The 2015 National Agenda for Digital Stewardship
The 2015 National Agenda for Digital Stewardship

What Works in Digital Preservation, and What is Needed

Presented by:

Micah Altman, escience@mit.edu
Director of Research, MIT Libraries
Chair, Coordination Committee, National Digital Stewardship Alliance

Michelle Gallinger, Gallinger Consulting, mgalinger@gallingerconsult.com

Trevor Owens, Digital Archivist, The Library of Congress, trow@loc.gov
DISCLAIMER

These opinions are my own, they are not the opinions of MIT, Brookings, any of the project funders, nor (with the exception of co-authored previously published work) my collaborators

Secondary disclaimer:

“It’s tough to make predictions, especially about the future!”

Attributed to Woody Allen, Yogi Berra, Niels Bohr, Vint Cerf, Winston Churchill, Confucius, Disreali [sic], Freeman Dyson, Cecil B. Demille, Albert Einstein, Enrico Fermi, Edgar R. Fiedler, Bob Fourer, Sam Goldwyn, Allan Lamport, Groucho Marx, Dan Quayle, George Bernard Shaw, Casey Stengel, Will Rogers, M. Taub, Mark Twain, Kerr L. White, etc.
Collaborators & Co-Conspirators

• The over 160 institutional members of NDSA, and the tens of thousands of hours contributed by their representatives to NDSA working groups, meetings and reports

• National Agenda Authors:
  Jefferson Bailey, Karen Cariani, Jim Corridan, Jonathan Crabtree, Blaine Dessy, Michelle Gallinger, Andrea Goethals, Abigail Grotke, Cathy Hartman, Butch Lazorchak, Jane Mandelbaum, Carol Minton Morris, Trevor Owens, Meg Phillips, Abbie Potter, John Spencer, Helen Tibbo, Tyler Walters, Kate Wittenberg, Kate Zwaard
Who needs digital preservation?
No one needs digital preservation... for its own sake

- And no one needs drills ...
- ... microscopes ...
- ... fiber optic cable ...
- ... TCP/IP protocols ...
- ... portland cement ...
- ... the world-wide-web...
Preservation & Stewardship = Durable, Long-Term Access

• The rule of law require maintaining authentic public records
• Scientific advances rely on a cumulative, traceable evidence base
• Art, history, culture require durable access to national heritage information
• Our nation needs durable access to a strategic information reserve
• Humanity needs durable long-term access information in order to communicate to future generations
A roadmap towards coordinated stewardship of digital information
The 2015 National Agenda for Digital Stewardship

ndsa.org/nationalagenda/
The Process

• 2nd Edition
  -- refined priority areas
  -- updated content
  -- action steps

• Input from > 160 NDSA member institutions
  -- comments and contributions directly and through working groups
  -- synthesized by working group chairs, coordinating committee & secretariat

• Integrates comments from 10 expert external reviewers
What’s New: Extending the Roadmap to Actionable Opportunities

• First edition of Agenda identified
  – Key trends
  – Key challenges

• 2015 edition calls out
  – Action steps
Summary: What’s Working

- Increasing attention
  - externally to information and stewardship
  - within organizations to access

- Content
  - more content being preserved
  - successful preservation of some common collections

- Technology
  - Understanding of formats
  - Prioritizing replication
  - Prioritizing file fixity

- Successful broad institutional collaborations
  - on practices, projects, and tools

Where we need to go...

- Systematic measurement and evaluation of success
- Adapting business models to sustain collaboration
Core Recommendations: Digital Content

• 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
National Agenda for Digital Stewardship

Identifies high-impact opportunities for the state of the art, practice & collaboration

Aimed at senior institutional decision makers.

- Digital Content Areas
- Organizational Roles, Policies, and Practice
- Technical Infrastructure
- Research
Digital is Different

• Digital Content is Different
  Accessible; Replicable; Computable; Changeable

• Why organizations no longer own or manage all of their critical information assets...
  – Too much information produced and published
  – Increase in collaborators, learners, information consumers
  – Increasing attention to information transparency & access
  – Information increasingly is embedded in a larger context necessary for understand it
Core Recommendations: Digital Content

• **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.
Digital Content: Areas for Targeted Development

• Content selection at scale
• Connection to researchers
• Connection creator community
• Software preservation
Digital Content: Approaches to Selection at Scale

- Large scale data doesn’t fit easily into existing acquisition processes.
- Strategies are needed for actively acquiring unique born-digital materials strengthens special collections.
- Need to articulate how digital collecting supports the organizational mission.
Digital Content:
Connection to Researchers

- Usability is a driver for digital preservation
- Models for Access are being developed. More are needed.
- Increasing demand for enhanced use options and tools
- Increasing demand for data sets at data analysis level -- researchers want raw data
- Analytic tools

The 2015 National Agenda for Digital Stewardship
Digital Content:
Connection to Creator Community

- Incentives to preserve
- Expanding knowledge of digital preservation in creative content communities
- Coordinated national approach with deeply engaged stakeholders
Digital Content: Advancing Software Stewardship

**Ongoing Work and Work to Date:**
- Preserving.exe summit/report
- Internet Archive JSMESS
- KEEP & Olive Executable Archive

**Work to Watch:**
- Software Preservation Network
- CLIR Software Curation Post Docs
- Emulation work at Rhizome & Yale Libraries
Digital stewardship has yet to receive a high priority in many organizations.
Organizational Policies & Practices

“Digital is not important”

Evidence of the value of digital preservation

- data availability drives innovation
- reporting collection usage
- sharable metrics about quality and success of preservation
Organizational Policies & Practices

“All digital content is important”

Make appropriate resource decisions that lead to action

- Never enough resources. “Save everything” can be paralyzing.
- Advocate for appropriate resources
- Reallocate resources as necessary
- Prove value of holdings through use
- Refine cost estimates and contribute to better costing models
Organizational Policies & Practices

Organizations should share information on staffing practices

- Digital holdings will grow substantially
- Organizations will need more staff working with digital
- 75% of responding organizations are retraining staff
Core Recommendations: Organizational Policies and Practices

- **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.
Core Recommendations: Technical Infrastructure Development

• 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.
Technical Infrastructure: Targeted Development

• Software preservation
• Coordinating distributed service ecosystem
• File formats
Applied Technical Development: Software Preservation

• Organizations need to identify what stake they have in software collecting/preservation.

• Need for engagement with the software industry in these problems and issues.

• Investments in research and tool development for virtualization and emulation of computing environments (Olive, KEEP, JSMESS).

• Basic research is still required in preservation of entire computational environments and software that runs as web applications.
Applied Technical Development: Coordinating an Ecosystem of Distributed Services

• Examples of Successes:
  – DuraSpace/Duracloud,
  – Archive-It
  – Portico
  – Bagit

NDIIPP Architectures of Collaboration (ca. 2009)

The 2015 National Agenda for Digital Stewardship
Applied Technical Development: Ecosystem Challenges & Recommendations

• a significant challenge for digital stewardship is the continued reliance on infrastructures subject to the vagaries project oriented funding and a lack of effective coordination in ensuring that community-developed tools have models for sustainability.

• Stewardship organizations and funders need to focus on identifying and implement processes to maintain key software platforms, tools, and services.

• Now that stewardship organizations are amassing large collections of digital materials, it is important to shift from more abstract considerations about file format obsolescence to develop actionable strategies for monitoring and mining information about the heterogeneous born-digital files the organizations are managing, especially the formats that don’t result from digitization activities.
Applied Technical Development: Next Steps in File Formats

• Stewardship organizations should document and share info on file formats they currently manage to inform research and development.

• Stewardship organizations should prioritize the development of file format action plans that most appropriately reflect the kinds of content they are actually managing.

• To inform acquisitions, institutions should use of comment on the Library of Congress Recommended Format Specifications.
Core Recommendations: Research

• **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.
Research Priorities

• Strengthening the Evidence Base for Digital Preservation

• Stewardship at Scale
  – “Big Collections”
  – Systematic Value Models for Selection at Scale

• Targeted Applied Research
Research Priority: 
*Strengthening the Evidence Base*

– Prioritize programs that systematically contribute to the overall cumulative evidence base for digital preservation practice and outcomes.

– Prioritize rigorously integrate research and practice.

– Go beyond case-studies in their approach, and include replicable methods to support systematic inference.
Research Priority: 
*Stewardship at Scale*

– Prioritize programs and projects that increase the scalability of digital stewardship.

– Analyze challenges of “big” collections that go beyond size and storage, to dealing with the variety, and velocity of big data and big collections across all phases of the curation lifecycle.

– Develop new selection and appraisal models that are evidence based, non-labor intensive, and evaluate portfolios of information.
Research Priority: 
Applied Research

Prioritize small number of targeted applied areas of research that constitute special opportunities for improving the *reliability and efficiency of preservation practice*

– Cost modeling
– Computable information equivalence
– Trust frameworks and policy systems
About the NDSA

NDSA

Content  Standards  Infrastructure  Innovation  Outreach

The 2015 National Agenda for Digital Stewardship
NDSA Mission

To establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations.
(Brief) History the NDSA –

• The *National Digital Stewardship Alliance* (NDSA) was launched in 2010 as an initiative of the *National Digital Information Infrastructure and Preservation Program* (NDIIPP), with a core of NDIIPP partner organizations.

• Over the last four years, the NDSA has grown rapidly, to include 165 organizations that represent all sectors, and include universities, consortia, professional associations, commercial enterprises, and government agencies at the federal, state, and local levels.
Organizational Collaboration is Vital for Effective Stewardship

• Effective stewardship requires *knowing what information is at risk*...
  – What information is already being stewarded by other organizations
  – What critical tools, standards and policy are in development

• Effective stewardship requires *understanding what works*...
  – Integrating research from many fields
  – Discovering, adapting, and implementing good practice
  – Practice, research, tool and standards development is spread across...*Disciplines; sectors & communities of practice*

• Without collaboration, substantial uncontrolled risks remain:
  – External risks to organizations:
    Third party attacks; shifts in sfunding; Change in legal regimes
  – Internal organizational failure:
    Unintentional curatorial modification; Loss of institutional knowledge & skills; Intentional curatorial de-accessioning; Mission change
How NDSA Works

• **Based** on collaborative community effort and partnership -- there are no fees for NDSA membership.

• The **Library of Congress** has provided essential organizational support, logistical support, and substantive collaboration and expertise.

• Each NDSA member organization joins by being vetted, committing to the NDSA principles.

• Each member *contributes* efforts to working groups, reports, surveys, meetings and other NDSA initiatives.

[digitalpreservation.gov/ndsa/memberslist.html](http://digitalpreservation.gov/ndsa/memberslist.html)

The 2015 National Agenda for Digital Stewardship
NDSA Initiatives

**Working Groups**
- **Content**: selection and acquisition of digital collections
- **Standards**: digital formats and best practices
- **Infrastructure**: development and maintenance of tools
- **Innovation**: cutting edge research and development
- **Outreach**: relationships and messaging

**Recent Outputs**
- **Extending Knowledge**
  - Preservation Storage Survey
  - Web Archiving Survey
  - Preservation Staffing Survey
  - Geospatial Selection & Appraisal report
  - Content case studies
  - NDSA Interview Series
- **Tools for Practice**
  - Levels of Preservation
  - Digital Preservation in a Box
  - Digital Preservation on Wikipedia
- **Dissemination**
  - National agenda for digital stewardship
  - NDSA Innovation Awards
  - NDSA Social Media
  - Digital Preservation pioneers

The 2015 National Agenda for Digital Stewardship
Want more information?

Contact NDSA for...

• Information on becoming an institutional member
• Briefings, webinars, and consultations on the Agenda or other NDSA work
• Assistance in gathering comments on National policies and programs
• Assistance in recruiting experts for review and discussion panels
• Referrals to content stewards in specific areas
More Information

Visit the website: ndsa.org

Contact me: <escience@mit.edu>