EMORY LIBRARIES & INFORMATION TECHNOLOGY | OpenEmory

Networked Approaches to Preservation: Learning from Collaborative Digital Preservation Efforts

Moriah Neils Caruso, *University of Washington*
Simon F. O'Riordan, *Emory University*
Erin Wolfe, *University of Kansas*
Liz Woolcott, *Utah State University*
Jennifer Mullins, *Dartmouth College*
Drew Krewer, *University of Arizona*

Book Title: Digital Preservation in Libraries Preparing for a Sustainable Future

Publisher: American Library Association Editions

Publication Place: Chicago

Publication Date: 2018

Type of Work: Chapter | Final Publisher PDF

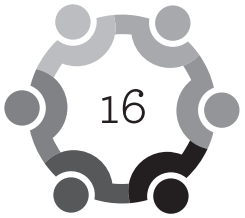
Permanent URL: https://pid.emory.edu/ark:/25593/tjbdm

Final published version:
https://www.alastore.ala.org/content/digital-preservation-libraries-preparing-sustainable-

Copyright information:

© 2019 American Library Association.

Accessed March 31, 2023 9:37 AM EDT



Networked Approaches to Preservation

Learning from Collaborative Digital Preservation Efforts

Moriah Neils Caruso, Simon O'Riordan, Erin Wolfe, Liz Woolcott, Jennifer Mullins, and Drew Krewer

In an ever-evolving technological landscape, it is difficult to know what a data recovery scenario might look like 20 or 50 years down the road. Open-source digital asset management systems are making progress toward supporting the Semantic Web; cultural heritage institutions are both producing and acquiring more digital assets than ever before; and erratic climate trends are reminding information professionals of the importance of replication and geographical distribution. How does one even begin planning for long-term preservation, access, and use in the face of all this rapid change?

Many of the foundational principles of digital preservation, such as the OAIS model and the concept of the Trusted Digital Repository, date back to the late 1990s and early to mid-2000s. When looking to address the preservation challenges presented by current day bitstreams, data, and systems, one is challenged to find authoritative, canonical guidance that holds a similar weight to these early underpinnings. Instead, we more often find that our practice is evolving in the arena of institutional documentation of efforts; we see that

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

cultural heritage communities are learning from each other and working in coordinated ways to address emerging preservation needs.

In this chapter, we will explore one such collaborative effort—-the work of the Digital Preservation Network's (DPN's) Preservation Metadata Standards Working Group (PMSWG). The PMSWG emerged from one of the first national membership organizations that focused on addressing the preservation storage needs of the cultural heritage sector, and it works on collaboratively developing metadata standards and best practices to support the preservation efforts of DPN's membership. After establishing a use case to guide its work, the PMSWG leveraged its members' institutional metadata approaches to develop a minimal "core" record designed to provide a baseline amount of information that can aid in understanding recovered content. To further address a broader range of institutional types and directly assess the needs of the DPN community, the working group also conducted focus groups, which resulted in the drafting of documents intended to address the metadata needs and concerns expressed during that inquiry process. Over the course of this chapter, we will share the current preservation trends and concerns that we uncovered during this process, as well as discuss how we developed practical recommendations intended to address the present-day directions of the digital preservation community.

About the Digital Preservation Network

The Digital Preservation Network began in January 2012 when the founding members, led by James Hilton, university librarian and dean of libraries and vice provost for academic innovation at the University of Michigan, met in Washington, DC, to discuss building a preservation network which would ensure that the most valuable at-risk content could be saved against technical failure, natural disaster, and institutional failure.[1] By the next month, February, twenty-five organizations had joined a launch team and committed funding to establish DPN. By 2013 an initial business model had been developed, a board of directors had been established, and the number of charter members had expanded to sixty institutions.[2] At this point, DPN had become visible and prominent enough to be endorsed by a number of Association of American Universities presidents. Having developed a more mature architecture, system protocol, and service model, DPN was able to make an initial launch in 2013.[3]

The DPN service is built upon the concept of depositing content through an initial node, and then replicating that content across two other nodes in the

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

DPN network. As the content is checked for fixity over time, the nodes communicate with each other and "heal" corrupted files if needed. Additionally, if institutions that serve as nodes permanently suspend services, the architecture allows for content to be redistributed to other active nodes in order to ensure that policy is followed and the content is secure.[4]

As of 2018, DPN has established itself as a leader in the field of collaborative digital preservation efforts and long-term storage services. Deposited files, along with accompanying metadata, are guaranteed to be secure for at least twenty years into the future. The security of the materials is ensured by the replication strategy previously mentioned, as well as by the fact that nodes are separated geographically, are characterized by distinct technological architectures, are established on a diverse range of hardware platforms, and are supported by a variety of organizational structures.[5] Additionally, when members join DPN, they are asked to enter into a succession agreement that will allow either DPN or another member institution to take custodial responsibilities of the materials, should the initial member be unable to fulfill its duties as a steward. DPN's approach is comprehensive, and represents how collaborative efforts can ensure the ongoing accessibility of the collective digital cultural record.

The Importance of Collaboration

Diverse Membership

The Preservation Metadata Standards Working Group came together in December 2015 as one of a number of members' working groups that had been formed to provide guidance on DPN services and best practices, as well as to establish networks of collaboration between members. The group's broad initial charge to "outline the metadata standards that DPN will follow for digital preservation"[6] provided significant leeway in determining a path forward. The group's members came from different types and sizes of institutions, with varying levels of preservation infrastructure, staffing, and support. Each member of the group also brought a variety of practical experience, with job responsibilities in digitization and digital library development, metadata and cataloging leadership, and digital preservation management. Realizing that the diversity of our organizations and expertise was an asset that could help ensure that our work would be more broadly useful, the group committed to collaborative efforts.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Although we recognized the strength of our diversity as a working group, it also made fulfilling our charge more difficult. An early deliverable for the group was to "define a set of fields that will remain constant and available with all deposits made into DPN." We quickly realized that this goal was problematic. DPN's membership is not comprised of a homogeneous block of academic institutions. There is a wide variety in institution type and size, from large research universities to smaller cultural organizations, and DPN membership promises to become even more varied over time. Furthermore, there is the larger challenge faced by digital preservation as a field: the types of content to be preserved are manifold and are continuing to grow. There is a variety of content types (video, audio, still image, document, software, data), a variety of formats (PDF, CSV, FLAC), and a variety of structures (complex multipart objects, flat objects, and the relationships between them). Furthermore, the size of the content continues to increase at a rapid rate, both in individual files and collections. Creating new metadata standards that would cement any particular set of fields with every deposit or hardwiring requirements into the architecture of DPN didn't seem useful or realistic. As a result, the PMSWG began to explore options for doing useful work with existing or developing tools in the communities where members were situated.

Making Recommendations, Not Requirements

Because the DPN did not take its first deposits from members until mid-2016, the early efforts of the PMSWG were based on group members' experience at their own institutions preparing content for deposit into DPN, as well as on a use case developed to elucidate the requirements of metadata in a DPN deposit. The PMSWG quickly realized that the diversity of our own experience and institutions did not point to an elaborate, one-size-fits-all, comprehensive metadata solution, but rather to a set of recommendations that could be adapted to the local realities at each institution. We began to work on a recommended core record that represented a minimum level of metadata needed to provide contextual information that would result in the successful retrieval of preserved digital content.

Additionally, while PMSWG members were from different-sized institutions and had different experiences to bring to bear, we also acknowledged that there were likely many additional institutional situations to consider that were unrepresented in our group. We needed to leverage the collective wisdom of the DPN membership. Network-wide collaboration, then, was the most efficient and intelligent way to meet our goal.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

The PMSWG decided to use focus groups to acquire this necessary additional context, to hear what other members were struggling with, and to learn what solutions they might already be employing in order to further refine our own thinking. In response to the needs identified in the focus groups, two additional documents were created for the benefit of the DPN membership: (1) "Metadata Considerations for Deposits," which is intended to provide an overview of the types of metadata that may be encountered and considered while preparing deposits for DPN, and (2) "Evaluating Your DPN Metadata Approach," which is designed to help guide members through a number of questions to facilitate the development of a sound approach to preserving data as well as to prepare for its deposit into DPN. These outputs and documents are described in greater detail below, and while the work of the PMSWG was intended for use by DPN's membership community, the group hopes that this work has broader utility.

Use Case

DPN is designed to be a tertiary dark storage environment. The use case for institutions accessing their content from DPN would most likely be due to catastrophic data loss from local systems. Content deposited within DPN will presumably be untouched for a substantial period of time. Because of this, any metadata added to DPN-bound content needs to be understandable and verifiable when retrieved later on. Further complexities include technological obsolescence of the systems or software to display, interpret, or run the files that are deposited in DPN. Likewise, it could be that the depositing institution (and any accompanying institutional knowledge, including metadata stored in other systems) may not ultimately be the one recovering the data, due to the succession agreement outlined above.

Considering these constraints and realities, the PMSWG used the following use case to guide our thinking; this case was originally authored by Jennifer Mullins to aid Dartmouth College's utilization of DPN:

> To fulfill the goal of having geographically distributed copies of preservation master files, the Library decides to deposit copies of materials in an off-site dark storage environment, such as the Digital Preservation Network. Once deposited, materials cannot be changed or removed. Files need to be packaged so that, when retrieved—whether in one, twenty, or fifty years' time—they can be understood, verified, and used.
>
> - The goal of understanding the files would be met if both the content and context were discernible.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

- The goal of verifying the files would be met if there is proof that the files are identical to the ones initially deposited.
- The goal of using the files would be satisfied if the file's content could be rendered (through current software or emulation), or if the file could be verifiably related to a copy in a current file format, with changes to the original well-documented, as well as documentation that the file's significant properties have been maintained in the transformation.
- Meeting these goals would rely on producing metadata to be packaged with objects before off-site deposit occurs, as well as managing metadata created locally throughout the life cycle of the object and metadata created by the storage system.[7]

Reading through the use case, one quickly realizes that the group's purview was moving beyond the administrative and technical metadata that is most often associated with digital preservation. In particular, the demands of the first clause, understandability, requires the kind of contextual information that is gleaned from descriptive metadata. The second goal, the verification of files, which is generally achieved through generating and tracking checksums, is already an integral part of the DPN technical architecture. The goal of using preserved files is often the hardest to achieve, especially given the challenges outlined above: long time lines, quickly evolving technological environments, and collections that may span not only time and technology, but also institutional change.

Recommended Core Record

The PMSWG gathered minimal metadata records from our own institutions in order to find common metadata schemas and elements in use. We looked at overlap and divergences, reviewed best practices, and sought further feedback and clarification from colleagues when needed. Through this process, we came to define the core record—the minimum level of information needed in order to uniquely and persistently understand and contextualize digital assets at a later date.

This recommended core record includes the following elements of descriptive and administrative metadata:

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Title—The name of the resource being described
Creator—The name of the person(s) or organization(s) with primary responsibility for creating the content
Date—Date information significant to an event in the life cycle of the original content, such as creation, publication, or issue date
Description—A summary description of the content, such as an abstract
Rights Statement—Information about rights held in and over the resource
Access Rights—Information about who can access the resource or an indication of its security status
Identifier—A unique identifier for a digital object (either a local identifier from your organization or a formal standard identifier issued and maintained by an external organization)
Format (original)—The format of the original item represented in the digital surrogate
Format (digital)—The format of the digital file or digital surrogate

As a *minimum* recommended core record, this does not constitute a one-size-fits-all solution that will apply to all digital objects or to all institutions. Multipart items may require additional structural metadata describing the relationship of the object to other objects. Where versioning is important, preservation metadata may be included to identify transformations or other events in the object's history. Specific formats may require technical metadata to record the essential characteristics or the software environment that is needed to render the object.

Beyond these elements, additional administrative metadata may be needed to contextualize the object. The core record as described here was designed for DPN users. As such, we have omitted certain elements that are added by DPN at the point of deposit, such as an institutional identifier, authenticity elements (i.e., checksums), and basic technical metadata (e.g., file size, mime type, etc.).

In addition to the inclusion of these elements, we recommend the use of a well-documented metadata schema that is widely adopted and both human and machine-readable. Table 16.1 shows the recommended fields, along with mappings to a few commonly used schemas.[8]

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Table 16.1 • Recommended core record

FIELD NAME	SIMPLE DC	QUALIFIED DC	MODS	VRA CORE	PB CORE	EAD
Title	title	title	<titleinfo> <title>	<title>	<pbcoretitle>	<unittitle> <title> <titleproper> <subtitle>
Creator	creator	creator	<name> <namePart> with optional attributes	<agent role=""> with <agent name="">	<pbcreator> <creator> <pbcreator> <role>	<name> <origination/ persname> <origination/ corpname> <origination/ famname>
Date	date	date dateSubmitted issued created	<originInfo> <dateIssued> Or <datecreated>	<date type="">	<pbcoreAssetDate>	<publicationstmt/ date> <unitdate>
Description	description	description	<abstract>	<description>	<pbcore Description>	<abstract> <scope content> <notestmt> <note> <physdesc>
Rights Statement	rights	rights	<accessCondition type="use and reproduction">	<rights>	<pbcoreRights Summary> <rightsSummary> <pbcoreRights Summary> <rightsLink>	<userestrict>
Access Rights	rights	accessRights	<accessCondition type=" restriction on access">	<rights> <note>	<pbcoreRights Summary> <rightsSummary> <pbcoreRights Summary> <rightsLink>	<accessrestrict>
Identifier	identifier	identifier	<identifier type="local">	<location type= "repository/> <refid type="">	<pbcoreIdentifier> <source>	<unitid> <eadid>
Format (original)	type	type	<typeOfResource>	<worktype>	<pbcoreAssetType>	<physdescstruc tured> <unittype>
Format (digital)	format	format	<physicial Description> <internetMedia Type>		<pbcoreinstantiation Digital>	

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Focus Group Observations

In December 2016, the PMSWG conducted focus groups with eleven different participants from DPN member institutions during two different telephone calls. The goal of the focus groups was to collect information about how member institutions were approaching their preservation ingest process, including information on the workflow, systems, metadata schemas, and anticipated use cases. The discussion was designed to be reflective of the long-term uses for preservation files and metadata, rather than simply surveying participants about the metadata practices they currently employed. Focus group participants were asked five questions:

1. Tell us a little bit about yourself, what types of metadata schemas you use, and what types of content you anticipate depositing in DPN.
2. What roadblocks have you encountered that have had a direct impact on your ability to deposit content into DPN?
3. Have you thought about situations that could prompt you to retrieve content from DPN? What would your institution need to successfully restore this content?
4. How are you deciding what to put into DPN? What policies or philosophies do you have around this selection process?
5. Is there anything additional you would like to contribute to the conversation at this time? Are there things you were hoping to get out of the call that were covered?

These five questions went beyond the scope of establishing a common set of metadata fields, and they had the intention of discussing the entire workflow for preserving digital files. Many factors play into the decision to keep specific kinds of metadata. For instance, the types of objects and collections will help determine the metadata fields needed to ensure long-term preservation. Likewise, the organization of files affects the structure and depth of metadata required to make sense of the content that is housed in DPN. The conversation elicited from these questions provided a snapshot of the diverse methodology and institutional needs for preservation workflows. When analyzing the data at broader levels, the similarities and differences among the institutions was illuminating.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Repositories and Tools

The participants in the focus groups used a variety of repositories, including Fedora-based repositories, Rosetta, DSpace, CONTENTdm, and locally developed repositories. They also used tools such as Archivematica and BagIt for staging or preparing their files for ingest to DPN. In addition, the participants discussed different methods of deposit into the DPN architecture. Some institutions were already participants through APTrust, and others planned to use the newly developed DuraCloud Vault tool. This wide range of repositories and tools creates a unique challenge for recommending metadata practices to DPN member institutions. The organization, structure, and built-in metadata schemas for each repository system have substantial differences that directly affect digital production workflows, as well as workflows for packaging content and depositing it into DPN.

Metadata

For preservation metadata, most participants reported using metadata standards such as PREMIS, FITS, or system-supplied metadata. Descriptive metadata varied the most, with standards such as Dublin Core or MODS featuring the most prominently. Institutions also reported using MARC or MARCXML, EAD, VRA Core, CDWA Lite, PBCore, TEI, MEI, or unspecified XML structures. System-supplied metadata or metadata exported directly from repositories was used in some cases.

There was also an inconsistent pattern of associating metadata with the preservation files. Some institutions included descriptive metadata for each file, while others bagged groups of files together with a single metadata record describing the overall contents of the bags rather than the individual files. Some institutions did not include descriptive metadata in the preservation process at all, but kept them separately in their access repository.

Content

Focus group participants reported that the file types prepared for long-term preservation were usually image, sound, video, and text-based files. In most cases, the files were high-resolution or master copies, with a few institutions also choosing to preserve derivative or access copies of files. For one institution, the majority of content was born-digital items with no physical source material. None of the focus group participants reported considering file types with emerging challenges, such as research data or software, at this time.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

Roadblocks

Most of the focus group participants reported that determining which material should be preserved posed the greatest problem for their institution. Given that the yearly deposit allocation for DPN members is 5 terabytes, some institutions felt that they had too much content to select from, while others felt they did not have enough, leading to many of the prioritization challenges discussed further below. Building on the difficulties of selection, participants reported that preparing digital collections and objects for preservation storage had a definite learning curve, since many institutions were packaging materials for this type of preservation repository for the first time. This challenge led to discussions about the difficulty of developing institutional-level workflows for preservation, including preparing and packaging files, establishing a consistent metadata structure for packaged files, and tracking which materials have been preserved in DPN. Most participants felt that they were inventing the process as they went along, and they were very interested in how other institutions were handling the workflow. Additional challenges were reported in dealing with the sheer size possible with digital collections: either processing large individual files like those produced from audiovisual collections, or dealing with collections containing a large number of files.

Content Retrieval Use Cases

When asked what types of scenarios or use cases were envisioned for retrieving files preserved in DPN, all of the respondents cited large-scale catastrophe scenarios, such as natural disasters, as a reason why they would try to retrieve data from DPN. Two institutions also mentioned organizational failure as a potential reason to retrieve content from DPN. Participants noted that in such cases, they would need to be able to track DPN's UUID and fixity information on the content. They also envisioned needing to build in regular maintenance checks on the material and were interested in being able to see a standardized report on the current status of deposited items.

Prioritization

When deciding on what content to put into DPN, participants reported that their institutions were focused on identifying material that was mandated by their universities for long-term retention and access (often theses, faculty archives, or other institutional priorities), as well as valuable, irreplaceable, or unique materials first. Some participants noted that they were also interested in

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

identifying the "low-hanging fruit," that is, collections that would not need a lot of remediation before they were ready to be deposited. As is mentioned above in the "Roadblocks" section, in many instances, participants noted that their institutions were having trouble identifying material beyond this starting scope and were putting together committees to review their collections for deposit.

Metadata Considerations for Long-Term Preservation and Access

Throughout these focus group conversations, a thread emerged pointing to challenges with preservation metadata that is used to address three areas: authenticity and integrity, contextual information, and significant properties. Focus group participants commented that these areas, especially the latter two categories, were particularly challenging because they usually require some degree of institutional discernment, which can vary quite a bit between contexts. The answers at one institution might differ a great deal from those of another institution. In addition, there are multiple approaches an institution can take, and often there is no single correct answer or right way.

Authenticity and Integrity

Two of the most important concepts in any digital preservation workflow are authenticity and data integrity. *Authenticity* is the idea that an object is really what it says it is. That is, if an object says in the metadata that it was created on a certain date and is a specific file type, it actually is. *Data integrity* is the concept that an object has not been corrupted over time while being preserved, and that it is exactly the same as when it was first deposited in a preservation environment. While this information is not novel, it still bears repeating due to the paramount importance of this type of metadata in the preservation of digital objects, regardless of the storage medium.[9]

Steps that can be taken to check and maintain authenticity and data integrity include the following:

- Use checksums to verify fixity (that an object remains unchanged since a point in time). There are multiple times and ways to use checksums in a preservation workflow.
- To ensure authenticity, tools (i.e., FITS, exiftool, or JHOVE) can be used to extract and validate technical information about an

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

object, including the file format and the software name and version. This information can be stored in a metadata standard like PREMIS in order to provide authentication for the object.

PREMIS events can be used to document the chain of custody of a digital file or collection, as well as record any derivatives or format migrations to also aid the verification of authenticity.

Contextual Information for Accessibility and Understanding

An object being preserved needs to have information about it to be useful, otherwise it would only be a collection of bits—inaccessible, undiscoverable, and fairly useless without significant efforts to re-contextualize it. Increasing the length of time between deposit and retrieval, as described in the DPN use case above, makes this provision of metadata even more difficult, since any institutional memory that could determine the nature of the object might be gone. There are also considerations of what "designated community" will need to access and understand the digital data in the future. Depending on what assumptions are made about these users, different metadata will be necessary.

Descriptive metadata that describes an object is vital to provide context. This could include information such as the creation date, content type, or creator.

Similarly, descriptive metadata can be added to a collection, providing contextual information for the objects within.

Accessibility is increasingly defined not only as having the bits available, but in the broader sense of a file being accessible to populations that may need additional information to use a file. For example, an audio file that includes a transcript in the metadata, or the close-captioning of a film, can be essential for those who have hearing disabilities to access and understand a file. The "designated community" for a file should be considered when developing metadata practices—but of course, balancing the amount of metadata with the time spent in generating it must be a factor as well.

Significant Characteristics

Significant characteristics (also known as significant properties) are defined by Andrew Wilson as "the characteristics of digital objects that must be preserved

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record."[10] Two important points about significant characteristics are that (1) not every characteristic is significant, and it is neither necessary nor practical to preserve every single bit of information regarding a collection of objects; and (2) different content types will have different significant characteristics. For example, for audio and video files, it is usually important to record the bit depth, but for a word-processing file, the significant characteristics could be the word count, the font, and the language. Again, significant properties can and will vary depending on the designated community and institutional context.

Developing Guiding Documents

In order to address the challenges that arise with these kinds of preservation metadata, and building on what we learned in developing a minimal recommended core record, two documents were created for the benefit of the DPN membership that address metadata issues for digital content deposited into DPN: "Metadata Considerations for Deposits," and "Evaluating Your DPN Metadata Approach." These documents were made freely available on DPN's website, and were published by the PMSWG on July 27, 2017.

The purpose of these two documents is to provide general metadata guidance to the members of DPN regardless of the type of institution or the material being deposited. However, the PMSWG hopes that the lessons learned and information gathered during this process are more broadly useful beyond DPN members. A short summary of each document is provided below to help illustrate their potential utility.

"Metadata Considerations for Deposits"

The first document, "Metadata Considerations for Deposits,"[11] is intended to provide an overview of the types of metadata that may be encountered or considered while preparing deposits for DPN. The first section describes the metadata elements that are required for deposit within DPN according to their BagIt specification.[12] The second section contains more information about metadata generated via the DuraCloud Vault deposit process, which is used by nearly all DPN members unless they are also a member of the APTrust. The third section includes the recommended core descriptive metadata record, with definitions and crosswalks to common metadata schemas. The fourth section

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

provides general guidance for DPN members in determining the significant properties of the content to be deposited in DPN. In particular, it provides guidance on a select number of significant properties by file type, like duration, frame rate, and color space for audiovisual files.

"Evaluating Your DPN Metadata Approach"

The second document, "Evaluating your DPN Metadata Approach,"[13] was created to help guide members through questions to facilitate the development of preservation metadata as well as to prepare for deposit into DPN. Examples of questions include "What information is needed to understand and contextualize an object?" "If data must be migrated, how will the essential characteristics of the original be known?" and "Are there formats whose essential characteristics you might see as challenging to capture/understand?" For each question, there is guidance or advice for differing scenarios in order to account for the variety of institutions and use cases. For example, one question could have different advice for descriptive metadata and for structural metadata, while another question could have advice for applying metadata to a single object versus an entire collection.

Conclusion

As acknowledged above, the field of digital preservation is based on a number of foundational ideas and documents, including the OAIS model, the concept of a Trusted Digital Repository, and the PREMIS metadata schema. However, as digital preservation efforts continue to mature and evolve in different institutions, librarians on the ground are making practical decisions in response to real-world constraints and priorities. While the rules of the road have been laid down by these fundamental and important documents, it is the many small decisions that practitioners have to make every day that are providing the direction of the future. As we come together to share challenges and discuss solutions, we learn from each other's diverse contexts and approaches. Through such collaborative work, the Digital Preservation Network's PMSWG has attempted to formulate useful outputs to help guide members when considering metadata in the context of long-term access and use. The future is uncertain and the field of digital preservation, while almost as old as digital files themselves, is still young. By sharing information, collaboratively iterating through practical solutions, and encountering the occasional but inevitable failure—in

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

short, by actually doing digital preservation work—we can make progress towards the goal of preserving the myriad of human intellectual output in digital form for future use.

NOTES

1. Kthanos, "2/23/12 Launch Team Notes," Digital Preservation Network, February 23, 2012, https://digitalpreservationnetwork.wordpress.com; Internet Archive, https://web.archive.org/web/20120414010115/https://digitalpreservationnetwork.wordpress.com/.
2. James Hilton and Mary Molinaro, "DPN Update," ARL Association Meeting 2016 (Spring), Vancouver, BC, Canada, www.arl.org/about/arl-in-transition/4010 as mm16sp_Hilton.pdf, p. 5.
3. Ibid., p. 7.
4. "The Digital Preservation Network Explained," Digital Preservation Network, www.dpn.org/dpn-admin/resources/anintroductiontodpn.pdf.
5. Ibid.
6. "DPN Members Committees, Preservation Metadata Standards Committee," Digital Preservation Network, https://dpn.org/members.
7. Jennifer Mullins, "DPN Use Case for Preservation Metadata" Message to Preservation Metadata Google Group, January 29, 2016, e-mail.
8. Chart also available via the Digital Preservation Network, https://dpn.org/dpn-admin/resources/dpnmetadataconsiderations.pdf.
9. For more information on data integrity and authenticity, see Clifford Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," in *Authenticity in a Digital Environment,* CLIR 2000, www.clir.org/pubs/reports/pub92/lynch.html.
10. Andrew Wilson, "Significant Properties of Digital Objects," National Archives of Australia, p. 2, April 7, 2008, JISC Significant Properties Workshop, British Library (FIX) slide 15.
11. Available via the Digital Preservation Network at http://dpn.org/dpn-admin/resources/dpnmetadataconsiderations.pdf.
12. Available via the Digital Preservation Network at https://docs.google.com/document/d/1JqKMFn9KfeIMAAEdOGQr6LZPqNWx8Qubi12uoUXi2QU/edit.
13. Available via the Digital Preservation Network at http://dpn.org/dpn-admin/resources/evaluatingyourdpnmetadataapproach.pdf.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.