Building Value-added Services for Institutional Repositories (IRs): Modeling the Rutgers Experience

Myoung C. Wilson, Rutgers University
Ronald C. Jantz, Rutgers University

International Federation of Library Associations
Social Science Libraries Section, Satellite Conference
Social Science Libraries: A Bridge to Knowledge for Sustainable Development

Biblioteca Nacional de Cuba José Martí, Havana, Cuba
8-10 August 2011

Sponsored by:

United Nations Educational, Scientific and Cultural Organization – Cuban Office
Biblioteca Nacional de Cuba José Martí
Springer Publishing
Center for Global Studies (CGS) at the University of Illinois
University of Illinois Library
International Federation of Library Associations

This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.
Introduction

With the launch of DSpace in 2002 by the MIT University Libraries in collaboration with the Hewlett-Packard Corporation, there was a widespread expectation that institutional repositories (IRs) would be able to capture the rapidly growing digital assets of an institution. Initially, the primary purpose of IRs was to increase access to those research materials that had been produced in digital formats at home institutions and, in the process, to preserve and highlight the intellectual capital of an institution. IR developers also hoped that one consequence of IR development would be the promotion of open access scholarship that would reduce the increasingly unsustainable cost of accessing research information.

The developers of IRs, typically a collaboration between librarians and repository developers at large research universities, envisioned that by utilizing Web 2.0 technology, the consumers of IR information would also be the producers of information. This type of interactive system, where users both supply information and disseminate “user supplied content,” is the model for the success of social media ventures such as YouTube, Wikipedia, FB, Twitter, and EBay. The hope that was nurtured by the developers of IRs was that institutional repositories, too, would become “participatory information ecosystem[s]” (Purcell, 2010). However, faculty members, who have traditionally served as both consumers and producers of research information, did not, for a number of reasons, prove to be willing suppliers of their research into IRs. Basefsky (2009) argues that this is largely due to IRs’ narrow focus on capturing intellectual output rather than offering IRs as a “social academic research service. The slow acceptance by faculty members of IRs as a place to deposit their work has prompted many academic librarians to re-evaluate the role of IRs as a component of academic research.
Building Value-added Services for Institutional Repositories (IRs)

Literature Overview

Three areas of research on IR development are highlighted with selected examples. These areas are: (1) the status of IR development in different institutions; (2) the underperformance of IRs as repositories for faculty work; and (3) the development of IRs as a critical scholarly node in inter- and intra-institutional networks fostered by the addition of new functionalities and services.

Research on the Status of IR Development.

In contrast to earlier studies where researchers utilized a case study methodology to report on IR development in individual institutions (Baudoin & Branschofky, 2003), a number of studies have also investigated the status of IR development across academic institutions (Jantz & Wilson, 2008). A recent representative of these studies was a census of IRs carried out in 2007 by a team of University of Michigan researchers (Markey, Rieh, St. Jean, Kim & Yakel, 2007). They conducted a survey of 2,147 college and university libraries in the United States to determine the status of planning and implementation of IRs. The study reported that, among 446 respondents, almost 53% had done no IR planning, 20.6% were contemplating IRs, 15.7% were actively planning and pilot testing IRs and about 10.8% had already implemented an operational IR. This survey clearly revealed that although IRs may be gaining approval overall, acceptance is far from universal. DSpace was identified as the lead platform for launching an IR followed by ContentDM, bepress and FEDORA. (Markey, et al., 2007).

Status of IRs as Repositories for Faculty Research.

The second area of IR research focuses on the continuing underperformance of IRs as repositories for faculty research. Several recent studies suggest that faculty deposits in IRs are

extremely low for a variety of reasons. For example, a survey on the use of PocketKnowledge, a
digital repository at Teacher’s College, Columbia University found that “students contributed
the bulk of user-generated content to PocketKnowledge (89%) while instructors (faculty,
adjuncts and instructors) self-archived only 2%” (Asunka, Chae & Natriello, 2010). Jantz and
Wilson (2008) also reported that one third of the IRs represented in the Association of Research
Libraries (ARL) websites was empty of faculty deposits. However, another recent study on the
use of digital materials deposited in an IR found that the most heavily used materials were the
articles (presumably deposited by faculty and graduate students) and undergraduate theses
(Connell, 2011). The latter category then affirms the 2010 survey of SPARC (Scholarly
Publishing and Academic Resources Coalition) attendees in which 86% indicated that
showcasing student research in an IR will become a growing trend in the future (Bankier, 2010).
Three types of studies are worth noting with regard to faculty deposit practices (or non-
practices) and the relationship of these practices to the status of IRs as repositories for faculty
research. The first of these types of studies investigated faculty self archiving practices. For
example, Kim (2010) identified five different venues that were used by faculty to self archive
their research; the most frequently used venue was faculty personal web pages (66.7%) while
IRs were the least frequently used (22.7%). The other venues employed were research group
websites (51.5%), departmental websites (41.7%), and disciplinary repositories (28.3%). This
survey also revealed that younger faculty members with more proficient technical skills were
most involved in self archiving in contrast to others where the extra time and effort inherent in
self-archiving represented barriers (Kim, 2010). Kim also noted that rights management issues,
such as obtaining copyright clearances, remain a huge barrier to faculty self-archiving. In order
to alleviate copyright complications, Hanlon and Ramirez (2011) report that a majority of IR managers offer a “mediated deposit model” in which library personnel, rather than faculty authors, seek publishers’ permission. In an earlier study, Davis and Connolly (2007) conducted similar interviews with the Cornell University faculty members regarding their low use of DSpace and found similar responses as those that were found in Kim’s survey. One surprising finding of the Davis and Connolly study was that the faculty members at Cornell considered “access” to research materials, one of the pillars of IR foundational goals, a non-issue for them.

A second type of study explored the perceptions and experiences of IR faculty end-users. Here, researchers found that there is no common understanding of the term “institutional repository” on the part of IR faculty end users. Institutional repositories are often confused with other licensed library databases where material is stored or misunderstood as one of the crowd sourced online products such as Wikipedia (St. Jean, Rieh, Yakel & Markey, 2011). While this study also reported that the most common navigation path to IRs is through the library homepage, an earlier study of academic library IRs noted the lack of uniform navigational paths to IR sites from the library homepage, implying the lack of a common understanding of what an IR is, even among IR developers (Jantz & Wilson, 2008).

In a third type of study that analyzed the citing patterns of Web resources by faculty, Zhang (2011) revealed that, while the overall citation rates of Web resources has increased from 10.1% in 2001 to 19% in 2007, scholars generally do not view materials found on the Web of equal value with those that appear in peer reviewed publications. This despite the fact that the repository databases were largely produced by institutions and government agencies and made up about 50.5% of the Web resources that were cited (Zhang, 2011). The author suspects that
Building Value-added Services for Institutional Repositories (IRs)

scholars tend to cite a print version that appeared in a peer reviewed journal rather than its online version which they actually used. Hence, in addition to the barriers identified in other types of studies (self archiving practices, name confusion, and misperception), there still appears to be an initial hesitance on the part of faculty to accept information from the Web as equal to sources from peer reviewed journals.

**Institutional Repositories as Nodes in Scholarly Networks.**

The third area of research on IR development was undertaken, in part, as a response to low faculty interest and participation in IRs. This area of research also addresses issues that emerge as IRs mature in their content and services. Work of this type addresses how the initial role of IRs, as static repositories of digital intellectual assets, can be transformed. In this newer conceptualization, IRs move beyond a static role to become a dynamic node in scholarly networks where academic collaboration is fostered and where new research is reported. An earlier study that aimed to identify the strategies and conditions that influence and advance IR development involved an in-depth, micro-level analysis of IRs in three large research universities in the United States (Palmer, Teffeau & Newton, 2008). In-depth interviews were conducted with IR developers, liaison librarians and faculty members. The study concluded that there are three objectives related to IR development that were sought after by all three institutions: (1) improving problem-solving (i.e., managing digital assets and making them accessible), (2) enhancing collaboration, and (3) strengthening intellectual property management (Palmer, et.al., 2008). This study also emphasized IR support of faculty research “upstream” rather than focusing on the final products of scholarship.
A similar study was conducted using a comparative case study of five IRs in colleges and universities. (Yakel, Rieh, Markey, St. Jean & Yao, 2009). This study concluded that the success of an IR should be measured by both internal indicators (e.g. the quality and extent of content) as well as external impact factors such as how well the IR fulfills its goal in terms of service to the academic community. This study also suggested that the goal and mission of IRs must be broadened by “inserting the library into the researcher’s workflow” and, as noted earlier, IRs must support faculty research “upstream”. Further, Bankier and Smith (2010) argue that the IR’s mission is to support the business of the university in its broadest sense. Work along these lines is already taking place at the University of Rochester where a recent Webinar is dedicated to “reengineering IRs.” As will be explored in this paper, by adding services such as customized portals, export capabilities, and research data management services, the Rutgers University Community Repository (RUcore) can now be viewed as a powerful system that facilitates scholarly collaboration and communication. What follows below provides background information about the origins of RUcore and the activities that have addressed many of the issues discussed above.

**Background to the Development of RUCore**

Earlier case studies on IR development revealed different strategies for marketing IRs to university communities, especially strategies that would attract faculty participation in populating IR content (Jantz & Wilson, 2008). The Rutgers Libraries’ initial effort to develop a suitable strategy and to reach out to its research community began in 2000 with the creation of an ad hoc Scholarly Communication Steering Committee (SCSC). The goal was to lead a university wide dialogue to promote a better understanding of the evolving scholarly
communication milieu while exploring proactive approaches, such as encouraging Rutgers journal editors to consider moving their publications from high-priced commercial houses to lower-priced venues supported by professional societies (Dess & Wilson, 2003). In the initial year of 2000-2001, a three pronged action program was planned; the first was to hold a series of focus group luncheons for faculty to ascertain their attitudes and the extent of their knowledge about the emerging models of digital publishing and the impact on scholarly communication; the second was to develop an information brochure and a website highlighting the current troubled state of scholarly communication; and the third was to organize a symposium, national in scope, that would invite speakers from outside Rutgers. During these conversations and activities, Rutgers faculty were adamant that any future system of scholarly communication must be able to sustain the highest quality of scholarship, firmly grounded in and reinforced by a suitable system of refereeing. It was also stressed that any new system must be compatible with the faculty tenure and promotion process (Dess & Wilson, 2003).

The ad hoc committee that was created almost a decade ago at Rutgers has now become a standing committee of the library faculty governance structure. The initial website that began in 2001 continues to exist to this date under Scholarly Communication and Open Access: Research and Publication in Flux and highlights the issues and activities of the open access movement, author’s rights, as well as the content and services of RUcore. The 2009 symposium, jointly hosted by the Office of University Vice-President for Academic Affairs and the University Libraries invited national leaders in this area such as David Shulenburger and Karla Hahn. Attending Rutgers faculty members were asked to submit their areas of interest regarding
RUcore, the Rutgers institutional repository. A majority expressed interest in learning more about RUcore and author’s rights. A few expressed interest in how to manage datasets.

Liaison librarians are considered an essential part of the human infrastructure that helps sustain and populate IR content. A special brochure, “Introducing RUcore” was prepared by the University Libraries’ marketing team for distribution by liaison librarians to their respective academic units. Selected liaison librarians attended faculty departmental meetings in order to share information regarding RUcore development and to listen to faculty concerns and comments. Rutgers Libraries is in the process of re-defining the role of liaison librarians on campus, emphasizing their mission to “engage” the academic community by looking “outward” and to “listen” to faculty concerns regarding RUcore, among other matters.

**RUcore: A Flexible Platform and Portal Architecture**

From the outset, Rutgers University Libraries (RUL) viewed the IR in its broadest sense – a place to capture the output of the university. This broad goal provided the motivation and impetus to construct a flexible platform that could evolve with the changing needs of users and the rapidly evolving technological infrastructure. A flexible digital repository should allow users to store all types of digital objects along with the appropriate descriptive and administrative information. A digital object might be an electronic journal article, a digitized image, digital video, a dissertation, or a complete book in digital form. Further, the repository architects should not make too many assumptions about the types of users and applications that might use the repository. The basic set of assumptions includes fundamental concepts to keep in front of us when developing an institutional repository. These concepts include sustainability – is the architecture such that it can be easily modified to adapt to new user requirements and
Building Value-added Services for Institutional Repositories (IRs)

new technology? – and scalability – will the architecture accommodate a growing number of objects (into the millions) and an increasing size of a single object (into multiple gigabytes)? These fundamental concepts have enabled the generic RUcore architecture to remain constant while relatively “thin” software applications provide the necessary unique features for each end user service.

Figure 1 provides a high level view of the conceptual architecture of RUcore – Rutgers’ institutional repository. The IR obtains basic repository functions (e.g. ingest of objects) from FEDORA (Staples, Wayland, & Payette, 2003) – a flexible library platform. Referring to Figure 1, major components developed by Rutgers University Libraries include the Workflow Management System (WMS) and services that focus on the user, collections, and preservation of the digital object. For sustainability of the IR infrastructure, all components are structured modularly, enabling ease of upgrading or replacing software for specific services. Of particular note is the portal architecture which enables the IR manager to easily configure content, format, or user-specific portals and attach RUcore services such as a search engine and statistics reporting. Given the multiplicity of access points in today’s complex infrastructure, export services insure that IR content is also available from the library OPAC, Google Scholar, and via various harvesting methods such as OAI-PMH.
Early Content Development

Although the architecture of RUcore embodied many of the principles that are important to institutional repositories, early content development was not exclusively from that of a single institution. Three major portals characterize both the flexibility and content diversity that have been a hallmark of RUcore since its inception. The New Jersey Digital Highway (Rutgers University Libraries, 2011) – an Institute of Museum and Library Services (IMLS) grant-funded project – is an immersive, user-centered portal that continues to grow through collaborations with the state’s cultural heritage organizations. As a first major project, the New Jersey Digital Highway (NJDH) verified many of the important initial objectives for providing access to and preserving many different format and subject collections. The portal architecture, added more recently, enabled institutions to maintain local ownership and branding while offering shared services.
Building Value-added Services for Institutional Repositories (IRs)

access to their unique resources. A collection or a single resource can be a member of many different portals. A typical example includes the Peter Still Papers – a saga of Peter Still and his family who fled slavery in Maryland for freedom in New Jersey. This collection appears as part of NJDH and is also accessible through the University Special Collections and Archives portal.

A second major grant funded project focused on providing statewide access to commercial and public videos. NJVid (NJEDge, 2011) was funded by IMLS and provides a statewide digital video portal and central storage repository. This project demonstrated important technological capabilities including the streaming of large video files whose archival masters could range up to 20 GB. Authorization and authentication services using Shibboleth and XACML were proven out on the NJVid project and have significant potential for supporting multi-university grant funded projects within the institutional repository. Given the flexible architecture, it is noteworthy that the RUcore software was easily ported to a separate server and data center for the NJVid project.

As a final step in early content development, the migration of a project based on legacy technology has demonstrated the flexibility that is required of the IR architecture. The New Jersey Environmental Digital Library (NJEDL) – a repository of New Jersey environmental research and reports – has been funded for many years by the New Jersey Department of Environmental Protection. Using metadata mapping and WMS batch capabilities, this valuable resource of grey literature consisting of over 5000 digital objects has been easily moved to RUcore with an associated NJEDL custom portal.

The Institutional Repository and Value Added Services
Building Value-added Services for Institutional Repositories (IRs)

Viewing the IR from the architectural or platform perspective is not particularly inspiring or revealing. As Bankier and Perciali (2008) have indicated, the IR community has struggled to represent IRs as vibrant and responsive – too often leaving the impression that the IR is a “vault where papers go to die after the exciting work of creation is done.” Within RUcore, a key objective has been the creation of a “research place” by focusing on three main components of the university scholarly environment: faculty research, dissertations, and interactive collaborative spaces. These objectives have been accomplished by librarians serving in new liaison roles and the development of custom services to support the scholarly environment.

**Faculty Services (Deposit, Grant Support, Author Portals).**

Faculty want to maximize access to their scholarly output while minimizing their involvement in the mechanical processes that provide enhanced access. In RUcore, faculty can create their own collections, deposit articles through an easy to use web-based interface, and attach their collections to a personal webpage. To minimize personal involvement in the mechanics of submission, an academic department can identify a trusted proxy agent who can deposit articles for all of the researchers in the department.

Faculty frequently create an online identity and profile via personal web pages. The RUcore portal concept allows a faculty member to attach their articles in the repository to their own personal website with associated services such as full text searching, preservation support, and statistics reporting. Through the portal concept, a faculty member can maintain control of their online presence while reaping the benefits of a powerful IR platform. Since much research is conducted with grant funds, grant assistance in various forms is also being provided. Recent regulations require NIH grant recipients to make their articles accessible in PubMed Central.
Building Value-added Services for Institutional Repositories (IRs)

With RUcore’s submission service and the assistance of the repository manager, submission to PubMed Central is greatly simplified while also providing another access point for the author’s article.

Special features of export services significantly enhance access to research. A Google Scholar sitemap is updated nightly and contains all published articles and dissertations, insuring that these important resources are available via Google Scholar. For dissertations and other resources in the repository that require item-level access, the repository manager can export MARC records to the library OPAC. In addition, any specific collection or set of collections can be harvested via the OAI-PMH protocol in order to provide access to RUcore content from other repositories.

**ETDs (Theses and Dissertations).**

A significant amount of the research generated by universities is found in student theses and dissertations. ETD repositories benefit students and universities by enhancing graduate education, expanding graduate research, increasing a university's visibility, and informing students, faculty, and the administration about the benefits of digital technology (Yiotis, 2008).

At Rutgers, there are seven graduate schools that require the submission of a thesis or dissertation. In addition to the basic repository functions including access, searching, and preservation, the RUcore ETD system (Mills, 2010) is customized to support graduate school processes and policies that are required of the student as part of the submission process. Additional micro-services enhance the value of dissertations for both students and academic departments. Many students expect to publish their dissertations in book form and do not want to grant public access for several years. As part of the submission process, a student can
request an embargo period of up to two years. After the embargo period expires, the dissertation becomes automatically available without human intervention. Using a dynamic ETD portal, the repository manager can create a special portal for those dissertations from a specific department or school. This portal can be placed on the website for the academic department and is automatically updated whenever a new dissertation is deposited in RUcore. Rutgers University Libraries launched the ETD service in late 2007; at this juncture there are over 2000 dissertations in the repository – most of which are available to the public.

**Collaborative Spaces.**

An attractive and emerging feature of IRs is the collaborative workspace where discipline-specific tools, data, and metadata are available to faculty. RUcore developers working in collaboration with faculty have implemented the VMCAnalytic (Agnew et al, 2010) and integrated the capability with the RUcore infrastructure. The VMCAnalytic provides an interactive collaborative space where education faculty researchers and practicing teachers can analyze and annotate videos to improve the individual teaching of mathematics. While customized to a specific discipline and format, the VMCAnalytic is unique in drawing upon the RUcore generic services such as search, retrieval, access control, and metadata. It is expected that the VMC model can be readily extended to science data.

**Science Data.**

Much of the research underlying science publications is in the form of data – research that is infrequently published. Science is international and global cooperation is imperative to spur innovation and economic growth. Scientists and granting agencies are addressing the need to increase access to research data. The U.S. National Science Foundation (NSF) is now requiring a
data management plan to ensure that science data is usable, broadly available and preserved for long term access. RUresearch, a data portal of RUcore\textsuperscript{vii}, meets all the requirements identified in the NSF guidelines for data management, including preservation, sustainability and metadata for both data sets and accompanying research products. In this emerging service, library data specialists provide assistance and advice to researchers who want to design a data strategy and management plan that is NSF compliant. Although much of the RUcore infrastructure is mature and has been heavily used, science data offers new challenges. In a recent survey (Science staff, 2011), respondents indicated that the lack of common metadata and established archives represent major obstacles to archiving and sharing data. The size of science datasets also represents significant access and storage challenges. The response to the Science survey indicated that 20% of the respondents regularly use or analyze datasets that exceed 100 gigabytes. However, almost 50% of the respondents work with data sets that are one gigabyte or less – a sizable market that can be readily addressed by research libraries. The data service not only provides a place to archive data but also provides users with the ability to locate, identify, and cite research datasets with confidence.

**Summary and Conclusion**

This paper has suggested that institutional repositories should be conceived of as an essential infrastructure of the 21\textsuperscript{st} century university and one that has the capacity to grow and evolve in support of the complex enterprise of scholarly communication and publishing. Those individuals who were involved in early IR development could not have anticipated the variety of content to be deposited or the impact of the rapidly evolving technological infrastructure.

Beyond the basic capabilities of access, archiving, and preservation, repositories must offer
Building Value-added Services for Institutional Repositories (IRs)

integrated services and tools that support research, communication, and publishing. Major components that will make it possible to achieve this comprehensive mission are a flexible repository architecture, new library liaison roles, and integrated collaborative workspaces where discipline-based researchers can take advantage of value added services and tools. This paper has explored a number of IR services at Rutgers that, viewed together, have resulted in significant value-added services while also suggesting a different way for libraries to conceive of and market the IR. It is expected that this renewed focus will not only lead to a successful IR, but also result in more compelling and relevant roles for the research library.

Acknowledgement

It is impossible to acknowledge all the contributors to RUcore. Within Rutgers University Libraries, many librarians and staff from technical and public services units have contributed to the development, ongoing support, and marketing of the repository. Beyond the Libraries, the collaboration with Rutgers’ Office of Information Technology, especially in the area of services for authentication and authorization, must also be acknowledged. Finally, the continuing development and support of FEDORA – open source software for managing and preserving digital content – has been a critical and essential element for the success of RUcore.
References


Building Value-added Services for Institutional Repositories (IRs)


This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License. To view a copy of this license, visit [http://creativecommons.org/licenses/by-nc-sa/3.0/us/](http://creativecommons.org/licenses/by-nc-sa/3.0/us/).

---

i RUcore is available at [http://rucore.libraries.rutgers.edu/](http://rucore.libraries.rutgers.edu/).

ii The Workflow Management System software developed by Rutgers University Libraries is offered as open source to other interested parties. See [http://rucore.libraries.rutgers.edu/open/](http://rucore.libraries.rutgers.edu/open/).

iii The Shibboleth System is a standards based, open source software package for web single sign-on across or within organizational boundaries.

iv XACML (eXtensible Access Control Markup Language) is an XML access control policy language and a processing model for the interpretation of policies.

v The OpenETD software developed by Rutgers University Libraries is offered as open source to other interested parties. See [http://rucore.libraries.rutgers.edu/open/](http://rucore.libraries.rutgers.edu/open/).

vi The VMC website is available at [http://videomosaic.org/](http://videomosaic.org/).

vii The RUresearch Data Portal is available at [http://rucore.libraries.rutgers.edu/research/](http://rucore.libraries.rutgers.edu/research/).