



U.S. Department
of Transportation

Office of the Secretary
of Transportation

The National Transportation Library

1200 New Jersey, S.E.
Washington, D.C. 20590

National Transportation Library Metadata Policy

October 24, 2024
Version 1

Policy Scope

Metadata is the information used to organize, describe, index, locate, distribute, preserve, and use a resource. This policy establishes a framework of principles for creating and managing metadata at the National Transportation Library (NTL). It pertains to all types of metadata applied to digital resources in ROSA P, the designated public repository for U.S. DOT-funded research.

Introduction

The National Transportation Library (NTL) is administered by the Bureau of Transportation Statistics, part of the United States Department of Transportation (U.S. DOT). The Library provides national and international access to transportation information; coordinates information creation and dissemination; and provides reference services for DOT employees and public stakeholders.

Established in 1998 by the [Transportation Equity Act for the 21st Century](#) (TEA-21; P.L. 105-178), NTL's authorized role was expanded in 2012's [Moving Ahead for Progress in the 21st Century](#) (MAP-21; P.L. 112-141). Since 2016, NTL has led the implementation of the Official DOT Public Access Plan issued in response to the February 22, 2013 White House Office of Science and Technology Policy (OSTP) Memorandum, "[Increasing Access to the Results of Federally Funded Scientific Research](#)". NTL incorporates updated public access policy guidance pertaining to metadata from the August 25, 2022, White House OSTP memo, "[Ensuring Free, Immediate, and Equitable Access to Federally Funded Research](#)" (the Nelson Memo) in this policy.

NTL provides open access to full-text digital publications, datasets, and other resources in the Repository and Open Science Access Portal (ROSA P). NTL collects resources from across all modes of transportation with a specific focus on U.S. DOT information products and U.S. DOT-funded research. See NTL's [Collection Development Policy](#) for information about acquisitions methods, selection criteria, formats, retention and de-selection, and access and rights.

Collection Management, Discovery, and Access

Metadata is important for trustworthy public access to research funded and produced by the U.S. DOT. All resources added to ROSA P are cataloged at the item-level. In addition to descriptive metadata, each record includes either the full-text, open access files or, when stored

elsewhere, a link to the full-text, open access files. Resources included in ROSA P are made freely available to the public under the [U.S. DOT Public Access Plan](#).

Each metadata record and the resource(s) it describes are added to one or more collections in ROSA P. Collections contain resources grouped around a similar topic, transportation mode, operating administration, or other shared attribute. NTL creates collection-level metadata that may include content, scope, spatial coverage, temporal coverage, and background information. Collections are assigned Digital Object Identifiers (DOIs) to ensure persistent access to their landing page as outlined in [NTL's Digital Curation Policy](#).

Metadata records are published in ROSA P, then indexed in databases and search engines including [Transport Research International Documentation](#) (TRID), [Science.gov](#), and [Google Scholar](#). ROSA P provides simple and advanced searching, browsing, and metadata harvesting functionality through its [REST API and OAI-PMH API](#).

Although rare, it is sometimes necessary to remove resources from ROSA P. Removal requests are only considered for items that meet the criteria described in NTL's Collection Development Policy, Section 6: Retention and De-Selection Policy. Removal requests must be made in writing, submitted to NTLDigitalSubmissions@dot.gov, and approved by NTL's Director. In cases where resources are removed, NTL maintains the metadata record in ROSA P. The resource itself is withdrawn from public view and replaced with a removal notice. Maintaining the metadata for withdrawn resources promotes transparency and is in accordance with the [FAIR Principles](#), U.S. DOT Public Access Plan, the Nelson memo, and [CoreTrustSeal Trustworthy Repository requirements](#).

Metadata Creation and Standards

Section 4 of the Nelson memo specifically mandates that federal agencies should, "Collect and make publicly available appropriate metadata associated with scholarly publications and data resulting from federally funded research, to the extent possible at the time of deposit in a public access repository." The Nelson Memo requires persistent identifiers for federally funded researchers (ORCID iDs); research organizations (ROR IDs); sources of funding and awards (Award DOIs); and all research outputs (DOIs). Additionally, the Nelson Memo requires that metadata include all author and co-author names and affiliations; the date of publication; and the status of peer review.

Cataloging of new submissions is processed in the order in which it is received. U.S. DOT-funded research submitted to NTL is prioritized for cataloging before retrospective works or digitization projects of physical resources.

Funded research must meet several requirements when submitted, [including Section 508 Accessibility](#) compliance and completion of a Technical Report Documentation Page (TRDP). The TRDP is a source of user-supplied metadata including authors, author affiliations, funding, title, publication date, abstract, keywords, and PIDs associated with the research. Librarians use the "item in hand" – the digital resource and its TRDP, cover, title page, table of contents, front matter, etc. – to create descriptive metadata. Publications lacking required documentation or failing Section 508 accessibility are returned to the submitter for remediation.

NTL adheres to national and international metadata standards. NTL's metadata scheme is based on a foundation of [Dublin Core Metadata Initiative terms](#) with additional metadata elements specific to the repository, the transportation community, and Federal Government guidelines. For example, Funding and Publication/Report Number are two metadata elements

used by NTL to capture local and Federal requirements. A minimum set of core metadata elements is required for all resources included in the repository. Datasets have additional required elements. See the Metadata Tables in Appendix A for specific guidelines about NTL's required and recommended metadata. This policy applies to descriptive, administrative, and structural metadata necessary for managing digital objects in the repository.

Metadata Quality

NTL strives to ensure that metadata records are accurate and complete. The cataloging system provides several built-in checks for data entry: input of required metadata; spelling and grammar checks on free-text fields; search and select boxes for controlled vocabularies; and drop-down menus for controlled fields such as Language, Format, and Resource Type. Metadata records are reviewed and approved by a senior cataloger for quality assurance before being published in ROSA P. We perform routine updates and database maintenance projects as we become aware of new or corrected information, and we repair metadata when errors are observed or reported.

Metadata at NTL is created primarily by humans with some machine-generated data. Examples of machine-generated metadata include accession numbers, URLs, file size, and checksums. Use of Artificial Intelligence (AI) to support metadata creation may be explored in the future. Any potential use of AI will involve human review and include a public note on the metadata record indicating the use of AI.

Appendix A: Metadata Tables

Required Metadata

Minimum descriptive information required for all metadata records. Datasets have additional requirements (see Required for Datasets table).

Element Name	Definition	NTL Usage
Title	"A name given to the resource." Dublin Core	Titles are transcribed from the resource's Technical Report Documentation Page (TRDP) using Title Case format following the Chicago Manual of Style. When a TRDP is not available, titles are transcribed from the title page, cover, or other documentation on the resource itself.
Creator	"An entity responsible for making the resource." Dublin Core	<p>Creators are individuals (persons) <i>primarily responsible for making the content</i>.</p> <p>Corporate Creators are used for organizations, including:</p> <ul style="list-style-type: none"> • The institutional affiliation of authors. • Institutions responsible for making the content, such as works for hire as defined by the U.S. Copyright Office (Circular 30). • Performing organizations listed in Box 9 of the TRDP.

		List Creators in the order that they appear on the TRDP. When a TRDP is not available, list creators in the order that they appear on the title page, cover, or first page on a resource.
Contributor	“An entity responsible for making contributions to the resource.” Dublin Core	<p>Contributors are individuals (persons) <i>involved with making contributions to the content</i>. People who are acknowledged in works for hire may be included as Contributors.</p> <p>Corporate Contributors are used for organizations that funded or sponsored the research. Sponsoring organizations are listed in Box 12 of the TRDP.</p> <p>List Contributors in the order in which they appear on the TRDP. Contributions – whether by individuals or organizations – include the range of roles defined in ANSI/NISO-Z39.104-2022: https://credit.niso.org/</p> <p>Sources of funding are displayed as digital persistent identifiers in the Funding field in ROSA P.</p>
Publication Date	“Date of issue of a work as designated by the publisher of the work.” ISO 10161-1:2014(en) 3.6.123	Use the date from the resource itself. If the item is not born-digital (such as a print resource that was digitized), use the date the item was first made available in its original format. The year is required. Included the month and day of publication if available.
Subject/TRT Terms	“A topic of the resource.” Dublin Core	<p>Use the Transportation Research Thesaurus (TRT), a controlled vocabulary for cataloging and indexing transportation-related research. Add additional keywords provided by creators in the TRDP or other controlled vocabulary terms such as Library of Congress Subject Headings (LCSH) to enhance discoverability.</p> <p>A minimum of two TRT terms is required. Best practice is to include 4-6 TRT terms. TRT Terms, keywords, and subjects display in one field in ROSA P.</p>
Publisher	“An entity responsible for making the resource available.” Dublin Core	Usually a Corporate Publisher. Enter in order of organizational hierarchy with the top-level parent organization first. Local guidelines allow for repeating a name in the Publisher and Creator field to accommodate author affiliations in the Corporate Creator field.

Format	“The file format, physical medium, or dimensions of the resource.” Dublin Core	Input the digital file format of the resource. For example: PDF, CSV, MPEG, JSON, ZIP. When ingesting multiple files to a metadata record, select the format of the primary (main) file. NTL accepts multiple digital file formats; see the Collection Development Policy , Section 3 for details.
Resource Type	Type is “The nature or genre of the resource.” Dublin Core	NTL uses a combination of local, DCMI Types , and COAR Resource Types . Resource type is determined by the content of the resource.
Collection	Resources grouped around a similar topic, transportation mode, operating administration, or other shared attribute.	All resources must be assigned to a collection. Resources may be assigned to multiple collections.
Geographical Coverage	Coverage is “The spatial or temporal topic of the resource, spatial applicability of the resource, or jurisdiction under which the resource is relevant.” Dublin Core	NTL uses this field for geographic locations that the item is about, areas where the research was conducted, or regions that may be impacted by the research. Enter “United States” if the research was conducted in a lab or lacks a specific city, state, or U.S. region. NTL maintains a local controlled vocabulary for Geographical Coverage locations.
Language	“A language of the resource.” Dublin Core	The language the resource is presented in.
ORCID	A “unique, persistent identifier (PID) for individuals”. ORCID	Implementation of a persistent identifier field for creators is pending a repository system upgrade.
DOI (Digital Object Identifier)	An Identifier is “An unambiguous reference to the resource within a given context.” Dublin Core	Unique digital persistent identifiers in the form of DOIs are assigned to research outputs in ROSA P.
Funding	The source(s) of funding for the research identified by a contract or grant number.	NTL uses this field to reference the identifier for the source of funding. Enter the contract or grant number from Box 11 of the TRDP. The funding agency is entered in the Contributor field.
Abstract	The description or summary of the resource. Also known as the Description, defined by Dublin Core as, “A summary of the resource.”	Transcribed from the resource or created from the contents of the resource when no abstract is clearly defined. Usually included in the first few pages of the document or in Box 16 of the Technical Report Documentation Page (TRDP).

Required for Datasets

The following fields are requirements for all datasets cataloged by NTL. All data in ROSA P will be accompanied by a [DCAT-US v1.1](#) .JSON metadata file, which is the federal standard for data search and discovery, in accordance with the US DOT's Public Access Plan.

Element Name	Definition	NTL Usage
Content Notes	A note added to give context or additional information valuable for using the resource.	In addition to the uses described above, the Content Notes field for datasets is currently where the data curator describes the level of curation performed on the dataset in according to the CoreTrustSeal Levels of Curation .
Abstract	The description or summary of the resource. Also known as the Description, defined by Dublin Core as, “A summary of the resource.”	Below the abstract, the NTL data curation team describes the size and contents of the final data package. This includes the size of the total package in the largest byte unit (MB for megabytes, GB for gigabytes). This space is also used to describe each file included in the final data package, how to open it, and where to find more information about that final format. The complete list of NTL defined file formats can be found in our File Format Dictionary .
References	Relation is “A related resource.” Dublin Core	The “References” metadata field for datasets always links to the corresponding report, brief, or summary created alongside the dataset. If no document is associated with the dataset, this field is left empty.
Download URL	The name of the file, including the file extension, as a uniform resource locator linked directly to the file.	While all records lead to a document located in the Primary URL, for datasets usually 3 Download URLs are created: one for the dataset README file, one for the DCAT-US JSON file, and one for the final dataset package. This package includes the README, DCAT-US file, the final dataset, supplemental code, tables, charts, codebooks, data dictionaries, data management plans, and any other supporting files or documentation.

Recommended Metadata

Based on the individual resource and the cataloger’s judgement, the following metadata fields are recommended when applicable.

Element Name	Definition	NTL Usage
Series	Is Part Of: “A related resource in which the described resource is physically or logically included.” Dublin Core	NTL uses the ISO definition of a monographic series: “A group of separate resources related to one another by the fact that each resource bears, in addition to its own title proper, a collective title applying to the group as a whole.” The Series field links to resources sharing a series title.

Content Notes	A note added to give context or additional information valuable for using the resource.	When applicable, enter as free text. Examples of Content Notes include but are not limited to licensing information, harmful language statements, multimedia descriptions, notes about the quality and completeness of files, etc.
Alternative Title	“An alternative name for the resource.” Dublin Core	Enter an alternative title when appropriate to improve discoverability. This field is most often used when the title on the TRDP and the research output’s cover page are different.
Edition	The version of the content of the resource.	When applicable, enter as free text. An edition statement may be found in the TRDP, cover, title page, or other documentation on the resource itself.
Publication/Report Number	A numeric or alphanumeric string used to identify a resource. Publication and report numbers are assigned by the Creator(s) of the content.	This field functions as an Identifier in ROSA P as defined by Dublin Core : “An unambiguous reference to the resource within a given context.” Enter the report number from Box 1 and/or Box 8 of the TRDP.
Right Statement	Rights: “Information about rights held in and over the resource.” Dublin Core	NTL uses this field for a copyright or intellectual property statement when applicable. Licensing refers to how a copyrighted resource can be used and is entered in the Content Notes field.
Related Documents	Relation is “A related resource.” Dublin Core	Related Documents link research outputs such as datasets, publications, and technical briefs that resulted from the same source of funding. Related Documents have reciprocal relationships and are identified by Uniform Resource Identifiers for each record.

System-generated Metadata

Element Name	Definition	NTL Usage
Download URL	The name of the file, including the file extension, as a uniform resource locator linked directly to the file.	The ROSA P record identification displays in the file name portion of the URL.
File Size	Extent is “The size or duration of the resource”. Dublin Core	File size displays in ROSA P with the File Type field.
ID#	The ROSA P record identification number.	
Main Document Checksum	“A value computed on data to detect error or manipulation.” NIST	