Developing a Digital Scholarship Ecosystem

Dr. Ray Uzwyshyn, Ph.D.
Texas State University Libraries
San Marcos, Texas, USA, 78666
1-512-245-5687
ruzwyshyn@txstate.edu

ABSTRACT

This research presents elements necessary to develop a Digital Scholarship Research Ecosystem for a university, college or research institution. Software systems, hardware, human resources and timelines are outlined with brief theoretical overviews and a pragmatic focus on 'open-source' (freely available) software, best-in-class applications and global best practices. Major digital scholarly system components in a larger digital ecosystem are discussed: Online Institutional Collection Repositories (D-SPACE), Online Research Data Repositories (DATAVERSE), Identity Management Systems (ORCID), Electronic Thesis and Dissertation Management Systems (VIREO), Academic Journal Systems (OJS3), Digitization Labs, User Interface Software (OMEKA). System assessment, synergistic possibilities and future directions are reviewed. This research arises from a successful five-year phased implementation of such a digital ecosystem for Texas State University Libraries, a large US university research library system. This scholarly ecosystem is suitable for any university, college, research institution or academic research library interested in setting up or building on such an infrastructure and enabling faculty and graduate students with their scholarly research online.

CCS Concepts

Applied computing → Education → Digital libraries and archives; Information systems → Information retrieval → Document representation → Document collection models; Applied computing → Computers in other domains → Digital libraries and archives; Information storage systems → Storage management → Information lifecycle management; Information systems → Data management systems → Database administration → Database utilities and tools.

Keywords

Digital Libraries; Digital Archives; Digital Scholarship; Information Repositories; Information Retrieval; Digital Research Collections; Research Data Repositories; Identity Management Systems; Electronic Thesis and Dissertation Management Systems; Digitization Labs; User Interface Software; Open-Source Software; Open Access Journal Software; Online Academic Journals; Research Information Systems.

ICEIT 2020, February 12–14, 2020, Oxford University, UK.

DOI: http://dx.doi.org/10.1145/12345.67890

1. INTRODUCTION

1.1 Overview

In our new millennia, faculty and graduate student research has quickly migrated online. Infrastructures are needed to support this new scholarly research cycle. Digital infrastructures may be profitably developed through low-cost open-source software methods. Components are readily available and may be placed together into digital ecosystems, opening synthetic possibility and unique synergies for dissemination of scholarly research on previously unparalleled networked and global scales. The open-source aspects of such ecosystems allow a wider global group of research institutions to take advantage of such infrastructures' online possibilities, especially where budgetary considerations form larger factors. The term 'digital ecosystem' draws on the language of ecology to describe and analyze digital scholarly research systems from vantages that consider component relationships and how these relationships influence the larger environment [1]. The concept of a digital information ecology opens notions of dynamics among ecological ideas and properties of networked, global information environments [2]. This research views the digital information ecosystem as part of a larger ecological whole which may synergistically function as a digital ecosystem. These new ecosystems situate academic research and new digital artifacts created from research within socially distributed online networks for discovery and progress in the research enterprise.

1.2 Rationale and Hypothesis

By systematically organizing graduate students' and university research faculties' scholarly output online, new research synergies are enabled and processes engendered to create a digital scholarly ecosystem which may become a standard scholarly model for any university, college or research institution in the 21st century. Placing digital scholarship components within an ecosystem paradigm usefully guides larger evolutionary possibilities for researcher information systems, digital component development and global human researcher communities. By undertaking a phased systematic approach to building such systems, we may better improve access and retrieval possibilities and circulation of research towards new discoveries and insight. Collocating these components in larger digital ecosystems may also unearth unexpected network effects allowing better progress and efficiency for scholarship and research in the 21st century.

2. SYSTEM SOFTWARE

This section outlines major classes of open-source scholarly software and hardware components needed for a digital scholarship research ecosystem. Components are defined in general terms. Best-in-class open-source examples and source downloads are given so anyone interested in replicating such a system can do so. All software is freely available and links to resources are given in footnotes. Elements necessary to develop a digital scholarship research ecosystem for a university, college or research institution are

presented. Software systems, hardware, human resources and timelines are outlined with a pragmatic focus on 'open-source' (freely available) software, best-in-class applications and global best practices. Major digital scholarly system components in a larger digital ecosystem are discussed: Online Institutional Collection Repositories (D-SPACE), Online Research Data Repositories (DATAVERSE), Identity Management Systems (ORCID), Electronic Thesis and Dissertation Management Systems (VIREO), Academic Journal Systems (OJS3), Digitization Labs and User Interface Software (OMEKA) and synergistic possibilities are reviewed. This research arises from a successful five-year phased implementation of such a digital ecosystem for Texas State University Libraries, a large US university research library system. This scholarly ecosystem is suitable for any university, college, research institution or academic research library interested in setting up such an infrastructure and enabling university faculty and graduate students with their scholarly research online.

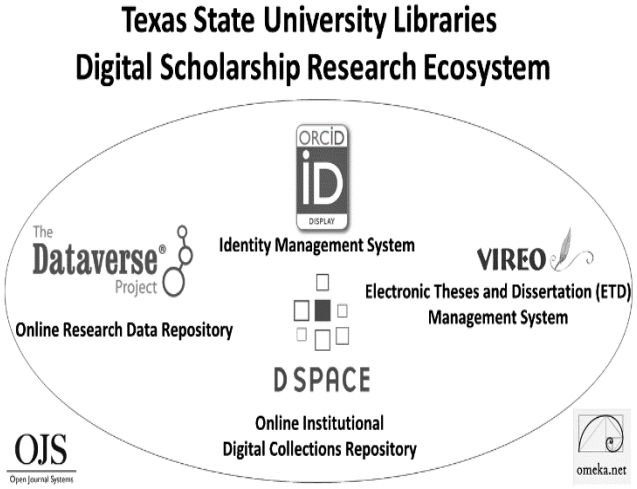

Figure 1. Texas State University Digital Scholarship Research Ecosystem

2.1 Institutional Digital Collections Repository

An online digital collections repository is a digital infrastructure which provides open access to scholarship and research produced by a university, college or research institution. A digital repository organizes, centralizes, preserves and makes accessible research and knowledge generated by the institution's research community. The application centers on text-based media artifacts including preprints, faculty publication, white papers, conference presentations, graduate student theses and dissertations. Research centers within an institution such as Specialized Research Groups/Centers, Special Collections and University Archives may also avail themselves of an institutional repository's possibilities [3]. This digital collection repository gives open web-centric visibility to the institution's research output and opens the institution to the possibilities of a global scholarly networked environment. Texas State University Digital Collections Repository [4] reconfigures the originally MIT-developed open-source product, D-SPACE [5].

2.2 Online Research Data Repository

Working closely in concert with any institutional repository is its younger cousin, the online research data repository. A scholarly data research repository is a data-focused repository where researchers deposit datasets from their research and experiments. The repository is online and the data is made available for open public access and re-use. Each dataset includes citation information and a DOI (Digital Object Identifier), or UNF (Universal Numerical Fingerprint), to facilitate attribution and usage tracking [6]. The repository provides a home to link researchers' data directly to their research publications either in journals, monographs or online presentations available from the repository. Data research repositories play important roles in fulfilling funding requirements and providing transparency for ensuring reproducibility of data to forward the scientific research enterprise. Texas State University reconfigures Harvard's Institute for Quantitative Social Science Dataverse open-source software [7] as part of a larger Texas consortia universities' network [8]. This implementation lends itself to the added advantage of sharing and collaboration between geographically co-located academic institutions (22 state research institutions) for aggregating similar disciplinary data and research data projects. The application allows researchers easy collaboration and comparison of results and data [9].

2.3 Researcher Identity Management System

With the proliferation of researchers globally, any digital research ecosystem would do well to utilize a 'Researcher Identity Management System' to connect and disambiguate scholars' names. These systems disambiguate common researchers' names such as James Smith, Maria Hernandez, Mustafa Abad or Zhang Wei. That is, the system differentiates the work of James Smith, the MD, from the astrophysicist, James Smith, or the scholarly work of the botanist, Maria Hernandez, from the biochemist, Maria Hernandez. Both may have extensive lists of journal publications. An online 'Research Identity Management System' gives a researcher a contributor ID, a unique number, so that all publications from a single researcher can be easily found, linked, organized and aggregated across multiple information systems. Texas State utilizes the open-source ORCID system [10] to connect and organize unique researchers' publications and disambiguate similarly named researchers. ORCID ID's may also be used to link together a researcher's entire corpus of publications and profile, helping researchers better manage their research environment and providing authority control for their research corpus [11].

2.4 Thesis & Dissertation Management System

An Electronic Thesis and Dissertation Management System addresses intermediary steps in the Electronic Thesis and Dissertation process, bridging student thesis/dissertation submission with graduate school review, online publication and ETD preservation. Texas State University has customized and utilizes, VIREO, an open-source middleware which connects theses, dissertations and graduate students with the institution's graduate school, submission process and university digital repository [12]. Through VIREO, a student's thesis or dissertation is seamlessly transferred to the university institutional repository for preservation and access and searchability on the internet [13]. The more immediate online publication of graduate student research enables access to research worldwide with indexing metadata and efficient retrieval of research. This immediate accessibility also provides a leg up for students as they move forward from graduate school to research academic related careers.

2.5 User Interface Software

As digital projects become more demanding and specialized, user interface software becomes a requirement to provide both an elegant portal or gateway entrance to specialized research projects and a necessary middleware to seamlessly connect the various software and media components mentioned above. For raising the quality of online exhibitions to research portals utilizing the digital and data

repositories, Texas State University Libraries uses OMEKA [14]. OMEKA is a flexible open-source web-publishing platform allowing streamlining of the front end of larger research projects, scholarly collections and 'special center' scholarly research. The software can connect to an institutional repository and data repository for linkages to text, image, media collections and datasets [15]. For more advanced imaging and image-centered projects, specialized software such as the International Image Interoperability Framework (IIIF) may be used [16]. This framework allows progressive image download and magnified zooming for ultra high-resolution images ranging from art history-based research to cellular biology, pulmonary pathology and manuscript research. Any research scholarship, predicated on the need for rich access to high-resolution image-based resources, may usefully take advantage of this tool. With this tool geographically dispersed collections may be digitized, aggregated and hosted for scholars online for consultation and work with these high-resolution artifacts and manuscripts regardless of physical location globally [17].

2.6 Academic Journal & Conference Software

An integral part of a digital scholarly ecosystem is the option for a university department or specialized research center to have its own online research journal and access to conference/colloquia management software. Texas State utilizes Open Journal Systems (OJS 3), a freely available software for the management and creation of peer-reviewed academic journals [18]. Open Journal Systems enables faculty editorial workflows for the open-access academic journal publication process from article submission and assigning referee review to metadata and indexing [19] and later, online publication. Texas State University hosts OJS3 through the Texas Digital Library, a consortium of 22 Texas University Libraries so that these research journals may also be easily aggregated and organized [20]. While beyond the scope of this paper, web-publishing tools are also available to create online infrastructures for scholarly conference and colloquia. These digital systems allow the creation of academic conference websites, from automation of presentation and article submission to organizing conference proceedings for search and retrieval [21].

3. HARDWARE, THE DIGITIZATION LAB

As university faculty, departmental and research center projects become more complex, specialized digitization equipment and staff facilitation will be needed. This ranges from more robust photographic digitization equipment to audio and video digitization equipment to high volume scanning equipment. A digitization lab will also be useful for historical analog media formats. This ranges from archival manuscript digitization and various textual documents to previous era's audio and video formats and larger project, more rapid text and image digitization needs (i.e. rapid text and slide scanners, large format newspaper scanners, etc.). While a full discussion of this hardware is beyond this paper's scope, an overview of basic types of basic equipment needed and further links to projects possible is provided on the Texas State Digital and Web Services Site [22].

4. IMPLEMENTATION TIMELINES

The above-described digital scholarly ecosystem may appear daunting to set up. This does not have to be, if a phased implementation with measured benchmarks process is followed. Research institutions will be at various levels of development for such ecosystems. It is best to start simply or continue from where an institution currently stands and build strong digital foundations from there. If starting from scratch, start simply by setting up an online digital institutional repository (D-SPACE). Populating the content of this repository may begin with local faculty and departmental research papers, preprints and graduate student theses and dissertations. Once initial workflows are operating well and faculty and student demand rising, a user interface application, such as OMEKA, can be added and larger image/text-based research content projects may be tried. Faculty and research centers will request more complex projects as initial projects are successfully completed. Depending on an institution's orientation, a multimedia equipment direction (see section 2.5 Digitization Lab) or, alternatively, data-oriented direction (see section 2.2 Data Repository) may also be pursued. After these applications have been launched and this workflow infrastructure realized, projects connecting institutional repository, data repository and user interface applications may be piloted to begin to explore synergies and more complex possibilities. If the institution is in possession of a graduate school, VIREO ETD middleware may be pursued earlier in the benchmarking process, connecting graduate school, students and digital repository. Increasingly, theses and dissertations rely on voluminous datasets, so it is beneficial if a data repository and institutional repository are set up closer together. An online academic journal system may also be added at this time working out connections and administrative editorial responsibilities with faculty groups. Once all these major components are functioning well together, more advanced hardware may be added to the digitization lab. Higher level imaging software may be pursued (i.e. IIIF Framework) if projects warrant. Farther afield, more leading-edge R&D connections may also be prototyped to build out the digital ecosystem's possibilities. These connections range from adding a university library GIS specialist and tying 'cognitive mapping' possibilities to digital research and data, to enabling data and research through data visualization and machine-learning possibilities. A good goal trajectory may be planned along a one to five-year phased timeline for getting the infrastructure outlined above started. Benchmarking implementations and milestones should follow a phased project management program [23], with easy wins, as the complexity of systems deepens.

5. HUMAN RESOURCES

As with any larger IT project, dedicated human resources will be needed. It is best to develop such project teams so that there is a partnership between the university's libraries and the university's IT division. A system administrator will be needed to set up the open-source server infrastructure (i.e. LAMP: Linux, Apache, MySQL, PHP and Microsoft) for software beginning with the institutional repository unless hosted options are pursued [24]. A Digital Collections Librarian will be needed for day-to-day administration, marketing and user-support of the repository. A Metadata Librarian, or expert versed in various specialized academic metadata schema or, at the least, Dublin Core [25], will ensure accessibility for efficient search and retrieval and search engine optimization of all research that will be placed online. The Digital Collections Librarian may also double as the Data Repository/Open Journal administrator/faculty liaison and Online Researcher ID system support as these are set up. These duties will more than encompass a full-time employee's time. As demand grows, these positions may be divided into two or three. Once the User Interface Application (OMEKA) has been set up and these research scholarship projects begin, a dedicated project manager, with the PMP certification (Project Management Professional) will be useful to manage a growing projects' list and faculty/graduate student projects which require facilitation [23]. As the breadth of material increases, a digitization specialist will also be needed to run a digitization lab as scholarly digitization needs increase. This project manager will also coordinate workflows among various stakeholders as the list of projects and staff increases. As demands and

project complexity increases, further positions may be added. For later phases, a Collections Analytics and Data Visualization Specialist can enable insight into academics' and graduate students' data research collections. A GIS Specialist can enable faculty and students with GIS possibilities for their research. An extra programmer will also be useful to integrate connections among the digital systems' evolving component ecology with API's of legacy, existing and new university systems.

6. ECOSYSTEM SYNERGIES

As the system matures, more complex ecosystem synergies will naturally arise. Research papers in the digital collection repository may be customized to link with associated datasets in the data repository. Dissertations and theses housed in the collections repository may be associated with datasets in the research data repository. Once a paper is published through the Open Journal Systems online journal, the associate dataset and antecedent thesis/dissertation may be referenced directly through the data repository and digital collections repository. All papers/data and other intellectual output of a researcher may then be aggregated through a researcher profile found through the identity management system. All of these offer a paradigm shift for research efficiency and efficacy. This digital ecosystem allows easy online search, retrieval, aggregation and organization of academic research on global levels. Through structured metadata schema application, search engine optimization is achieved for higher level accessibility. Through the digital collections and data repositories, online research may be naturally linked with associated primary research data on faculty and graduate research levels. Entire research corpora may be easily aggregated through the online identity management system providing additional connective threads among systems. Multimedia websites and portals with robust backend storage also become possible with the user interface software and digitization lab for next level online research possibilities. All may also be referenced directly through open access online journal options, a next stage of global scholarly communications. Complex networks here become possible among collaborators, stakeholders, staff, faculty, students, research workgroups and teams. Digital scholarship ecosystems are in their infancy but with manifold possibility. Prospects are rich and, yet, largely unexplored for future development potential and further synergies with research scholarship and technology. New connections and intuitions should be pursued to empower researchers and research on global networked scales and, as yet, unseen dimensions.

7. RESULTS AND ASSESSMENT

The core scholarly research ecosystem outlined above was developed and implemented for Texas State University Libraries, 2014-2019. The system grew organically and utilized a variety of open-source digital scholarship components. This ecosystem ultimately consisted of Online Institutional Collection Repository (D-SPACE), Online Research Data Repository (DATAVERSE), Identity Management Systems (ORCID), Electronic Thesis and Dissertation Management Systems (VIREO) and Academic Journal Systems (OJS3). The larger ecosystem is currently utilized successfully by research faculty and graduate students daily and has been continually assessed through biannual LibQUAL qualitative surveys, 2015-2019, and annual statistical reports (See Table 1). System usage patterns, downloads, ORCID ID adoption and open-source new academic journal development have shown continual growth [26].

Table 1: Texas State University Digital Scholarship Ecosystem Annual Growth

System Downloads	2015	2016	2017	2018
DSpace	330,668	396,650	656,778	1,015,314
ETDs	158,240	200,373	328,420	470,437
Dataverse	N/A	N/A	455	3,451
ORCID ID's	190	316	438	545
OJS Journals	1	2	2	3

8. FUTURE DIRECTIONS

Metaphors of ecological diversity and evolution are key for the health of any digital research information ecology. Digital scholarly ecosystems focus attention on relationships among technological tools, researchers, research and practices. The ecosystem presently described may be scaled from university to consortia to state, national and international levels. Laterally, disciplinary and interdisciplinary directions for digital scholarly ecosystems are also possible. Here, the concept of local information ecologies and networks of relationships involves active participation and practices of researchers engaged in the research process [27]. Concepts of interpenetration, interdisciplinarity and cross-fertilization of ideas among researchers should be embedded on application levels. Enhancing social network possibilities are fertile directions to augment the present ecosystem described. These social network and connective possibilities form a next stage of inquiry and build. Much of the research system infrastructures overviewed herein also asks for next levels of application of data visualization and machine learning, especially with regards to the data repository and institutional digital collections repository. The application of these possibilities can provide further layers of insight and paths of connectivity among disparate bodies of research. There is also much room for connecting a wider range of currently unconnected global research institutions into these open-source digital research ecosystem paradigms on local, consortia or other levels. Possibilities are manifold.

9. CONCLUSIONS

Setting up a Digital Scholarly Research Ecosystem should be an imperative for any research institution globally in the 21st century. These are no longer optional in our globally connected village but core to the academic research enterprise - present and future. By taking a phased, project management approach, any institution can be successful in building such infrastructures, enabling their faculty, students and institution on a global stage and, hopefully, also building on and improving the digital scholarly research ecosystem presented in this research model.

10. ACKNOWLEDGEMENTS

I wish to acknowledge and thank my colleagues, employees and administration at Texas State University Libraries without whom none of these infrastructures would have been possible: University AVP and University Librarian Joan Heath, Todd Peters, Laura Waugh, Misty Hopper, Ginger Williams, Jeremy Moore, Jason Long, Erin Mazzei, Mary Aycock, Gina Watts, Stephanie Larrison, Nathanial Dede-Bamfo, The Wittliff Collections staff and director, David Coleman, and University Archivist, Kris Toma, Texas State University Archives. I would also like to acknowledge the scholars,

research faculty and graduate students who continue to contribute to the content of our digital ecosystem and utilize these infrastructures daily.

11. REFERENCES

[1] Nardi, B. and O'Day V. 1999. *Information Ecologies.* MIT Press, Cambridge, Mass.

[2] Davenport, T. *Information Ecology: Mastering the Information and Knowledge Environment*. 1997. Oxford University Press, Oxford.

[3] Texas State University. 2019. Digital Collection Repository Communities. Retrieved from https://digital.library.txstate.edu/

[4] Texas State University Libraries. 2019. What is Texas State University's Digital Collections Repository? Retrieved from https://digital.library.txstate.edu/

[5] Lyrasis Duraspace. 2019. DSPACE website and Download. Retrieved from https://duraspace.org/dspace/

[6] Uzwyshyn, R. 2016. Research Data Repositories: The What, When, Why and How. *Computers in Libraries* (April 2016). Retrieved from http://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml

[7] Harvard Institute for Quantitative Social Science. 2019. The Dataverse Project. Retrieved from https://dataverse.org/

[8] Texas Digital Library. 2019. Texas State University Data Research Repository. Retrieved from https://dataverse.tdl.org/dataverse/txstate

[9] Uzwyshyn, R. 2018. Research Data Repositories: Developing and Implementing Infrastructures for Institutional and Consortial Environments [PPT presentation]. *Coalition for Networked Information Biannual Conference*. San Diego, CA. April 12-13, 2018. Retrieved from: http://rayuzwyshyn.net/TXU2018/DataResearchRepositoriesCNIPresentation.pptx

[10] ORCID. 2019. ORCID: Connecting Research and Researchers. Retrieved from https://orcid.org/

[11] Texas State University Libraries. 2019. Developing a Researcher Profile & Managing Your Research Environment: ORCID. Retrieved from: https://guides.library.txstate.edu/researcherprofile/orcid

[12] GITHUB. 2019. VIREO 4: Github Download. Retrieved from https://github.com/TexasDigitalLibrary/Vireo

[13] Texas State University Graduate College. 2019. Thesis & Dissertation Information. Retrieved from: https://www.gradcollege.txstate.edu/students/thesis-dissertation.html

[14] Roy Rosenzweig Center for History and New Media. 2019. OMEKA Website and Source Code Download. Retrieved from https://omeka.org/

[15] Texas State University Wittliff Collections. 2019. The Making of Severo Perez's *And the Earth Did not Swallow Him* (OMEKA Example). Retrieved from: http://exhibits.library.txstate.edu/thewittliffcollections/exhibits/show/severo-perez/

[16] IIIF Consortium. 2019. International Image Interoperability Framework. Retrieved from https://iiif.io/

[17] Texas State University Wittliff Collections. 2019. Cabeza de Vaca Digital Library (IIIF Viewer Example). Retrieved from: https://exhibits.library.txstate.edu/cabeza/

[18] Public Knowledge Project. 2019. Open Journal Systems. Retrieved from https://pkp.sfu.ca/ojs/

[19] Texas Digital Library. 2019. *Journal of College Academic Support Programs* (OJS3 Example). Retrieved from: https://journals.tdl.org/jcasp/index.php/jcasp/index

[20] Texas Digital Library. 2019. Open Access Journals. 2019. Retrieved from: https://www.tdl.org/journals/

[21] Public Knowledge Project. 2019. Open Conference Systems. https://pkp.sfu.ca/ocs/

[22] Texas State University Libraries. 2019. Digital and Web Services Equipment and Projects. Retrieved from https://www.library.txstate.edu/about/departments/dws.html

[23] Uzwyshyn, R. 2012. Can We Get Some Order Here? The Application of Principles of IT Project Management for Online Library Projects. *APUS Internal Whitepaper*. Retrieved from http://rayuzwyshyn.net/APUS2013/ITProjectManagementArticleUzwyshyn3.pdf

[24] Duraspace. 2019. Hosted DSpace. Retrieved from: https://duraspace.org/dspace/resources/hosted-dspace/

[25] American Society of Information Science and Technology. 2019. Dublin Core Metadata Initiative and Standards. https://dublincore.org/

[26] Texas State University Libraries. 2019. LibQUAL+ Survey Presentation (ppt.) and Annual Reports: 2014-2019. Texas State University Internal Documents

[27] Nardi, B and O'Day, V. *1999 Information Ecologies: Using Technology with Heart.* First Monday. Retrieved from: https://firstmonday.org/ojs/index.php/fm/article/view/672/582