COMPLETE COVERAGE OF LIBRARY INFORMATION TECHNOLOGY

COMPUTERS IN LIBRARIES

BY LIBRARIANS, FOR LIBRARIANS, BY LIBRARIANS, FOR LIBRARIANS
OUR 40 YEAR OF SERVICE

NOVEMBER 2020
VOLUME 40, NUMBER 8
ISSN: 1041-7915; USPS: 730-510

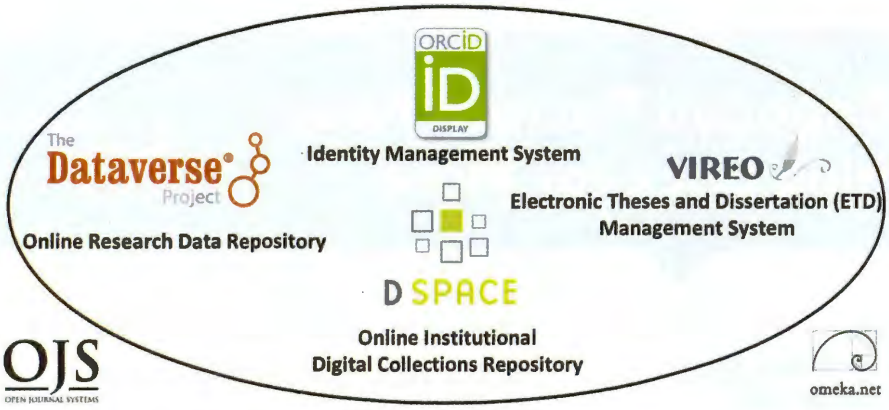
AGILE LIBRARIES

FEATURING:
How to Structure a Digital Research Ecosystem

Open Digital Research Ecosystems:

How to Build Them and Why

By Ray Uzwyshyn



The open source components for Texas State University's digital research ecosystem

Placing digital scholarship components in this ecosystem paradigm sets baseline expectations and provides a road map for future development.

Many universities and college libraries now place their academic research online through open digital ecosystems. These new infrastructures make research readily discoverable, opening doors for collaboration and quickening scholarly progress. This article explores these research tools and their potential for enabling researchers to connect, build on each other's work, and empower discovery.

WHAT IS IT?

A digital scholarly research ecosystem is a network of software components that enables faculty and student research, raises research profiles, and opens possibilities for global collaboration. The emergence of open source software tools supports the development of ecosystems to empower research institutions and academic libraries around the world. This software makes it possible to build systems

without starting from scratch because of the open components and active developer communities that are able to customize and link the networked components. Such ecosystems increase discovery and the possibility for making previously hidden connections.

The digital research ecosystem at Texas State University's libraries consists of six main software components: a digital collections repository, a research data repository, an identity management system, an electronic thesis and dissertation (ETD) management system, user interface software, and open source journal software.

Collection and data repositories, identity management systems, and ETD systems enable the dissemination of knowledge. They better identify information and allow curation, aggregation, and the online publication of research. Network effects also produce large gains for researchers in terms of retrieval and research visibility. By collocating open source components in a networked ecosystem, the whole becomes greater than the sum of its parts. Research that otherwise would be buried or not easily accessible behind database paywalls can be more easily found. Plus, digital ecosystems provide the means for speeding up the academic research cycle, facilitating communication and collaboration. These systems enfranchise the discovery of information. They help in the gathering and analysis of data and enable peer review, the online publication of research, and sharing. Together, these abilities extend and empower the impact of research.

PRIMARY COMPONENTS

Each of the digital research ecosystem's components serves a specific purpose and need, enabling the larger system. To get started in building and customizing an ecosystem that's best-suited to your institution's needs, let me highlight what each component does and where it fits in an ecosystem, starting with the repositories that form the system's core content engines.

The Digital Collections Repository

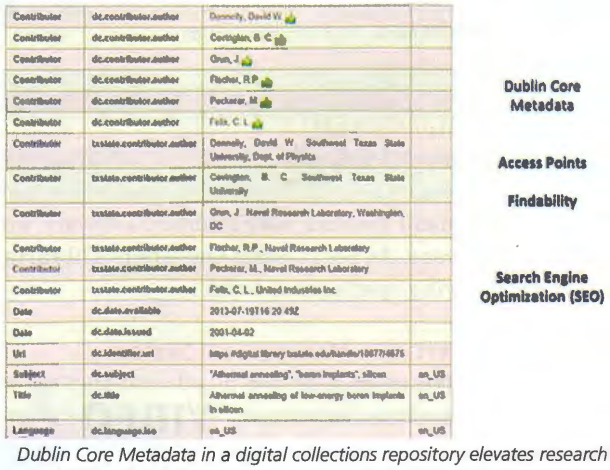
An institutional digital collections repository organizes, centralizes, and makes accessible information, research, and knowledge generated by an institution's research community (faculty members and graduate students). This material includes preprints, faculty publication white papers, conference presentations, and graduate student theses and dissertations. (Note that while a digital collections repository may be used for a spectrum of media formats, it is best used for textual content.)

A collection repository delivers value through the application of structured metadata schema for SEO, which helps to heighten retrieval possibilities. Our Texas State University's digital collections repository uses the open source platform DSpace.

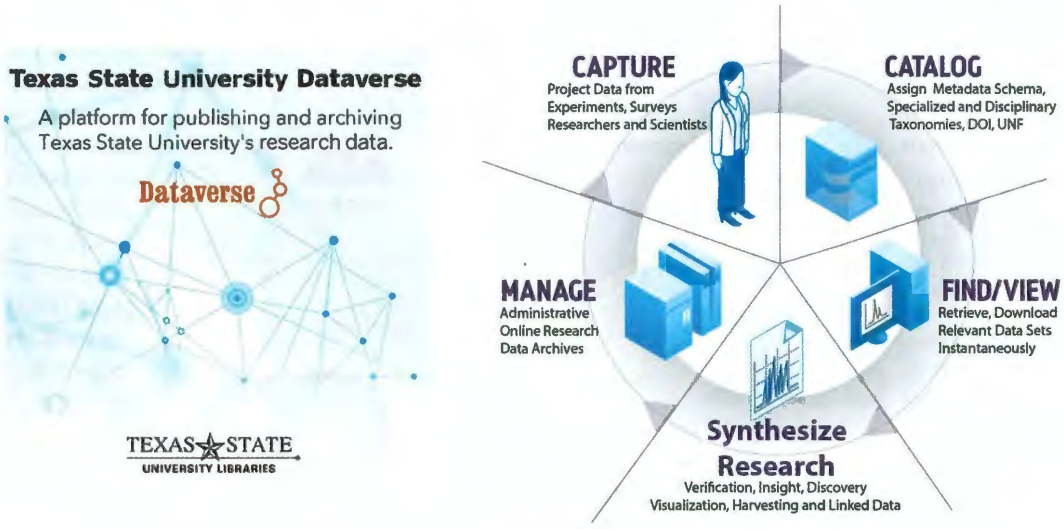
The application of structured metadata for textual academic research opens discoverability. Metadata enables multiple points of subject access. These effects translate to increased article citations through the more precise availability of relevant research through online search engines.

Since most publishers today allow digital archiving in one form or another, research data can be made not only more discoverable but more retrievable. OA publishing models have also contributed to an increase in citations in many fields, demonstrating larger network effects (slideserve.com/kasimir-buckley/emerging-trends-in-scholarly-communication-heather-joseph-executive-director-sparc).

DSpace, the digital collections repository at Texas State University's libraries; digital.library.txstate.edu and duraspace.org/dspace



Dublin Core Metadata in a digital collections repository elevates research via SEO; digital.library.txstate.edu/handle/10877/4675?show=full



Texas State University uses Dataverse to facilitate the data research cycle; dataverse.tdl.org/dataverse/txstate

Research Data Repository

The second primary component of a digital research ecosystem is a research data repository, the data analogue of a text-centered collections repository. Data repositories are specifically tailored for publishing and archiving research data. They allow a researcher to capture, upload, and assign metadata as well as retrieve and download datasets.

In the last 30 years, multi-university collaboration has increased exponentially in science, engineering, and social science disciplines. A data repository may be configured as a single instance or as a consortial model. Texas State University's libraries constitute one of several individual instances of the larger Texas Digital Library Dataverse project. The state consortium allows researchers to share, publish, and archive their data but also search. Research for similar datasets can be collocated with other consortium members, enabling opportunities for discovery, collaboration, and comparison of data results.

SECONDARY COMPONENTS

Secondary components of a digital research ecosystem include ETD management and research identity management systems, user interface software, and OA journal software. At Texas State University's libraries, we use the following open source tools.

Thesis and Dissertation Management System

The Vireo ETD management system addresses intermediary steps in the ETD process. This software bridges student thesis/dissertation submission with graduate school review, online publication, and ETD preservation. It connects with the collections repository and the data repository. Students can publish and link theses and dissertations with data and other text-based research materials. Vireo was developed by the Texas Digital Library and Texas state universities libraries and is open source software that's freely available to all institutions.

VIREO
Submit Graduate School Theses and Dissertations

ORCID online identity management system hub; orcid.org

Online Research Identity Management System

ORCID is a widely used research identity management system that gives researchers a unique number, an ORCID iD. This ORCID iD disambiguates scholar names globally and allows publications to be found, linked, and aggregated across multiple information systems. Papers in the collection repository and datasets in the data repository may be associated with ORCID iDs for aggregation of research profiles. ORCID itself may also act as a hub connecting the research landscape. This hub is a network within a network to aggregate from several further sources and connect researchers (see image).

Open Source User Interface Software

Open source user interface software is utilized in a digital research ecosystem to allow an elegant portal or gateway entrance for larger research projects, digital collections, and data repositories. Texas State University uses Omeka to link research, text, images, media, and datasets. Omeka acts as a front end to connect component networks. It can be configured for individual projects and as a shareable resource across multiple sites. This class of software provides a publishing platform for sharing digital research

and creating multimedia-rich front-end online exhibits. Omeka may be linked to the previously described back-end data, digital library collection repositories, and online archives of source materials. Also, the Avalon Media System should at least be noted here. It is an open source research service specifically for managing and providing access to very large collections of digital audio and visual research media (video) should this be needed to further expand system multimedia content capabilities.

Academic Journal Software

OA academic journal software mirrors the process of print publishing. Texas State University uses Open Journal Systems (OJS) to facilitate our refereed journal workflow for online publishing. As with all online journal management software, OJS provides the potential to create deeper connections to source data and background research. The software supports embedded links to content in the digital collections and data repositories, thus improving discovery.

OJS
OPEN JOURNAL SYSTEMS

IMPLEMENTATION

A journey of a thousand miles begins with a single step. While the list of software and projects may sound daunting, there are many paths to take with regard to the implementation of digital research ecosystems. Let me focus on two major considerations in developing your implementation strategy, staffing requirements, and assessment.

Human Resources

A human resource infrastructure may be set up with as few as two staffers. A systems administrator will be needed for server infrastructures and to handle basic maintenance and customization. A digital collections librarian/specialist will oversee the administrative side of the various software, including researcher support, marketing, and instruction on chosen systems. Every institution or library will also have unique research needs and possibly a special focus (textual content, data, multimedia, dissertation archiving, etc.). Implementation and human resource infrastructures should be tied to institutional directions and profiles. A research-focused science and engineering university will clearly have different needs from a 4-year liberal arts-focused college. Software should be chosen accordingly.

Add-Ons: The Digitization Lab

As digital ecosystem software needs are met, adding a digitization lab and the associated hardware can expand possibilities for faculty-member and graduate-student research projects—including the digitization of books, manuscripts, and journals—and audiovisual and visualization digitization technologies. Combining research ecosystem components opens amazing possibilties for digital scholarship and research collaboration opportunities.

Courtesy of Texas State University's Digitization Lab, library.txstate.edu/about/contact/departments/digital-web-services.html

Digitization labs open various possibilities, such as the digitization of this 500-year-old Cabeza de Vaca manuscript.

System	2015	2016	2017	2018	2019
Downloads					
DSpace	318,742	385,163	341,224	972,359	1,010,349
ETDs	158,240	200,373	328,420	470,437	505,658
Dataverse	n/a	n/a	455	3,451	2,043
Items Added					
DSpace	1,437	1,546	1,660	2,135	2,720
ETDs	1,174	1,326	1,581	1,789	2,218
Dataverse	n/a	n/a	28	33	53
ORCID IDs					
ORCID	190	316	438	545	669
Hosted Journals					
OJS	1	2	2	3	4

Texas State University libraries' annual digital ecosystem usage growth: downloads, number of items, ORCID iDs, and hosted journals

Open Source Software Links and Downloads

Avalon Media System
avalonmediasystem.org

DSpace
duraspace.org/dspace

Dataverse
dataverse.org

Omeka
omeka.org

Open Journal Systems
pkp.sfu.ca/ojs

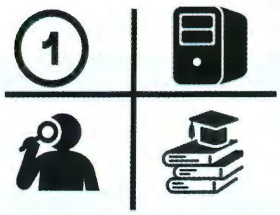
ORCID
orcid.org

Vireo
tdl.org/etds

Connecting Librarians Discord
discord.gg/j5VCXgn

Seeing the Global Network Potential

It is not difficult to see that the future possibilities for digital research ecosystems are rich. These ecosystems enable scholarly research on unprecedented levels and empower a global research network. Prospective projects include allowing easier international research collaborations and better aggregation as well as review and tracking of data through stronger research communities. Presently, there are 236–300 "very high" and "high" research-activity institutions in the U.S. and Canada (Carnegie Classifications I and II). Beyond North America, there are approximately 1,000–1,250 research-focused universities and institutions worldwide. Why not empower these top research institutions globally with these exceptional possibilities?



Future possibilities for 2020–2024: one server per research institution

It would be easy enough to equip research universities that do not possess such systems and configure servers with open source scholarly research software components. One server per institution should be the goal. Mirror sites can be easily set up globally with fractional server space models. Webinar training could take place over five continents and analytics later assessed. While such worldwide research infrastructure initiatives do not yet exist, it would not be difficult for one of the more forward-thinking, global grant-funding agencies that are looking for new paradigm developmental initiatives to take a leadership role.

As system and faculty/student needs expand, a web developer, project manager, metadata librarian, digitization specialist, GIS specialist, and data specialist may be added depending on institutional directions. Basic systems with digitization labs can be set up, and expansions can be managed in 1–5 year project timelines.

Assessment

A discussion of any digital ecosystem should include both assessment infrastructures and analytics. Qualitative and quantitative measures are needed to track, improve, and report results. Statistics from the Texas State University digital ecosystem (see image on page 23) show that downloads from the libraries' digital collections repository and ETDs have grown five-fold in the past 5 years. The later-implemented data repository is also on target to grow exponentially.

Both analytics and open-ended comments are essential to track your results, see where needs are, and find out where the system may be further improved. We determine faculty and student perceptions through biannual LibQual surveys. For Texas State University, user opinions have been very positive, supporting future growth and system expansion.

SUMMARY REFLECTIONS

Setting up digital scholarship ecosystems is easily achievable within most research institutions in the U.S. and Canada. All of the open source software reviewed here is mature. The network infrastructures are worked out and functioning well. The associated research collaborative communities are established and robust. Placing digital scholarship components in this ecosystem paradigm sets baseline expectations and provides a road map for future development. Locally, these types of systems open global possibilities. This new digital framework accelerates discovery for researchers worldwide and enables future progress in our now networked global village.

See Also

Uzwyshyn, R. "Digital Scholarship Research Ecosystems" (presentation PDF). Association of South Eastern Research Libraries. April 2020. aserl.org/wp-content/uploads/2011/08/DigitalScholarshipResearchEcosystems.pdf.

Uzwyshyn, R. "Developing an Open Source Digital Scholarship Ecosystem." ICEIT Conference Proceedings, Oxford University, U.K., 2020. dl.acm.org/doi/abs/10.1145/3383923.3383926.

Uzwyshyn, R. "Research Data Repositories: The What, When, Why, and How." *Computers in Libraries*, April 2016. infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml.

Ray Uzwyshyn (Ph.D., M.B.A., M.L.I.S.) is the director of collections and digital services for Texas State University's libraries.