

U.S. Digital Preservation Collaborations from 1994-2014 State-by-State Overviews

Christina Drummond

Katherine Skinner

2015 January 28

Version 1.0

Publication Notes:

Title: U.S. Digital Preservation Collaborations from 1994-2014: A State-by-State Overview

Authors: Christina Drummond and Katherine Skinner

Publisher: Educopia Institute, 1230 Peachtree Street, Suite 1900, Atlanta, GA 30309.

Copyright: 2015

This publication and the data referenced herein are covered by the following Creative Commons License:

Attribution 4.0 International (CC BY 4.0)

You are free to copy, distribute, and adapt this work under the following conditions:



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The above is a human-readable summary of the full license, which is available at the following URL:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Acknowledgements

We wish to thank members of the research team and the NDIIPP and NDSA representatives (2012-14) who provided editorial feedback throughout this research project (especially Michelle Gallinger and Abby Potter). We also would like to thank our team of project advisors and consultants, including Liz Bishoff, Matt Cooper, Aaron Trehub, and Christopher (Cal) Lee, whose contributions were vital to the success of the project. Finally, we especially would like to thank Anjum Najmi and Tre Wilson for their data collection efforts and Nick Krabbenhoeft for tireless hours of data indexing and scrubbing, without which we would not have quality data to underscore this report.

Our analysis is based on Version 1.0 of the ICONC U.S. Digital Preservation Collaborations 1994-2014 Dataset. Any omissions or oversights within the ICONC dataset and this report are entirely our own. We recognize there has been a vast amount of activity in the past decade, and we look forward to working with the greater digital preservation community to continue building upon this data in the future. We hope this data will be used to foster a greater understanding of the diversity of perspectives, voices, and activities underway today—and that that understanding will help us to build stronger partnerships across a national network of practitioners.

Christina Drummond and Katherine Skinner

Related Datasets and Interactive Visualizations

ICONC source data was collected, categorized, and published as a part of the ICONC research project. The resulting datasets are integrated with multiple interactive visualizations to allow users to create their own views of the data with user-defined subgroups of organizations or activities. Interested parties can download a .zip file including the raw dataset worksheet with accompanying pivot table report worksheets (.xls), the data entry specifications codebook that governed data inclusion and coding, and the Creative Commons license governing this dataset. This raw data package for this project is published on Educopia's website under the following Creative Commons License:

Attribution- 4.0 International (CC-BY 4.0), and is accessible at <http://educopia.org/research/iconc>.

To facilitate and encourage broader community use of the data, interactive data visualizations were also created and published within the Tableau Public platform. These interactive visualizations allow users to click or manually filter views to measure and view the collaborative organizations and efforts within the ICONC dataset. The ICONC data and related Version 1.0 ICONC dashboards are thereby also available on the Tableau Public platform at <http://educopia.org/deliverables/pilot-iconc-dashboards>.

TABLE OF CONTENTS

1	Background	1
1.1	Data Inclusion.....	2
1.1.1	Activities	2
1.1.2	Organizational Collaborators	2
1.2	Information Sources	3
1.3	Dataset Peculiarities.....	5
2	Digital Preservation: Trends and Highlights.....	6
2.1	Active Collaborative Digital Preservation Activities	6
2.1.1	Activities with Multi-state Engagement.....	6
2.2	Digital Preservation Efforts: 1994-2014.....	7
2.2.1	Subset of Activities Benefiting from NDIIPP Funding	7
2.3	Organizational Diversity of Collaborators.....	8
2.3.1	Organizational Engagement by Activity Type	8
2.3.2	Sector Engagement by Geographic Region	9
3	Regional Activity Highlights	12
3.1	NORTHEAST - New England Region (Census Region1-Division1)	12
3.1.1	Regional Summary	12
3.1.2	Maine	12
3.1.3	New Hampshire.....	12
3.1.4	Vermont.....	12
3.1.5	Massachusetts.....	12
3.1.6	Rhode Island.....	12
3.1.7	Connecticut.....	13
3.2	NORTHEAST - Mid-Atlantic (Census Region1-Division2)	13
3.2.1	Regional Summary	13
3.2.2	New York	13
3.2.3	Pennsylvania.....	13
3.2.4	New Jersey.....	14
3.3	MIDWEST - East North Central (Census Region2-Division3)	14
3.3.1	Regional Summary	14

3.3.2	Wisconsin	14
3.3.3	Michigan.....	14
3.3.4	Illinois	14
3.3.5	Indiana.....	15
3.3.6	Ohio	15
3.4	MIDWEST - West North Central (Census Region2-Division4)	15
3.4.1	Regional Summary	15
3.4.2	Missouri.....	15
3.4.3	North Dakota.....	15
3.4.4	South Dakota.....	15
3.4.5	Nebraska	16
3.4.6	Kansas	16
3.4.7	Minnesota	16
3.4.8	Iowa.....	16
3.5	SOUTH - South Atlantic (Census Region3-Division5)	16
3.5.1	Regional Summary	16
3.5.2	Delaware	16
3.5.3	Maryland	17
3.5.4	District of Columbia.....	17
3.5.5	Virginia	17
3.5.6	West Virginia	17
3.5.7	North Carolina	17
3.5.8	South Carolina	17
3.5.9	Georgia.....	17
3.5.10	Florida	18
3.6	SOUTH - East South Central (Census Region3-Division6)	18
3.6.1	Regional Summary	18
3.6.2	Kentucky	18
3.6.3	Tennessee	18
3.6.4	Mississippi	18
3.6.5	Alabama	19
3.7	SOUTH - West South Central (Census Region3-Division7).....	19

3.7.1	<i>Regional Summary</i>	19
3.7.2	<i>Oklahoma</i>	19
3.7.3	<i>Texas</i>	19
3.7.4	<i>Arkansas</i>	19
3.7.5	<i>Louisiana</i>	19
3.8	<i>WEST - Mountain (Census Region4-Division8)</i>	20
3.8.1	<i>Regional Summary</i>	20
3.8.2	<i>Idaho</i>	20
3.8.3	<i>Montana</i>	20
3.8.4	<i>Wyoming</i>	20
3.8.5	<i>Nevada</i>	20
3.8.6	<i>Utah</i>	20
3.8.7	<i>Colorado</i>	20
3.8.8	<i>Arizona</i>	21
3.8.9	<i>New Mexico</i>	21
3.9	<i>WEST - Pacific (Census Region4-Division9)</i>	21
3.9.1	<i>Regional Summary</i>	21
3.9.2	<i>Alaska</i>	21
3.9.3	<i>Washington</i>	22
3.9.4	<i>Oregon</i>	22
3.9.5	<i>California</i>	22
3.9.6	<i>Hawaii</i>	22
4	<i>Future Opportunities</i>	23
4.1	<i>Engage and Build Upon Priority Activity Footprints</i>	23
4.1.1	<i>Digital Preservation Training and Knowledge Sharing</i>	24
4.2	<i>Connect, Strengthen, and Build Upon Existing Sector Associations</i>	26
4.3	<i>Develop Strategic Informatics Tools</i>	27
5	<i>Appendix: Data Entry Specifications</i>	29
	<i>ICONC Project Source Data Entry Specifications</i>	29
	<i>Document Introduction</i>	29
	<i>Activity Related Field Definitions</i>	30
	<i>Activity Related Field Definitions Cont.</i>	31

<i>Activity Related Field Definitions Cont.</i>	32
<i>Activity Related Field Definitions Cont.</i>	33
<i>Organization Related Field Definitions</i>	34
<i>Organization Related Field Definitions Cont.</i>	35
<i>Digital Preservation Network Related Field Definitions</i>	36

1 Background

In 2012, the Educopia Institute was approached by the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) to identify and document collaborative digital preservation activities in the U.S., across relevant memory sectors (e.g., academic, government, archival, museum, nonprofit, and commercial). Using these "state profiles", Educopia was also asked to help NDIIPP identify future opportunities for collaborative activities. The Educopia Institute proposed and was funded by a Cooperative Agreement to conduct this research as part of the Identifying Continuing Opportunities for National Collaboration (ICONC) project (September 2012-December 2013).

As Educopia scoped this research, the project team determined that the data it would gather might have multiple use cases within the memory community and beyond. A well-constructed relational dataset resulted, comprised of information about collaborative digital preservation activities and the domestic organizational collaborators behind such activities. Activity data captured included important information such as start and end dates, descriptions, NDIIPP funding status, URLs, and founding or host institutions. Data about collaborating organizations included their general location, sector, and organizational focus. Most importantly, the relationship between activities and organizations was preserved, so that one could view both the activities a specific organization is involved in and the organizations involved in a given activity.

Rather than producing a stand-alone report, the ICONC team has produced three deliverables. First, an open (yet beta) relational dataset contains 211 collaborative digital preservation activities and 1856 related collaborating organizations to support a wide range of research questions. Second, a set of initial interactive data-driven dashboards serve to allow NDIIPP and other groups to begin viewing custom defined subsets of the data underlying this report in maps, bubble charts, and other graphical means. Third, this narrative overview both highlights research findings and activities underway in each of the 50 states and the District of Columbia, and recommends future directions based on the initial analyses of the 3,298 records in this dataset.

Due to time and budget constraints, this dataset is only a pilot, and should be thought of as a beta version. While the aim was to be exhaustive in capturing activities and collaborators, due to the sheer volume of information and the manual data mining, items may not yet be reflected in this dataset. There are also two specific areas involving taxonomy classifications where the pilot dataset can be refined. To facilitate searching for collaborators hailing from a particular industry sector (e.g. video production or news media), preliminary organizational coding occurred. However, taxonomy development requires more time and community feedback than what was available within the pilot period. The initial basic taxonomy classifying the activities within the dataset¹ should also be revisited and refined by the

¹ Activities were manually classified as one of the following eight activity types: Research, Digital Preservation Network, Open Repository/Digital Library, Closed Repository, Conference, Training, Service/Software, Other

broader community, adding additional information about the targeted users or beneficiaries of specific activities when known.

1.1 Data Inclusion

1.1.1 Activities

The pilot dataset spans the years 1994-2014. For an activity to be included within the dataset, it had to meet three criteria.

- 1) Collaborative: The activity must have had more than one institution collaborating on producing—and not merely benefiting from or using—the activity.
- 2) Stated Digital Preservation Aim: The activity must have had a preservation aim clearly stated within publically available materials. Efforts that indirectly involve preservation, or that clearly did not exist for preservation purposes were excluded.
- 3) U.S. Based Collaborator: While a wealth of digital preservation activities occur internationally, this pilot dataset has an inherent U.S. focus. Only those activities that include at least one U.S.-based collaborator are represented within this dataset. If domestic organizations were involved in international efforts, both those efforts and their domestic collaborators are shown. Of the total 1856 organizations within the pilot dataset, 582 are international collaborators that are omitted from the following analyses and related data-driven dashboards.

Both ongoing and concluded activities are represented within the dataset.

The collaborations represented within the dataset include a wide range of activity types. To facilitate analysis, activities were labeled with one of eight activity type descriptive categories: software/service, research, preservation network, training and knowledge sharing, standards and policies, advocacy, conferences, and digital archives/repositories.

1.1.2 Organizational Collaborators

To be included in the pilot dataset, an organization or its representative had to be on a public list of collaborators for an activity that met the above criteria. International collaborators fell outside the scope of this pilot and are not referenced in this overview or the accompanying data dashboards. For inclusion, organizations also had to be *collaborating* on the production of the effort or activity and not merely acting as a beneficiary, user, or member. The organizational collaborators included in the dataset represent a range of players and roles: host organizations, founding organizations, partner organizations, and the organizational homes of those involved in planning or steering committees. Home organizations of speakers and attendees at conferences or other events are not captured within the dataset. Similarly, organizational users of products or services are excluded. Finally, it is important to note that because every individual person collaborating on a project is represented in aggregate under their organizational home, the dataset does not represent how much of an organization is engaged in a given collaborative activity, e.g. whether one person or an entire department is involved.

1.2 Information Sources

The dataset was compiled through manual data mining of the following published reports and publically available websites.² Data collection requirements and data coding guidelines were used to ensure consistent data handling. (Appendix A)

Published Reports

- *Humanities Collections and Reference Resources - an evaluation 2000–2010*. Division of Preservation and Access, National Endowment for the Humanities. (2013)³
- *The State Electronic Records Initiative - Phase I Report from the Council of State Archivists' State Electronic Records Initiative (SERI) Committee*. (2012)⁴
- *States of Sustainability: A Review of State Projects funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP)*. Christopher Lee. (2012)⁵
- *Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report*. Library of Congress. (2010)⁶
- *Status of the Preservation of Electronic Records by State Archives Report*. Committee on Electronic Records and Information Systems (CERIS) (2004).⁷
- *Connecting the Archival Community - Final Report of the Archival Education and Information Web Needs Assessment Project*. (2002).⁸

Websites and Published Data

- Library of Congress' Digital Preservation Partners⁹, NDIIPP Initiatives¹⁰, and Digital Preservation Tools¹¹ sites

² A number of additional sources were reviewed without resulting in any data records, such as CoSA's *The State of State Records: A Statistical Report on State Archives and Records Management Programs in the United States* (June 2013) (<http://www.statearchivists.org/reports/index.htm>) and the NHPRC 2003 *Electronic Records Agenda Report* (<http://www.mnhs.org/preserve/records/eragenda.html>).

³ http://www.neh.gov/files/divisions/preservation/hcrr_evaluation_report_2013.pdf

⁴ <http://www.statearchivists.org/seri/SERI%20Phase%20One%20Report%20-%20final%20review%20draft%20-%202012-06.pdf>

⁵ http://www.digitalpreservation.gov/multimedia/documents/ndiipp-states-report032612_final.pdf)

⁶ http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf

⁷ <http://www.nagara.org/associations/5924/files/CERIS%20Survey%20Report.pdf>

⁸ <http://www.statearchivists.org/reports/WebNA2002/WebNAExecSumm.pdf>

⁹ <http://www.digitalpreservation.gov/partners/>

¹⁰ <http://www.digitalpreservation.gov/about/initiatives.html>

¹¹ <http://www.digitalpreservation.gov/tools/>

- Conference websites for the Best Practices Exchange¹², iPres¹³
- Websites for the State Archivists Association (SAA), the State Electronic Records Initiative (SERI), the National Association for State Chief Information Officers (NASCIO), the National Association of Government Archive and Records Administrators (NAGARA), the Legal Information Preservation Alliance¹⁴, OCLC Research¹⁵, and the Repository Exchange¹⁶
- The LOCKSS Global & Private LOCKSS Networks site¹⁷
- Consortial IR Instances. Orbis Cascade Alliance.¹⁸

Surveys: The research team partnered with the National Association of State Chief Information Officers (NASCIO) and the National Association of Government Archives and Records Administrators (NAGARA) to query their memberships on existing collaborative projects.

Grant awards: The research team also manually mined data from grant award information posted by IMLS, NEH, NHPRC, NSF (with a focus on DataNet), NIH, DOE, and the Andrew W. Mellon Foundation. Search terms used included the following: “digital preservation”, “records retention”, “long term preservation”, “digital stewardship”, “digital curation”, “Digital asset management”, “research data management”, and “research data curation.”

All data required a publically accessible URL source to validate the data¹⁹. For historical data, the Internet Archive’s WayBack Machine²⁰ was utilized to find preserved copies of historical websites.

Due to the limited nature of this pilot, organizational classification taxonomies require further refinement to allow for stronger analysis. Additionally, the research team recognizes that given the sheer volume of data, it is possible that collaborative, digital preservation efforts may yet be missing from this initial pilot data set.

¹² <http://www.bpexchange.org/>

¹³ E.g. <http://ipres2013.ist.utl.pt/>

¹⁴ <http://lipalliance.org/resources/state-initiatives/>

¹⁵ <http://www.oclc.org/research/>

¹⁶ <http://www.re3data.org/>

¹⁷ <http://www.lockss.org/community/networks/>

¹⁸ http://webcache.googleusercontent.com/search?q=cache:HppqQl6gyOM4J:www.orbiscascade.org/index/cms-filesystem-action/groups/irtf/ir_shared_repositories.xlsx+&cd=2&hl=en&ct=clnk&gl=us

¹⁹ This public URL requirement means that the dataset exclusively represent efforts with open collaborator lists, to the exclusion of efforts like the Storage Networking Industry Association, Long Term Retention working group (<http://www.snia.org/ltr>).

²⁰ <http://archive.org/web/>

1.3 Dataset Peculiarities

The research team made several structural decisions when compiling the dataset that impact the dashboards and visualizations. These decisions on how the data was captured and represented in the dataset, and the resulting impacts on the visualizations are noted below.

To allow for future analysis into academic engagement, parent organizations such as a university system are represented as individual subunits. For example, if both a library and an information school of a university were involved in collaborations, each unit is reflected as an independent organization for the purposes of this report. Therefore, a simple count of organizations within a given state could be an invalid measure of institutional involvement if a number of subunits are involved at participating institutions.

When reviewing activity level data, it is also important to remember that a single organization may have participated in multiple activities. For this reason, organizations may be represented multiple times within a single graphic in this report, if they were involved in multiple activities. (For an example, see Figure 4.)

Activities do not have singular locations, but instead are represented by networks of collaborating organizations nodes. Such network maps may be viewed for a particular activity within the dashboards that accompany this report.

2 Digital Preservation: Trends and Highlights

While there is a wealth and depth of information that can be gleaned from examining only currently active initiatives, the majority of this information is left within the interactive data-dashboards for on-demand custom querying. The following analyses within this report primarily represent the entire dataset, providing an overview given the past 20 years. Exceptions to this are noted within the text below.

2.1 Active Collaborative Digital Preservation Activities

The dataset includes 106 collaborative digital preservation activities that were **ongoing** as of March 1, 2014, with a related 1,064 collaborating organizations. Included within this are:

- 24 software or service activities, engaging 128 collaborating organizations;
- 21 digital archives or repositories, engaging 340 collaborating organizations;
- 18 research activities, engaging 180 collaborating organizations;
- 15 standard or policy development efforts, engaging 60 collaborating organizations;
- 14 digital preservation networks, engaging 310 collaborating organizations; and
- 10 training or knowledge sharing efforts, engaging 281 collaborating organizations.

Figure 1: ACTIVE DIGITAL PRESERVATION ACTIVITIES



2.1.1 Activities with Multi-state Engagement

Only seven activities currently engage collaborators in over half of the U.S. states. Of these, Phase 1 of the State Electronic Records Initiative (SERI) has collaborators in 49 states, and the NDSA has organizational members in 43 states. The Legal Information Preservation Alliance has the next largest geographic spread, engaging organizations across 37 states.

2.2 Digital Preservation Efforts: 1994-2014

The pilot dataset includes both ongoing initiatives and historical efforts that ceased operations over the past decade. Inclusive of both ongoing and historical activities, there are 210 activities in the dataset with 1,276 related domestic collaborating organizations. The number of research-oriented collaborations show the most dramatic reduction over time, while collaborations around two areas—standards and policy development, and digital archives and repositories—markedly increased.

Figure 2: 1994-2014 Collaborative Digital Preservation Activities



2.2.1 Subset of Activities Benefiting from NDIIPP Funding

This dataset reflects 42 NDIIPP funded projects that clearly engaged 390 collaborating organizations, plus an additional 171 collaborative digital preservation activities that did not show any clear indication of NDIIPP funding. Just over half of all activities are ongoing in nature. Overall, slightly more of those that received NDIIPP funding continue today under the same name (57%), as compared to those that did not receive NDIIPP funds (52%).²¹

However, a more nuanced story is hidden within the project continuation rates for activities that began between 2001 and 2009, when the bulk of NDIIPP direct appropriations for digital preservation projects were expended.²² (Figure 3) The NDIIPP funded efforts for training and knowledge sharing, standard and policy development, and software/service development projects all continue today. For digital preservation research efforts, 42% of the NDIIPP supported activities continue today, compared against 22% of those funded through other means. Remarkably, only 44% of the NDIIPP supported preservation networks continue today, which is half the continuation rate of preservation networks funded through other means (88%). Similarly, only 43% of NDIIPP supported digital archive or repository projects are ongoing, compared to an 85% continuation rate for collaborative digital archives or repository projects that started between 2001 and 2009 without NDIIPP support.

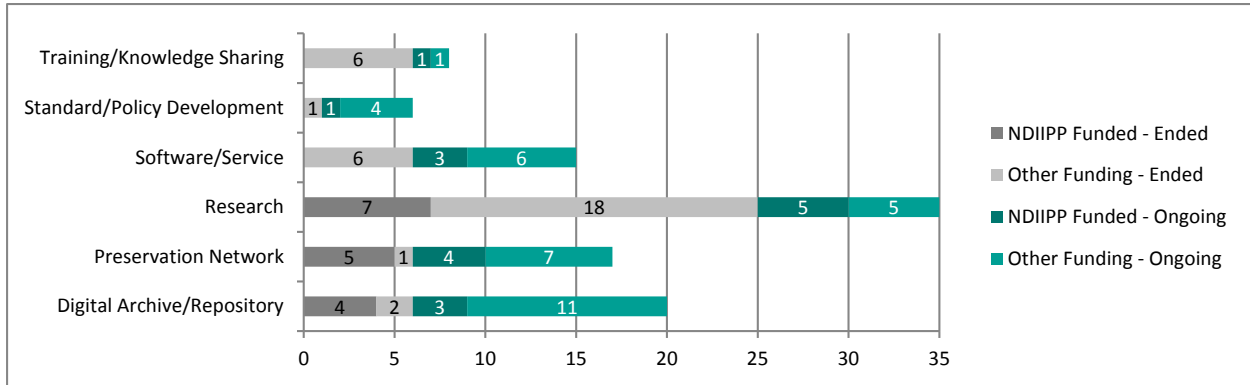
These continuation rate comparisons are intriguing. If one considers research and development activities (including standards development) as temporal in nature, it would be appropriate to see fewer of these activities continue over time, resulting in lower continuation rates. Further examination of the 35 research activities might explain why 42% of the NDIIPP funded activities continue today. It's even more critical to understand the success stories behind the high continuation rates of the non-NDIIPP funded archives, repositories and preservation networks, as compared to their NDIIPP counterparts.

²¹ It was outside of the scope of this pilot to track project lineage, noting when initiatives evolve from or split out of related projects. Such data would allow a better approximation of funding impact. The mere fact that a project ends only tells part of the story.

²² Advocacy and Conference activities were omitted from this analysis as no such activities were funded by NDIIPP between 2001-2009

Given the notable difference between continuation rates, further research is recommended to identify factors and practices that positively influenced the non-NDIIPP supported preservation network and digital archive/repositories sustainability. There is an opportunity to create a companion report to Christopher Lee's 2012 *States of Sustainability Report* that can similarly go in-depth to uncover the secrets to on-going success within these additional initiatives.

Figure 3: Continuation Status of Collaborative Digital Preservation Activities that began between 2001-2009



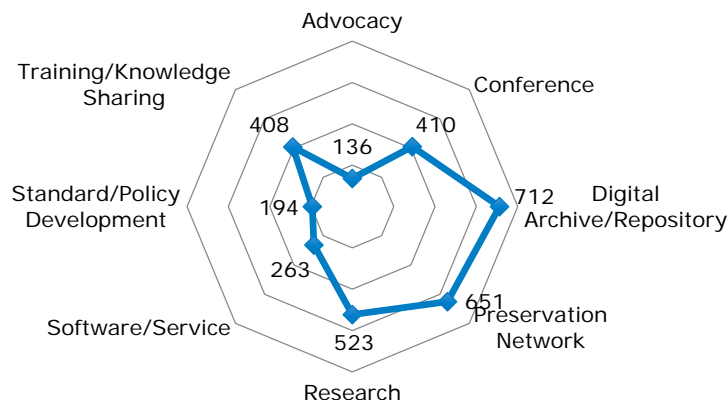
2.3 Organizational Diversity of Collaborators

The 1,276 domestic organizations represented within the dataset hail from multiple sectors. Just over half of these collaborators represented academic organizations (709). Nonprofit organizations make up 13% of represented collaborators (169), while the 143 private enterprises represent 11%. Federal government agencies represent 4% of the collaborating organizations, with involvement spanning across 50 agency sites (engaging agencies such as the Department of Energy, the National Institute of Justice, and National Park Service.) The 23 local governmental entities were engaged across 16 activities, most commonly training and knowledge sharing. Due to engagement in efforts such as the State Electronic Records Initiative (SERI), 157 state government entities make up 12% of the dataset. Finally, 2% of the organizations were themselves collaborative in nature, not clearly defined by a single sector (23).

2.3.1 Organizational Engagement by Activity Type

The highest organizational engagement levels within this dataset are within digital archive and/or repository efforts (712 organizations) and preservation networks (651 organizations). Advocacy and standard and/or policy development activities engaged the fewest collaborating organizations. (Figure 4)

Figure 4: Number of Organizations Engaged by Activity Type



2.3.2 Sector Engagement by Geographic Region

Academic organizations make up the majority of collaborators in all regions, to varying degrees. (Figure 5) Yet, nonprofit and private enterprise engagement also has occurred across all US geographic regions.

Private enterprise engagement was heaviest in the Northeast and West of the country (Figure 6); although, no private enterprise collaborators were based in either Alaska or Hawaii.

Nonprofit engagement is most heavily concentrated in the Northeast. (Figures 7-9)

Academic collaborators are most heavily concentrated in the Eastern half of the country and along the West Coast, with clusters in Texas, Utah and Colorado representing notable exceptions. (Figures 10-12)

Figure 5: Regional Sector Diversity among Collaborating Organizations

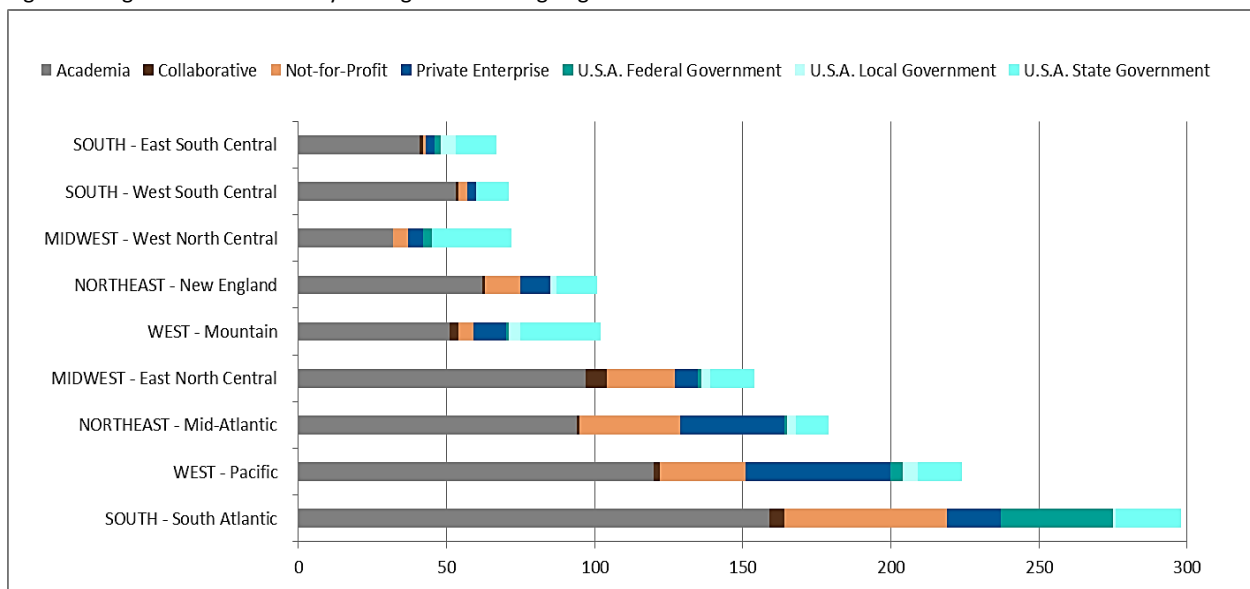
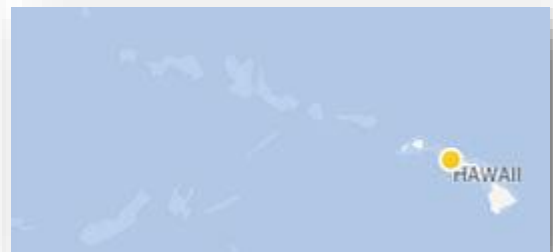


Figure 6: Geographic Distribution of Private Enterprises within Dataset



Figures 7-9: Geographic Distribution of Not-for-Profit Organizations



Figures 10-12: Geographic Distribution of Academic Organizations within Dataset



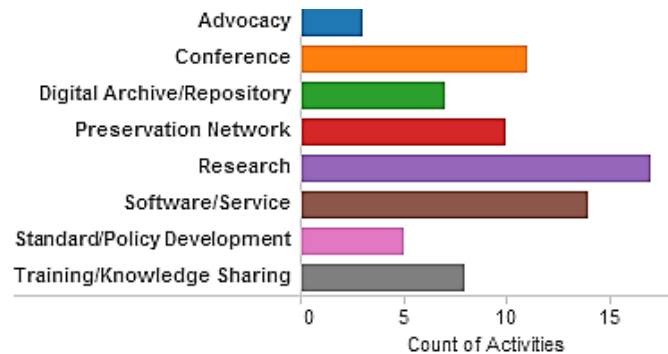
3 Regional Activity Highlights

3.1 NORTHEAST - New England Region²³ (Census Region1-Division1)

3.1.1 Regional Summary

Within our dataset, New England was the geographical host to 101 collaborating organizations, who engaged in 75 unique activities.

Figure 13: Spread of 75 Activities in New England



3.1.2 Maine

Organizational collaborators hail only from the southern half of the state, with 9 organizations participating in 8 activities. The Maine State Archives, the Maine Office of GIS, and the University of Maine - Stillwater play the most active roles, participating in four, two, and two activities respectively. Five organizations are involved in either the CLOCKSS or GeoMAPP digital repository efforts.

3.1.3 New Hampshire

The three collaborating organizations in New Hampshire are exclusively involved in digital preservation networks, ranging from the LOCKSS Alliance, to HathiTrust, to the Digital Preservation Network initiative (DPN). All continue to act as a member or partner to one of these networks (or two in the case of Dartmouth College).

3.1.4 Vermont

Five activities engaged nine organizations across the state. Of the nine, only the Vermont State Archives and Records Administration and the Vermont State Department of Information and Innovation have been involved in multiple efforts. Seven organizations were involved in *A Model Technological and Social Architecture for the Preservation of State Government Digital Information* project, which ended in 2012.

3.1.5 Massachusetts

While 60 collaborating organizations hailed from across the state, much of the involvement is heavily concentrated around the Boston region. The 60 unique activities taking place spanned all activity types. Fourteen research projects occurred over time, engaging 17 organizations. Eighteen organizations from across the state have also collaborated on preservation networks.

3.1.6 Rhode Island

Collaborators hailed from the American Mathematical Society, the Rhode Island State Archives, or Brown University, or Brown University's Library. Of the group, Brown University was the most engaged,

²³ Census bureau regional description titles are used to group each of the states.

participating in five of the seven activities. Rhode Island also is notable in rating as one of the top four states with the most activities occurring with the fewest number of collaborators.

3.1.7 Connecticut

The 16 organizations collaborating on 33 activities are clustered in the middle of the state, with a concentration around the Hartford region. Over half of the organizations are connected with either The University of Connecticut (7 of the organizations), or Yale University (4 organizations.) However, Xerox collaborated on the IS&T cross sector digital preservation for A/V conference from 2005-2010. Finally, it is worth noting that Connecticut also has the most activities occurring with the fewest number of collaborators among the states²⁴.

3.2 NORTHEAST - Mid-Atlantic (Census Region1-Division2)

3.2.1 Regional Summary

The Mid-Atlantic region is home to 179 organizations collaborating on 115 digital preservation activities.

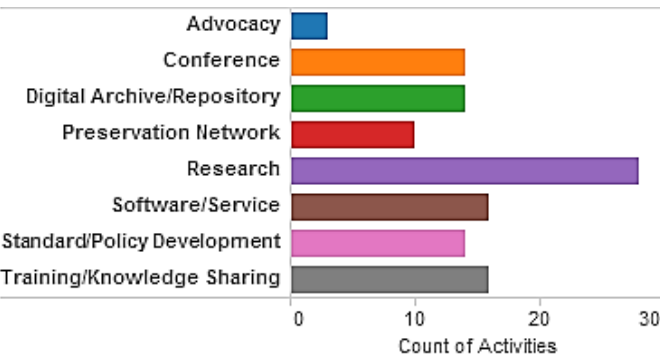
3.2.2 New York

Within our dataset, New York State has one of the highest activity levels (99), second only to California. The 116 collaborating organizations hail from across the state. Organizations collaborating on multiple activities include the Brooklyn Historical Society, the AudioVisual Preservation Solutions - AVPS, which is also a founding institution behind the AVCC - *Audio Visual Community Cataloging* project, and the Council of State Archivists. The rich variety of organizations includes multiple professional associations such as the Association for Computing Machinery and the Association of Jewish Libraries, broadcast media firms (ABC News, HBO), and technical media firms (IBM, Eastman Kodak.)

3.2.3 Pennsylvania

With 26 activities represented, involving 47 organizations, Pennsylvania’s collaborators are located across the state. Pennsylvania is notable for having the second fewest relative number of activities engaging the highest number of collaborators among the states. Twenty organizations are involved in collaborative digital repositories or archives, while 14 are involved in preservation networks. The American Society for Media Photographers has participated in three activities, including playing a leading role in the *Metadata Standards and Tools for Digital Photography* effort. Penn State University and its campus-based entities are singularly engaged in the broadest set of activities compared among the Pennsylvanian organizations, with the University of Pennsylvania and the University of Pittsburgh also well represented.

Figure 14: Spread of 115 Activities in Mid-Atlantic



²⁴ Calculated as the ratio of number of unique activities to number of organizational units participating within a state.

3.2.4 New Jersey

Sixteen organizations engaged in 20 activities, primarily across the mid-section of the state. Princeton, Rutgers and their subsidiaries are the only entities involved in multiple activities. Four distinct research activities touched the state, with JVC America continuing to be involved in CineGrid and the New Jersey State Archives continuing their involvement in SERI.

3.3 MIDWEST - East North Central (Census Region2-Division3)

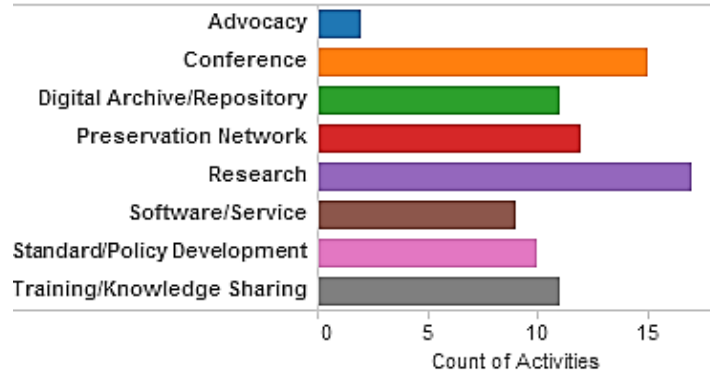
3.3.1 Regional Summary

The East North Central Midwest region accounts for 156 collaborating organizations that have participated in 87 activities.

3.3.2 Wisconsin

Fourteen organizations participated in a span of 20 activities. CLOCKSS, GeoMAPP, and PDB preservation networks had organizational involvement. The Wisconsin Historical Society was engaged across knowledge sharing efforts like the NDSA, the DPOE training, and the state-archive oriented STEP training. While the University of Wisconsin - Madison and its affiliated units represent nearly half of Wisconsin's activity in our dataset, the Wisconsin Historical Society represents nearly a quarter of the activity, participating across seven activities over time.

Figure 15: Midwest - East North Central Spread of 87 Activities



3.3.3 Michigan

The majority of the activity was based around Ann Arbor and the academic hub of the University of Michigan (67%), with Michigan State (11%) and Eastern Michigan (3%) universities notably less engaged.²⁵ Outside of academia, organizational involvement included state agencies like the state archives, state library and the Department of History, Art and Archives. In total, Michigan had 41 activities involving 28 collaborating organizations. The diversity of collaborative engagement is reflected in organizational involvement in efforts ranging from the EVIA effort to preserve ethnographic field videos and the Open Language Archives Community aiming to develop best practices and a repository network for languages, to the USDocs and Cinegrid efforts.

3.3.4 Illinois

Collaborative efforts span the state, with a concentration around the Chicago area. A total of 46 activities were captured in our dataset, engaging 61 organizations. Training or knowledge sharing, and digital archive activities engaged 25 and 15 organizations respectively. The Center for Research Libraries is playing a role in multiple standards development efforts, while the Illinois State Archives and Illinois State Library have played diminishing roles over time, now engaged in SERI and DPOE respectively. Northwestern University and its affiliated units are actively participating in nine different activities

²⁵ Inclusive of their affiliated units or programs

ranging from the Legal Information Preservation Alliance, to the Variation on Video project and Preserving Digital Still Images project. A number of professional associations based in the state are also involved in collaborative efforts, including the Accreditation Council for Graduate Medical Education, the American Medical Association and the American Academy of Pediatrics, and the American Association of Law Libraries.

3.3.5 Indiana

Across the state, 31 activities engaged primarily academic organizations. Indiana University and its subunits account for 44% of the activity in our IN dataset, followed by Purdue (15%) and Notre Dame (15%). The State Archives, the Commission on Public Records, and the not-for-profit Legal Information Preservation Alliance also participated in multiple activities, inclusive of the NDSA, SERI, and the Program for Electronic Records Training, Tools and Standards, and the Legal Information Archive.

3.3.6 Ohio

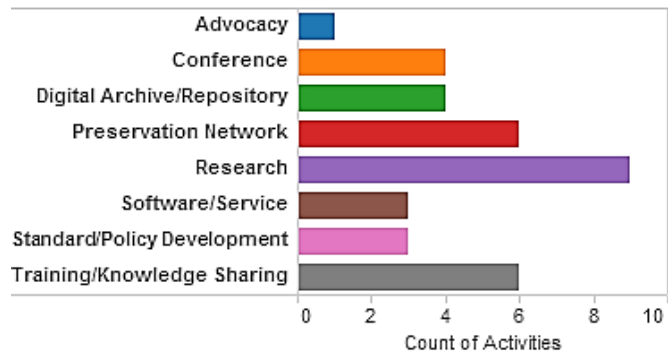
While 32 activities have engaged organizations from the Southwest to the Northeast of the state, the majority of organizations collaborated on only one or two activities. OCLC is a notable exception, as it has participated in 14 activities, seven of which are ongoing. Also in central Ohio, the not-for-profit Ohio Historical Society has participated in six activities ranging from the ongoing SERI and DPOE efforts to concluded efforts including the Persistent Archive Testbed and Preservation Training.

3.4 MIDWEST - West North Central (Census Region2-Division4)

3.4.1 Regional Summary

The West North Central Midwest region accounts for 72 collaborating organizations having participated in 36 activities.

Figure 16: West North Central Spread of 36 Activities



3.4.2 Missouri

Across the midsection of the state, 15 organizations have collaborated on 13 activities, all but three of which are ongoing. With the exception of the University of Missouri and the Washington University in St. Louis, most organizations were engaged in only one or two activities.

3.4.3 North Dakota

Seven organizations, with all but one from the state government sector, were engaged in six activities. Organizations were involved in a single activity, with the exception of the State Historical Society of North Dakota, which is continuing engagement in multiple activities.

3.4.4 South Dakota

Two collaborative organizations represent the state, the South Dakota State Archives and the South Dakota Bureau of Information and Telecommunications. Both participate in the SERI initiative, while the State Archives also participates in DPOE.

3.4.5 Nebraska

Concentrated in the Southeast quadrant of the state, 11 organizations have collaborated on 13 activities. The University of Nebraska-Lincoln and Creighton University (and their affiliated programs), complement the range of state government involvement. Multiple organizations are active in the SERI initiative, the Legal Information Preservation Alliance, and the State Government Durable Records Working Group.

3.4.6 Kansas

Within the Northeast quadrant of the state, 10 organizations collaborate in 13 activities. The Kansas Historical Society has been involved in seven activities, such as SERI, DPOE, and GeoMAPP, while both the University of Kansas and Kansas State University have each been involved in four, such as the WEST and DPN networks.

3.4.7 Minnesota

Clustered around Minneapolis, 17 organizations have collaborated on 23 activities. The Minnesota Historical Society has been incredibly active and diverse in its activities, participating in 14 activities, spanning from the NDSA, GeoMapp, and SERI to the Program for Electronic Records, Training, Tools and Standards project and the Chronicles of Preservation: Distributed Preservation of Born-Digital and Digitized Newspaper Collections project. The University of Minnesota and its affiliates are also well represented in the dataset, representing 29% of the activity in the state.

3.4.8 Iowa

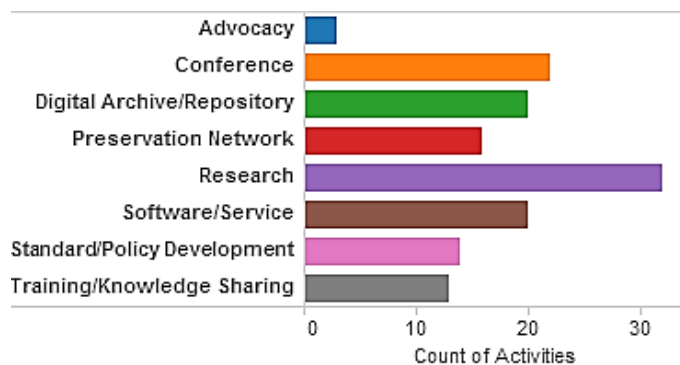
Spread across the state, 10 organizational units have participated in 10 activities. The University of Iowa or its affiliates have participated in a wide variety of activities, ranging from networks like CLOCKSS, DPN, and HathiTrust, to the Crisis, Tragedy and Recovery Network (CTRNet). Iowa State University is also engaged in the DPN and HathiTrust networks, in addition to the Western Regional Storage Trust and LOCKSS Alliance. The MechDyne Corporation is engaged in research through the CineGrid initiative, while the State Historical Society is involved in SERI.

3.5 SOUTH - South Atlantic (Census Region3-Division5)

3.5.1 Regional Summary

The South Atlantic region accounts for 140 activities engaging 299 organizations.

Figure 17: Spread of 140 Activities in the South Atlantic Region



3.5.2 Delaware

Situated at both the North and South ends of the state, four organizations have collaborated on six activities. The Delaware Public Archives has played a leading role in the Delaware State Government e-Records Strategy effort, while also participating in SERI and another initiative. The University of Delaware's collaborations are reflected in both the Protein DataBank initiative and HathiTrust.

3.5.3 Maryland

Primarily clustered mid-state, 42 organizations have collaborated on 52 activities. Some private entities like the Center for Home Movies and the Highlands Group are collaborating on the Personal Digital Archiving initiatives, while a number of professional associations are engaged with the CLOCKSS activities. Multiple federal agencies are represented here, such as the Department of Agriculture, NIST and a few of the Institutes of Health. The University of Maryland, the Johns Hopkins University and their affiliated programs have also engaged in a variety of activities.

3.5.4 District of Columbia

The District of Columbia surfaced as regional location undertaking the most activities (93) with the fewest number of engaged organizations (73). Individual organizations are engaged in more activities compared to organizations in other states. This is understandable given the role of federal agencies in this region, with entities like the Library of Congress, Government Printing Office, National Archives and Records Administration and Smithsonian Institute each collaborating on numerous efforts.

3.5.5 Virginia

Spread across the state, 57 organizations have collaborated on 56 activities. Within proximity of the DC region, numerous professional associations ranging from the Aerospace Medical Association, to the American Diabetes Association and American Society of Civil Engineers contribute to the CLOCKSS effort. The Consultative Committee on Space Data Systems plays a leading role in multiple standards development initiatives.

3.5.6 West Virginia

The West Virginia State Archives representatives have collaborated on the SERI initiative, while the West Virginia University was involved in the former DCAPE - Distributed Archival Custodial Preservations Environments project.

3.5.7 North Carolina

Concentrated within central North Carolina, 46 organizations collaborated on 51 activities. The academic hubs of UNC-Chapel Hill, Duke, and North Carolina State University represent much of this activity. Multiple geospatial initiatives are active in the state, including GeoMAPP and the North Carolina GeoSpatial Data Archiving Project. Multiple organizations from NC are also involved in the DataNet Federation Consortium and the NC Docks effort.

3.5.8 South Carolina

Spread from the Northwest to the Southeast of the state, 9 organizations have collaborated on 12 activities. The South Carolina Department of Archives and History has contributed to five activities, including PeDALS, the NDSA, LOCKSS Alliance and SERI. The University of South Carolina, Clemson University, and their affiliates are also involved in multiple activities.

3.5.9 Georgia

Across the state, 34 organizational units have collaborated on 29 activities. Very active organizations in the state, inclusive of their affiliates, include the Georgia Institute of Technology, University of Georgia, Emory University, and the not-for-profit Educopia Institute. The Georgia Archives and the Georgia

Geographic Information Systems Coordination Committee have collaborated on the TIGER/Line File Preservation effort.

3.5.10 Florida

Spread out across the state, 32 organizations have collaborated on 26 activities. The Florida Virtual Campus at the Florida Center for Library Automation has been involved in a span of 10 activities, with some ongoing like the NDSA, Bag-IT, the PREMIS project and the Florida Digital Archive. Academic institutions and their affiliated programs, including the University of Florida and University of Miami, have also collaborated on multiple activities across their institutions.

3.6 SOUTH - East South Central (Census Region3-Division6)

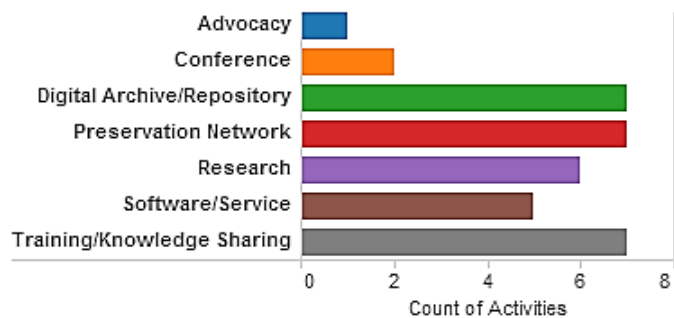
3.6.1 Regional Summary

The East South Central region represents 67 organizations collaborating on 35 activities.

3.6.2 Kentucky

Across most of the state, 28 organizations have collaborated on 21 activities. The Kentuckiana Digital Library collaboratively engaged 21 organizations of varying forms, inclusive of public libraries, two and four year colleges, and historical societies. The Kentucky Department for Libraries and Archives is quite active, having participated in 10 activities over time, six of which are ongoing efforts such as the NDSA, SERI, and GeoMAPP. The University of Kentucky and its affiliates participated in eight activities, ranging from USDocs and the Kentuckiana Digital Library to the *Strategies of Data Archiving for Cultural Anthropology, Using Gwembe Tonga Research Project Data* effort.

Figure 18: Spread of 35 activities in East South Central



3.6.3 Tennessee

Within Central and Eastern Tennessee, 16 organizations have collaborated on 17 activities. Both the University of Tennessee and Vanderbilt University, and their affiliated programs, have engaged in 11 and six activities respectively. Both institutions are contributing to the ongoing Federated Archive Cyberinfrastructure Testbed project. Three state agencies collaborated with others on *A Model Technological and Social Architecture for the Preservation of State Government Digital Information*. It is also worth noting that Oak Ridge National Lab and UT-Knoxville are also collaborating with others on the *DataONE - Data Observation Network on Earth* project.

3.6.4 Mississippi

Across the state, nine organizations have collaborated on 12 activities. The Mississippi Department of Archives and History has been active in seven efforts, all but one of which are ongoing. In 2012, the library at the University of Southern Mississippi led a team of collaborators at the Mississippi State University and University of Mississippi on a *Mississippi Digital Library Preservation Assessment* project.

3.6.5 Alabama

Spread across the state, 14 organizations have collaborated in 15 activities. Eleven of the organizations spanning academia, state agencies, and public libraries have collaborated on the Alabama Digital Preservation Network. The University of Alabama and Auburn University also have engaged in multiple initiatives.

3.7 SOUTH - West South Central (Census Region3-Division7)

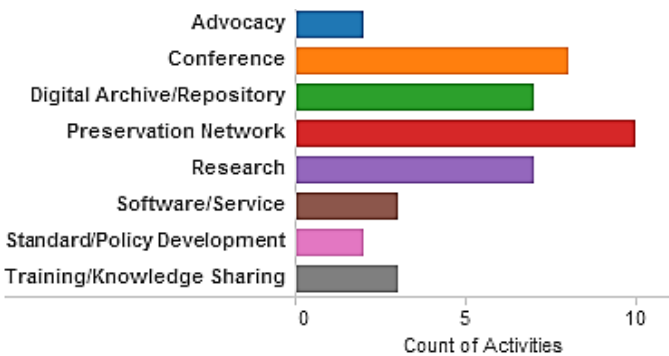
3.7.1 Regional Summary

The four states in the West South Central account for 71 organizations engaged in 42 activities, with Texas playing a dominant role.

Figure19: Spread of 42 activities in West South Central

3.7.2 Oklahoma

Concentrated around three locations, seven organizational units collaborated on eight different activities in the state. The University of Oklahoma and its affiliated units collaborated on five activities, ranging from the Western Regional Storage Trust to the DPOE and Legal Information Preservation Alliance initiatives.



3.7.3 Texas

Spread across Central and Eastern Texas, 46 organizational units participated in 37 activities. Twenty-seven organizations were engaged in one of many digital preservation networks, such as the Texas Digital Library, DPN, LOCKSS Alliance, or the Western Regional Storage Trust. The University of North Texas is the most engaged, involved in 15 activities (leading the *CyberCemetery* effort and an ETD Lifecycle Management project.) Similarly, UT-Austin and its affiliated units were engaged in 16 varied activities, ranging from acting as a founding institution behind the aforementioned Texas Digital Library to collaborating on ongoing efforts like PermaCC and the UniProt Consortium.

3.7.4 Arkansas

Concentrated in three locations, eight organizations collaborated on seven unique activities. Three state agencies, the Arkansas Bureau of Legislative Research, the Arkansas History Commission, and the Arkansas State Library, collaborated with others on *A Model Technological and Social Architecture for the Preservation of State Government Digital Information* effort that concluded in 2012.

3.7.5 Louisiana

Located in the Southern half of the state, 10 organizational units have collaborated on 8 activities. Only Louisiana State University and its affiliates (5 activities) and Tulane University (4 activities) have collaborated on multiple initiatives. Two academic and one state body are collaborating in the Legal Information Preservation Alliance, while two universities are collaborating on the USDocs initiative.

3.8 WEST - Mountain (Census Region4-Division8)

3.8.1 Regional Summary

Together, the eight Mountain states represent 102 organizations engaged in 60 activities.

3.8.2 Idaho

On the Western end of the state, six organizations have collaborated on eight activities. The Idaho Historical Society is the only organization collaborating on multiple ongoing initiatives, namely the DPOE and SERI efforts.

3.8.3 Montana

Six organizations concentrated in two locations, have participated in 10 activities.

The Montana State Library has collaborated on three efforts, two of which are ongoing including GeoMAPP - in which it plays a leading role. Similarly, the Montana Historical Society has been involved in three initiatives, with both SERI and the PERTTS program ongoing.

3.8.4 Wyoming

In the Southeast corner of the state, six organizational units have collaborated on six unique activities. The University of Wyoming and three of its affiliated programs represent the bulk of this activity, including digital archive or repository efforts, such as the Alliance Digital Repository, GeoMAPP, and CLOCKSS. The State Archives are collaborating on SERI while the state's Department of Enterprise Technology is also engaged in a state-wide effort around government records.

3.8.5 Nevada

Within the bottom half of the state, five organizational units collaborated on eight activities. The University of Nevada Las Vegas and its Law Library represent half of this activity, being engaged in the Western Regional Storage Trust, the Legal Information Preservation Alliance, CLOCKSS, and PermaCC. The Nevada State Library and Archives is also involved in multiple initiatives, including the State-based Mid-term Archiving Solution effort it is collaborating on with the Nevada Office of the CIO.

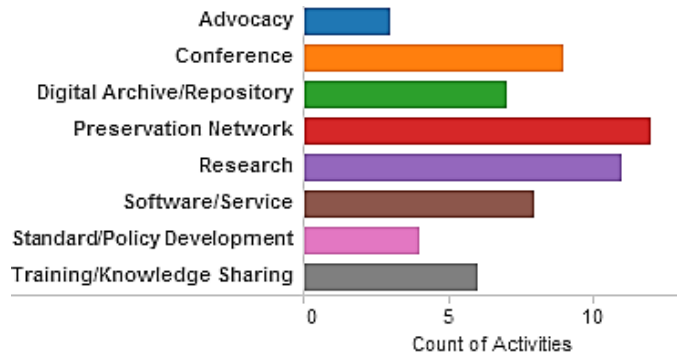
3.8.6 Utah

Within the Northern half of the state, 14 organizational units have collaborated on 21 activities. The Genealogical Society of Utah - FamilySearch has consistently collaborated on the IS&T efforts since 2009. Brigham Young University and its affiliated units have collaborated on six efforts, including the ongoing Western Regional Storage Trust, Legal Information Preservation Alliance, LOCKSS Alliance, and DPOE effort.

3.8.7 Colorado

Within the mid-section of the state, 33 organizational units have collaborated on 26 activities. The University of Colorado at Boulder and its affiliates collaborated on seven activities, spanning domains from the PermaCC effort to the Research Data Alliance's Preservation e-Infrastructure Interest Group

Figure 20: Spread of 60 activities in West Mountain Region



and the Western Regional Storage Trust. The University Corporation for Atmospheric Research has engaged in five efforts over time, three of which are ongoing (NDSA, Data Conservancy, and Chronopolis.) Seven Colorado based institutions have collaborated through the Alliance Digital Repository consortium.

3.8.8 Arizona

Within the mid-section of the state, 20 organizations have collaborated on 24 activities. The Arizona State Library, Archives and Public Records has been the most active, participating in at least 10 distinct activities including the ongoing USDocs, CLIR Fellows, SERI, GeoMAPP, and PERTTS efforts. It also acts as a lead organization to the State-based Trusted Electronic Records Repository effort. Within academia, both Arizona State University and the University of Arizona (and their affiliates) represent a large variety of activity engagement, ranging from the Open Language Archives Community to the *tDAR - The Digital Archeology Record* project.

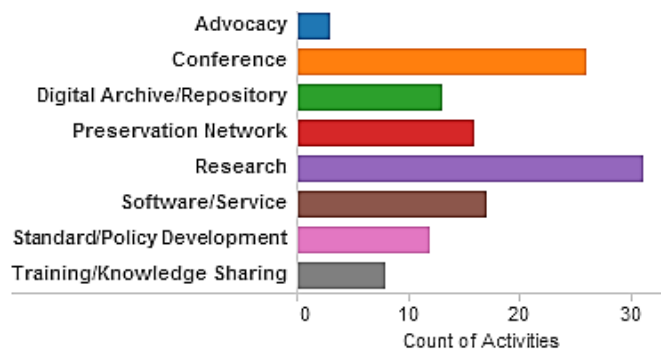
3.8.9 New Mexico

In upper Central New Mexico, under a dozen organizational units have collaborated on 15 activities. The Los Alamos National Laboratory Research Library has participated in six activities, acting as a lead institution behind the ongoing *Electronic Journal Metadata & Transfer Project* and *Memento* - a web archiving incentive program. The New Mexico Commission on Public Records has also participated in multiple efforts, three of which are ongoing (SERI, the Centralized Electronic Records Repository, and PERTTS). The Centralized Electronic Records Repository acts as a centralized electronic repository for the state agencies of New Mexico, engaging both the State Records Center and Archives and the state's Department of Information Technology.

3.9 WEST - Pacific (Census Region4-Division9)

3.9.1 Regional Summary

The three mainland states, plus Alaska and Hawaii, account for 225 organizations engaged in 126 activities. California dominates this activity, accounting for 71% of the organizational involvement, although each state has collaborating organizations.



3.9.2 Alaska

Alaska is remarkable not only for having the fewest activities engaging the highest number of collaborators among the states, but for the wide geographical representation by organizations. Spread across the state's vast geography, 18 organizational units - inclusive of many nonprofits and local libraries, are engaged in 8 activities. Fifteen of these organizational units collaborated on the DPOE training initiative reaching the far corners of the state. The Alaska State Library is engaged in multiple efforts, including the ongoing LOCKSS Alliance, USDocs effort, and the previously referenced DPOE. The Alaska State Archives has collaborated on two efforts, including the ongoing SERI.

3.9.3 Washington

In seven hubs across the West and East edges of the state, 28 organizational units have collaborated on 33 activities. Microsoft Research has participated in a wide variety of activities, including the ongoing Data Conservancy, Open Planets Foundation, PDF/A and PhotoMetadata.org efforts. Also in the Seattle region, the University of Washington and its affiliated units have participated in 11 efforts ranging from InterPARES to CLOCKSS and Terra Populus. Five organizations inclusive of network interconnection exchanges, commercial research labs and academia are collaborators on the CineGrid collaborative research project focused on digital media on photonic networks. The Washington State Archives is also notably leading a new effort around Cloud-based Records Management for Washington State.

3.9.4 Oregon

Primarily concentrated in Northwest Oregon, 17 organizational units have collaborated on 14 activities. Three universities and the not-for-profit Orbis Cascade Alliance collaborate on the Western Regional Storage Trust, while the Oregon State Archives and the Oregon State Library had previously collaborated on the Multi-State Preservation Consortium (which ended in 2011.) The Oregon State Archives are also collaborators on DPOE, SERI, and the aforementioned Cloud-based Records Management for Washington State project.

3.9.5 California

Of all U.S. States, California has the highest level of collaboration, both in terms of the 117 activities represented and its 159 collaborating organizations. A few institutions have historically had remarkably high activity levels, such as Stanford University (and its affiliates), engaged in 47 activities, the California Digital Library, engaged in 16 activities, the San Diego Supercomputing Center, engaged in 16 activities, and the Internet Archive with 12 activities. Across the UC System, activities abound.

Similar to Washington State, research activities are prevalent, with over 60 organizational units engaged in digital preservation research efforts touching upon motion pictures, recorded sound, virtual worlds, anthropology data, and born-digital public records. The WEST and HathiTrust, and DPN preservation networks all have numerous collaborators in the state.

The engagement of diverse organizations in terms of content area and sector is also remarkable. The Academy of Motion Picture Arts and Sciences has engaged in four efforts, three of which are ongoing. The nonprofit Bay Area Video Coalition plays a leading role in two ongoing collaborative efforts, the Quality Control Tools for Video Preservation and the AudioVisual Artifact Atlas. Hewlett-Packard Laboratories have collaborated on five efforts, including the ongoing DSpace and LOCKSS efforts. The Internet Archive has participated in a wide variety of topical areas, with collaborations ranging from the Integrated Digital Event Archive and Library - IDEAL, to Personal Digital Archiving 2014, to the Crisis, Tragedy, and Recovery Network (CTRnet), DDP Frameworks, and the International Internet Preservation Consortium (IIPC).

3.9.6 Hawaii

Three organizations have participated in four activities. The Hawaii State Archives has participated in both the NDSA and the SERI efforts. The University of Hawaii has been a member of the Western Regional Storage Trust, while the Pacific Resources for Education and Learning has participated in DPOE.

4 Future Opportunities

By synthesizing and viewing the universe of collaborative digital preservation information, it is possible to identify opportunities through this dataset. The following initial suggestions to strengthen digital preservation practices and collaboration in the U.S. have resulted from ICONC efforts, but serve as illustrative possibilities of how the data can inform opportunity analysis. The aim is to provide direct access to interactive data analysis tools that can support planning for increased cross-sector work (including around defining terminology, sharing workflows and “good practice” documentation, and sharing tools) and increased attention to states as is appropriate given the realities of that state environment.^{26,27}

4.1 Engage and Build Upon Priority Activity Footprints

The occurrence of a collaborative activity leaves a footprint of established professional connections, working relationships, and communications pathways that result from a history of working together. When such a history of collaboration is present, it can inform and facilitate both the development of new activities within the collaborating group, and the process of introducing new members into the group. Collaborators behind activities also can act as strategic partners and key stakeholders for future endeavors, bringing their experience, materials, and connections to the effort.

Maps of existing collaboration networks can inform outreach strategies and alliance building activities. The NDIIPP could use the NDSA’s *2014 National Agenda for Digital Stewardship* (National Agenda) to identify priority needs within the field of digital preservation, and from these needs, identify desired activity characteristics or specific communities of interest. Once such are defined, NDIIPP can use the dashboards to hone in on what has and is occurring within such priority collaboration communities. For illustrative purposes, we highlight below a few examples of how this pilot ICONC dataset and the related pilot ICONC interactive dashboards could be used to inform strategic planning building from the National Agenda.

²⁶ It is important to note that the low activity count or a low organizational collaborator count within a state does not conclusively indicate a lack of digital preservation activity. Since this dataset is focused on collaborators behind the production of effort, users and attendees are not reflected. Organizations or their staff may use established products, participate in activities, or participate in trainings or conferences. They just may not be collaborating with others in doing so.

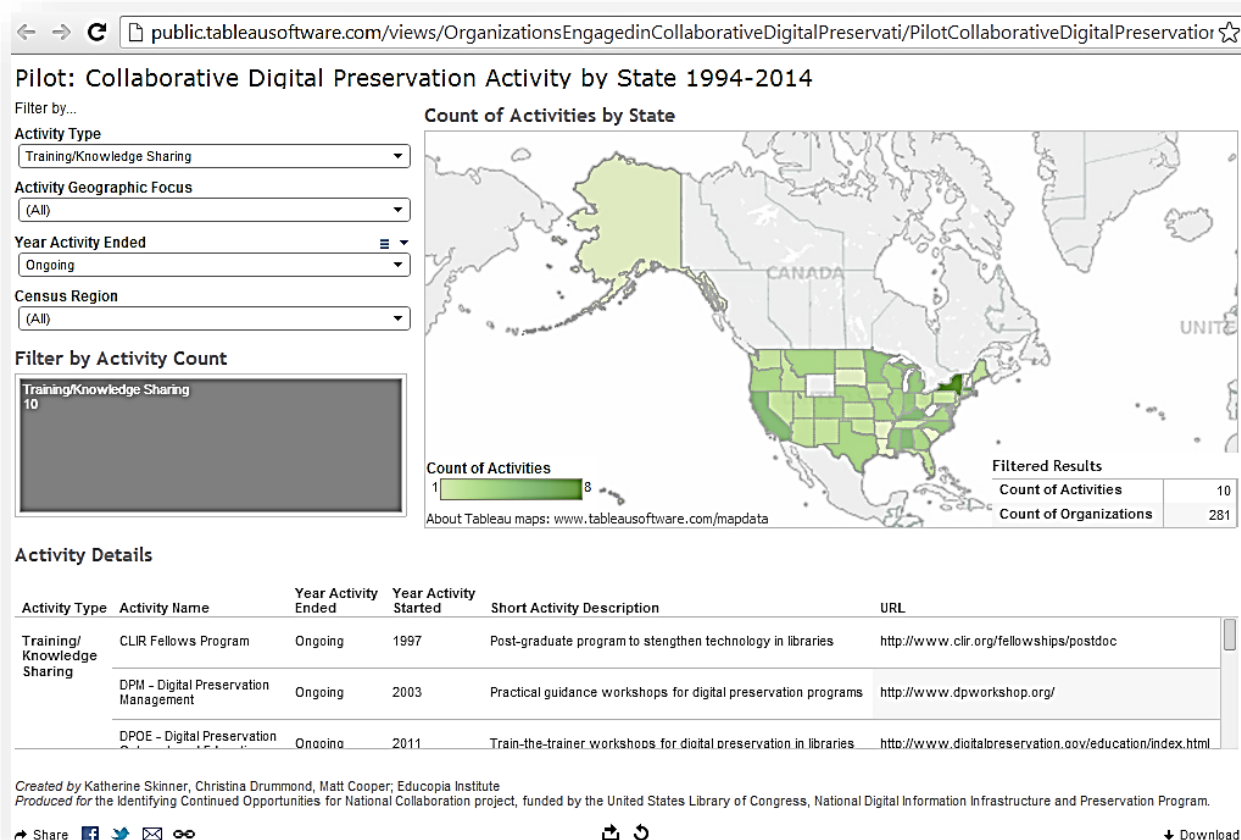
²⁷ It is no surprise that geographic hotbeds for technology industry and technology R&D are also well represented within this dataset. The most useful metric developed within this project for comparing states in terms of collaborative activity was a ratio of the number of activities to the number of collaborators in a given state. This rough measure allows you to quickly identify outlier states that either have the most activities going on with the fewest collaborators (Connecticut, Rhode Island, Montana, Nevada), or the fewest activities with the most collaborators (Vermont, Pennsylvania, Alaska). However, without going into the specifics of a type of activity or genre, this high level information is of limited usefulness

4.1.1 Digital Preservation Training and Knowledge Sharing

One opportunity highlighted within the National Agenda is for the community to develop and share training resources for content curators.²⁸ By using an ICONC interactive dashboard driven by the pilot ICONC dataset, one can view both known training and knowledge sharing activities and the organizations that have collaborated on producing said efforts.

As Figure 22 shows, one can view a density map of organizations that have actively engaged to collaborate and produce digital preservation training or knowledge sharing. Specific details about the 10 ongoing training or knowledge sharing efforts are displayed under this map, with a short description and URL. It is important to remember that this information provides a first glance both at existing initiatives, and the players behind those initiatives - as the organizations represented are the planners behind the efforts - not the participants themselves. Such information can help NDIIPP and the NDSA to strategically plan and align activities, while leveraging existing materials and engaged participants.

Figure 22: Pilot ICONC Dashboard Filtered for Active Digital Preservation Training and Knowledge Sharing Activities
Interactive version online at: <http://educopia.org/deliverables/pilot-iconc-dashboards>



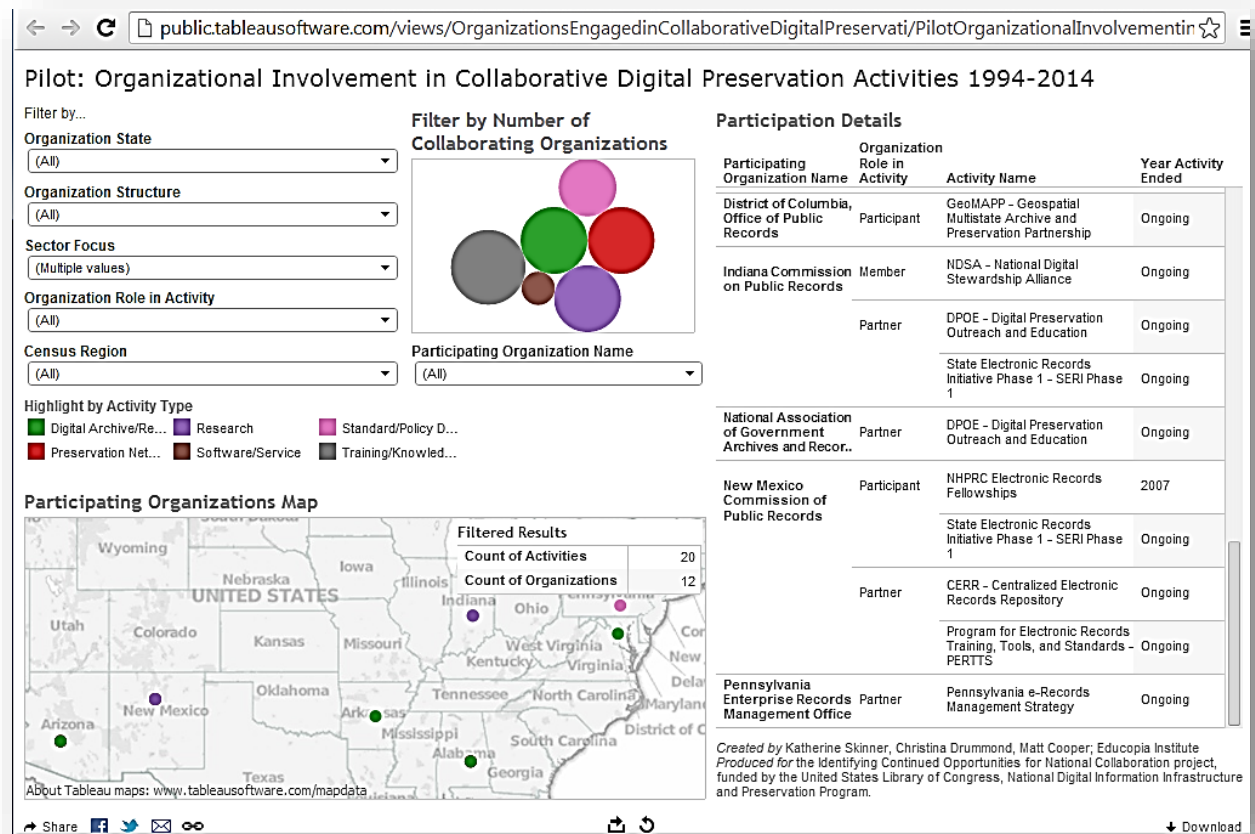
²⁸ 2014 National Agenda for Digital Stewardship. NDSA. Pg. 9.
<http://www.digitalpreservation.gov/nds documents/2014NationalAgenda.pdf>

4.1.1.1 Electronic Records

To identify potential stakeholders and identify regional leaders who can inform conversations around Electronic Records retention, it is possible to use the ICONC dashboards to view the collaborations of the subset of organizations with a government records or state records mission focus. While the sector focus taxonomy used within this pilot dataset is still under development, preliminary coding allows one to select organizations with a particular focus to be viewed. As a result, information is presented for the 12 organizational collaborators within the dataset coded as having a government or state records focus, noting specifics about their activity engagement.

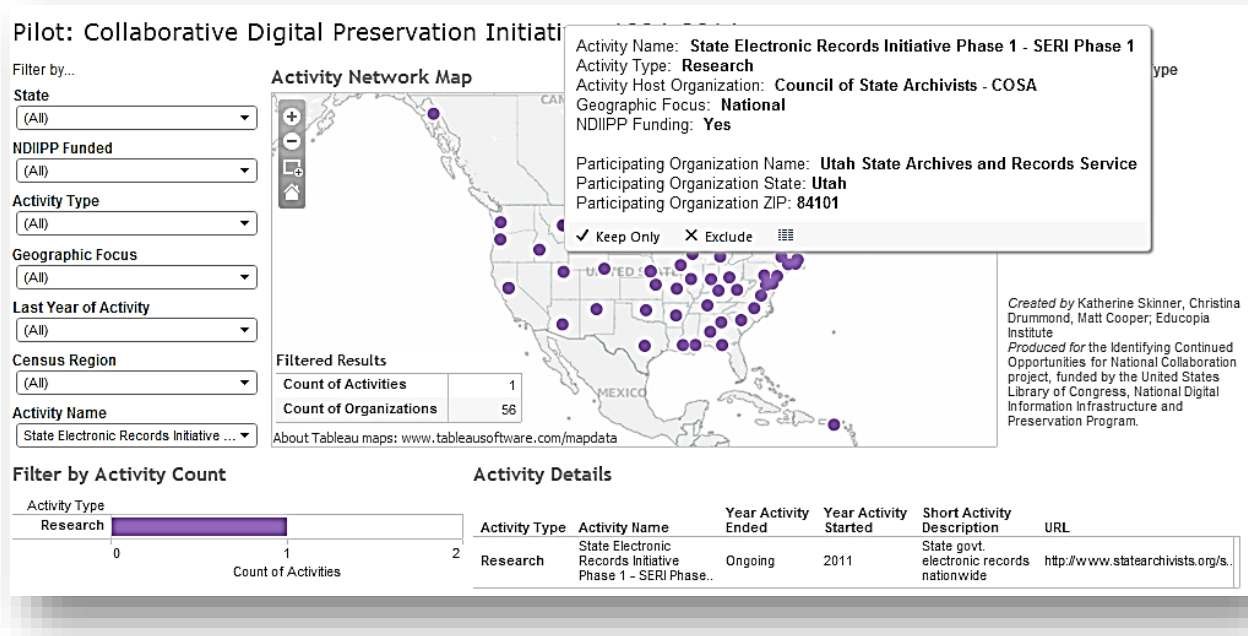
Figure 23: Pilot ICONC Dashboard Filtered for Government and State Records Oriented Collaborators

Interactive version online at: <http://educopia.org/deliverables/pilot-iconc-dashboards>



Another way to view this data would be to look at all of the collaborators engaged within each activity tied to the topic of Electronic Records. Yet another dashboard allows us to view all of the collaborating organizational nodes engaged in The State Electronic Records Initiative (SERI). (Figure 24) Clicking one of the organizational points on the map brings up details about that organization.

Figure 24: Pilot ICONC Dashboard Filtered for Organizations Collaborating on SERI

Interactive version online at: <http://educopia.org/deliverables/pilot-iconc-dashboards>

A future area we hope to explore to extend the ICONC dataset is to develop a classification taxonomy for the activities that identifies either content or thematic focus. While this can currently be (roughly) done for organizations, it would be useful to view together the network maps for all activities within the dataset that are related to research data or records retention or images (for example). This has been identified as a future area of work for ICONC.

Informatics dashboards such as the examples above allow leaders to ask questions dynamically of the data as they develop their strategic plans. While it is hard to extract organizational priorities and local realities from the dataset, this toolset can enhance such knowledge that program staff and community leaders can bring to the table.

4.2 Connect, Strengthen, and Build Upon Existing Sector Associations

In order to leverage expertise that has developed in parallel in siloed sectors, there are numerous opportunities for NDIIPP to develop strategic relationships with targeted sector- or genre-focused communities. The ICONC dataset can help leaders to find the “needle in the haystack” when trying to identify the organizations and individual contacts who can act as connectors between sectors and communities on specific projects. The networks of collaborating organizations for given activities can be compared against each other, to identify organizations that participate across specified efforts. Similarly, such mapping can provide insights into the geographic location of these organizations, informing decisions on who might be best positioned to host on-the-ground meetings.

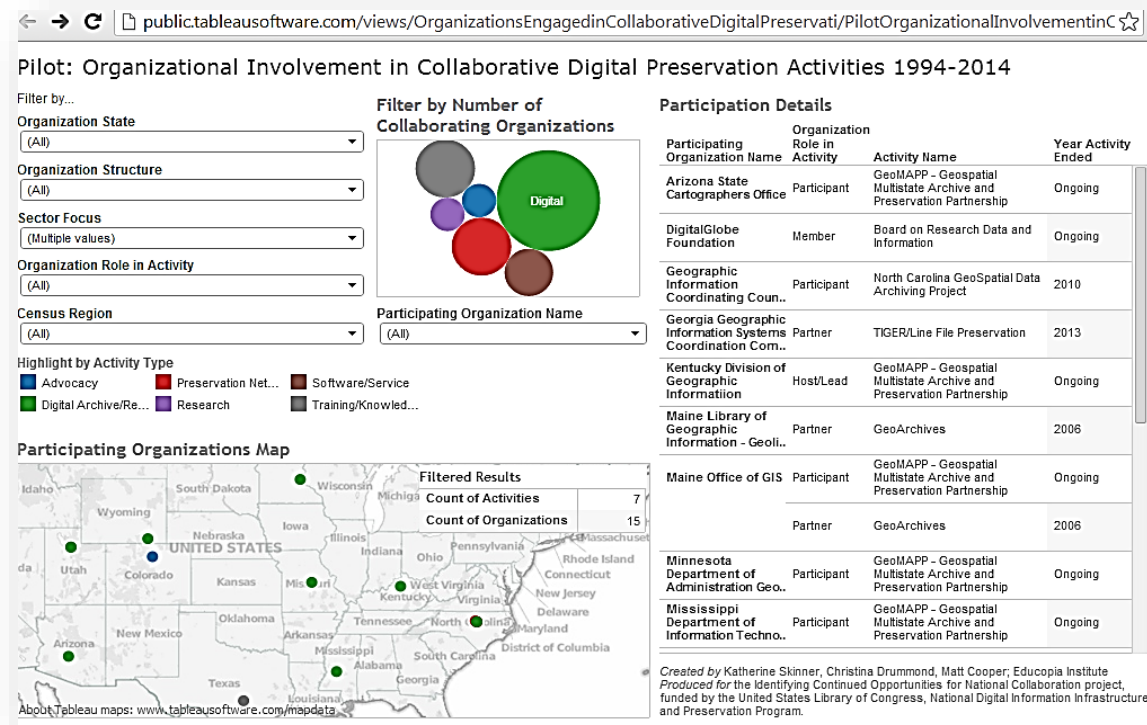
The existing pilot dashboards can show the organizations that engage in a specific activity network. As one example, records retention efforts could bridge and leverage the communities and experience within the Legal Information Preservation Alliance, NAGARA, and NASCIO. With respect to stewarding GIS records, an opportunity exists to engage the state-based Geographic Information Coordinating

Councils, even if such organizations are known by another name. (Figure 25) For a final example, potential partners in the area of audio and visual content when looking at the collaborator networks involved in efforts such as the CineGrid project or the Metadata Schema Development for Recorded Sound project.

While the dashboards do not yet support the functionality to look at multiple layers of network nodes nested on top of one another, this functionality is possible. In the meantime, the data for such analysis resides within the full raw dataset.

Figure 25: ICONC Pilot Dashboard Filtered for GIS Related Organizations

Interactive version online at: <http://educopia.org/deliverables/pilot-iconc-dashboards>



4.3 Develop Strategic Informatics Tools

As is case when building software, piloting a beta version helps users fully understand what would be most beneficial to have in a final product. ICONC's pilot efforts to date, both around collating a dataset and developing beta dashboards, provide the digital preservation community with initial tools to support strategic planning and partner identification. However, like any development project, ICONC's pilot implementation of informatics around collaborative digital preservation has demonstrated to the research team where further refinements could be made.

Building upon this pilot, the research team hopes to further develop informatics tools for the digital preservation community to help partners to find each other and to allow planners to track progress and inform new projects. In a commercial setting, such 'business intelligence' tools are often guarded to protect an organization's own interests and provide a competitive advantage. However, if the aim shifts

from benefiting a single institution to benefiting a sector writ large, or expanding a practice or developing a field, tools that aggregate and present public information clearly become instruments for the public good. By providing deeper and more holistic information about past and present efforts within a practice or discipline, tools such as those seen here can enable community leaders to incorporate data-driven analysis into their planning.

Funding contingent, the ICONC project is already looking at ways to extend beyond this pilot to capture dynamics such as activity lineage (how an activity impacts or evolves into activities that follow) and individual engagement. It is also looking to refine the existing pilot data and tools for several targeted communities, including funders, administrators, and planners.

The model utilized within this ICONC project surfaces, prioritizes, and publishes the actual data and visualization toolsets alongside the traditional narrative report. Such direct access to usable data may prove more valuable than the report itself due to the ways that the open, data driven toolsets can be repurposed and built upon in the future. A leader in digital preservation or curation could fund the creation of analogous tools for specific communities or data elements to inform decisions on how to move an effort forward over time.

For example, the Organizational Roles, Policies and Practices section of the NDSA's National Agenda points to a number of variables that could be modeled as dashboards. In this case, staffing and infrastructure survey data could be aggregated and mapped to provide a view of where collaborators reside with unique insights into these variables. In fact, prior surveys in the field have captured some of these elements, such as the Association of Research Libraries, Digital Library Federation²⁹, DPOE³⁰, MetaArchive, and the NDSA itself. For another example, a companion dataset of publically available professional individual information such as VIVO could be tied to a dataset like the one presented herein to assist the field in identifying experts, potential collaborators, and change agents, recognizing their approximate locations and past engagement history.³¹

²⁹ See e.g. <http://old.diglib.org/roles/survey5a.htm>

³⁰ See e.g. <http://www.surveymonkey.com/s/D6RF8RZ>

³¹ When encountered in the data collection phase, individual level data helped support the creation of the ICONC pilot dataset, but was purposely abstracted to the organizational level.

5 Appendix: Data Entry Specifications

ICONC Project Source Data Entry Specifications

Document Introduction

Critical Data Entry Specifications:

1. *Each row in the 'Source Data' worksheet should represent a unique activity-organization-pair.*
2. *For each activity-organization pairing, a row will list activity info (identical across all organization pairings for the given activity), and organization activity (unique for each row.) Columns I to R will list information for each organization involved in an activity.*
3. *After completing data entry for a single activity and all of its related organizations, you should have multiple rows with identical activity data posted in columns A to H and S to Y.*
4. *Care must be taken with respect to spelling, spacing, and text formatting. Data is pulled directly from the Source Data spreadsheet into reports.*



Omission of any of the above will result in low-quality data of limited to no use, which would have to be re-entered.

Data Inclusion Specifications:

Activities must meet two criteria to be represented in this data set:

- 1) *Digital preservation/curation/stewardship must be a core component of the activity.*
 - *Often this will be specified on the website for the activity, if not apparent from the name itself.*
- 2) *The activity must be collaborative in nature.*
 - *Multiple organizations (i.e. institutions) must be leading, contributing, building, or participating in the production or provision of the activity.*
 - *If activity is centralized, i.e. one institution controls all services or governance, then it is not collaborative and would be omitted.*



If you are unsure whether an activity should be included, err on the side of inclusion and check in with Christina.

Source Data Worksheet Notes:

- *Yellow column headings pertain to activity data fields.*
- *Green column headings pertain to organization data fields.*
- *Orange column headings pertain to data relevant for activity-organization pairings involving an activity type of 'Digital Preservation Network'*

Activity Related Field Definitions

A) Activity Name

- Spreadsheet location: Column A
- Data definition: the name of the collaborative digital preservation activity (e.g. ANADP Conference or ADPNet)
- *Data formatting notes:*
 - Use title case style and capitalization, as this data will pull directly into reports.
 - If the name is an acronym (e.g. ADPNet), follow the acronym with a hyphen and the full name of the initiative (e.g. ADPNet - Alabama Digital Preservation Network)

B) Activity Status

- Spreadsheet location: Column B
- Data definition: classification of the time-bound status of a collaborative digital preservation/curation/stewardship activity.
- *Restricted data field options: Project, Program, Other*
 - *Project:*
 - *Should be time limited, i.e. have an end date, either in the past or future.*
 - *e.g. projects tied to grant awards*
 - *Program:*
 - *Often a matured project.*
 - *Could have an end date if the program ended.*
 - *If ongoing, it should not have an end date (if so, classify as a project.)*


C) Activity Type

- Spreadsheet location: Column C
- Data definition: classification of the type of collaborative digital preservation/curation/stewardship activity.
- *Restricted data field options: Research, Digital Preservation Network, Open Repository / Digital Library, Closed Repository, Conference, Training, Service/Software, Other*
 - *Research:*
 - *Only include if digital preservation/curation/stewardship oriented.*
 - *Includes grant awards, e.g. the ETD project.*
 - *Digital Preservation Network*
 - *Must have primary purpose of preservation (not public access or something else. If focus is elsewhere, use a different activity type.)*
 - E.g. Portal to Texas History - would not be classified as Digital Preservation Network, as primary focus is access. Would be classified as Digital Library.
 - *If multiple institutions are merely contributing content (as opposed to running the network), then classify as a Digital Library/Repository.*
 - *Nodes hosting content for the network should be found on the network's website. Otherwise, enter N/A*
 - *Multiple institutions acts as nodes, hosting all of the content.*

Activity Related Field Definitions Cont.

- *Open Repository / Digital Library*
 - *Multi-institution repositories should be included if explicit emphasis on preservation/curation.*
 - *If multiple institutions are contributing content, this counts as multi-institution.*
 - *e.g. Portal to Texas History*
 - *Implicit, non-stated preservation activities are not enough to include a digital library or knowledgebase. These entries should all have an explicit focus on long-term management and preservation.*
 - *Single institution repositories should not be entered, as they are not collaborative.*
 - *Public content access (i.e. open access) is often a marker of the repositories.*
- *Closed Repository*
 - *Closed repositories do not have access as a primary goal. Instead, these are often focused on disaster recovery and or data retention.*
 - *Multi-institution repositories should be included if explicit emphasis on preservation/curation.*
 - *If multiple institutions are contributing content, this counts as multi-institution.*
 - *Single institution repositories should not be entered, as they are not collaborative.*
 - *Includes Dark Archives if multi-institutional*
 - *Single institution acts as the node, hosting all of the content.*
- *Conference*
 - *Primary focus of the conference must be digital preservation/curation/stewardship.*
 - *Cannot be included if digital preservation is only a thread, or theme.*
 - *Collaboration may be in the form of putting on the conference (organizing/planning), or be a result (e.g. CurateCamp)*
 - *Don't organizational attendees as organization/activity pairs.*
 - *For each conference, do capture:*
 - *Organizing/Program committee members & institutions:*
 - *Host/lead organization*
 - *Founding institution*
- *Training*
 - *Note: this category includes collaborative programs, initiatives (e.g. DPOE, SAA DAS), and training events (e.g. week-long institutes, pre-conference workshops). As long as they are explicitly preservation focused and collaborative.*
- *Service/Software*
 - *Must be explicitly preservation/curation/retention focused*
 - *Must be collaborative in nature (hosted, run, or produced by multiple institutions.)*

Activity Related Field Definitions Cont.

- *Standard / Policy Development*
 - *Development effort for a common standard (technical, policy, etc.)*
 - *Must be at a co-investigator level or management level to be listed as a partner. Others can be classified as participants.*
- *Data formatting notes:*
 -  *If activity fits the definition of more than one activity type, enter text in the format: 'Multiple: <activity type1>, <activity type2>.'*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.

D) Activity URL

-
- Spreadsheet location: Column D
 - Data definition: url for activity website. If activity has been suspended, a link to website referencing final project reports, description, partners, which maybe provided by LoC if project was NDIIPP supported.
 - *Data formatting notes:*
 - Include _____ in address.
 - If entering multiple urls, separate with a comma and space. Enter multiple urls if a good summary page on the LoC's website, in addition to project page.
 - If none available: leave blank.

E) Activity Short Description

-
- Spreadsheet location: Column E
 - Data definition: general description of activity, of eight words or less. Words should not reference information that is captured elsewhere in the spreadsheet (i.e. NDIIPP funded, geographic focus, current operational status). Consider a functional definition, which may include a focused community reference.
 - *Data formatting notes:*
 - Use phrases in sentence case capitalization.
 - Omit periods.

F) Was Activity NDIIPP Funded

-
- Spreadsheet location: Column F
 - Data definition: answer of whether NDIIPP provided direct funding to activity in grant form.
 - Restricted data field options: Yes, No
 - *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

Activity Related Field Definitions Cont.**G) Year Activity Started**

- Spreadsheet location: Column G
- Data definition: enter first year of activity operations.
- *Data formatting notes:*
 - Enter year in four digits, *e.g.* 2013
 - If unknown, leave blank.

H) Year Activity Ended H

- Spreadsheet location: Column F
- Data definition: enter last year of activity operations.
- *Data formatting notes:*
 - Enter year in four digits, *e.g.* 2013.
 - If activity has not ended, enter 'Ongoing'.

I) Activity Geographic Focus

- Spreadsheet location: Column I
- Data definition: classification of the intended regional scope of the collaborative activity.
- Restricted data field options: International, National, Regional, State, Local, None Specified.
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

J) Current Host/Lead Organization for Activity

- Spreadsheet location: Column I
- Data definition: Name of the organization that is currently hosting and/or leading the activity. This should be explicitly listed on the website.
- *Data formatting notes:*
 - Capitalize as a proper noun.
 - Multiple organizations could have this designation, but it has to be documented online.
 - If unknown, leave blank.

K) Founding Institution of Activity

- Spreadsheet location: Column K
- Data definition: Name of organization that founded the activity in the first year. This needs to be explicitly stated.
- *Data formatting notes:*
 - Capitalize as a proper noun.
 - If unknown, leave blank.

Organization Related Field Definitions

L) U.S. State/Territory of Participating Organizations

- Spreadsheet location: Column L
- Data definition: name of U.S. State or Territory where participating organization is located.
- Restricted data field options: List of U.S. States and Territories
- *Data formatting notes:*
 - If participating organization is based outside of the U.S. states and territories, enter 'International'.
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

M) Participating Organization Name

- Spreadsheet location: Column M
- Data definition: full name of participating organization. If name is an acronym, begin the entry with the acronym followed by a hyphen and then the full name.
 - For academic institutions: individual schools, departments, centers, and institutes will be collapsed and represented solely as the university.
 - For state governments: state agencies will be represented individually. E.g. Secretary of State and State Archivist would both be included as organizations if they were involved, even if the Archivist office was nested within the Secretary of State's office.
 - For international partners: only include an international organization if they are a host or founder of the activity. International organizations acting in the other roles will not be represented in this version of the data set, given time limitations and the domestic focus of the project.
- *Data formatting notes:*
 - Capitalize as proper nouns.

N) Organization Role in Activity

- Spreadsheet location: Column N
- Data definition: classification of the organization's role(s) in the collaborative activity. If multiple, state, 'Multiple: <role 1>, <role 2>
- Restricted data field options: Founding Institution, Host/Lead, Partner, Sponsor, Participant, Developer, Member Organization
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.

O) Organization Address Line 1

- Spreadsheet location: Column O
- Data definition: First line of listed address of organization, inclusive of street number and street.
- *Data formatting notes:*
 - *Enter St., Rd. and other common street abbreviations in abbreviated form.*

Organization Related Field Definitions Cont.**P) Org Address Line 2**

- Spreadsheet location: Column P
- Data definition: Second line of listed address of organization, inclusive of street number and street. Only use if needed.
- *Data formatting notes:*
 - *Spell out Suite (instead of Ste.)*

Q) Org City

- Spreadsheet location: Column Q
- Data definition: City of listed address of organization.

R) Org State/Province (incl. country if international)

- Spreadsheet location: Column R
- Data definition: State or territory for listed address of organization. If organization is based outside of the U.S., enter the nation instead (e.g. Spain).
 - *Note, while this information mirrors the data in column I, it must be formatted differently to enable address Geolocation.*

S) Org Zip

- Spreadsheet location: Column S
- Data definition: Postal code for listed address of organization.

T) Organizational Contact Name

- Spreadsheet location: Column T
- Data definition: Primary contact at this organization for the given activity, if known. For activities that have ceased, enter the person of record at the given organization. If you cannot find this information, enter, "None specified".
- *Data formatting notes:*
 - *Include both first and last names.*
 - *Do not include titles.*

U) Organizational Contact Email

- Spreadsheet location: Column U
- Data definition: email address for primary contact listed in column Q, if known. If this information is not readily available from activity website, leave data blank.
- *Data formatting notes:*
 - *Use @ within the address, (e.g.)*

Digital Preservation Network Related Field Definitions**V) If Activity is Digital Preservation Network: List Member/client/participant organizations**

- Spreadsheet location: Column V
- Data definition: a listing of all participating organizations in the network. Note: all organizations within this list should also have their own rows tied to the activity.
- *Data formatting notes:*
 - *Enter each organization as a Proper Noun*
 - *Enter each new organization on a new line within the cell. Press 'Alt' and 'Enter' simultaneously to add line break within the cell. Alternatively, use notepad to create or clean the list from the web and then paste the clean list into the cell.*

W) If Activity is Digital Preservation Network: Is the network a membership organization?

- Spreadsheet location: Column W
- Data definition: classification of membership organization status. Data should only be entered for activity type pairings (rows) where activity type= 'Digital Preservation Networks'.
- Restricted data field options: Yes, No, N/A
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.

X) If Activity is Digital Preservation Network: List node hosting organizations

- Spreadsheet location: Column X
- Data definition: a listing of all organizations hosting nodes on the network. Note: all organizations within this list should also have their own rows tied to the activity. This would be explicitly listed on the website.
- *Data formatting notes:*
 - *Enter each organization as a Proper Noun*
 - *Enter each new organization on a new line within the cell. Press 'Alt' and 'Enter' simultaneously to add line break within the cell. Alternatively, use notepad to create or clean the list from the web and then paste the clean list into the cell.*

Y) If Activity is Digital Preservation Network: Technical Administration Organization - Technical Platform

- Spreadsheet location: Column Y
- Data definition: Composite data element listing both the organization taking care of the technical administration of the network, and the technical platform being used to support the network.
- *Data formatting notes:*
 - *List both organizations and technical platforms as proper nouns, separated by a hyphen, e.g. Educopia - MetaArchive.*