

Project management and institutional repositories: a case study at University College Dublin Library

Author:

Joseph Greene
Institutional Repository Project Manager
University College Dublin
325 James Joyce Library
Belfield
Dublin 4
Ireland

Joseph.Greene@ucd.ie

Abstract

This paper describes University College Dublin Library's participation in a series of parallel projects including building a national open access portal, Rian.ie; developing an international subject based portal, EconomistsOnline.org; and the planning, development and management of a university institutional repository (IR) service. Particular emphasis is placed on the use of the PMBOK® project management methodology. While much of the literature on IRs concentrates on critical success factors, only a few papers suggest applying standard methodologies to IR project planning, and very few papers detail the complex process of planning an IR using these methodologies. This paper addresses this gap in the literature and describes the practical experience of participating in two OAI-PMH harvesting projects at national and international levels and the effect that this has had on local IR development. Participating in the two services can be shown to have had a positive effect on all aspects of project management.

Introduction

In early 2007 the Irish Universities Association Librarians' Group initiated a project to build an Open Access institutional repository (IR) of research publications at each of its seven member universities. This project culminated in a national open access research portal for Ireland, Rian.ie, which harvests bibliographic metadata from all higher education institutional repositories via the OAI-PMH protocol (IREL-Open).

At about the same time, the Nereus Consortium began planning a project to build its own Open Access portal (Nereus). Nereus is a group of mainly European academic libraries and research institutions dedicated to improving access to economics resources. The consortium's flagship project, the Network of European Economists Online (NEEO), set out to create an open access research portal for economics materials that would include full publication lists and profiles of 500 or more economics authors, a multi-lingual search interface, automatic assignment of JEL classification codes, single click links to over 71,000 full-text economics publications, primary datasets and other value-adding services (NEEO).

Committing to participate in both of these projects, University College Dublin (UCD) Library faced three challenges:

- (1) To build an open access institutional repository service as required by the IREL-Open project and as necessitated by the NEEO project
- (2) To collect and archive a body of content to be harvested by both services
- (3) To make the technical modifications needed to become compliant with the two harvesters.

Participating in two harvesting projects while researching and developing an IR service clearly indicated the need for a project management methodology capable of organising and balancing multiple strands of development.

This study will endeavour to further the work of Cervone and Afshari and Jones by describing the application of our chosen project management methodology, the Project Management Body of Knowledge (PMBOK¹), to plan and implement the UCD Library IR service. It will also demonstrate some of the effects on project complexity and management requirements of participating in national and international OAI-PMH portal projects.

Following a brief literature review, we will describe the phases of the UCD Institutional Repository project. We will then detail how the five project process groups have been applied in this project. Finally, we will describe how each of the nine project knowledge areas described by the PMBOK Guide have been affected by participating in the two harvester projects.

Literature review

¹ PMBOK is a registered trademark of the Project Management Institute, Inc.

There is a growing body of professional literature specific to building IRs and other kinds of open access repositories. Much of the literature concentrates on success factors; notable examples include Hoorens and Villalba-Van Dijk, and Westell, whose recommendations converge at several points. Kim and Kim take this line of enquiry a step further, providing a quantitative diagnostic evaluation framework for IRs.

Success factors and performance indicators provide both direction for planning and methods for measuring success, but do not elucidate the 'how' of planning and implementing an IR. Not surprisingly, Campbell-Meier's case study on IR development in six US doctoral universities found that '[p]roject planning was an area that did not get much attention from the project teams (128).' One of three primary findings of the study is the importance of incorporating project management practices into IR development (Campbell-Meier 145).

Addressing this need for project management practices, the Library and Information Technology Association (LITA) National Forum Conference Report provides an introduction to the PMBOK project management methodology (Riggs 15). A contributor at the conference, Cervone describes the PMBOK methodology and its potential for being applied to digital library development.

A case study by King describes a project carried out by the Decision Science Support department of a large multinational corporation (10-24). The study provides an informative insight into the planning and execution of a repository project in a sector where project management techniques are common practice (King 13).

While Cervone illustrates a methodology and King details the inputs and outputs of a repository project, neither fully link a specific methodology and its application to IR planning and development. This seems to have only been addressed to date by Afshari and Jones, who give a detailed description of a project in the higher education sector which made use of the PRINCE2 project management methodology to create an integrated institutional repository system. The paper points out that using a standard methodology increased buy-in within the institution, stating that it 'put the repository team in an excellent position having both top-down and bottom-up support for the endeavour' (Afshari and Jones 341).

Project phases

As defined by the PMBOK Guide, 'a project is a temporary endeavour undertaken to create a unique product, service, or result' (Project Management Institute 5). Projects can be divided into broad phases and each phase broken into subprojects for better management control (Project Management Institute 17-19).

As the UCD Institutional Repository project evolved, three distinct phases emerged. The first phase will be the main focus of this study and could be considered a research and development phase.

Phase one consisted of project planning, selection of and technical modifications to the repository platform, securing departmental, author and publisher agreements to openly archive a large body of content (1000+ items) and metadata cataloguing activities. The workflow in this phase was geared towards the retrospective collection of the entire publication outputs of thirty-two academic economists in order to fulfil the requirements of the NEEO project. As such, the work of locating full-text

documents, gaining permission from publishers to reuse the works and uploading and cataloguing the documents fell entirely on library staff.

Phase two of the project is currently underway at the time of writing and consists of mainstreaming the service both within the library and throughout the university. This includes the establishment of a system for authors to self-deposit their works, a copyright information and clearance service and metadata quality control and cataloguing work, collectively known as the workflow.

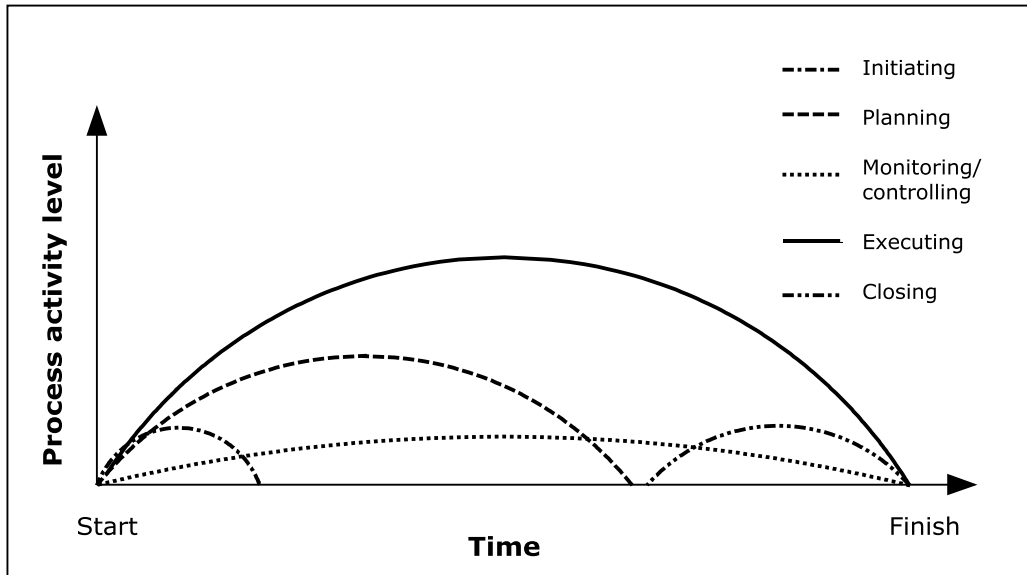
The third and as yet undefined phase of the project will be informed in part by the outcomes of phase two. Strategic library planning focusing on the UCD Digital Library as a whole may provide for the institutional repository becoming integrated with other digital library initiatives, other research support systems and potentially the addition of an e-thesis system. In effect, the third stage may signal the end of the UCD IR as an individual entity and the beginning of it being a facet of the broader UCD Digital Library.

Project process groups at different levels of application

The PMBOK Guide describes five project management processes that are present in all types of projects, across all industries (Project Management Institute 37). Summarising Cervone (31), the five project process groups are:

- (1) Initiating processes: gaining project or phase authorisation
- (2) Planning processes: defining project objectives, committing resources and planning work in a work breakdown structure (WBS)
- (3) Executing processes: putting the plan into action
- (4) Monitoring and controlling processes: measuring results and preventing variance from the plan
- (5) Closing processes: formalising project acceptance and ending project processes

Figure one describes the five process groups in terms of their interaction and level of activity over time.



Level of activity applied to each process group over time

Figure 1

The knowledge, skills and processes needed for successfully delivering a project are not, however, uniformly applied on all projects. It is the project manager's responsibility to determine the appropriate processes and to what extent they are applied to the project at hand (Project Management Institute 37).

Phase one

Throughout the UCD IR project lifecycle, the level and clarity of process group application has been varied. Phase one, which covers the majority of what was originally viewed as the IR project, was not explicitly planned using a standard methodology, though an *a posteriori* description of the five process groups at this level is still quite informative and shows how the process groups were at work, even if implicitly.

Repository development in 2007 was a relatively new field for UCD Library. A digitisation project begun in 2005, the Irish Virtual Research Library and Archive (IVRLA), provided information initially but was of a sufficiently different nature -- mass digitisation of primary source materials -- to leave many unknowns for planning an open access research repository.

The decision to build the repository and to participate in the NEEO project came from the senior management level and initiation work began as early as 2005. Funding from the national project gave leverage and provided the necessary impetus for the E-Services department to begin selecting and customising the appropriate software package. Initiation and planning were therefore a matter of course; having decided to participate in the projects, planning followed quite naturally, though not initially using a standard methodology.

Working with the liaison librarian to the School of Economics, the project recruited a group of economics scholars to participate. The Library hired a cataloguer whose time was allocated half to the IR project, half to the Library's Information Resource

Management department. Resources for copyright, data entry and technical support were reallocated from existing library staff members. The cataloguer researched metadata schemas, the team planned a basic workflow and data entry, cataloguing, copyright clearance and full-text acquisition began. These constituted the execution processes. Later in the first phase the Library hired a full-time project manager, who set about providing more detailed planning for individual subprojects and managed overall project execution.

The NEEO project required that each institution submit monthly statistical updates on total metadata records, total full texts, staff time and repository usage. Periodic meetings were held with team members to discuss progress. Together these activities formed the basis of the monitoring and controlling process. Figure 2 details a subset of this control data.

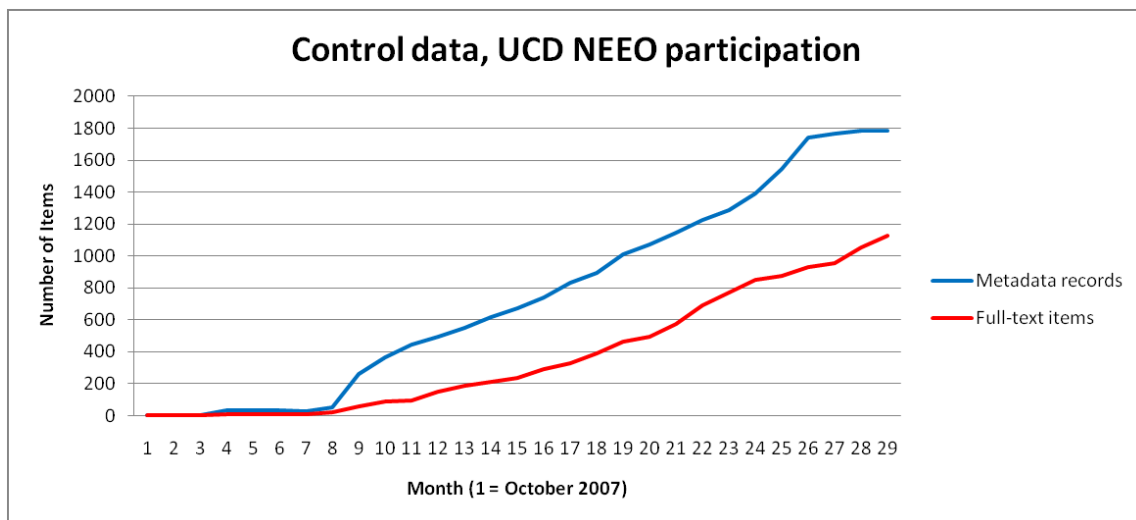


Figure 2

Since the IR project is ongoing, closing processes have not yet been fully implemented at the general level. However, this case study could be viewed as part of the closing processes of phase one in terms of administrative closure and as an update of organisational process assets (Project Management Institute 100-102). Closing the project phase should also include a hand-over of the new knowledge and processes developed in phase one to an expanded production team and steering group within the Library.

Phase one subprojects

In order to more carefully understand and manage the overall project, phase one was broken into approximately twelve subprojects. Figure 3 below, in the section on time management, details a high level summary of some of the major subprojects.

The five process groups were applied in varying degrees at the level of subprojects. In general, they were applied more or less rigorously depending on the level of prior knowledge in the area. For example, there was often little institutional knowledge in relation to research and development projects such as metadata schemas and data normalisation, copyright clearance and harvester compliance. In these cases the projects consisted of a period of research and experimentation followed by a review of the research and implementation of the findings. Usually a single team member

carried out the research and worked closely with the project manager during implementation.

The research reports themselves were major outputs of the research and development (R&D) subprojects. The reports provided the necessary knowledge to inform development. The copyright R&D project produced control variables to be measured and suggestions for further risk management, in addition to the main output of clearing over one thousand full text publications for use in the IR. The metadata R&D project produced a metadata schema and a data normalisation plan. Technical R&D projects specified detailed requirements and produced technical solutions for harvester compliance.

As subprojects began to involve more staff and risk increased, project planning and project process application became more explicit. Early in the project a new research centre became involved and acted as the subject for a pilot of a self-submission process, where the head research administrator would collect and submit previously published or in-publication articles to the repository using the built-in submission system. Initiation consisted of discussions with one of the research centre's principle investigators and the senior research administrator, internal discussions with library management, and the IR project team meetings to define the project. These discussions also involved obtaining authorisation at each level.

Planning the self-submission pilot consisted of a project team meeting to detail required preparatory actions and project execution responsibilities. The project manager set up email notifications for each item deposited in order to monitor the process flow. In addition, a new method of managing and monitoring the copyright process was developed for the project based on the submission of the postprint (authors' peer-reviewed final draft) version of papers, instead of requesting permission from publishers to use published versions. Apart from this, very little was needed to monitor the project apart from regular discussions with team members.

Once the research centre's first year's worth of publications had been processed, the project team met to discuss the outcomes of the project, the data was analysed and the project closed. The arrangement with this research centre is now an ongoing service. In addition to the collection of open access publications, the copyright recording instrument and the data produced during the project became outputs that would later inform project planning for the full scale university service.

Project processes were more explicit in a scalability project. This was designed to incorporate a second Library section into the IR project and to rapidly process a collection of seventy working papers. The scalability project involved five library assistants who had not previously worked on digital projects in the Library.

Because of the increased level of inter-departmental collaboration, process groups were carefully controlled in the scalability project. A relatively detailed project plan documented six of the nine knowledge areas related to the project. The plan was produced as an output of successful initiating and planning processes involving department heads, line managers, library assistants and IR project team members.

In this project, the IR project cataloguer monitored time spent by each library assistant and the items submitted. Once all execution processes were complete, managers involved in the project were informed and a closing meeting was held with the library assistants to evaluate the project from their perspective, resulting in a list

of lessons learned. These lessons learned were used to improve the process and the project was repeated with a second set of content.

Phase transition

The final subproject of phase one will be the project to mainstream the IR as a full university-wide service, in essence a transition from phase one to phase two. This project, in planning at the time of writing, has required a careful initiation phase even though the overall IR project has been running since 2007. The need for an additional initiation process between project phases is not unusual and is described by the PMBOK Guide (Project Management Institute 43).

Initiation consisted in the formation of an IR steering group made up of senior Library management, department heads and the IR project manager. Clearly defining the project goals, delimiting risks and managing the expectations and influence of stakeholders were crucial elements for laying the grounds for the planning processes.

Planning for the mainstreaming of the IR service took into account the level of exposure the project would place on the Library; the project plan was accordingly detailed. The five project process groups were carefully considered and each of the nine knowledge areas were researched, debated by the project team and steering group and detailed in the project plan. All of the previous research and development projects will have served as inputs for each of the five project process groups. Variables to be monitored during execution will include the following:

- (1) Number of newly registered depositors
- (2) Number of copyright queries from authors
- (3) Number of new item types and formats required by authors
- (4) Number of deposits per day/week
- (5) Number of items in each workflow stage per day/week
- (6) Number of items per copyright action category
- (7) Number of embargoed items
- (8) Rate of throughput (deposits/copyright/quality control/cataloguing)
- (9) Frontline staff time spent on project execution processes

If the project successfully proceeds as planned, the closing process should culminate in a report that analyses the control data, summarises discussions with frontline staff and selected researchers who will have availed of the service, suggestions for process improvements and lessons learned. It is expected that the main project deliverable will be a complete self-submission open access institutional repository service, fully integrated within UCD Library's model of service provision as defined in the Library's 2007 Operational Plan.

Project knowledge areas and the effects of participating in harvester projects

The PMBOK Guide provides a detailed description of 44 project management processes that make up the process groups detailed above, organising them into nine project management knowledge areas (Project Management Institute 77-298). Rather than repeat or attempt to summarise the PMBOK Guide, we will provide

examples of how participating in harvester projects informed each of the nine project knowledge areas.

Integration management

The PMBOK Guide states that 'project management is accomplished through processes, using project management knowledge, skills, tools, and techniques that receive inputs and generate outputs' (Project Management Institute 37). A primary output or deliverable of phase one was to bring the IR into compliance with the two OAI-PMH harvester projects. This meant both technical compliance as well as compliance with the stated goals of collecting a body of open access content.

The harvester compliance requirements served as inputs for research and development projects, which themselves output the necessary knowledge for compliance. In addition, the knowledge created during phase one has had a major influence on planning for phase two, the mainstreaming of the IR service. The planning of the IR phases and subprojects around harvester project goals provided the basis for project management planning, one of the seven integration management processes.

Furthermore, as detailed above, monthly statistics required by the NEEO project constituted the main source of data for monitoring and controlling processes in phase one, another integration management process.

Though never formally described, the project team has adhered to a *de facto* change control procedure. This procedure was only required when the IREL-Open working group's chair requested that the UCD IR project expand beyond economics to include the research publications of a new inter-university project in computer science. This was incorporated into phase one as the self-submission pilot subproject described above.

Scope management

By the time the full-time project manager joined the project team, the project scope had been clearly defined. The first work of the project manager was to begin creating a work breakdown structure (WBS). This was accomplished by first considering the requirements of the two harvester projects. Working outward from harvester requirements, it was possible to determine what tasks were required to achieve harvester compliance. For example, in order to send MODS bibliographic data via OAI-PMH to the NEEO portal, first the local metadata would have to be correctly structured, then the software to crosswalk the local qualified Dublin Core would need to be written and installed. Using this decomposition method, a complete list of tasks and dependencies was constructed, constituting the WBS, a key component of scope management.

Time management

Deadlines set by both of the harvesting projects provided targets for activity sequencing and schedule development. Despite limited abilities in estimating the duration of R&D projects, it was possible early on to at least provide an accurate activity sequence based on task dependencies revealed in the creation of the WBS. Subprojects and tasks that had been detailed in the WBS were placed into a highly

speculative Gantt chart which was updated regularly as more became known about each task. A summarized final Gantt chart is shown in Figure 3.

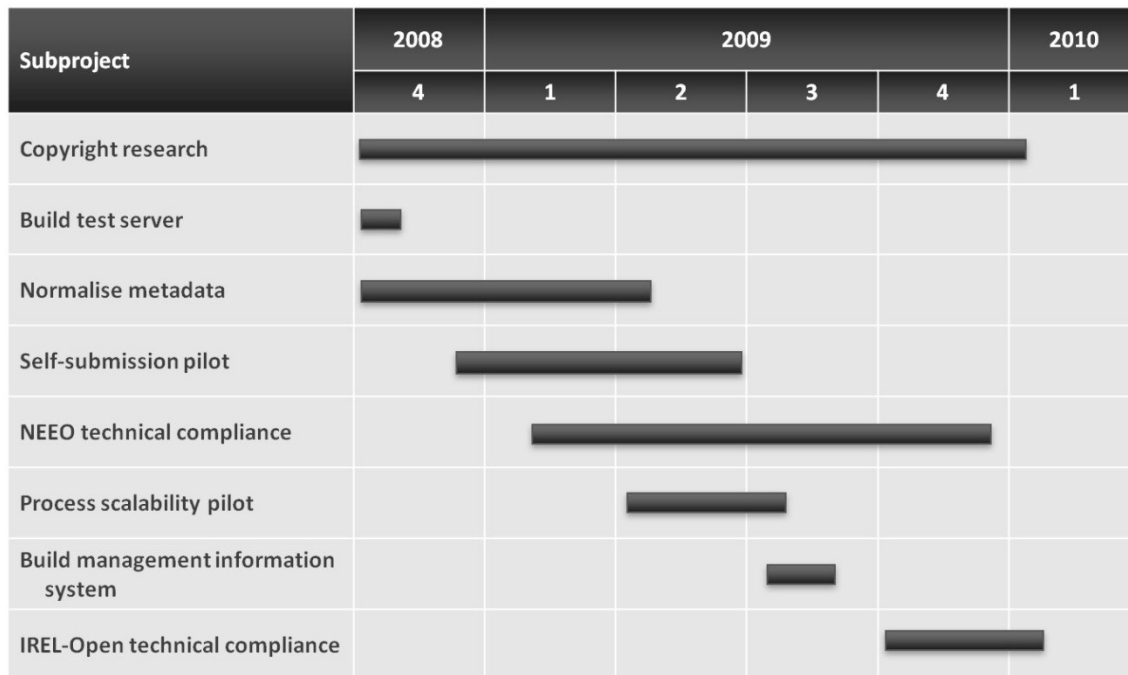


Figure 3

Cost management

Staff time spent on the project was by far the highest cost. In addition to the human resource capital provided by the Library, the harvester projects provided funding to offset the direct marginal costs of participation. The NEEO project aimed to provide 50% funding for staff time allocated to project activities. The IREL-Open project funded the equivalent of one half time dedicated staff member, 50% funding for other staff members' contribution and funding for equipment.

Quality management

The main focus of quality management has been on metadata quality. Metadata requirements made by both harvester projects guided the implementation of the local metadata schema and resulted in a robust metadata schema. In order to be able to export the local qualified Dublin Core data as MODS for the NEEO project, the project team performed a detailed review of the metadata schema and normalised the use of metadata fields, improving existing metadata and defining future metadata quality requirements.

Some modifications were made to the local schema in addition to data normalisation, particularly with author and editor data. In the NEEO MODS bibliographic metadata implementation, name data was required to be exported with first name and family name in separate fields, along with a role code and digital author identifier (DAI). This required the use of an automated subfielding technique in an additional name field, so that these elements could be parsed by a metadata crosswalk program

when sending the metadata via the OAI-PMH interface. This work would not have been possible without name authorisation, another metadata quality process.

Human resource management

The need for staff members to work in copyright clearance and cataloguing was clear from the outset of the project. The scale of the NEEO project required that this work be shared by several staff members; these staff members formed the core IR project team. Several subject librarians assisted in recruiting participants and liaising with the schools that were involved in the NEEO project. The library also hired a full-time project manager in order to successfully deliver the IR service and achieve the goal of complying with the two harvesters.

Communications management

Scheduled project meetings for the two harvester projects provided invaluable opportunities to meet with IR teams and stakeholders from higher education and research institutions all over Ireland, the European Union, the United States and Australia. The influence of this face-to-face knowledge sharing on the UCD Library IR's development cannot be overemphasised. Without the personal contacts made at project meetings, many of the IR's goals could not have been met. In addition, project meetings provided for performance reporting and information distribution.

Communicating harvester project requirements and developments within UCD Library encouraged participation in the project at many levels of the Library; marketing the two projects communicated benefits and incentives to stakeholders outside of the Library including heads of schools and the UCD Research Office.

Risk management

Apart from the usual risks of running a Web based service, copyright is the greatest risk when openly archiving research papers, many of which are published by very large and wealthy corporations.

Both harvester projects provided a great deal of information related to copyright through research and knowledge sharing. The task of the NEEO project, to archive the life's work of each participating researcher, required extensive work in relation to copyright that is beyond the usual remit of IRs.

Many publishers allow authors to openly archive the final peer-reviewed draft of their papers (the postprint) in their local IR. Since most of the papers to be included in the NEEO project had been published years before the project began, an intensive process of requesting permission to use published versions of the majority of papers produced strong practical knowledge and a large dataset related to copyright, item types and publisher policies, above and beyond the usual tools such as the SHERPA-RoMEO database (RoMEO).

The knowledge and dataset built around the NEEO project copyright clearance activities provided for the categorisation of copyright related risks, data on individual publishers' behaviour, and new approaches to this high risk, politically charged field of service provision. Analysis of the data revealed the number of items by publication year where publishers had never responded to requests to reuse items. Another analysis of the data showed how many queries had been ignored by what we

considered low risk publishers. These data mining exercises may inform quantitative and qualitative risk analyses in future, allowing the IR to carefully take more risks in order to increase access to publicly funded research.

Procurement management

There were two major purchases made for the IR project. The first and largest was the contract awarded to an external company to customise the repository software platform. The second purchase was for a metadata crosswalk that would output DIDL wrapped MODS records from the native qualified Dublin Core via the OAI-PMH interface for harvesting by the NEEO portal.

The Library purchased the crosswalk in partnership with another NEEO partner, the European University Institute (EUI) in Florence, Italy. At the time, EUI were using the same IR platform and version. In addition, the two institutions' NEEO compliance developments were at about the same stage. Purchasing the crosswalk together was mutually advantageous: UCD provided metadata crosswalk information that was easily adapted by EUI. EUI located the specialist software developer, drafted the contract and negotiated a discount group rate. The software developer needed to write code and documentation only once, making minor adjustments for each institution.

Discussion

We have seen how project process groups have been applied at the level of project phases and subprojects of the UCD Library IR project. Participating in harvester projects as an integral part of project planning has also been shown to have had a positive effect on all project management knowledge areas.

In addition, the collection of over 1,100 full text open access documents generated during harvester project work has positively influenced end user buy-in, both those who deposit their research and those who download documents from the IR. The evidence for depositor buy-in included a documented increase in demand for the service and even a perceived competition for inclusion by individuals and academic units prior to mainstreaming. On the other hand as demonstrated in Figure 4, downloads have steadily increased over time, disproportionately to the number of documents archived.

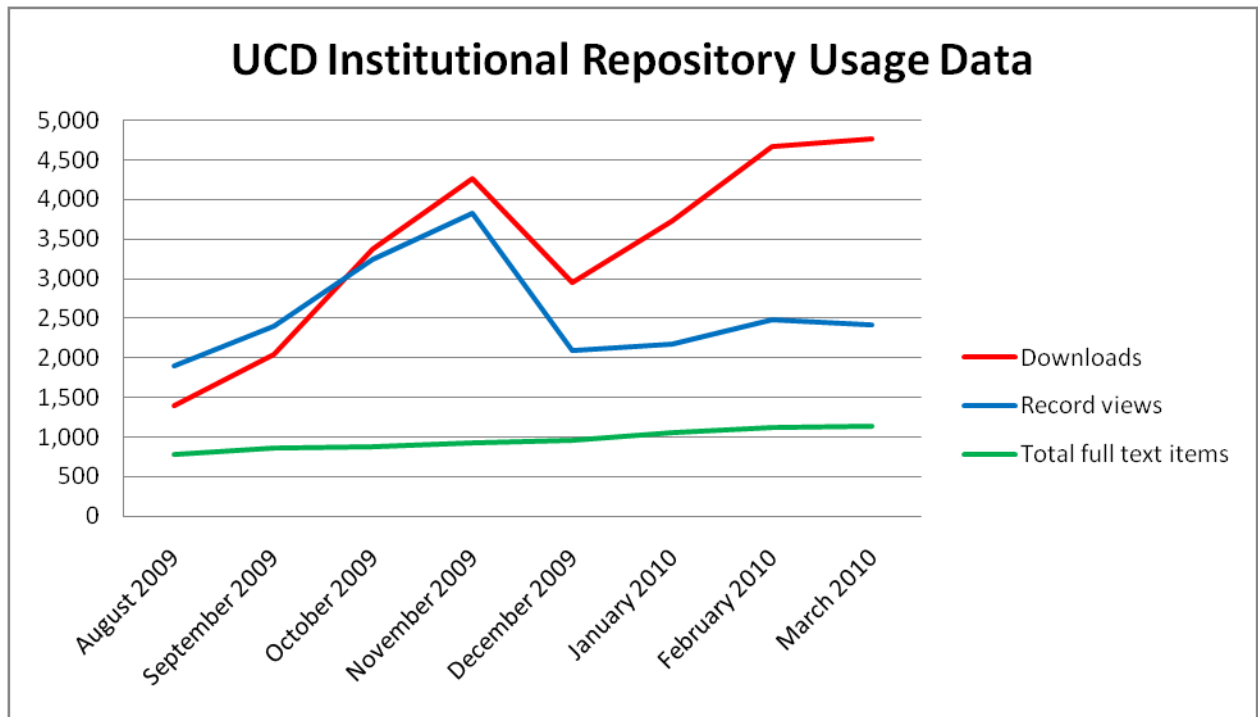


Figure 4

Furthermore, the use of a standard project management methodology, in combination with harvester project involvement may have also assisted UCD Library in realising critical success factors for embedding the IR into the wider university. UCD Library's IR meets many of the points raised by Hoorens and Villalba-Van Dijk: providing direction to developing a vision and strategy, communicating benefits and providing incentives (Hoorens and Villalba-Van Dijk 1-4).

Conclusion

It could be argued that the use of a standard project management methodology has provided the tools and techniques while participating in harvesting projects at national and international levels has provided the direction and impetus for successfully building an open access institutional repository service at UCD library. The results of phase two of the overall project will confirm or contradict this and can be demonstrated by the rate of uptake by UCD researchers over time (i.e. the number of deposits), the level of uptake across disciplines and in usage information such as patterns in copyright queries and download rates.

Areas of improvement include streamlining workflow processes, systems interoperability, an improved deposit interface, a usage statistics notification service for depositors and heads of schools, platform considerations and long-term preservation of materials. Many of these points are likely to be addressed in the third phase which could see the IR becoming part of the development of a comprehensive digital library service.

There is evidence both in the literature (Afshari and Jones 341) and in our practice that the use of a standard project management methodology can increase internal stakeholder buy-in, from frontline staff, line managers, middle management and

senior management. In a 2009 study of academic librarians' attitudes and corresponding actions in relation to open access, Palmer et al. (327) point to a discrepancy between librarian support and actions taken that reflect the stated support of open access. It is the author's hope that the current study and studies like it that seek to explicate the planning and implementation of an open access repository service through standard methodologies may assist in the move from support in principle to support through action.

References

- Afshari, Fereshteh, and Jones, Richard. "Developing an integrated institutional repository at Imperial College London." *Program*. 41.4 (2007): 338-352. Web.
- Campbell-Meier, Jennifer. "Case studies on institutional repository development: Creating narratives for project management and assessment." Ph.D. thesis. University of Hawai'i, December 2008. Web.
- Cervone, Frank H. "Standard methodology in digital library project management." *OCLC Systems & Services: International Digital Library Perspectives*. 23.1 (2007): 30-34. Web.
- Hoorens, Stijn, and Villalba-Van Dijk, Lidia. "Embracing the future: Embedding digital repositories in the University of London." *Rand Europe*, Cambridge: TR-625-UCL, 2008. Web.
- IREL-Open. Web. 19 March 2010 <<http://www.irel-open.ie/>>
- IVRLA. Irish Virtual Research Library and Archive. Web. 19 March 2010 <<http://www.ucd.ie/ivrla/>>
- Kim, Yong Ho, and Kim, Hyun Hee. "Development and Validation of Evaluation Indicators for a Consortium of Institutional Repositories: A Case Study of dCollection". *Journal of the American Society for Information Science and Technology*. 58.8 (2008): 1282-1294. Web.
- King, Michael A. "The Implementation of DSpace at the InterContinental Hotels Group: A Knowledge Management Project Success." *Journal of Cases on Information Technology*. 10.2 (April-June 2008): 10-24. Web.
- NEEO. Economists Online Version 2.0. Powered by the NEEO Project. Web. 19 March 2010 <<http://www.neeproject.eu/>>
- Nereus. Web. 19 March 2010 <<http://www.nereus4economics.info/>>
- Palmer, K. L., et al. "Where There's a Will There's a Way?: Survey of Academic Librarian Attitudes about Open Access." *College & Research Libraries* 70.4 (July 2009): 315-35. Web.
- Project Management Institute. *A Guide to the Project Management Body of Knowledge*. 3rd ed. Newton Square, PA: Project Management Institute, 2004.

Riggs, Colby. "2006 LITA National Forum: Conference Report". Nashville, Tennessee, 26-29 October 2006. Library Hi Tech News. 10 (2006): 15-17. Web.

RoMEO. SHERPA/RoMEO Home - Publisher copyright policies & self-archiving. Web. 19 March 2010 <<http://www.sherpa.ac.uk/romeo/>>

Westell, Mary. "Institutional repositories: proposed indicators of success". Library Hi Tech 24.2 (2006): 211-226. Web.