the University of Hull's RepoMMan project: that of a 'my repository' space feeding materials through a staging area into a public repository. A member of academic staff, working in a 'my repository' area might create a record with basic metadata and attach the media. This object then passes to a staging area where a member of the SWS team or GCU librarian checks and enhances it. The object is then made publicly available. This mediated deposit would allow a check on metadata and rights. SWS need to ensure that rights and authorisation are fully implemented at object level; possibly the rights agreements need to be in machine readable format (ODRL)?

There was a suggestion that users of SWS ought to have the ability to 'feed back' user-generated metadata into an object, this would probably take the form of tagging done in addition to the authoritative record *[cf CLARET Project]*. This provoked discussion of different levels of authorisation for different levels of annotation and where third-party annotations might be stored; no decisions have yet been taken. *[Shibboleth implications?]*

There was also discussion about depositors being able to check records before going live e.g. a discipline specialist creates an initial 'pre-record' consisting of a limited set of metadata fields; an information specialist then adds/checks metadata and rights clearance issues. The initial depositor is then given the option now or in the future to change metadata in the basic fields (this would also generate an alert for the record to be checked again). The depositor would only be allowed to change a limited number of fields.

2.7.5 Alerts - general

Discussion of alerts (messages from the repository to a user or manager), led to a number of general points:

- There are at least three types of alert: date driven (five years has elapsed since..., the monthly scan reveals that...), event driven (the metadata on this object has been automatically updated), and alerts that may come in the form of a status report rather than a simple message (for instance a periodic scan and count of formats held, the most/least popular item, ...) *[is this a third type of alert, or is it a version of date-driven?]*

- There may be classes of alert for classes of people which generate classes of response. Classes of alert may have default actions associated (including, for instance, hiding an object if an alert isn't dealt with).

- Alerts need to have an associated priority level: high, normal or low. Take down issues, for instance, would be high priority; status reports would probably be low. A manager, in particular, may get a number of reports each day, it is important (s)he be able to prioritise them.

- Messaging needs to be a two way process so that the system knows a user has responded appropriately to an alert, and so clears/resets it, or that the user has deliberately chosen not to react (I don't care that the object is five years old, tell me in

another five). There was discussion of 'Jabber' (XMPP) as a means of managing these mechanisms; this would allow for two-way interaction which RSS/Atom would not.

2.7.6 Specific alerts in the workflow

At the time of the two meetings covered by this report, SWS are still developing their ideas for an enhanced workflow covering RMDP and a Fedora repository; however some specific alerts were discussed and the idea that when a 'proto-object' is created in the repository, one of its datastreams should map out a skeleton workflow - this probably implies a form of metadata to describe the workflow and events within it. Examples of alerts included:

Date

- it is four weeks since you requested the media for this proto-object
- this (proto-)object appears to have stalled in the workflow
- revision/review/update overdue
- key date in lifecycle reached
- dark archiving suggested (objects with a finite useful timespan)
- object in restricted security space too long (one made available on a limited basis for a particular reason)
- share open too long (similar to above, but a share on a restricted item for a specific, limited purpose)
- format check due

Event

- media you requested has arrived at SWS
- media you requested is now available for private use
- media you requested is publicly available

Status

- the repository now contains nnn objects
- the repository contains nnn objects of the following file type

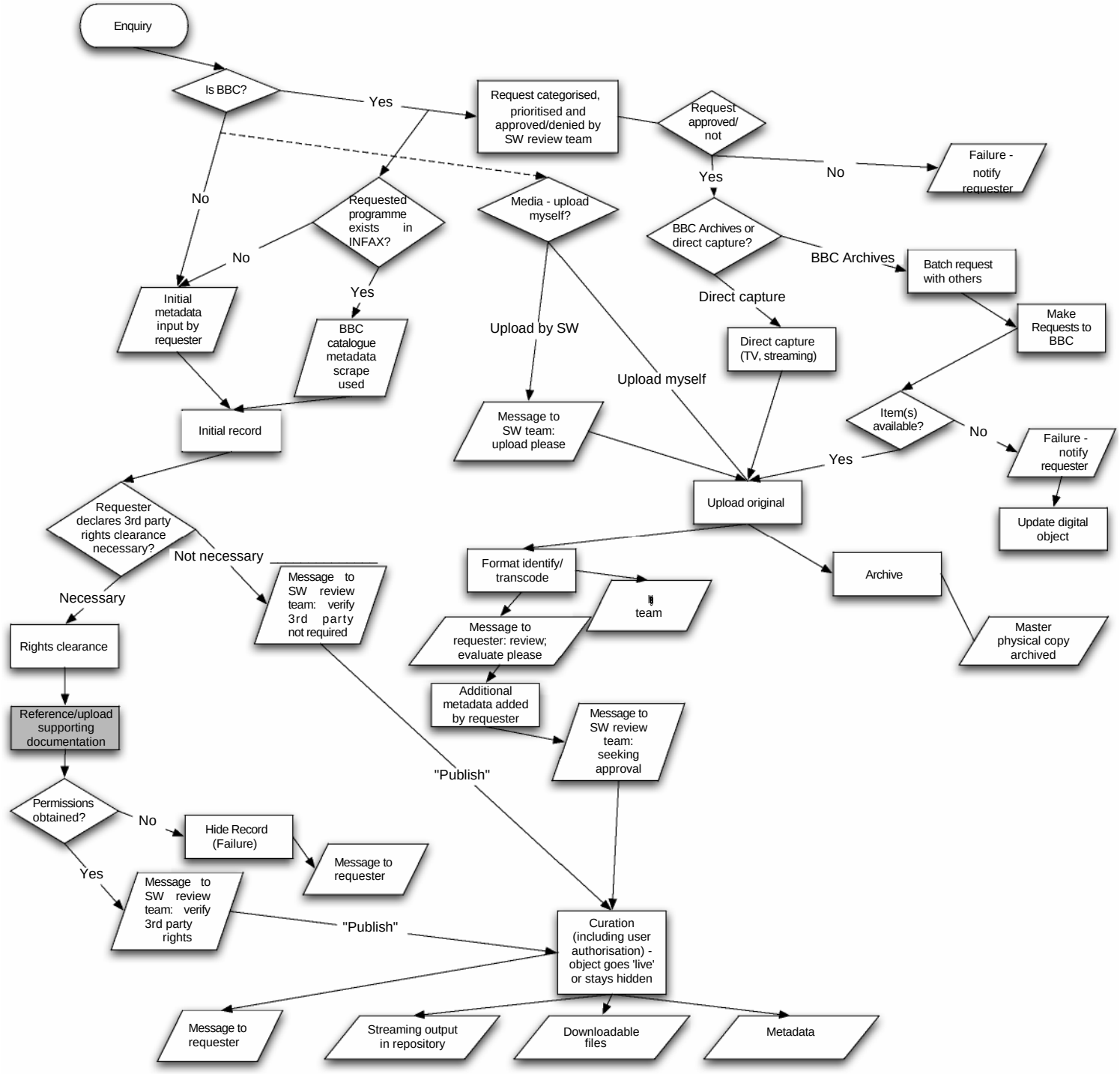
2.7.7 Other thoughts

In no particular order

- what form of identifier will the 'new' SWS repository use?
- if it is to be kept up to date, the metadata for the object needs to be updated each time it is used. Is there an overlap here with RIDIR - tracking use backwards and forwards?
- how are clips formally identified? Standards? RIDIR again?

2.7.8 Idealised workflow

The diagram overleaf shows an idealised messaging-enabled workflow for the improved SWS system:


Enquiry
Is BBC?
Yes
No
Request categorised, prioritised and approved/denied by SW review team
Request approved/ not
Yes
No
Failure - notify requester
Requested programme exists in INFAX?
No
Yes
Media - upload myself?
BBC Archives or direct capture?
BBC Archives
Batch request with others
Direct capture
Make Requests to BBC
Initial metadata input by requester
BBC catalogue metadata scrape used
Upload by SW
Upload myself
Direct capture (TV, streaming)
Initial record
Message to SW team: upload please
Item(s) available?
No
Yes
Failure - notify requester
Update digital object
Requester declares 3rd party rights clearance necessary?
Not necessary
Necessary
Upload original
Format identify/ transcode
team
Archive
Message to SW review team: verify 3rd party not required
Message to requester: review; evaluate please
Master physical copy archived
Rights clearance
Additional metadata added by requester
Message to SW review team: seeking approval
Reference/upload supporting documentation
"Publish"
Permissions obtained?
No
Hide Record (Failure)
Message to requester
Yes
Message to SW review team: verify 3rd party rights
"Publish"
Curation (including user authorisation) - object goes 'live' or stays hidden
Message to requester
Streaming output in repository
Downloadable files
Metadata

2.8 Development of the RepoMMan tool to 'publish' materials

The RepoMMan tool allows users to interact with the University's repository to use it as a workspace for developing and storing materials. The finished tool will have a 'publish' button which allows a user to start the process of moving an object (simple or compound) from their private space to some level of public exposure. It seems reasonable that REMAP should, at that point, start interacting with the object.

When a user clicks 'publish' the object will be cloned from their private space and the copy will be placed in a quarantined area of the repository. The process will generate a new internal identifier for the cloned object and the object itself will be 'owned' by the repository and not by its original author. Although the author can have no editorial interaction with the new object, he or she will want to be reassured that it is making its way towards the publication they have requested.

The new object will pass through a number of procedures during which it will be manipulated so that it conforms to one of the University's formal digital content models, including preservation metadata and management datastreams. Ultimately it will emerge from the processes, ready to be posted in a public, or semi-public, space. At this stage in the development of the services that will be involved it seems likely that the object as finally posted will have yet a third internal identifier within the repository.

It is suggested that the following events should be handled by REMAP.

(1) Confirmation message to the author that a copy of their material has been successfully received by the repository for 'accession' and is being dealt with.

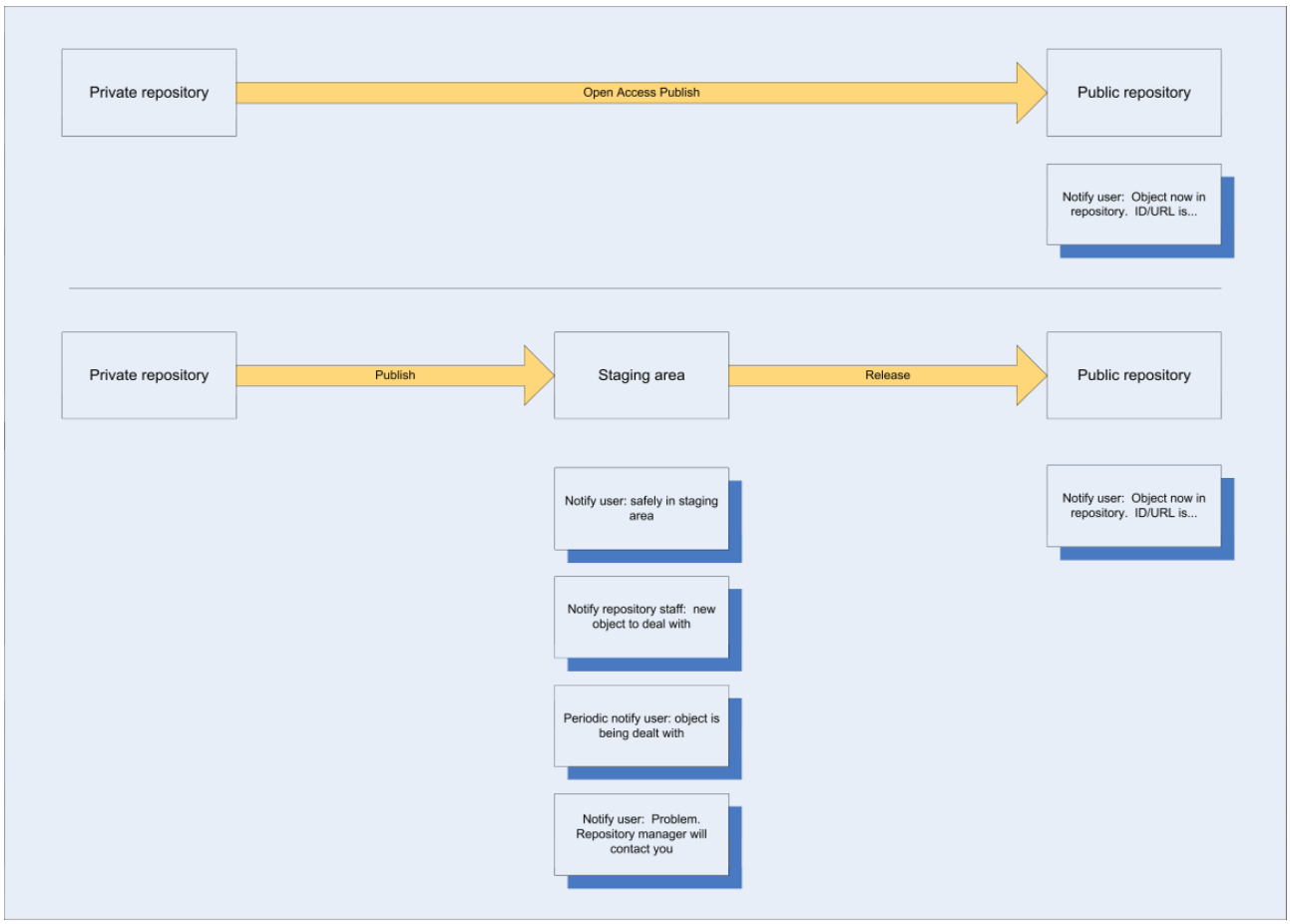
(2) Notification to repository personnel that the object needs action

Initially it is likely that repository staff will manually initiate and check the conversion of the object and check (including a rights check) the resulting output which will be a third object with a new internal identifier. They will then 'publish' this object into the (semi-)public repository. In the longer term, the conversion could be automatic but is still likely to require a human check before publication is permitted. An exception may be a facility for the author to do 'open-access' publishing themselves, without an intermediate check, but in this case the user will need to indicate that they take responsibility for the publishing process and any consequences that may flow from it.

(3) The original author should now receive confirmation that the object is published and (probably) its new identifier. The identifier, or the URL associated with it, may be useful as a hyperlink - perhaps in an author's own website.

(4) The conversion process should set object RMDP flags, depending on the type of object. Some of these flags may need to be set up by the author or at the checking stage.

Flags and alerts mentioned elsewhere will be useful, for instance "this object has been in the queue for a long time".



2.9 Development of repository functionality to support electronic theses and dissertations (ETD)

The process for supporting ETD will probably be established in two phases. In the early years of ETD support it is likely that electronic versions of theses etc will be gathered by the Graduate School and passed to the University Library alongside the bound, printed copy that will still be required. Library staff will then deal with the ETD using the CD-ROM or DVD provided. Once procedures are better established it is possible that in a second phase postgraduate students may be allowed to submit the ETD directly to the library using the RepoMMan deposit tool or something similar. At this stage it is possible that the requirement for a bound, printed copy of the document may be removed although such a copy could have merits from a long-term preservation point of view.

However the ETD makes its way into the repository staging area, the needs in terms of messaging and RMDP are much the same. As far as the publication process is concerned this will follow the outline in 2.8 above; it should be noted, however, that the process will need a specific message or message component to acknowledge that an embargo has been set should this be a condition of submission.

Once in the repository the ETD becomes subject to the normal requirements of RMDP within it.

2.10 Development of procedures specifically to support digital archives at the University of Hull

Use cases relating to digital archive were developed with Chris Awre, as part of the REPOMAN project:

- ARCH001: Transfers to digital archive as part of digital object records management

- ARCH002: Transfers to digital archive following independent review (eg against classification list) and decision

These analysed processes involved in accessioning and managing archives in an analogue environment and mapped corresponding pre-conditions, triggers and steps in a digital environment as a first stage towards identifying user requirements.

Based on OAIS functions, these user requirements include:

Ingest

An enhanced ingest process is required to treat differently objects being added to the digital archive. Ideally this additional interactivity would be available only to individual users authorised by the archivist. Additional metadata is required, to ensure authenticity and integrity of the object; to record the context of its creation in detail and to enable preservation requirements to be determined. Such ingests should trigger a message alerting archive staff that material has been added and requiring a response from them to 'accession' the material. This would include metadata checks, and the addition of metadata to manage preservation and to record the accessioning process itself. A facility for users or archive staff to ingest series of objects/documents together and maintaining the links between them would be desirable.

NB in the two use cases developed within REPOMAN the digital object would already be in a digital repository. This would apply in the case of Committee papers or academic programmes (use cases described above), where documents stored in the repository as part of records management were then 'transferred' into the digital archive for permanent preservation and open access. In these cases the requirements for additional metadata, messaging and batch processing described above would still apply, even though they would not be part of the ingest process.

Storage

Following ingest the preservation version of an object should be stored in a 'dark archive' with a replica as an access copy if appropriate. File conversion may be required at this stage, to a recognised preservation format. Some degree of structure within the dark archive would be desirable, to reflect macro-cataloguing (see below). The storage medium should be robust - eg with no single point of failure.

Access

Requirements for access (ie to replica access copies) are largely the same as for other objects within a digital repository. In some cases a messaging facility would be desirable, for potential users to request access to individual objects or series of objects and for archive staff to grant or refuse this access. Similarly the capability to require users to read terms and conditions (eg re copyright or confidentiality) and to indicate acceptance of them prior to access to certain objects would be desirable.

Preservation

A digital archive requires more robust and detailed preservation planning, including a migration strategy to combat obsolescence of software or hardware and a system of regular audits of the status of individual items, so as not to rely on user access to detect errors

Data management

Requirements for data management specific to a digital archive reflect the need to record the hierarchical structure of records series and the interrelatedness of individual documents. Requirements might include the use of a classification scheme to list or categorise objects and

link them in a structure; and both micro- and macro- cataloguing (within repository metadata, or ideally via links with CALM - existing software used for archive cataloguing).

Administration

Administrative requirements for a digital archive would be broadly similar to other use-cases for a repository. Requirements linked to records management would include messaging triggered by flags attached to objects already stored in the repository (ie not ingested as archives) alerting owner and archivist that review is due for possible 'transfer' into digital archive. An additional requirement would be the creation of report logging all activity relating to objects within the repository designated as archives.

2.11 Development of an integrated web presence for the newly formed Centre for Spirituality Studies

The Centre for Spirituality Studies is new at the University of Hull for the academic year 2007/08. Their need for an integrated web presence offers the REMAP team the potential to design an RMDP-enabled solution starting with the rare 'blank sheet of paper'.

3 User needs

Objects

The examples outlined above identify two types of object that will be held in a repository and form the basis for various types of RMDP.

The simplest of these is a digital object in the normal sense that most people probably think of it: some sort of digital entity that embodies binary content, such as a document or image, and other information probably in the form of metadata that provides added value around that content.

All repositories are likely to contain collections of objects and therefore a 'collection object' acting as the parent, however the examples above have also shown the potential usefulness of having these parent collection objects serve as the management point for a number of similar and/or related child digital objects. The parent objects would have no binary content, in the manner described in the previous paragraph, but could contain common metadata for all the objects in the collection and, importantly, could contain RMDP flags and information for its child members as a group. It is worth noting that there may be occasions where a collection of one object is desirable and that, at times, there may be 'empty' collection objects containing RMDP triggers for relating to child objects yet to be created.

Taking all the examples above it is possible to envisage a range of RMDP alerts. The following represents an attempt to categorise these, together with examples relevant to the needs gathering work:

Events

- information only: an event has taken place
    - *generally: FYI, this has just happened*
    - the object you requested from xxx has been deposited (administrative)
    - the item you requested is now available in the repository (end user)
    - the item you requested cannot be made available (end user)
    - the item you submitted for publishing has been successfully placed in the queue
    - the item you submitted for publishing is now available at: URL
    - the item you submitted for publishing is now successfully in the repository and will be released from embargo on dd/mm/yy
    - the repository has taken the following default action on object xxx
    - an open-access deposit has been made at: URL

- requiring action
    - *generally: 'this' has just happened, 'that' now needs to be done (workflow sequence)*
    - an item has been deposited in the accession queue for attention
    - an item submitted on dd/mm/yy is still in the publishing queue
    - the following item is still in the publishing queue awaiting action by its author requested on dd/mm/yy

Dates

- information only
    - it is now xx days/weeks since you requested deposit of xxx (administrative)
    - document xxx has not been deposited. The deadline was dd/mm/yy (administrative)
    - the object abcd appears to have stalled in a workflow since dd/mm/yy

- requiring action

  - revision/review/update due
  - specified lifespan reached. Hide?
  - limit of embargo reached. Expose?
  - object exposed in restricted security space too long. Hide?
  - restricted share open too long. Remove?
  - format check due

Status

- the repository contains nnn objects
- the repository contains nnn objects of each of the following file types
- examination papers have been deposited for xxx, xxx but not for yyy, yyy

Other issues

Particular notification or messaging process(es) may need the following:

- take default action
- snooze for xx days
  - this suggests the need for two dates, the original notification date and the deferred date
  - perhaps it should be possible to set a maximum on the cumulative length of snooze
- actions required by a collection trigger need to show granularity
  - hide these but not those
  - yes to all
- messages and alerts should have an 'importance' attached to them so that busy administrators know which are the most urgent

4 Identified use cases

Following the needs gathering process, the REMAP team have considered what use cases should be taken forward to use as exemplars within the project and have attempted to assign a priority order to them.

The need for a priority order arises from the fact that JISC funding for this project must finish on 31st March 2009 and so there is no scope for a project extension. The idea of prioritisation is that we should aim to demonstrate a range of potentially important and useful outcomes of our approach to repositories and RMDP rather than take fewer, more complex, examples and so limit the range of what we might be able to show.

The sections that follow are a brief overview of our intentions which will be specified much more fully in the follow-up document which provides workflows for these REMAP processes.

4.1 RepoMMan

The RepoMMan needs will be the first addressed. In the first instance these needs call for pure notification (passive messaging) but require a Fedora object to be built with a 'flags' datastream. Later stages in the process (eg, 'this object has been waiting for accessioning for a long time') offer the possibility of needing to deal with 'snooze' functionality.

4.2 ETD and Past undergraduate examination papers

These two user needs studies between them call, in the first instance, for relatively straightforward extension of the processes developed at 4.1. Specifically they may require secondary actions to take place after a period of time. The ETD example may require the ability to deal with lifting an embargo after a period of time, whilst the relevance of older examination papers will need review after a period of time. The examination paper example also offers the possibility of dealing with flags at a collection, rather than individual object, level.

4.2.1 Electronic Dissertations and Theses (ETD)

When an ETD is deposited there is the possibility than an embargo is associated with the document. If this is the case, a message will need to be generated at the time of the embargo release to query whether it should go ahead. This will probably be an alert with a default action (release) but offering the possibility to extend the embargo. Either way, this alert will generate a two-way process because repository staff need to respond to it in some way before any action is taken.

4.2.2 Past undergraduate examination papers

At a superficial level this is a similar example to ETDs. It is anticipated that examination papers should be reviewed five years after they were set to check that their content is still relevant to current students; it is likely that most will be hidden from students at this point whilst still remaining visible to staff. However the example is somewhat more complex than the ETD case. Examination papers will be grouped (collected) by examination season. To prevent a departmental secretary being faced in summer 2012 with alerts from all their subject's past papers from summer 2007, we shall investigate having the collection object send a single alert for all the papers within it. This raises the need for some discrimination in the user response which may need the user to be able to select a number of papers to hide whilst resetting the review date on others.

4.3 Register of Policies and Procedures (RPP), Committee Section and Quality Handbook

All these three needs studies have similar components and potentially have a need for more complex interaction than the preceding use cases. The RPP requires periodic review and the process of review itself is a workflow that may be capable of oversight through the use of alerts and responses. However, the involvement of a group of people in the process will require the system to be able to forward alerts for information.

The particular setup at the University of Hull means that the RPP is managed by one of the project's contributors and is thus an ideal test ground for this more complex procedure. Once we are satisfied that the system is robust it can be tried by Committee Section and by the Quality Office who have very similar needs.

4.4 Archives

The preceding examples are substantially to do with records management and preservation rather than the specific needs of archiving. In the later stages of the work dealing with them we will consider what it means when a digital object is transferred from being 'live' in the repository to being an archived object. A number of potential issues arrive which will be addressed in a pragmatic way as the project develops. An object may possibly be archived out of public view but more likely in public view. In this latter case action will need to be taken to make the object visible through archives cataloguing. In some cases the transfer to archive status will take place according to a predictable timeline, in other cases not. Materials transferred to the archive are there for potentially very long-term preservation and we will need to consider whether additional preservation flags need to be activated at this point. It is perhaps the case that some compound objects should be considered for 'bursting' so that, in the archive, their different components can be handled separately.

4.5 Spoken Word Services

Spoken Word Services have a needs set that sits alongside all the preceding cases. It ranges from passive messaging to complex messaging involving interaction with repository objects and, ultimately, even archiving. SWS will act as a 'sounding board' for approaches developed at Hull to ensure that they are more widely applicable than just the University campus.

4.6 Centre for Spirituality Studies

As noted above the Centre for Spirituality Studies represents a blank canvas for RMDP. The ideas developed in the previous use cases will be tested with staff in the Centre to ensure that they cover the needs of an 'unseen' user group as well as the needs of those specifically involved.