Developing an integrated institutional repository at Imperial College London

Fereshteh Afshari and Richard Jones

Authors:
Fereshteh Afshari, IT Officer (Applications and Projects), Central Library, Imperial College, London. E-mail: f.afshari@imperial.ac.uk
Richard Jones, Web and Database Technology Specialist, Imperial College, London. E-mail: richard.d.jones@imperial.ac.uk

Research Paper

Purpose

To demonstrate how a highly integrated approach to repository development and deployment can be beneficial in producing a successful archive.

Design/methodology/approach

Imperial College London undertook a significant specifications process to gather and formalise requirements for its repository system. This was done through an initial proposal stage, and then the engagement of groups of College members with interest in the project to elucidate the requirements and allow the specification of a system that would be of genuine benefit. Then, using well understood technology for distributed systems, such as web services, and a well understood repository platform (DSpace), it was possible to undertake that work inside a structured project.

Findings

Demonstrates the advantages of producing integrated systems, especially with regard to lowering adoption barriers through easing academics' deposit workflows, introducing strong administrative tools for library administrators, and making research available in open access repositories in a well engineered environment.

Research limitations/implications

The service produced by the project is relatively new, and the long term benefits or failings cannot yet be enumerated. The paper looks primarily at the management and organisational issues but does not deal with the technical details to any great extent.

Practical implications

A useful source of information for institutions considering heavy integration work and the use of the PRINCE2 methodology for engaging institutional support.

Originality/value

This paper introduces a heavily integrated repository system within UK higher education. A lack of literature on this topic suggests that this paper could be beneficial for others considering the same route.

Keywords:

Imperial College; Institutional repository; Integration; Interoperability; DSpace; Web services

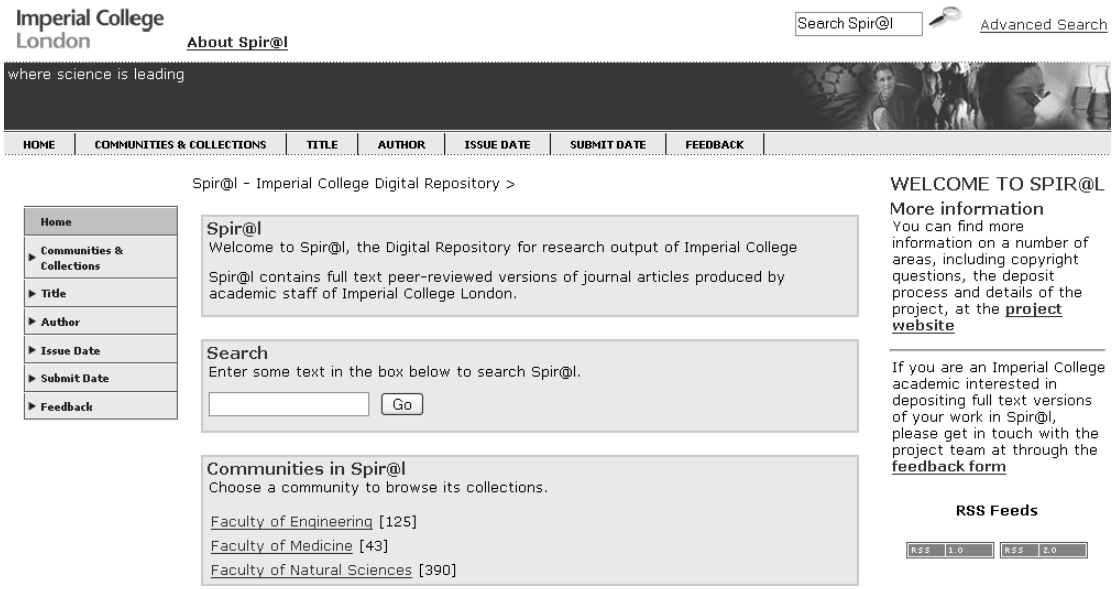
Word length: 6096

1. Introduction

In July 2007 Imperial College (IC) London made public a fledgling institutional repository (IR), Spir@l, of its academics' research publications (http://spiral.imperial.ac.uk), to coincide with its 100th anniversary as an institution and its new found independence from the University of London. Imperial College is one of the highest ranked research institutions in the world, with around 8,000 students and 3,000 research and related staff. It is exclusively focussed on science and technology, with over 50% of the institution being given over to medicine. Figure 1 shows the opening page of Spir@l.

Take in Figure 1.

Figure 1. Opening page of Spir@l



Like many institutions developing repositories, IC faces the challenges of adoption and content acquisition that can limit the speedy growth of the repository. The experiences in developing institutional repositories at Southampton University are described by Simpson and Hey (2006) and at Edinburgh University by Jones and Andrew (2005). This article looks at the choices that IC Library made before going into development and then production with this new service, and explains why, although a comparative latecomer among research institutions into the field, the techniques and technologies employed might be more beneficial in the long run. The article looks primarily at the management and organisational issues as well as the development process, but does not deal with the technical details to any great extent.

The principal objective during the creation of this archive was to ensure adoption by embedding the repository so deeply in the institution's working practices, that barriers to adoption were significantly lowered, if not removed. There were several drivers affecting this solution:

- the College already had a central Publications system holding over 130,000 bibliographic records of various different publication types (journal articles, research reports, book chapters etc.) from its researchers;

- there had been a strong and long-standing demand from academics to make full text available in their Professional Web Pages (PWPs) (see section 3.2 for more on PWPs).

It was therefore an obvious strategy to bring these existing systems together to provide an integrated solution for the repository. As an additional advantage, the

Library had previously taken part in the consortial London E-prints Access Project (http://www.sherpa-leap.ac.uk/about.html) in order to identify issues relating to setting up a repository. This original repository was based on the Eprints software (http://www.eprints.org/software/) from Southampton University, but for Spir@l, the decision was taken to use DSpace for its perceived better flexibility in terms of the code itself.

2. Repository project

2.1. Project aims

In 2005 the Library put forward a case for receiving central College funding for establishing an institutional repository to provide an archival location for all the College's research output and to make these open access. This was to include journal articles, conference papers, book chapters and all other content types available in the Publications system. The stated aims and objectives of this project were:

- to implement a quality controlled repository of academic publications;
- to provide an open access repository (complaint with the standards of the Open Access Initiative - Protocol for Metadata Harvesting (OAI-PMH)) of research publications that will raise the visibility of IC's research;
- to ensure that the content of the repository complies with copyright requirements;
- to draw on the bibliographic data already available in IC's publications system when creating metadata for repository content;
- to develop procedures for ongoing collection of research publications for deposit in the repository;
- to establish links from IC's PWPs to the full text of academics' publications in the repository, thus giving higher visibility to research output.

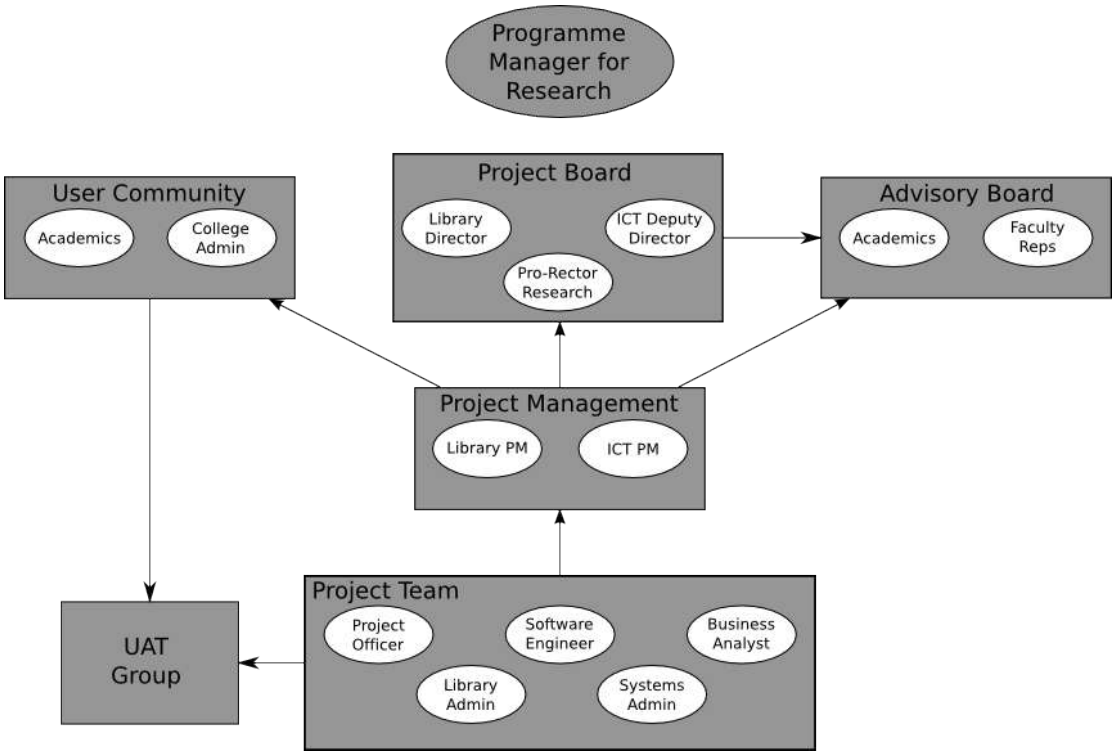
A three year project was funded as a consequence of this proposal and a team formed to do the initial analysis, design and development work. It was to be a collaborative work between the Library and the Information and Communication and Technologies (ICT) department.

2.2 Project management

IC currently requires new projects to be managed under a PRINCE2 (http://www.prince2.org.uk/web/site/home/home.asp) project management methodology. It therefore has a very well defined structure and series of groups of people involved in giving advice to the project team. Figure 2 gives details of the groups involved.

Take in Figure 2

Figure 2 Schematic view of Project Team structure



The Library and the ICT department each provided a Project Manager (PM) who worked together to develop either side of the project. Each PM was jointly responsible for the project team. The ICT team consisted of a Software Engineer, a Business Analyst, and some Systems Administration time, while the Library side consisted of a Project Officer, who was in charge of researching requirements, as well as several administrators who would aid in testing and who would be the final end-user administrators of the system. Above the Project Team sat several groups whose members helped in steering the project, and acted as a point of escalation should problems arise. The Advisory Board was a group consisting of senior academics and faculty principals, as well as the PMs, the Project Director and representatives from ICT management. The User Community were consulted on the actual requirements

of the project as practising academics and administrators of varying levels (some very senior) were involved. They would also offer feedback on the current state of the system, and propose future work. The Project Board acted as the prime steering mechanism in terms of College policy, and consisted of senior staff members such as the Library Directory (Project Director), ICT management, and the College's Pro-Rector for Postgraduate Studies.

The project obtained a lot of backing from the institution by using the PRINCE2 framework. This put the repository team in an excellent position having both top-down and bottom-up support for the endeavour. This was a significant advantage because it is often seen that the Library is the 'sole' creator, administrator and policy maker of the repository (Lynch and Lippincott, 2005). It may be a consequence of IC's relatively recent arrival on the institutional repository scene, as is evident from its standing as a research institution, and the rate of IR adoption indicated in Lynch 2005, that the benefits of institutional repositories are now more apparent, and some of the hardest lessons have already been learned. These are issues such as:

- how should the idea of the IR be sold both to the institution and to the academics?

- the lowering of the technical barriers to producing IRs, allowing for more time to be spent on yet more difficult areas;

- the paradigm of mediated deposit has been formed and tested.

3. Systems at IC prior to the integrated system

In order to give a better understanding of the implementation of the integrated repository at IC it is important to know the topology of the original systems prior to set up.

3.1 The Publications system

IC's Publications database includes bibliographic details of academic works in scholarly journals which are automatically and regularly trawled from online databases such as PubMed and Web of Science. It is not just a system for storing data, but also provides several services to the IC management to enable it to assess and monitor the level of research performance and activity within departments. In addition to holding information about existing staff publications, as soon as new

academics join IC their publications are automatically retrieved and placed in their workspace. Equally when they leave, their publications are removed from the system. Although the Publications database is an automated system for gathering material, the onus of managing lists of publications themselves is on the authors. The system sends e-mail notifications to academics when it finds new articles for them. This is to:

- inform them that new information has arrived in the database;
- ask them to 'approve' the article to ensure that it really is theirs.

Academics log in using their IC username and password and find a list of items awaiting their approval. If the article does not belong to them, they can 'decline' it. This process of 'approve' and 'decline' helps the system determine which publications may later be available in PWPs, and provides that extra human check on whether the author in the metadata really is the IC staff member. Figure 3 shows the output an example academic might receive on logging in to the Publications database.

Insert Figure 3

Figure 3. Output an academic receives from the Publications database

Academics can also manually enter details of publications which have been missed out by the automatic trawling, or which have not yet been picked up in the online databases yet, such as pre-prints.

The Publications database is also a core system for feeding other systems such as the PWPs, the Research Assessment Exercise (RAE) database, the grant application system and now also the repository. Because of this, academics are encouraged to keep their lists up to date, and to ensure that they have accepted or declined the relevant records in their workspace.

Another feature of the Publications database is its ability to identify, with a reasonable degree of accuracy, when two publication citations from different online databases are actually the same article – as can be seen in Figure 3 the "Red deer stags..." article has two entries under one record. The system uses a variety of fuzzy matching techniques to achieve this, and then marks the two records as 'joined' in the academic's workspace. This means that these records will be treated as effectively a single unit, unless the academic elects to 'split' them into two unique ones. Likewise, if there are two records which are actually the same, they can be 'joined' manually if the computer has not spotted the match.

This system is in active use by all academics at IC. If they are too busy to deal with their publications themselves, they can nominate someone else to manage them on their behalf. This nominated person may be a secretary or administrator who then can 'impersonate' one, or several people, in a department.

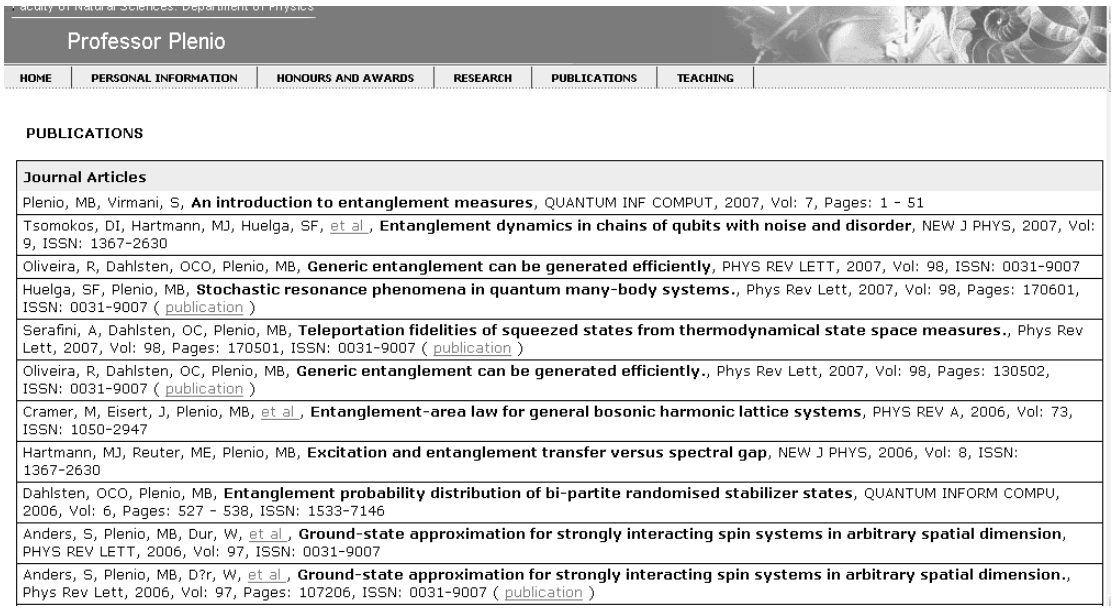
As a whole, the metrics available to management, as discussed at the start of this section show that the system is very popular among IC academics. They see it as a very valuable tool and essential to their professional activities, and which ensures that references to their work are always as up to date as possible. It is clear, therefore, why integration with this Publications database would be a benefit to the repository.

3.2 Professional Web Pages (PWPs)

One feature of IC's Content Management System is the PWPs. This is a 'site' which can be set up by academics to contain their personal information, research interests, lists of publications, any commercial activities, and so on. PWPs are divided into six sub-sections (Home; Personal information; Honours and Awards; Research; Publications; Teaching) and provide a standard, uniformly formatted personal website for each academic. Figure 4 shows part of the publications section of the PWP of one IC professor.

Take in Figure 4

Figure 4. Extract of the publications section of one academic's PWP



Faculty of Natural Sciences: Department of Physics

Professor Plenio

HOME | PERSONAL INFORMATION | HONOURS AND AWARDS | RESEARCH | PUBLICATIONS | TEACHING

PUBLICATIONS

Journal Articles

Plenio, MB, Virmani, S, **An introduction to entanglement measures**, QUANTUM INF COMPUT, 2007, Vol: 7, Pages: 1 - 51

Tsomokos, DI, Hartmann, MJ, Huelga, SF, et al , **Entanglement dynamics in chains of qubits with noise and disorder**, NEW J PHYS, 2007, Vol: 9, ISSN: 1367-2630

Oliveira, R, Dahlsten, OCO, Plenio, MB, **Generic entanglement can be generated efficiently**, PHYS REV LETT, 2007, Vol: 98, ISSN: 0031-9007

Huelga, SF, Plenio, MB, **Stochastic resonance phenomena in quantum many-body systems.**, Phys Rev Lett, 2007, Vol: 98, Pages: 170601, ISSN: 0031-9007 (publication)

Serafini, A, Dahlsten, OC, Plenio, MB, **Teleportation fidelities of squeezed states from thermodynamical state space measures.**, Phys Rev Lett, 2007, Vol: 98, Pages: 170501, ISSN: 0031-9007 (publication)

Oliveira, R, Dahlsten, OC, Plenio, MB, **Generic entanglement can be generated efficiently.**, Phys Rev Lett, 2007, Vol: 98, Pages: 130502, ISSN: 0031-9007 (publication)

Cramer, M, Eisert, J, Plenio, MB, et al , **Entanglement-area law for general bosonic harmonic lattice systems**, PHYS REV A, 2006, Vol: 73, ISSN: 1050-2947

Hartmann, MJ, Reuter, ME, Plenio, MB, **Excitation and entanglement transfer versus spectral gap**, NEW J PHYS, 2006, Vol: 8, ISSN: 1367-2630

Dahlsten, OCO, Plenio, MB, **Entanglement probability distribution of bi-partite randomised stabilizer states**, QUANTUM INFORM COMPU, 2006, Vol: 6, Pages: 527 - 538, ISSN: 1533-7146

Anders, S, Plenio, MB, Dur, W, et al , **Ground-state approximation for strongly interacting spin systems in arbitrary spatial dimension**, PHYS REV LETT, 2006, Vol: 97, ISSN: 0031-9007

Anders, S, Plenio, MB, D?r, W, et al , **Ground-state approximation for strongly interacting spin systems in arbitrary spatial dimension.**, Phys Rev Lett, 2006, Vol: 97, Pages: 107206, ISSN: 0031-9007 (publication)

The PWPs are highly customisable and allow academics to display information for different audiences. For example, they can be configured to show only certain information to the public, while allowing College users access to a greater amount of information. Often, the content of most sub-sections of these pages are populated automatically from other systems. The publications sub-section, as shown in Figure 4, is populated directly by a controlled feed from the academic's workspace in the Publications database. These pages have become very important to their owners, as they represent their professional profile, are effectively an online CV, and the most likely point of contact with their name through a search engine.

The PWPs provide a unified 'look and feel' to the academics' websites and conform to certain web and accessibility standards. Because central web servers at IC host PWPs, departments do not need to use their own servers to host these nor to provide facilities for backing up and maintenance. The information on PWPs is always up to date and, because they are managed centrally and looked after by a team of experts, links to these pages are always live.

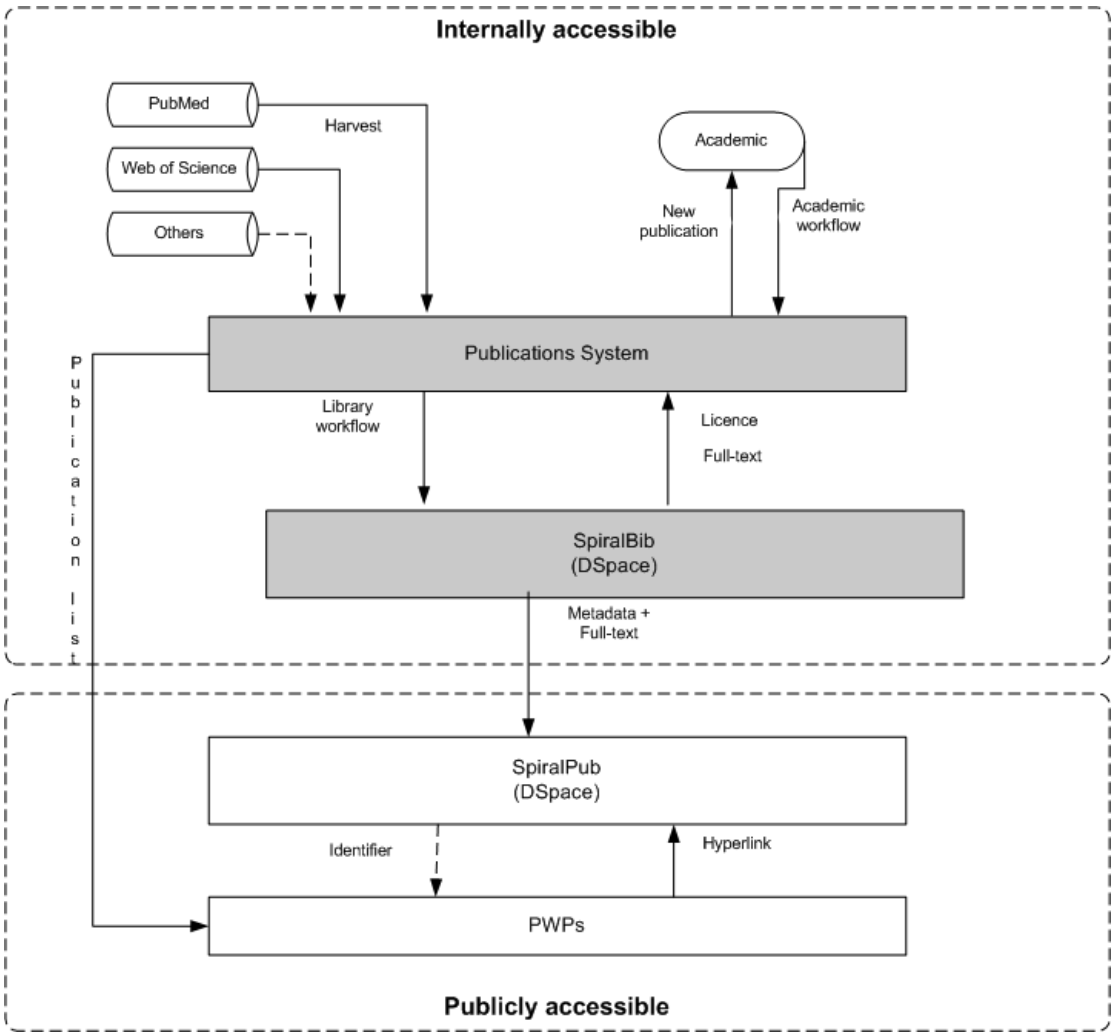
4. Spir@l: the integrated IC repository solution

4.1 General overview

Spir@l comprises four distinct components as shown in the schematic diagram in Figure 5.

Take in Figure 5.

Figure 5. Schematic of Spir@l repository solution



The four components include:

- the Publications system – to provide a unified interface for the IC academics. This also includes the bibliographic details of all publications and the academics' interface for self deposit.

- Spir@l Administrative Repository (Spir@lBib) – to deal with the administrative workflows used by the library staff. This stores the file content of the full text of the publications as well as the licence information.

- Spir@l Public Repository (Spir@lPub) – to provide public access to content which has passed successfully through the deposit and publication workflow.

- PWPs – which are populated with links from the Publications database and Spir@lPub.

Spir@lBib and Spir@lPub are both based on the DSpace repository platform (http://www.dspace.org). The work of the project was to implement this set of connections and responsibilities and an explanation of how this was achieved follows.

4.2 The seamless interface

Academics interact with Spir@l via the Publications system, as this enables them to access the repository facilities from within an interface which is popular and familiar to them, as well as in regular use. The intention was to create an environment where academics could not only manage their publications, but also could provide full-text files and check their copyright status too.

This environment was designed to interact with the repository in a seamless fashion, where the deposit workflow was simplified, and from the academics' perspective the interface would simply involve uploading their full-text files and granting a licence without needing to know what went on behind the scenes. This addresses one of the most significant problems in repository adoption and widespread use, which is that filling in metadata forms is tedious and off-putting; the Spir@l overcame this with its automated approach.

To design the 'seamless interface', the behaviour of academics was investigated and the kind of functionality they needed for the deposit workflow was identified. This included requests like:

- everything should be available on one page;

- there should be no delays in interaction;

- it should be in a familiar style;

- it should be informative and responsive.

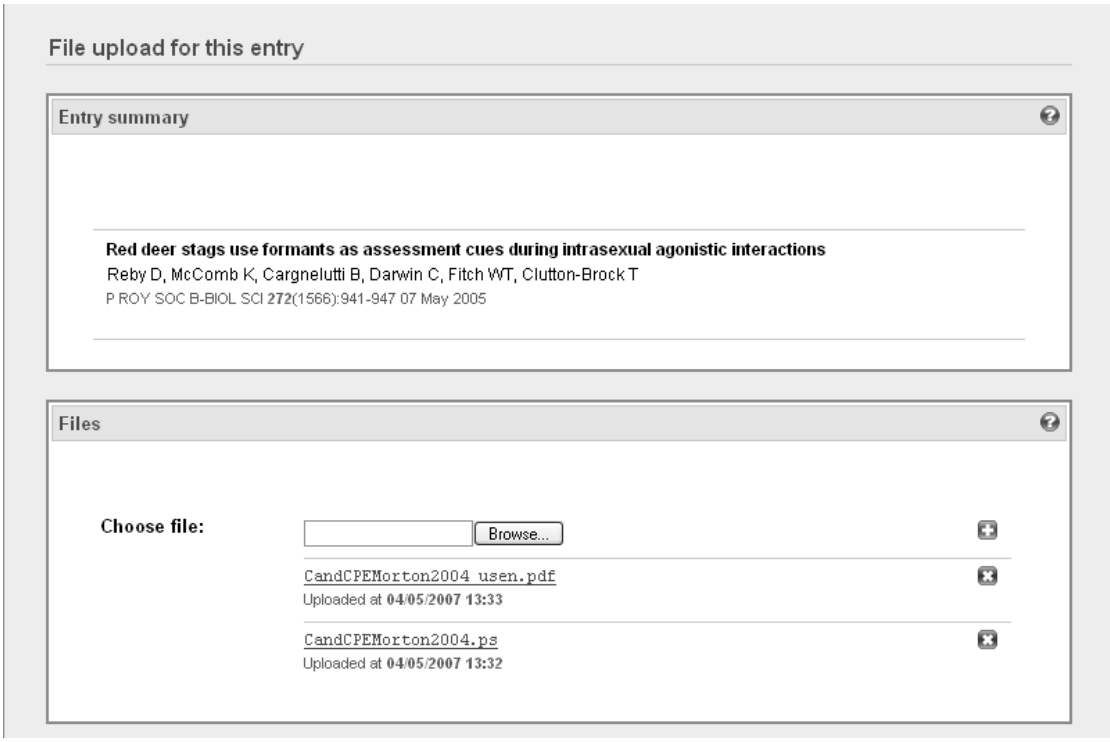
The requests were analysed and the processes was designed to provide the 'seamless interface'. Throughout the design process the User Community was consulted – these people worked closely with project staff to ensure that the repository met the academics' requirements. After an academic has logged on to the system and can see a list of automatically harvested journal articles, a button linking to the repository management page is shown. The following sections describe this page in a little detail, paying particular attention to how this is directly related to the repository management.

4.2.1 Uploading full-text files

Once in the 'seamless interface', the academic sees a brief description of the publication. Below this, there is a section for uploading as many files as are necessary. As soon as a file is uploaded, a copy of it goes to the repository and a hyperlinked file name appears below the upload box; a copy of the file is at no point stored inside the Publications database. At any point the academic can click on the file name and view the content of the file. If the file is uploaded in error, or needs to be removed for any other reason, this can be done here too. There is no specific limitation on the number or type of files uploaded, leaving it to the academic to decide on whether to include different versions of the work, addenda, corrections and any other related material. Figure 6 shows part of the upload interface for academics.

Take in Figure 6

Figure 6. File upload screen segment



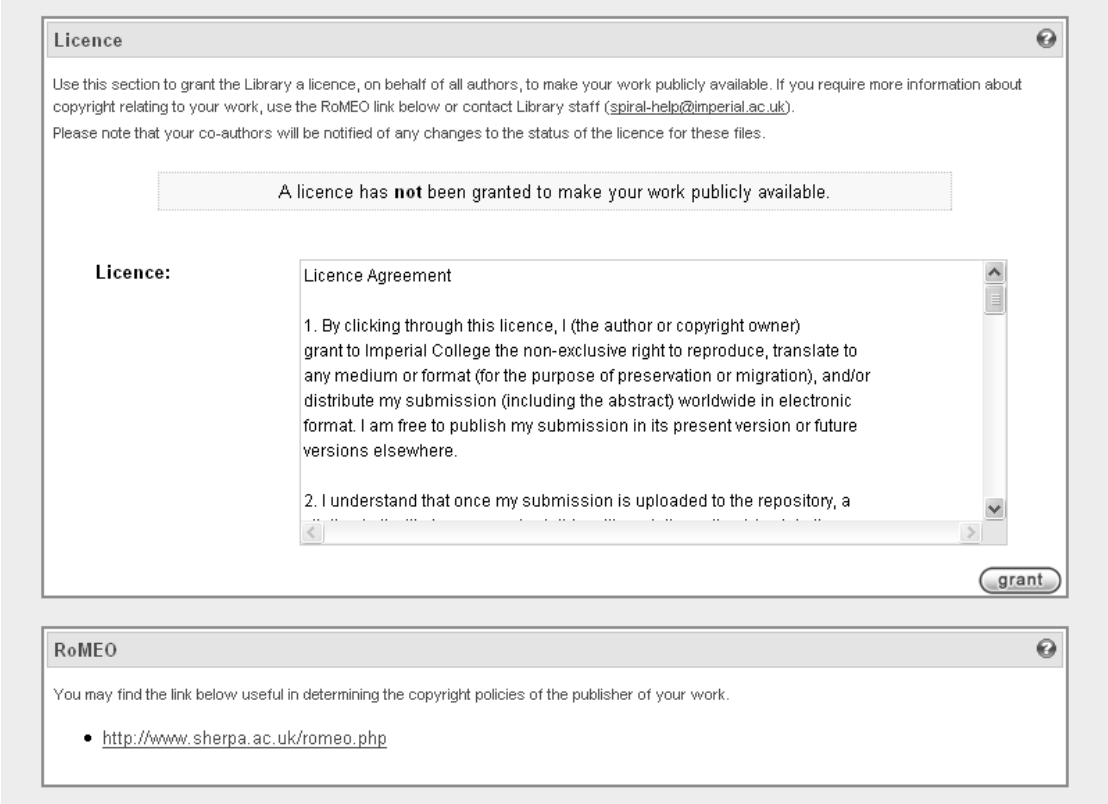
4.2.2 Licensing

The Library at IC needs to confirm that academics have permission from the relevant publisher to make work publicly available. There is a fairly standard licence available in the 'seamless interface', which consists of clauses covering the institution's rights to store, disseminate, and preserve the work, as well as providing assurance that copyright is owned by the author or permission has been given by the owner. This licence is combined with a Creative Commons Attribution, Non-Commercial, Share-Alike usage agreement (http://creativecommons.org/licenses/by-nc-sa/2.0/uk). To mitigate against future changes in the length of the licence having an adverse effect on the user interface, the licence is displayed in a scrollable text box, part of which is shown in Figure 7.

Take in Figure 7

Figure 7. Screenshot showing part of the Licensing process

Since licensing and copyright issues would normally be addressed by repository staff in their advocacy programme, it was necessary to be especially careful with sending the correct message to the academics. Clear instructions and contact details on this page were provided so that academics know who to turn to with any queries. It was understood that while preparing a work for deposit, the licence may be removed from a publication at any time, therefore it was necessary to have this functionality embedded in the system. This is because in the deposit workflow at IC, uploading files and granting the licence can happen in any order or combination. Academics can assign a licence to their publication and can also withdraw it at a later date if necessary; this is due to the collaborative nature of the environment (see section 4.2.5) and thus the need to allow previously performed actions to be undone if taken in error.



4.2.3 SHERPA RoMEO and journal policies

The final section in the seamless interface introduces a link to the SHERPA RoMEO (http://www.sherpa.ac.uk/romeo.php) website. This enables academics to check the copyright status of the journal in which their work is published or the publisher of the work. In discussions it was clear that academics in IC were concerned about copyright issues. The link to the SHERPA RoMEO website enables academics to check the copyright details of their publications before submitting the full-text file. The link to RoMEO opens a new window in the browser, so that the academic is not taken away from the repository environment. Irrespective of this, though, library administrators also check the copyright of every item as it appears in their workflow, to reduce the risk of publication in error.

4.2.4 Web services

In order to make the seamless interface work using Spir@lBib as its storage provider, it was necessary to design a set of web services which could transmit the file and licence content from one server to another, and simultaneously manage all of the additional workflow and status information that was necessary for administrators, but had to be kept obscured from the academic. It was decided that operations between the two systems would be synchronous (i.e. in real time, with no caching or delayed effects), to improve the robustness of the information model.

The advantage of using web services is that all communications are performed using XML over HTTP, and these common and well understood standards allow very different systems, such as the Publications database and DSpace, to communicate in a common language, and without creating undue dependencies.

4.2.5 Collaborative environment

An interesting consequence of the repository design, and the seamless interface is that it provides a collaborative deposit environment for IC academics, so they can see all the actions relating to their own work as well as that of their co-authors. For example, if a publication is produced by three IC academics and all three have approved this work in the Publications system when one of these co-authors uploads the full-text of the publication and then the other two co-authors log in to the system, they can also see that the publication now has a file attached to it. Equally, if a file has been queried by Library staff, any one of these authors can satisfy the query and deal with it. This functionality is a consequence of the Publications information model, which offers a many-to-many mapping between the list of publications and the list of IC staff.

4.3 Spir@lBib and Spir@lPub

4.3.1 General overview

The repository policy for deposit is both self-deposit and mediated by the Library staff. Self-deposit would take place via the 'seamless interface', as described in the previous section, and the Library would do the mediated deposit via the Spir@lBib interface. We wanted to take advantage of the quality and consistent metadata held in the Publications system and use that to provide the basic bibliographic information in Spir@lBib, and later Spir@lPub. In this way, the Library staff did not need to manually enter metadata. It was agreed that staff involved in the Publications system would provide an interface to bulk ingest their records of approved journal articles, which, when all filtering considerations are taken into account, amounted to 80,000 bibliographic records at the start of 2007.

After this initial import, the Publications system would regularly feed the repository system with any updates, additions and deletions of this bibliographic data, and Spir@lBib would be free to issue 'time-boxed' bulk update requests as a safeguard against network failure. In order to deal with ingest at this magnitude, it was necessary to consider a different approach to the structure of the repository and its deposit workflow, as the default DSpace workflow

(http://dspace.org/technology/system-docs/functional.html#ingest) was inadequate for our needs; it is too rigid, and is not designed to scale to such a large degree of throughput.

Once the bulk data had arrived, the Library staff needed to differentiate between records with full text and those without (the majority of records in the system). To do this, every record would have a series of status codes associated with its current overall state. Upon arrival, every record would be assigned default status codes, and these would be updated as actions were performed on the items. For example, when a file is added through the seamless interface the general status becomes 'updated', and a field which monitors the state of files is updated to reflect the action. Where a publication has a file and licence associated with it, its overall status would be 'Completed', assembled from the status of all its component parts. A set of 'saved searches' exist to allow Library staff to navigate the large database of items in their various conditions, and carry out the daily routines of dealing with incoming deposits. The main activity for Spir@lBib is to deal with these 'Completed' publications; the Library staff check the details of the uploaded file against its bibliographic details, then check its copyright status in RoMEO. Once happy with all details, Library staff assign the work to a Collection (or set of Collections) in the public repository and then ask the system to deposit the publication in Spir@lPub. If a RoMEO record does not exist for the publisher, the library administrator must go directly to the publisher's website for the self-archiving policy.

As soon as a full-text publication arrives in Spir@lPub, details of its URL would automatically be sent to relevant PWPs via Publications, so that the list of publications on PWPs would also have a link to the repository record. To allow for multiple files, and to ensure that any relevant copyright statements are seen by users following the links, the URL goes to a repository page which represents the item, rather than the content directly.

Throughout this many staged deposit process, it was felt to be important to keep the authors up to date with the state of their items in the workflow, and also to have a point where they can be informed of any actions that they should take themselves. To do this an e-mail notification workflow has been built that operates over a web service. Whenever an identified action takes place, all authors of an item would receive a system generated e-mail, in a digest form, summarising the Institutional Repository activities during the day. These e-mails would include information about publications which had successfully been added to Spir@lPub or any which had been queried (see below). As a consequence of the deposit interface being a collaborative environment, the e-mails also work towards the same feature. Not only the author who was responsible for uploading the file or agreeing to the licence is informed, but

all co-authors would receive this e-mail too. Effectively any of the co-authors could deal with a query, for example. Note that in the case of co-authors, we refer explicitly only to authors based at Imperial College; collaborators from external organisations are expressly not dealt with by the system, no information other than their name in the metadata is held about them, and the terms of the licensing indicates that IC authors have sought their agreement for deposit of the material in the repository.

Having described the 'basic' workflow some of the complexities that might arise are now discussed. For instance, an academic may upload into the repository a file which is not acceptable for the public repository for a number of reasons. The most likely of these is that the publisher's copyright agreement does not permit deposit of that particular version. Another reason for lack of acceptability into the public repository is that IC policy might restrict some research outcomes. In these cases, the Library can mark a particular file with a 'query', and have that query appear within the Publications interface. Using the notifications mechanism discussed above, all authors are alerted to the query, and upon visiting the supplied link will find all the details they should need to diagnose, and hopefully fix, the problem.

Individual files containing the full text of publications, therefore, take on a role within their own micro-workflow inside the larger deposit process, which takes full advantage of the power of the web services that manage interactions between these systems. Further to querying files, they can be marked as 'deleted' without physically removing them, or marked as 'generated' if there is, for example, a PDF produced from another format.

4.3.2 Metadata exchange with MODS

Spir@lBib receives an initial data feed from the Publications system to form the basic metadata in the repository system. In order to do this, it was necessary to analyse the database data structure at both ends and map the internal metadata onto our desired structure, through an intermediate format. The Metadata Object Description Schema (MODS - http://www.loc.gov/standards/mods/) standard was used for this conversion, where Publications creates the record, and Spir@lBib just needs to know how to interpret it. Any changes on either side of the system would not result in a problem for the other side, provided that this intermediate format mapping is maintained. MODS offers a rich, hierarchic schema into which it is possible to encode the potentially complex data that is moved between systems.

In addition to the standard bibliographic details, a number of additional administrative metadata fields are received, such as the IC usernames of the authors, the internal

identifiers for the records, and the source database of the record imported. These are also encoded into the MODS record for the items which is sufficiently flexible and comprehensive to support this additional data.

4.3.3 Communities and collections

DSpace structures its archive based on the idea of Communities, Collections, Items and Bitstreams (files). Institutional repositories developed using DSpace are using these semantics in a variety of ways. At IC, Communities and Collections are used to represent the College's organisational structure and group publications in their related subject areas. Faculty staff were consulted to identify the level of detail required for the Communities and Collections. In some faculties the structure remained at the department level, whereas in some others the research groups were also included.

With this structure, it is then relatively straightforward to take the incoming usernames and match their organisational details with the institutional identity management system, and use this to determine the likely deposit location for incoming publications. This is carried out during the metadata ingest process from Publications, and allows a single record to appear in as many collections as are necessary for the collaborating authors.

4.3.4 Management and administration

With an integrated system in mind a pool of people with different roles and responsibilities was needed as described earlier in Figure 2. In the ICT team, someone who could be responsible for the co-ordination of the server set-up, supporting applications, network and security issues was needed as well as a Software Engineer to carry out the bulk of development work on DSpace. A Business Analyst was engaged to capture requirements from all parties and translate those requirements to functional specifications of the system. During the final stages of the system development, it was also necessary to have a professional tester to examine the system thoroughly before its release to production.

From the Library side of the project, a Project Officer was needed to carry out the initial research and investigation on different aspects of the repository. These duties included:

- investigating copyright issues;
- defining day-to-day administration processes and general working principles;
- performing ongoing testing;
- recommending new features;

- identifying metadata requirements;
- maintaining project web pages.

Several repository administrators were later assigned to the project to carry out the tasks relating to copyright checking and publishing work to Spir@lPub.

The repository administrators are also responsible for the quality of the metadata that comes in from Publications. Despite coming from authority online sources or added manually by academics, there are still potential problems with the data, such as missing, inaccurate or incomplete details. In these cases it is sometimes necessary for the source itself to be corrected (for example, the PubMed record), as Spir@lBib is not the authority control for the metadata itself.

The system was also required to provide administrative reports to the Library staff. In addition to the complex workflow management, the ability to track exactly what had happened to all items during their life-cycle was required. This is, at least in part, because the workflow is about the present, and sometimes there is interest in the past. Therefore, audit trails of all actions on significant parts of the item that were identified are kept inside the metadata itself, for easy interaction by the Library administrators. This includes times, dates and users for events such as file upload, metadata modification, deposit and removal from the public repository and many other operations.

Another aspect of working in an integrated system is to appreciate the fact that the system is no longer stand-alone, but is a cluster of systems. Therefore, any developments or enhancements to the system require careful attention to other systems with which it interacts. If this is overlooked, the consequences of modifying or damaging a process which has implications across the distributed network can result in service interruptions, delays, extra demands on resources and even additional costs.

5. Conclusion

It is early days to assess whether this approach will be a genuine success or not, but this does not deter an initial evaluation of it in its own right. The workflow for academics has been simplified by offering them a seamless interface and allowing them to interact with the repository from within the existing Publications system.

However, this doesn't mean that it was simplified for the Library staff. Considering that there is still a large gap in what they need and what the technology can offer, the provision of a smooth and flawless interface for them is still some way away. There are a series of complex processes happening within Spir@lBib to cater for the Library staff needs, and it has no comparison to the ease of use supplied for the academics.

But as the processes in Spir@lBib are considerably more complex to achieve, a bulk of future development will be focussed around perfecting this solution.

By having provided an easy to use and attractive system at the front end it is hoped that this will encourage academics to deposit more of their publications and achieve a higher volume of full-text publications in the repository. By having a critical mass in an OAI-PMH compliant repository, such as Spir@lPub, search engines and aggregators will trawl the content of the repository and IC's research output will have a higher visibility.

Academics also see that links made to their full text in PWPs are an important feature in terms of visibility. By doing this the direct needs of the academics have been addressed and persistent and reliable links to the full text are provided in an open access manner.

Designing, developing and implementing an integrated repository has been a massive learning curve for all those involved in the project. Despite the fact that several teething problems were encountered, we believe that the benefits brought by this approach outweigh its challenges. Initial signs show that academics are supportive and see the repository system as a crucial and beneficial feature and, encouragingly, find it easy to use. We hope that by having reduced barriers inhibiting academics from using the system, we have implemented a means whereby they are enabled to take an active role and provide more of their work. The seamless interface has enhanced the Publications system and provided an environment which is attractive and already familiar to academics, and which does not require them to fill in the details of their publications. PWPs are an important part of our academics' professional activities, and we added value to these pages by enabling full-text links to their publications lists. The integrated repository, Spir@l, is now part of the College's information architecture and infrastructure.

We managed to achieve this because we received full support from the Library and ICT management. We had both top-down and grass-roots support from faculties and academics respectively, and we had access to the skill sets required to complete such an endeavour. We have been fortunate that the College has understood the importance of having an institutional repository, and we hope that in the medium term future that this system will be a leading example of integration and interoperability at just the right level to be truly productive.

References

Jones, R. and Andrew. T. (2005), “Open access, open source and e-theses: the development of the Edinburgh Research Archive”, *Program: electronic library and information system*, Vol. 39 No.3, pp. 198-212.

Lynch, C, Lippincott, JK. (2005), “Institutional repository deployment in the United States as of early 2005”, *D-Lib Magazine*, Vol.11 No. 9. Available at : http://www.dlib.org/dlib/september05/lynch/09lynch.html

Simpson, P. and Hey, J. (2006), “Repositories for research: Southampton’s evolving role in the knowledge cycle”, *Program: electronic library and information system*, Vol. 40 No.3, pp.224-231.