This infoKit reflects the increasing use of repositories using the documentation, guidance and expertise built up during the Repositories Support Project (RSP). This has been augmented by Lou McGill.

Contents

What is a repository?

A digital repository is a means of managing, storing and providing access to digital content. Repositories can take many forms, and all sorts of websites and databases could be considered to be repositories. For the purposes of this infoKit, we will primarily be looking at institutional repositories.

"What goes into a repository is currently less an issue of technological or software ability, and more a policy decision made by each institution or administrator."

Putting digital content into an institutional repository enables institutions to manage and preserve it, and therefore derive maximum value from it. A repository can support research, teaching, learning, and administrative processes. Although many institutional repositories are primarily established for the benefit of the organisation and its users, there is an increasing movement towards 'open access' to the wider community, sometimes in a global sense. The JISC/HE Academy UKOER (Open Educational Resources) Pilot Programme[1] included several projects that utilised institutional repositories to make their learning and teaching materials openly available. Issues around this are discussed in the OER infokit[2] which also includes links to the outputs of these projects.

Digital repositories may include a wide range of content for a variety of purposes and users. What goes into a repository is currently less an issue of technological or software ability, and more a policy decision made by each institution or administrator. Institutional repositories are often referred to in strategy and policy documents as they can support key institutional aims and objectives. Typically content can include research outputs such as journal articles or research data, e-theses, learning and teaching materials, and administrative data. Some repositories only store particular items (such as theses or journal papers), whilst others seek to gather any credible scholarly work produced by the institution; limited only by each author's retained rights from publishers. However, some more complex objects such as websites, content-packaged learning objects, 3D topographical representations and other data sets may present technological and management challenges.

One of the advantages of a repository is that each piece of content can be described in some detail via the input of associated 'metadata'. This acts much like a catalogue record in a library management system and allows searching across items within the repository. If the repository has implemented an appropriate metadata exposure method (such as Open Archives Initiative Protocol for Metadata Harvesting and/or RSS) the metadata can then be harvested by external services and exposed to the wider world. Repositories use open standards to ensure that the content they contain can be searched and retrieved for later use. The use of these agreed international standards allows mechanisms to be set up which import, export, identify, store and retrieve the digital content within the repository.

Types of Repositories

Digital Repositories

When it comes to managing electronic and online resources there are alternatives to repositories, such as virtual learning environments (VLEs), wikis and other informal content sharing applications, but the focus of this infoKit is on formal repositories. In simple terms, a formal digital repository is where digital content, or assets, are stored and managed to facilitate searching and retrieval for later use. A repository supports mechanisms to import, export,

[1] http://www.jisc.ac.uk/oer

[2] http://openeducationalresources.pbworks.com/

identify, store, preserve and retrieve digital assets. It is also increasingly being recognised that repository use can be encouraged through additional services to encourage community sharing and exchange of both practice and content.

The type of content held in a repository can have significant impact on the way it is both designed, managed and used. There are some very active communities working around different types of repository - each of which has its own history, culture and terminology. Where possible this infokit will highlight sections that refer to specific repository types.

Institutional Repositories

Repositories can be many shapes and sizes, from small specialist collections to national or international services. The focus of this InfoKit is on repositories managed by Further and Higher Education Institutions, at departmental or institutional level, which implies a certain level of commitment and intention to embed repository use and management into everyday work. Many institutional repositories initially focussed on research outputs and some still limit their collections to this type of content. Others have started to widen the original remit to include learning and teaching materials. Whilst institutional VLEs have, to some extent acted as stores for learning and teaching materials, they tend not to support the search and retrieval functions required for a repository. Making this content more open, even within the institution, presents challenges for institutions with a commitment to open up their resources.

Open Access Repositories

Repositories are increasingly being made more 'open' to make content accessible to wider user groups, sometimes at a global level. Not all repositories are open: some are designed to support sharing within a specific group and are sometimes described as 'closed'. These repositories often require authentication and some have varying levels of access and 'degrees of openness'. Most of the advice in this infoKit is relevant to open and to "closed" or authenticated repositories. The advice in this infoKit is also not limited to managing research papers online, but applies to learning resources, research data sets, committee papers and other content types.

"Open access" is a term that is used in a specific sense and most often used in relation to collections of research papers. Open access is discussed further in the **drivers** section. The concept of open access repositories is closely linked to open access research papers, but the idea has wider currency and is linked to concepts of open educational resources (OERs) and open content licensing.

Benefits of a Repository

Digital repositories have great potential for value added services and offer a range of benefits to researchers, teaching academics, learners, institutions, the global research community and the wider world. Some of these benefits will have been key drivers for the development of a repository whilst others may be unanticipated. Different repository models (research only repository, learning and teaching materials repository, mixed content repository, open repository) will have different benefits for each stakeholder group.

For example: Open research repositories offer additional advantages by taking the results of research that has already been paid for and making it freely available online. This process can have significant advantages for individual authors, for researchers, for institutions and for the process of research generally by allowing improved management of intellectual outputs and freeing up the process of dissemination.

The following table identifies possible benefits for different groups:

Research outputs repository	Learning and Teaching materials repository
Benefits for the global community (open): Assists research collaboration through facilitating free exchange of scholarly information Aids in the public understanding of research endeavours and activities	**Benefits for the global community (open):** Supports re-use and re-purposing Supports community input to metadata through tagging, notes, reviews Supports development of effective retrieval through professionally created metadata Ensures trust through appropriate licensing Supports the sharing and re-use of individual assets Supports the sharing and re-use of complex learning resources Helps to develop critical mass of materials in particular subject areas Cost efficiencies Decrease in duplication Provides access to non educational institutional bodies such as employers, professional bodies, trade unions, etc.
Benefits for the institution: A repository can interoperate with other university systems and maximise efficiencies between them by sharing information A repository can increase the visibility and prestige of institution (depending on content contained) Repository content is readily searchable both locally and globally Allows an institution to manage their intellectual property by raising awareness of copyright issues and facilitating the recording of relevant rights information A repository that contains high quality content could be used as 'shop window' or marketing tool to entice staff, students and funding Repositories can store other types of content that isn't necessarily published, sometimes known as 'grey literature' A repository may be an important tool in managing an institution's research assessment or quality assessment submission Repositories could provide cost savings in the long run provided that a significant amount of content is deposited in them Offers greater flexibility over websites with better security and preservation of various kinds of digital materials through the collection of standardised metadata about each item	**Benefits for the institution:** Maintaining and building on institutional reputations nationally & globally Attracting new staff and students to institutions Increased transparency and quality of learning materials Supports sharing across/between departments within institutions and interdisciplinary cross-fertilisation Shares expertise efficiently within institutions Supports modular course development Supports storage, management, preservation, attribution and retrieval of student content Easily incorporated with institutionally-owned technologies Supports the altruistic notion that sharing knowledge is in line with academic traditions and a good thing to do Likely to encourage review of curriculum, pedagogy and assessment.

	Supports preservation of learning resources. Facilitates presentation of resources for accreditation bodies Enhancing connections with external stakeholders by making resources visible
Benefits for the researcher: Increased visibility of research output and consequently the department and the institution Increased impact of publications. Research made freely available can be disseminated more widely and have greater impact. Work done on **citation analysis**[3] has demonstrated that research that is made freely available will be easier to cite Offers usage metrics so researchers can determine hit rates on specific papers Creates the potential to undertake citation analysis through following links to papers held in other repositories In fast moving subjects such as Electronics, researchers can make preprints (as opposed to peer reviewed papers) available via a repository, to establish that they were first and to get feedback Helps researchers manage and store digital content connected with their research, including the underlying research data Helps researchers manage any requirements of funding bodies for publications to be made available in a repository Provides the possibility to standardise institutional records e.g. an academic's CV and published papers Allows the creation of personalised publications lists	**Benefits for those supporting teaching & learning:** Supports sharing of knowledge and teaching practice Encourages improvement in teaching practice Offers one-stop access point for staff Encourages multidisciplinary collaboration and sharing Supports CPD and offers evidence of this Increased visibility within the institution and possibly their subject discipline community Reward and recognition from the wider community if made open
Benefits for learners: Provides access to the latest research (especially useful at postgraduate level)	**Benefits for learners:** Increased access options for students enrolled on courses (particularly remote students) Increased access for non-traditional learners (widening participation)

[3] http://opcit.eprints.org/oacitation-biblio.html

Content Types

Digital repositories may include a wide range of content for a variety of purposes and users. The focus of each repository, and, therefore, what content it will store, will depend upon the policy decisions made by each institution or administrator. To give an idea of the type of content currently held in repositories worldwide, the **Directory of Open Access Repositories- OpenDOAR**[4] has identified the following as the most common content types in repositories:

- Journal articles
- Bibliographic references (metadata only)
- Books, sections and chapters
- Conference and workshop papers
- Theses and Dissertations
- Unpublished reports & working papers
- Datasets
- Content-packaged learning objects
- Multimedia and audio-visual materials
- Software
- Patents
- Other special items

Within a repository's Content Policy there may be a desire to clarify the content types to be included as there may be areas of overlap between the types listed above. The list below, adapted from the default content type list provided with EPrints software may be a starting point for such clarification.

Publications

- **Articles** - Articles in journals, magazines, newspapers. Not necessarily peer-reviewed. May be an electronic-only medium, such as an online journal or news website
- **Books** - Complete books or conference volumes
- **Book Sections** - Separately-authored chapters or sections in books
- **Conference or Workshop Items** - Papers, posters, speeches, lectures or presentations given at a conference, workshop or other event. If the conference item has been published in a journal or book then please use 'Book Section' or 'Article' instead

[4] http://opendoar.org/

Theses & Dissertations

Student theses and dissertations submitted to an institution as part of the requirements for a degree, including:

- **Doctoral Theses** - PhD
- **Masters Theses & Dissertations** - MSc, MA, MBA, etc
- **Bachelors Dissertations** - BSc, BA, etc

UK Examples

Aberystwyth Theses Guidance
http://www.aber.ac.uk/en/quality-assurance/pg-research-handbook/section-8/section-8-2/

International Examples

Utah State - **Approval Form**, **Checklist**

http://bit.ly/qGPdaB

Resources used to support teaching and learning

- **Curricula and syllabuses**
- **Course validation documents**
- **Course materials** - learning resources, lecture notes, learning exercises
- **Assessment materials** - tests, exam papers
- **Study skills, support and revision materials**
- **Student-produced work** (including all the content types listed here)

Audio Visual Items

- **Images** - Digital photographs or visual images
- **Video** - Digital video
- **Audio** - Sound recordings
- **Show/Exhibition** - An artist's exhibition or site-specific performance-based deposit
- **Artefact** - An artist's artefact or work product. Could also apply to archaeological finds
- **Performance** - Performance of a musical event
- **Composition** - A musical composition

Data

- **Datasets** - Bounded collections of quantitative data (e.g. spreadsheet or XML data file)
- **Experiment** - Experimental data with intermediate analyses and summary results

Grey Literature

- **Patents** - A published patent. As-yet unpublished patent applications should never be included in a repository, because to do so might disqualify the application

The following are grouped together by EPrints as Monographs, although many people use 'monograph' as a synonym for 'book'

- **Technical Reports**

- **Project Reports**
- **Documentation & Manuals**
- **Working Papers & Discussion Papers**

Other

Something within the scope of the repository, but not covered by the other categories.

Accessing Content in Repositories

A key factor in understanding your institutional repository, and the benefits it offers, is to gain an awareness of searching for repository content. Most open repository content can be found by conducting a simple internet search. 'Closed' repository content may be accessed through an authenticated access mechanism. However, a number of key services have been developed to support searching across repositories. These typically utilise the **OAI-PMH** protocol which enables more sophisticated searching and more relevant results than is possible from general online search engines. By providing access to high-level functions that span entire networks of repositories services such as these facilitate deposit and reuse of content.

- **BASE**[5]: is a multidisciplinary search engine based in Germany that allows users to search for materials in OA research repositories, OA journals and other resources worldwide. All repositories and websites are reviewed to ensure that only high quality open access scholarly resources are included in the BASE search
- **OAIster**[6]: is another multidisciplinary search engine of OA research repositories, journals and other resources worldwide, based in the U.S. All repositories and websites are reviewed to ensure that only high quality open access scholarly resources are included in the OAIster search
- **OpenDOAR Search**[7]: OpenDOAR uses Google to provide a simple full-text search of all repositories listed in OpenDOAR. The research repositories have been quality-checked for academic relevance
- **DART-Europe**[8]: the DART-Europe search service currently provides access to theses from over 20 European institutions
- **Intute Repository Search**[9]: searches across approximately 89 UK Further and Higher Education repositories. Searches may be limited to a subset of these repositories for more specific searching. These repositories include learning and teaching content as well as research.

[5] http://www.base-search.net/

[6] http://www.oaister.org/

[7] http://www.opendoar.org/search.php

[8] http://www.dart-europe.eu/

[9] http://www.intute.ac.uk/irs

Exporting and reusing repository data

Once a repository holds items, it may be the case that the information can be reused by other systems and for other purposes. By reusing the data already entered, cost savings may be made both for the host institutions and those accessing the data/resources. In order for the data to be reused it first needs to be exported.

Exporting the data

There are different ways that the information can typically be exported:

- **OAI-PMH**: The standard method for remote systems to harvest the contents of a repository is by using the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH). This is explained fully in **OAI-PMH Harvesting**
- **RSS**: Repository content can be made available to others using RSS exporting tools. Whenever you add or modify content in your repository, your RSS feed is automatically updated and thus the changes are syndicated to all relevant parties. RSS formatting tools can then be to incorporate the repository data into the destination location
- **Repository platform export tools**: Each repository platform will have an export feature that allows the data to be exported. These tools will vary from platform to platform, but will typically allow the bulk export of all or some items, including their metadata and files
- **Custom tools**: If you have specific requirements that cannot be fulfilled by either OAI-PMH or the standard export tools, with suitable development effort you may be able to export the data in a different manner

Whichever way data is exported for reuse, it is likely that some data manipulation scripts or effort will be required to convert the data into the required format.

Uses for the data

There could be many different uses for the data contained within a repository. Some examples include:

- **Publications/bibliographic databases**: Many institutions maintain publication information databases. The data held within the repository can be used to populate these
- **Personal web pages and automated publication lists**: Some institutions generate publications web pages for their members on the basis of content in the publications database. Repositories can support this function. Consideration needs to be given as to whether full-texts and metadata that are made available from staff home pages should link back to the repository in the first instance
- **Virtual Learning Environment**: The repository could be used to store and manage learning resources which can then be referenced from within the VLE. Increasingly institutions are investigating the transfer of some learning materials from the VLE to an institutional repository. This makes the content more open and accessible either within the institution or in a fully open way to the wider community (depending on policies relating to openness of learning and teaching materials).
- **Marketing Systems**: Directed advertising on the repository pages can attract new PhD applicants and, potentially, new staff

Interoperability and Integration

Interoperability and integration are terms that describe the ways in which repositories work with other systems using common standards and protocols. It is important that each repository interoperates with other systems in order for the institution to reap all of the rewards that come from the sharing of information. Integration refers to the process of setting up your repository to work with other systems, so that data can flow from one system to another. Without common standards, it would be difficult to integrate different systems. Information flows between interoperable systems by using interfaces which provide routes into, and out of, information systems. Sometimes these interfaces are used directly by people (e.g. web user interfaces or RSS feeds) and sometimes they are used by machines (e.g. **OAI-PMH** and **SWORD**). Interfaces used by machines are sometimes referred to as m2m (machine-to-machine) interfaces.

There are different ways in which your repository can interoperate with other systems. Three types of integration are:

1. Integration with external systems to get items in to a repository: While repositories are often populated with items that have been submitted using the repository software, there are many cases where the information can be gathered from external systems. A common use-case is to populate the repository from an institutional publications database. Some institutions are investigating ways to populate their repository with learning and teaching materials from their VLE in an effort to make them more accessible both within the institution and sometimes in a more open way with the wider community. Another way of working is to provide depositors with desktop-based smart deposit tools that integrate with their working environment to help capture their work as it is created. The most widely adopted standard for depositing items into a repository is **SWORD**

2. Integration with systems to get items out of a repository: Once a repository contains a useful corpus of items it can be integrated with other systems that want to use that data. These may be local systems such as institutional search engines or researcher web pages, national systems such as **EThOS**[10], or international systems such as **Google Scholar**[11] or **OAIster**[12]. One of the most common methods for extracting the structured metadata of the items in repositories is 'harvesting', with the standard protocol being the Open Archives Initiative Protocol for Metadata Harvesting (**OAI-PMH**). **RSS feeds**[13] are another standard mechanism that allow repositories to provide information to other systems; in this case, RSS feed readers.

3. Integration with systems that provide services to a repository: Repository software specialises in storing items and metadata,

Support for each of these integration options varies with each repository software platform, but the majority now support integration with the following standards:

- **OAI-PMH**
- **SWORD**
- **RSS/Atom feeds**
- **Shibboleth**
- **LDAP/Active Directory**

[10] http://www.ethos.ac.uk/

[11] http://scholar.google.com/

[12] http://www.oaister.org/

[13] http://en.wikipedia.org/wiki/RSS_%28file_format%29

but can often work more effectively if it makes use of services provided by other systems. One common system that repositories are often configured to work with is local authentication systems such as **Shibboleth**[14], **LDAP**[15] or **Active Directory**[16]. These services allow the repository to look up usernames, passwords, and user details (name, email, telephone number etc) from a centrally managed system. Other systems that may provide services to a repository could include file format validation (**JHOVE**[17]), virus scanning of ingested files, or external cloud storage of files

Alternatives to Repositories

The decision of whether to use a repository for managing digital content is often challenging as it is highly likely that digital assets already reside in many types of systems. These range from locally-written closed systems to the open web. In the modern Higher Education information environment there are now typically many systems which each provide similar functionality and it is often a difficult decision to decide which system is the most suitable home for different sets of information.

One issue that may prevent busy academic staff depositing learning and teaching content is the fact that they may be required to deposit in many places, particularly if being mandated to deposit in a national and/or institutional repository. Academics may be more inclined to deposit in a subject community repository or system or just make their content available on the open web through a web site. Using an institutional repository or even a VLE that can feed content into other systems presents a more compelling reason to deposit. Any system needs to integrate with others in a way that results in 'once only' deposit.

Requirements Analysis

While content management systems (CMS), portals, Virtual Learning Environments or collaborative working environments could effectively fulfil some repository requirements, they do not typically provide the ability to expose digital content to the wider world. In addition, the ability to export and import bibliographic data with such systems is often poor or non-existent. Dedicated repository software provides a complete set of features. Before examining possible institutional information architectures and the potential inclusion of repositories into that architecture, it is useful to consider the varied *functional requirements* of an institutional repository. Different drivers have an impact on this and may even compete with each other.

> "Existing systems such as CMSs, portals, or collaborative working environments can fulfil some of the requirements [of a repository]. However they do not provide the rich tools for ongoing access and discovery found in repository software..."

For example there may be an institutional commitment to publish the intellectual output of the University on the open web which may clash with a desire to remain competitive and protective in terms of course content. Some institutions keep these two functions apart in different systems but more are taking the challenge of bringing the two together. Specific functionality that may be required by an institutional repository may include:

[14] http://en.wikipedia.org/wiki/Shibboleth_%28Internet2%29

[15] http://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

[16] http://en.wikipedia.org/wiki/Active_Directory

[17] http://hul.harvard.edu/jhove/

- **Open Access**: Some items may need to be made available online. This includes both the metadata, licensing/use information and the item in an appropriate digital format. Where items cannot freely be made available online, the facility to restrict access is required. Institutions may need to include different 'degrees of openness' for different items.
- **Search engine indexing**: Sites that store materials should be accessible to crawlers and indexers in order for items to be found by conventional search engines
- **Bibliographic metadata**: Items often come with associated metadata (title, author names, bibliographic citation etc). These details need to be stored along with the item, in a recognised metadata schema such as Dublin Core1 or MODS2. The ability to 'crosswalk' between **metadata schemas** is also important
- **Export functionality**: A common aim of repositories is to export citation details on a per-author basis to automatically generate researcher CVs or RAE forms, or more generally to export data into bibliography manager or data analysis tools. Good export functionality can be necessary to fulfil this requirement
- **Import functionality**: Some institutions have publication management systems or traditional websites and want to import that data into a repository for public access
- **Metadata harvesting**: Search system providers such as **Intute Repository Search**[18] and **OAIster**[19] use a specialist protocol known as the **Open Archives Initiative Protocol for Metadata Harvesting** (OAI-PMH) rather than normal web crawling, which cannot accurately harvest metadata. In order for a site to be indexed by specialist search services such as these, an OAI-PMH interface is required
- **Persistent identification**: In order to facilitate ongoing access to resources within a repository, repository software utilises tools to assign persistent identifiers. Some repository platforms make use of external identifier resolution services to ensure the identifiers would even persist if the domain name of the server or institution changed
- **Licensing section**
- **Content packaging**

Integration with other systems

Repository software solutions need not be standalone systems, and can work effectively when integrated with other systems in the information architecture of an institution. Two typical examples of repositories working alongside other systems might be:

- *Using a collaborative working environment such as **SharePoint**[20],* authors can create, manage and apply version control to work. The work can then be published along with its metadata in the repository for final dissemination

[18] http://www.intute.ac.uk/irs/

[19] http://www.oaister.org/

[20] http://sharepoint.microsoft.com/

- If an institution has web pages in its CMS about members of staff, research CVs (publication lists) can be automatically generated from items in the repository and imported into the CMS

Conclusion

Existing systems such as CMSs, portals, or collaborative working environments can fulfil some of the requirements identified in the list above. However they do not provide the rich tools for ongoing access and discovery found in repository software, nor do they typically provide OAI-PMH interfaces. Furthermore, the ability to export and import bibliographic metadata is often poor or non-existent.

Given that these characteristics are increasingly perceived as essential to making academic output widely available over the web, a strong case can be made for integrating a repository into institutional information architectures, or even using a repository to replace existing systems, as only dedicated repository software provides this complete set of features. The key to making this decision lies in a comprehensive assessment of functional **requirements** and comparative analysis of existing systems with the intended **repository solution**.

Drivers

The drivers for research repositories and learning & teaching repositories are quite different and have shaped the resulting services considerably. This can be quite challenging for institutions who aim to bring the two functions together into one repository. This is especially true if one or other such repository already exists in an institution. Preservation of research outputs may be a significant driver for research repositories but managing changing knowledge practices is vital for current learning and teaching repositories.

"Repository administrators need an understanding of the various drivers and their practical implications to enable them to pursue and promote the repository agenda within the institution."

Repository administrators need an understanding of the various drivers and their practical implications to enable them to pursue and promote the repository agenda within the institution. This will also allow them to engage with the wider academic community and prepare them to be able to respond to individual queries and concerns as the repository is established. This section considers separately the drivers for each kind of repository as repository managers are likely to be focussing on one particular form of repository.

Open Access Research Repositories

Scholarly Communication

Scholarly communication is a term used to describe the ways in which researchers publish and disseminate the outcomes of their research to make them available to their colleagues, peers, the wider academic and education community and beyond. Scholarly communication can be thought of as the creation, transformation, dissemination and preservation of knowledge related to research. It was once thought of simply as another way to describe the process of publishing a research monograph or an article in an academic journal, but recent developments have meant that the term is now used to signal a number of issues. These include author rights & copyright, open access, institutional repositories and preservation.

The traditional method of scholarly communication is by writing up the findings of research in to an article to be published in an academic journal or by publishing a book, book chapter or conference paper.

The term 'Scholarly Communication' is believed to have been in use for over 30 years, but recent developments (primarily the way that we disseminate and access information in the age of the internet) have meant that new methods of scholarly communication have come into being. The 'scholarly communications crisis' (also known as 'the journals crisis') has also been a driver supporting a move towards open access, with new options being born, such as publishing in subscription-free peer-reviewed open access journals, or disseminating versions of published papers through digital repositories.

Repositories and Research Management

Research Management

Improving research quality is often a key strategic driver for academic institutions and much effort is frequently placed in the administrative processes which gather and submit information about research for quality assessment exercises. The institutional repository is well placed to add value to this process of collecting information about

research, in particular being able to offer a detailed and systematic method of collating bibliographic information about research publications.

The RSP has found that discussions are often ongoing within an institution about whether to invest in a centralised current research management system, or CRIS, in order to better manage the substantial amount of research being undertaken. While repository cannot normally fulfil the role of collecting information about the entire research process, it is well placed to be involved in the discussions about introducing such a system. Integrating the repository with any research management system will ensure a long term future for the repository within the institutional information management landscape.

Research Information Network Report

Communicating knowledge: How and why UK researchers publish and disseminate their findings (PDF)

http://bit.ly/quiKvi

In addition to the potential research management benefits of the repository it should also be emphasised that placing research in an open access repository has been shown to increase the **visibility and impact** of the work. This is likely to become a measure of research quality for the future. Although much of the detail has yet to be agreed it is already clear that the new **Research Excellence Framework** (REF)[21], which replaces the RAE, will place a much higher emphasis than hitherto on bibliometric indicators of quality. It is consequently essential that institutions make early plans for collecting, recording and monitoring research publication data.

Preservation

Digital preservation involves the management of digital information over time. It is generally considered that the preservation of digital information could require more constant and ongoing attention than preservation of other types of media (**Lifecycle Information for E-literature**[22], 2006). A lot of time, money and effort would need to go into preserving the huge amount of data created during the explosion in use of different IT systems over the last few decades. So while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost, creating a *'digital dark age'*.

There is much discussion about, and currently no clear answers to, issues surrounding digital preservation in repositories. There are, however, a number of points that are worth bearing in mind when considering preservation in relation to your repository. Firstly, simply having even the most basic awareness of digital preservation is a step in the right direction. Try to consider developing a **preservation policy** for your repository.

You might even consider that the repository is about access and therefore not the location to carry out digital preservation activities. Do think about preservation and consult with other agencies for advice.

Preservation Strategies

In 2006, the **Online Computer Library Centre**[23] (OCLC) developed a four-point strategy for the long-term preservation of digital objects that consisted of:

[21] http://www.hefce.ac.uk/research/assessment/reform/

[22] http://eprints.ucl.ac.uk/1854/

[23] http://www.oclc.org/

- Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications
- Evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied
- Determining the appropriate metadata needed for each object type and how it is associated with the objects
- Providing access to the content

Online Computer Library Center, Inc. (2006). **OCLC Digital Archive Preservation Policy and Supporting Documentation**[24], p. 5

There are several additional strategies that individuals and organisations may use to actively combat the loss of digital information.

Refreshing is the principle of transferring data between two types of the same storage medium. For example, you might copy data from an old, worn CD to a new one. This would need to be done periodically because of the potential for physical deterioration of the media.

Migration involves transferring of data to newer system environments, for example making a move from MS Windows to Linux. It might also be converting Microsoft material such as word or excel files to a PDF.

Replication is the idea of creating duplicate copies of data on one or more systems. You might consider backing your repository up at another institution.

Emulation is replicating of functionality of an obsolete system. This might involve re-designing and building some hardware to access data on obsolete storage media.

Learning and Teaching Repositories

Drivers for Learning and Teaching Repositories

Many of the early formal teaching and learning repositories in the UK were not institutionally based, but were:

- national (such as **JORUM**[25] or **NDLR**[26])
- regional (such as Staffordshire's **SURFWBL**[27])
- sectoral (such as the Scottish FE collection on **COLEG**[28]), or

[24] http://bit.ly/qXxMwE

[25] http://www.jorum.ac.uk/

[26] http://www.ndlr.ie/

[27] http://www.jisc.ac.uk/whatwedo/programmes/x4l/surfwbl.aspx

[28] http://www.coleg.org.uk/coleg/69.html

- subject discipline based (such as **IVIMEDS**[29] or **IRISS**[30])

Many of these repository services were driven by substantial national funding initiatives which intended to create a core collection of materials that improved access, sharing, cost efficiencies and prevent duplication. The JISC funded **Good Intentions Report**[31] provides a useful history of learning & teaching repositories.

With the growth of the web, and particularly Web 2.0, many academics 'published' or shared their learning materials online in an open and informal way which contrasted significantly with the growth of institutional VLEs. Learning materials in VLEs were often 'hidden' behind authentication systems that resulted in content not being shared across departments and were only accessible to tutors and students on each course. Institutional repositories present a way of bringing the two together and can offer different degrees of openness so that academics can choose how widely to share their materials.

It is important to note that drivers for institutions to establish and maintain repositories are not necessarily the same as for the individuals that create the content. There are numerous stakeholders both in terms of deposit and use of materials.

Sharing Learning and Teaching Resources

Many of the nationally funded content initiatives (such as **Exchange for Learning Programme (X4L)**[32] had the original intention of encouraging sharing as well as reducing costly duplication. **JORUM**[33] - a national UK HE and FE learning and teaching repository was developed to facilitate this. Despite this investment several barriers prevented widespread sharing. JISC has commissioned a number of studies into the **'sharing'** of learning and teaching resources which highlight the cultural, pedagogical, technical and organisational barriers to sharing.:

- Community Dimensions of Learning Object Repositories **CD LOR**[34],
- Trust in Digital Repositories **TRUST DR**[35],
- West Midlands Share - Promoting shared use of digital content across the region **WM-Share**[36],
- Repository Metadata and Management project **RepoMMan**[37]**,**

[29] http://www.ivimeds.org/

[30] http://www.iriss.ac.uk/learnx/

[31] http://bit.ly/pS58w3

[32] http://www.jisc.ac.uk/whatwedo/programmes/x4l.aspx

[33] http://www.jorum.ac.uk/

[34] http://www.academy.gcal.ac.uk/cd-lor/

[35] http://trustdr.ulster.ac.uk/

[36] http://www.jisc.ac.uk/whatwedo/programmes/edistributed/wmshare.aspx

[37] http://www.hull.ac.uk/esig/repomman/

- **Rights and Rewards in blended repositories survey**[38],
- **Sharing e-learning content**[39] - a synthesis and commentary,
- **Good Intentions report**[40]: improving the evidence base in support of sharing learning materials

It could be argued that sharing implies an open model (sharing with all) but many individuals have little incentive to share in this way. The FE sector, with shared curricula and assessment frameworks, does have a culture of sharing within the sector, but individuals in HE institutions may be more likely to share within their subject discipline community that with colleagues in their own institution.

This lack of incentive for individuals to share and lack of recognition by institutions of the value of sharing learning & teaching materials meant that most institutional repositories focussed primarily on research outputs. This tied in with the move by institutions to invest in VLEs which had access control and authentication as a significant function. Closed VLEs do not encourage sharing and for many institutions these were seen to be the main mechanism for managing access to their learning and teaching materials.

Web 2.0 and social software applications have facilitated the sharing of practice and content both for individuals and subject communities. This, combined with the Open Movement, has gone some way to contributing to changes in cultural and perceptions around sharing. Another significant driver in this cultural shift is the perceived simplicity of **Creative Commons licensing**[41].

Open Educational Resources

The global open education movement has become a significant driver for individuals and institutions as they realise the potential benefits to opening up their teaching and learning materials. These benefits are different for each stakeholder group - these are usefully described in the **Open Educational Resources infokit**[42].

Widely available learning content, and informational content, is fundamentally changing the relationship between students and their institutions as sources of expertise. This presents a challenge to existing models of the production of academic knowledge and the role of the institution in supporting student learning. Coupled to this, funding models in Higher Education are changing in response economic pressures, resulting in many institutions re-examining their own business models in response.

> The ongoing JISC/Academy Open Educational Resources Programme (https://openeducationalresources.pbworks.com/Open-Educational-Resources-Programme) is addressing many issues around open institutional repositories.

As institutions consider new open models they are increasingly seeing the potentials offered by a managed repository (rather than a closed VLE which may contain student data). They need to balance the new drivers for openness with their existing drivers

[38] http://bit.ly/pyX9wX

[39] http://bit.ly/ouvwDk

[40] http://ie-repository.jisc.ac.uk/265/

[41] http://creativecommons.org/

[42] https://openeducationalresources.pbworks.com/Why-OER-Stakeholders-and-benefits

for competitiveness, confidentiality, and integrated institutional systems. Options for institutions include:

- creating a learning and teaching repository which can offer open access (which may have varying degrees of openness)
- adapting an existing open access research repository to include learning and teaching content
- utilising the national open repository - **JORUMOpen**[43]
- utilising an open VLE system (such as Moodle) as a repository

Individuals may have a mandate to deposit their learning & teaching materials into JORUMOpen (through funded projects), or their institutional repository. They may also have an option of **depositing into subject community environments**[44] or just publishing openly on the web if they own the content they develop for a particular course (many academics in HE institutions do not own such content). Individuals may need to be presented with an option where they deposit only once (perhaps into their institutional repository) which can feed into other channels.

Developing Teaching and Learning Resources

Formal learning and teaching repositories may be called 'Learning Object repositories (LORs)'. The Community Dimensions of Learning Object Repositories CD LOR[45] study (2007) usefully describes a Learning Object (LO) and LORs

The format of learning and teaching materials can shape the resulting repository models and the concept of 'packaged' learning objects is still a current and valid model. There is also a recognition that 'unpackaged' materials are equally useful (and may even be much easier to re-purpose and re-use). There is an ongoing debate around granularity of learning materials and how far the pedagogic context should be integrated into resources, or wrapped around it (pedagogic wrappers).The model that supports multiple formats and multiple approaches to deposit is likely

> *The term 'Learning Object' has come to represent the concept of a highly granular digital resource developed to meet a single learning objective. Such LO's may be aggregated to form larger units, and may be incorporated into a range of learning activities, be they entirely online, within blended learning, or classroom based. Key to the idea of LOs is that they should be durable, interoperable, reusable and shareable. Learning Object (LO) Repositories have emerged in recent years to support the storage, management, sharing and reuse of teaching and learning resources within and across learning communities.*

to be less about changing the repository structure or format, and more about ensuring integrated services around repositories and less formal stores of learning materials.

The increasing recognition that content created by students during learning activities could be included in repositories is likely to raise significant issues around ownership and management. Student access to institutional, national and other repositories is also likely to shape repository models. Repositories will need to be flexible and responsive to respond to these changing drivers.

[43] http://open.jorum.ac.uk/xmlui

[44] http://wiki.cetis.ac.uk/Feed_deposit

[45] http://bit.ly/oEW8rg

One of the challenges for any repository of learning and teaching materials is the management of version control due to the very significant driver of changing knowledge practices. Repositories are often seen as a place to store and preserve content in a static way, but many subject disciplines have to constantly update learning content. Medical education is an area where the importance of this is clear - changes in treatment practice must be reflected in teaching materials and storing 'out of date' materials may prove dangerous.

Repositories: A Management Framework

This section of the infoKit covers a range of topics to guide a repository manager through the initial planning process for establishing an institutional repository. These include the planning approach, making a business case, identifying key stakeholders, the costs and benefits of a repository, risk management, strategic planning and a selection of practical planning tools covering repository staffing and training.

Scope

To ensure consistency it is important to clearly define the content and scope of a repository.

- Is the repository going to have a disciplinary or institutional focus?
- Is the repository going to hold research outputs, data and/or learning resources, or a combination of content types?
- Will both published and unpublished items be collected?
- Will the repository need multiple levels of access - open, authenticated or both?
- Will it only hold actual content (such as full-text items, learning objects, individual learning assets) or will metadata-only records also be held?

When defining the content and scope of a repository the current needs of the institution and its researchers, teachers, learners, the resources available to the repository, as well as future plans for the repository's development should all be taken into account. Existing institutional strategies and policy will also shape the scope of the repository. It is also important to remember that the level of work required to develop and maintain a repository will depend on the scope and extent of the repository.

Strategic Plan

Related Resources

Strategy infoKit
http://www.jiscinfonet.ac.uk/infokits/strategy

Once your scope has been defined and funding has been agreed to establish a repository you need to devise a strategic plan as to how the repository will be implemented. Within the strategic plan you need to identify the vision for the repository; the mission of the repository; the goals and objectives to be achieved by the repository; and an action plan of how those goals and objectives will be met. The mission statement should identify the scope and purpose of the repository and what it is hoping to achieve; a summary of the goals and objectives may be useful here. The goals should be statements of what the repository is looking to achieve, i.e. "the repository will enhance the visibility of the institution by showcasing its research outputs (or learning materials) to the widest possible audience; facilitating both its discovery and delivery." The objectives should be **SMART**[46]: specific, measurable, assignable, realistic and time-related- allowing for their subsequent evaluation; and should relate to the goals set, i.e. "10 new deposits will be made to the repository every month." It is also important to detail how the achievement of the objectives will be evaluated, which critical success factors will be used and what measurements will be taken.

[46] http://www.jiscinfonet.ac.uk/tools/smart-targets

A useful tool to help identify and develop the goals, objectives and performance targets for a repository over the course of its lifetime is PLATTER- **Planning Tool for Trusted Electronic Repositories**[47], devised by Digital Preservation Europe (DPE). The focus of the tool is on the process by which the repository organization sets and manages its objectives and is designed to complement existing audit and certification tools.

In terms of the PRINCE2 Project Management method, the process of strategic planning begins with the creation of a **Project Initiation Document**[48]. This document should relate to the **Business Case**, and builds upon the analysis of any risks conducted for the purposes of this document. It is emphasised that thorough analysis of the resources available, any constraints in place and any associated risks will help inform a sound project plan, with a realistic timeframe and room for contingency planning.

Related Resources

Project Management infoKit
http://www.jiscinfonet.ac.uk/infokits/project-management

Planning Approach

The implementation of a repository is usually considered as a project and should be managed as such. There are a number of different project management methodologies available which offer proven techniques for planning and managing projects. However, a methodology commonly used is **PRINCE2**[49] (PRojects IN Controlled Environments). A useful resource to help guide you through the Project Management process is our **Project Management infoKit**[50] which has used PRINCE2 to create a user-friendly project management framework.

Projects can be seen to have three distinct phases. These are:

- initiation and start-up
- running and management
- closure and evaluation

There are different aspects to consider within each phase; and different documentation should be produced and should support the project along its various stages. Within the first phase of the project the Business Case is the most important document as it will not only sell the idea of a repository to the decision makers in the first instance but will also act as a reference document for the development of the Strategic Plan/Project Initiation Document. The Strategic Plan/ Project Initiation Document, which will include the Project Plan and the Project Schedule, uses the information gathered from the Business Case to inform all of the stakeholders involved with the repository's implementation as to the Whys, Hows, Whos and Whens of the project. Although created in the first phase of the project, this document acts as the baseline for the running and management of the project, and in turn is adapted as the project is executed. The Strategic Plan/Project Initiation Document also has a place in the **final phase of project management**[51] as the

[47] http://www.digitalpreservationeurope.eu/platter/

[48] http://www.jiscinfonet.ac.uk/infokits/project-management/project-initiation

[49] http://www.ogc.gov.uk/methods_prince_2.asp

[50] http://www.jiscinfonet.ac.uk/infokits/project-management/index_html

[51] http://www.jiscinfonet.ac.uk/infokits/project-management/closing-a-project

evaluation and closure of the project will be dictated by the Project Plan and the goals and objectives contained within it.

Business Case

Institutional repositories affect many people and there are a considerable number of **stakeholders** that have to be accommodated within the design, establishment and expansion of an institutional repository. Experience from those who have established and run institutional repositories in the past shows that before moving forward it pays to think things through and to perform some standard stakeholder and risk analysis. Creating a business case will address many of these issues.

To make a case for implementing a repository it must be made clear what the **benefits** will be to the institution or group; how it will help address the **strategy** of that institution or group; and what the **alternative** would be if the repository was not implemented. Not only do you need to promote the added value that the repository would bring to the 'business' and the community it serves but also give a realistic idea of what the costs of implementing and running the repository would be, along with any associated **risks**. It is important to consider the effects the implementation of a repository will have on current processes and what **resources** will be needed to implement the repository, including human resources.

As a general guideline, it is best to make a clear case with the more clearly identified costs presented up front, with a persuasive and clearly articulated set of strategic benefits that match institutional priorities. It is better to clearly identify costs without attempting to short-change development resources - it might be tempting to squeeze development work through on the cheap, but the price will be high expectations based on insufficient resources. A repository is a strategic investment and should be funded accordingly.

Key Stakeholders

There are many individuals and groups within an institution who can benefit from a repository and who therefore, can be identified as stakeholders within it. Take the time to make sure you've identified everyone with an interest, and establish the **key benefits** that will apply to each group. This ensures your institutional repository (IR) will begin on a positive note and have wide-reaching support.

There are two critical reasons to identify stakeholders:

- **To promote -** The repository manager needs to persuade others of the benefits of an institutional repository. The aims may include obtaining backing, funding or agreement in principle to the installation of an IR. There is a need to support the concept of the repository and to encourage use of it when it is available
- **To consult -** The repository manager needs to engage local stakeholders in discussions to identify their requirements, interests or concerns regarding the service

Some key stakeholders of a repository have been identified below:

> Further advice for constructing a Business Case can be found within the business case http://www.jiscinfonet.ac.uk/infokits/project-management/business-case page of our Project Management infoKit.

University administrators, senior management and policy makers

This group of stakeholders represent the institution as a whole. An individual or group of people with the original idea will usually be required to sell the concept of the repository to these key decision makers. The aim of initial

communication with this group will be to get the concept of a repository on the institution's strategic agenda and to facilitate discussions about how it might be integrated into existing information management frameworks. Subsequently, this group of stakeholders will perform the important task of approving and backing the IR and committing staff time and resource.

Academics as authors and researchers

Academic staff play various roles related to the repository: contributing content, searching and discovering content, sometimes managing versions or updating content. Persuading and supporting academics to submit content to the repository is one of the most significant challenges a repository manager will face.

Library staff

Discussions about a repository often originate from within the library of the organisation and repository managers often come from a library background. Working with a repository suits the professional skills found within the library and strengthens the library's role within the wider organisation. Library staff often become involved in the processes and workflows within the repository; typically they engage in promotion of its use through their strong links with the academic community.

Technical support staff

Technical staff fulfil vital roles of installing, configuring and maintaining repository software and hardware. Repository managers need to develop clear lines of communication with technical colleagues, who need to understand of the purpose of the repository and how it fits in with the institution's strategic vision.

Other support staff

These can include staff managing the research outputs of the institution; departmental administrators and secretaries (who may organise content on behalf of academics); copyright officers (who will be a source of advice and expertise); or staff development officers (who can help deliver the repository message to the wider academic community); educational developers who may provide pedagogic support to staff who want to take advantages of the affordances offered by digitising learning and teaching materials.

Identifying the Benefits

Once the main **stakeholders** have been identified it is useful to map out their needs and aspirations in key areas:

- Information management
- Research management
- Assessment and REF
- Institutional profile
- Personal profiles
- Marketing and publicity
- Competitor parity

The next task is then to clarify how the repository addresses these stakeholders and their needs in the local and national context. Prioritise this information to make your case by identifying who is important in your own institutional context and what they need, to show the strategic value for your repository:

- Show relevance to stakeholder needs
- Show relevance to stakeholder aspirations
- Show an account for stakeholder concerns
- Identify any short-term returns
- Describe long-term benefits in added-value and strategic support
- Quantify costs without undercutting true investment

Financial benefits

Characterising the benefits of the repository in **financial terms** is very difficult, as the benefits are more obviously strategic rather than financial. The financial effects themselves are more easily analysed on the macro scale rather than the institutional scale. Houghton and Sheehan (2006) considered the financial benefit of enhanced access to research findings in terms of percentage increases in access and efficiency of national R&D investment.

If any institutional advocate is to use an approach of this sort, then it may be best to enlist knowledgeable assistance in financial analysis, as fixing financial gains by an individual institution is complex and investigative. It can be seen that using any predictive financial model in negotiation immediately ties the repository into being seen as a resource stream, rather than anything else. Once money is brought into a discussion it does tend to dominate any other consideration or argument. It could be said that in the world of institutional politics and finances where there are many competing resource issues, that any financial cost-benefit analysis is the wrong approach and one which should be avoided.

While questions about cost and value-for-money are reasonable, the benefits of the repository are more clearly seen in the strategic advantage and benefits that it gives in information management, research profile, author service, public and professional reputation and more. Adopting this approach requires sensitivity, as replying to questions about money by appealing to high-level strategy can look like an attempt to sidestep financial issues!

Comparators

In making a case, as general comparators other university information systems can be used: for example, what is the 'cost' of the university email system? What is the financial benefit of the email system? The first question can probably be estimated, given time: the second question goes beyond a spreadsheet answer. There are many other comparably strategic or infrastructural systems - the institutional website; timetabling system; library catalogue, or the e-learning environment. In each case, the overall benefits can be judged against high-level targets and infrastructural needs, but in addition it can be seen that there would be costs and competitive disadvantages if the institution did NOT have the system.

Integration within the institution

The repository is best seen in this way and is another reason for repository development to be seen as working to **integrate** the repository into the mainstream of institutional information systems and management. If the repository is seen as a stand-alone project, then not only will costs and benefits be assessed at a financial level as a self-contained unit, but the repository itself will not be able to provide the benefits that it should.

Integration with other institutional systems is essential. In making a case, link the repository directly into the strategic aims of your particular institution: these are often in the area of research development, information management, public profile, etc all of which relate directly to repository use. Making a case in this way will be intimately connected with each individual institution's needs, political priorities and current concerns and is best constructed locally with that knowledge.

Houghton, J. and Sheehan, P.(2006). The economic impact of enhanced access to research findings. CSES Working Paper no. 23. Melbourne: Centre for Strategic Economic Studies, Victoria University. [Online]. Available at: **http://www.cfses.com/documents/wp23.pdf**. Accessed September 27, 2010.

Costs

Although a range of open source repository software is available for free, there are a number of other costs that need to be considered when planning the implementation of a repository. It is not simply a case of purchasing the hardware and installing the software - when it comes to implementing and running a repository, there is a high level of human resource cost that needs to be considered as well. If a repository is to be successful it must be considered as part of the institution's long term strategy, which therefore requires planned investment in the short, medium and long term. Therefore, to both create and maintain an institutional repository, financial, staffing and time resources will need to be invested.

Typical hardware costs

Informal evidence collected by the RSP[52] indicated that the amount being spent on hardware for repositories varied greatly, ranging from virtually nothing for pilot services, through to a few thousand pounds, up to top end of the scale where thousands and thousands of pounds were being invested in computing hardware for the repository. The Welsh Repository Network[53], a JISC funded Repository Start-Up Project, provided each of the higher education institutions within Wales with £4000 for hardware costs to support an institutional repository. The project enabled every institution to host an institutional repository, many of which having had no repository in place at the beginning of the project.

Staff Costs

The human level of resource needed for a repository is variable, although experiences in the sector have demonstrated that a greater resourcing tends to result in a more rapid ingest of materials.

Start-up costs

- Advertising for, and appointing, a suitably qualified repository administrator
- Time needed with staff to choose the appropriate software
- Expertise of technical Staff to install the software

[52] http://www.rsp.ac.uk/

[53] http://whelf.ac.uk/wrn/index.shtml

- Time needed to administer it, make changes and amendments
- Developing repository policies and an overarching strategy
- Establishing a steering group or advisory panel
- Training support staff, researchers, teachers and other end users
- Adapting existing learning and teaching materials for deposit - this could range from simple digitisation of existing materials to redeveloping materials, adapting pedagogic contexts and stripping inappropriately licensed content
- Time spend integrating repositories with other institutional systems
- Publicising the repository as well as full-blown advocacy campaigns
- Investigating legal issues such as licences, copyright etc and possibly establishing appropriate licences for different content

Ongoing demands

Certainly the task of promoting the benefits, while at the same time reassuring staff on their concerns about using a repository, is a complex, resource intensive task. It is not an area which can be ticked off a list as ever being entirely complete. It is also an aspect that may ebb and surge unexpectedly due to developments within the institution or across the scholarly world. However, the one constant is that there will always be some area of the repository work with which staff will be required to engage; such as:

- Ongoing support needs of end users
- Creating content records (metadata) - some argue that this is a relatively straightforward process where anyone can do it, while others argue that it is a highly specialised skill - many adopt multiple approaches to adding metadata which may involve several staff through a complex workflow
- Ensuring that staff are placing appropriate content into the repository
- Assessing quality of learning and teaching content
- Supporting digitisation/packaging of learning materials
- Considering the implementation of a mandate (requirement) for all staff to deposit into the repository
- Educating and investigating rights of who owns what
- Assisting academics in recovering their rights to deposit research materials through direct interaction with their publishers
- Running usability testing with the user community and revising the site interface as a result
- Allaying fears from academics on issues such as whether or not their publishers will allow them to store their publications in a repository, if their learning materials are of sufficient quality to release openly, or if people will misuse their learning materials
- Assisting staff to select the most appropriate level of openness and licence if options are provided

Risk Management

In terms of repositories there are risks associated with content if ownership is unclear or if content has not been copyright cleared. If someone questions any content held in a repository the generally held view is that a take-down policy should adopt a safety-first approach and have as a first action removal of the item, or at least removing it from public view. A more considered evaluation of possible copyright infringement or other legal difficulty can then be made, secure in the knowledge that the institution has at least made some immediate response. After consideration, where the process might take some time to resolve, the item can then be removed completely or re-exposed accordingly.

Prompt action by the institution would demonstrate that some care and responsibility has been taken. If the item remains visible while the difficulty is identified and resolved, then any problem will have been compounded by its remaining visible throughout. A court may decide that this behaviour merits punishment in itself, quite apart from the original transgression.

A take-down policy should state what action should be taken, by whom, and in what time frame, to secure the situation. This might be something as simple as saying that any complaint will be passed on within 'x' hours of receipt on a working day to be assessed for removal of the public record within, say, 24 hours or shorter. There would then need to be some defined process where the complaint could be properly assessed and an appropriate course of action taken.

Repository Policy Framework

A comprehensive policy framework is a vital tool through which to establish the operational boundaries within which a repository will function. It supports day to day management of the repository whilst simultaneously taking the longer-term vision of future operational activities into account. Supportive and integrated policies are also indicative of an organisation's level of commitment and are essential if a repository is to be successfully embedded into the wider organisational framework. A clear and visibly supported repository policy framework does more than simply provide abstract high-level support. It also:

- **Facilitates stakeholder understanding:** Responsibilities and lines of communication are clarified; processes are clearly defined; and equity, standardisation and consistency are promoted. These all contribute towards increased and easier use of the repository
- **Ensures support for the repository:** The decision-making process is formalised and sign off by senior management is facilitated. These engage senior management in supporting and contributing to the repository from a longer-term perspective
- **Helps with the planning and decision-making processes:** Risks are better understood and managed, IPR are easier to comply with, and the implications of dealing with different types of resources are exposed. These outcomes contribute towards ongoing and day to day management of the repository

Like all projects there are a range of risks that could adversely affect the implementation of an institutional repository. **Project risk management** http://www.jiscinfonet.ac.uk/infokits/project-management/managing-risks is covered in detail in the Project Management infoKit.

Some, but not all, policies will need formalising as legal agreements. Specifically this involves definition of a deposit licence and a usage licence which users will need to agree to. Licences are a useful tool through which to manage the risks associated with maintaining a repository.

To ensure the policy framework is comprehensive, repository managers should seek to integrate a number of policies into the wider organisational framework. These can be classified into two main streams: strategic, and operational and the subsequent sections of this infoKit will deal with each of these in turn.

Strategic Policies and Mandates

Policies for repositories do not exist in isolation. They need to be created in harmony with the wider strategic policies of the institution. An important first step is to find out about existing policies. It is likely that an institution will have high-profile vision statements and defined agendas in areas such as research, teaching, theses and intellectual property. Do not underestimate the usefulness of existing policies. They can help align and embed a repository with the institution. In contrast, if such policies do not exist then the repository can act as a catalyst for their creation.

One of the most powerful policies is a mandate that makes it compulsory for every relevant member of staff to ensure that their research outputs are deposited in the repository. Many funding councils are considering making deposit a condition of continued research funding and, internationally, consortia of institutions are in the process of ratifying their mandates. It could be argued that mandates are the single most effective way to ensure success.

It is less common for learning and teaching material deposit in repositories to be mandated, although there may be mandates to deposit within an institutional VLE. Repositories can be linked to VLEs to harvest material, particularly if there is a policy to make learning and teaching materials more openly available through institutional repositories.

Further Information

The OpenDOAR policies tool http://www.opendoar.org/tools/en/policies.php is a simple way for repository administrators to formulate and/or present their repository's policies. It provides a series of check boxes and pick lists for all key policy options.

Repositories handling audio-visual and learning materials can find a wealth of policy advice and templates in the Strategic Content Alliance IPR Toolkit:

http://bit.ly/q5C7RG

UK Examples of Policy Documents

Oxford Research Archive Policies:
http://www.ouls.ox.ac.uk/ora/ora_documents2/ora_policies

Oxford Research Archive Notice and Take-down procedure:

http://bit.ly/pMo1qI

PhilPapers Terms of Service:
http://philpapers.org/help/terms.html

University of Leicester Policies:

http://bit.ly/nnJCWX

Leeds Metropolitan University Operational Policies:
http://repository.leedsmet.ac.uk/main/policies.php

International Examples of Policy Documents

Georgia Tech Mission and Collection Development Policy: http://smartech.gatech.edu/help/policy.html

Max Planck Digital Library Concept Paper:
http://bit.ly/oP4MYu

OpenMed Submission Policy:
http://openmed.nic.in/information.html#submission

University of Leicester Examples

Research:
http://www2.le.ac.uk/offices/researchsupport/ref

Theses:
http://www2.le.ac.uk/departments/gradschool/current/theses/submission

Operational Policies

Strategic policies are concerned with the bigger picture and their consideration will begin to inform the day to day operation of the repository. The details of day to day operations are defined by operational policies. Operational policies are essential tools for a wide variety of people involved in the practical aspects of running a repository.

Operational policy areas include:

Submission policies

- Who can deposit?
- What type of materials can they deposit and in what format?
- Are there any policies relating to content packaging?
- What level of moderation is required for checking deposits, for example:
    - At what stage is IPR checked to ensure legality of deposited content and who does this (particularly relevant to the use of third party content in learning and teaching materials)
    - Are any quality checks included (in relation to formats such as sound quality, or in relation to pedagogic quality) and who does this. It may be that existing quality processes are felt to be sufficient.

Collection policies

- Will the repository focus on a specific discipline, or will it reflect the entire academic output?
- What types of materials are sought?
- What metadata must be collected?
- What versions are acceptable?
- Should peer or quality reviews be implemented?

Preservation policies

- For how long will the repository aim to preserve deposits?
- Can this be guaranteed?
- What formats should be used for preservation purposes?
- How will learning and teaching materials be updated - particularly when content becomes obsolete or wrong

Usage policies

- What can end-users and services do with repository metadata and content?
- How should publishers restrictions or embargoes be managed?
- At what level should usage be permitted, e.g. on an item by item level?
- Is there a take-down policy to respond to copyright or other infringement?

Notice and Take-Down policies

A take-down policy serves to try and minimise institutional risk from inappropriate material being made available through the repository. In the same way that institutions rarely police the content provided by an author on other

webpages, repositories can provide a facility for authors to expose their work on their own cognisance. This forms part of the concept of the Deposit Licence, in asking authors to accept copyright and content responsibility for the material that they deposit. However, what happens if there is a complaint by a reader? Complaints may be genuine, erroneous or frivolous and it might take some time to decide which. It is important to know in advance what measures should be taken to handle the situation.

The OER (Open Educational Resources) infoKit includes a section on legal aspects of OERs: https://openeducationalresources.pbworks.com/Legal-Aspects-of-OER and also on IPR: https://openeducationalresources.pbworks.com/Intellectual-Property-Rights-considerations

A take-down policy should state what action should be taken, by whom, and in what timeframe, to secure the situation. This might be something as simple as saying that any complaint will be passed on within 'x' hours of receipt on a working day to be assessed for removal of the public record within, say, 24 hours or shorter. There would then need to be some defined process where the complaint could be properly assessed and an appropriate course of action taken.

Preservation Policy

It is important to think about the long term future of the material entrusted to your repository. You should ensure that it continues to be readable as technology and file formats change, and as your organisational circumstances change. It may be necessary to establish different policies for different types of materials. Some content may need to be kept as deposited whilst other materials may need to reflect change. Learning and teaching materials in some subject areas are, by nature, more likely to be updated.

Examples

Web2Rights:

http://bit.ly/qMCERc

U-Now takedown policy (Nottingham University):
http://unow.nottingham.ac.uk/td_policy.html

UAL Take Down Policy (Open Educational Resources (ADM-OER) Project):

http://bit.ly/nywCVa

In the longer term, you may wish to use a preservation service, such as the one operated by the Centre for e-Research (CeRch)[54] at King's College London (formerly AHDS).

The main elements of a preservation policy are:

Retention Period

- How long do you undertake to retain items for?
- Indefinitely or not?

Functional Preservation

- What are your intentions to ensure continued readability and usability of the items in your repository?
- What technical steps are you taking to fulfil your intentions, either by yourself or with partners?

File Preservation

- How are you backing up your repository files, in what form, and how often?

Withdrawal Policy

- Do you allow items to be withdrawn?

[54] http://www.kcl.ac.uk/iss/cerch/

- If so, what reasons are acceptable?

Withdrawn Items

- How are items withdrawn?
- Are they deleted entirely, or do you just remove them from public view?
- Do the original URLs remain valid, and if so, do they point to 'tombstone' citations or to replacement items?

Version Control

- Do you allow items to be changed after they have been committed to the repository?
- Do you allow multiple versions?
- Can *addenda* and *corrigenda* be accommodated?

Closure Policy

- What would happen to material in your repository, should it be closed down for some reason?

Backup and File Restore

It is important, at the very least, to ensure that regular backups are made of the repository. This is usually carried out by IT systems administrators; a full backup should be made at least every week, with daily 'sinces' in between. Repository managers should also ensure that effective file restore procedures are in place with acceptable turnaround times. Ideally verification of backups should also be carried out.

Legal and Policy Framework

The legal aspects of maintaining a repository require serious consideration at a senior level within the institution, in particular relating to the correct risk management of the copyright held within repository content. This section outlines the key considerations to be taken into account. In order to manage the risk associated with copyright and other legal issues which might arise from use of the repository it is essential that these are incorporated into the wider policy framework.

Copyright

One of the key legal issues that will impact on the development and use of a repository is copyright. Copyright protects the expression of ideas when fixed within an original work. Works can be in any medium, for example, literary, dramatic, musical, artistic, and all such formats may be included within a repository. Copyright allows the rights holder to define if and how others can use their their work. This therefore affects how a repository's contents can be used and distributed. Copyright is often perceived as being complicated and as presenting barriers to using the work of others.

Learning and teaching materials often include 'third party' content (content produced and owned by someone else). When these materials are deposited in repositories the provenance of the content needs to be identified and cleared for appropriate use which can be very difficult and time consuming. This can be the first time that learning and teaching materials have been through this process and can present a steep learning curve for academic staff, who may need a lot of support. Many of the final reports[55] from JISC/HE Academy UKOER pilot programme projects, which

[55] https://openeducationalresources.pbworks.com/Pilot-Programme%3A-OER-Release-Outputs

sought to encourage open release of existing content, found that this process was too costly. They suggested that open release should be incorporated into newly developed learning and teaching materials, rather than applied to existing content. These projects produced a wide range of materials around Copyright[56] that will be of use to any repository manager who is considering open access for learning and teaching materials.

The sections below give a brief introduction to the copyright issues to be considered within the more common types of work deposited within a repository. Further information about copyright and copyright in regards to repositories can be found at the Intellectual Property Office[57] and the JISC Legal Information Service[58].

Journal articles

The first copyright of a literary work, such as a journal article, is held by the author(s) of that work. However, when an article is published, typically the author assigns copyright, or gives a copyright licence to the publisher. Although the majority of publishers and journals allow authors to subsequently archive that work under certain conditions, other publishers are more restrictive. Depending on the particular agreement that is signed, the author retains more or less rights to use the article. Some agreements forbid the author from photocopying the article, using it in teaching, or putting it online. Other agreements are more liberal and allow the author to retain rights to use the article as he or she wishes.

The SHERPA service RoMEO[59] lists journal publishers and their associated copyright agreements, detailing how the publisher will generally allow an author to use a work on the web and where different, within a repository. In most cases a differentiation is made between a pre-print and a post-print of a work. These terms are often used to describe successive stages in the development of an article. Unfortunately, the terms are used to mean different things by different people and this can cause some confusion and ambiguity.

Pre-prints vs post-prints

One usage of the term pre-print is to describe the first draft of an article, before peer-review, even before any contact with a publisher. This use is common amongst academics, for who the key modification of an article is the peer-review process.
Another use of the term pre-print is for the finished article, reviewed and amended, ready and accepted for publication, but separate from the version that is typeset or formatted by the publisher. This use is more common amongst publishers, for whom the final and significant stage of modification to an article is the arrangement of the material for putting to print.

Within the RoMEO listing however, pre-prints are characterised as being the version of the paper before peer review; and post-prints as being the version of the paper after peer-review, with revisions having been made. This means that in terms of content, post-prints are the article 'as published'. However, in terms of appearance this might not be the same as the published article, as publishers often reserve for themselves rights in their own arrangement of typesetting and formatting.
Some publishers will insist that authors use the publisher-generated PDF - often because the publishers want their material to be seen as a professionally produced PDF that fits with their own house-style. However, such a formatted

[56] https://openeducationalresources.pbworks.com/Pilot-programme-outputs%3A-Legal-issues

[57] http://www.ipo.gov.uk/types/copy.htm

[58] http://www.jisclegal.ac.uk/ipr/IntellectualPropertyPub.htm

[59] http://www.sherpa.ac.uk/romeo/

file is the copyright of the publisher and cannot be used without explicit permission. Typically, this means that the author cannot use the publisher-generated PDF file, but must make their own PDF version of the content for submission to a repository, or consider submitting a version in another format.

Requesting to use published material

A request template can be used to form a letter to a publisher asking for permission to mount material on a repository on behalf of an academic author. Some publishers insist on the author writing or emailing them directly to request permission to mount copies in a repository. In such cases, it may be useful to provide the author with an alternate template to help them construct his or her request.

Books & Book chapters

Permissions normally differ for the inclusion of book content from that of journal articles even within the same publishing office, and RoMEO does not cover book copyright agreements. Therefore it is best to review the copyright transfer agreements signed by the author or any equivalents available on the publisher's website, with the author contacting publishers directly if necessary.

An example of such a template is available from SHERPA: http://www.sherpa.ac.uk/documents/request_template.htm

Assisting authors with publishing agreements

The Copyright Toolbox[60] developed through JISC, SURF and RoMEO research provides advice for both authors and publishers on publishing agreements and licences. The aim of the resource is to assist authors and publishers to achieve a balance between granting maximum access to a journal article and financial compensation for the publication by the publisher of the article. Within the toolbox you can find both an introduction to publishing agreements and licences as well as sample wording for both document types. In particular, there is a Licence to Publish[61] which any author can send to a publisher to consider instead of the publisher's own agreement, to make clear the author's desire to put a version in their institutional repository.

Multimedia

There can be many layers of copyright within multimedia deposits as well as other legal considerations such as performer's rights and data protection. As such, each deposit should be considered on a case by case basis to ensure all aspects have been covered. Materials produced for learning and teaching are often owned by the institution, although teachers may not be aware of this. Clarification of ownership should be made clear in staff contracts. Many learning and teaching materials contain a variety of formats and some are 'packaged' in some way to provide a pedagogic context. What is considered as one resource in terms of deposit may actually contain materials owned by several different people. For example one powerpoint presentation could contain still images, sound files, and textual content from a range of authors.

Still images

As with an original literary work, the first copyright of an original still image lies with its creator. For example, in the case of a drawing, the first copyright will be held by the artist; in the case of a photograph, first copyright will be held by the photographer. However, in the case of photographs, care needs to be taken that the image does not include a

[60] http://copyrighttoolbox.surf.nl/copyrighttoolbox/

[61] http://copyrighttoolbox.surf.nl/copyrighttoolbox/authors/licence/

work of somebody else's copyright, also known as third party copyright, i.e. the photograph is of somebody else's painting or sculpture that is not on public display.

Consideration of data protection, obscenity and related areas including the protection of children, and libel may also be relevant laws that need to be considered before an image is deposited and distributed via a repository.

Moving images

Within the Copyright, Designs and Patents Act 1988[62] *'any medium from which a moving image may by any means be produced'* is considered as a *'Film'*. Included within this category therefore, are videos, animations and even computer game footage. Within films the first copyright is held by the Producer, and if he or she is a separate individual, the Principal Director. However, copyright may also be held within the story that is being performed, any dialogue that is performed, and/or any music or sound played or performed. Performers' rights may also need to be considered.

Sound recordings

The copyright considerations with regard to sound recordings (which can include music, speech, or song) are fairly similar to those of films. The first copyright is held by the Producer, yet other copyright can be held in any music performed, any lyrics performed or any prose recited. Performers' rights may also need to be taken into consideration.

Performers' Rights

If a work contains a performance (which can include acting, playing an instrument, dance, song or mime) the individual giving that performance has rights within the recording, filming and distribution of that performance. A performer has to give their express permission for the performance to be either filmed or recorded in the first instance; and where a performance has been filmed or recorded with permission, further permission has to be granted by the performer for that recording to be copied or distributed. A performer's rights are infringed by anybody who carries out any of the above acts without the performer's consent. A performer's rights are also infringed by anybody who plays or shows an illicit recording of a performance to the public; or by anybody who distributes an illicit recording to the public. Therefore, if a performance item is deposited and made accessible via a repository, and it is known that the performers' permissions were not sought to either record or show the performance, whoever is responsible for the repository will be liable to an infringement along with those who made and copied the performance recording in the first place. For information on performers' rights in relation to recorded lectures please see the JISC Legal guidance[63].

Repositories handling audio visual and learning materials can find a wealth of policy advice and templates in the Strategic Content Alliance IPR Toolkit:

http://bit.ly/q5C7RG

Theses and Dissertations: Fair dealing and third party copyright

There are certain instances of 'fair dealing' where third party copyrighted material can be used without the rights holder's permission. The two instances which are most significant within academic work are use for: non-commercial research and private study; and criticism, review and reporting current events. Fair dealing has no strict definition but

[62] http://www.opsi.gov.uk/acts/acts1988/UKpga_19880048_en_1.htm

[63] http://www.jisclegal.ac.uk/ManageContent/ViewDetail/tabid/243/ID/1608/Recording-Lectures-Legal-Considerations-28072010.aspx

can be assessed by considering the economic impact to the copyright owner by the use of their work in that particular way. If there is a low or minimal economic impact to the copyright holder, it can be considered as fair dealing. Whenever copyrighted material is used in these instances it is necessary for the owner of the copyright to be sufficiently acknowledged.

It can be quite common within work created by students, such as theses or dissertations, for third party copyrighted materials to be included without the copyright holder's permission under fair dealing. However, depositing and distributing a work which includes such copyrighted material via the repository could be considered beyond fair dealing and therefore, would infringe that persons copyright. If a repository's collection policy is looking to include theses and dissertations, how to accommodate instances of third party copyright within them needs to be considered before these works can be included and accessed. Ideally, any new students whose work could potentially be included within the repository should receive copyright training and should actively seek to obtain permission for any included works. In some instances it can be possible to obtain permission to use a copyrighted work retrospectively. However, trying to obtain permission at a later date can be difficult, particularly where contact details are no longer available or accurate.

Repository Steering Groups

A well-chosen, well-informed and committed Steering Group can make an important contribution to the sustained success of a repository. A repository Steering Group (or Project Board, Management Committee etc.) undertakes the high level management of a repository on behalf of an educational institution. The Steering Group helps to ensure that the strategic objectives of a repository are aligned with those of the parent organisation, and to ensure that the necessary resources are made available for the repository to be effective. It is a forum for discussion and decision-making, for the approval and signing-off of plans and actions, for arbitration and solution-finding; it should also champion all the publicity and dissemination activity linked to the repository. A Steering Group may be formed with a limited lifespan, to oversee a fixed-term repository project, or it may have a remit for a repository as a longer-term, independent concern (with or without a project element).

Terms of Reference

Terms of reference help participants and others to understand the role of a Steering Group. HEIs often have standard terms of reference for their Project Boards and Committees. If so, these would normally be used as a starting point, with adaptation where appropriate to reflect the repository remit. The terms of reference for a typical repository Steering Group might embrace some or all of the following themes:

Overseeing the implementation and operation of the repository

- Policy-making and strategic decision-taking
- Monitoring budgets and staffing resources: ensuring that necessary resources are committed to the repository, and ensuring that any plans are achievable
- Monitoring progress against plans and agreed objectives; evaluation
- Overseeing and championing advocacy of the repository at all levels within the HEI, and maximising opportunities for publicity and dissemination

Specific advice on open access to theses and dissertations is available from the EThoS Toolkit: http://ethostoolkit.cranfield.ac.uk/tiki-index.php?page=toolkit

- Reviewing outcomes and their impact within the HEI and elsewhere
- Maintaining an awareness of similar activities and initiatives, and funding opportunities, nationally and in the wider world
- Regular reporting, either to a named senior individual, or to a parent committee
- Securing the long-term sustainability of the repository, ensuring that it is firmly embedded in the strategic thinking of the institution

Issues associated with the day-to-day management and population of the repository might include:

- The respective merits of mediated deposit and true 'self-archiving'. Is one or other more likely to result in deposited content, and what are the resource implications of success?
- Copyright policies and institutional risk tolerance
- The collection policy. What types of research output should be eligible? What types of learning materials should be included? What levels of metadata are required by depositors? What range of licences for use should be available? How open should the repository be?
- The advocacy programme. When to target a particular audience, and how best to gain their support
- The evaluation of opportunities for external project funding

Issues relating to the place of the repository in the wider institutional strategic framework might include:

- The potential for the repository to serve as a comprehensive publications database
- Supporting an institution's preparation for, and success in, external research assessment
- Supporting learning, teaching and assessment strategies and objectives
- Raising the profile of the institution and showcasing high quality learning materials
- The potential for the repository to contribute to institutional performance in national and international web rankings and league tables

Examples of longer-term strategic issues for the HEI, which might be influenced by a repository Steering Group, include:

- The managed curation of digital objects other than research outputs and learning & teaching materials, such as primary datasets and departmental image collections
- Institutional digitisation strategy
- Institutional strategy on digital preservation
- Institutional approach to openness of learning and teaching materials

For the Steering Group to be influential, and for it to keep in touch with wider institutional strategy, it should be chaired by a senior member of staff - ideally, the Vice-Principal/Deputy or Pro Vice-Chancellor/Director responsible

for Research or for Learning, Teaching & Assessment, or the chair of the 'parent' committee (if applicable). In addition to the Chair and the Repository Manager, the Steering Group might include:

- Chief Librarian, and relevant Library staff
- Senior computing support staff
- Research Administration staff
- Head of Educational development
- Head of Learning Technology
- Staff with research assessment responsibilities
- Representation from the Corporate Communications/Marketing department
- Academic researchers from different disciplines
- Copyright Officer

Seniority is less important than an ability to report the needs of researchers and teachers, and to champion the repository within a given constituency. The Arts research practices and formats, particularly in the Fine Arts, can present some interesting challenges, and should not be overlooked in a research repository!

The Steering Group needs to meet with sufficient frequency to support continuity, without meeting so often that it becomes immersed in day-to-day detail. In an HEI, termly meetings may suffice. Some considerations when planning the meetings structure are the:

- Institution's usual pattern of committee meetings
- Meetings schedule of the 'parent' committee, where applicable
- Timing of internal funding rounds
- Demands and timescales of any specific projects for which the Steering Group is responsible

Staffing Repositories

Staffing levels

Estimating the time required for administration is highly variable and dependent on your deposition rates and approach to content mediation. Regardless of the approach used it is likely that staff time will be required for fielding practical queries from authors/depositors, quality checking, adding metadata or even making deposits. Depending on mechanisms for developing learning and teaching content, and where responsibility sits within the institution, repository staff may or may not be involved in developing, reformatting or packaging content, and applying appropriate licences for use and re-use. Checking copyright ownership of materials which are made up of a range of items can be very time consuming and should be done before the deposit stage, however checks at the deposit stage may also be required.

Advocacy materials often state that it only takes 10 minutes to deposit one simple item such as a pdf document. Complicated items, however, can take much longer (for instance, if you need to merge separate files for figures in with the body of a paper). For a research repository with fully mediated deposition, 30 minutes per item would also make allowance for other administrative activities such as copyright checking. This estimate can perhaps only be halved for an author self-archiving system. Deposit in a learning and teaching repository can take longer if you include content packaging and copyright checking.

In addition to administering deposits, consideration should be given to promotional activities.

Repository Roles

Staff requirements for a repository vary greatly between institutions depending on the remit of the repository and on existing and available resources. In some repositories the skills, knowledge and abilities required may be expected of a single repository post with the assistance of general IT personnel. However, many institutions spread the work over two main posts:

1. A Repository Manager - who manages the 'human' side of the repository including content policies, advocacy, user training and a liaison with a wide range of institutional departments and external contacts
2. A Repository Administrator - who manages the technical implementation, customisation and management of repository software, manages metadata fields and quality, creates usage reports and tracks the preservation issues
3. Other institutions spread the work over several posts or over several departments; typically including library cataloguers, subject librarians, other library, teaching and administrative staff, learning technologists and educational developers, copyright officers as well as IT services

A **study conducted by SHERPA:** http://bit.ly/puD5nh

identified the skills, knowledge and abilities required of staff to develop and manage a successful institutional repository. The study did not try to identify the skill set required of a particular post as it was recognised that the skills and responsibilities of individual repository roles within an institution would be dependent on each repository's requirements and scope. Summarised below are the main skills identified

Repository skills set

Management

The ability to manage the set-up and development of the repository including strategic and financial planning and liaison with relevant groups and individuals.

Software

Familiarity with relevant web-based systems and repository software along with the ability to implement and modify systems and software to meet the needs of the repository and institution.

Metadata

Familiarity with relevant metadata standards and the ability to monitor and ensure metadata quality is maintained.

Storage and Preservation

An awareness of current best practice and the ability to liaise with other departments to ensure storage and preservation procedures meet best practice.

Content

Familiarity with current IPR issues along with the ability to develop content policies and engage with key stakeholders to maximise quality and quantity of content.

Liaison

The ability to liaise with various groups, departments and individuals both within the institution and externally to promote the deposit and use of items in the repository.

Download the toolkit and all associated resources:
http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/recruitment.aspx

Advocacy, Training and Support

The ability to meet the needs of the repository and its users in terms of advocacy of OA and the repository, training in the deposit and use of the repository and support for users requiring assistance or information.

Current Awareness and Professional Development

Familiarity with current trends in the repository and research community and an awareness of developments in repository software and associated technologies.

Recruitment Toolkit

Many repository projects are JISC funded and recruiting staff for such projects can be difficult and can pose severe risks to project timescales and outputs.

To help with this problem JISC commissioned Brunton Consultancy Ltd to prepare a toolkit for recruiting to JISC projects with a specific focus on the digital repository area. The toolkit deals with the whole recruitment process from planning through advertising to interviewing and the job offer. Any JISC project dealing with digital repositories planning to recruit should read the toolkit as it is full of good advice, useful resources and templates. While the toolkit is focused on digital repositories, most of the advice offered will apply to every JISC project that is planning to recruit.

Professional library organisations such as **CILIP:** http://www.cilip.org.uk/ also offer frequent training courses on issues such as metadata and copyright, as well as other 'soft' skills such as management, change management and assertiveness, which can be applied to benefit the role of the repository manager.

Training Repository Staff

As with all jobs, the roles of the Repository Manager and Administrator will require a certain amount of training, depending on individual experience and qualifications.

Experience of library systems, metadata and copyright are all an advantage to a repository manager or administrator. Technical skills and knowledge of repository systems are also an advantage. Repository software suppliers (open-source or commercial) should be able to provide technical training, and projects such as the **Repositories Support Project** (RSP)[64] and the **United Kingdom Council of Research Repositories** (UKCoRR)[65] can provide support and share best practice.

Training should enable:

[64] http://www.rsp.ac.uk/

[65] http://www.ukcorr.org/

- staff to use the repository
- authors to deposit material
- repository staff to design and specify the repository
- repository staff to manage the repository
- technical support staff to implement designs and requests
- technical support staff to integrate the repository with your local information environment

Staff who will be contributing learning and teaching materials to an institutional repository are likely to require training in content development, metadata, Intellectual Property Rights, licensing, deposit processes and workflow. Repository staff may need to develop a good understanding of the teaching and learning process, and the pedagogical context of materials. An effective two way communication flow between academics and repository may be supported by in-house training that takes account of institution specific policies and approaches.

UK Examples

NECTAR Briefing Sheet: http://bit.ly/nTsasz

NECTAR Leaflets: http://nectar.northampton.ac.uk/information.html#leaflets

University of Bolton: http://digitalcommons.bolton.ac.uk/ubiradmin/8/

University of Leicester - **Blog Posts:** http://uollibraryblog.wordpress.com/category/leicester-research-archive/ ,**Leaflets:** http://www.le.ac.uk/li/research/archive.html#publicity

International Examples

University of British Columbia – **Posters:** http://bit.ly/o5gxar **Leaflets:** http://bit.ly/prJsAC

Advocacy

Embedding a repository successfully as part of the institution requires significant **cultural change**, including integration within workflows of deposition and operational procedures. Promotion is one of the most powerful tools to achieve this. For effective promotion it is best to adopt more than one **approach**, ensuring **stakeholders** hear about the service from as many different directions as possible. There is a need for bespoke adaptation to take account of each institution's particular social, political, strategic and operational environments.

It is of key importance to remember that promotion must be a sustained ongoing effort. This should be budgeted for accordingly, both in terms of financial and staff resources. While institutions have noted that it has been possible to speak and interact with all departments in a matter of months, do not underestimate the time it will take to embed the repository in an institution. New staff and priorities arise, and there will be a need to repeat much of the effort. It is advisable to make use of other staff (e.g. liaison librarians, departmental and research administrators) to spread the load. Equipped with the core message they should be able to keep the repository **message** firmly on the agenda.

Keep talking

Do not be afraid of repeating and reiterating. Find new contexts for and continually adapt your message. Ultimately the target being aimed for is to give institutional repositories the same weight as other academic activities and priorities (e.g. examinations and applying for grants). Linking promotion to evaluation activities within the institution can ensure a two way dialogue which signals a willingness to listen and respond to concerns or suggestions.

Culture Change

It is widely agreed that the **technical implementation** of a digital repository is the easy side of the equation and that the most significant challenge facing institutions is the need to promote and drive cultural change. A key first step to

successful culture change is to define the **message** and motivation for repository development at your particular institution. Motivations for setting up repositories vary depending on communities but include:

- The Publications/Serials Crisis
- **Scholarly communication** and sharing
- Improving **research management** and reporting
- Supporting teaching and learning, including **re-use and re-purposing**
- Management of digital assets and **preservation** requirements
- The **open educational movement**[66]

Changing the culture of an institution is not an overnight job, nor is it one that can be achieved without a significant application of resources. The key to success of a repository is that its principal contributors, the academic community, understand what the driving forces are, why such a resource is needed, and what their role is. Thorough planning for the repository will help define these elements.

Cultural change is classically commonly brought about through engagement with academics on one of three levels:

- Intellectual (conceptual, logical, the greater good)
- Emotional (hearts and minds, enlightened self-interest)
- Political (necessity, external drivers, compliance with policy)

Intellectually repository administrators can expound the professional virtues and **benefits** brought about by the repository, persuading their academic community of the need for their engagement. Alternatively they can advocate the personal value to individuals of repository use, and seek to achieve engagement in this manner. Or finally they can drive the community through the introduction of mandated or strongly-recommended **policies** for the deposition of material, a process that may take some time to accomplish, uncover simmering issues of academic freedoms but ultimately prove an exceptionally successful approach to repository enlargement. Institutions may be able to include some aspects of reward and recognition for deposit into staff development or performance review mechanisms. This will require strong institutional level support and is likely to be linked to strategy and policy.

A more in-depth consideration of **culture:** http://www.jiscinfonet.ac.uk/infokits/change-management/culture/index_html and the effect this has on change management is available in our Change Management infoKit. The issues around culture change is also discussed in relation to **Open Educational Resources** in our OER Infokit: https://openeducationalresources.pbworks.com/Pilot-programme-outputs%3A-Cultural-issues

Core Message

Each contact, briefing and event helps to build an awareness of the repository, forming part of the overall **promotional framework**. Repository managers will usually promote the intellectual **benefits** of the repository,

[66] https://openeducationalresources.pbworks.com/What-are-Open-Educational-Resources

The repository collates, preserves and makes readily available to the global community the most valuable output of this institution - its research.

but should not neglect the emotive aspects, such as personal reward. It is beneficial to address the issue of *'What's in it for me?'* to **stakeholders**. Most audiences have limited time and attention. Potentially, all they will initially remember is the strapline.

A core message or strapline is therefore essential for promotion and this should reflect the repository's ethos and tangible advantages. This message should be incorporated into all promotional literature, every essential talk and be embraced by repository staff. There is a need to tailor aspects of your message to particular audiences so make them challenging, engaging and informative.

Simple examples of a strapline are:

An expansion of this idea is to consider reducing the central message or unique selling point (USP) of your repository to just 25 words. It is a powerful focusing technique, and may well help you to identify or refine your core message, offering a snappy but accurate response to the chance encounter with a key academic staff member who asks *'So what does this repository do?'*. For example:

Advocacy Options

A number of different advocacy strategies can be used, including top-down and bottom-up approaches, alongside blanket and targeted activities. The resources available will need to be considered and will affect the scope of the advocacy strategy. Some suggested approaches could be:

Top-down

- Explore institutional requirement for deposit (mandates)
- Obtain supporting statements from the very highest level of the institution
- Invite stakeholders to join repository steering groups to assist in exploring unique institutional challenges; influencing the strategic position of the repository
- Keep the Pro-Vice Chancellors and key committees informed of developments and successes. This ensures the repository is embedded in the organisation

Increase the visibility of your work - Deposit in UniRepoPrints!

The success of the university is built upon its research - build a firm foundation today with UniRepoPrints

Think local, reach global.

- Include repository deposit and use within formal staff development and performance review mechanisms

Bottom-up

- Locate repository champions. Enthusiastic early adopters can act as change agents, taking your messages out on a peer-to-peer basis

- Demonstrate how new researchers/teachers can contribute, and gain a flying start to their careers. Repository usage statistics can provide powerful encouragement
- Engage students, especially graduates, by promoting the use of open access research material. In turn they will influence their peers and mentors
- Inform and involve support staff, ensuring they understand the importance of the repository to the institution's strategy

Targeted

- To ensure there is some initial content in order to encourage more deposits, institutional repository managers and librarians can identify so-called 'green' publishers - those who allow self-archiving in any form - and then asking the academics who have published in those journals for permission to deposit those papers in the institution's institutional research repository. To check the list of publisher copyright policies on self-archiving, visit RoMEO[67].
- Work with departments most likely to benefit from the repository, such as:
  - those reviewing research management/reporting processes
  - subject areas with Funder Mandates[68]
  - those who's academics publish in wide range of journal publications
  - subject areas where academics already publish to Open Access friendly services such as PubMed Central[69] and Arxiv[70]
  - teaching departments with strong partnerships outside the institution can see the benefits of making resources open through the managed repository
  - articulate the benefits of managing versions, metadata and usage statistics that a repository brings for those academics who already publish learning and teaching materials on the open web

Example of a **supporting statement in the form of a podcast from the Vice Chancellor of Nottingham University:** http://unow.nottingham.ac.uk/about.html

Advocacy can also use an intellectual, emotional or political message to achieve its intended results.

Other approaches to encourage self-archiving practices include presentations at events; the use of posters; and organising workshops and training. A starting point could be to organise a departmental staff briefing and present the launch of the repository, outlining its key features, benefits and future plans. This can be extended to a bigger event to invite key institutional staff such as Heads of Department, Heads of Research and senior administrative staff.

[67] http://www.sherpa.ac.uk/romeo.php

[68] http://www.sherpa.ac.uk/juliet/

[69] http://www.pubmedcentral.nih.gov/

[70] http://arxiv.org/

It is clear that having a repository is useless without having a significant amount of content in it. The main problem faced by most repository managers is how to get staff to deposit content in a repository. It seems that just asking them to do this voluntarily produces little content. Some research suggests that having some sort of publications policy or 'mandate' in place (i.e. staff are obliged to deposit) results in a higher rate of self-deposit. However, not all institutional environments are comfortable with this perceived level of interference with academic freedoms and it remains an area of much debate.

Advocacy Activities

Events

Repository promotion is often conducted at an individual or small group level. Larger events can also make for effective promotion. It is important to reach out to as many members of the institutional community as possible. You may wish to enlist a head of service to approach senior staff directly. Events should be informative, digestible and wherever possible tailored to the attendees' professional interests. Repository team members should attend with the express purpose of engaging individuals. Capitalise on success by rapidly following up on leads and contacts generated.

Making Presentations

Promoting the development and use of an institutional repository inevitably involves making presentations to a wide variety of audiences. The repository manager should take all opportunities that are available to them for presenting information about the repository to the various stakeholders within the institution. This can range from high profile dedicated repository launch events, more formal opportunities - for example at departmental meetings, research committees and other internal seminars - through to simply taking the time to discuss the repository informally at a chance meeting over coffee. Keeping the **messages** simple and highlighting the benefits (as opposed to simply the features) of an institutional repository are key promotional tools. The content of each presentation will be defined by local circumstances and it is worth trying to tailor the content to the audience concerned, for example an audience of academics may appreciate most hearing about funder mandates and the requirements of their grants to place a copy of their work in an open access repository.

Faculty/Departmental Campaigns

Presentations promoting the new repository are also useful at departmental meetings. Securing an invitation to a departmental research committee is a very good opportunity to advocate the benefits of Open Access and self-archiving. Presentations may be more effective if they are delivered separately for academics, library staff and senior management.

However, institutional staff such as managers, librarians and repository administrators should be aware of concerns often voiced by academics regarding the deposit of their work. Examples of such concerns include the disruption of the traditional publishing model, the actual quality of work found in repositories, the fact that it could be time consuming and issues of work preservation. Anyone involved with advocating the repository should make sure they are comfortable with likely concerns and have answers prepared. The concerns and benefits for research repositories are quite different to those for learning and teaching repositories and anyone involved in promotional activities needs to be fully aware of these, particularly if the repository covers both functions.

Workshops & Training

Once the repository is in place running specific workshops and **training on the use of the repository** is a logical next step. FTaking authors through the processes of submission systems, or covering topics such as copyright, or the

principle of open access are worthwhile sessions to run. It is also worth trying to get the repository represented on existing training programmes running within the institution. This prevents duplicating effort and there are several benefits to tapping into existing training opportunities. For example, postgraduates often receive a variety of training sessions in preparation for writing their theses and dissertations and this would be an ideal opportunity to introduce them to the details of submitting an e-thesis to the repository.

Further Resources:

Planning Checklist

Posters

Printed media such as posters, leaflets and postcards can also be a useful way of promoting the repository to faculty. These can be distributed at events and presentations about the repository or can be sent out as an internal mailshot to academics. Again, keeping the message simple and offering well designed materials with a high visual impact will ensure the repository is presented in the best possible light.

Using the Media

A final opportunity for internal promotion of the repository is to use existing sources of internal media. Placing articles, news items and profiles in internal newsletters can improve the visibility of the repository. Internal media for staff, students or alumni can be considered and both hard copy and online options should be utilised. Email campaigns to either a small group of academics, or to the wider university community may also be worth considering. As your repository service develops, don't underestimate the power of a good news story: '500th item deposited', or 'geographers go open!', or 'university's top research theses now available to the world'.

Using social networking technology

A range of technologies can be utilised to publicise new deposits including syndicated feeds (such as RSS), twitter or networking sites such as facebook. This is particularly useful to get open educational resources into the public eye and most repositories can utilise feeds. Some of these activities can be time consuming and may need to rest with the original depositor but those that can be automated within the repository deposit workflow should be investigated. Several of the projects in the **JISC/ HE Academy UKOER Pilot Programme have produced guidelines on how to get the best of such technologies**[71] and this **blog post from JISC Cetis highlights which web 2.0 tools**[72] **have been used to support a** variety of functions, including publicity.

Technical Framework

In order to create an effective digital repository it is important that the technical set-up process is planned in detail. This section forms a technical framework, guiding you through the key decisions, options and processes involved in the creation of a repository infrastructure, such as:

- Defining **requirements**. Without a requirements specification informed decisions cannot be made relating to choices of repository platform and environment

[71] https://openeducationalresources.pbworks.com/Pilot-Programme%3A-OER-Release-Outputs

[72] http://blogs.cetis.ac.uk/johnr/2010/03/30/the-use-of-web-20-tools-in-the-ukoer-programme/

- The **installation** of a repository platform which may require the purchase of hardware and software, or could involve negotiating a hosting contract
- **Integrating** the repository with other systems such as local authentication systems
- **Testing** the repository to ensure that it works as expected, and fulfils all the criteria set out in your requirements specification
- Creation of technical policies for long-term aspects such as **metadata**, **workflows** and **file formats**
- Technical promotion of the repository. This is important to ensure that other systems such as external **search engines** index or **harvest** content properly.

Platform Choices

Deciding which repository platform to use is often a daunting task. The choice can be broadly split into six options:

1. Open source hosted locally
2. Open source hosted locally, with help from a service provider
3. Commercial software hosted locally
4. Open source hosted commercially
5. Commercial software hosted commercially
6. Create your own repository software

The choice is often dictated by local policies and practice, but it is worthwhile taking the time to define a set of specifications that you would like the solution to meet. Having a detailed set of **requirements** will help you to judge the suitability of each option.

A list of common repository platforms is available on the RSP website: http://www.rsp.ac.uk/repos/software#packages

Running open source software is often seen as a cheap solution as there are no costs in terms of software purchasing or on-going licensing. There are, however, costs involved in terms of staff development time, ongoing systems administration and upgrade efforts, and debugging problems. Often this relies on the knowledge of local staff, and the good will of members of the relevant repository community's technical support email list. In contrast, hosted solutions can be easier to budget for since they normally have fixed costs, and all maintenance and upgrades are usually included in the package.

Open source solutions can offer more scope to customise the software to local requirements. Customisations other than simple branding and workflow settings are often either unavailable from commercially hosted vendors, or are expensive custom developments. It should be noted that extensive local developments can sometimes be costly in the long-term as they will often need to be rewritten for each subsequent version of the software.

If you wish to run an open source repository platform but would like some support, then it is possible to buy hosting services for popular repository platforms such as DSpace and EPrints, or to buy support services for the software. Often the companies that provide commercial support and hosting for open source solutions are part of the

platform's development community, so by making use of these services you are helping to pay for future developments of the software.

Typical reasons for using external services may include specific expertise as well as systems infrastructure management, scalability and resilience, an institutional preference for outsourcing, or because a cost analysis shows it would be more cost effective than providing a similar service in-house.

Considerations when choosing a repository platform

The following list of bullet points may help you to define your requirements:

- Does the solution fulfil identified functional requirements of what it should do? Try to create a list of essential and desirable features as a baseline for comparison
- How much does it cost? Consider the initial purchase cost and ongoing annual service agreements, and compare it to local staff and equipment costs
- Are there many others users of the product? What do they say about it? Can you visit them and talk through their experiences with the software?
- What are the ongoing development plans for the product? Is there an exit strategy available to move the repository elsewhere should this be desired?
- What statistics are available from the product?
- Can the software be integrated with local systems, for example authentication systems?
- What training is offered for the system? Are there any additional costs associated with it? Training is often essential for effective use of the system

If considering a hosted option:

- What service level does the company offer in terms of 'uptime' (the availability of the service) and speed of response when reporting bugs or service issues? What hours of service does the help desk operate? What backups are taken, and what are the possibilities for restoring old items?
- What limits does the provider place on use of the product? Is there a limit on storage or traffic, and if so, what happens when these are exceeded? Consider institutional requirements of the product today, next year, and in future years
- How much can the product be customised, and how much of this can only be done by the service provider? Are there any costs for customisations?
- Does the service provider offer the opportunity for a test account to try the service before buying?

If considering an open source option:

- Does the software architecture fit that of other systems that you run, or your staff are knowledgable about? For example DSpace and Fedora are Java-based systems, whereas EPrints is a Perl-based system
- Is there a local user community for your platform that can help with any issues?

Software Skills Needed

The **installation, customisation** and ongoing **maintenance** of a digital repository generally requires the same technical skill set, and are usually performed by the same staff. In principle, installing and customising a repository is a one-off task, requiring a relatively short burst of intense activity by staff with appropriate IT skills. It is therefore advisable for the relevant staff to be assigned full-time during this phase of the project. The period for which they will be required depends on the software being used, the degree of customisation, and system architecture. The closer to an out-of-the-box product you go, with its default configurations and recommended platforms, the quicker the process will be.

Ideally, technical staff who have experience with the software and hardware platforms required by your **choice of software** should be engaged or assigned to the project. If no one suitable is available, time and resources for relevant training or recruitment should be allowed. Software providers often run technical courses on their products so it may well be worth sending staff on them.

The popular software choices tend to be fairly quick to install, taking the order of a couple of days for a standard configuration. This assumes you do nothing in the way of customisation more than the basics of adding institutional name and logo, the name of the repository, and the name and email address of the administrator. More time should be allowed if the repository is to be matched with an institutional or departmental web look and feel. This could be anything from a week to a month. As stated elsewhere, it important to note the changes made, because they may need to be repeated them during a future software upgrade.

If the repository is to be pre-loaded with material, for instance extracts from PubMed Central or some other resource, time should also be allowed for this.

Much of this commitment of resources can be avoided by outsourcing, either by hosting the repository with a suitable service provider or bringing external people in to do the in-house setup. However, time for setting specifications, liaison and project management should still be allowed for, and consideration should be made of sending staff on any product-specific courses for the sake of ongoing maintenance.

Specific skills

The specific skills required by technical staff will vary depending upon the choice of hardware platform, operating system, and repository platform. Local technical staff will probably have preferences over hardware and operating system, so it is advisable to make use of existing skills and use these systems.

There are generic skills that apply to all repository platforms. These include:

- HTML (HyperText Markup Language)
- Web page design
- CSS (Cascading Style Sheets)
- SQL (Structured Query Language)

Specific repository platforms will have their own requirements depending upon which programming language they are written in. DSpace and Fedora are written using Java, while EPrints is written using Perl. The technical documentation provided by the chosen repository platform should list the technical requirements it has, and therefore the skills required.

Hardware and Operating Systems

Hardware

In order to run a repository locally (as opposed to buying a hosted option) hardware is required. Repositories are typically run on servers housed in an institution's computer room which will be specially designed for the job (air conditioning, networking etc). Repositories can be installed on dedicated servers, shared servers (perhaps shared with a web server) or as virtual machines on large cluster. Institutions should have a policy for how they prefer to host services and servers. This policy may take into account storage and processor requirements, expected load on a server, or resiliency requirements (e.g. does it need to have a guaranteed 24x7 99.999% uptime?).

Most initiations, especially in the first few years when their repository is only sparsely populated, will find that a basic or moderately specified server will perform sufficiently for them. When considering hardware requirements, consider:

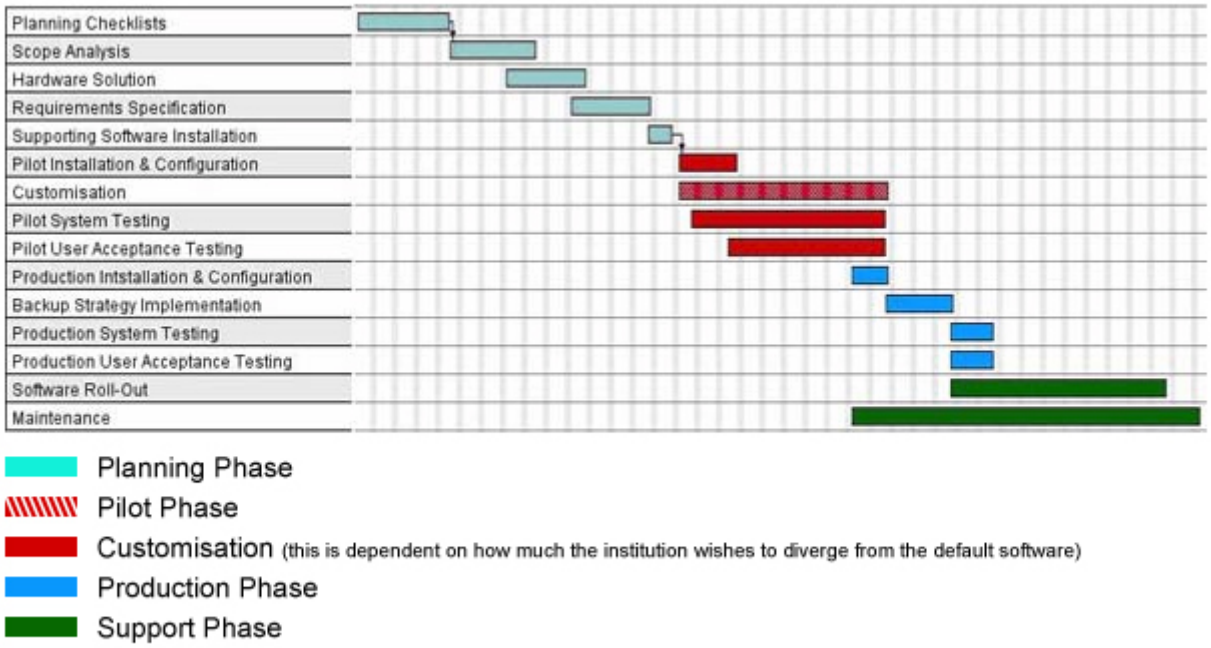
- How long you wish the server to last (3-5 years?)
- How many items are likely to be deposited in that time period
- What the average size of each item will be (to calculate disk requirements)
- What hardware specifications the chosen repository platform recommends

Maintenance of the server should be a standard part of IT services. Consideration should be made as to what will happen if the server breaks down or needs to be shut down for maintenance. Ideally, everything should fail-over to another machine.
Server, operating system, and systems software upgrades should all be budgeted for. If a rolling programme of upgrades is already in place it should be ensured that notifications well in advance of other upgrades that could have implications for your repository software. Additionally, upgrades may need to be initiated an upgrade if performance is degrading to an unacceptable level due to growing use of the repository.

Installation and Customisation

Installation

Although software installation procedures can vary between **repository platforms** many of the initial steps prior to installation are similar. Technical staff and repository administrators should be aware of steps such as checking technical requirements of the software, understanding installation instructions, user and system testing, migration from development to production platforms and backup strategies. This list is by no means exhaustive but highlights some areas that should be considered. The Gantt chart below gives an indication as to some of the areas a repository team should consider throughout a repository's lifecycle giving approximate timescales for each phase.



Gantt chart showing relative time scales for introduction of repository software service

Planning Checklists

Planning checklists are used to identify the key phases in a project lifecycle. The use of a planning checklist allows the team to focus on what goals/targets need to be completed in order for a project to run successfully. The above Gantt chart is a common method that is used by managers to illustrate the phases, activities and dependencies of a software project's work breakdown structure.

Project Scope Analysis

Defining the scope of a project is often a critical part in the project lifecycle. The scope defines what the project aims to achieve throughout its lifecycle and can be used as a marker at the end of a project to determine if the project outcome has been successful. It is often useful to define a mission statement which defines areas like; project deliverables, project objectives, project assumptions and project constraints.

Hardware Solution

Choosing the correct **hardware** for a repository installation is often dependent on a number of factors. Some of the overriding factors that will often determine the hardware route an institution takes are listed below:

- Current hardware set-up at the institution. IT departments will often prefer to choose similar hardware to the current systems. It is not uncommon for the repository to share resources with other institutional systems
- Available funding for the repository will have an effect on what hardware can be purchased
- Estimated usage of the repository. Is there a large storage requirement? Is there a need for a high performance, resilient solution?

Requirements Specification

Requirements specification involves identifying the conditions that the repository solution must meet to satisfy the various users or stakeholders of the system. This can be a very time-intensive phase in the repository lifecycle. It is important to capture the requirements precisely during this phase of the project as the success of the repository will often depend on the accuracy of the requirements. In many software projects requirements analysis can be broken down into:

- Requirements Gathering (determining the requirements of the users of the system)
- Requirement Analysis (analysing these requirements, resolving issues, determining use cases)

Supporting Software Installation

Repository software such as DSpace and EPrints often require additional third party software to be installed on hardware before installation of the repository can take place. It is important to schedule supporting software installation into the software lifecycle and make provision for any issues you may encounter while installing these supporting systems.

Backup Strategy Implementation

A good backup strategy is essential to ensure the repository is protected against unforeseen events whether this is for disaster recovery or to restore files which may have been accidently deleted or corrupted. For many organisations backing up the repository may be a simple case of incorporating the repository backup into the institutional backup strategy. Whatever method is chosen it is important that regular checks are performed to ensure data can be recovered in the event of a problem.

Repository Roll-Out

Providing all the phases leading up to the repository roll-out have been carefully considered and completed successfully, the rollout of the repository across an institution should be free from issue. Repository administrators and IT staff should be prepared to field questions regarding the repository and in many cases my need to provide hands on support and training.

Maintenance

Throughout the project lifecycle the repository will require **maintenance**, whether this be to fix a bug, install a new patch or feature request or upgrade to a new version of the repository software. It is important that major changes be made to the development system before modifying the live system. In this way, if any changes cause issues in the development system, the live system will still be operational. System maintenance is part of any software lifecycle and IT staff need to be aware that support for the repository does not end when the software is rolled out.

Production Customisation

It is desirable, given the efforts that have been made on its behalf to retain a portion of the institutional IPR, for institutional repositories to be clearly branded as belonging to a particular organisation. Basic issues include picking an appropriate name, including the institutional logo and visual identity, and in some cases incorporating the local standard full web template.

Customisation challenges

There are risks when **customising** your repository interface such as losing track of changes or problems encountered when upgrading your software. To avoid these kinds of issues the best advice is meticulous documentation both within code or scripts, and, if at all possible, in offline documents as well.

In many cases the introduction of a standard operating practice when it comes to introducing changes will aid in the integration of new technical or administrative staff, as well as protecting against issues around staff succession and replacement. Hosting a mirror test server where new implementations or changes can be introduced and any issues explored before going live is a simple but effective route to trouble free upgrading.

Configuration and Development

Configuration and development are two methods by which you can customise a repository to ensure that it fits requirements. Configuration refers to options that can be changed in repository platform settings, whereas development refers to larger changes that require development time and effort.

Configuration

Most repository platforms contain basic configuration settings. These typically include settings such as:

- **Look and feel/branding:** Colour scheme, logos, graphics, fonts
- **Metadata fields:** Which **metadata** fields are used to store which bits of information
- **Search options:** Which fields to enable for search and browsing
- **Authentication methods:** In-built username and password system, or an external provider such as Shibboleth or LDAP
- **Email settings:** Allows email server configuration and the address from which system-generated emails should be sent
- **Submission forms:** Change the metadata elements that are collected during the submission of an item
- **Workflows:** Adjust the stages and users responsible for different parts of the **deposit workflow**

Configuration settings are usually set in configuration files on the server or can be set via an administrative interface. It is useful to record configuration changes in a changelog so that the can be re-applied them if necessary after a system rebuild or upgrade.

Development

Sometimes there may not be a configuration option available to achieve the requirements. In this situation there are two options: change your requirements, or perform some custom development. For example requirements may dictate that PDF files should be viewed in an embedded frame within the repository with the metadata showing in a side panel. If this is not an option provided by the repository platform and requirements cannot be changed, then that feature will have to be developed.

If a proprietary/closed-source repository platform has been chosen, then the only option will be to contract the work to the commercial company or one of their approved developers. If the company feels that this is a useful feature that will be of use to other customers they may charge little or no money for the development, or ask that the institution sponsor the development. However if they feel that no other customer will require that customisation, then they will likely charge the full development cost.

If you are using an open source repository platform, then you will have the ability to perform the development work in-house. Most repository platforms provide either professional development services themselves, or have a list of recommended service providers.

You may also find other users of the software are interested in developing the same feature and would be happy to collaborate in order to reduce the development cost. If a new feature is to be developed it is worth informing the community of your intentions in order to find potential collaborators, and to see if it is possible that the development efforts may be included in a future release of the software. The ideal solution is that the new feature is general enough to be included in the next release of the software as this will reduce ongoing development costs. By consulting with other members of the user community, plug-ins, upgrades or workarounds that could be used to achieve the same outcome without having to perform any development work can often be found.

Each open source repository platform has its own method of suggesting new features and contributing code, so if it is worth finding out how the chosen repository platform software community works and how best to work with it.

The probability of having to undertake development work can be minimised by ensuring that the **requirements specification** is complete for your repository. This ensures that the best match for the specified requirements is found. Where there are differences, it will assist in creating estimates of the development work required as the features that require development work will be identified.

Comparison

Configuration of the repository platform is often quick and easy in comparison to developing new features. When new features are developed both the initial and ongoing developments should be budgeted for. If the development can be contributed back to the community so that it becomes part of the software, then this can reduce that cost. However if a specialised development is required then it is likely that the development will have to be continued over future versions of the software. If this is not budgeted for the institution may be locked into an increasingly old version of the repository software as upgrades cannot be afforded because of the implications for development work.

Pilot and Test Services

When **installing** and **configuring** a new repository, it is often good practice to have a pilot installation that can be used as a test bed for a repository development without adversely affecting a live installation. At project conception, a pilot system will be used to change and configure the software, perform system testing and initial user acceptance testing until the repository is at a stage to be used as a production system. In latter stages of the repository lifecycle, the initial pilot installation will often grow into a development platform which can be used for testing upgrades and patches before applying them to the live institutional repository.

Pilot System Testing/Functional testing

System or functional testing exists to ensure the system performs functionally as expected, and to iron out any bugs which may be prevalent in a new repository installation before it becomes a live service. It is sensible to dedicate a significant amount of time to this type of testing to ensure the service offered to users is as trouble free as possible. Bugs that are found in live systems will often cause annoyance to users and can harm a system's reputation.

It is useful to create test plans that exercise each function of the software to ensure that it performs correctly. If a problem is identified later, the test plan can be updated to ensure this bug is tested for. The test plans can be re-used after each upgrade to ensure that the system continues to function as expected, and that old bugs have not been reintroduced.

Pilot User Acceptance Testing

In contrast to system testing, user acceptance testing on a repository is performed by the people who will use the system on a day to day basis, ensuring it meets with their requirements. Feedback from a subset of the users who will

use the repository can be used to ensure that flaws and missing features can be rectified before the system becomes live.

Production System Testing and User Acceptance Testing

Once the production system is live it is still important to perform both system testing and user acceptance testing to ensure bugs haven't crept into the repository when moving from the pilot system to the production system. As the repository becomes live and is opened up to a greater user audience, there is a good chance that bugs will be found which may not have been identified in the pilot system.

Cover Sheets

Many visitors to repositories come via search engines to full-text files or learning packages, thus bypassing repository home pages and metadata records. A cover sheet is therefore often the only way to inform visitors of the repository's existence as the source of the information they have found. It is an opportunity to demonstrate the value of repositories to researchers and teachers, and a chance to market the institution. Cover sheets can contain branding such as logos or images: although it may become labour intensive to replace cover sheets if the repository and/or the institution are rebranded. The cover sheet can also be used as a way to link to the repository home page, and/or the metadata record for the item in the repository. The link to the metadata record may be all that needs to be considered, as the other information that could be included on the cover sheet might be contained in the metadata record. Cover sheets typically contain information about:

Meeting copyright holders' policies

If cover sheets are to be used, then displaying a copyright statement is highly recommended. Some publishers allow post-prints to be made available online whilst insisting upon standard phrases being displayed along with the text: the cover sheet is one way to meet those requirements. Even when publishers do not require such statements, a standard statement referring to copyright demonstrates respect for rights holders, just as others who are reading the content should show. A simple statement outlining the copyright terms under which the item is available in the repository can be included on a cover sheet. Consider that this policy might change over time so a standard statement referring to a policy on the institution's website may be preferable. Different items in the repository may have different use/re-use or repurposing options, with some being more restrictive than others (i.e. some learning materials may be fully 'open' whilst others may be for use within the institution only) and the cover sheet is one place that this can be recorded.

Information about which version is presented

In relation to research outputs there is often little information in the metadata record as to which version of the article is held in the repository, even whilst the final, published version is referred to. The cover sheet is an opportunity to redress this, but even if the metadata record does make it clear which version the full-text represents, the metadata record may not have been the reader's entry point to the text. It is often difficult to describe the version held in the repository in either a cover sheet or the metadata, owing to the confusion over terminology and authors' imprecise records. However, referring to the published version elsewhere can make it clear that yours is not the published 'version of record'. It may also be worth pointing out that access to the published version may require a subscription, which explains the reason for making an alternative version available.

In relation to learning materials it may be important to provide version material in your cover sheet and metadata because in some subject disciplines this information may date very quickly and, as in the case of medical subjects may

even be dangerous if out of date. A versions policy and preservation policy (LINK TO PAGE) is crucial to deal with this and there may be a policy to remove out of date material and replace it with new content.

Pedagogic context for learning materials

Some institutions prefer to utilise 'pedagogical wrappers' to accompany learning materials within repositories as they feel that the context (intended use) should be incorporated to guide other users. This 'wrapper' itself may constitute the main item/record within the repository or be held elsewhere with links to the items within a repository. The Open University Cloudworks service is currently being investigated as a **potential 'pedagogical wrapper'**[73]. This kind of link could be incorporated into a cover sheet.

Helping others to trace the published version of research outputs

It is the published version of the research output, the "version of record" that anyone wishing to cite the article is likely to want to cite (see page 9 of the **VERSIONS toolkit**[74]), and the cover sheet can be used to give the reader the information needed to trace and cite the published article. It would be a shame for an author to lose a potential citation because the repository version of the article itself did not contain enough information to make it easy for the reader to cite. If possible, it would be ideal to put a full reference for the repository version and for the published version onto cover sheets, making it extremely easy for their works to be cited.

How to attach/display cover sheets

Further information on creating cover sheets on the fly at point of user download

iText[75] is built into your own applications so that you can automate the PDF creation and manipulation process.

EPrints extension CoverPDF[76] - automatically adds cover pages to PDF documents. Cover pages are generated when eprints are uploaded and contain a mixture of set phrases and text derived from the eprint metadata. Whether a page should be generated and the detail of what should appear on it are configurable for different types of eprint. These cover pages are then stored alongside the documents, and merged with them at delivery. **Bepress Digital Commons**[77] software also does this.

Example cover sheet[78]

A cover sheet might be page 1 of the file itself, added by authors before/whilst depositing, or by repository staff after deposit. Alternatively, the cover sheet might be generated 'on the fly' at the point the document is requested by the reader, rather than being a part of the file itself. An automatically-generated cover sheet would not interfere with any automated metadata creation tools, nor with authors' wishes to have the front page of their work displayed in rollover images just as they intended it to look, and it would mean no extra work for depositors or repository administrators in attaching such cover sheets. Plus, any later branding or policy changes can be altered in the source

[73] http://cloudworks.ac.uk/cloud/view/4186

[74] http://www2.lse.ac.uk/library/versions/

[75] http://www.lowagie.com/iText/

[76] http://files.eprints.org/465/

[77] http://www.bepress.com/ir/features.html

[78] http://oro.open.ac.uk/17386/

information for such cover sheets. Another way to potentially automate cover sheet creation would be to offer an 'add-a-cover-sheet' option to authors as they deposit and upload the file. In this way, repository staff are not involved in lengthy processes, authors are prompted to include appropriate information in the files themselves, and authors get to choose whether they want a standard cover sheet or not.

Conclusion

There are many purposes that cover sheets can serve in a repository, but it is not necessarily certain that they should be used. Having given consideration to all the issues around cover sheets, you will need to balance your ideal scenario against technical capacity and workflows, to decide on what it is actually going to be possible. Cover sheets are something that could be re-addressed in the future as the repository matures and technological solutions are developed.

Integration

More information is available from the SWORD website: http://www.swordapp.org/

It is important to consider how the repository will integrate with other institutional systems as there is often shared data to consider. Repositories containing learning and teaching materials may have to integrate with the institutional Virtual Learning Environment (VLE).

It is also important to think about how external services will be able to integrate with the repository - particularly important to assist harvesting of your content and pushing information/content out to the wider global community. Standards and interoperability are crucial to facilitate this. Social networking services present excellent opportunities to 'advertise' and promote your content, but can be time consuming.

SWORD

In order to make repositories interoperable with each other, and with external systems, they must adhere to agreed protocols. SWORD is a common interface for depositing items into a repository.

What is SWORD?

SWORD stands for Simple Web Service Offering Repository Deposit. At its basic level SWORD allows two types of interaction with a repository:

1. Query a repository to find out what collections a user can deposit items into
2. Perform a deposit into a repository collection

Rather than implementing a wholly new standard, SWORD is a profile of the **Atom Publishing Protocol**[79] (AtomPub). AtomPub is often used by blogging software to allow the remote posting of items into a blog, and also forms the basis of other well-known standards such as **GData**[80], the Google Data API.

SWORD uses AtomPub but extends it with repository-specific extensions that adapt it to make it fit with the way that we use repositories. This brings the benefit of using a widely-adopted standard, but with the specifics required by the repository domain.

[79] http://atompub.org/

[80] http://code.google.com/apis/gdata/

How does SWORD work?

A typical SWORD deposit will take place in two steps.

1. The first step will typically require the user to provide their username and password that allows the repository software to construct a service document. The service document describes the collections into which a user may deposit items
2. Once a user has decided which collection they wish to deposit an item into, they send a file to the deposit URL of that particular collection. The repository will then ingest the item and put it through any workflows that it is set up for

Users do not interact with SWORD directly; rather they make use of a client tool, in much the same way as we interact with websites through a web browser. SWORD clients can be:

- Standalone applications integrated into web browsers
- Built in to publishers workflows to deposit items automatically into an author's institutional repository
- Incorporated into other software products such as word processors
- Included in websites such as social networking sites where a deposit could also trigger alerts being sent to colleagues alerting them to the new item

One of the benefits of using a standardised protocol such as SWORD for repository deposit is that a user is free to choose which SWORD client they use depending upon their working preferences, and an institution is free to choose which repository they implement as all clients and repositories will interoperate. A list of clients and code libraries to write your own clients are available from the **SWORD website**[81].

How can SWORD be used?

There are several scenarios where SWORD could be used:

- **A desktop deposit tool:** Rather than interacting directly with a repository, authors could deposit via a user-friendly desktop application
- **'Save-As' in a word processor:** Authors could deposit an item directly into a repository by using a 'Save-As' plugin for their word processor that deposits the item in a repository rather than saving it to a disk
- **Multiple simultaneous deposit:** If a user needs to deposit their work in both an institutional and a funder's repository, they could deposit once and SWORD could be used to perform the other deposits
- **Deposit by machines:** Laboratory equipment might deposit experiment results into a repository without requiring human intervention

What is the difference between SWORD and AtomPub?

SWORD has added a number of extensions to AtomPub to allow it to fit in with the way repositories work. The extensions are:

[81] http://www.swordapp.org/

1. Allowing the depositor to deposit on behalf of another user. By configuring an optional setting, the depositor can say that they are performing the deposit on behalf of another user. Only if the user has the right to deposit on behalf of that user will the deposit be accepted. This facility may be used where an agent (person or machine) is depositing for another user
2. Rather than depositing single files, SWORD allows the deposit of a package, along with a description of the package. A Package might be a zipped file of metadata and content files, such as an IMS package
3. Extra metadata such as collection policies can be described in the service document
4. There are developer-friendly extensions to request verbose output about what the server has done, and to perform 'noOp' (no operation) deposits which are only tests, and not actually ingested into the repository

For a more detailed overview of harvesting protocols and related issues, see the **Linking UK Repositories report** (Swan & Awre, 2006)

http://bit.ly/nDvuIc

OAI-PMH Harvesting

In the repository domain there are several standards that are widely implemented and ensure interoperability, the most well known being Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) which is used as a common interface for harvesting **metadata** from repositories.

What is metadata harvesting?

When external search services want to be able to index a website, they often crawl it by following links in order to find each web page, and then extract the text from the pages they find. This process works well but lacks the ability to make use of structured metadata and advanced search facilities based on metadata fields, or more advanced interrogation through methods such as data mining. It is also hard to identify new content without either another complete crawl of the website, or by making use of sitemaps.

The alternative to web crawling is harvesting. Harvesting involves making queries to the repository about its content, and receiving replies that contain lists of items, and item metadata. The **Open Archive Initiative Protocol for Metadata Harvesting**[82] (OAI-PMH) is a machine-to-machine (m2m) interface that is specifically designed to facilitate the harvesting of metadata from Open Access repositories, and the vast majority of repositories provide this option.

Most repositories provide an OAI-PMH interface and are designated 'OAI-compliant'. When a repository adheres to this protocol, some or all of the metadata that it holds for all the items in its collection is exposed for harvesting by service providers. The returned metadata usually includes a URL for the item, and for each full-text file which can then also be processed if required.

Principles of OAI-PMH

Systems that provide information via OAI-PMH are known as data providers, and systems that harvest information using OAI-PMH are known as service providers (as they provide new services with the data).

[82] http://www.openarchives.org/pmh/

OAI-compliant repositories have a base URL in addition to the URL for human users. For instance, Aberystwyth University's **CADAIR repository**[83] has the **OAI Base URL**[84]. On its own, an OAI Base URL simply returns XML containing an error message. This is because the protocol expects instructions in the form of a 'verb' and other arguments to be appended to the URL.

OAI-PMH Verbs

The simplest case is the verb '**Identify**[85]', which returns identity information about the repository.

Altogether, there are six OAI-PMH verbs, some of which require additional arguments:

Identify	Returns information about the repository Example - **RRP – Roehampton : http://rrp.roehampton.ac.uk/cgi/oai2.cgi?verb=Identify**
ListMetadataFormats	Lists the metadata formats supported by the repository. The minimum requirement is oai_dc (Dublin Core) Example - **CADAIR - Aberystwyth University: http://cadair.aber.ac.uk/dspace/**
ListSets	Lists the sets provided by the repository (e.g. departments, subjects) Example - **e-Prints - University of Southampton: http://eprints.soton.ac.uk/perl/oai2?verb=ListSets**
ListIdentifiers	Lists record identifiers, dates and any other headers for each deposited item. Requires the argument 'metadataPrefix'. Results can be limited to specified subsets using the 'set' parameter. Results can be limited to specified time periods by using the 'to' and 'from' parameters. Example - **ePublication Library - Chilbolton, Daresbury, and Rutherford Appleton Laboratories: http://epubs.cclrc.ac.uk/oai?verb=ListIdentifiers&metadataPrefix=oai_dc**
ListRecords	Harvests metadata records from the repository. Requires the argument 'metadataPrefix' - metadataPrefix=oai_dc should suffice. Results can be limited to specified subsets by using the 'set parameter'. Results can be limited to specified time periods by using the 'to' and 'from' parameters. Example - **ePrints – Nottingham: http://eprints.nottingham.ac.uk/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc**
GetRecord	Gets an individual metadata record from the repository. Requires the arguments 'identifier' and 'metadataPrefix'. Example - **ePrints - White Rose Research Online: http://eprints.whiterose.ac.uk/cgi/oai2?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:eprints.whiterose.ac.uk:937**

Note: The verbs and their associated arguments are case-sensitive.

When the results of an OAI-PMH query are large, they are often split into chunks of records. Each chunk ends with a 'resumption token' that can be used to retrieve the next chunk.

OAI-PMH installations can be set up to return results using a variety of metadata schemas. As a minimum, all OAI-PMH servers must be able to return results using the unqualified simple Dublin Core (oai_dc) schema, and this is all that many repositories or packages offer. However, they can provide as many or as few additional schemas as they

[83] http://cadair.aber.ac.uk/

[84] http://cadair.aber.ac.uk/dspace-oai/request

[85] http://www.british-history.ac.uk/oai/oai.aspx?verb=Identify

wish. For example EPrints and DSpace both support the United Kingdom Electronic Thesis and Dissertation Dublin Core (uketd_dc) metadata format required by **EThOS**[86].

OAI Registration

While a repository can be harvested simply by providing an OAI-PMH interface, registration provides a useful means of promoting the visibility of your repository to service providers for harvesting. The Open Archives Initiative provides a service allowing your repository to be registered as a data provider in the OAI registry. The registry is a publicly accessible list of all OAI conformant repositories which allows easy discovery of data providers by service providers. When registering your repository, the OAI service will perform conformance testing to ensure your repository complies with OAI-PMH. If validation is successful, your repository will be added into the registry. The OAI will also periodically test your repository for conformance. If the analysis fails, your repository will be removed, and a notification email sent to the administrator detailing the reason for removal. This ensures the integrity of the OAI registry and your repository interface. Information on registering your repository with the OAI can be found at the OAI website. The OAI-PMH interfaces provided by the major open source repository platforms are created in such a way to ensure that they are always compliant.

Although registering with the Open Archives Initiative will assist in increasing the visibility of your repository to other service providers, **direct registration** with these services is also possible to guarantee your repository is harvested by them. Service providers which require additional registration are Intute Search, OAIster and OpenDOAR.

Useful links

Open Archives Initiative: http://www.openarchives.org/ - The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content

Open Archives Forum: http://www.oaforum.org/ - The Open Archives Forum provided a Europe-based focus for dissemination of information about European activity related to open archives and, in particular, to the Open Archives Initiative. They have a useful online tutorial

OAI Registration: http://www.openarchives.org/data/registerasprovider.html - Information on registering with the OAI as an OAI-PMH conformant data provider

Open Archives Initiative - Repository Explorer: http://re.cs.uct.ac.za/ - This site presents an interface to interactively test archives for compliance with the OAI-PMH

RSS/Atom

RSS and Atom feeds are both standards for providing feeds of information from repositories. Both these standards enable internet users to keep up to date with new content on numerous websites without having to visit each individual site. Repositories can be configured to provide information to users about new repository content via RSS or Atom.

How do RSS and Atom work?

Both RSS and Atom are XML-based standards that work by pulling information together for the user, for display remotely in feed readers. Websites can offer feeds for the entire site or for specific parts of the site. For example, in a repository, there may be a feed to display any new article that gets added, and there may also be feeds that only show items for a specific author or a specific department.

Software and services designed to manage and display RSS and ATOM feeds are available as desktop software, web-based services and services or software for mobile devices. Software or services that

[86] http://ethos.bl.uk/

manage RSS and Atom are commonly called feed readers or aggregators.

In practice a feed is an alternative version of a webpage that displays all of the latest additions to the website in RSS or Atom XML. The URL for this XML page can be added into feed readers which will display the XML in a user-friendly way and update users about any new content on the website.

Websites that offer **RSS**[87] or Atom feeds for their updates normally display a link saying RSS, Atom or Feed or the RSS logo: somewhere on the relevant page.

RSS and Atom feeds usually contain metadata about new content such as title, url, author, abstract etc, but they can also be used to contain the full text of the content, this is most common in RSS and Atom feeds from blogs.

The feeds do not work in real-time, there is a delay between updates on a site and the item appearing in a feed.

What's the difference between RSS and Atom?

RSS and Atom are competing standards and there are technical differences in the ways that the two standards work. However this should not be a major concern to repository managers since on the surface, both do a largely similar job and sites can offer either or both as options for people who want feeds of the latest content.

The Atom feed standard is related to the Atom publishing protocol used by SWORD.

How can RSS and Atom be used?

At its simplest RSS and Atom feeds can be used to update users on change to a website via software such as feed readers. However RSS and Atom are very versatile and can be used in a number of ways both by end-users and by other pieces of software.

Here are some examples of RSS and Atom being used in interesting ways for repositories:

- Feeds can be embedded in other websites. So for example a departmental website can display all the latest content added to the repository. Feeds can also be embedded as plain text or can be used to display more attractive visualisations of repository content. See **Les Carr's blog**[88] as an example
- Feeds can be combined using a service such as **Yahoo! Pipes**[89] so that feeds from a number of places can be offered as one feed. This may be useful for combining repository updates with updates from other institutional or subject sources
- Feeds can be used to pull content into other places. Les Carr has an example of this with **MS PowerPoint**[90]

To find out more about the differences between RSS and Atom, visit the **Atom Wikipedia page:** http://en.wikipedia.org/wiki/Atom_(standard)

[87] http://en.wikipedia.org/wiki/RSS

[88] http://repositoryman.blogspot.com/2008/11/more-things-to-do-with-repository-feed.html

[89] http://pipes.yahoo.com/

[90] http://repositoryman.blogspot.com/2008/11/more-repository-value-for-users-making.html

- RSS feeds can be used to push content into other places as well, some people have chosen to link their repository feeds to twitter accounts so that a message is sent to twitter every time a new item is added. This could obviously be done using subject feeds rather than the whole repository feed
- Services such as **Tabbloid**[91] can turn feeds into a PDF newsletter which can then be emailed to people
- The JISC/HE Academy **UKOER pilot programme**[92] investigated the possibility of using RSS feeds to submit items to the **Jorum Repository**[93] (the national service for learning and teaching materials). This provided an interesting testbed and is discussed in **John Robertson's JISC Cetis blog**[94].

These are just a few examples, there are a lot of ways RSS and Atom feeds can be used to provide useful services to the end-user.

Registering Repositories

A good way to let the world know about the repository is to register it with directories of repositories such as **OpenDOAR**[95] and **ROAR**[96]. The benefits of listing a repository are:

- Gain exposure: make other repository managers aware of it
- Increase visibility of content so that it can be used, cited and linked to
- Make use of the statistics that registries such as ROAR can provide
- Allow repository details to be included in sites that make use of the data in these registries, such as the **Repository66**[97] map mashup
- Some other indexing and search services will use the registries as a source of repositories that they will then harvest and expose

Specific sites that you may wish to register with include:

- **DSpaceInstances**[98] - This is page in the DSpace wiki, and lists repositories that run DSpace software. To add to the DSpace repository, click on the 'Edit tab' against the relevant alphabetical section to edit the page, then copy the format of adjacent entries

[91] http://www.tabbloid.com/

[92] http://www.jisc.ac.uk/oer

[93] http://www.jorum.ac.uk/

[94] http://blogs.cetis.ac.uk/johnr/2010/02/04/rss-for-deposit-jorum-and-ukoer-part-1-review/

[95] http://www.opendoar.org/

[96] http://roar.eprints.org/

[97] http://maps.repository66.org/

[98] http://wiki.dspace.org/index.php/DspaceInstances

- **OAIster**[99] - A search service for OAI-compliant digital resources, not just open access repositories. It harvests metadata for the content of these resources using OAI-PMH. OAIster's list of data contributors is used by Google Scholar to control its crawling of open access repositories
- **openarchives.org**[100] - This registry of OAI Base URLs is hosted by the Open Archives Initiative. One advantage of registering is that the functionality of your OAI Base URL is automatically validated
- **OpenDOAR**[101] - Directory of Open Access Repositories - OpenDOAR is a quality-controlled worldwide directory of academic open access repositories maintained by SHERPA, University of Nottingham. It has strict criteria for inclusion, the key conditions being that the repository must contain some full-text items, and it must be accessible without the need for a username or password. Each repository is visited by OpenDOAR staff for indexing purposes, rather than relying on automated analysis. As well as providing a simple repository list, OpenDOAR lets you search for repositories, or search repository contents. OpenDOAR data is used to control harvesting by Intute Repository Search and other search services
- **ROAR**[102] - Registry of Open Access Repositories - This worldwide registry is operated by the University of Southampton. Its strong points include statistics on the growth rates of individual repositories, cumulated globally. These are compiled using OAI-PMH. Associated services include statistics on file formats, and ROARMAP - a database of institutional and funders' Open Access archiving policies

Case Studies

Googlizing a Digital Library: http://journal.code4lib.org/articles/43 The Code4Lib Journal, 24th Mar.2008, No.2 (Jody L. DeRidder, 2008)

SHERPA's **Ways to snatch Defeat from the Jaws of Victory:** **http://www.sherpa.ac.uk/documents/ways-to-screw-up.html** has a list of common mistakes

Search Engines and Repositories

Search Engines

Search engines such as Google, Google Scholar, Yahoo! Search and Live Search use robots called **web crawlers**[103] to index web pages. They find web pages by following links, so it is therefore wise to check that the repository is browsable both by robots and people, and to check that its structure is suitable. If necessary, it is also possible to block robots from indexing certain pages or groups of pages.

If you are using one of the popular repository software packages such as DSpace or EPrints optimising the repository for search engine robots is less of a concern because of good default set-ups. This section will therefore mainly be of interest to those who have developed their own software or site, used an uncommon package, or heavily customised an installation.

[99] http://www.oaister.org/

[100] http://www.openarchives.org/Register/BrowseSites

[101] http://www.opendoar.org/

[102] http://roar.eprints.org/

[103] http://www.rsp.ac.uk/usage/robots#crawlers

How Web Crawlers Work

Search engines index websites using special programs called 'robots'. They usually have their own names. For instance, Google's main robot is called 'Googlebot'. Other terms you may encounter are 'web crawler' and 'spider', which reflect the way these robots work. Starting from a specific web page, the crawler follows all the page's hyperlinks to index other pages on the website, and often external pages too. In this way, the whole website is covered, although it may take some time to reach the lower levels of the page hierarchy, and some search engines do not guarantee to index every page. Some technical features of a website could prevent pages from being crawled. Also, there are ways of blocking certain pages from being indexed, if desired. These are discussed later.

General advice on making websites friendly to search engines can be found in Google's Webmaster Guidelines: http://www.google.com/support/webmasters/bin/answer.py?answer=35769 , and Peter Suber has prepared more specific advice on how to optimise repositories for Google crawling: http://www.earlham.edu/~peters/fos/googlecrawling.htm

Ensuring Browsability

Links are the key to successful web crawling. More specifically, it is static links that are the key. Dynamic links that are, for example, generated by an interactive search are likely to be unreachable by a robot. A related point is that URLs using arguments (i.e. the URL contains a '?') may be bypassed by some robots because they may regard them as transient dynamic content. Ideally, the links should be text links, although linked images are also usually acceptable. Links that use buttons to run JavaScript, PHP or other programmed functions will normally be ignored.

For effective web crawling, it must be possible to visit every page and document in the repository just by clicking on hyperlinks - without ever needing to type in text or to use buttons.

More information is available in the Web Server Administrator's Guide to the Robots Exclusion Protocol: http://www.robotstxt.org/wc/exclusion-admin.html and in the relevant Google Help Page: http://www.google.com/support/webmasters/bin/answer.py?answer=40360

See the HTML Author's Guide to the Robots META tag: http://www.robotstxt.org/wc/meta-user.html or Google's Help Page: http://www.google.com/support/webmasters/bin/answer.py?answer=61050&query=robots+meta&topic=&type= for more information on 'robots' meta tag options.

Website Structure

As mentioned earlier, web crawlers may take a while to reach the lower levels of a web page hierarchy, and some robots may only dig down so far. It therefore helps to keep hierarchies relatively shallow. This also helps usability for humans. A typical structure would have a set of 'Browse by...' options on the home page that link to lists of documents, thence to metadata pages for individual items, and finally to the full text.

Example - Browse by Year > 2007 > [List of titles] > [Metadata page] > [Full text PDF]

Performance can be partially improved by listing some items on the home page - typically 'Recent Additions', 'Popular Papers', etc. Such lists primarily provide a useful 'come back' feature that encourages people to revisit the repository, but fortuitously also causes the listed papers to be indexed sooner by web crawlers.

Blocking Robots

There are cases where robots need to be prevented from indexing a particular page or a group of pages. A depositors' login page would be a typical example. There are two methods of doing this that are accommodated by nearly all of the reputable search engine spiders:

'robots.txt' Files

This approach is the best method for blocking groups of pages, although it can also be used to block single pages. A plain text file named robots.txt is placed in the website's root directory that contains a set of instructions about robots to be excluded and/or pages to be ignored. Each block of instructions starts with a line specifying the 'User-Agent' to which the block applies, followed by one or more lines indicating the files or directories that are to be 'Disallowed'.

User-agent: *	The * indicates the instructions apply to all robots
Disallow: /login.php	Robots should not to index the file login.php
Disallow: /restricted/	Ignore all files in the /restricted/ directory tree

'robots' Meta Tags

This method can be used to block a single web page, although it may not be as reliable as the previous approach. Meta tag elements in a page's HTML <head> division provide information for robots such as authors' names, keywords, description, etc, and are not displayed on the visible web page. One meta tag - 'robots' - controls whether or not a web crawler (a) indexes the page, and/or (b) follows the links on the page. The example below blocks both indexing and link-following:

```
<meta name="robots" content="noindex,nofollow" />
```

...which can be shortened to...

```
<meta name="robots" content="none" />
```

Sitemaps.org

'**Sitemaps**[104]' turn the traditional relationship between search engines and websites on its head, by allowing websites to tell search engines what pages are in their site, rather than the conventional method of a search engine having to crawl the whole of a website to find new content.

A sitemap is a collection of XML files which, in their simplest form, can tell search engines what pages exist on a website, and when they were last updated. These files are typically built each night, and then compliant search engines can be 'pinged' (by visiting a specially formed URL) to inform them that the site map has been updated. Check with the repository software documentation to see if it supports sitemaps.

Technical Maintenance

A stable operational repository with minimal local customisations is likely to have fairly modest technical maintenance requirements. Probably the most important factor is to check the arrangements for backups, and ensure that the repository is backed up on at least a daily basis. Consideration should also be made as to how material would be restored from backups and what the turnaround time would be.

If the repository is maintained by a different party, for example your university IT department, then for this and other ad hoc troubleshooting and support it tends to be advantageous to secure a named contact - ideally the person who

[104] http://www.sitemaps.org/

installed the repository - rather than having to use a general IT helpdesk. There may also be advantages in negotiating a service level agreement to ensure prompt resolutions to any problems.

Software Upgrades

Software upgrades have larger technical requirements. Upgrades may require considerable effort, especially if major changes have been made to the software and/or the associated database, or the site is heavily customised. The timescale is very much case-dependent, but you may possibly have to think in terms of a week to a month, or more, of full-time technical work. It can help to do a preliminary scoping exercise to determine the tasks that need to be completed, and estimate the effort required. As with installation, if possible, it is beneficial to assign technical staff full-time for the duration of the upgrade.

There are three sorts of software upgrade:

- **Software patches** - These amend the installed software and usually fix specific bugs or security issues
- **Minor upgrades** - These typically correct bugs and/or add minor functional enhancements, and will usually incorporate any earlier patches. In many cases, nothing more is required than downloading and installing the upgrade file
- **Major upgrades** - These may entail changes to the data model (i.e. the way information is stored in the database), and/or radical changes to the user interface. Consequently, existing data may need migrating into the new data format, user interfaces re-customised, and possibly workflows modified along with advice to depositors.

File Formats FAQs

More information about file formats and repositories:
http://www.jiscinfonet.ac.uk/infokits/repositories/technical-framework/ff-faq

Whatever the type of upgrade, it should be handled systematically. Documentation will reveal possible implications such as software re-customisation and potential outages in service. For all but the simplest of patches and upgrades, a trial run on a development installation may need to be performed. This is where any customisation would be redone, along with development and testing of data migration scripts, and finally the running of acceptance tests. Using this experience planning can be made for the steps required for the final changeover. This will ensure that it can be completed in the quickest possible time with minimum disruption to end-users.

If re-customisation of software is required, the notes made upon original installation the software will be indispensible. It is good practice to thoroughly document the changes made for future reference.

Technical Policies

As with all repository policies, the technical policies should reflect the visions and aims outlined in strategic documents. Technical policies should align with other management policies and are likely to be shaped by decisions made at the requirements specification stage.

File Formats

The subject of file formats is an important topic for repositories, and as such, creating a robust **policy** at the inception of the repository service is important. However, file formats are constantly changing and policies therefore need to be flexible. This section contains good practice guidelines about considering and creating file format policies.

Why are file formats important?

Repositories are living archives. In terms of the support it must provide for stored files, it must take into account two important functions of the files it holds:

1. **Access:** The files are held so that users can access them. This means that they must be stored in formats that can be used by today's intended audience
2. **Preservation:** The files are held so that users in 5, 10, 50, or more years can still access them. This means that they must be stored in formats that can be used by future audiences, or in formats that can easily be migrated

These two considerations are not always complementary. A file format that is good for access today may not be a format that is easy to migrate, but a format that is easy to migrate may not be easy to read. Take, for example, a journal article. It could be argued that a good format for easy access today is a PDF file. However, were PDF files to drop out of fashion, it could also be argued that the PDF specification is not easy to migrate in the future to a new file format. On the other hand, storing an article in marked-up xml is good for long-term access as it can be easily converted to a presentation format. It does not, however, in itself make a good format for easy access by end users.

Useful tools

PRONOM: http://www.nationalarchives.gov.uk/PRONOM/ - run by the UK National Archives and provides a registry of information about file formats

DROID: http://droid.sourceforge.net/ - also created by the UK National Archives and allows the batch identification of file formats

JHOVE: http://hul.harvard.edu/jhove/ - a tool to assist with the identification and validation of different file formats

PLANETS: http://www.planets-project.eu/ - The EU funded PLANETS project produces tools to assist with the long-term preservation of digital content

Guidelines

Here are some general guidelines to help in the creation of a file format policy:

- **Collection or deposit policy:** If the decision is made that the repository will only accept deposits where the deposited files meet a strict file format criteria, then some users may be put off depositing their content. Ensure that users can deliver files in the preferred format, otherwise the risk is run of not being able to collect files. If users are not able to deliver the preferred file formats, consider putting workflows in place that allow the repository administrators to perform the file conversions. It is unlikely that this type of approach would be appropriate for learning and teaching materials as these could cover many formats, even within one package. Guidance about file formats and preferred formats could still be offered to depositors.
- **Store multiple versions:** As mentioned previously, file formats that are good for access today may not be good for access in the future. Consider therefore storing multiple versions of the file. If, for example, a file created by a word processing program is deposited a plain text version and a PDF version could also be included.
- **Ensure you know what file formats you hold:** If files are stored which are specialised and cannot be converted into more generic formats, ensure that precise details of the file format are archived along with the file. Such metadata might include the name and version number of the software used to create the file. If the software is specialised, archiving a copy of the software along with the item is good practice. If this is required, the same principles apply to the software. Ensure a copy of the operating system that is required to run the software remains accessible.

- **Consider your repository in the wider information environment:** Any new repository will typically sit amongst many other platforms run by institutions. These may include content management systems, documentation systems, virtual learning environments, video streaming servers and file servers. It is likely that the repository is intended as an archive for materials from many of these systems, however it may not be the most appropriate place from which people access materials. Consider a video file. It would be prudent to store an uncompressed archival version in the repository, but access to that video today may be better provided by making use of a compressed streamed version from your video streaming server

According to NISO: http://bit.ly/nh3SIV

Metadata is 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource'.

- **Offer multiple levels of support:** It may only be practical to support a few different file formats. This can be catered for by offering varying levels of support. Designating a few formats as 'supported' means that an attempt, or guarantee, will be made (depending upon your policy) to migrate these over time to ensure continuing accessibility. Other formats may be designated as 'known' which means they are 'recognised' but are not guaranteed to be migrated over time. Other formats may be described as 'unsupported'. In this case the guarantee is to preserve the file as it was when deposited

- **Preservation policy:** A file format policy should form part of a wider preservation policy. The preservation policy may influence your choice for file formats if it dictates time durations during which files must still be able to be read, versus a policy which just recommends 'best endeavours'

- **Data files:** If data is to be stored, ensure that the data is adequately described. For example if storing results from a questionnaire, it may be obvious that a column of data marked 'sex' which contains responses of either 'm' or 'f' relate to 'male' or 'female', however a column marked 'socio-economic status' with values 'a', 'b', 'c', 'd', 'e' and 'f' is not so obvious. Ensure a data dictionary, ontology, or coding scheme is stored with the data

- **Plan for change:** The file formats in use change over time, as do versions of software. It is important that the file format policy adapts over time to ensure that it stays up to date with current file formats and versions

- **Be practical:** Being overly-strict about file formats may mean collecting no files leading to an empty repository! A sensible approach must be used that weighs up the cost and benefits of different file formats and the effort required to convert between them.

Metadata

etadata is information about information or data about data.

Metadata is a vital element of any repository and objects stored in an institutional repository are always accompanied by metadata records. Metadata supports day to day management of repository content, keeps the repository usable, and allows information about the resource to be easily shared with other systems. Metadata can be collected at various stages during the ingest of an item into the repository.

Other pages in this resource outline the purpose of metadata in the repository, explore further information about the key metadata types and metadata standards that a repository needs to comply with and give summaries of more detailed metadata standards and application profiles which are available for repository managers to use.

Metadata and the Repository

In the context of an institutional repository, metadata is needed to facilitate discovery of your repository content. Assigning relevant criteria to content items in the repository:

- helps users identify resources
- helps repository managers organise content
- brings similar resources together
- distinguishes similar resources
- gives location information
- is essential to facilitate harvesting of repository content by external systems
- supports archiving and preservation

Repository managers need to consider how they intend to gather or create metadata and what it will be used for:

Metadata creation

Metadata can be collected in many ways; these include:

- Added by the author during the submission process (through their user profiles and submission forms)
- Derived from the deposited content using automated tools
- Added or edited by the repository manager or a cataloguer
- Added by users of the content

The metadata that needs to be created manually by the repository manager or by a cataloguer is typically administrative and bibliographic information (such as a subject heading).

Metadata use

Once the metadata is collected, it must be reliably stored and is used to manage the resource (for example, it can be used to review the strength of a repositories collection in a particular domain or to check the currency of the file formats). Certain technical protocols or standards, such as **OAI-PMH** can be used to enable external services to access metadata records and facilitate resource discovery by a much broader audience.

It is important to consider these two contexts together. Strong communication between repository administrators and technical teams is needed to ensure the requested metadata input matches the requirements for the local management of the repository's items and enables the repository to be harvested for external use.

Metadata Types and Schemas

Types of metadata

In the context of repositories, there are three main types of metadata:

- **Descriptive:** Facilitates resource discovery and identification; includes elements such as item titles, authors and keywords
- **Structural:** Describes how items relate to one another; particularly important for recording versions of an item over time and for dealing with complex items such as chapters within a book or learning packages
- **Administrative:** Helps manage the resource itself; it includes rights management and preservation metadata that records how files were created and the rights associated with them

Defining your metadata schema

Metadata schemas are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. Most schemas will cover all three types of metadata identified above. Repository administrators will need to consider their metadata schemas at an early stage of repository implementation. The schemas being used will vary depending on the types of content being stored, the repository's audience, and the intended use of the repository contents. An institutional repository manager is likely to need to define schemas for everything from the relatively simply text-based materials being received through to more complex multimedia objects. Specialist collections within the repository may benefit from support for a more detailed subject classification.

Further Information

Draft IRIScotland metadata agreement:
http://bit.ly/nEC1Si

JISC CETIS - Metadata and Digital Repositories:
http://wiki.cetis.ac.uk/Metadata

When defining your schema it is important to consider local needs such as departmental and research structures and any local decisions needed about subject fields. Most people will also need to extend their schemas to new types of materials as the repository grows. Local metadata schemas are often based around one or more agreed **metadata standards** and may conform to common **application profiles**. Realistically and in the first instance most people will work with the metadata scheme which comes with the out of the box installation of their chosen software, for example, out-of-the-box **DSpace**[105] uses. The default scheme should be carefully checked to ensure it meets and complies with basic requirements. Extra fields can be added and schemas customised over time but any change may involve retrospective editing of existing records and there is a need to retain the mapping to the export standards to aid interoperability and integration.

Metadata Standards

A range of metadata standards have been developed by the repository and archives community that can guide those wishing to establish a schema. Wherever possible, schemas should be open and comply with widely recognised standards to facilitate interoperability and broader reuse of stored resources. The most commonly used standard, and the one required for interoperability and harvesting via the OAI-PMH, is **Dublin Core**[106].

[105] http://dublincore.org/documents/dcmi-terms/

[106] http://www.dublincore.org/

Dublin Core

The Dublin Core standard arose from a 1995 workshop held in Dublin, Ohio. Although there are now a larger number of metadata terms defined by the Dublin Core Metadata Initiative (DCMI), the Simple Dublin Core involves the 15 elements given below. Each is optional and repeatable, and may appear in any order the creator of the metadata wishes.

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Simple Dublin Core is widely used, simple and easily adopted. It is built into most repository software.

DCMI has, however, now developed an Abstract Model to support the development of **Application Profiles** and allow richer use of metadata. As such Simple Dublin Core is now regarded as a legacy use of the standard. These 15 elements have been reincarnated as part of the DCMI Metadata Terms. DCMI Metadata Terms is built around this element set but supports other terms which add further levels of detail to metadata; for example to distinguish between additional or translated titles, or define different types of dates or contributors. Many of the additional DCMI Metadata Terms refine the 15 simple elements.

Other Metadata Standards

Other metadata standards are available for fulfilling more complex purposes, for example for collecting preservation metadata. These standards include

MODS

MODS[107] (Metadata Object Description Schema) as an XML schema is 'intended to be able to carry selected data from existing **MARC**[108] 21 records, as well as to enable the creation of original resource description records'. This schema includes elements of MARC fields and uses language tags, rather than numeric ones.

PREMIS

PREMIS[109] is an XML schema which supports implementation of version 2.0 of the PREMIS Data Dictionary

IEEE LOM

Institute of Electrical and Electronics Engineers Standards Association[110] (**IEEE**)[111] 1484.12.1 - 2002 Standard for Learning Object Metadata *'is a data model, usually encoded in XML, used to describe a learning object and similar digital resources used to support learning.'* It is often used in the context of online learning management systems (LMS) but is also used by some repositories that hold learning materials.

Some metadata standards are designed to describe aggregations of objects.

METS

METS[112] (Metadata Encoding & Transmission Standard) is maintained by the Network Development and MARC Standards Office of the Library of Congress.

DIDL

The **Digital Item Declaration Language**[113] (DIDL) is a metadata markup standard developed for the accurate description of multimedia objects. It was developed by the Moving Picture Experts Group (MPEG) as an ISO standard.

IMS CP

IMS Global Learning consortium[114] - Content Packaging Specification *'provides the functionality to describe and package learning materials, such as an individual course or a collection of courses, into interoperable, distributable packages. Content Packaging addresses the description, structure, and location of online learning materials and the definition of some particular content types.'*

Application Profiles

An application profile defines a metadata schema for a repository that draws on one or more metadata standards. It can be defined as a declaration that specifies which metadata terms an organisation, information provider, or user

More information on Application Profiles is available on Ariadne: http://www.ariadne.ac.uk/issue25/app-profiles/v

[107] http://www.loc.gov/standards/mods/

[108] http://www.loc.gov/marc/

[109] http://www.loc.gov/standards/premis/

[110] http://standards.ieee.org/

[111] http://wiki.cetis.ac.uk/What_is_IEEE_LOM/IMS_LRM

[112] http://www.loc.gov/standards/mets/

[113] http://bit.ly/pNo3mZ

[114] http://www.imsglobal.org/content/packaging/index.html

community uses in its metadata schema. It may also make declarations about how the metadata terms should be used. An application profile selects terms from one or more existing standards and may opt to refine the use of any of those terms (for example by mandating the use of a particular controlled vocabulary). It may not add terms to an existing standard. To add new elements to an application profile which are not in a referencable standard (or other namespace) a repository manager must first define their own namespace and then reference that namespace in the application profile.

Several application profiles are being developed around the Dublin Core standard to meet particular shared needs within the repository community. They include a **scholarly works application profile**[115] (SWAP), an **e-thesis application profile**[116], and an **images application profile**[117]. The Scholarly Works Application Profile is based on Dublin Core and is intended to support the description of scientific or scholarly research texts. The e-thesis Application profile is based on Dublin Core and is intended to support e-theses produced in UK institutions. The Images Application Profile is based on Dublin Core and is intended to support describing images. There are a number of application profiles available of the IEEE LOM including the **UK LOM Core**[118]. This application profile was intended to support the sharing of learning materials in the UK.

If a community of practice shares an application profile it can promote richer interoperability and allow more refined search services but this may increase the metadata creation workload. The best advice is to keep a watching brief on how others are managing their repository metadata: regularly review good practice and implement improvements.

Subject Classification

The use of a defined subject classification schemes in institutional repositories is optional and an interesting debate has emerged as to the value of doing so. Some question spending time on classifying content within repositories when the full text of the items being included will be indexed? Some question assigning free-text keywords in metadata when the indexing mechanism will do the job automatically. Alternatively others may argue that the use of an official classification scheme will improve subject discovery of content, in particular offering better ways of browsing items within the repository. There is a stronger argument for investing time to subject classify media such as images, sound files and video as full text searching is not going to be useful for these formats. The choice of whether to use an official scheme lies with the institution itself and will largely depend on resources available to spend time inputting metadata and the level of mediation planned in the content ingest workflow. At the root of this are questions regarding whether or not academics have the time, or inclination, to use an official classification scheme, and, if it is a valuable use of the repository administrator's time to classify all incoming content.
In practice, there is a range of ways of indicating the subject matter of the content. The most common approaches are:

[115] http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile

[116] http://ethostoolkit.cranfield.ac.uk/tiki-index.php?page=UKETD_DC%3A+The+metadata+core+set+recommended+by+EThOS

[117] http://www.ukoln.ac.uk/repositories/digirep/index/Images_Application_Profile

[118] http://zope.cetis.ac.uk/profiles/uklomcore

- **Subject classification** - using a library-based scheme such as **Dewey Decimal**[119] or **Library of Congress Subject Headings**[120]. Within particular subjects there are detailed classification schemas such as **Medical Subject Headings**[121] (MeSH)
- **Course classification** - using a scheme such as **Joint Academic Coding System**[122] (JACS)
- **Departmental classification** - assigning repository items to department/research groups
- **Informal classification** - allowing uncontrolled vocabularies/'folksonomies' /user-generated tagging

These approaches can be used in combination, so that for example, a contributor tags their content with whichever terms they think are useful, and the content inherits the contributor's department as an additional classification. A repository officer might then assign a JACS term to it as well.

[119] http://www.oclc.org/dewey/

[120] http://www.loc.gov/catdir/cpso/lcco/

[121] http://www.nlm.nih.gov/pubs/factsheets/nlmclassif.html

[122] http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=158&Itemid=233

Workflows

A repository workflow is a breakdown of the administrative tasks involved. They allow the various activities involved in the running of the repository to be assigned to the individuals or groups who are best able to deal with them. The process of defining workflows is closely aligned with a stakeholder analysis of those involved with the repository: once the benefits and issues relevant to each stakeholder group have been identified, the next step is to ask exactly how users are going to interact with the repository software, as well as the specific tasks they are going to complete.

Workflows are *'a depiction of a sequence of operations, declared as work of a person, a group of persons'*

Workflows – Wikipedia:
http://en.wikipedia.org/wiki/Workflows

There are several types of workflow in a typical repository. These include workflows to manage user registration and administration; workflows to manage authorisation and permissions within the repository; and various administrative workflows to allow for maintenance and software updates. However, the most significant workflow focuses on the submissions process. This workflow is crucial as it will be used regularly by a wide variety of depositors.

It is through understanding the workflows of potential contributors to the repository that you can understand the best points at which deposit of content might happen, and how to make that deposit as easy as possible. For example, the process of an individual academic producing a doctoral thesis in history involves different activities to the process by which a team of researchers collaborate on a paper in experimental physics. The copyright, ownership of data, timescales and publisher involvement will vary, and consequently the potential workflows and interaction with the repository will vary.

UK Examples

NECTAR Workflow Document:
http://nectar.northampton.ac.uk/details.html

Oxford Research Archive Deposit Guides: http://bit.ly/rqFMpa -
Article, **Theses:** http://bit.ly/pzG3aq

International Examples

PARADISEC: http://bit.ly/rjWQzh

Further Resources

Further information about workflows and process review is available via our **Process Review infoKit:**
http://www.jiscinfonet.ac.uk/InfoKits/process-review

Submission Workflows

Submission (or 'deposit') workflows define the steps involved in adding content to the repository; gathering the necessary **metadata**, **permissions** and files associated with the content; and doing all the necessary checks on these elements before making the item available to the wider world.

There are several benefits to creating good submission workflows within the repository, namely they:

- **Streamline the deposit process** - Comprehensive submission workflows minimise effort and simultaneously ensure capture of all required information without duplication of effort or heroic measures
- **Encourage user deposits** - User-friendly submission workflows can encourage academics to deposit more items

- **Integrate quality assurance** - Building checking stages into workflows allows items or metadata to be double-checked for accuracy and consistency early in the life of the item
- **Add value** - Workflows can add value to a collection or process for example, by adding subject classification to an object, or by triggering other actions such as submission to publishers or other repositories
- **Facilitate administration** - Once content starts flowing into the repository sound workflows enable the repository administrator to manage new deposits, track objects through each stage, and address any problems that may arise

Typical tasks within a submission workflow can include:

- Acknowledging newly-submitted items (if this is not done automatically by your software)
- Checking the eligibility of depositors and/or the types of item being deposited
- Verifying, and if necessary querying copyright permissions. For instance, there may be a need to check which **version** of the item is being deposited i.e. preprint; author's peer-reviewed version; published version
- Validating metadata
- Approving the submissions i.e. making them publicly visible
- Releasing **embargoed** full-texts when the relevant period has expired - if your software does not do this automatically

Planning a submission workflow

A simple submission workflow has three basic elements: metadata; permissions; and file management. The following questions must be addressed to define an effective and comprehensive workflow:

Metadata input

- What metadata is going to be gathered from the authors?
- What metadata (if any) will be generated automatically?
- What metadata (if any) are administrators or other repository staff going to add to each record?
- What are the options for minimising free text fields?

Permissions/copyright and licence handling

- Who is responsible for checking the copyright of each submission?
- At what stage in the process is this check completed and how are the decisions recorded in the metadata?
- When will the depositor sign a deposit agreement or license?
- How will embargoes be dealt with?

File management

- What files will be requested from authors?
- What **formats** will be requested?

- How will associated files be identified and stored in the repository?
- How will different versions of deposits be managed?

Spend some time thinking about and discussing submission workflows at an early stage in repository development, at the same time as considering **repository policies**. Workflows will impact on the configuration and customisation of your repository software in terms of the permissions set on certain stages of the submission workflow process, and included elements within item submission forms. Once draft workflows are in place it is recommended to test them with a group of your users and to remain flexible to allow your workflows to adapt over time. Several different workflows may be required for different categories of materials - for example if your institutional repository has research materials and learning and teaching materials, or if some materials are open and some closed access.

Handling Embargoes

Some of the items to be submitted to the repository may be subject to an embargo which states that the material within the item may not be publicly distributed for a certain period of time. In the case of published items, embargoes may be put in place by the publisher to protect their commercial interest. In the case of theses or dissertations an embargo may be set by the institution as the work contains sensitive material; or by the student as he or she is looking to publish some or all of the work in the near future. It is important to work out what position to take on embargoes and to have strategies to deal with them in place at an early stage; the imposing of an embargo can come at a crucial stage of the academic community's engagement with the repository and there is a need to have a simple and clear path to take when this scenario presents itself.

Storing embargoed full-texts and alerting systems

Where an academic is keen to deposit his or her text (although it is found to be subject to a publisher's embargo); it may be preferable to create an item record within the repository and accept the full-text at the first offering, rather than going back to the academic at the end of the embargo period. This is consistent with advocacy efforts to show deposit is important, and it captures the academic's involvement at a relevant point in their workflow. Another example would be in an institution where an e-submission and repository distribution mandate is in place for theses and dissertations. An electronic copy of the work will be sent to repository staff despite any embargoes it may be subject to. In such situations a system needs to be put in place for repository staff to be able to store the full-text item for the embargo period; and then to be alerted as to when the embargo has finished and the item can be included within the repository.

The mechanism for how repository staff will store the embargoed full-text and how they will be alerted to the end of an embargo period will depend on the individual software and local set-up. Some repository software has embargo alerting systems in place allowing repository staff to add an embargo end date to an item on creation; sending an alert to the repository administrator once the embargo period is complete. Up until that point the item was stored in the repository but under restricted access. Once the repository administrator has received the alert the access permissions on the item can be altered. Other software systems do not offer an alerting system but do allow the full-text of an item to be subject to restricted access, only displaying the item metadata to the public. If a reminder is then set on a separate electronic diary system, for example, the repository administrator can then remove the access restrictions after the appropriate date. Some repository administrators may choose not to use the repository to store the full-text at all and other methods and systems can be used depending on local resources.

Alternative access to a full-text item

If a metadata only record for an embargoed item is included within a repository, it may be appropriate to include the item's Digital Object Identifier (DOI) or a URL to the publisher's website where a copy of the full-text is available. This is sometimes a condition a publisher sets to allow an item to be included within a repository and it will enable those users who have subscription access to the journal to view the full-text during the embargo period.

Another strategy that can be adopted within some repository software as a holding measure for embargoed item full-texts is to make use of an item request button. This is a button placed in the metadata record when an embargo is in place which allows the user to contact the author of that paper and request a one-off copy via e-mail or in hard copy. The sharing of research in this manner does not break copyright law (or at least offers a defence of 'custom and practice') and it is a well established practice in most academic communities, pre-dating the arrival of online distribution systems. This does mean that a user searching for an embargoed item is initially denied access to the full-text but does have a chance, at least, of obtaining the full-text subject to a delay by approaching the author direct.

For more information regarding individual funding council mandates please see **SHERPA-JULIET:** http://www.sherpa.ac.uk/juliet/

Funder mandates vs embargoes

Increasingly, funding councils are requiring that the outcomes of the research that they fund are made publicly available at the earliest possible date. Conditions of funding state that one of the methods that can be utilised to do this is to deposit the research within an open access archive such as a repository. However, instances may occur where an academic's research council mandate contradicts the embargo put in place by the publisher. Although publishers may be contacted on an individual basis to request the inclusion of papers, it is the publisher's embargo which is usually upheld.

Further information

Version Identification Framework:
http://www.lse.ac.uk/library/vif/

VERSIONS Project:
http://www.lse.ac.uk/versions (including user study, November 2007)

Versions

As repositories have grown in size and scope, the problems caused by versioning have become more prominent. It is now an accepted problem; only 5% of academics and 6.5% of information professionals find it easy to identify versions of digital objects within institutional repositories. Across multiple repositories this becomes 1.8% and 1.1% respectively. Version identification is not about identifying the 'best' or 'right' version of an object. The Versions Identification Framework (VIF) aims to make the version status of a research object and the relationships between linked objects clear to end-users so they can identify which is the most appropriate version for their purpose.

Draft Version

Early version circulated as work in progress

Submitted Version

The version that has been submitted to a journal for peer review

Accepted Version

Author-created version incorporating referee comments and is accepted for publication

Published Version

The publisher-created published version

Updated Version

Updated since publication

Incorporating different versions of learning and teaching materials proves equally challenging, presenting a barrier to deposit: many academics are concerned that people may use earlier versions of their content that may not still be accurate (this is more apparent for those subject disciplines where course content regularly changes and particularly if old information could be dangerous - as in medical disciplines). Version policies which recognise these fears and provide clear guidelines around deposit, updating and using content are essential to gain trust and ongoing deposit. A learning and teaching repository could chose to hold only the most recent version, or could hold previous versions only for legacy purposes and not make these discoverable. This relies on good communication and clear workflows to ensure that new versions are deposited appropriately.

Versioning problems are minimised if some or all of 5 key pieces of information that exist at the time of creation of the object are captured, defined and made sufficiently visible to the user, to a deposit mediator or made machine-readable:

- Defined dates
- Identifiers
- Version numbering
- Version labels/taxonomies
- Text description

It is important to include versioning information both in metadata and within an object itself and one of the following solutions should be used systematically to store version information within a repository:

- The filename
- A watermark
- A coversheet - link to coversheet page
- An ID tag or property field

Advising academic authors about versions

Academic researchers typically produce many revisions and versions of a research output, up to 60 in some cases, and it increases when working with co-authors. Authors vary in how many of these versions they keep: either all, milestone only, or just the latest. They tend to want to limit the number of versions disseminated.

A key message for repository managers to convey is that, because of standard agreements between authors and publishers, keeping only the Published Version is likely to limit the author's ability to make their work accessible in future. As a minimum, keep milestone versions like an author's Accepted Version. Half of authors are not satisfied with the way they manage their own personal collections of digital objects. An institutional repository provides academic staff with a managed environment to deposit their milestone versions, where they can be easily located

and retrieved. 81% of authors surveyed stated they would deposit their final Accepted Versions of journal articles in an institutional repository 'if invited to do so'. Adding version information to authors' papers, with clear links to published versions, helps allay concerns about the quality of accepted versions and about potential loss of downloads for published versions. Institutional repository managers can help inform academic authors about publishers' access policies. Researchers remain uncertain and unaware of possible negotiable alternatives. **SHERPA RoMEO:**[123] and **SPARC**[124] are two excellent resources for providing detailed advice to authors on 'permitted' versions.

The **VERSIONS toolkit**[125] offers 5 top hints for authors, derived from feedback by active researchers:

- Plan how personal versions of files will be stored and named
- Keep separate, permanent author-created Submitted Versions and final author-created, Accepted Versions of research publications
- Add the completion date to the first page of any versions, especially milestone versions
- Consider carefully how to disseminate work before signing any agreements with publishers
- Keep a copy of signed agreements. Deposit your work in an open access repository and guide readers to latest and published versions

[123] http://www.sherpa.ac.uk/romeo/

[124] http://www.arl.org/sparc/

[125] http://www2.lse.ac.uk/library/versions/

Disclaimer

We aim to provide accurate and current information on this website. However, we accept no liability for errors or omissions, or for loss or damage arising from using this information. The statements made and views expressed in publications are those of the authors and do not represent in any way the views of the Service.

The JISC infoNet Service offers general guidance only on issues relevant to the planning and implementation of information systems. Such guidance does not constitute definitive or legal advice and should not be regarded as a substitute therefor. The JISC infoNet Service does not accept any liability for any loss suffered by persons who consult the Service whether or not such loss is suffered directly or indirectly as a result of reliance placed on guidance given by the Service.

The reader is reminded that changes may have taken place since issue, particularly in rapidly changing areas such as internet addressing, and consequently URLs and email addresses should be used with caution. We are not responsible for the content of other websites linked to this site.