LEVELS OF DIGITAL PRESERVATION

A TOOL FOR MITIGATING TECHNICAL DIGITAL PRESERVATION RISKS

OUR GOAL

To develop concise and easy to use rubric to help organizations manage and mitigate digital preservation risks.

PROJECT BACKGROUND

This project is chartered as a National Digital Stewardship Alliance action team for the Content, Innovation, Infrastructure, and Standards working groups.

ACTION TEAM MEMBERS

- Andrea Goethals, Manager of Digital Preservation and Repository Services, Harvard University
- Abbie 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
- Robin Ruggaber, Director, Online Library Environment, University of Virginia Library

OUR SCOPE

This project 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.

WHAT IS THIS GOOD FOR?

- This is useful for developing plans -- not a plan in itself
- These levels are non-judgmental:
- These levels can be applied to collection(s) or system(s)
- This is designed to be content and system agnostic

USEFUL FOR…

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.

USEFUL FOR…

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.

USEFUL FOR…

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

USEFUL FOR…

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.

FOUR LEVELS

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)

We have attempted to fit these technical features into four different levels. Each level adds new layers of risk mitigation by roughly related to the level’s name.

SIX AREAS

At each of the four levels we organized considerations in six areas.

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

STORAGE AND GEOGRAPHY

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)
• 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

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)
▪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

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)
• 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

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)
▪ 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

Level One (Protect your data)	Level Two (Know your data)	Level Three (Monitor your data)	Level Four (Repair your data)
• 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

	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

NEXT STEPS

We are in a public request for comments phase. After public comment we intend to revise and finalize the document and publish and disseminate it.