New Opportunities for Research Libraries
Open Source & Open Access Software
and Digital Scholarly Ecosystems

Raymond Uzwyshyn, Ph.D. MBA MLIS, ruzwyshyn@gmail.com
Associate Dean, Research Library Collections and Strategy
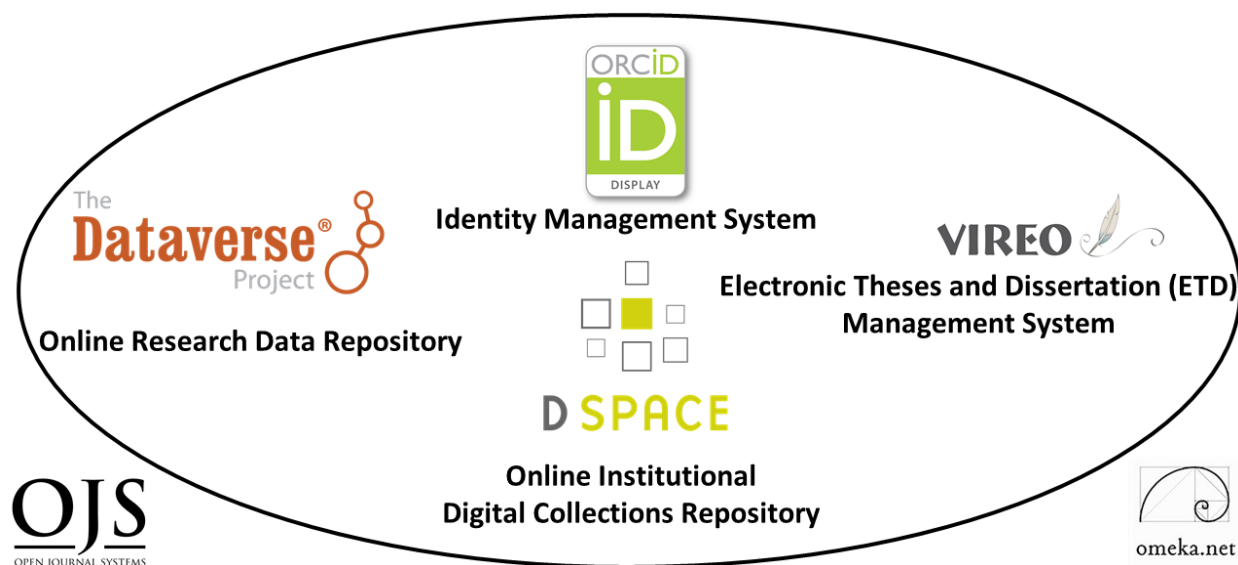Mississippi State University Libraries, USA

_____

ABSTRACT:

Leading University research libraries and institutions around the globe have begun to place academic research online through open digital ecosystems. These open source and open access infrastructures are available to any university and college. They make research more readily discoverable opening doors for collaboration and quickening further progress. This article overviews setting up these new possibilities for institutions globally wishing to enable researchers and university faculty and quickly connect them to the current state-of-the-art global network research possibilities.

INTRODUCTION: Online Digital Scholarly Research Ecosystems

A open access digital scholarly research ecosystem is an open source network of software components enabling faculty and student research, raising research profiles, accessibility and possibilities for global collaboration. This recent class of open-source software for library digital research ecosystems empowers research institutions and academic libraries globally. Open access means openly available globally. Open source means freely available. These systems are easily configurable by all university and college systems globally today.   The larger idea is that collocating open source digital components in a networked research global ecosystem creates previously untrodden connections and larger network effects towards innovation and discovery globally.

Texas State University Libraries Digital Scholarly Research System

  Characteristics of any digital scholarly research ecosystem include open source software, customizable components and active technology developer communities to customize and link components locally and globally. Texas State University Libraries digital research ecosystem consists of six main software components that are divided into content and communication.

Content: A Digital Collections Repository and a Research Data Repository;
Communication: An Identity Management System, an Electronic Thesis and Dissertation Management System, User Interface Software and Open Source Journal Software.

By collocating these open source components within a networked ecosystem, the systemic value of compatibly communicating components expands connecting researchers, systems and innovative possibility.

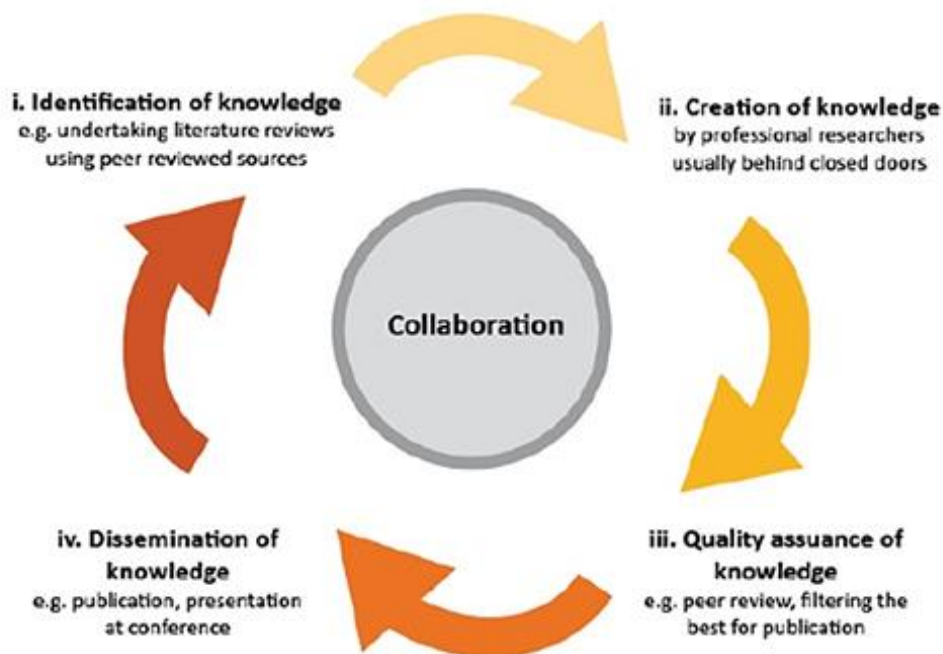**The Big Picture: Global Possibilities**

  In a global networked environment like the internet, network effects lead to large gains for researchers. This may be characterized in terms of ease of research retrieval, findability and research visibility (SEO or search engine optimization) . Research that otherwise would be buried or not easily accessible behind database paywalls on buried in long Google lists can be found more easily through library metadata application. Digital ecosystem components together provide the means for speeding up the academic research cycle (see figure below). This enables communication and collaboration for the 21st century. On pragmatic levels, digital ecosystem components empower the discovery of information: the gathering and analysis of data, online publication of research and sharing. All of this together extends and gives power to the impact of research.

The Academic Research Cycle and Enabling Power of Research Software Components

On abstract levels, digital ecosystem components enable the quality assurance of our collective open science, large data and knowledge production capacity through online peer review of data and experiments. The filtering that libraries empower through metadata application and placing on the open web's search engines allow the accurate retrieval of relevant information and research.

The academic research cycle

i. Identification of knowledge
e.g. undertaking literature reviews
using peer reviewed sources

ii. Creation of knowledge
by professional researchers
usually behind closed doors

Collaboration

iv. Dissemination of knowledge
e.g. publication, presentation
at conference

iii. Quality assuance of knowledge
e.g. peer review, filtering the
best for publication

The Academic Research Cycle: Collaboration and Knowledge Management

These systems enable the dissemination of knowledge through online publication and the identification of knowledge. This occurs through the aggregation and curation of knowledge that online Identity Management systems, Electronic Theses and Dissertation systems and data/collection repositories together allow.

Digital Research System Primary Components: Content and Communication

Preliminarily, it is important to have at least a general idea of what various digital library research ecosystem components do best. Where and why they should be used within a digital research ecosystem? This allows staff to customize and tailor ecosystem possibilities to specific institutional needs. Each of a digital research ecosystem's components serve a specific purpose and need, enabling the larger online research system.

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 and graduate students). This material includes pre-prints, faculty publications white papers, conference presentations, field notes, graduate student theses and dissertations. While a digital collections repository may be used for a spectra of media formats, it is best used for 'textual' content and depth structure linking. Depth structure linking is linking to a deeper set of documents that other researchers interested in diving deeper into a topic may wish to explore for further development, research and exploration.

D Space, Texas State University Libraries Digital Collections Repository**.**
https://digital.library.txstate.edu/

A primary use case value for an institutional collection repository is the application by library metadata catalogers or technicians of structured metadata schema for search engine optimization. This enables and heightens retrieval. Texas State Digital Collections repository utilizes the open source platform, DSpace for these purposes.

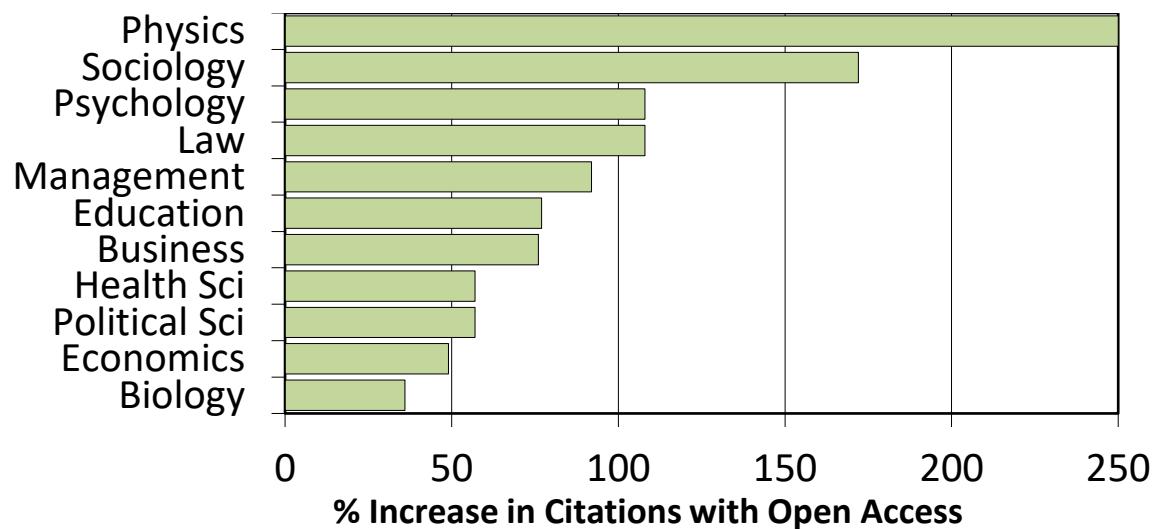| Contributor | dc.contributor.author | Donnelly, David W | |
Contributor	dc.contributor.author	Covington, B. C.	
Contributor	dc.contributor.author	Grun, J.	
Contributor	dc.contributor.author	Fischer, R.P.	
Contributor	dc.contributor.author	Peckerar, M.	
Contributor	dc.contributor.author	Felix, C. L.	
Contributor	txstate.contributor.author	Donnelly, David W., Southwest Texas State University, Dept. of Physics	
Contributor	txstate.contributor.author	Covington, B. C., Southwest Texas State University	
Contributor	txstate.contributor.author	Grun, J., Naval Research Laboratory, Washington, DC	
Contributor	txstate.contributor.author	Fischer, R.P., Naval Research Laboratory	
Contributor	txstate.contributor.author	Peckerar, M., Naval Research Laboratory	
Contributor	txstate.contributor.author	Felix, C. L., United Industries Inc.	
Date	dc.date.available	2013-07-19T16:20:49Z	
Date	dc.date.issued	2001-04-02	
Uri	dc.identifier.uri	https://digital.library.txstate.edu/handle/10877/4675	
Subject	dc.subject	"Athermal annealing", "baron implants", silicon	en_US
Title	dc.title	Athermal annealing of low-energy boron implants in silicon	en_US
Language	dc.language.iso	en_US	en_US

Dublin Core
Metadata

Access Points

Findability

**Search Engine
Optimization (SEO)**

Rich Dublin Core Metadata in a digital collections repository elevates research for Search Engine Optimization (SEO)

The application of structured metadata for textual academic research opens accessibility and multiple points of subject access. Later, these effects may translate to increased article citations through the more precise availability of relevant research through online search engines.
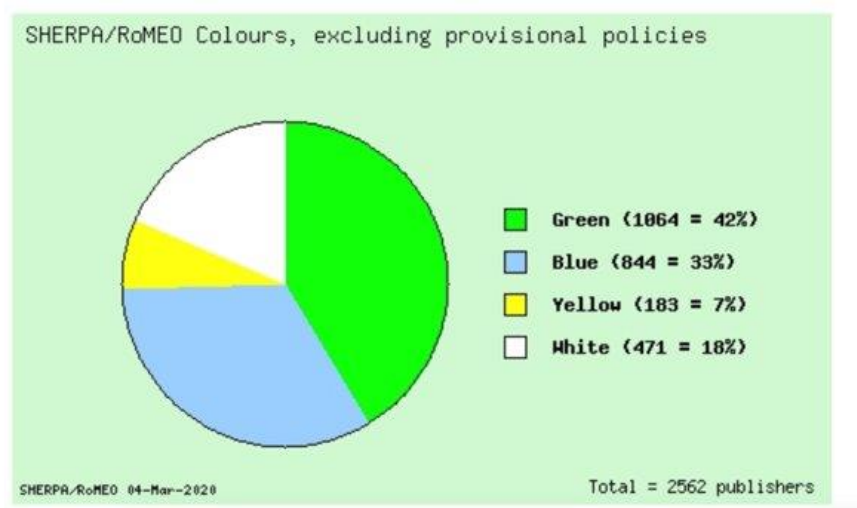
Percentage Increase in Citations with Open Access (Data: Stevan Harnad and Heather Joseph, 2014)

Most publishers today globally also allow digital archiving in one form or another. In a recent Sherpa/Romeo Copyright Policies and Self Archiving Survey, at least 82% of 2562 global publishers allowed digital archiving whether pre or post prints, final drafts or works with formal peer support.

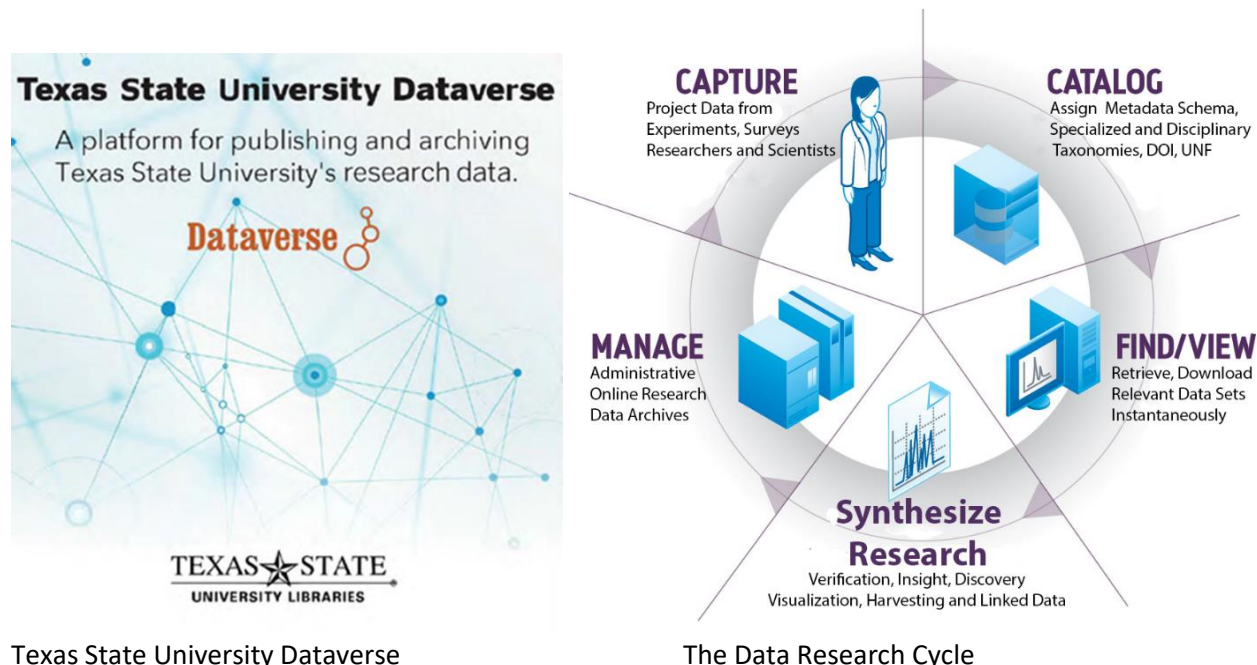| RoMEO colour	Archiving policy	Publishers	%
green	Can archive pre-print and post-print	1064	42
blue	Can archive post-print (ie final draft post-refereeing)	844	33
yellow	Can archive pre-print (ie pre-refereeing)	183	7
white	Archiving not formally supported	471	18

Summary: **82%** of publishers on this list formally **allow** some form of self-archiving.

Sherpa/Romeo/Copyright Policies and Self Archiving, 2020. Source: www.sherpa.ac.uk/romeo/statistics.php?la=en&fIDnum=|&mode=simple

Research Data Repository

A research data repository is the data analogue of a text-centered collections repository. It is specifically suited for publishing and archiving research data. A data specific repository allows a researcher to capture, upload, assign metadata schema, retrieve and download datasets.

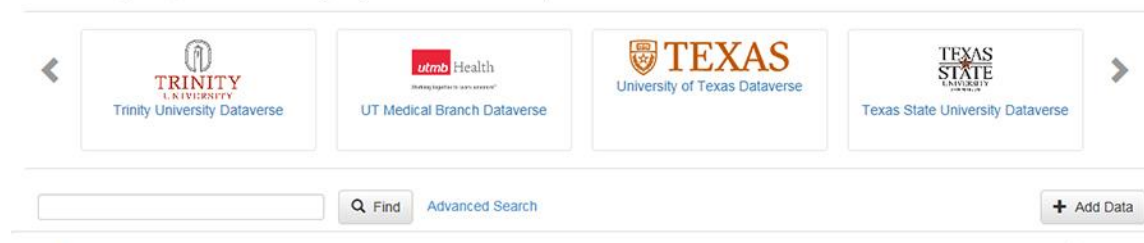Texas State University Dataverse   The Data Research Cycle

A data repository may be configured as a single instance or as a consortial model for universities that may be located throughout a state, region, country or set of countries. In the last 30 years, the rise in multi-university collaboration has increased exponentially in the sciences, engineering and social science disciplines. Texas State University Libraries is one of several state individual instances of the larger Texas Digital Library Dataverse. This state consortium allows researchers to share, publish and archive their data but also search across research for similar research from other consortium members. Opportunities for collaboration and comparison of research are enabled.

Texas State Digital Library Consortial Model Searches across Texas Universities

Secondary Digital Research System Components

Secondary components of a digital research ecosystem include Electronic Thesis and Dissertation Management and Research Identity Management Systems, User Interface Software and Open Access Journal Software.

Vireo Open Source Electronic Thesis and Dissertation Management System

Vireo ETD Management System, https://www.gradcollege.txst.edu/docs/Thesis_Diss_Submission.pdf

The Vireo ETD management system addresses intermediary steps in the Electronic Thesis and Dissertation processes. This software bridges student thesis/dissertation submission with Graduate school review, online publication and ETD preservation. The software connects graduate schools, Honors colleges, the library and library technology infrastructures. This includes the library collections repository and data repository. Students can publish and link theses and dissertations with data and other text-based research materials and workflows. Communication and workflow from deans to theses and dissertation advisors to library technologists and university copyright officers becomes seamless. Vireo is developed by the Texas Digital Library and Texas state universities libraries and is open source software freely available to all institutions.

VIREO   THESIS & DISSERTATION
                                                                           SUBMITTAL SYSTEM
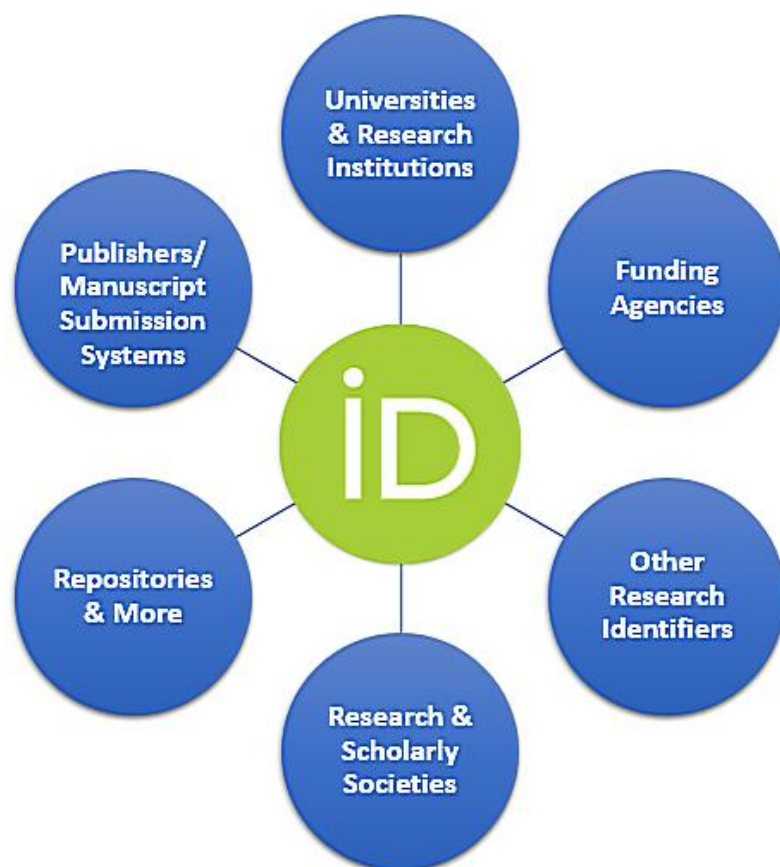
Welcome to Vireo the Thesis & Dissertation Submission System at Texas State University

Once your committee has approved your final thesis/dissertation and signed the Thesis/Dissertation Committee Approval Form, you are ready to submit your document to the Graduate College for review. This submission process is fully electronic, and is made through an online application developed and maintained by the Texas Digital Library, in conjunction with the Texas A&M, MIT, and UIUC.

To get started with your submission, click the link below. You will be asked to authenticate using your Texas State NetID:
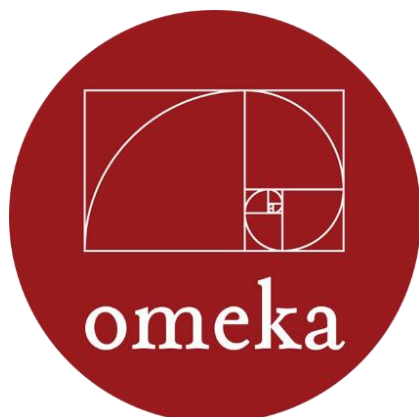
Login

ORCID: Online Research Identity Management System

ORCID is a hub connecting the research landscape

ORCID Online Identity Management System, https://orcid.org/

ORCID is a widely utilized Research Identity Management system which gives researchers a unique number, an ORCID ID to connect. This 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 ID's for aggregation of research profiles. ORCID itself may also act as a hub connecting the research landscape. This hub may act as a network in a network. This serves to aggregate from several sources and connect researchers on wider levels. It is very useful, especially to a region, country or state as it is able to disambiguate common names in regions. For example the biochemist Ang Lee may be differentiated from the mechanical engineer Ang Lee and their respective research may be easily obtained and aggregated.

OMEKA: Open Source User Interface Software

https://omeka.org/

Omeka open source User Interface Software is used in a digital research ecosystem to allow an elegant portal or gateway entrance for larger research projects, digital collections and data repositories. Texas State uses Omeka linking research, text, images, media and research datasets. Omeka acts as a front end elegant user interface to connect component networks. Omeka can be configured for both individual projects and as a sharable resource across multiple sites. This class of software provides a publishing platform for sharing digital research and creating multimedia rich front-end online exhibits linking to the previously described larger depth structure of backend data and deep collection digital library repositories, further research and online archives of source materials housed in the institutional repository of Dspace. A similar open source product that may also be noted here, Islandora.

OJS Open Journal Systems Academic Journal Software

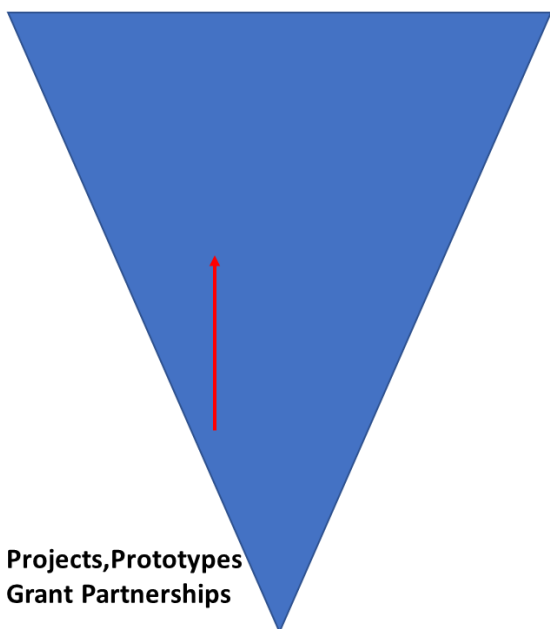**https://openjournalsystems.com/ojs3/**

Open access Academic Journal Software is similarly used to its physical counterparts. This software is best used for refereed journal workflow for online publishing articles and linking experimental data sets with source research articles. In addition, online journal management software allow deeper connections to source data and background research to provide links and further depth of content to primary research sources housed in collection and data repositories. In this way, it allows reproducibility, transparency and integrity towards the larger research enterprise.

The Digitization Lab

As digital ecosystem software needs are met a digitization lab and associated hardware can also expand possibilities for faculty and graduate student research projects from the digitization of books, manuscripts and journals to audiovisual and visualization digitization technologies. Combining research ecosystem components opens amazing possibilties for digital scholarship and research collaboration opportunities on a number of levels (See Diagram)

Cognitive Cartography/Multimedia Archives
(Video, Text, GIS, Images, Field Notes)
Dick Reavis: National Tour of Texas

Multimedia, Digital Archives/Retrospective ETD Projects
(Digital video, online exhibit images, text, digital archives)
Severo Perez: And the Earth Did not Swallow Them

Online Exhibits/Digital Archives/ Online Academic Journals
(images and text, Omeka front end/Database back end, IIIF)
Cabeza de Vaca La Relacion Digitization
 Santiago Tafolla: Mexican Amer. Confederate Soldier
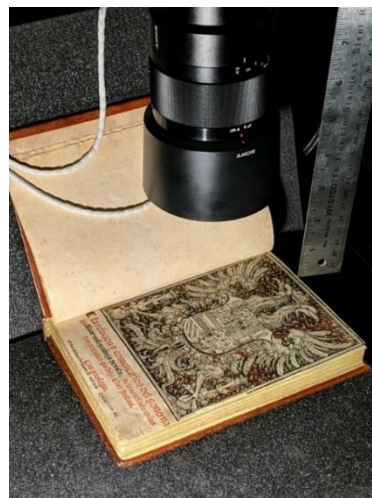
Interactive Image Archives/Data & Research Projects
(Image libraries, Interactive Commenting/Metadata)
Texas State Flickr Commons

Digital Libraries Archiving & Documentation Projects
(Text, Metadata, OCR, Search, Zoom ability, Page Turning)
 Pedagogs University Yearbooks

Faculty Digitization Proposals/Partnerships

Projects,Prototypes Grant Partnerships

Digitization projects and Digital Research Ecosystems

The Digitization Lab Opens Various Technologies and Possibilities.

Implementation Paths: Evolution of Library Infrastructures

While the list of software and projects may begin to sound daunting, a human resource infrastructure for a digital research ecosystem may be set up with as little as two staff. 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 softwares. This includes researcher support and marketing and instruction on chosen systems. Every institution or library will also have unique research needs and focus (i.e. textual content, data, multimedia, dissertation archiving etc.).

Implementation and human resource infrastructures should be strongly tied to insitutional directions and profiles. A research focused science and engineering university will have different needs from a liberal arts college. Software should be chosen accordingly. As the system and research faculty/student needs expand a web developer, project manager, metadata librarian, digitization specialist, GIS Specialist and Data specialist may be added. Basic systems with digitization labs can be set up and planned expansions managed in one to five year project timelines.
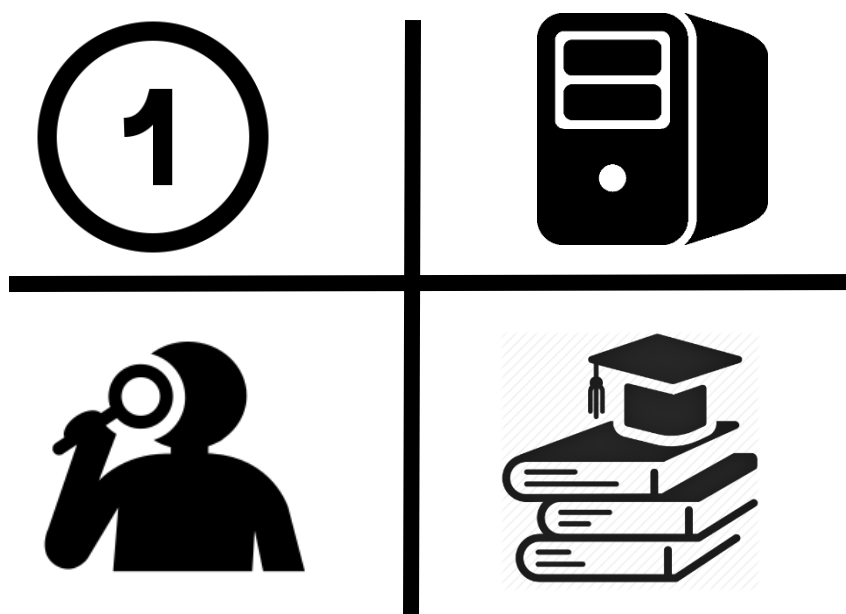
Assessment and Results

A central discussion of any digital ecosystem also surrounds assessment. Both qualitative and quantitative measures are needed to report out, track and improve results. Taking the example of annual Texas State Digital ecosystem statistics below, downloads from the libraries' digital collections repository and ETD's have grown five-fold in the past five years. The later implemented Data Repository is also on target to grow exponentially. Faculty and student perceptions can also be tracked through qualitative perceptions through biannual LibQual surveys. Analytics and comments will be essential to track results, see where needs are and where the system may yet be further improved.

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

Example Annual Digital Ecosystem Usage Growth: Downloads, Number of Items, ORCID ID's and Hosted Journals

Future Possibilities

It is not difficult to see that future possibilities for digital research ecosystems are rich. These ecosystems enable scholarly research on unprecedented network levels. They empower a global research community and connect universities, regions, countries and global areas. Prospective projects range from enabling easier international research collaborations to better aggregation, review and tracking of data through stronger global research networks.   Presently there are two hundred and sixty-six to three hundred very high and high research activity institutions in the US and Canada (Carnegie Classification I & II). Beyond North American borders there are approximately 1000-1250 Research focused universities and Institutions worldwide. Currently, less than the top 0.5% of global institutions possess all of these open source ecosystem components. Why not at least enable and empower the top 2-3% or 1000 research institutions globally with these exceptional possibilities?

One Digital Library Scholarly Ecosystem Server Per Research Institution: 2022-2027 Future Possibilities

Servers may be configured with open source scholarly research software components. Mirror sites around our global village may be set up with fractional server space models. Webinar training could take place over five continent and analytics later assessed. While such global research infrastructure initiatives do not yet exist, this would not be difficult for more forward thinking agencies or institutions looking for worthy new millennia far reaching developmental initiatives. This would not be overly prohibitive for even larger areas or zones such as North America, Europe, Eurasia or South East Asia and the Middle East. The technology is available, open source and open access for those with the desire, needs and acumen.

Conclusions

Digital scholarship research ecosystems are currently being set up increasingly in research library institutions in the US and Canada. The open source software overviewed here is mature. The network infrastructures well worked out and functioning. The associated library and university research communities necessary are also robust and in place around the globe. Placing digital scholarship components within an ecosystem paradigm also sets both baseline and roadmap for the further development of research scholarship. These new models enable a successfully tested core digital research paradigm for further evolutionary possibilities and development. These types of systems open a range of global possibilities including accelerating discovery for researchers and enabling future progress of knowledge in our new millennia.

Bibliography and Further Sources

Open Source/Open Access Software Links and Downloads
Dspace https://duraspace.org/dspace/
Dataverse https://dataverse.org/
Omeka https://omeka.org/
Open Journal Systems https://pkp.sfu.ca/ojs/
ORCID https://orcid.org/
Vireo https://www.tdl.org/etds/
Further Background, Articles and Examples
Uzwyshyn, R. Online Research Data Repositories: The What, When, Why and How. Computers in Libraries. 36:3, April 2016. pp. 18-21. Uzwyshyn, R. Developing an Open Source Digital Scholarship Ecosystem. ICEIT Conference Proceedings, Oxford University , United Kingdom, 2020. https://dl.acm.org/doi/abs/10.1145/3383923.3383926
Texas State University Libraries Data Repository & Digital Ecosystem Example Set-Ups Texas State Data Research Repository Texas State University Libraries Digital Scholarship Ecosystem. Texas State Digital Collections Repository Texas State Online Research Identity Management System (ORCID) Open Journal Systems @ Texas State