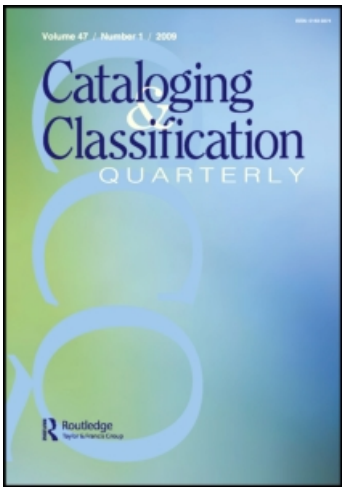
This article was downloaded by: *[University Of Utah]*
On: *5 May 2009*
Access details: *Access Details: [subscription number 909990238]*
Publisher *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cataloging & Classification Quarterly
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t792303976

University Scholarly Knowledge Inventory System: A Workflow System for Institutional Repositories

Anne Morrow [a]; Allyson Mower [a]
[a] University of Utah, Salt Lake City, Utah, USA

Online Publication Date: 01 April 2009

To cite this Article Morrow, Anne and Mower, Allyson(2009)'University Scholarly Knowledge Inventory System: A Workflow System for Institutional Repositories',Cataloging & Classification Quarterly,47:3,286 — 296
To link to this Article: DOI: 10.1080/01639370902737307
URL: http://dx.doi.org/10.1080/01639370902737307

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Cataloging & Classification Quarterly, 47:286–296, 2009
Copyright © Taylor & Francis Group, LLC
ISSN: 0163-9374 print / 1544-4554 online
DOI: 10.1080/01639370902737307

Routledge
Taylor & Francis Group

University Scholarly Knowledge Inventory System: A Workflow System for Institutional Repositories

ANNE MORROW and ALLYSON MOWER
University of Utah, Salt Lake City, Utah, USA

The University Scholarly Knowledge Inventory System (U-SKIS) provides workspace for institutional repository staff. U-SKIS tracks files, communications, and publishers' archiving policies to determine what may be added to a repository. A team at the University of Utah developed the system as part of a strategy to gather previously published peer-reviewed articles. As campus outreach programs developed, coordinators quickly amassed thousands of journal articles requiring copyright research and permission. This article describes the creation of U-SKIS, addresses the educational role U-SKIS plays in the scholarly communication arena, and explores the implications of implementing scalable workflow systems for other digital collections.

KEYWORDS digital repositories, workflow management, copyright permissions, digital collections, collection development

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

INTRODUCTION

In 2005, the campus libraries at the University of Utah collaborated to establish initiatives, best practices, and a collection development policy for an institutional repository. Librarians began outreach efforts around campus, educating faculty on retaining rights to their publications and recruiting their participation in the repository (named USpace). Participating

Received July 2008; revised September 2008; accepted November 2008.

The authors, along with Scott Cowley, Application Programmer at Marriott Library, created the University Scholarly Knowledge Inventory System.

Address correspondence to Anne Morrow, Digital Initiatives Librarian, J. Willard Marriott Library, University of Utah 295 S. 1500 E., Salt Lake City, UT. 84112. E-mail: anne.morrow@utah.edu.

faculty were assured the repository coordinators would first research an article's copyright and verify that the original publisher would permit submission to a long-term, open access archive. The careful consideration that had been given to developing the institutional repository (IR), and the energetic efforts to recruit faculty, led to a large influx of potential content early in the repository's formation. The most critical issue quickly became developing a workflow that would scale to meet USpace's growing needs.

At the outset, coordinators consulted both SHERPA/RoMEO and individual publisher Web sites to determine archiving rights. If SHERPA/RoMEO did not include a publisher, staff would search for the publisher on the Web to locate a policy. As repository coordinators found policies, they began storing them on shared network drives. Although it was helpful to have somewhat ready access to a publisher's policies, the set- up required consulting at least two sources to determine rights. In addition, SHERPA/RoMEO, while informative on self-archiving, is not necessarily tailored to address institutional archiving; any information collected needed to be checked thoroughly on a case-by-case basis to determine if institutional archiving was one of the provisions. Moreover, publisher Web sites were inconsistent; policies were in some cases not immediately available or decidedly vague and ambiguous. Over the course of several months, coordinators contacted publishers directly via e-mail to find out their policies or to clarify existing ones. This resulted in an additional source to consult when verifying archiving rights. It became clear that a single rights source integrated into other IR work such as assignment of metadata and uploading of files would be an efficient approach to prepare material for USpace.

Ultimately, USpace required a tool that could store items, policies, communications, and author information. Coordinators wanted to be able to report on progress made from multiple points: date, authors, colleges, departments, and so on. Most importantly, a system was needed that could separate content into queues according to publishers permissions. Coordinators looked at open-source as well as commercial software, but nothing met all of the requirements and, as a result, a system was built in-house. The tool—University Scholarly Knowledge Inventory System (U-SKIS)—includes all the functionalities listed earlier and provides an informational foundation from which to educate faculty on copyright and publisher's archiving positions. In addition, the tool has the potential for assisting in the workflow of other digital collections dealing with rights management, communications, authors, and creators.

METHODS

U-SKIS began as an MS Access database created by the repository coordinator at the health sciences library with the following mission statement:

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

> "The workflow database tracks University of Utah authors, their publications, contacts made by library staff and others regarding publishers' copyright policies for submission to the institutional repository and PubMed Central (PMC) as well as author's permission to submit to the IR or PMC."

From this, the coordinator developed a simple, five-table database (authors, publications, publishers, library staff, contacts). The mission statement also defined the specific fields required for each table. Once completed, library staff at the health sciences library could access it via a local, shared drive. It quickly became obvious that the database could not handle multiple users within the health sciences library let alone the other IR team members located in the main library on another part of campus. In order to accommodate multiple users in disparate locations, the database needed to be Web accessible. A rough HTML interface was created and the data in the existing MS Access tables were sent to an application programmer at the main library. The programmer set up a similar structure in MySQL and used Perl 5.5 for the programming. The coordinators and the programmer spent several months testing the system for usability.

To go along with the system, coordinators developed workflow teams (Figure 1) as well as a series of training documents for each team. These major components intentionally reflected traditional library workflow. The IR team felt that treatment of electronic information should similar to analog/print information. Coordinators gathered IR material either via faculty submission from the USpace Web site or by talking directly with faculty members about consulting their curriculum vita (CV) to archive their material. Once received, items are assigned to staff to begin the process in U-SKIS. Bibliographic investigation includes the discovery of the item's publisher, acquisitions involves obtaining the item itself if authors have only provided IR managers with a citation on a CV, and permission entails sending direct communication to either a publisher or an author.

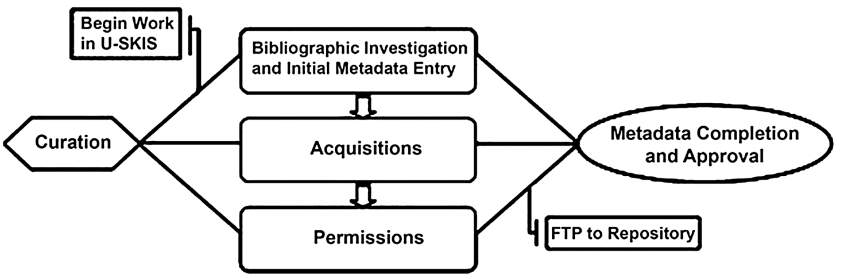


FIGURE 1 Workflow Teams for Processing IR Material.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

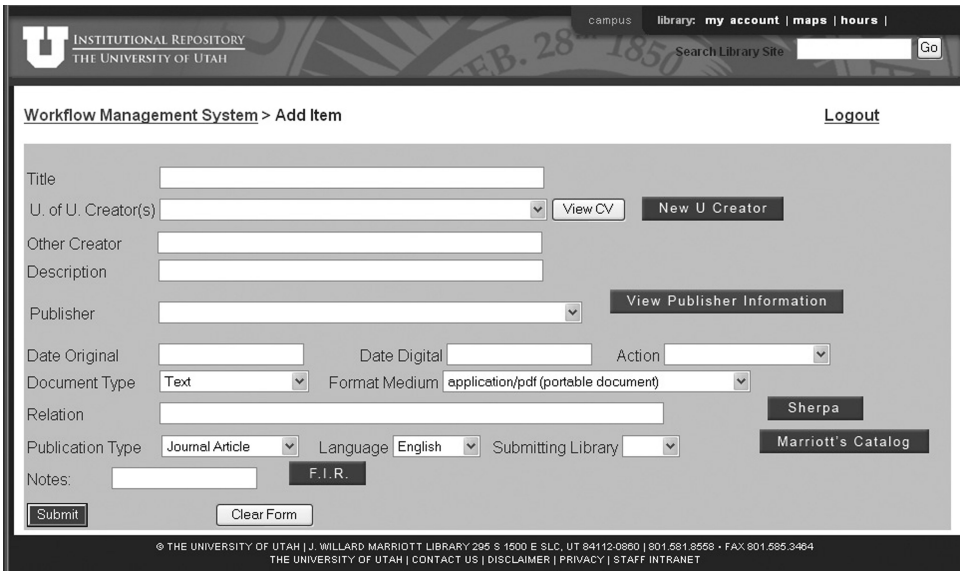


FIGURE 2 Dublin Core Fields Used to Describe an Item. Fields are Mapped to Those in the Repository and are Populated at the Time of FTP.

The system tracks an item and/or citation from initial ingest through the permissions and acquisitions process until it becomes fully prepared for IR inclusion. U-SKIS follows the Dublin Core standard to apply metadata, which are then re-used once the item is ready to be added to the repository (Figure 2). System administrators utilize FTP to send files and metadata to the repository's CONTENTdm software (Figure 3).

Items move through the system depending on 'publishers' archiving policies (Figure 4). For each item, a publisher is selected, which opens up its record showing if the item can be archived and which version can be archived (author's manuscript or 'publisher's PDF). The record also indicates what the 'publisher's policy is regarding deposits to PubMed Central. Information on respective embargo periods is also included. Items that fall within an embargo period remain in an embargo queue and can be sent to the repository once they have been cleared. Publisher records are updated to reflect any changes, but mechanisms are also in place to save permission history so that it is possible to point to permission granted at a certain period in time (Figure 4).

Records are created for each faculty member involved in USpace (Figure 5). These records can be used to compile statistics regarding the number of people involved and from what college and/or department. The system also houses authors' curriculum vita so that repository staff can work through past publications.

U-SKIS tracks communication between repository staff, publishers, and content creators. It records when a contact is made, who made the contact

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

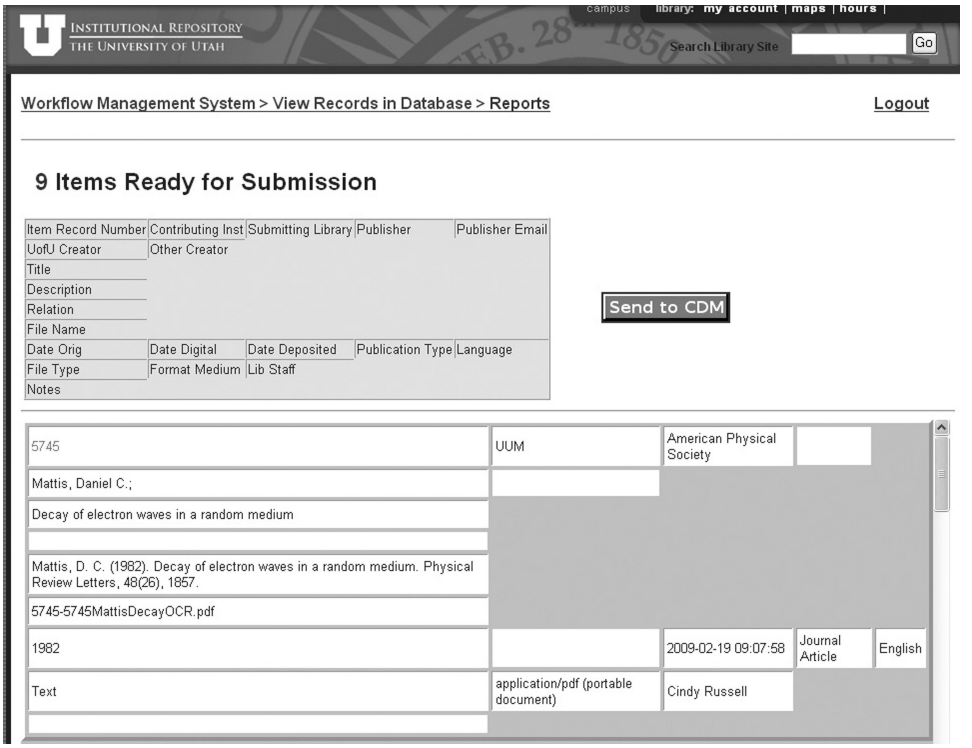


FIGURE 3 Completed Items are Sent to the Repository via FTP.

and by what means (phone, e-mail, in-person). The system also includes searching and reporting features that assist in the permission and acquisitions process and in collection development decisions (Figure 6).

RESULTS

The University of Utah Libraries have used U-SKIS since July 2007. IR coordinators and staff have added over 3,800 items and 400 publishers. Approximately 400 articles have been uploaded to USpace via U-SKIS. Another 1,800 items are eligible once publisher requirements have been met. U-SKIS was fully completed in December 2007 and added to sourceforge.net in February 2008. Since then, approximately 70 institutions have downloaded U-SKIS. The Libraries of Claremont Colleges have fully implemented it as part of their IR work. Other digital collections at the University of Utah, namely the Utah Artists Project and Western Soundscape, have customized U-SKIS to suit their project needs.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

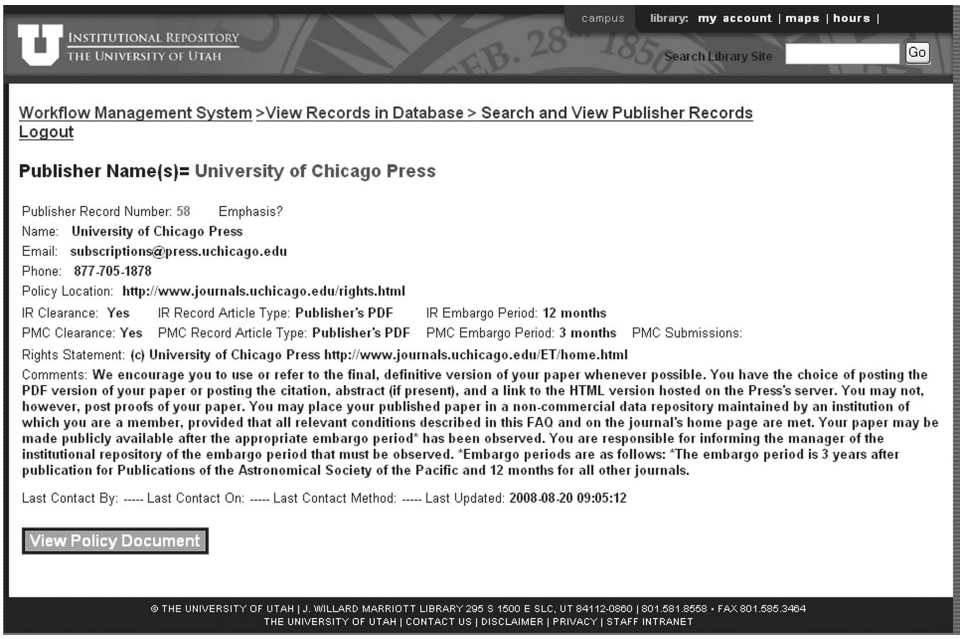


FIGURE 4 Publisher Records Indicate IR Clearance and Required Article Type. The Record also Includes the Policy Document that Coordinators Referred to at the Time of the Record's Creation. This Provides a Permission History.

DISCUSSION

Academic libraries have always preserved access to scientific and creative research literature. This traditionally has come in the form of subscribing to, cataloging, and shelving journal volumes and books. This function—to

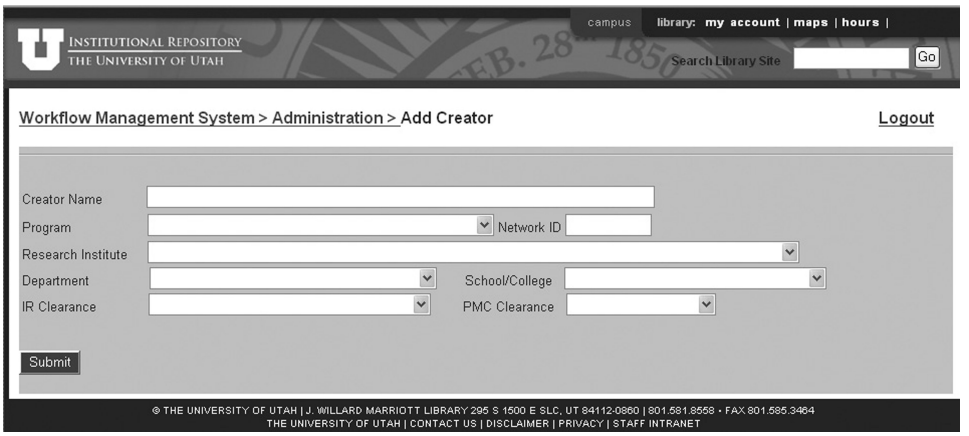


FIGURE 5 Creator Record.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

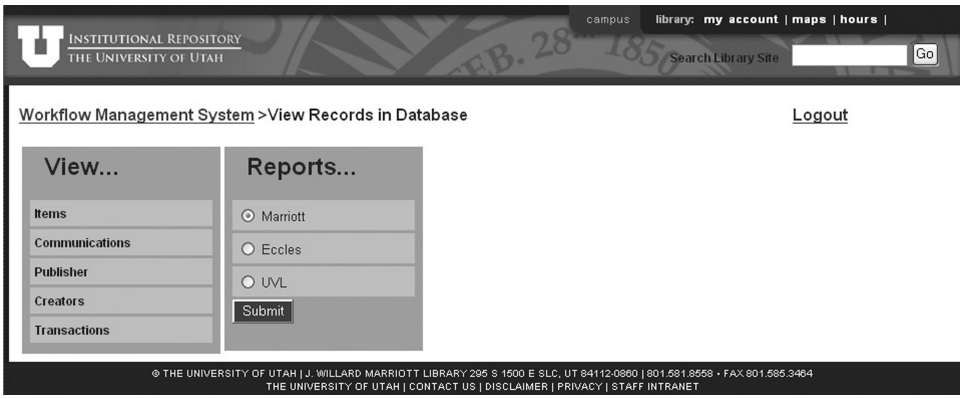


FIGURE 6 Reports.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

curate, describe, and maintain access to information—is changing in response to both the rise in digital information and the rise in journal costs. A campus community has the potential to create millions of megabytes of digital information that remain unpublished, uncollected, or unmanaged.[1] Many institutions built repositories thinking content providers would flock to them, but it has been quite the opposite.[2] Repository managers have approached the low deposit rate in various ways. Loughborough University developed relationships with key stakeholders such as the research office, the information services committee, and the program development and quality team.[3] University of Massachusetts Medical School raised their profile by adopting a hosted content site that provided them with robust marketing tools.[4] The University of Cambridge looked at the deposit habits of early adopters (as well as non-adopters) to develop a deposit approach tailored to chemists.[5]

The IR coordinators at the University of Utah hoped to address low participation rates by developing library workflow practices that would facilitate the archiving of the opus of a faculty member's academic work and not restrict it to recently published material. As an archive, USpace stores full-text searchable documents in addition to the metadata record. This goal of maintaining access to full text documents became a key issue to resolve in terms of both archiving standards and copyright. The full text of an article needed to be stored on local servers (and not linked) so the campus libraries could provide a persistent URL. Therefore, only a limited number of previously published scholarly articles could be added to the collection due to publishers' archiving policies.

However, this became a valuable opportunity to explain to authors the significance of retaining certain rights in the future to ensure their interests remained protected. As coordinators worked with the curriculum vitae of participating faculty, they also updated them on the archiving position of

publishers. Working with the complete collection of an author's publications represents a meaningful position from which to educate faculty on their rights as authors. The effort is certainly labor-intensive, but this kind of service opens up opportunities for libraries and librarians to interact with departments and faculty in nontraditional ways. One department in particular at the University of Utah provided IR managers with all of the curricula vitae ($n = 40$) of their faculty. Although not all rights were cleared, the department faculty became more acutely aware of their copyrights and the ramifications of transferring them. As a result, the department has begun discussing the possibility of a departmental process to archive research results.

Although the role of U-SKIS is to track items, it also assists coordinators in educating faculty on copyright and archiving. Of the 110 publishers in U-SKIS that support deposit in an open access institutional repository, the majority of them do not allow deposit of the publisher's PDF. Approximately 1,020 articles currently in U-SKIS could be added to the repository if the authors had retained their final, peer-reviewed manuscripts (commonly referred to as the author's manuscripts). Part of the educational efforts of the IR team focuses on the importance of authors retaining the peer-reviewed manuscript. U-SKIS give coordinators concrete evidence to point to in conversations with departments and faculty, especially in terms of which publishers do not support archiving.

Moreover, IR managers have utilized U-SKIS to examine strengths and weaknesses in the repository's outreach program. With U-SKIS, IR managers can determine the level of participation across departments. For example, the philosophy and ophthalmology departments have the majority of their faculty participating, yet only a small percentage of the biology department has material in the repository. The tool helps us frame questions about nonparticipation and the relationship between repositories and the communication practices of individual academic disciplines. This, in turn, contributes to broader scholarly communication initiatives. As Davis and Connelly have noted, faculty members have a "diversity of attitudes, motivations and behaviors regarding the use of digital repositories"; their motivation for participation in IRs is in part affected by the patterns of research and publication within their discipline.[6]

While informing scholarly communication programs, U-SKIS also provides practical means for discovering ways to achieve mass IR deposits. U-SKIS centralizes publisher policies and groups them based on their permissions status. This grouping, in conjunction with other external tools such as bibliographic management systems, can assist in identifying "low-hanging fruit." For example, coordinators employ EndNote to search across bibliographic databases based on publishers that support the archiving of the publisher's PDF version. This represents additional (and, perhaps, more effective) means for coordinators to populate the repository.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

In the process of implementing U-SKIS for the institutional repository, the issue of workflow for digital collections in general emerged. The tendency has been to focus more on developing content than building an integrated workflow infrastructure to support the digital content stored by libraries. As workflows have generally remained localized, U-SKIS is an attempt to apply to digital content the same kind of integrated library systems that have supported the workflows for analog materials. While U-SKIS does provide an opportunity for digital collection administrators to customize and adapt the tool to suit the needs of their collection, pursuing further adaptation in order to improve digital collection workflow management in general is also merited.

FUTURE DIRECTIONS

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

A relatively limited number of previously published articles can be deposited because of faculty publication habits and policies of publishers. As a result, an author's most seminal works may be absent from the repository. The most attractive solution would be the initiation of a mandatory deposit policy by the institution. However, such a policy may not include, much less have any legal ground to enforce, the deposit of previously published articles. The only way to capture the scholarly works already disseminated is if the publisher has some degree of acceptance of institutional archiving and the author and repository are able to satisfy publisher conditions. According to the DRIVER study conducted on repositories in the European Union, the "situation with regard to copyright of published materials and the knowledge about this among academics . . . " is a concern shared by more than half the institutions surveyed, and is viewed as a roadblock that may interfere with development of digital repositories.[7] An important litmus test for U-SKIS will be whether administration of it includes ongoing maintenance to assure the veracity of the information it stores.

It is hoped the development of the Open Access Directory wiki will contribute to the effort to centralize information on publisher policies. In the spring of 2008, Arta Dobbs, Molly Keener, and Scott Lapinski issued this announcement on behalf of the Open Access Directory wiki:

> Following the recent launch of the Open Access Directory http://oad.simmons.edu/ wiki by Peter Suber and Robin Peek, work has started on a list collecting publisher policies on NIH-funded authors. Recognizing that many of our colleagues have been working on such lists individually, we invite you to share information, including existing lists, with us so we may create as comprehensive a resource as possible. While this list is certainly not meant to serve as a replacement for SHERPA/RoMEO, much interest has been expressed in its creation.

IR coordinators have contributed publisher information housed in U-SKIS to this effort and hope the directory can serve as a comprehensive resource. It will be important that the administrators of U-SKIS develop an ongoing process of ingesting and updating information on publishers, otherwise the value of the information on publishers will become inconsistent, irrelevant, and inaccurate.

Even though there are many published works still under copyright, it does not necessarily mean repository managers should abandon the logic of building institutional repositories that reflect the body of a researcher's work. As libraries strengthen their scholarly communication outreach programs and as scholars become more involved in open access movements at their institutions, institutional repositories will be in a better position to deposit work produced in the future.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009

CONCLUSION

As the amount and type of information in digital form increases and the scholarly communication arena evolves, librarians find themselves in the midst of a transitional phase. Getting university authors and creators to self-archive in open access repositories will understandably be a lengthy process. As Dorothea Salo has noted, researchers in the pre-digital age rarely archived material on their own; the scenario in the digital age is not much changed.[8] Maintaining a proactive exchange with faculty with the intent of increasing their understanding of scholarly communications issues will, the authors hope, increase the number of publications that can be added to the repository in the future. U-SKIS helps address low deposit rates and workflow challenges involved in managing an institutional repository.

NOTES

1. John F. Gantz, David Reinsel, Christopher Chute, Wolfgang Schlichting, John McArthur, Stephen Minton, Irida Xheneti, Anna Toncheva, and Alex Manfrediz, *The Expanding Digital Universe: a forecast of worldwide information growth through 2010* (Framingham, MA: IDC, 2007), http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf.

2. Nancy Fried Foster and Susan Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories," *D-Lib Magazine* 11, no. 1 (2005), http://www.dlib.org/dlib/january05/foster/01foster.html.

3. Joanna Barwick, "Building an Institutional Repository at Loughborough University: Some Experiences," *Program: Electronic Library and Information Systems* 41, no. 2 (2007), http://hdl.handle.net/2134/2640.

4. Mary E. Piorun, Lisa A. Palmer, and James F. Comes, "Challenges and Lessons Learned: Moving From Image Database to Institutional Repository," *OCLC Systems & Services* 23, no. 2 (2007), http://works.bepress.com/piorunm/3.

5. Peter Morgan, "Facilitating the Deposit of Experimental Chemistry Data in Institutional Repositories: Project SPECTRa" (paper presented at the annual meeting for the International

Association of Technological University Libraries, Stockholm, Sweden, June 11–14, 2007). http://www.dspace.cam.ac.uk/handle/1810/195216.

6. Philip M. Davis and Matthew J.L. Connolly, "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace," *D-Lib Magazine* 13, no. 3/4 (2007), http://www.dlib.org/dlib/march07/davis/03davis.html.

7. Maurits van der Graaf, "DRIVER: Seven Items on a European Agenda for Digital Repositories," *Ariadne* 52 (2007), http://www.ariadne.ac.uk/issue52/vandergraf/.

8. Dorothea Salo, "What You're Up Against" (paper presented at the NISO/PALINET Conference, Washington, DC, December 3, 2007), http://minds.wisconsin.edu/handle/1793/22081.

Downloaded By: [University Of Utah] At: 00:05 5 May 2009