Levels of Digital Preservation:

A tool for mitigating technical digital preservation risks

The goal of this document is to provide a **basic tool for helping organizations manage and mitigate digital preservation risks.** This document 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).

- **This is useful for developing plans -- not a plan in itself:** This is not a digital preservation cookbook, what we detail here is necessary but not sufficient for ensuring digital preservation.
- **These levels are non-judgmental:** Organizations have different resources and priorities, and as a result need to think about how to best allocate those resources to meet their specific needs.
- **These levels can be applied to collection(s) or system(s):** These levels function coherently with everything from individual case by case collection level decisions as well as issues for an entire centralized repository
- **This is designed to be content and system agnostic:** This is only about generic issues. Specific kinds of content (e.g., documents, audio interviews, video, etc.) are likely to have their own nuances, but these levels and factors are generic enough that they are intended to apply to any digital preservation situation.

Each level begins to address a new area. Level 1 addresses the most likely risks in the short term. As you progress through the levels they address mitigation of risks over the long-term.

Project Background:

There is both very basic digital preservation information, like NDIIPP's personal archiving materials, and extensive and substantial requirements for being recognized as a trusted digital repository. However, the working group felt there was a lack of solid guidance on how an organization should prioritize its resource allocation between these two ends of the spectrum. The goal of this project is to develop a tiered set of recommendations for prioritizing enhancements to digital preservation systems (defined broadly to include organizational and technical infrastructure). This is defining targets for at least three levels of criteria for digital preservation systems, at the bottom level providing guidance to "get the boxes off the floor" and at each escalating level offering prioritized suggestions for how organizations can get the most out of their resources.

Project Team

- Andrea Goethals, Manager of Digital Preservation and Repository Services, Harvard University
- Abie Grotke, Web Archiving Team Lead, Library of Congress
- Amy Kirchoff, Archive Service Product Manager, ITHAKA
- Kris Klein, Digital Programs Consultant, California State Library
- Jane Mandelbaum, IT Project Manager, Library of Congress
- Trevor Owens, Digital Archivist, Library of Congress
- Meg Phillips, Electronic Records Lifecycle Coordinator, National Archives
- Shawn Rounds, State Archivist, Minnesota Historical Society
- Jefferson Bailey, Fellow, Library of Congress
- Linda Tadic, Executive Director, Audiovisual Archive Network

Levels of Digital Preservation: A tool for mitigating technical digital preservation risks

	Level One (Protect Your Data)	Level Two (Know Your data)	Level Three (Monitor Your Data)	Level Four (Repair Your Data)
Storage and Geographic Location	• Two complete copies that are not collocated • For data coming in on heterogeneous media (optical disks, hard drives, floppies) get the digital content off the medium and into your storage system	• Three complete copies • At least one copy in a different geographic location • Document your storage system(s) and storage media and what you need to use them	• At least one copy in a geographic location with a different disaster threat • Start an obsolescence monitoring process for your storage system(s) and media	• All copies in geographic locations with different disaster threats • Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems.
File Fixity and Data Integrity	• Check fixity on ingest if it has been provided with the content • Create fixity info if it wasn't provided with the content	• Check fixity on all ingests • Use write-blockers when working with original media • Virus-check high risk content	• Check fixity on all transformative acts • Check fixity of sample files/media at fixed intervals • Maintain logs of fixity info; supply audit on demand • Ability to detect corrupt data • Virus-check all content	• Check fixity of all content in response to specific events or activities • Ability to replace corrupted data
Information Security	• Identify who has read, write, move, and delete authorization to individual files • Restrict who has those authorizations to individual files		• Maintain logs of who has accessed individual files	• Maintain logs of who performed what actions on files, including deletions and preservation actions • Perform audit of logs
Metadata	• Inventory of content and its storage location • Ensure backup and non-collocation of inventory	• Store administrative metadata • Store transformative metadata and log events	• Store standard technical and descriptive metadata	• Store standard preservation metadata
File Formats	• Encourage use of limited set of known and open file formats and codecs	• Inventory of file formats in use	• Validate files against their file formats • Monitor file format obsolescence threats	• Perform format migrations, emulation and similar activities