Good Migrations: A Checklist for Migrating Your Digital Preservation Infrastructure

AUTHORS (2020)
Sibyl Schaefer
Matt Schultz
Linda Tadic
Nathan Tallman
Paige Walker

AUTHORS (2015)
Karen Cariani
Nick Krabbenhoeft
Kevin McCarthy
Trevor Owens
Leah Prescott
Abbey Potter
Sibyl Schaefer

This document was started by a working group in 2015 and resumed by a later working group in 2020. Additionally, many ideas from NDSA community members were incorporated into the final product after a call for feedback.

About the NDSA

Founded in 2010, the NDSA is an international membership organization that supplies advocacy, expertise, and support for the preservation of digital heritage. The NDSA promotes a vision in which all digital material fundamentally important to our cultures receives appropriate, effective, and sustainable stewardship care from the international preservation community to protect and enhance its persistent value, availability, and (re)use. NDSA member institutions represent all sectors, and include universities, consortia, non-profits, professional associations, commercial enterprises, and government agencies at the federal, state, and local levels.

More information about the NDSA is available at https://www.ndsa.org.

Introduction

The goal of this document is to provide a checklist for what you will want to do or think through before and after moving digital materials and metadata forward to new digital preservation systems/infrastructures. Migration may involve changes to all or part of your digital preservation technology stack (various layers of hardware, software, databases, storage arrangements, etc.), so some checklist items may not be applicable to all organizations. In some cases, organizations have adopted turn-key solutions whereby the requirements for ensuring long-term access to digital objects are taken care of by a single system or application. However, in many cases, organizations use a combination of both homegrown and off-the-shelf applications, as well as storage solutions, that together function as a preservation system. The checklist uses the technical framework of the functional areas of the Levels of Digital Preservation.

Scope

Specific strategies around normalization of content or conversion of metadata formats are out of scope for this document. This document is strictly focused on working through issues related to **moving fixed digital materials and metadata** forward to new infrastructure. This document is also not a project plan. It is useful to reference the large body of work related to project management when migrating part of your digital preservation infrastructure.

Planning the Migration

General

- ❏ Assemble a planning team, identify roles, establish and document a shared vocabulary. Develop a timeline and a roadmap for accomplishing the migration.
- ❏ Conduct an environmental scan, looking for examples of other organizations who have completed a migration similar to yours for lessons learned (e.g., with similar technology stacks or storage appliances).
- ❏ Document the reasons for migrating and the decision making process, including the person or group making the decision, the context of the decision, and the benefits to be realized for migrating. Refer to any relevant institutional documentation such as the business case, budget, preservation policy, and risk overview.
- ❏ Identify any dependent systems that will need to be updated, such as those that make API calls, and make plans to update or retire them, or otherwise notify users of changes.

- ❏ Document feature parity (e.g., support for content types and file formats, preservation actions) as well as any differences between the new and old system, and make plans to revise and refine workflows or processes.
- ❏ Note any digital objects, unneeded versions, or metadata that should not transfer to the new system (old test records, duplicates, etc.). Document your decision and criteria. For an audit trail, keep a record of content that was not transferred, and why it was not.
- ❏ Develop new documentation and/or training for users to transition from the old to the new system. Consider the following types of documentation: (1) documentation for those actually performing the migration (steps to take to migrate), (2) documentation for using the new system, and (3) documentation for any transition steps that users of the system need to take to move from old to new.
- ❏ Notify users of the date and time the system will be down and not accepting new records or objects. If the process will take some time, provide users with clear communication on what levels of service can be expected.
- ❏ Plan a location for temporary storage for content that will be migrated, whether it be new content that is awaiting ingestion or a staging server for intermediary storage of migrated content.
- ❏ Create and implement a testing plan to ensure that the functions in the new system work as expected. This can include testing your requirements for stability, responsiveness, reliability, and security, as well as benchmarking performance against your current system. Address concerns that would negatively impact a successful migration.
- ❏ Test migration workflows to make sure there are no issues – both single item and bulk batches of varying content types, file sizes, and number of files, as well as metadata. Document the tests you used to evaluate the workflow and any exceptions that might affect the migration timeline. Confirm that relationships between metadata and content are correctly migrated in the tests. This data could be used to create a project timeline.
- ❏ Document what you did and how you approached the migration to provide provenance information about the migration of the materials. Include information about any failures or issues that arose, and how they were addressed, in the documentation.
- ❏ If applicable, use transfer methods that preserve file system attributes like creation dates, owners, groups, and executability.

Storage

- ❏ Plan on what to do with your old storage media/systems. You might decide to keep them, reuse them for some other purpose, or destroy them in accordance with

existing information security and environmental sustainability policies. The decision should be deliberate, documented, and communicated.

- ❏ If the old storage media/systems stored sensitive or protected data (PII, HIPAA, FERPA, etc.), securely erase the content from the media to prevent recovery of the data.

Integrity

- ❏ Make sure you have fixity information for your objects and a plan to incorporate that information into your new system. Different systems may use different algorithms or instruments for documenting fixity information, so make sure you are comparing the same kinds of outputs.
- ❏ If the level of fixity (object, block, media) is different between the old and the new system, plan and document how fixity will be verified post-migration.
- ❏ Check/validate additional copies of your content stored in other systems; you may need to rely on some of those copies for repair if you run into migration issues.
- ❏ Try to use a transfer method that provides some level of integrity verification, such as checksumming packets during the transfer. Post-migration verification of file-level checksums may obviate this need. If using a proprietary transfer tool, document information about why the tool was used and how it works.

Control

- ❏ Review identity and access management policies and permissions, including a list of agents (people and software), and user roles that have access to the digital preservation system.
- ❏ Assess user accounts and their corresponding permissions and revise as necessary, with particular attention to administrator (e.g., sudo) access. Ensure that green-lighted accounts will continue to have access to the new system.
- ❏ If encryption is used, identify and document all the encryption keys and key escrows that need to be migrated. Plan for any changes in key management and how the ability to decrypt post-migration will be verified.
- ❏ Maintain logs of the migration so it will be easier to identify at what points any errors occurred.
- ❏ Ensure the restricted digital objects remain inaccessible to unauthorized users during and after the transfer. This may require the use of an encrypted transfer layer, such as SSH or HTTPS.

Metadata

- ❏ Review the state of metadata in the current system, and ideally, clean up any metadata inconsistencies or issues that are likely to create problems on migration.

- ❏ Plan a migration path for all types of metadata, including descriptive, administrative, technical, and preservation metadata that may not be stored or exported as discrete metadata files. Review any event metadata and/or audit logs captured by the current system and decide which to carry over to the new system.
- ❏ Document key information (database naming conventions, nuances and idiosyncrasies in system/data structures, use metrics, etc.).
- ❏ Identify the storage locations of all the metadata for your objects. If you are transferring this metadata, plan to verify the fixity of metadata records before and after the migration. This may involve generating checksums for metadata files if they don't yet exist.
- ❏ Ensure that relationships between metadata and objects are documented and will migrate to the new system. If the relationships do not automatically migrate, make a plan to re-establish them.

Content

- ❏ Review the state of files in the current system, clean up any data inconsistencies or identify issues that are likely to create problems on migration.
- ❏ If filenames must be changed prior to migration, document in the new system the original filename, new filename, and why it was changed.

After the Migration

General

- ❏ Notify your users of the change and again provide them with new or revised user documentation.
- ❏ Assemble all documentation generated and keep with other system information for future migrations. Be sure legacy documentation is clearly marked as superseded.
- ❏ Establish timeline and process for reevaluating when future migrations should be planned (if relevant).

Storage

- ❏ Record what is done with the old storage media/systems after migration and if any secure erasure was applied.
- ❏ Implement your plan to decommission infrastructure that is no longer needed.

Integrity

- ❏ Check that all files selected for migration in the old system have equivalents in the new system and vice versa.
- ❏ Verify and document fixity of your digital content and metadata to ensure that your new system has all your objects intact.

- ❏ If any objects did not migrate successfully, as identified by comparing fixity values, then repair or replace the objects with other copies. Ideally, log this kind of information as an event in preservation metadata.

Control

- ❏ Determine that all attributes, such as permissions and creation dates, have transferred or have been created as planned.
- ❏ Document identity and access management policies and permissions, including a list of agents (software and people) and user roles that have access, especially administrator access, to infrastructure.
- ❏ Ensure that identity and access management has migrated as expected and retired and temporary accounts are appropriately restricted or removed.
- ❏ Verify that encrypted content can be decrypted.

Metadata

- ❏ Check to make sure all your metadata has migrated and spot check to make sure it has not been mangled.
- ❏ Verify that any required event metadata for the migration was created as expected.

Content

- ❏ Perform quality control on a sample of digital objects to ensure the migration was successful. Consider manual verification of integrity and visual inspections. Document this process.