Documentation for institutional repositories

Documentation for Institutional Repositories

Steve Probets and Celia Jenkins
Department of Information Science, Loughborough University, Loughborough, Leics, LE11 3TU

Email: s.g.probets@lboro.ac.uk; c.d.jenkins@lboro.ac.uk

Abstract

In order to identify best practice, the documentation of seven academic institutional repositories (IRs) was compared and contrasted. This was followed by semi-structured interviews with six practitioners experienced in the set-up, management and maintenance of IRs, including representatives of three JISC FAIR projects.

The aim was to identify the requirements of policy documentation provided by IRs. Although many issues were found to be handled differently depending on what IR software was used, or the stage of development of the IR, several common factors emerged. These included the importance of developing the documentation in collaboration with individual academics, departments and senior management whose views and needs are central to the success of the IR. Additional findings were that policies should be formulated only when the purpose and aims of the IR have been clearly defined and that the IR documentation itself should be concise and easy to understand, with the rights and responsibilities of stakeholders clearly presented.

1 Introduction

Open access to research literature has recently been receiving considerable interest world-wide [1, 2]. There are two basic routes for achieving open access to this literature. The first is by publishing in a journal which provides open access to all (the so-called 'gold route'). The second is to publish in a conventional journal but to mount a copy of the article on an open access repository (the 'green route'). These digital archives can either be subject based, as is the case with the physics arXiv [3] repository, or, in a growing number of cases, they can be hosted at specific research institutions such as Universities. The self-archiving of research articles in institutional repositories (IRs) has been the focus of much work in the UK, with the Joint Information Systems Committees' (JISC) *Focus on Access to Institutional Resources* (FAIR) programme [4] funding many projects to develop and promote the technologies and procedures to enable self archiving to become commonplace. The number of IRs seems likely to expand given the interest shown by many institutions, as well as the recommendation made in a recent report from the House of Commons Select Committee on Science and Technology that research councils should mandate that research output be open access and that HE institutions should establish IRs to house their published output [1]. Although the government's response to this report was to take a 'hands-off' approach, there were indications that they "recognise the potential benefits of Institutional Repositories and see them as a significant development worthy of encouragement." [5]. There are many IRs coming online and the Directory of Open Access Repositories [6] will shortly list all open access repositories.

IRs are maintained by individual institutions, and so, in order to support their interoperability, many are built using technologies such as the *Open Archive Initiative Protocol for Metadata Harvesting* (OAI-PMH) [7]. This enables metadata describing resources maintained in the repositories to be

harvested and brought together into centralised services to facilitate retrieval and access to these resources. Technical solutions for creating IRs and making the metadata available have also been developed, with ePrints, DSpace and CDSWare all widely used. In addition, various models for sharing metadata descriptions of the contents of repositories and for providing usable services to interface with these archives have been investigated [8]. However, although the ability to create services is important, the success of IRs ultimately relies on the amount of material that they contain. One of the most important issues is ensuring that they are as fully populated as possible [9,10]. This will only occur through establishing IRs - it is through the setting up of initiatives and projects that problems can be identified, solutions found, and lessons learnt.

Many IRs have been set up worldwide, some of these are pilot projects and some are now becoming full IR services, and it is at this stage that they need to start formulating policy documentation, or at least formalising their policy decisions. It is important that the necessary policies and procedures are in place to ensure that the responsibilities for content, management and maintenance of the repositories are accurately specified. Clarification and presentation of the roles and responsibilities of the various parties may help with promotion of a repository and may also help to encourage the population of the IR. Therefore, it seems appropriate at this time to ensure that policy documents are formulated correctly.

This paper describes a short research project, which aimed to determine best practice in developing documentation, policies and procedures for IRs. In particular it looks at the approach that should be taken in forming policy documentation and the general issues that should be covered in the documentation.

1.1 Methods adopted

The main objectives of this research were:

- To discover the factors affecting the content of IR policies;
- To establish what issues are covered and how these issues are handled in existing IR policy documentation;
- To discover practitioners' views and experiences concerning IR policy decisions and policy formulation; and
- To make recommendations in order to help develop and maintain IR policy documentation.

In order to fulfil these objectives, two complementary approaches were taken. Initially an analysis of existing IR policy and documentation was undertaken, this was followed by semi-structured interviews with experts and practitioners involved in developing and advocating IRs.

1.1.1 Document research

This involved analysing the publicly available documentation provided by seven existing IRs. The documentation was downloaded and analysed during Summer 2004. The IRs were:

- The California Digital Library eScholarship Repository (CDL) [11]. (Using ePrints software);
- MIT DSpace Repository [12]. (Using DSpace software);
- Queensland University of Technology E-print Repository for Research Output (QUT) [13]. (Using ePrints software);
- Hong Kong University of Science and Technology Institutional Repository (HKUST) [14]. (Using DSpace software);

- The University of Melbourne – Eprint Repository (UMER) [15]. (Using ePrints software);
- The Open University ePrint Pilot Service (OU) [16]. (Using ePrints software); and
- The University of Southampton Electronics and Computer Science (ECS) ePrints Service [17]. (Using ePrints software).

A mixture of IR services, IR pilot projects, and one departmental repository were chosen to give a broad picture of how various issues are presented in the documentation. In some cases the available documentation was not necessarily a formal policy document but consisted of one or more documents describing the roles and purposes of the IR. When choosing IRs to consider it was necessary to consider repositories based on the DSpace software developed by MIT, as well as those using ePrints software, developed by the University of Southampton. Although both are freely available open source applications which allow IRs to be set up and maintained, DSpace was created in order to capture, describe, and distribute digital works, as well as preserve them [18], whereas ePrints was created "more specifically for institutional or disciplinary repositories of papers" [19].

The purpose of the document research was not to determine which policies are more comprehensive, but to identify issues of importance to future developers of policy documentation. In order to analyse the documentation, specific criteria were developed through an analysis of literature, meetings and workshops that discussed the practical implications involved in setting-up and maintaining IRs and how such issues should be addressed. Twelve main criteria were established and sub-categories were developed for each of these criteria. The main criteria included i) whether an overview of the documentation was available; ii) content and collection policy; iii) details of management and administration; iv) start up procedures; v) submission procedures; vi) metadata standards and quality; vii) legal issues; viii) user policy; ix) privacy policy; x) funding; xi) preservation policy; xii) backup and recovery. Space restrictions prevent reporting the full analysis, however the main findings in each of these areas are presented in Section 2 below.

1.1.2 Semi-structured interviews

Through the examination of the IR policies further questions came to light. This was especially the case when some, but not all, of the institutions covered a particular issue, or did not cover it in the same way. It was therefore felt that the best way to answer these questions was to discover the views and experiences of actual practitioners in the field, through conducting a series of semi-structured interviews.

Interviews were completed with four representatives from three projects funded under JISC's FAIR programme. These projects were chosen as they have experience in setting up IR services, and are now looking at creating and formalising their documentation. Three more interviewees were chosen from institutions, which have either set up an IR pilot project, or are looking to set one up. These interviewees were predominantly from a librarian background and had responsibilities in areas such as the management of collections, electronic resources and support. All interviewees had a thorough knowledge of IRs and were involved in the development of an IR at their institution.

2. Content of the Documentation/Policies

The policy documentation was analysed and is discussed below under section headings that correspond to the evaluation criteria. There were differences between the content and organisation of the different policy documentation, and the interviewees occasionally differed in their opinions towards how the documentation should be maintained. Some interviewees felt that the policy documentation should be a series of policy documents, whereas others saw it as being one

overarching policy document, with procedural documents linked to it. However, the organisation of the documentation is perhaps not of such great importance, as long as the policy decisions are clearly identifiable and easy to locate.

In the discussion below the opinions of the interviewees on the issues under discussion are also considered. The interviewees all indicated the need for the policy documentation itself to be short and understandable, and for the academic to be central to the process.

2.1 Overview of the policy/documentation

An overview of the IR policy acts as an introduction to the area, it often informs readers of the issues involved and may also encourage authors to deposit their work. All of the policy documentation analysed provided definitions of terms and explanations of the aims of the IR as a way of conveying its benefits. Some explicitly listed the benefits of the IR to academics and the institution itself, though this tended to be found more in the documentation of the pilot projects. It should be noted that the two UK-based repositories stated the benefit of linking IRs to the UK Research Assessment Exercise (RAE), with ECS declaring that "with all our research output accessible online we will be able to respond to the RAE and other administrative initiatives with minimal input and effort from individual staff" [17]. It has been proposed that in future it would be possible for submissions to the RAE to require deposition of research articles into an IR. Raising awareness of this possibility, which is discussed in more detail by Day [20], may encourage an acceptance of the principles of self-archiving by academics.

Two policies (QUT and ECS) also explained the function of the IR by providing a glossary of terms, including 'self-archive', 'preprint' and 'postprint'. In the early stages of IRs, when cooperation from academics is needed, it may be necessary to define unfamiliar terms and ideas to academics. Another way of making it clear how the IR will operate and how the stakeholders will interact with it, was to state the roles and responsibilities of the stakeholders. This was particularly evident in the MIT documentation.

The interviewees felt that explanations and definitions of IRs should be found in supporting documentation. Nevertheless, the function of the IR should also be mentioned in any collections policy, as it has a direct impact on collections guidelines. It is, therefore, necessary for an IR to have a clear idea of its purpose and aims, and for the documentation and policies to help to communicate these ideas to individual academics and departments.

2.2 Content and collections policy

A content and collections policy is likely to be a central part of an IR policy as it defines what content is to be accepted and how this is to be organised. An IR could include different types of material such as pre-prints, post-prints (peer-reviewed research articles), book chapters, working papers, and theses. All these types of document would need to be integrated into the IR. Many of the policies (including QUT, HKUST, UMER and OU) specified what types of materials are to be accepted, and three stated what document formats are actually allowed. The choice of document format, which may include postscript, PDF, ASCII, and HTML, may be more important than at first thought as it has a direct effect on submission procedures and preservation issues. (See sections *3.1 Submission/deposit procedures* and *2.11 Preservation policy*.)

Listing both type and format of content would make the submission process clearer for academics. In the current climate where academics are not familiar with IRs, anything that simplifies the process and makes it clearer for academics to submit should lead to more fully populated repositories. Only Southampton's ECS departmental repository explicitly stated that it is mandatory to submit work into the IR, but it still encouraged deposition by informing academics of IR issues and offering help and support. QUT has also made deposition mandatory, yet this is not explicitly stated in its policy, preferring to downplay this aspect and concentrate firstly on communicating the explanations and benefits of the IR to academics [21].

All of the policies stated that work must be research-based and from that institution. However, two policies (CDL and UMER) explicitly mentioned the possibility of including work of authors who are not members of the IR institution. CDL stated that "content does not have to be authored by UC faculty to be included in the eScholarship Repository" [11], whereas UMER stated that "contributors may include outsiders if they are co-authoring with University of Melbourne authors or are affiliated closely with the University" [15]. This issue needs to be addressed in the policy documentation as much research output is written collaboratively between researchers in different institutions.

Two policies (MIT and CDL) specifically mentioned the possibility of depositing a series of materials. MIT encouraged any series to be deposited, "so that DSpace can offer as full a set as possible" [12]. However, it is not immediately clear what is meant by 'series'. A series could include work from an author not from the IR institution. Therefore, a definition or explicit statement may need to be given in the documentation on whether material not authored by members of the IR institution is allowed as part of a series, so as to avoid confusion. CDL was more specific, stating that it offered a mechanism by which groups from the institution may form a series from a particular type of appropriate content, such as working papers, peer-reviewed content, and, most recently, seminar material.

Four policies explicitly stated how their collections are organised, with MIT, HKUST and CDL all organising content under departments or units, and QUT using subject categories. MIT focused on the term 'DSpace Community', which is "an administrative unit … that produces research, has a defined leader, has long-term stability, and can assume responsibility for setting Community policies" [12], whereas CDL talked of "contributor units" which can be "any University of California research unit, institute, center, or department" [11]. This implies that there is flexibility in the way that the content of IRs can be organised.

Each 'community' may have the ability to make their own decisions on issues such as the type of material to be deposited, and the submission procedures. This was particularly evident in the MIT policy, which also considered possible problems, stating what happens if a DSpace Community or collection ceases to exist. Organising content in this way (under 'communities') has the advantage of providing departmental flexibility. (See *How to achieve departmental, or institutional, flexibility* in section 3.1)

The interviewees felt that a collection policy was an important part of the IR policy documentation, as not only does it give guidelines to submitters as to the kinds of things that they can expect to submit, but also helps the repository's administrators know what content can or should be maintained. It was felt that the collections policy should include the policy on withdrawal, as well as state any quality control procedures that are in place.

2.2.1 Withdrawal of items

Another issue that is of concern to all involved was whether or not to allow withdrawal of items. Although one definition of IRs is that they should be cumulative and perpetual [22] this view may not be shared by all IR managers. But, even if it is, there are some circumstances which would warrant the withdrawal of material, especially if it involves legal dispute. A policy should clearly set out under what circumstances items may be withdrawn, and how this may be carried out, for example if withdrawal means complete deletion of a metadata record as well as removal of the offending resource. The majority of policies covered the withdrawal of items, but with the removal of items being generally discouraged. A majority of the policies stated that there will always be a record of the resource, even if it is removed. This is to ensure that it is possible for a user of the IR to know what has happened to that record. Two of the policies (MIT and UMER) explained the circumstances under which items may be withdrawn. The situations in which withdrawal is accepted were given by UMER as "if the journal in which the paper is formally published requires it; if the paper proves scurrilous, plagiarizes, is libellous or breaches copyright; and if the academic author decides to remove it" [15]. It is interesting to note that although some of the other policies accept an author's request to withdraw an item, they do not give an idea of any of the circumstances which would warrant this.

Four of the interviewees explicitly stated that policy documentation should formally specify what happens if material needs to be withdrawn, in what circumstances withdrawal is allowed, and who has the responsibility to make these decisions. The need to still have a record of the withdrawn item was also recognised by the interviewees. This will ensure that once an item is withdrawn its bibliographic details, and the reason for its withdrawal, are still available in the long-term.

2.2.2 Quality control

With the exception of UMER, which stated that "most working papers receive informal review from the Dean or Head of Department to protect the University/department from embarrassment" [15], there was no obvious mention of quality control procedures in the documentation. Research-based work does not mean that it is of a high quality, and this must be kept in mind, especially where pre-prints are concerned.

The interviewees also raised this issue, especially as academics may have concerns with having their high-quality post-print works maintained in repositories containing pre-prints, which may not be of such a high quality. In general, post-prints have already been through quality control procedures as they have been peer-reviewed. One approach has been to segregate pre- and post-prints into separate repositories, as in Glasgow. Another approach is to divide the repository into two sections, that of pre-print and post-print. A further possibility mentioned by two interviewees is that departments employ a kite mark, stamp system or metadata entry to show the level of quality of individual works in the IR. As quality control may be of concern to academics, mechanisms for clear labelling of pre- and post-prints are likely to be important [30], and so any procedures that are in place (such as the indication of pre- or post-print status of submissions) should be clearly stated in the IR policy documentation.

2.3 Management/administration of the IR

Academics need to know who is responsible for running the IR and how to get in contact with them in case they need help or advice. This also makes the IR a more approachable service, and one in which academics are encouraged to collaborate.

Four of the policies (CDL, MIT, QUT and HKUST) clearly stated who was responsible for managing the IR - usually the library. However, only the QUT policy explicitly stated that the IR and the associated policy would be reviewed, with the next date of review given as 01/01/2006. Reviewing the IR is an important process, as it ensures that the IR is continuously meeting the needs of the academics and the institution. IRs are still a relatively new approach and the documentation and policies are often in the early stages of development. The interviewees felt that periodic reviews of both the IR and its associated policy documentation were a good idea, as this would ensure that they evolve together, and would encourage the adoption of best practice. It was felt that including details of the review process in the policy documentation would make this a formal decision and would communicate commitment to IR policy development. It would also hopefully ensure that the review process is not forgotten or overlooked.

It is, perhaps, surprising that there was not more coverage in any documentation as to how the IR integrates with existing electronic resources offered by the institution. The IR should work with the other ways that departments already report their research, whether it is through academics' home pages, departmental publications database or existing subject-based archives.

2.4 Start-up procedures

Start-up procedures give details on what to do to become a part of the IR, clearly setting out how to prepare to join, and what is offered in the way of help to academics and their departments.

Two policies (MIT and CDL) covered start-up procedures comprehensively, having specific documents dealing with, among other things, training and customisation (e.g. for a specific department's output) of some of the IR pages. Other policies were less complete in this area, with two policies not mentioning these issues at all. It may be that these institutions do not enable configuration at a departmental level or it may be that they provide informal start-up procedures involving visiting departments and individual academics personally. Whatever approach is taken, the issue of educating departments on how to use or configure the IR may need to be considered. It should be noted however, that information on the physical formalities (such as how individual academics can register or submit work) is covered by several of the policies in their supporting documentation.

2.5 Submission/deposit procedures

Submission/deposit procedures can be on two levels. The first is a step-by-step practical guide on how to deposit. The other is a description of submission workflows, detailing such things as who is to deposit content and who reviews it.

A decision has to be made as to the level of responsibility to give academics in the submission process. Only one policy (QUT) provided academics with clear steps on how to convert their papers into a suitable electronic format for submission. However, all policies gave the academic a choice of how to submit their work. Either the library offered a mediated service, or academics deposited their work themselves. This is important as it accommodates both 'early adopters', i.e. those who already have experience of self-archiving in a repository, and those who need to be encouraged to deposit. Three policies (CDL, QUT and OU) gave further details of where to get further help and advice on submission if required by the academic. IR practitioners have found that offering a mediated service may be preferable in the short term, as many academics are unsure of the submission process [23]. Be that as it may, consideration has to be given as to whether this will be sustainable in the longer term due to the inevitable time and money constraints involved if self-archiving of every pre-print,

post-print or working paper is to become part of normal working practice. Even if academics are to have more responsibility there should still be some documented mechanism in place for the library to provide some level of assistance, especially at the start.

The interviewees agreed that, on a practical level, submission procedure guidelines are needed so as to provide step-by-step deposit instructions to authors. Documentation describing the different submission workflows was also deemed very useful, especially when, as is often the case, there is a mixed economy model, i.e. deposition of content by author as well as mediated and/or departmental self-archiving. Different workflows may impact on other IR issues, for example if a third party deposits on behalf of the author then additional clauses may be necessary in any warranties provided by the depositor.

2.6 Metadata standards and quality

Good quality metadata is extremely important to support a level of service, which enables content to be found quickly and easily, particularly if metadata from IRs is to be harvested into centralised services. It may, therefore, be necessary to have a policy, which states what metadata is required, what level of quality is necessary and who is to author the metadata. It is, therefore, quite surprising that metadata was so poorly represented, both in the actual policy documents and supporting documentation. This may be due to the fact that guidance from service providers as to their requirements may not yet be available or it may be simply that in many IRs the academics themselves are **not** expected to provide metadata (or are only expected to enter simple fields such as author, title and key words). Even though metadata entry may be the role of the library, only one policy (UMER) has an actual section on metadata, which stated that it uses simple Dublin Core (DC), but that thesauri (such as the ERIC thesaurus) may also be used. It may be useful for policies to indicate whether specific classification schemes will be used within the repository.

Of course, one reason for the lack of discussion of metadata in the documentation may be that, in the early stages of development, repository managers may have different priorities and that configuring the service and populating the repository are deemed of greater importance. However three of the interviewees indicated that metadata of a sufficient quality is essential for any IR, not least because it will enable service providers to harvest the metadata and build appropriate value-added services. Metadata should, therefore, be considered when setting up an IR. However, only one interviewee felt that there was a need for a specific policy on metadata. In the short term, in order to encourage population of IRs, it may be better to either get academics to enter simple metadata, such as author name, title and keywords, or have a mediated service where the administrator adds the metadata. This would ensure a basic level of metadata without interfering with the submission process. All repositories that conform to the Open Archives Initiative must contain Dublin Core metadata records, though IRs themselves could consider whether to use specific classification schemes or controlled vocabularies in the assignment of subject metadata provided within the Dublin Core scheme. If these decisions are taken at the individual repository level (an alternative would be for service providers to provide this added level of metadata) then these decisions could be explicitly described in the IR documentation. As the generation of metadata is likely to be linked with the submission process, it may therefore be covered within the submission procedures documentation, with a formal metadata policy being drawn up later, in the full IR service stage.

2.7 Legal Issues

The area of intellectual property rights (IPR) is a complex area and one that is of great concern to academics and institutions alike. Therefore, all legal issues need to be clearly explained to the

academic, and their position explicitly stated, especially as they must agree to the terms given in any IR deposit agreement.

Only two policies explicitly discussed who actually owns the copyright to IR material, with both CDL and HKUST stating that the authors of the work retain copyright (assuming they own the copyright in the first place). This is a key point to make, as there could be confusion as to whether the academic is assigning the copyright to the IR on deposition.

None of the policies required authors to own the copyright to their work. It was sufficient for the author just to have the right to self-archive, or to have an agreement with a publisher who permits self-archiving. However, only two policies made it clear that they actually ask the author to warrant that they have the intellectual property to the work. A majority of the policies required authors of the content to grant the IR institution non-exclusive rights to distribute their work. This means that the author may still retain copyright, and may be able to publish the work elsewhere. However, only two policies (CDL and MIT) explicitly considered long-term accessibility by asking for the right to make a copy of the work for backup purposes, and to translate the work into other formats if necessary for preservation.

Certain IR policy issues, such as content guidelines and submission procedures, can have a direct impact on legal matters. This is of importance if the submission workflow is to allow deposition 'by proxy'. An example of this would be where a person is depositing on behalf of the author. In cases such as this, consideration needs to be given as to how appropriate warranties can be obtained and how decisions such as this may affect other documentation and policy decisions.

The IPR of metadata was also poorly covered, although QUT in its 'ePrints Deposit Agreement' specifically covered the copyright issues surrounding both submitted works and the associated bibliographic metadata. Future problems may occur if it is not clear who owns the copyright to IR metadata. Recent guidelines are now available which explain how the IPR of metadata can be specified within the OAI-PMH [24], and it may be worth institutions considering these issues and familiarising themselves with these guidelines.

The extent to which academics will have responsibility for legal matters also affects the documentation. One example is whether they need to know the copyright status of their work at the time of deposition. Three of the policies (CDL, MIT and QUT) all stated explicitly that it is the academic who is responsible for ensuring compliance with publisher copyright agreements. This may initially seem to be a sensible approach that would save the IR administrator time and effort, however the legal implications of this should be considered. The issue of vicarious liability arises here. If proven that putting the onus on an employee of an institution to accurately and precisely identify the copyright situation is an unreasonably complex task, then it is likely that the institution could be liable in any legal action. Following on from this, a decision must be made on how the IR will help academics in legal matters. In consideration of how complex legal issues are, some of the policies (including QUT, HKUST, OU and ECS) offered practical suggestions on how authors can retain copyright. It may be that due to the sensitive area of copyright, some IRs are reluctant to raise some of these issues. However, IRs could make use of other work in this area and, indeed, six policies provided links to other resources, such as the Rights MEtadata for Open archiving (RoMEO) Project [25] for further help and advice on copyright matters. The RoMEO database (now supported by the SHERPA project) describes publisher's policies on self archiving and will shortly be extended to provide a simpler interface for obtaining the policy of individual journals towards self archiving [26]. Making authors aware of any institutional resources or support to help authors discover the copyright status of articles should also be specified in documentation.

All interviewees indicated that IRs should have both copyright guidance and a deposit agreement/license for academics to sign. In the deposit license itself the IR institution would require the author to grant to the IR a non-exclusive right to distribute copies and to make copies for backup and preservation purposes. There was a feeling that the rights and responsibilities of the IR should also be stated so that depositors know what to expect from an IR and how their submission will be managed.

Some interviewees indicated that it should be explicitly stated as to who owns the copyright to the work in the IR, i.e. whether it is the University or academic. This would make the legal position of academics clear to everyone. However, this situation is complicated by the copyright transfer agreements that authors may have already signed with publishers. If authors have already transferred the copyright to a publisher, the situation for post-prints is far from clear. As legal issues are such a concern to academics and institutions alike it is imperative that all IR policy decisions are examined to see whether they impact on IPR and copyright, and that legal issues are thoroughly checked and clearly explained in the policy documentation.

2.8 User policy

A user policy states the ways people may use the information they find in the IR and whether there is any access control to content. This may be important in the beginning stages of the IR "when open access is premature or otherwise not desirable" [27]. None of the seven institutions had such a policy. As a result it was not clear what users can actually do with the research output they find on the IR. Although the concept of OA encompasses no barriers to access, authors or institutions may wish to restrict the use to which the material can be put - it is therefore important to state explicitly what people can actually do with the content of the IR. The RoMEO project considered what authors expect to be able to do with research material made available on IRs [28, 29].

The MIT policy explicitly mentioned access control, stating that "a DSpace community retains the right to limit access to content at the item level either to MIT only or to specific individuals or groups" [12]. Controlling access may seem to be opposed to the basic OA principle behind the development of IRs, yet there may be certain circumstances when control is necessary, such as wanting only members of the IR institution to access learning materials.

Three of the interviewees said that user policies should be covered in the documentation. One interviewee in particular realised the importance of having a formal user policy as they had been asked by an end-user of an IR what they were allowed to do with the paper once they had found it on the IR. Explicit policies stating the way content can be downloaded and re-used was, therefore, felt to be important.

2.9 Privacy policy

There is also the question of privacy concerning any personal details given by depositors as part of the submissions process or by users of the IR service who may want to register for alerting services. Three of the analysed policies had associated privacy policies in place for users of the IR. These were directly linked from the IR policy. Two of these provided a direct link to their university-wide privacy policy, whereas MIT had an IR-specific one. It is not clear whether there needs to be a privacy policy specifically for the IR, but it may make the issues clearer as it concentrates on a particular area.

Many of the interviewees had not considered a need for a privacy policy, but none saw it as a particular requirement. Although agreeing that a policy may be needed if the IR were to provide an alerting service, it was felt that a university-wide privacy policy would to be adequate enough for the purposes of an IR.

2.10 Funding

How the IR is sustained financially is a possible issue to consider, as it is important to look at the long-term and at how the IR will be maintained for the future. Perhaps not surprisingly the funding model of the IR was not covered in any detail by the policies, and it did not seem to be a requirement to include this. However, one policy (MIT) did state that it is possible for MIT Libraries to charge for IR services, and another (CDL) that its business model "may change over time" and that it will "work cooperatively with all University units to determine an equitable long-term funding model" [11]. This statement effectively outlined the plan to make the IR sustainable over the long term, which is perhaps a major concern of those considering implementing an IR.

The interviewees also felt that details on how the IR will be funded were not important enough to appear in the IR policy documentation. As one of the interviewees said, "Academics just want to know that the IR is sustainable in the long-term. It certainly does not need to be covered in any great length."

2.11 Preservation policy

This is another possible major issue of concern to academics and IR managers. IRs must make sure that the research output will be accessible in the long-term. Making sure research content is preserved is important, though some do not see preservation as a priority of IRs and feel that population of the IR is of greater importance. Arguments for and against formal preservation of IR content have been proposed by Pinfield and James [30] and can also be found in a recent report published by JISC [31], with the arguments against preservation primarily being that the content of self-archived post-prints is likely to have been preserved through the preservation of the journal in which the work appears. However preservation issues are still a major concern for many and one of the IRs (MIT) provides a very comprehensive preservation policy. This is the only institution which had a separate document covering this. It included a table of file formats and the level of support given to each one, i.e. whether the format will be preserved, recognized but not guaranteed to be preserved, or not recognized at all. Another IR (CDL) covered some of the preservation issues, though in less detail. The method of preservation was also mentioned in some of the documentation, with CDL stating that content will be preserved through migration, and MIT stating they will use whatever technique is appropriate, such as migration and emulation. The other policies did not cover preservation to any great depth. Given the arguments against preserving the content of IRs, at this stage it could be argued that population of IRs is of higher importance.

Despite this, all interviewees felt that preservation is an important aspect of the IR, not least because academics want to know that their work will be available in the long-term. However the interviewees did not see preservation as an issue, which needs to be addressed in the pilot project stage. It was mentioned that the choice of document formats allowed must also take into account preservation issues, as some formats may be easier to preserve than others.

Perhaps preservation should not be regarded as simply an IR issue but one of the larger information society, with the prospect of a university-wide preservation policy being suggested by two interviewees.

2.12 Backup and recovery

Even if a long-term preservation policy is not present, it is important to ensure that content is permanently accessible in the short term. There must be a backup and recovery mechanism in place so as to make sure IR content is not lost in the event of a system breakdown. The two policies of MIT and HKUST specifically stated that a backup and recovery mechanism was in place, with MIT also requesting authors in its 'Non-Exclusive Distribution License' to "agree that MIT may keep more than one copy of this submission for purposes of security, back-up and preservation" [12]. The fact that back-up and recovery is not specifically covered in all policies may be because this procedure is expected with any computer system that is run by a university. Nevertheless, the case may be put forward that policies should explicitly state the obvious, particularly when it can reassure potential submitters.

2.13 Differences in the documentation

The analysis of the documents and policies has identified many areas for thought when devising IR policy documentation. As stated earlier, the aim of the research was not to compare and contrast different documentation but to consider the issues covered and bring these issues to the attention of future policy makers. As such it should be noted that all the policies and documentation have a slightly different focus. The policies of those IRs that use DSpace software seem to differ slightly compared to those using ePrints software. This may be due to some of the IR documentation being derived from standard documentation provided with the software, but may also be because DSpace software concentrates more on the roles and responsibilities of the DSpace Communities, and also on the long-term issues concerning IRs. It is therefore not surprising that MIT covered these two aspects comprehensively in its policy documentation. Although HKUST also uses DSpace its IR is at a much earlier stage, and so its documentation did not cover these issues in the same depth.

However, it must be noted that IRs using ePrints software (such as QUT, and CDL), also provided comprehensive policies, though perhaps not to the same degree as MIT. QUT did cover submission/deposit procedures and IPR/copyright issues in depth and CDL covered some preservation issues, including the use of migration as a preservation strategy and stating that it will preserve digital material in the repository.

As expected, policy documentation of the more established IRs is more detailed than those of the pilot projects, with either a main policy document and linking supporting documentation, such as QUT, or a series of policy documents, like MIT. However, the pilot project documentation of OU and HKUST still covered a good range of issues such as content guidelines, submission/deposit procedures and copyright, which may be seen as the most important in the early stages of the IR. Some of the documentation is principally advocacy based, with more space being devoted to the overview issues. One particular issue that stands out is that of the level of help offered to academics, especially where submission procedures and copyright are concerned. HKUST in particular provided a high level of assistance to academics, which is not surprising as it is very much in the early stages where encouragement to deposit is essential.

The policy of Southampton University's ECS departmental repository was strong on encouraging self-archiving by explaining its benefits and focusing on copyright issues, which is of major concern to academics. In general terms a departmental policy does not need to cover as broad a scope as an institutional repository, as it has to meet the needs of a narrower range of people, i.e. those of one particular department.

Comparing the documentation has shown that there appears to be a process of progression, as an IR moves through stages in its development. The documentation must reflect this and should develop alongside the IRs themselves.

Some issues are not covered to any great depth by any of the policies, including those of the more established IRs. These issues include:

- Metadata;
- Quality control procedures;
- User policy; and
- The funding model.

Of these, the User policy was deemed by the interviewees to be the most important in the short term.

3 Policy Formulation

The interviews with IR practitioners also brought up some issues surrounding how policies should be formulated. There was a consensus amongst interviewees that the needs and opinions of individual academics and their departments are central to the success of the IR. This seemed to be the most important point to come out of the interviews. It was felt that working closely with stakeholders would ensure:

- Their existing research reporting practices are identified, and taken into account when implementing the IR;
- They know and talk about the issues surrounding IRs, especially those of greatest concern to them, such as copyright;
- They are given the chance to provide valuable input, so that, in recognising their needs, the IR will be somewhere where they will want to deposit work;
- They feel a sense of reassurance as they are being included in the IR implementation process.

Above all, academics must see the IR of benefit to themselves, and not just something they are forced to contribute to or which adds more to their workload.

It was also felt that there must be close collaboration with the senior management of the academic institution, especially as they have a key role to play in the whole institution's information strategy, and can have a direct influence on the purpose of the IR. Having a policy endorsed by senior management would also add greater weight to the IR cause.

3.1 Policy considerations

Following on from the need to cooperate closely with academics was the feeling that their needs must be reflected in the policy documentation. This involves decisions on:

- What content is to be accepted.

Some subject disciplines have a pre-print culture, whereas others produce predominantly peer-reviewed post-prints. This, in turn, has an effect on the policies to be formulated.
In many cases there are different repositories for different types of content, such as at the University of Nottingham, where there is one for research material, one for e-theses and one for work for which the University owns the copyright. However, these repositories may further need to accommodate different types of content. One interviewee gave the example of a Chemistry Department who wanted to deposit posters alongside their research papers, stating that instead of splitting these up

according to content type, they can be put into sub-sets. These requirements need to be seriously thought out by each current or prospective IR in advance to ensure the needs of their academics are met.

- How to achieve departmental, or institutional, flexibility.

All interviewees recognised that departments and even institutions have their own needs, which should be central to the IR, and which should be reflected in IR policy documentation. Interviewees had views on how this flexibility could be achieved: To avoid the confusion of having different policy documents for different departments one possible answer would be to formulate policy documents on a generic level, with possible additional documentation created specifically for individual departments. This solution has two advantages. Firstly, the documentation can be tailored to the needs and experiences of the department, making it more focused and easier to understand. Secondly, it is a way of supporting the devolving of responsibility to departments or 'communities' in such matters as types of content and submission workflows.

In addition to departmental flexibility, there may be a need for institutional flexibility. Having one IR for multiple institutions, as is the case for the White Rose Consortium (the Universities of Sheffield, York and Leeds), is an interesting situation, as all the institutions involved have to agree on decisions affecting their joint IR. This makes collaboration between the IR institutions even more important. Again, as for departmental flexibility, having separate policies for each institution, rather than just the one with supplementary documents, may make maintaining the repository difficult.

- Whether to make it mandatory to deposit

Making deposition mandatory would ensure that IRs become fully populated, especially if it is made a compulsory part of the RAE process. However, until external sources dictate that deposition in an IR is mandatory, many academics will continue to communicate their research in the ways that they are used to, such as in traditional toll-based journals and on the author's own web page, especially if they do not see the benefits for themselves of IRs. An alternative option is not to mandate in the short-term, but to work towards this goal gradually, through advocacy. The aim should be to make depositing work in the IR part of academics' standard working practice. The question of whether to mandate or not is, at the time of writing this paper, the subject of much debate. In June 2005, RCUK – a partnership between the UK's eight research councils, released its proposed position on Open Access. It proposed to make deposition of papers resulting from research funded through the councils mandatory "in an appropriate e-print repository (either institutional or subject-based) wherever such a repository is available to the award-holder. Deposit should take place at the earliest opportunity, wherever possible at or around the time of publication" [32]. However if copyright transfer agreements prohibit this, or if an appropriate repository is not available, then it could be possible that some research papers may not be deposited. Despite this potential loophole, for those institutions with repositories, the proposal should ensure that self archiving becomes part of the standard working practice for academics holding research council grants. Whether deposition is mandatory or not, once self archiving becomes commonplace it seems likely that authors will not want the loss of exposure and impact that may accompany non archived papers. It is therefore vital that academics' needs and rights are clearly represented in any IR policy documentation.

3.2. Process of policy development

All the interviewees agreed that IR policy issues should be worked through in the pilot project stage. However, even if formal policy documents are not drawn up straight away when the IR becomes a full service, IR policy decisions still need to be formalised and communicated to the academics and departments, in the form of advocacy materials and supporting documentation. One of the

interviewees considered the supporting documentation, such as 'about' pages, as being, in many ways, more important than the formal policy documents. It is this documentation that is used to encourage the population of IRs. The interviewees also made the point that policies are constantly being updated, and so should not be formulated in the early stages of IR implementation. Core policy decisions need to be made and then revised and reviewed as IRs develop.

4. Recommendations

Although the policies of individual IRs are likely to change from institution to institution, the areas that these policies should cover are likely to be similar. This work has identified these common areas. As a result of this work, the following recommendations can be made:

Approach to policy formulation

1. There should be close collaboration between the IR management/administration and individual academics, their departments and the institution's senior management. The needs and rights of the academic are central to the success of the IR.
2. The purpose and functions of the IR should be clearly defined before formulating any policies
3. IR policy decisions should be worked through and formalised in the pilot project stage, before formal policies are to be drawn up.
4. Efforts should be made to ensure that all documentation is factually correct (especially in the case of legal documentation). To ensure this it may be necessary to direct readers to sources of information that may be of benefit (such as the Sherpa/Romeo copyright database)
5. Mechanisms should be put in place to enable any guarantees to be realised.
6. Before formulating policies the following should be considered:
    - What content is to be accepted
    - How to reflect the needs of individual departments.
    - What level of responsibility is given to academics, especially concerning the checking of copyright and the deposition of their work.

Organisation And Content

7. Policy documents should be clear, concise and easy to understand.
8. There should be a formal review process put in place, so as to ensure the policy documentation develops in tandem with the IR.
9. Policy documentation should cover the following issues:
    - An overview of the IR/advocacy materials
    - Collections policy – including a withdrawal policy and any requirements relating to the quality or status of acceptable material.
    - Metadata
        - The metadata schemes to be used;
        - The level of metadata quality expected;
        - The process of metadata creation.
    - IPR and copyright (including copyright guidance and deposit license).
      The deposit license needs to contain the following:
        - A non-exclusive right to store and distribute authors' work;
        - A non-exclusive right to make copies for backup and preservation;
        - The legal rights and responsibilities of each of the stakeholders;
        - A statement identifying the copyright owner of resources in the IR;
        - A statement identifying the copyright owner of metadata in the IR.

- Submission/deposit procedures
    - Details of submission workflows;
    - Step-by-step instructions.
- Preservation policy
- User policy.

10. IR policy issues should be easy to identify and locate in the policy documentation. They do not necessarily need to be organised as separate documents.
11. The roles, rights and responsibilities of the IR, academics/departments and the institution itself should be clearly defined and set out in the documentation.

It should be realised that IRs are likely to develop and expand rapidly and that their policy documentation will have to develop alongside them. Institutions need to make decisions as to the types of content accepted, the way to introduce departmental flexibility, and the choice of IR software. These issues may affect the formulation and content of any policy documentation. It is hoped, however, that the recommendations presented in this paper will act as a basis from which IR policies, procedures and guidelines can be successfully formulated.

Acknowledgments

The authors would like to thank the John Campbell Trust for helping to cover the expenses involved in undertaking this research.

References

1. UK House of Commons Science and Technology Select Committee Recommendations. 2004. http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39903.htm

2. National Institutes of Health recommendations. 2004. http://thomas.loc.gov/cgi-bin/cpquery/?&db_id=cp108&r_n=hr636.108&sel=TOC_338641&

3. arXiv.org e-print archive. http://arxiv.org/

4. Focus on Access to Institutional Resources (FAIR) Programme. http://www.jisc.ac.uk/index.cfm?name=programme_fair

5. *House of Commons Select Enquiry, Third Special Report – Government Response*. Feb 2005. http://www.parliament.the-stationery-office.co.uk/pa/cm200405/cmselect/cmsctech/249/24904.htm

6. Directory of Open Access Repositories. 2005 http://www.opendoar.org/.
7. Open Archives Initiative Protocol for Metadata Harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html

8. Swan, A., Needham, P., Probets, S.G., Muir, A., Oppenheim, C., O'Brien, E.A., Hardy, R., Rowland, J.F.B. and Brown, S., Developing a Model for e-Prints and Open Access Journal Content in UK Further and Higher Education, *Learned Publishing* , 18(1) , 2005, pp. 25-40

9. Mackie, M. Filling Institutional repositories: practical strategies from the DAEDALUS project. *Ariadne,* 2004: 39. http://www.ariadne.ac.uk/issue39/mackie/

10. The 2004 IR Workshop Program. SPARC group. November 2004.
http://www.arl.org/sparc/meetings/ir04/ir04speak.html

11. California Digital Library eScholarship Repository.
http://repositories.cdlib.org/escholarship/policies.html

12. MIT DSpace Repository. http://libraries.mit.edu/dspace-mit/mit/policies/index.html

13. Queensland University of Technology E-print Repository for Research Output.
http://www.qut.edu.au/admin/mopp/F/F_01_03.html

14. Hong Kong University of Science and Technology Institutional Repository.
http://library.ust.hk/info/repository.html

15. University of Melbourne – Eprint Repository.
http://www.lib.unimelb.edu.au/eprints/collectionpolicy.htm

16. Open University ePrint Pilot Service. http://libeprints.open.ac.uk/about.html

17. University of Southampton Electronics and Computer Science (ECS) ePrints Service.
http://www.ecs.soton.ac.uk/~lac/archpol.html

18. Yeates, R. Over the horizon: institutional repositories. *VINE: The Journal of Information and Knowledge Management Systems*, 2003: 33(2), 96-100.

19. Lynch, C.A. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL Newsletter*, 2003: 226. http://www.arl.org/newsltr/226/ir.html

20. Day, M., Institutional repositories and research assessment. 2005.
http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/rae/rae-study.pdf

21. Callan, P. The development and implementation of a university-wide self-archiving policy at Queensland University of Technology (QUT): insights from the frontline. 2004.
http://www.arl.org/sparc/meetings/ir04/presentations/callan_files/callan.pdf

22. Johnson, R.K. Institutional repositories: partnering with faculty to enhance scholarly communication. *D-Lib Magazine* 2002, **8** (11).
http://www.dlib.org/dlib/november02/johnson/11johnson.html

23. Pinfield, S., Gardner, M. and MacColl, J. Setting up an institutional e-print archive. *Ariadne*, 2002: 31. http://www.ariadne.ac.uk/issue31/eprint-archives/

24. Lagoze, C., Van de Sompel, H., Nelson, M. and Warner, S. Conveying Rights Expressions about metadata in the OAI-PMH Framework. 2004.
http://www.openarchives.org/OAI/2.0/guidelines-rights.htm

25. RoMEO Project. http://www.lboro.ac.uk/departments/ls/disreseach/romeo/index.html

26 Publisher Copyright Policies and Self Archiving. http://www.sherpa.ac.uk/romeo.php

27. Fyffe, R. and Forrest Warner, B. Scholarly communication in a digital world: the role of an institutional repository. 2003.
http://kuschlarworks.ku.edu/retrieve/186/Institutional+repository+White+Paper.doc

28. Gadd, E.A., Oppenheim, C. and Probets, S.G., Romeo Studies 2: How Academics Want to Protect Their Open-Access Research Papers, *Journal of Information Science* , 29(5) , 2003, pp. 333-356 .

29. Gadd, E.A., Oppenheim, C. and Probets, S.G., Romeo 3: How Academics Expect to Use Open-Access Papers, *Journal of Librarianship and Information Science* , 35(3) , 2003, pp. 171-187 .

30. Pinfield, S. and James, H., The Digital Preservation of e-Prints, *DLib Magazine,* **9**(9), 2003. http://www.dlib.org/dlib/september03/pinfield/09pinfield.html

31. Swan, A., Needham, P., Probets, S., Muir, A., O'Brien, A., Oppenheim, C., Hardy, R. and Rowland, F. Delivery, management and access model for e-prints and open access Journals within further and higher education. 2004. http://www.jisc.ac.uk/journals_work.html

32. Research Councils UK: RCUK Announces Proposed Position on Access to Research Outputs. June 2005. http://www.rcuk.ac.uk/press/20050628openaccess.asp