

2015 NATIONAL AGENDA FOR DIGITAL STEWARDSHIP

NDSA NATIONAL DIGITAL STEWARDSHIP ALLIANCE

September
2014

A report on the challenges, opportunities, gaps, emerging trends, and key areas for research and development that support the national capacity for digital stewardship.

Authored by the NDSA Coordinating Committee and NDSA Working Group co-chairs



This work is licensed under a [Creative Commons Attribution 3.0 Unported License](http://creativecommons.org/licenses/by/3.0/).

Persistent URL: <http://hdl.loc.gov/loc.gdc/lcpub.2013655119.1>

EXECUTIVE SUMMARY

The “National Agenda for Digital Stewardship” provides funders, decision-makers, and practitioners with insight into emerging technological trends, gaps in digital stewardship capacity, and key areas for research and development to support the work needed to ensure that today's valuable digital content remains accessible, useful, and comprehensible in the future, supporting a thriving economy, a robust democracy, and a rich cultural heritage.

This “Agenda” integrates the perspective of national experts and leading institutions. The “Agenda” is released by the National Digital Stewardship Alliance, a membership organization of leading government, academic, nonprofit, and private sector organizations with digital stewardship responsibilities. Members of the NDSA collaborate to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations.

The “Agenda” outlines the challenges and opportunities related to digital preservation activities in four broad areas: *Key Issues in Digital Collection Building*, *Organizational Policies and Practices*, *Technical Infrastructure Development*, and *Research Priorities*. Each section articulates priority challenges, and then offers a set of *Actionable Recommendations* to address the challenges.

Changes in the Climate for Stewardship

The last ten years have seen strong global trends in the production and use of digital content. The theme of the decade has been *more*: more information being produced; more content being published and shared; more forms of publication and filtering; more public access to information; and more collaborators coming together to learn, use, and create new content. There is increasing recognition by businesses, research institutions, policy makers, and funders that digital content, thoughtfully managed, not only supports a thriving cultural heritage sector, but also contributes more broadly to positive job creation and international competitive advantage. More has been the theme of digital stewardship as well. More work is being done to steward digital content than ever before. “Digital preservation makes headlines now, seemingly routinely. And the work performed by the community... is the bedrock underlying such high profile endeavors.”¹ A recent example being the appearance of a deleted blog post in the Internet Archive Wayback Machine that could be evidence in the MH17 plane crash in Ukraine.² Even with examples of digital stewardship successes in the short-term, a major long-term challenge is developing approaches to managing , in a transparent and authentic way, support and context for the massively increasing volume of digital content at levels of rapid upward scalability.

Last year the White House issued a major directive requiring agencies to increase open access to publications and data from federally funded research;³ the National Institutes of Health, the world’s largest public funder of research, launched a major new program focused on the use and management of big data, and appointed its

¹ Kirschenbaum, M. (2014, July 22). *Software, It's a Thing*. Presentation at Digital Preservation 2014. Retrieved from <https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3>

² Taylor, N. (July 28, 2014). The MH17 Crash and Selective Web Archiving. The Signal Blog. Library of Congress. Retrieved from <http://blogs.loc.gov/digitalpreservation/2014/07/21503/>

³ Holdren, J. P. (2013). "Increasing Access to the Results of Federally Funded Scientific Research." *Office of Science and Technology Policy*. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

first head of data science.⁴ The long-running case addressing Google's scanning of millions of books was also decided, allowing scanning, searching, and other "fair uses" to continue unimpeded.⁵ Commercial services such as Amazon Glacier, Rosetta, Preservica and community-based platforms such as LOCKSS, DuraCloud, and the Digital Preservation Network continue to develop substantial functionality in support of medium and long-term stewardship. In the research area, EU funding of projects such as SCAPE and 4C are advancing our knowledge of effective digital preservation execution and planning.

The first iteration of the "Agenda"⁶ was released for 2014. It identified opportunities and recommendations for addressing the most pressing technical, institutional, legal, and economic challenges faced by the digital preservation community. The 2015 edition of the "Agenda" builds on the earlier work, updating the 2014 text, and identifying high-level action recommendations, directed at funders, researchers and organizational leaders, that will advance the community capacity for digital preservation, the evidence base for efficient and reliable practice, and the network of durable content that is available to the nation. The 2015 "Agenda" is another step on a continuum towards successful stewardship. It includes specific actions that can be taken now, recognizing that it will require reflection, iteration, and refinement to identify comprehensive and effective interventions for the entire breadth of systems involved in stewardship.

Key Issues in Building Digital Content Collections

Much of the investment and effort in the field of digital preservation has been focused on developing technical infrastructure, networks of partnerships, education and training, and establishing standards and practices. Little has been invested in understanding how the stewardship community will coordinate the acquisition and management of born-digital materials in a systematic and public way. A gap is starting to emerge between the types of materials that are being created and used in our society and the types of materials that make their way into libraries and archives. The stewardship community must recognize this gap, understand why it exists, and determine how it could be addressed at local, regional, and national levels.

Overarching Challenges of Digital Content

Both born-digital and digitized content present fundamental new issues to stewards tasked with ensuring meaningful long-term access to content. The need for effective digital stewardship is urgent, because content and the context that makes it meaningful is changing rapidly. Moreover, effective digital stewardship requires collaboration and coordination, because organizations rely on information beyond their institutional boundaries. Stewardship of the content each organization uses is impossible for any single organization to do on its own, and avoiding the risks of loss depends on actions taken across the community.

The stewardship community needs to develop a broad evidence base describing the *practice and content of stewardship*, including identifying the types and collections of content that are being used by its members, who

⁴ NIH Office of the Director. (2013 December 9). NIH Names Dr. Philip E. Bourne First Associate Director for Data Science." *News and Events*. Retrieved from <http://www.nih.gov/news/health/dec2013/od-09.htm>

⁵ Miller, C.C., Bosman, J. (2013 November 14). Siding With Google, Judge Says Book Search Does Not Infringe Copyright. *New York Times*. Retrieved from <http://www.nytimes.com/2013/11/15/business/media/judge-sides-with-google-on-book-scanning-suit.html>

⁶ National Agenda for Digital Stewardship. (2014) National Digital Stewardship Alliance. Retrieved from <http://www.digitalpreservation.gov/ndsa/nationalagenda/index.html>

is taking responsibility for these collections, what organizations have the capability to take on additional stewardship, what organizations can provide long-term access to preserved materials, and identifying what stewards are doing to reduce the risk of loss inherent to digital information. This evidence base could be used to identify where we are doing well, where the gaps are, where the single points of failure are, and where the opportunities are to coordinate to risk reduction to important collections.

Approaches to Content Selection at Scale

Both libraries and archives have established concepts of selection and appraisal that are meant to guide curators in making these often subjective decisions. It is difficult to evaluate how well libraries, archives, and museums are collecting and preserving the large amounts of digital data that their users, patrons, researchers, and institutions rely on. Traditional forms of scholarship like articles, edited volumes, and monographs (and their digital equivalents) are fairly well-understood in terms of how they fit into an institution's collection strategy. The non-traditional forms of evidence, like much of the data found on the open web, do not easily fit into existing acquisition processes. In addition, the usage data and logs that augment the open web data are becoming just as significant to researchers. The following recommendations and the discussion later in the document are meant to advance a fuller understanding of approaches to selecting in the digital environment.

Core Digital Content Recommendations:

- *Build the evidence base for evaluating at-risk, large-scale digital content for acquisition.* Develop contextual knowledge about born-digital content areas that characterizes the risks and efforts to ensure durable access to them.
- *Understand the technical implications of acquiring large-scale digital content.* Extend systematic surveys and environmental scans of organizational capacity and preservation storage practices to help guide selection decisions.
- *Share information about what content is being collected and what level of access is provided.* Communicate and coordinate collection priority statements at national, regional, and institutional levels.
- *Support partnerships, donations and agreements with creators and owners of digital content and stewards.* Connect with communities across commercial, nonprofit, private, and public sectors that create digital content to leverage their incentives to preserve.

Organizational Policies and Practices

Despite continued preservation mandates and over ten years of work and progress in building a comprehensive practice around digital preservation, the community still struggles with advocating for resources, adequate staffing, and articulating the shared responsibility for stewardship. Underlying all of these challenges is a lack of prioritization of digital preservation programs. Integrating digital stewardship practice and thinking across an entire organization is a core challenge, especially in a time of restricted resources. Part of the challenge is giving decision makers the information they need to make informed decisions and manage organizations that steward digital materials. A significant part of digital stewardship, especially work in standards and practices, is related to and directly impacts private industry—which owns or creates a large amount of digital content created today. Engagement with the commercial sector is important to ensuring the preservation of contemporary culture.

Efforts in the area of organizational roles and policies for digital stewardship should be focused on the following objectives. These are actions for which practitioners, managers, stakeholders, and funders can advocate and implement, as they work toward an environment where the mandate and need for digital preservation are matched with the resources, staffing, and an effective professional community prepared to meet those mandates and needs.

Core Organizational Policies and Practices Recommendations

- *Advocate for resources.* Share strategies and develop unified messages to advocate for funding and resources; share cost information and models; and develop tools and strategies that inform the evaluation and management of digital collection value and usage.
- *Enhance staffing and training.* Explore and expand models of support that provide interdisciplinary and practical experiences for emerging professionals and apply those models to programs for established professionals. Evaluate and articulate both the broad mix of roles and the specialized set of skills in which digital stewardship professionals are involved.
- *Foster multi-institutional collaboration.* Foster collaboration through open source software development; information sharing on staffing and resources; coordination on content selection and engagement with the development of standards and practices; and identify, understand and connect with stakeholders outside of the cultural heritage sector.

Technical Infrastructure Development

Broadly speaking, the infrastructure that enables digital preservation involves the staff, workflows, resources, equipment, and policies that ensure long-term access to digital information. This section focuses specifically on the technical component of that infrastructure. Technical infrastructure can be generally defined as the set of interconnected technical elements that provide a framework for supporting an entire structure of design, development, deployment, and documentation in service of applications, systems, and tools for digital preservation. This includes hardware, software, and systems. Organizational policies, practices, and regulations inform many of the observations and recommendations for the development of digital stewardship technical infrastructure.

Core technical Infrastructure Recommendations:

- *Coordinate and sustain an ecosystem of shared services.* Better identify and implement processes to maintain key software platforms, tools and services; identify technologies which integrate well to form a sustainable digital workflow; and identify better models to support long-term sustainability for common goods are needed.
- *Foster best practice development.* Give priority to the development of standards and best practices, especially in the areas of format migrations and long-term data integrity.

Research Priorities

Research is critical to the advancement of both basic understanding and the effective practice of digital preservation. Research in digital preservation is under-resourced, in part this is because the payoff from long-

term access occurs primarily in the medium-long term and tends to benefit broad and diverse communities. Investments in core research will yield large impacts.

Core Research Recommendations:

- *Build the evidence base for digital preservation.* Give priority to programs that systematically contribute to the overall cumulative evidence base for digital preservation practice and resulting outcomes—including supporting test beds for systematic comparison of preservation practices.
- *Better integrate research and practice.* Give priority to programs that rigorously integrate research and practice or that increase the scalability of digital stewardship.

Conclusion

A common challenge running through this report is the limited amount of empirical evidence available. The digital preservation community is beginning to develop a shared evidence base; however, studies must be broadened and repeated over time to establish a robust evidence base from which generalizable guidance can be drawn. Furthermore, decision makers should recognize that basic research needs to be paired with the development, support, and evaluation of infrastructure to support a burgeoning and complex digital content environment and attention to the organizations and policies that will support successful long-term digital stewardship.

ABOUT THE NATIONAL DIGITAL STEWARDSHIP ALLIANCE

Founded in 2010, the National Digital Stewardship Alliance (NDSA) is a consortium of institutions that are committed to the long-term preservation of digital information. NDSA's mission is to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations. The NDSA comprises over 160 participating institutional members. These members come from 45 states and include universities, consortia, professional societies, commercial businesses, professional associations, and government agencies at the federal, state, and local level. NDSA organizations have proven themselves committed to long-term preservation of digital information.

To learn more about the NDSA: <http://www.ndsa.org>

ABOUT THE AUTHORS

The joint leadership group of the NDSA authored the report and engaged in discussions to identify significant trends and challenges. The membership of the NDSA contributed to these discussions. This dialog was enriched by an extensive range of resources and current research. The joint leadership group is made up of the Coordinating Committee members, the Working Group co-chairs, and the NDSA facilitator:

Micah Altman (MIT), Jefferson Bailey (Internet Archive), Karen Cariani (WGBH), Jim Corridan (Indiana Commission on Public Records), Jonathan Crabtree (UNC, Chapel Hill), Michelle Gallinger (Library of Congress), Andrea Goethals (Harvard Library), Abbie Grotke (Library of Congress), Cathy Hartman (University of North Texas), Butch Lazorchak (Library of Congress), Jane Mandelbaum (Library of Congress), Carol Minton Morris (DuraSpace), Kate Murray (Library of Congress), Trevor Owens (Library of Congress), Megan Phillips (NARA), Abigail Potter (Library of Congress), Robin Ruggaber (University of Virginia), John Spencer (BMS/Chace), Helen Tibbo (UNC Chapel Hill), Kate Wittenberg (Portico).