



# Stories of Risk and Value: the National Digital Stewardship Alliance Content Case Studies

Abbie Grotke, Library of Congress

Margaret K. Maes, Legal Information Preservation Alliance

Linda Reib, Arizona State Library

Best Practices Exchange, December 5, 2012





- An initiative of the National Digital Information Infrastructure and Preservation Program (NDIIPP)
- A collaborative effort among government agencies, educational institutions, non-profit organizations, and business entities to preserve a distributed national digital collection for the benefit of citizens now and in the future.
- Community-driven, action-oriented





## Values

- Stewardship: Commitment to managing digital content over the long-term
- Collaboration: The strength of communities working together
- Inclusiveness: A broad range of participants and approaches increases the likelihood that content will be preserved
- Exchange: The open sharing of ideas, services and software





# Objectives

- The Library of Congress and Members of the NDSA are committed to serving as digital stewards of America's national digital collection and employing standards, systems, and cooperative relationships that advance digital stewardship.
- Members commit to pursuing and participating in collaborative stewardship endeavors for at-risk digital content.





# NDSA Working Groups

Members commit to participating in one of the five working groups of the Alliance:



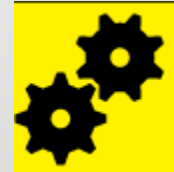
Content



Outreach



Innovation



Infrastructure



Standards & Practices





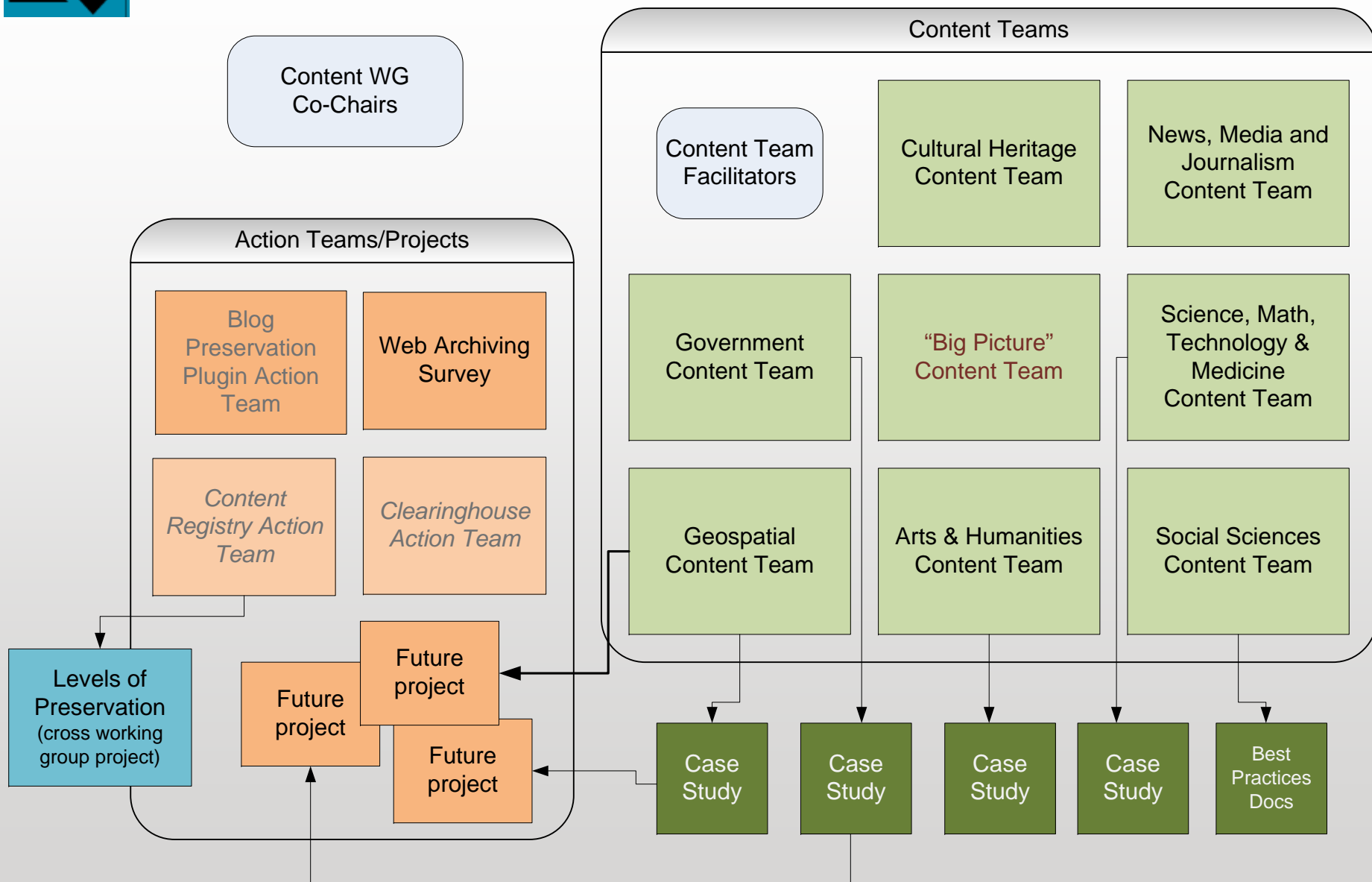
# Content

Identify content already preserved, investigate guidelines for the selection of significant content, enable discovery of at-risk digital content or collections, and match orphan content with NDSA partners who will acquire the content, preserve it, and provide access to it.





# Content Working Group



You are viewing an archived web page, collected at the request of National Library of Medicine using Archive-It. This page was captured on 13:57:11 Mar 29, 2012, and is part of the Health and Medicine Blogs collection. The information on this web page may be out of date. See all versions of this archived page. [MetaInfo](#)



**The Butter Compartment**  
WHERE DO YOU KEEP YOUR INSULIN?  
DIABETES BLOGGER & ART THERAPIST  
Advertisement

---

**Handful**  
Filed under: [Blogs](#), [Science](#)  
Blood Glucose

There's an in hope for a c see it. It's b never cured

**Greg Lang**  
Lewis-Sigler Institute for Integrative Genomics  
Princeton University

[Welcome](#)
[My Research](#)
[Papers](#)
[Strain Collections](#)

[Curriculum Vitae](#)
[Notebooks](#)
[Presentations](#)
[Protocols](#)



In front of a wall of used 96- and 384-well plates from a 1000-generation long-term evolution experiment, impressive looking, but foul smelling.

Like 4 people like this. Sign Up to see what your friends like.

Hello. This is my attempt at a semi-professional website describing my research.

My research is aimed at understanding the molecular basis of evolution. The yeast *Saccharomyces cerevisiae* is an ideal system for studying evolution in the laboratory: yeast is amenable to genetics and genomics studies, and many strains of *S. cerevisiae* as well as closely related species have been sequenced. These tools allow us to analyze historical events that led to the modern yeast species.

In addition to these case studies in evolution, I am interested in general principles of evolution such as the nature of beneficial and deleterious mutations and how parameters such as mutation rate, the distribution of mutational effects on fitness, and epistasis set the rate of adaptation.

In 2007 I completed my Ph.D. with [Andrew Murray](#) at Harvard University where I characterized mutation rate variation in yeast. Currently I am a Post-doctoral fellow in [David Botstein's](#) laboratory at Princeton University.

Lewis-Sigler Institute for Integrative Genomics Princeton University  
120 Carl Icahn Laboratory Princeton, NJ 08544  
Phone: 609.258.8044 FAX: 609.258.8020  
[glang@princeton.edu](mailto:glang@princeton.edu)

- Share compelling stories that demonstrate the value of digital preservation in our communities.
- Engage all members of the community in the preservation of content
- Encourage the cultivation of relationships that could enable preservation
- In conjunction with other working groups in the NDSA, identify actionable next steps to ensure important content is preserved.



ABOUT | FAQ | CONTRIBUTE

## Geospatial Data Preservation

**Featured Resource**

[Geospatial Data Transfer: Archival Data Transfer Validation and Dataset Functional Verification \[Web\]](#)

**QUICK LINKS FOR**

- Data Managers
- System Developers
- Researchers

**WHAT'S NEW**

[Suggest resources!](#)

[Advancing Strategic Science: A Spatial Data Infrastructure Roadmap for the U.S. Geological Survey](#)

[OGC Best Practices](#)

[Using Semantic Metadata for Discovery and Integration of Heterogeneous Ecological Data](#)

**Featured Practice**

### Format Descriptions for Geospatial Data

Explore descriptions of formats used for geospatial data and how to assess them for potential use.

Find resources via free text search on titles, authors, and key terms....

**EDUCATION & TRAINING** **TOOLS & SOFTWARE** **POLICIES & BENEFITS**

**BROWSE BY TOPIC** **BROWSE BY TYPE** **ADVANCED SEARCH**

**Featured Website**

### Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

Explore the results of the GeoMAPP project and learn about recommended practices for geospatial data stewardship.




- Focus digital preservation efforts on the content that is most valuable, so that material of historic or research importance is not lost
- Focus digital preservation efforts on the content that is most at risk, so that material in danger of disappearing for any reason is not lost
- Provide persuasive arguments for devoting scarce institutional resources to digital preservation activities
- Provide justification and background information for grant proposals and other fund raising efforts

- Community and Hyperlocal News
- Citizen Journalism
- Digital E-Prints of Newspapers
- Science, Medicine, Math, and Technology Blogs
- Science Forums

You are viewing an archived web page, collected at the request of [National Library of Medicine](#) using [Archive-It](#). This page was captured on 14:58:20 Dec 20, 2011, and is part of the [Health and Medicine Blogs](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)



HOME SUBSCRIBE CHANGE OF SHIFT ARCHIVES BLOGROLL CONTACT ME BLOGWORLD DISCLAIMER

November 16, 2011, 9:02 pm

## A Message From "Rank and File, RN"



Looks like Mr. Administrator has some 'splainin' to do! And I suggest he be straight with my colleagues.

**About Me**  
My name is Kim, and I'm a nurse in the San Francisco Bay area. I've been a nurse for 23 years, graduated in 1988 with my ADN. My experience is predominantly Pediatric and Adult Care.

**Direct apps for students \*\*\*\*\***  
**Nursing Essentials™** (iTunes & Android)  
**RN** "What's Hot" on iTunes! [View Now](#)



## the ann arbor chronicle

ANN ARBOR, MICHIGAN SUBSCRIBE TEXT SIZE: [A](#) [A](#) [A](#) Google™ Custom Search [Search](#)

### Around the Chronicle

- » FRONT PAGE
- » COMMENTS
- » STOPPED, WATCHED.
- » CIVIC NEWS TICKER
- » EVENTS

### More Local Coverage:

- The Ann Arbor Observer
- Lucy Ann Lance

### Civic News Ticker

**AATA Calls Special Board Meeting for Nov. 29**  
A special meeting of the Ann Arbor Transportation Authority board has been called for Thursday, Nov. 29, 2012 starting at 4 p.m. at the AATA headquarters, 2700 S. Industrial Highway. The meeting, which was announced via email on Nov. 21, did not have an agenda set until Nov. 28. The agenda includes a closed session and an item that would revise the AATA's advertising policy. [pdf of board packet, including revised advertising policy]

### Column: Brady Hoke's Sophomore Slump

Charming or cheesy? Too soon to grade Michigan's football coach

BY JOHN U. BACON NOVEMBER 30, 2012 at 9 am

Exactly one year ago, Brady Hoke was the darling of Michigan football fans.

He'd charmed even the doubters at his first press conference – where he coined his now famous phrase, "This is Michigan, for God's sake!" – then led a team that had averaged just five wins a season over the previous three years to a 10-2 regular-season record, including thrilling wins over Notre Dame, Nebraska and arch-rival Ohio State. Then he capped it all off with an overtime upset of Virginia Tech in the prestigious Sugar Bowl – Michigan's first BCS bowl victory since a young man named Tom Brady beat Alabama on Jan. 1, 2000.



John U. Bacon

The man could do no wrong. When Hoke started referring to injuries as "boo-boos" and Ohio State as "Ohio," fans did not think he was an ignoramus who knew nothing about the greatest rivalry in sports – as they surely would have if Rich Rodriguez had said the same things – but a motivational genius, who understood exactly what the duel was all about.

When fans noticed Hoke did not wear a headset during games – unlike just about every other coach in the country – they did not conclude he was an out-of-touch, glorified cheerleader, but a master delegator and teacher, trusting the play calling to his assistants while he focused on coaching his players.

When you're winning, everything's cool. But when you start losing, the same people who patted you on the back start questioning your quirks. [Full Story]

### Stopped. Watched. \*

Argo & Chandler  
Near Beckley Park, fancy new Popemobile street sweeper cleaning leaves from the gutters outside my window. [photo]

— PAID ADVERTISEMENT —

### Exceptional Students



**Ann Arbor Tech High School has the answer.**  
For A² Tech students – formerly Stone High School – credentials matter and the name of their school needed to better reflect the school's vision.  
A² Tech offers opportunities for students to complete...



# Case Studies

- Establish the value of the content and provide the rationale for selecting it for preservation.
  - What value does the content have? Is anyone collecting it?
  - What factors affect its risk of disappearance?

## *Science Blogs...*

- Reflect and may influence a wide range of viewpoints, and capture dissenting opinions and new trends not found in other publications
- Serve as a record of who did what and when and will serve as a data mine for understanding trends in scientific practices and attitudes, and how ideas evolve
- Provide insight into the environment of science, medicine, technology, and mathematics
- Show how science, medicine, technology, and mathematics is communicated among collaborators and broadly in the scientific community





# Case Studies

- Document recognized opportunities for preserving this content.
  - Are there workflows in the creation or distribution of content that present opportunities for preservation?

## *Digital e-prints...*

Newspapers produce a PDF printmaster to send to their printer. Most newspaper publishers neither preserve the PDF printmasters nor do they microfilm the printed issues, resulting in a loss of current newspaper content for future generations. The fact that the vast majority of current U.S. newspapers are printed from an "e-print" file offers a unique opportunity to work with publishers to capture and preserve that file.

## *Science forums...*

- Interest in born digital science content preservation among content creators evident in establishing sessions on preserving science blogs at Science Online
- Potential to incorporate this into the NDIIPP report on the Preserving Online Science meeting





# Case Studies

- Describe target audiences/  
stakeholders.
  - Who would find value in this content and how might they be engaged in the process of preservation?

## *Community and Hyperlocal News...*

These audiences share multiple levels of local, state, national, and international interest:

- Local communities: public libraries and historical societies; local newspaper publishers; genealogical societies; county and local governments; K-12 educators
- State-wide communities: Academic researchers; lay historians; university students and professors; archives
- National and International communities: Trending researchers; political scholars; economic analysts





# Case Studies

- Outline a plan for educating stakeholders.
  - How might NDSA or another organization raise the awareness of stakeholders including content creators, publishers, educators, libraries, researchers, or donors?

## *Citizen Journalism...*

- Content creators: Online campaign to bring awareness to importance of preservation; targeted messages. Blogger conventions.
- Publishing platforms: Build and promote easy-to-use plugins and tools that allow opt-in for preservation or better ways of personal archiving and instructions for delivery to preservation organizations (Wordpress, Tumblr, etc.)
- Funders: J-lab and Knight Foundation
- Journalism schools
- Public libraries and historical societies: Educate and partner with public libraries on local news preservation; workshop presentations at district library association meetings; conference calls with public library directors.
- Historical researchers, professors, and students: Connect at journalism and historical association meetings and conferences; panel presentations at archivist society conferences; vendor booths, brochures, and flyers about digital preservation at relevant conferences.



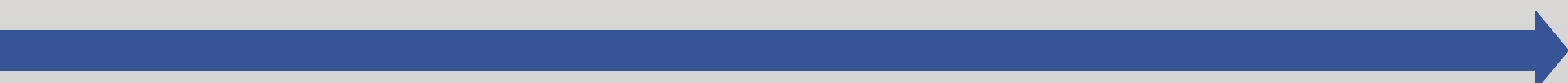


# Case Studies

- Describe potential obstacles or risk factors.
  - What barriers for users/creators/preservationists might be faced and what options are there for overcoming them?

## *Digital e-prints...*

- PDF printmasters are not routinely maintained by many publishers.
- Neither libraries nor publishers currently recognize PDF printmasters as preservation master copies that need to be actively preserved on a regular basis.
- Many newspapers don't have the technical expertise or the management policy to preserve their content before it is lost.
- One preservation solution does not fit all newspapers.
- Publishers and libraries do not have a highly successful track record of cooperation to date. Each have different motivations and do not always understand other's importance.
- Publishers are uncomfortable giving permission to third parties to make newspapers freely available online.
- Funding for digital preservation may not be readily available nor is there awareness of the urgency for this "at risk" content.
- Grant-funded staffing is common in most digital newspaper programs across the U.S. right now, and funding any preservation project on grant support raises sustainability concerns.







# Case Studies

- Develop actionable next steps.
  - What can we do next, as a community or as individual institutions to ensure that important content is preserved?

## *Science Blogs...*

- Create a registry of science, medicine, technology, and mathematics blogs and identifying which ones are currently being preserved.
- Promote principles of personal digital archiving. Are there clear and simple guidelines and best practices for blog authors to follow if they want to take action to preserve their blog?
- Collaborate with the producers of major blogging platforms, like WordPress, Blogger, and Moveable Type, with large numbers of subscribers, to promote best practices for personal blog preservation and potentially match the blog to a host institution for long term preservation.
- Create opportunities for bloggers to connect with organizations and institutions that can preserve their content.







# Gov. Content Team



Developing case studies that identify best practices for all phases of the lifecycle of state local and federal government records from creation, appraisal, guidance on formats, transfers of records along with their provenance and individual file content metadata, ingest and processing transfers into the archives, description of records, archival storage, versioning, risk assessment and planning, preservation actions like making transformations, and enabling access.





# Gov. Content Team

- The value of the content and the rationale for selecting it for preservation is established in records retention schedules. The content is at risk based on the inability of government institutions to preserve records due either to lack of resources or lack of understanding of best practice.
- By providing examples of best practice the team offers to other institutions the procedures and methods that some institutions believe will contribute to the establishment of trusted digital repositories based on International standards.
- Target audience: research community that use government records, and other stakeholders include both record creating and record preservation institutions, as well as the tax payers. Also, government institutions that will adapt the best practice documents to their institution.





## Gov. Content Team

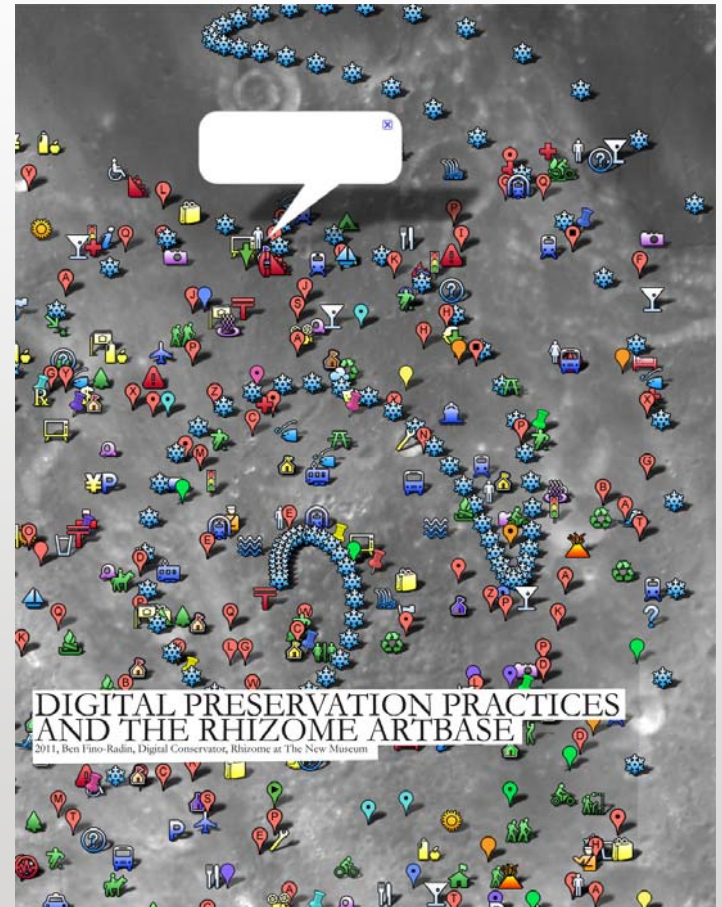
- Develop a list of commonly used electronic formats that are ranked according to three levels of preservation and access quality for each format type (video, audio, text, photo, etc.) that can aid in the development of effective transfer guidelines.
- Developed a list of agreed upon functional requirements for selecting preservation software based on the OAIS Reference model and serving governments at various levels.
- Provide sample preservation policy statements that can be used by government institutions with a similar environment.
- Provide studies of appraisal rationale and methods that facilitate the transfer of records in preservation ready and accessible formats by making records creators aware of preservation needs.





# Other Content Teams

- Geospatial: Working to address sustainability solutions for the [geopreservation.org](http://geopreservation.org) resource. Exploring opportunities to engage with the geospatial industry through invited speakers. Sharing partner geospatial collection development policies.
- Social Sciences: Working on a document: "How to provide guidance and assessment on the life of data. Not all data are meant to be kept forever."
- Arts and Humanities: brainstorming potential case studies and defining a model specifically for arts and humanities case studies.
- Cultural Heritage: Discussing closer collaboration with the Arts and Humanities Team. Working on draft case studies.





## Getting the Word Out

- Distribute and solicit feedback from NDSA members (beyond CWG)
- Publish case studies on [digitalpreservation.gov](http://digitalpreservation.gov) and/or other NDSA member websites
- Signal blog post
- Send announcements to listservs
- Present at conferences that members are attending





# NDSA

## Levels of Preservation





# Levels of Preservation

A basic tool for helping organizations manage and mitigate digital preservation risks.

Does *not* deal with broader issues related to collection development practices, critical policy framework decisions, general issues involving staffing or particular workflows or life cycle issues. Those are all critical, and in many cases are handled quite well by existing work (like the OAIS model, and the TRAC and TDR standards).





## Levels of Preservation

- Useful for developing plans — not a plan in itself
- Levels are non-judgmental
- Levels can be applied to collection(s) or system(s)
- Designed to be content and system agnostic







# Levels of Preservation

Four levels:

- One (Protect your data)
- Two (Know your data)
- Three (Monitor your data)
- Four (Fix your data)





# Levels of Preservation

Six Areas:

- Storage and geographic location
- File Fixity and Data Integrity
- Information Security
- Metadata
- File Formats
- Technology obsolescence





# Levels of Preservation

	Level One (Protect Your Data)	Level Two (Know Your data)	Level Three (Monitor Your Data)	Level Four (Repair Your Data)
<b>Storage and Geographic Location</b>	<ul style="list-style-type: none"> <li>Two complete copies that are not collocated</li> <li>For data on heterogeneous media (optical disks, hard drives, etc.) get the content off the medium and into your storage system</li> </ul>	<ul style="list-style-type: none"> <li>At least three complete copies</li> <li>At least one copy in a different geographic location</li> <li>Document your storage system(s) and storage media and what you need to use them</li> </ul>	<ul style="list-style-type: none"> <li>At least one copy in a geographic location with a different disaster threat</li> <li>Obsolescence monitoring process for your storage system(s) and media</li> </ul>	<ul style="list-style-type: none"> <li>At least 3 copies in geographic locations with different disaster threats.</li> <li>Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems.</li> </ul>
<b>File Fixity and Data Integrity</b>	<ul style="list-style-type: none"> <li>Check file fixity on ingest if it has been provided with the content</li> <li>Create fixity info if it wasn't provided with the content</li> </ul>	<ul style="list-style-type: none"> <li>Check fixity on all ingests</li> <li>Use write-blockers when working with original media</li> <li>Virus-check high risk content</li> </ul>	<ul style="list-style-type: none"> <li>Check fixity of content at fixed intervals</li> <li>Maintain logs of fixity info; supply audit on demand</li> <li>Ability to detect corrupt data</li> <li>Virus-check all content</li> </ul>	<ul style="list-style-type: none"> <li>Check fixity of all content in response to specific events or activities</li> <li>Ability to replace/repair corrupted data</li> <li>Ensure no one person has write access to all copies</li> </ul>
<b>Information Security</b>	<ul style="list-style-type: none"> <li>Identify who has read, write, move, and delete authorization to individual files</li> <li>Restrict who has those authorizations to individual files</li> </ul>	<ul style="list-style-type: none"> <li>Document access restrictions for content</li> </ul>	<ul style="list-style-type: none"> <li>Maintain logs of who performed what actions on files, including deletions and preservation actions</li> </ul>	<ul style="list-style-type: none"> <li>Perform audit of logs</li> </ul>
<b>Metadata</b>	<ul style="list-style-type: none"> <li>Inventory of content and its storage location</li> <li>Ensure backup and non-collocation of inventory</li> </ul>	<ul style="list-style-type: none"> <li>Store administrative metadata</li> <li>Store transformative metadata and log events</li> </ul>	<ul style="list-style-type: none"> <li>Store standard technical and descriptive metadata</li> </ul>	<ul style="list-style-type: none"> <li>Store standard preservation metadata</li> </ul>
<b>File Formats</b>	<ul style="list-style-type: none"> <li>When you can give input into the creation of digital files encourage use of a limited set of known open file formats and <u>codecs</u></li> </ul>	<ul style="list-style-type: none"> <li>Inventory of file formats in use</li> </ul>	<ul style="list-style-type: none"> <li>Monitor file format obsolescence issues</li> </ul>	<ul style="list-style-type: none"> <li>Perform format migrations, emulation and similar activities as needed</li> </ul>



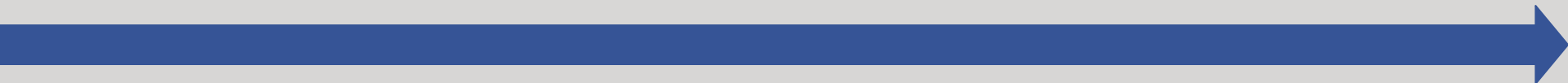
# Levels of Preservation

Next Steps for the Levels:

- Continue to solicit feedback via blog posts and talks like this
- Finalize “Release Candidate One” and publicize more broadly

Contribute your thoughts on the Signal:

<http://blogs.loc.gov/digitalpreservation/2012/11/ndsa-levels-of-digital-preservation-release-candidate-one/>





# Learn More & Get Involved

- Join the NDSA (it's free!)  
<http://www.digitalpreservation.gov/ndsa/membership.html>
- NDSA Membership Eligibility:
  - organizations or consortia;
  - demonstrated an active commitment to digital preservation;
  - share the stated values of the Alliance;
  - agree to maintain vendor neutrality;
  - agree to participate in one or more working groups.
- Read about NDSA work on *The Signal* blog  
[blogs.loc.gov/digitalpreservation/](https://blogs.loc.gov/digitalpreservation/)
- Follow on Twitter: [@NDSA2](https://twitter.com/NDSA2)
- For more information on CWG: [abgr@loc.gov](mailto:abgr@loc.gov)

