Published in final edited form as:

Med Ref Serv Q. 2018 ; 37(1): 31–42. doi:10.1080/02763869.2018.1404386.

DigitalHub: A Repository Focused on the Future

Violeta Ilik, MAM, MLIS,

Head of Cataloging and Metadata Services, Stony Brook University Libraries, 100 Nicolls Road, W2510 Frank Melville Jr. Library, Stony Brook, NY 11794. At the time of the poster presentation, she was Head of Digital Systems and Collection Services, Galter Health Sciences Library, Northwestern University.

Piotr Hebal, MIS [Senior Developer],

Galter Health Sciences Library, 303 E. Chicago Avenue Ward 2-212, Chicago, IL 60611.

Anton Olson, MALS, MS [Cataloging Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Susan Wishnetsky, MLIS [Assistant Collections Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Joelen Pastva, MA, MLIS [Metadata Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Ramune K. Kubilius, MALS [Collection Development/Special Projects Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Jonathan Shank, MLIS [Acquisitions and Electronic Resources Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Karen Gutzman, MSLS, MA [Impact and Evaluation Librarian],

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Margarita Chung [Administrative Director], and

Galter Health Sciences Library, 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Kristi Holmes, Ph.D. [Director]

Galter Health Sciences Library and Associate Professor of Preventive Medicine (Health and Biomedical Informatics), 303 E. Chicago Avenue, Ward 2-212, Chicago, IL 60611.

Abstract

The DigitalHub scholarly repository was developed and launched at the Galter Health Sciences Library for the Feinberg School of Medicine and the greater Northwestern Medicine community. The repository was designed to allow scholars the ability to create, share, and preserve a range of citable digital outputs. This paper traces the evolution of DigitalHub development and engagement activities, highlighting project challenges, innovations, success stories, and the team-based approach that was employed to successfully achieve project goals.

Keywords

Controlled vocabularies; DigitalHub; institutional repositories; interoperability; linked data; metadata; open access; scholarly communications

INTRODUCTION

The Feinberg School of Medicine (Feinberg) has a forward-looking vision to transform the library into an information hub where library and informatics services help to advance patient care, research, and education. To achieve this vision, campus leadership integrated the Galter Health Sciences Library & Learning Center (Galter Library) with the Northwestern University Clinical and Translational Sciences Institute (NUCATS) in July 2013. This unique relationship has developed into a strong partnership, with meaningful benefits in terms of how Galter Library is able to rise to meet the vision to improve service delivery and integration with the academic, research, and clinical enterprises.

The creation and establishment of a digital repository was viewed as a critical step to enable the discovery, open representation, and preservation of scholarly outputs and outcomes. The library was well-positioned to undertake this project given its role as an information hub on campus and its desire to provide new services to the Feinberg community. In early 2015, a small team was formed at Galter Library, consisting of the software application developer and the head of Digital Systems and Collection Services. This small, development-focused team was later expanded to a cross-library workgroup, the Digital Initiatives Working Group (DIWG), to achieve successful and sustainable implementation and adoption of the DigitalHub repository, as well as accomplish other digital projects at Galter Library. This paper traces the evolution of DigitalHub development and engagement activities, highlighting project challenges, innovations, success stories, and the team-based approach that was employed to successfully achieve project goals.

BACKGROUND AND LITERATURE REVIEW

To better contextualize the repository project, recent scholarship in the field of institutional and digital repositories was assessed. This was useful to define the scope and aims of the repository, as well as to inform the team about possible obstacles that might be encountered with populating the repository and achieving institutional adoption and acceptance.

Institutional repositories are developed for the management of an institution's digital assets, which can include digitized and born-digital text, audio, images, multimedia, and datasets supporting instruction, research, and study.¹ The content of a repository is typically scholarly in nature, with journal articles, conference papers, and theses and dissertations comprising the majority of managed resource types.^{2, 3} In his perhaps most widely cited and embraced definition, Lynch views an institutional repository not as a fixed set of hardware and software, but rather as a set of services offered to an institution to organize, disseminate, provide access to, and preserve digital materials.⁴ Repositories can increase an institution's visibility, aid in open access deposit requirements and funding mandates, improve publication dissemination, provide a central location for collecting and reporting research

outputs, ensure long-term preservation, and demonstrate research impact through usage metrics.^{2, 5–7} Many university departments are ill-equipped to manage their digital assets, with few formal mechanisms in place for preserving outputs or capturing associated metadata.¹

Repositories also provide a venue for the dissemination of teaching materials, grey literature, conference proceedings, presentations, and other forms of non-traditional scholarship often hidden from the scholarly community, but which can be valuable for educators, researchers, and students.^{2, 8} Clifford Lynch states that, the point of IRs, is not to disrupt the existing scholarly publishing system, but to allow it to be expanded and diversified by providing access and stewardship for material that mainly falls outside of the traditional scholarly publishing system as it exists today – both material created by faculty and material created by the institution, or departments or other groups within it.⁹

This aligns with the Confederation of Open Access Repositories (COAR) vision that repositories need to provide access to published articles as well as a broad range of artifacts beyond traditional publications such as datasets, pre-prints, working papers, images, software, and so on.⁵

Although the potential benefits of an institutional repository are aligned with institutional priorities, challenges need to be acknowledged and considered as part of the repository development and implementation process. Perhaps the biggest challenge faced by institutions seeking to implement a repository is integration with and adoption by the institution's larger community. Self-archiving by faculty is viewed as highly important to the success of a repository, but faculty submission rates are often lower than expected.¹⁰ Libraries, as the most common managers of institutional repositories, have long advocated for the open flow of information and dissemination of ideas, but this can run counter to some of the activities faculty engage in during traditional scholarly communication, creating obstacles for faculty participation.¹¹ A key challenge is aligning a repository's goals with the institution's research activities. Repository teams hoping to overcome integration and adoption challenges should have an idea of what the concept of value means for their institution.⁸

Communication and advocacy are essential for a successful repository. Librarians are particularly well-positioned to reach out to faculty as content management experts, partners, and collaborators to demonstrate and promote the benefits of a repository.^{7, 11} Repository advocates communicate tangible incentives and benefits of repository participation, such as increased usage metrics, mediated content uploads, enhanced metadata, and the other value-added services.^{2, 5} Repository teams should engage in proactive evaluation and continuous improvement so that services respond to changing user, institutional, and disciplinary needs.¹¹ Communication strategies should embrace champions and key stakeholders and shun discipline-specific jargon to broaden the message repository managers would like to convey.^{5, 7, 8}

PRIORITIES FOR THE DIGITALHUB REPOSITORY

As a service, DigitalHub was envisioned as facilitating the recognition of Feinberg scholars and their work by creating a single destination, or “living record,” to support the discovery, reuse, and acknowledgement of scholarly outputs. Galter Library was committed to following open access principles and hoped to use DigitalHub to help guide dissemination strategies. Additionally, a strong emphasis was placed on accommodating a wide range of digital works and research products to meet funders’ open access requirements, and on promoting non-traditional scholarly outputs. Non-traditional outputs were defined as items produced during the scholarly process but which are often not discoverable or made available for reuse, including measurement devices, patient education materials, curriculum materials, conference presentations, and community engagement materials.

As technology, the platform on which DigitalHub would be built needed to accommodate a variety of front- and back-end requirements. An open source solution was preferred because it offered more flexibility for local customizations. In order to select the most appropriate open source repository platform, Galter Library applied the National Library of Medicine’s master evaluation criteria as described by Marill and Luczak, considering factors such as functionality, scalability, interoperability, system security, platform and system support, stability, and strength of technology roadmap for the future.¹² Support for semantic web technology and machine-readable data was desired to better enhance the discoverability of repository contents. It was also important to provide an attractive, usable front-end that would require minimal instruction for new depositors and provide value-added features to encourage long-term, sustained contributions. Usage metrics, rich metadata, citation generation, DOI minting, and content sharing and reusability were all top priority features for DigitalHub.

DIGITALHUB DEVELOPMENT

Setting the Stage

After a review of repository platforms meeting initial criteria restrictions, the Galter Library repository team ultimately chose Fedora Commons <<http://fedorarepository.org>>, an open source project with an excellent application program interface (API) and a variety of frontend choices. Fedora was chosen for its ability to flexibly integrate with many types of enterprise and web-based systems, its scalability, and its durability. Fedora also provides the ability to express rich sets of relationships among digital resources and to query the repository using the semantic web’s SPARQL query language. Sufia <<https://github.com/projecthydra/sufia>> was chosen as a frontend, because of its interoperability with the library’s existing web infrastructure, its strong development community, and its customizability.

A variety of development activities were necessary to drive the project toward local goals. To better enable collaborative open development, a public GitHub repository was created and is still in use for the project <<https://github.com/galterlibrary/digital-repository>>, along with a ticketing system to document and manage all code changes, which is still actively used as the ticketing system for managing repository code work. Throughout the project, the

Galter Library team contributed its development work, including bug fixes and issue reporting, to the open source community. Once prototyping work was accomplished, the library formed the Digital Initiatives Working Group (DIWG) to help guide the project into production.

An Eye on Metadata

DigitalHub was designed to support comprehensive metadata for files and collections to allow for optimal discoverability. DIWG members selected the Dublin Core metadata schema as the basis of the descriptive metadata for DigitalHub, and additional elements were added for local purposes. In total, there are 26 descriptive metadata fields in DigitalHub, five of which are required for digital files (Title, Resource Type, Keyword, Rights, and Creator), two of which are required for collections (Title and Keyword). The number of required fields is minimal to reduce the burden on the depositor.

To ensure that the descriptive metadata is of the highest quality possible, different controlled subject vocabularies are used to enable autocomplete during entry. These include: National Library of Medicine Medical Subject Headings (MeSH) <<https://www.nlm.nih.gov/mesh>>, Library of Congress Subject Headings (LCSH) <<http://id.loc.gov/authorities/subjects.html>>, Library of Congress Name Authority File (LC/NAF) <<http://id.loc.gov/authorities/names.html>>, ISO639-2 for language <<https://www.loc.gov/standards/iso639-2>>, Geonames for geographic places <<http://www.geonames.org>>, and Creative Commons <<https://creativecommons.org/share-your-work>> and Open Source Initiative MIT license <<https://opensource.org/licenses/MIT>> for rights statements. Depositors were generally given a great deal of freedom and guidance for entering descriptive metadata. This was done to encourage the deposit of digital files without having depositors spend a lot of time and effort in supplying metadata for their files and collections. The interface is also configured to provide interactive help and metadata definitions to assist depositors in entering metadata.

Key Features of DigitalHub

A number of features were included in the scope of DigitalHub's development to enhance the user experience and produce a resource that would provide lasting value to users and the library. For example, Creator and Contributor fields were linked to the Northwestern University Lightweight Directory Access Protocol (LDAP) as auto-complete fields, enabling DigitalHub to find a person in Northwestern's LDAP Directory that matches the creator and use the name from the Directory in the Creator field. If no matching name is found in the LDAP Directory, then a name can be entered as free text. The Creator and Contributor fields were also integrated with VIVO <<http://vivoweb.org>> (an open source platform for research discovery), allowing for a link to the VIVO profile to be displayed for VIVO users.

An automatic Digital Object Identifier (DOI) assignment was added for all works deposited to the repository without a DOI specified in the metadata. The integration of DOIs into DigitalHub is a good example of a feature which strengthens Galter Library's commitment to non-traditional outputs. By receiving a DOI, objects placed in DigitalHub become more discoverable to web search engines, and can be more easily cited in manuscripts and grants.

DOIs also allow better tracking of dissemination and the subsequent impact of deposited materials as they are accessed and used further by others.

Other user-facing features added to DigitalHub included the addition of more informational content in “News” and “About” sections on the site. Enhancements were made to the tag cloud on the home page for improved browsing of repository content. Finally, the team incorporated an International Image Interoperability Framework (IIIF) server <<http://iiif.io>> to make image resources available to web applications hosted on other domains. This allows for richer image interactions and improved opportunities for sharing and collaboration.

DIGITALHUB LAUNCH

Pre-launch Activities: An Opportunity for Feedback-Driven Enhancements

Upon completion of the bulk of the development work, DigitalHub was rolled out for internal testing and a limited (soft) launch to one Feinberg department. Users were asked to perform uploads and supply accompanying metadata for materials such as professional and work-related presentation slides, articles, unpublished documents, photographs, and other materials. They reported any unanticipated error messages, and indicated their overall satisfaction with the upload process. They raised questions about specific metadata fields that they felt could be interpreted in different ways, and fields that they had wanted to see, but which were not present. The feedback that resulted from these initial launches was invaluable in identifying overlooked issues and lingering bugs. Based on this feedback, DigitalHub documentation was clarified, bugs were fixed, and technical improvements were implemented or prioritized as appropriate.

DigitalHub Launch

DigitalHub was finally ready for its official launch to the Feinberg community, right on time for Open Access Week 2015 (see Figure 1). It was successfully launched without major issues thanks to the team’s thorough preparation. The DIWG coordinated the marketing and outreach campaign with the Feinberg School of Medicine’s Office of Communications, which helped distribute mass email to all Feinberg and Northwestern Medicine community members. Additionally, the Galter Library communication team released a news item to announce the launch of DigitalHub a week prior to launch. The NUCATS Newsletter included a news story to communicate information about DigitalHub and its launch in its November 2015 newsletter.

DIGITALHUB ENGAGEMENT AND ADOPTION

Campus Engagement Strategies

Following the launch of DigitalHub, the DIWG employed a number of strategies to engage potential users wherever they may be in a “just-in-time” manner. For example, training and support materials were developed for users, including user guides and YouTube videos, and these were made available online, linked from the repository directly. A workshop, “DigitalHub: Preserving and Sharing Your Work Using NM’s Repository,” was developed and taught regularly by the Galter Health Sciences Library team, and the teaching materials

were distributed for use by liaison librarians to reach users who prefer face-to-face instruction, or who may have specific use cases or questions to discuss.

Liaison librarians play an important role in the success of almost every repository project. To better empower the liaison librarians to engage their departments and centers about DigitalHub, a DigitalHub in-service for the liaison librarians was developed. The goals for this in-service were multi-fold: increase familiarity with DigitalHub contents, uploading content, and use of lesser known features; explore examples of alternative outputs at Northwestern; support better understanding of the overall scholarly repository ecosystem; and prepare attendees to assist their groups with answering or anticipating questions from users about document types, enhancements, challenges, and so on. The key outcome for this in-service was better informed and experienced liaisons who could more easily incorporate DigitalHub in conversations, campus presentations, and liaison communications.

Galter Library liaison librarians collaborated closely with the DIWG to integrate DigitalHub messaging into communications with their assigned departments and programs on campus. The DIWG and liaison librarians collaboratively developed a list of challenges and responsive workflows for individuals who wished to deposit non-traditional scholarly works to better empower the liaisons to address questions related to file size and type, versioning, copyright, permissions levels and concerns about being “scooped,” storing and sharing research data as scholarship, using DigitalHub as a complement to other repository solutions, and DigitalHub as a collaborative resource for local and remote collaborators.

Finally, the DIWG engaged in purposeful messaging, aligning communications with key times in the scholarly calendar. The DIWG actively disseminated these messages in partnership with key stakeholders, such as the Northwestern University Clinical and Translational Sciences Institute and the Office of Research, to promote the repository to the campus and communicate enhancements, training opportunities, and project milestones. The library also established consultations, either one-on-one or group, and organized events that highlighted the current uses of DigitalHub and the many examples of the non-traditional scholarly works deposited to inspire more creative use cases.

Creative Use Cases

Currently DigitalHub has over 4,600 items publicly shared, and over 150 items privately deposited. DigitalHub has inspired a number of exciting use cases that highlight the institutional repository as a “hub” for a diverse range of scholarly activities on campus. Some of these are noted below:

- **StatTag** <<http://dx.doi.org/10.18131/G36K76>> is a free plug-in for conducting reproducible research and creating dynamic documents using Microsoft Word and R, SAS and/or Stata statistical software created by a team led by the director of the Biostatistics Collaboration Center and a Clinical Research Associate in the Department of Preventive Medicine.
- The **Statistical Considerations in Basic Science Sex Inclusive Research** <<http://dx.doi.org/10.18131/G3SG7F>> guide includes an explanation of five

scenarios pertaining to the sex of the cells to be used in the experiment along with a description of how an investigator should plan the experiment.

- The Center for Prevention Science Methodology Group has established a **Center for Prevention Implementation Methodology Collection** <<https://digitalhub.northwestern.edu/collections/6682x398c>> as part of making their research products openly available, per NIH open access requirements. The group's materials are a national-level resource of information in the field of prevention implementation methodology. They have indicated that they will be making their materials open access, most of which are in the form of audio files and videos.
- **Pediatric Neurology Briefs** <<https://digitalhub.northwestern.edu/collections/pedneurbriefs>> is an open access continuing education service designed to expedite and facilitate the review of current scientific research and advances in child neurology and related subjects. The DigitalHub team collaborated with the editor and assistant professor of pediatrics and neurology at Feinberg, to host the complete archives.
- DigitalHub has also partnered with Science in Society, Northwestern University's research center for science education and public engagement, to host images from its **Scientific Images Contest** <<https://digitalhub.northwestern.edu/collections/93c81706-47f8-49a1-88a2-3acbc971f4ed>> in a gallery. The contest has been held annually since 2010.

As these examples illustrate, DigitalHub has enabled Feinberg and other Northwestern scholars to create, share, and preserve attractive, functional, and citable digital representations of their teaching and research outputs.

CONCLUSION

Lessons Learned

After a year of DigitalHub use and library-sponsored events, the library gained a better understanding of how depositors used the platform, common questions that arose during use, and features users felt were missing. Scholarly work may be deposited either by content creators themselves, by their administrative support staff, or with the help of librarians. The following topics represent the most frequently asked questions related to the repository: rights of deposit (copyright laws); the type of Creative Commons License most appropriate for scholarly works; and the maximum size limit of the file that can be deposited. Users requested new features such as a proxy role to deposit on behalf of researchers; the ability to create a record for a resource without uploading an actual file; the ability to share private files with a collaborator; and the ability to acknowledge the funding for a project/research/publication. These questions and requests were useful in identifying and prioritizing future development directions. The DIWG also observed a lack of incoming user-supplied metadata, which Galter librarians now address by enhancing metadata for all public files following deposit as a value-added service to depositors.

The success of DigitalHub, from its development through its implementation and adoption, is largely due to the collaborative efforts of the librarians and staff at Galter Library. The platform is functional, attractive, and usable because of the skills of the development team and its response to feedback relayed from the DIWG and DigitalHub users. Early adopters and novel use cases resulted from the education and outreach efforts of the DIWG. Regular use and adoption in the Feinberg community are thanks to partnerships established by liaison librarians with their departments and other key stakeholders. As a new member of the COAR network, Galter Library hopes to elevate the profile of DigitalHub and maintain its early momentum to continue to serve the needs of the Feinberg community.

Future Plans

DigitalHub was developed with the vision of supporting the discovery, open representation, and preservation of scholarly outputs at Feinberg and Northwestern. With an understanding that research collections are at the heart of the research library, Galter Health Sciences Library is committed to maintaining, building, and implementing technical systems that will continue to enable and support the creation, curation, acquisition, and preservation of unique research collections. The library will continue its focus on non-traditional scholarly outputs and explore methods to demonstrate to the Northwestern Medicine community the benefits of utilizing the repository to share their work with other scholars, the public, and the wider community.

ACKNOWLEDGEMENTS

This publication describes efforts that were supported, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences grant number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors wish to extend special thanks to the entire Galter Library team and to campus partners for their wonderful support and assistance with the project.

REFERENCES

1. Cywin Allison J., Farrar Judy, Russo Tracy, et al. Digital Asset Management Team Report. Dartmouth, MA: University of Massachusetts Dartmouth, 2011 https://www.umassd.edu/media/umassdartmouth/collegeofvisualandperformingarts/resources/visualresourcecenter/DAMPT_report_final.pdf.
2. Crow Raym. The Case for Institutional Repositories: A SPARC Position Paper Washington, D.C.: Scholarly Publishing & Academic Resources Coalition, 2002 https://ils.unc.edu/courses/2015_fall/insl700_001/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf
3. Ganaie Shabir Ahmed, Jan Sumaira, Loan Fayaz Ahmad, and Refhat-un-nisa. "Current Trends of the Open Access Digital Repositories in Library and Information Science." *International Journal of Information Dissemination & Technology* 4, no. 4 (Oct-Dec 2014): 278–282.
4. Lynch Clifford. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL Bimonthly Report* 226 (2 2003): 7.
5. Confederation of Open Access Repositories. Incentives, Integration, and Mediation: Sustainable Practices for Populating Repositories. Göttingen, Germany, 2013 https://www.coar-repositories.org/files/Sustainable-best-practices_final.pdf.
6. Palmer Lisa A. "Cultivating Scholarship: The Role of Institutional Repositories in Health Sciences Libraries." *Against the Grain* 26, no. 2 (2014): 24–26, 28. http://www.against-the-grain.com/wp-content/uploads/2014/05/TOC_v26-2.pdf.

7. Sheret Larry, Walker Thomas, Beach Gretchen, and Zhang Jingping. "A Primer on How to Launch an Institutional Repository Successfully." *The Charleston Advisor* 16, no. 3 (2015): 48–55. doi: 10.5260/chara.16.3.48.
8. Palmer Carole L., Teffeau Lauren C., and Newton Mark P.. "Strategies for Institutional Repository Development: A Case Study of Three Evolving Initiatives." *Library Trends* 57, no. 2 (2008): 142–167. doi:10.1353/lib.0.0033.
9. Poynder Richard. "Q&A with CNI's Clifford Lynch: Time to Re-think the Institutional Repository?" Richard Poynder: Independent Journalist (website). September 22, 2016 http://www.richardpoynder.co.uk/Clifford_Lynch.pdf.
10. Lagzian F, Abrizah A, and Wee MC. "Measuring the Gap between Perceived Importance and Actual Performance of Institutional Repositories." *Library & Information Science Research* 37, no. 2 (4 2015): 147–155. doi:10.1016/j.lisr.2014.06.007.
11. Kutay Stephen. "Advancing Digital Repository Services for Faculty Primary Research Assets: An Exploratory Study." *Journal of Academic Librarianship* 40, no. 6 (2014): 642–649. doi:10.1016/j.acalib.2014.08.006.
12. Marill, Jennifer L, and Luczak Edward C.. "Evaluation of Digital Repository Software at the National Library of Medicine." *D-Lib Magazine* 15, no. 5/6 (May-Jun 2009). doi:10.1045/may2009-marill.

FIGURE 1.
Screenshot from the DigitalHub repository