TCNJ Library Institutional Repository
Proposal

Prepared by:
Library IR Working Group

TCNJ Library
December 2013
Executive Summary

An institutional repository (IR) is a digital service that an academic institution can offer to the members of its community and provides an online warehouse that is designed to archive, manage, organize, preserve, and showcase the intellectual output of students, faculty, and staff to a broader audience. Developments in digital technologies have affected many aspects of the academic enterprise, especially the way scholarly communication is being carried out. An IR supports the dissemination of knowledge by curating locally produced digital materials and creating institutionally unique digital collections to support campus scholarship, teaching, and learning. This document provides background information on the purposes and benefits of a repository service for The College of New Jersey community. This document also illustrates the IR project implementation plan recommended by the Library IR Working Group.

The Library IR Working Group recommends that the Library take action to implement an IR as a new service for TCNJ students, faculty, and staff. The key benefits that the IR service will bring to the TCNJ community include providing an intuitive mechanism to preserve, organize, and make accessible the digital scholarship TCNJ faculty and students produce, showcasing the accomplishments of the College’s scholarly community for a broader public. These benefits support the College’s strategic map that, in the transformed curriculum environment, places a strong emphasis on deep learning and intensive faculty-student collaboration. It is worth noting that among TCNJ’s seventeen peer institutions (“Comparator Schools”), thirteen institutions have implemented their IR service. The other four institutions are in the process of implementing one.

The IR Working Group further recommends that the Library focus on the development and implementation of a repository service using IR+, the open-source IR platform chosen for the pilot project conducted in 2011. This document also outlines specific actions and resources recommended for the campus community and the Library to develop the campus infrastructure for the successful implementation of a TCNJ institutional repository.
Purpose and Scope of the Document

We recommend that the Library move forward with implementing a digital institutional repository as a new service for TCNJ students, faculty, and staff. The purpose of this document is to provide background information about the purposes and benefits of a repository service for the College community. The document also outlines the tasks and support needed to create campus infrastructure for collecting, managing, and providing access to the intellectual output of the College in the digital environment.

Institutional Repositories: Definition

An institutional repository (IR) is a digital repository service that an academic institution can offer to the members of its community and provides an online warehouse that is designed to archive, manage, organize, preserve, and showcase the intellectual output of students, faculty, and staff to a broader audience. An IR supports the dissemination of the knowledge produced by the community by curating scholarly publishing, digital publishing, and the creation of unique, local digital collections to support campus scholarship, teaching, and learning.

Institutional Repository Development: General Background

Since the early 2000s, an increasing number of academic institutions have established IRs to collect and provide open access to their local digital content produced by their faculty, students, and staff. While initial growing pains prompted some questions about the viability of this new library service, IRs have matured and grown significantly in the past few years, with many repositories recording downloads in the millions.

Although the IR movement started with large research universities, the number of smaller institutions providing IR service has been steadily increasing over the years. It is also worth noting that all peer institutions for TCNJ (“Comparator Schools”) have implemented or plan to implement IRs to archive, organize, preserve, and provide permanent access to the digital content produced at their institutions (see Appendix for a list of “Comparator Schools,” with links to their IRs).

Institutional Repositories: Key Benefits
There is a compelling need to provide long-term storage and access to a number of categories of digital materials produced at TCNJ. The following list describes the major benefits that an IR service will bring to the TCNJ community:

*Providing open access and scholarly communication in the digital age (Link to TCNJ Strategic Map: B1)*

An IR will enable TCNJ faculty to promote their research and scholarship to a broader audience via a single digital portal. Like most academic institutions, the College uses a course management system to organize and deliver teaching materials. But there is no similar institutional infrastructure for faculty research materials. Some faculty members create personal web pages to distribute their scholarship in digital form. Others may leave their research materials scattered and largely hidden in commercial subscription databases. With a repository service such as an IR in place, the campus scholarly research will be centralized and better organized, and therefore more visible to a wider public and findable beyond the silos of subscription databases. An IR also offers increased support for TCNJ faculty scholarship, particularly as federal grant agencies increasingly mandate that researchers develop plans to manage and make freely available the published results and digital data resulting from funded research.

*Focus on student research and accomplishments (Link to TCNJ Strategic Map: A)*

An IR provides many benefits to current and prospective students, their families, and alumni. Undergraduate research is at the core of our unique educational experience at TCNJ. TCNJ students take pride in their active participation in academic research. An IR offers a centralized, institutional mechanism for highlighting the excellent work of TCNJ students and their learning outcomes—a very important service in the transformed curriculum that places a strong emphasis on deep learning and intensive faculty-student collaborative scholarly and creative activity. An IR will also be a great recruitment tool. Prospective students and their families will better understand the signature culture of the College by viewing the content in the IR and learn what students can expect to accomplish at TCNJ. Furthermore, archiving and providing permanent access to student-produced digital materials will help TCNJ graduates feel that they are an ongoing part of the College and will create stronger ties between alumni and their alma mater.

*Institutional identity and advancement & public accountability (Link to TCNJ Strategic Map: A, B, C)*

An IR collates and showcases the intellectual output of TCNJ students, faculty, and staff in one official portal, and thus helps to advance the institutional goals and mission of the College by highlighting TCNJ’s “excellence in teaching, creativity, scholarship, and citizenship,” as affirmed in the College’s mission statement. By exposing a wider public to the College’s intellectual output, the College’s digital repository will promote an institutional brand and
contribute to TCNJ’s overall reputation and prestige. An official digital repository will also enable the College to show a “return on investment” and enhance accountability to the general public, as it serves as an effective mechanism for the College to provide centralized access to the results of publicly supported faculty research and scholarship conducted on the campus. Furthermore, an IR can provide a mechanism to integrate, preserve, and provide access to historical materials associated with the College and facilitate the visibility of the College’s unique institutional history throughout the TCNJ community and beyond.

Library’s Role in Institutional Repositories

Responsibility for an operational IR is typically given to libraries because they have the required professional expertise in information organization and collection management. An IR will build on the traditional role of the Library as it provides new services to steward, organize, and manage collections of valuable digital content produced by the College community. Development of an IR will be aligned with the Library’s mission to “provide high-quality information resources, expertise and a learning environment that enhance the search for knowledge and understanding.”

Institutional Repository Development at TCNJ Library

A team of three librarians (Yuji Tosaka, Cataloging/Metadata Librarian, Cathy Weng, Head of Cataloging, and Sharon Whitfield, former Emerging Technologies Librarian) applied for and was awarded a MUSE project grant in 2011 to work with two computer science students and establish an institutional repository as a pilot project. For this pilot project, the MUSE team evaluated major open source platforms available for a local “proof of concept” implementation (i.e., DSpace, Greenstone, IR+, and Fedora), and had discussions with the NJVid (New Jersey Video Repository) staff about the feasibility of having a sandbox environment created as a hosted pilot repository. The hosted solution was not found to be workable due to NJVid’s current focus as a statewide digital video portal. Although highly flexible and extensible, Fedora was found overall to be technically too involved, as it would demand dedicated programming staff to customize, launch, and maintain a functioning IR. After evaluating the other open source solutions, the MUSE project team determined that IR+, an open-source system developed and used by the University of Rochester (https://urresearch.rochester.edu/), provided a cost-effective and yet feature-rich solution that met our needs for the pilot IR without extensive modifications exceeding the available local technical expertise. In terms of server support, while the IT Department offered to make a virtual server available for the pilot study, the MUSE team used an old Dell PowerEdge SC1425 server for ease of initial installation, testing, and experimentation during the pilot phase.
By the end of the eight-week summer program, the MUSE project team implemented a pilot IR running on IR+ and successfully tested the feasibility of developing an operational repository with limited staff time and resources. (The test repository is tentatively named TCNJ Digital Scholar, accessible within the campus network at http://libdev.tcnj.edu:8080/ir_plus/). Valerie Tucci, Physical Sciences/Engineering Librarian, worked with the MUSE team and helped to recruit content from the Chemistry Department. Some members of the Library faculty also agreed to deposit their research into the pilot IR. After the conclusion of the MUSE project, the same team continued to work and further develop expertise needed to support a permanent IR service, including formalizing the metadata application profile, establishing file submission processes, staff workflows, name authority control, rights management. As of November 2013, a total of 127 publications (39 from the Library and 88 from the Chemistry Department) have been deposited into the pilot repository.

Due to a change in personnel and the need for further developing the pilot IR to become a permanent library service, two additional library faculty members joined the original MUSE project team in the early fall of 2013 and formed a new working group. The current IR working group consists of Terrence Bennett, Business/Economics Librarian, Valerie Tucci, Physical Sciences/Engineering Librarian, Yuji Tosaka, and Cathy Weng.

IR Project Plan

We recommend that the Library move forward with full implementation of a new IR service for TCNJ students, faculty. The initial MUSE pilot team has identified several specific tasks within and outside the Library that will be essential to supporting the implementation of a digital repository service for the TCNJ community. The section below outlines what we will need to launch an operational IR as a new library service. It will be necessary for the Library to continue to assess additional resources such as technical or staffing needs as it builds up collections of digital content that are deemed to have significant institutional value. For long-term planning, in the ever-changing technology environment, regular environmental scans for better hardware and software solutions will need to be established. Shifts in user needs and preferences will also be monitored to ensure the relevance, growth, and sustainability of the TCNJ IR.

Campus Support

1. Technical support – A new server solution will be needed to host the proposed repository as a permanent library service. The test server used for the MUSE pilot is outdated and not suitable for long-term repository use. There are a few technical options for implementing the proposed repository service. One option is for the Library to make a formal request to the IT Department to create a virtual server for hosting and administering a production system.
Another option is for the Library to acquire and manage a physical server under its local control. A physical server also will be useful for ongoing repository development. Below is an outline of the suggested server specifications for an operational repository should the third option be chosen. Based on our communication with the IR+ developer at the University of Rochester, we believe that the suggested specifications and estimated costs below are scalable to meet the needs of a growing repository.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Low-end</th>
<th>Medium-end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server</td>
<td>Desktop or server</td>
<td>Server</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Core2 Duo</td>
<td>Intel Core i7, regular desktop processor</td>
</tr>
<tr>
<td>RAM</td>
<td>4 GB</td>
<td>6 GB</td>
</tr>
<tr>
<td>Hard disk type</td>
<td>5400 rpm SATA</td>
<td>7200 SATA or SCSI</td>
</tr>
<tr>
<td>Hard disk space</td>
<td>1TB</td>
<td>1TB</td>
</tr>
<tr>
<td>RAID5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Network</td>
<td>10/100 Mbit &amp; Wifi</td>
<td>2 x 100 Mbit</td>
</tr>
<tr>
<td>Backup media</td>
<td>External HDD</td>
<td>External HDD</td>
</tr>
<tr>
<td>UPS</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Networking</td>
<td>WiFi or wired UTP</td>
<td>Shielded UTP</td>
</tr>
<tr>
<td>Internet connection</td>
<td>2 MBit or Intranet if only lower bandwidth is available</td>
<td>5 MBit</td>
</tr>
<tr>
<td>Other</td>
<td>None</td>
<td>Standard (ufw)</td>
</tr>
<tr>
<td>Estimated cost</td>
<td>$2,800</td>
<td>$3,100</td>
</tr>
</tbody>
</table>

While the technical options above are based on the adoption of an open-source IR platform, a full institutional repository can also be implemented using a commercial IR solution, such as Digital Commons (http://digitalcommons.bepress.com/) powered by Bepress. The main benefit, especially if the repository is hosted by an IR vendor, is that all the technical aspects of the repository will be supported and maintained by the vendor, thus eliminating the burden of building local technical expertise for IR server administration. Based on an informal inquiry for an estimate from Bepress, the commercial solution will likely cost $15,000 or more annually for a small repository, which is higher than the cost of acquiring an open source solution. Should this solution be adopted, detailed plans and RFPs will need to be pursued.

A third option is to acquire a cloud server space to operate the new IR service. The cloud server can be used to connect to the existing campus IT infrastructure for server administration and public access. Based on an informal inquiry for an estimate from Amazon, the cloud server solution will likely cost $1,500–$2,000 annually to run a full digital repository for TCNJ.
The initial MUSE pilot team recommends that TCNJ adopt a cloud server solution for implementing an open-source IR platform. It is reliable, secure, and inexpensive, thus making it a suitable hardware solution. This option will provide a secure, flexible, and scalable computing, storage, and archiving platform. There will be no issue of server hardware getting outdated and needing regular upgrades as an initially small IR service starts to grow at TCNJ. However, the Library should consult and seek recommendations from the IT Department regarding the selection of a new server solution to support the proposed IR service.

2. Academic community support – To implement an IR as a permanent library service, we will need campus-wide support from the academic community (such as Academic Affairs, Faculty Senate, etc.). In light of the wider benefits of a repository service, our ultimate goal is to secure faculty buy-in and participation and promote open access as an institutional best practice at the College. Specifically, the focus should be on developing official guidelines to encourage faculty members to grant the College a non-exclusive license to distribute their publications and other forms of scholarly/professional/creative output (within their existing copyright restrictions) so that they can be permanently archived, stored, and made openly accessible.

Library Support

1. System migration – The project will require moving the pilot repository system from the current server to a new server to scale up the pilot experiment to a production system. The timeline for this task may depend to a large extent on what technical support can be provided by the IT Department as listed in Campus Support (1) above.

Based on the post-pilot IR evaluation, the initial MUSE pilot team recommends that the Library develop and implement a digital repository service using IR+, the open-source IR platform chosen for the pilot project. The pilot IR experience has shown that the software fully meets the essential requirements as the vehicle for TCNJ’s digital repository project. Developed at the University of Rochester\(^1\), IR+ is designed to provide an easily managed IR platform that accommodates a wide range of digital materials. While its implementation has been limited mostly to Rochester, it has been recognized that no other open source software currently can be installed out of the box with a clean, next-generation interface featuring faceted navigation and results-narrowing. IR+ has been tested to provide a scalable platform that effectively supports both large- and small-production repository systems, thus making it an ideal solution for institutions like TCNJ Library with limited IT staff and resources. IR+ also complies with international data protocols (Dublin Core, OAI-PMH, and the Handle

\(^1\) After adopting and implementing DSpace for a few years, the University of Rochester migrated from DSpace to IR+ to better meet institutional needs.
System), providing an effective mechanism for open data exchange and persistent, wider access. IR+ includes other useful features such as personal name management (especially useful for authors who publish under different names), usage statistics, researcher profile pages where users can collocate and showcase all their work, user workspace and collaboration, and version control.

The current IR proposal is independent of any digital initiatives that may be proposed by other working groups, such as the digital archives working group. While it is not inconceivable that the proposed IR service may be combined with other digital library initiatives because of their logical synergy, given the level of progress already made for the current IR project, we propose to proceed independently with implementing the IR service, building on the previous work that has been done and transitioning the current pilot IR into a production system. There will be no system dependencies necessary to run the proposed IR platform that the Library will create. The proposed IR system will be a standalone Web application, which will run on a standard Web server with a backend SQL server. This means that there is no need for the Library to wait to implement multiple digital initiatives concurrently. In the future, should the Library choose to adopt a new single software for the IR and digital archives, for example, it will be possible to easily migrate the IR content to a new platform using the batch export mechanism in IR+.

2. System maintenance – The project will require our continuing participation in the open source institutional repository community (currently IR+ community) and performing regular updates and enhancement of the repository software. This task can be accomplished on an ongoing basis with limited technical staff time in the Library unless the College decides to adopt open-source repository systems like Fedora that will require substantial investment of IT staff time and programming expertise for technical implementation and infrastructure support.

3. Metadata service – The project will require local guidelines and mechanisms to support metadata creation and ensure resource discovery, delivery, and access in the production repository system. Essential tasks include developing a local metadata application profile and authority control, routine metadata entry, and database maintenance mechanisms. Preliminary policies, procedures, and workflows for these tasks have been developed by the Cataloging Department. The IR metadata service will receive ongoing support from the Cataloging Department. We estimate that a full repository service will require additional paraprofessional staff (0.5 FTE) dedicated to metadata service. This additional staff is needed when the IR becomes a fully functioning production system. The need for additional staff will be periodically assessed
4. Copyright management – The project will require that copyright clearance policies and procedures be established and carried out to assure TCNJ faculty and students that their scholarly products are legally deposited and in compliance with any existing copyright restrictions. Preliminary policies, procedures, and workflows for these tasks have been developed by the Cataloging Department. The copyright management tasks need to be handled by a designated trained staff with professional oversight. We estimate that a full repository service will require additional paraprofessional staff (0.5 FTE) dedicated to copyrights management. This additional staff is needed when the IR service becomes a fully functioning production system. The professional oversight of this task can remain in the Cataloging Department or can be transferred to another library unit.

5. Library outreach and content recruitment – The IR project will not succeed without key content deposited into the IR, i.e., scholarly and creative works, etc. from students and faculty. It is crucial to have subject librarians’ collaboration and support to reach out to their individual departments to recruit the IR content. The Library may develop a set of common working principles and policies for content selection and workflow, in conjunction with campus faculty. Alternatively, this task may be best left to individual librarians, who will identify a baseline for their academic programs based on evaluation of their unique disciplinary needs.

6. Space requirements – Additional staff dedicated to the new IR project can be housed in the Cataloging Department, where there are two open cubicles available. One set of computer equipment will be needed, though, because only one cubicle is currently equipped with a computer workstation. Should the local hardware option be chosen for the proposed IR system, then additional space requirements in the Library would have to be established.

Repository Content

1. Faculty and staff work
The repository will provide an opportunity for faculty recognition and maximize research impact. It will archive, preserve, and distribute outcomes of faculty and staff scholarly/creative/professional activities. These include pre-prints (initial, submitted version of the article, e.g., prior to peer review – easiest to deposit due to no copyright restrictions by most publishers), post-prints (final, accepted, pre-pub version of the article, e.g., revised after peer review), and published version of the article, as well as working papers, research reports, local conferences and symposia, data sets, etc. that emanate from scholars and centers at the College, and multimedia content of creative works and performances.²

² It is our understanding that TCNJ uses Digital Measures to track faculty’s scholarly activities. However the proposed repository service will not duplicate Digital Measures functionalities and will serve a very different purpose. Digital
2. Student work
   The predominant types of materials in institutional repositories at smaller institutions are various categories of student work. These include undergraduate honors papers and award-winning papers, faculty-mentored research, individual capstone projects, poster presentations, essays, prose and poetry, student-created curriculum guides, student performances and creative works, and student portfolios. Given the College’s increasing focus on undergraduate research, to the point where it is now a mainstay of the curriculum in numerous disciplines, this is where the IR can have the greatest impact for this institution, as most student works are now lost permanently without institutional commitments to collect and preserve them for public access. The repository will offer an ideal institutional service for showcasing the large amount of undergraduate work that is being produced at TCNJ, including, but not limited to, MUSE and Celebration of Student Achievement presentations and capstone projects. (To showcase the best student work, the repository may need to develop an appropriate content selection policy, one that will only accept student work sponsored by faculty members in some form.)

3. College documents, archival materials, and historical publications
   The repository will allow the institution to preserve its institutional memory by collecting digital materials produced by various campus units, such as annual reports, newsletters, and College addresses and conference proceedings. Also, an important target area for a fully developed repository service will be digitized archival materials and special collections relating to the College’s history. The repository could be expanded to accommodate such digitized content, including historical photographs, The Signal, and other student publications. This proposal does not address any archival components because a separate digital archives project is currently underway in the Library for preservation and access. A separate proposal should be emanating from the digital archives working group.

Appendix: List of TCNJ Comparator Schools with Links to their IRs

Boston College: http://www.bc.edu/libraries/collections/eScholarshipHome/
Bucknell University: http://digitalcommons.bucknell.edu/
California Polytechnic State University, San Luis Obispo: http://digitalcommons.calpoly.edu/

Measures is a faculty tracking system for academic administration. Faculty members enter information on their teaching, scholarship, and service activities, which will be used for annual review, reappointment and promotion, accreditation, and reports to external agencies, etc. While publications can be attached to Digital Measures profiles, they are intended for internal use due to privacy and copyright reasons, among others. By contrast, a main purpose of an IR is to support the public dissemination of faculty work. (For that reason, exact copies of published articles often cannot be deposited into an IR without violating publisher restrictions.) Moreover, Digital Measures is not designed to fulfill the basic IR function of showcasing other types of materials, such as student work and institutional documents and historical materials.
Carnegie Mellon University: http://repository.cmu.edu/
Chapman University: http://www1.chapman.edu/library/archives/ir.html
College of William and Mary: https://digitalarchive.wm.edu/
Fairfield University: http://digitalcommons.fairfield.edu/
Loyola University (Maryland): http://www.loyola.edu/library/TechSupport/index.htm.htm
Miami University (Ohio): http://sc.lib.muohio.edu/
Saint Joseph’s University: http://scholarship.sju.edu/
SUNY Geneseo: http://dspace.sunyconnect.suny.edu/
University of Pennsylvania: http://repository.upenn.edu/
Villanova University: http://digital.library.villanova.edu/Collection/vudl:171664

IR Development Underway
  James Madison University
  Lehigh University
  Santa Clara University
  Truman State University