



National Transportation Library Digital Curation Policy

Version 1.3, January 22, 2018

Table of Contents

Background and Scope.....	1
Technology Platform.....	2
Curation Process.....	2
Description.....	2
Persistent Identifiers	2
Preservation and Backups.....	2
Format Migration.....	3
Content Alteration or Removal.....	3
Trusted Digital Repository Status	4
Review Cycle	4
Digital Curation Policy Update Log.....	4
Citations	4
Definitions.....	4

Background and Scope

Among other national library program duties, the National Transportation Library (NTL) is required by law to (1) acquire, preserve, and manage transportation information and information products and services for use by the United States Department of Transportation (USDOT), other Federal agencies, and the general public; (2) serve as a central digital repository for USDOT research results and technical publications; and, (3) provide a central clearinghouse for transportation data and information of the Federal Government (49 USC 6304). NTL’s *Repository and Open Science Access Portal (ROSA P)* (<https://rosap.ntl.bts.gov/>), is the digital library for open access transportation information and data.

To meet its legislative mandate to make collections publicly available over the long-term, NTL performs curation activities, including preservation, migration, and transformations, to ensure permanent access to its content. As stewards of digital collections, NTL staff use current, widely-accepted digital curation policies and practices where possible.



Technology Platform

Digital content in the National Transportation Library Digital Repository (NTLDR) is managed, preserved and hosted in the NTL cloud environment. Beginning in May 2017, publicly accessible digital content in the NTLDR has been uploaded into *ROSA P*, which is hosted on the Centers for Disease Control and Prevention's (CDC) Public Access Platform (CPAP) located in the Amazon Web Services cloud. The underlying technology stack for CPAP is Islandora (<https://islandora.ca/>), a combination of the Drupal content management system and the Fedora repository system.

As CPAP is a software-as-a-service (SaaS) solution, NTL staff work in partnership with CDC staff to perform and improve the curatorial functions. NTL staff also perform curation actions in-house as appropriate.

Curation Process

NTL's goal is to preserve all digital information at the bit level, at a minimum. This means that the NTL will protect digital information from bit rot and media failure, ensuring future devices will be able to faithfully reproduce the sequence of bits encoded in a digital information object. To achieve this goal, NTL employs the following practices.

Description

Digital resources in the NTLDR and *ROSA P* are fully described, in compliance with NTL's Metadata Guidelines (<https://ntl.bts.gov/policies>), fostering discovery, access, and re-use.

Persistent Identifiers

NTL is working towards assigning a digital object identifier (<https://www.doi.org/>), or DOI, as a persistent identifier to all information and data resource landing pages in *ROSA P*. This globally unique link will always lead users to a landing page containing the resource and its metadata, or a description documenting the alteration and/or destruction of the information source, if applicable. The DOI will always resolve to a landing page with metadata describing the curation of the resource, even if the resource itself is removed.

Preservation and Backups

In order to protect digital information and data from loss, NTL employs the "3-2-1" backup rule². NTL maintains

- three (3) copies of the electronic files;
- stored on two (2) different kinds of storage media;
- with at least one (1) copy stored in a different geographic and geologic region.

Currently, NTL maintains a copy of its repository content and metadata in the following locations:

- USDOT- managed Microsoft Azure cloud environment*
- CDC Public Access Platform (Amazon Web Services cloud environment)*
- Removable media (external drive)



Backups on the USDOT-managed Microsoft Azure cloud environment are in the disaster recovery site, in a different geographical area than USDOT headquarters.

Backups on the CDC Public Access Platform are in the disaster recovery (DR) site on the US West Coast, a different geographic area than CDC headquarters. The DR is updated daily. All daily backups of the staging server and weekly backups of the production servers are kept for 45 days.

The external drive housing *ROSA P* content is housed at USDOT headquarters in Washington, DC. This drive is updated quarterly.

Note on other digital versions: Because of the ease of copying electronic files, and the desire of NTL to share information and data freely, information objects housed in *ROSA P* may also be represented in other digital repositories. The NTL staff cannot guarantee the authenticity or provenance of files accessed from repositories outside our control. Whenever a question of authenticity or provenance arises, users should refer to the information item's landing page in *ROSA P* for the authentic version of an item. Further, the NTL staff will not be able to inspect electronic copies of information and data in other repositories for bit rot or corruption. However, if corrupted files are found in other repositories, NTL will work with these repositories to replace corrupted files with authentic copies of the original files.

Format Migration

Per NTL's Collection Development Policy (<http://ntl.bts.gov/policies/>), format preferences for content submitted to NTL are non-proprietary and open electronic file formats, such as .txt, .csv, .pdf, .tiff, as well as others as described by the Library of Congress in *Sustainability of Digital Formats* (<https://www.loc.gov/preservation/digital/formats/index.shtml>).

When content is migrated from one format to another, NTL will:

- Record the event in metadata
- Provide a description on the landing page
- Keep one (1) copy in the original format
- Maintain access to all versions.

Curatorial activities include migrating data from one format into another when earlier formats or devices become obsolete, and as NTL resources permit.

Content Alteration or Removal

Alteration or removal of resources, including publications and datasets, may be required if they contain data that is not publicly accessible. For example, material may be under copyright, may contain confidential information, or may compromise privacy or national security information. (See section 6 of the Selection Statement in the NTL Collection Development and Maintenance Policy for further detail: <https://ntl.bts.gov/policies/>.)

Any such changes made to a resource will be noted on the landing page or record in the database, which will remain accessible even if the data is no longer available.



Accessibility Statement

As a Federal Government resource, access to content and applications must comply with Section 508 of the Rehabilitation Act of 1973 (<https://section508.gov/>).

Trusted Digital Repository Status

To further its commitment to preservation and long-term access of items in its collection, NTL supports standards for repository trustworthiness. Assessment for trustworthiness is based on the International Standards Organization (ISO) *Reference Model for an Open Archival Information System (OAIS)* and ISO Standard 16363:2012 *Audit and Certification of Trustworthy Digital Repositories*. NTL is conducting a self-audit of the sixteen core requirements for a Trustworthy Digital Repository identified by CoreTrustSeal (<https://www.coretrustseal.org/why-certification/requirements/>). The results of this self-audit will be publicly available on the NTL website by the end of FY 2018.

Review Cycle

This policy is subject to a five-year review cycle. The policy may be reviewed, altered, and reissued as needed to meet changing needs and practices. Policy changes will be noted in the “Digital Curation Policy Update Log” section at the end of this document, and will be designated by incrementing the version number and updating the policy date.

Digital Curation Policy Update Log

- Version 1.0 finalized and published May 04, 2017
- Version 1.1 updated in October 2017
- Version 1.2, updated December 8, 2017
- Version 1.3, updated January 22, 2018

Citations

- 1) National Institute of Standards and Technology. (2015) Secure Hash Standards (SHS) (FIPS PUB-180-4). Accessed on 2016-12-16 from: <http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf>
- 2) Briney, Kristin. (2015). Data Management for Researchers: Organize, Maintain, and Share your Data for Research Success. “Chapter 8. Storage and Backups.” Exeter, England: Pelagic Publishing. <https://pelagicpublishing.com/products/data-management-for-researchers-briney>

Definitions

Definitions for certain terms used in this document are available in the Glossary for Policy Documents (<http://ntl.bts.gov/policies/>).