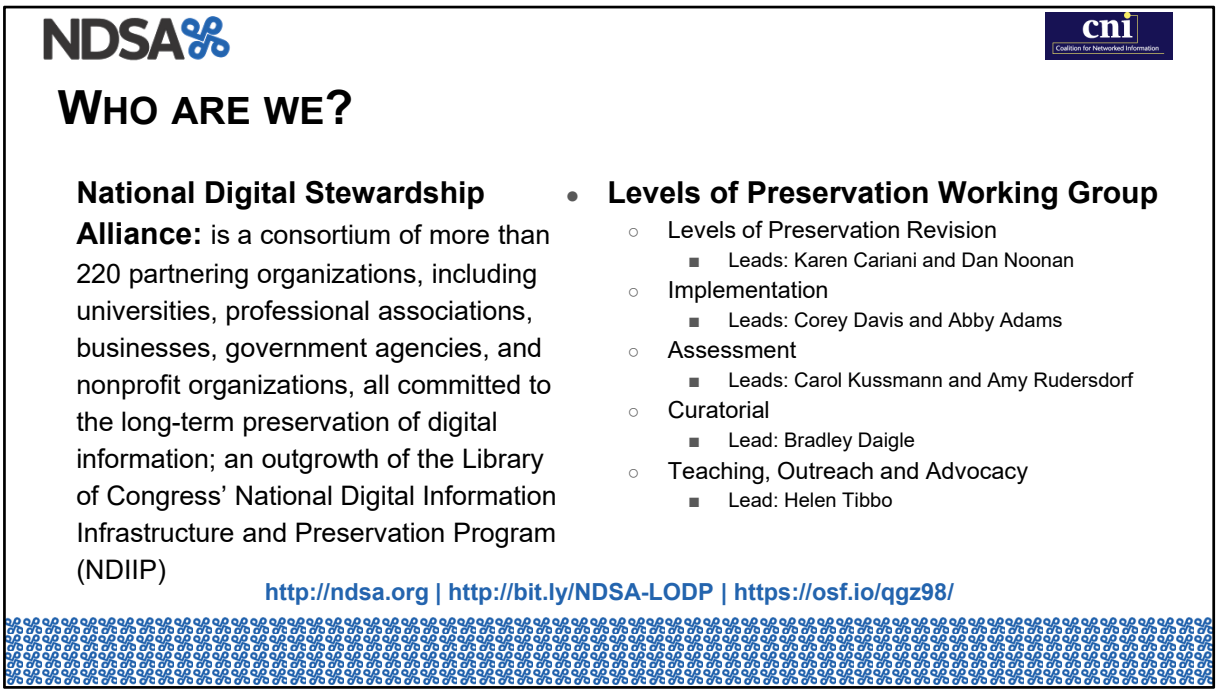
Evolution of Technological Guidelines:

NDSA's Levels of Digital Preservation Revision

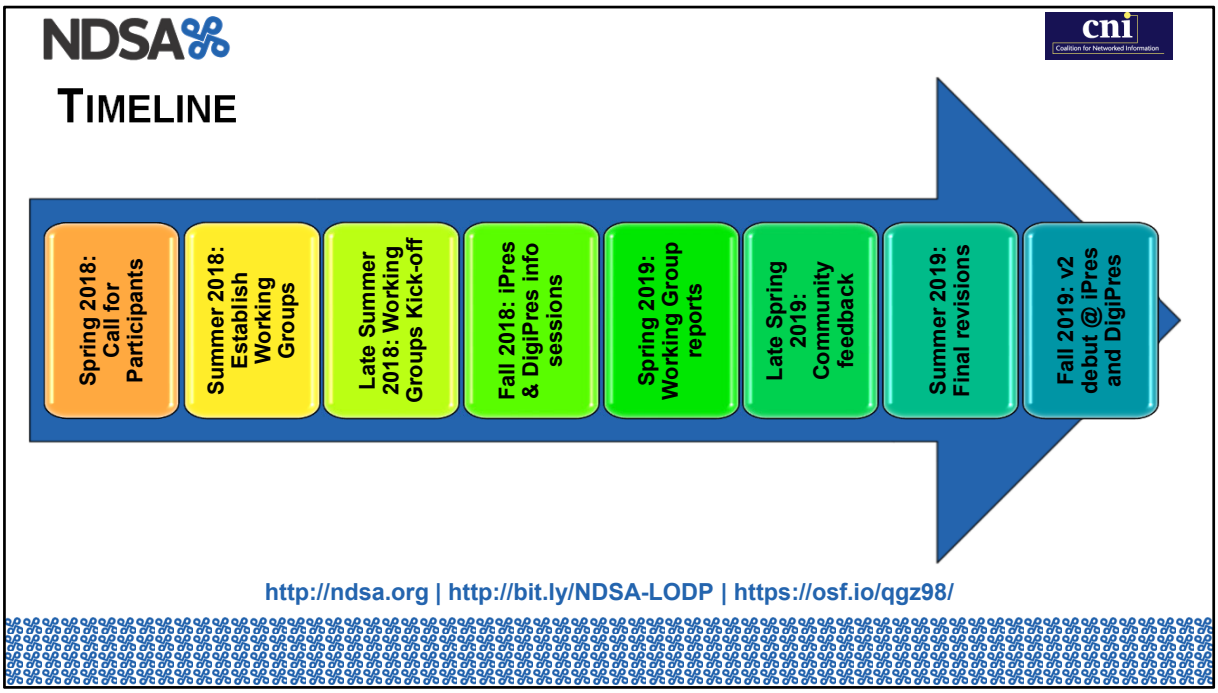
http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

Good afternoon, I am Dan Noonan, the Digital Preservation Librarian for The Ohio State University Libraries, and an incoming member of the NDSA's Coordinating Committee.



NDSA is a consortium of more than 220 partnering organizations, all committed to the long-term preservation of digital information.

These revision efforts were handled by a large cadre of volunteers via the sub-groups listed here.



From the call for participant's in early 2018 to the debut of version 2.0 at iPres and DigiPres this fall, this process took approximately a year-and-a-half.

Granted this does not include the pre-planning by NDSA's Coordinating Committee, nevertheless this is a pretty impressive feat from my experience in working on national and international workgroups.

NDSA

cni Coalition for Networked Information

GUIDING PRINCIPLES

This Photo by Unknown Author is licensed under CC BY-NC

- Cover all the key ***technological*** topics, but also be simple and practical as far as specific and immediate actions
- Agnostic approach that stresses common practice vs. local protocols
- Avoid details that require revision in the near-term
- Establish consistency
- Discuss, listen carefully, consider, decide, carry on…

http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

In 2013, with NDSA's unveiling of the Levels of Digital Preservation, Version 1—created through the work of a group that included our very own co-presenter Trevor Owens—Meg Phillips noted that the Levels are "…a tiered set of recommendations on how organizations should begin to build or enhance their digital preservation activities. A work in progress...."

I would stress the use of "a work in progress" as it is intended that this matrix is not a static document; they even called it version 1.

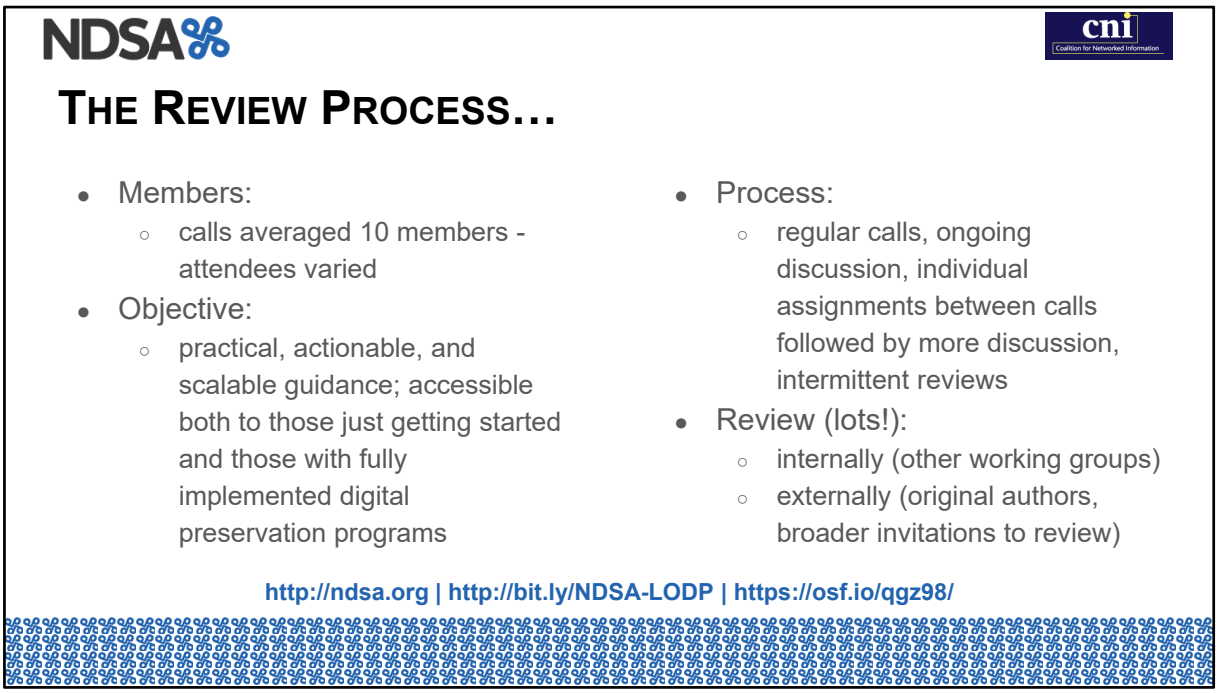
The Levels are meant to address the technological aspects of digital preservation, not digital preservation policy for example.

We stressed the need for an agnostic approach to the concepts used within the matrix that emphasizes common practices and understanding, and not particular local practices.

As an overarching strategy, we worked hard at bringing a sense of consistency across the functional areas and levels.

We entered this process with the expectation for much discussion, where folks

actively listened, were considerate of one another, and were making decisions, even if the decision was to move on and come back later.

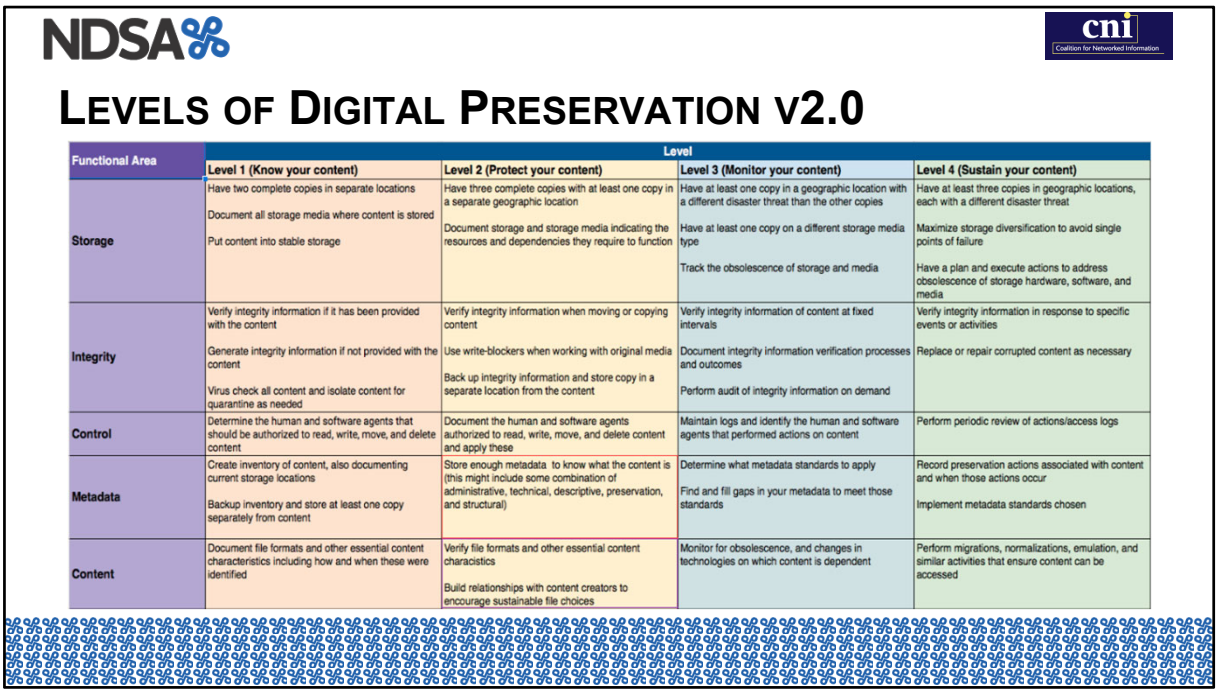


The Revision Subgroup, for which I was Karen Cariani's co-chair, was charged with the actual updating of the Levels' matrix.

The central objective was to make sure the actions identified in each cell of the matrix are practical, actionable and scalable guidance, not only for those with fully implemented digital preservation programs, but also (and maybe even more importantly) for those just getting started.

In addition to our regular virtual meetings, we spun-off some specific task groups that met between meetings to tackle various aspects of the work.

We reviewed LOTS:
- Amongst our team
- With other work groups
- Externally with broader community,
- And with the key original authors.



NDSA

cni Coalition for Networked Information

Levels of Digital Preservation v2.0

Functional Area	Level			
	Level 1 (Know your content)	Level 2 (Protect your content)	Level 3 (Monitor your content)	Level 4 (Sustain your content)
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content and isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characistics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

And here it is version 2.0 – and yes I expect y'all to be able to read this in detail – well not really and that is why I have the link provided on most of these slides.

REVISION PROCESS

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least 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 - Obsolescence monitoring process for your storage system(s) and media	- At least three 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 file 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 of content 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/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- 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	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

- Reviewed and updated column labels:
  - Protect>Know>Monitor>Repair Your Data ➔ Know>Protect>Monitor>Sustain Your Data
- Reviewed and updated row labels (Functional Area):
  - Storage and Geographic Location ➔ Storage
  - File Fixity and Data Integrity ➔ Integrity
  - Information Security ➔ Control
  - Metadata ➔ Metadata
  - File Formats ➔ Content

http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

While we were not starting with a "clean slate" we did struggle with where to start.
Do we start with the Levels and Functional Area labels? OR
Do we start with the content of the cells?

Ultimately, it was a little bit of back and forth. We did start with the labels, but then backed off as we felt we needed to address the content within the cells to see how they aligned to the labels.

This eventually lead to a re-alignment of the columns. In v1.0 the columns were Protect > Know > Monitor > Repair your data.

In v2.0 we felt it was more important to "know" your data before one is able to adequately "protect" it. Further, that from a preservation point of view we want to be able to not only repair our data, but sustain our data; therefore the progression is now Know > Protect > Monitor > Sustain Your Data.

We discussed at length the actual purpose of each of the Functional Areas, but feel we have adequately restructured them for simplicity, as well as extensibility. Therefore we made the following changes:

Storage and Geographic Location in now just Storage

- File Fixity and Data Integrity is now more fully encompassed within Integrity
- Information Security was really about Control of our content
- Metadata is well just Metadata
- And File Formats is more broadly conceptualized as Content

Revision Process

- Highlights:
  - Collaborative community effort
  - Relied on survey feedback
  - Considered adding documentation row
  - Vocabulary check
  - Metadata options - revisited multiple times
  - Alternative view

This Photo by Unknown Author is licensed under CC BY-SA

http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

As previously noted this was clearly a collaborative community process.

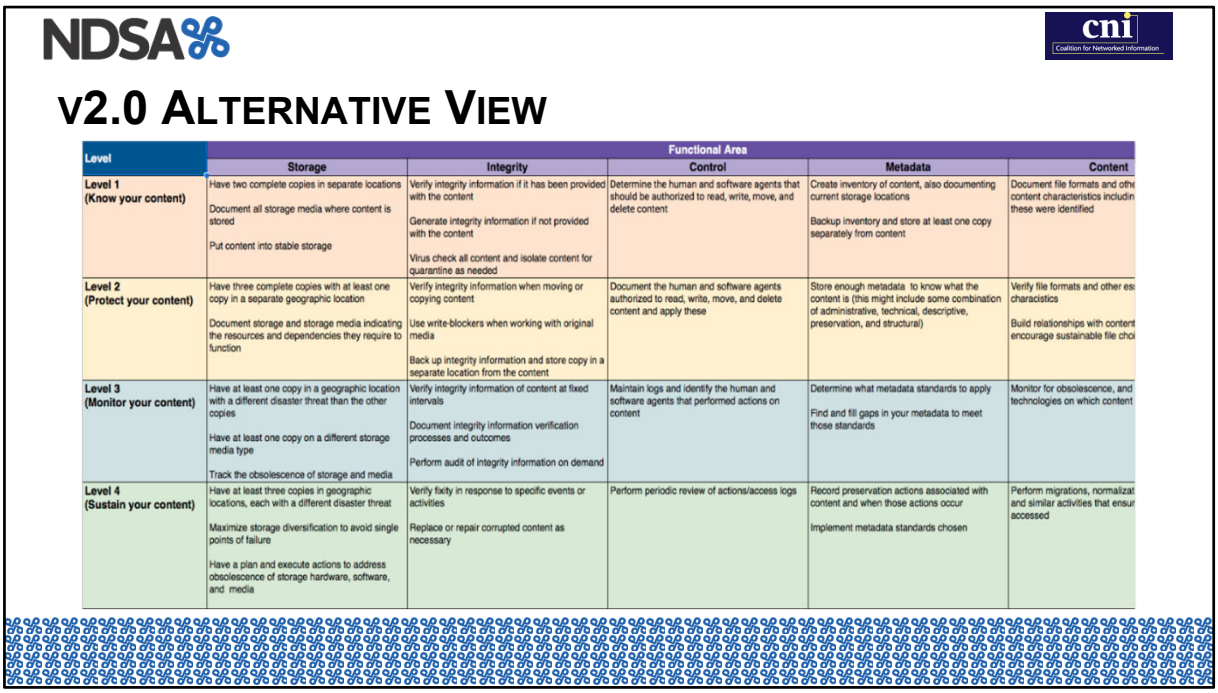
We relied upon survey feedback conducted by other workgroups, as well as our own.

We considered adding a functional area for "Documentation" but ultimately felt that it was out of scope for the matrix .

We constantly struggled with vocabulary. We'd make proposed changes, come back a week or so later, struggle some more over content, propose some more changes (or changes back); and then maybe go through the same process a few weeks later. Heck, I don't even remember how many times we revised the metadata row at this point.

It was clear that vocabulary and verbiage was important enough that we have created a set of working definitions. Even calling that "working definitions" vs "glossary" or some other term was up for debate.

Finally, we also created the matrix as a table that can be re-arranged, and have provide one alternative view.



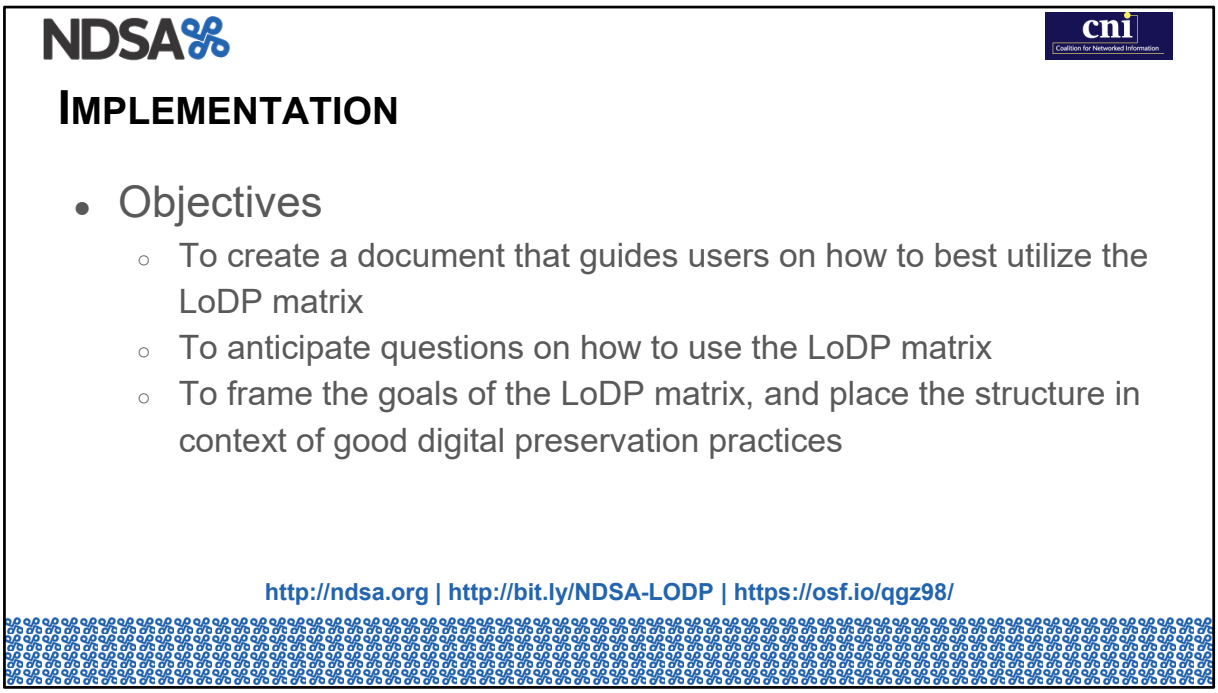
NDSA

cni Coalition for Networked Information

V2.0 ALTERNATIVE VIEW

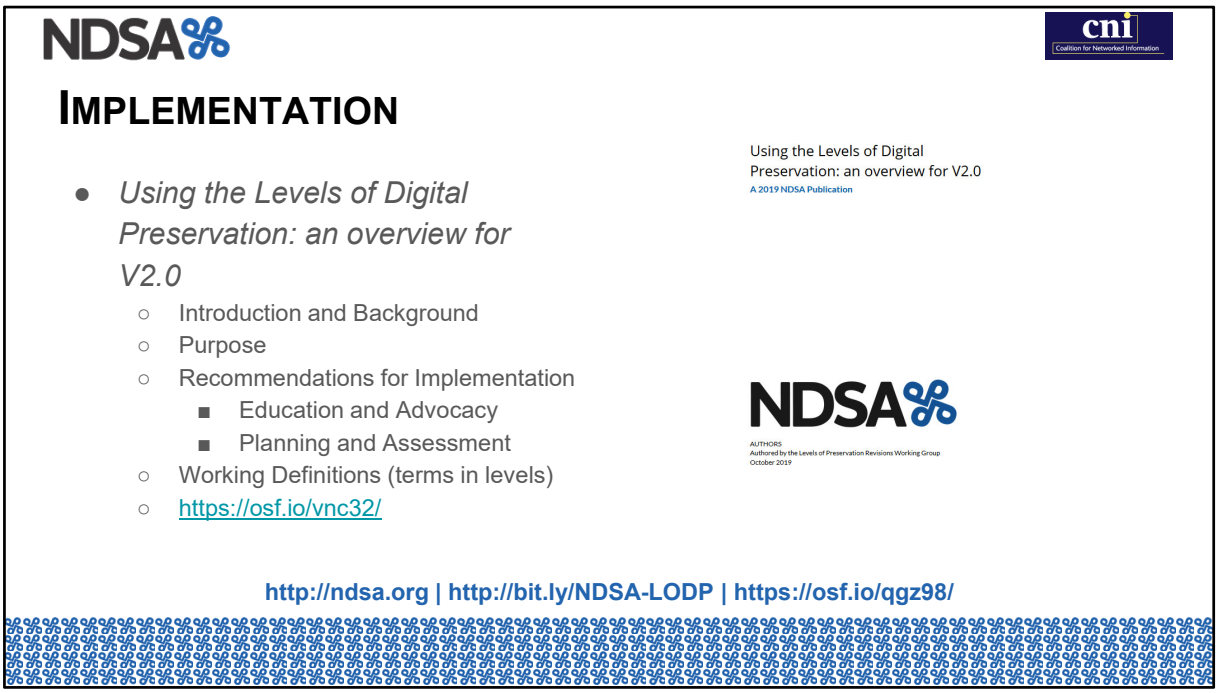
Level	Functional Area				
	Storage	Integrity	Control	Metadata	Content
Level 1 (Know your content)	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content and isolate content for quarantine as needed	Determine the human and software agents that should be authorized to read, write, move, and delete content	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Document file formats and othe content characteristics includin these were identified
Level 2 (Protect your content)	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Verify file formats and other es characistics Build relationships with content encourage sustainable file cho
Level 3 (Monitor your content)	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Maintain logs and identify the human and software agents that performed actions on content	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Monitor for obsolescence, and technologies on which content
Level 4 (Sustain your content)	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media	Verify fixity in response to specific events or activities Replace or repair corrupted content as necessary	Perform periodic review of actions/access logs	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen	Perform migrations, normalizat and similar activities that ensur accessed

This alternative view flips the columns and rows, where one can read across a row for all Functional Areas for a particular Level. In theory, this could be sorted with Level 4 at the top.



The Implementation Subgroup was created to explore the use and adaption of the Levels framework to implement or improve digital preservation infrastructure, administration, and maintenance.

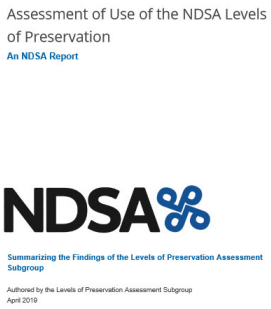
Further, the group was charged with developing guidance, based upon the results of this exploration and the revised matrix, for the use of the Levels.



To this end, eventually a collaboration between the Levels Revision and Implementation subgroups led to this document which provides an introduction and background to contextualize the Levels matrix. It goes on to provide recommendations for implementation, and supplies the working definitions for terms used within the matrix.

ASSESSMENT

- Reviewed how others were using the Levels as an assessment tool
- Found multiple ways organizations were using the levels as an assessment tool
- Documented findings in a report: https://osf.io/47kqm/

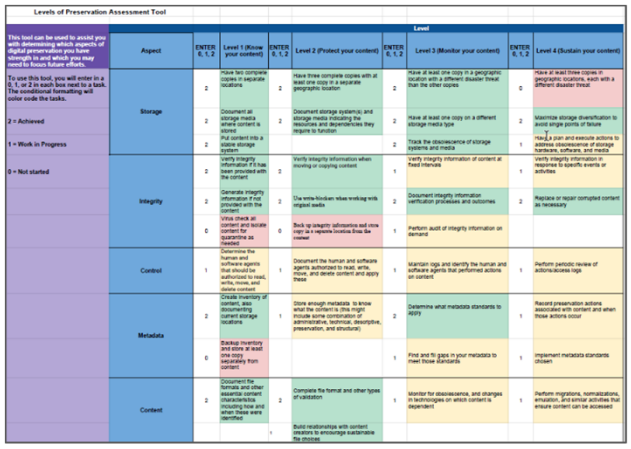


http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

The Assessment Subgroup was charged with investigating the use and adaption of the Levels as an institutional benchmarking tool for identifying gaps in practice, an assessment of digital preservation readiness, as well as to review progress.

Assessment Tool

- Created to provide users a template to use for assessing which components of the levels they have completed, are working on, or have not started working on.
- Is not the ONLY way to do an assessment, just one option.
- https://osf.io/m2fek/



http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

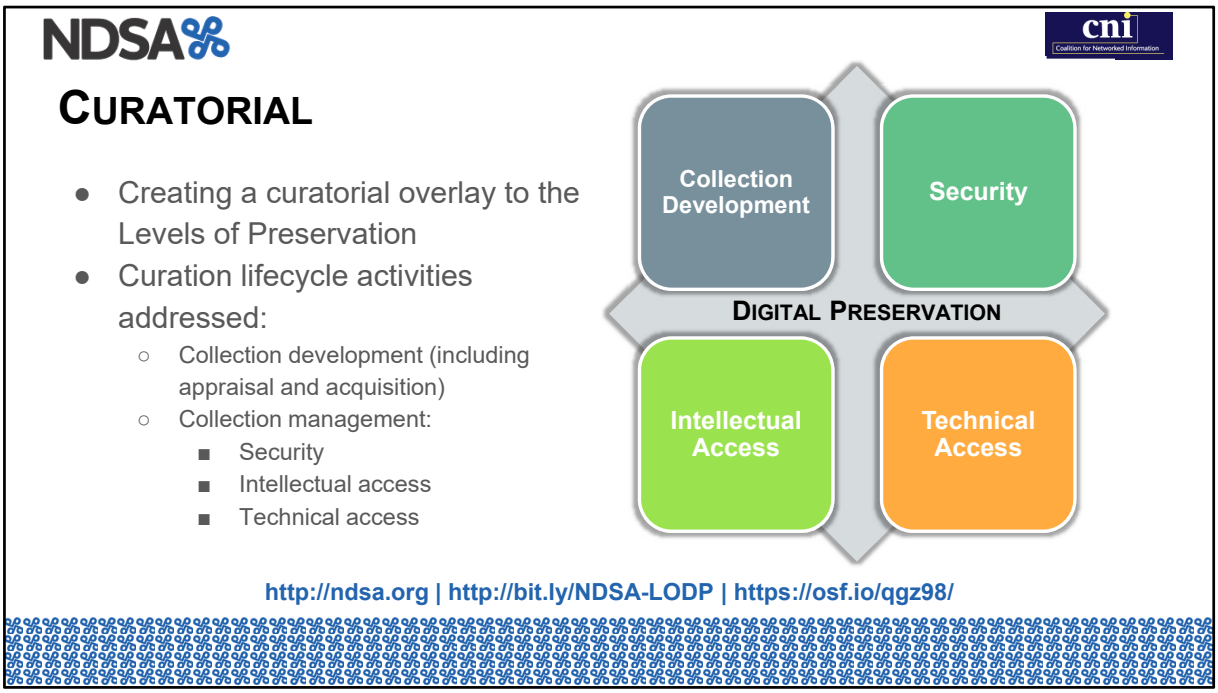
To that end, in addition to the report, the working group devised an initial spreadsheet based tool.

It is developed with conditional formatting to create a "heat map" type of result. The user enters

- 2 if the item has been Achieved
- 1 if it is a Work in Progress, or
- 0 if it has not been started

The results provide a visual representation of what is being done, and what might be something to focus next efforts on.

Keep in mind most institutions and organization will not necessary progress in lock step "up" the levels or even necessarily complete each aspect of a functional area before progressing to the next level. That is why this tool can be useful in easily visualizing those gaps.

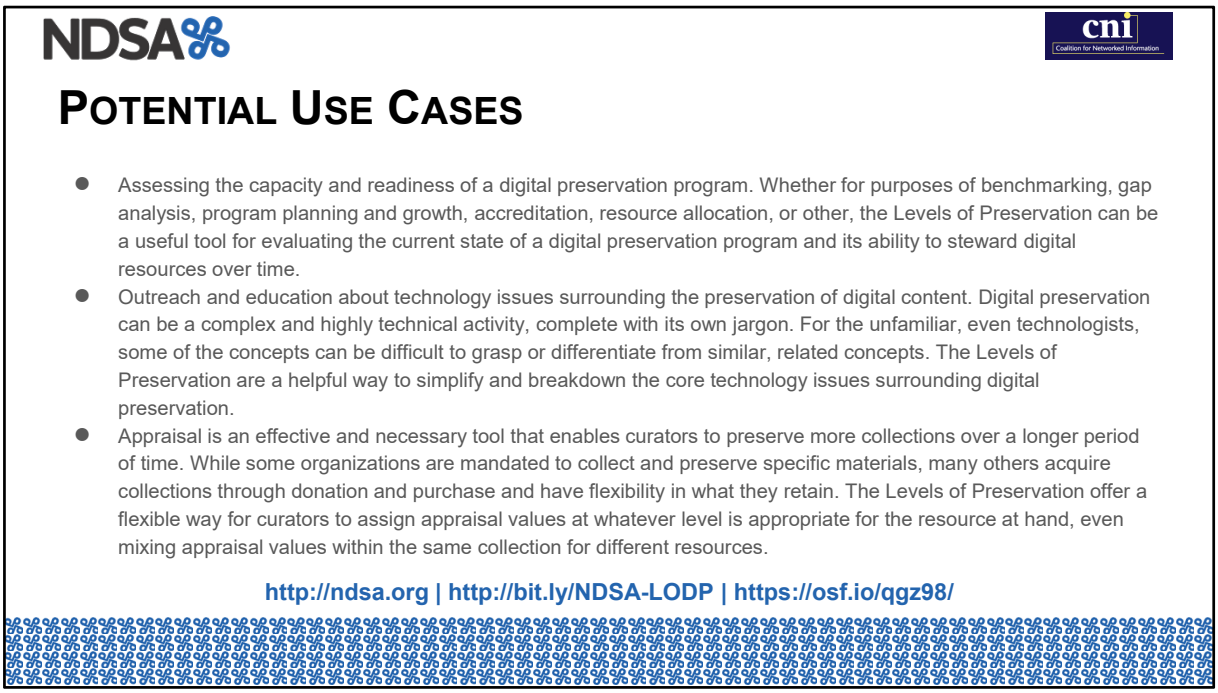


The Curatorial Subgroup was tasked with exploring a new angle on the use of the Levels, that of a non-technological decision-making overlay or layer, which will allow curators and collections stewards to make initial preservation decisions that map to the Levels.

They have had a heavy lift of defining what "curatorial" means in this situation and how to utilize it in this overlay of the Levels.

In the end the came up with four curatorial viewpoints
- Collection development (including appraisal and acquisition)
- And three for Collection Management:
  - Security
  - Intellectual access
  - Technical access

The application of the curatorial layer is not meant to be strictly linear, but is designed to be useful at any point in curation. The workgroup is in the process of developing exemplary workflows for consideration, which we do not have time today to demonstrate.



Based upon the existing work of the aforementioned work groups we have already begun to identify use cases for the Levels of Digital Preservation. From:

Assessing the capacity and readiness of a digital preservation program… TO

Outreach and education about technological issues surrounding the preservation of digital content… TO

Being a flexible way for curators to assign appraisal values at whatever level is appropriate for the resource at hand, even mixing appraisal values within the same collection for different resources…

Certainly, this list is not exhaustive.

NEXT STEPS

- Promotion
  - Conference/Meetings presentations
  - Webinars
  - Poster sessions
- Begin Feedback loop to gather suggestions future iterations of the Levels
- Teaching, Outreach, and Advocacy Subgroup

This Photo by Unknown Author is licensed under CC BY-SA

http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

Finally, it is time to begin to promote, as we are today, the revised Levels. In addition to presenting at conferences and meetings, we are developing a webinar and a strategy for poster presentations at conferences this next year.

It is also time to establish a regularized feedback loop to gather suggestion for future iterations and refinements of the Levels

To facilitate this work we are spinning up the latest subgroup, Teaching, Outreach, and Advocacy, which will also explore the use and adaption of the Levels as a teaching tool for understanding digital preservation concepts and pragmatic use, and to advocate for preservation resources.

Dan Noonan
Associate Professor
Digital Preservation Librarian
University Libraries | Digital Programs
The Ohio State University
320A 18th Avenue Library | 175 West 18th Avenue Columbus, OH 43210
614.247.2425 Office
noonan.37@osu.edu go.osu.edu/noonan

THE OHIO STATE UNIVERSITY
UNIVERSITY LIBRARIES

http://ndsa.org | http://bit.ly/NDSA-LODP | https://osf.io/qgz98/

Thank you!