nsdl_dc

NSDL_DC Metadata Guidelines

Note: These NSDL_DC metadata guidelines are were developed in 2007. Though somewhat aged, they are valuable for the detailed and precise nature of the information delivered. Some of the examples provided may refer to resource types or materials that are not currently in NSDL's scope (such as scholarly journals). **NSDL's current preferred metadata format for new collection contribution is the LAR metadata format**. (Jan. 2013)

I. INTRODUCTION

1. Role of the *Guidelines*
2. Users of the *Guidelines*
3. How to Use and Navigate the *Guidelines*
4. Metadata documentation versioning

II. GENERAL METADATA OVERVIEW

1. What is metadata?
2. Why use metadata?
3. What is NSDL_DC metadata? (The short answer)
4. How is metadata used in NSDL?
5. NSDL_DC metadata version

III. METADATA PLANNING

1. Overarching considerations and issues
2. Background knowledge
3. Deciding what to describe
4. Determining appropriate levels of granularity (collection versus resource)

IV. CREATING NSDL_DC METADATA

1. NSDL_DC metadata fields
2. Controlled vocabularies
3. XML and NSDL_DC
4. Metadata problems
   - confusing or ineffective data
   - embedded HTML
5. Item-level metadata checklist

V. TOOLS/METHODS FOR CREATING NSDL_DC METADATA

1. CWIS
2. NCS
3. Programmatic methods

VI. SHARING METADATA WITH NSDL

1. OAI method
2. NDR API method

VII. ADDITIONAL RESOURCES

I. INTRODUCTION

1. Role of the *Guidelines*

The *NSDL_DC Metadata Guidelines* are designed to provide cogent, succinct direction for National Science Digital Library partners, collaborators, and contributors as they work with NSDL staff to create resource and collections metadata, and share it with the NSDL Repository.

This document discusses metadata from a specialized point of view: how it can best be used for content discovery in the NSDL. This is a narrow view of the use of metadata, and is not intended as a complete overview of metadata best practice. Apart from coverage of the Dublin Core metadata element set and its role as the foundation for the NSDL_DC metadata framework, these *Guidelines* do not provide in-depth coverage of particular metadata formats beyond those that directly inform the NSDL metadata profile.

Those seeking further information and background on metadata theory, creation, formats, mapping/crosswalking, and other related issues should conduct appropriate web searches on those issues. There is an Additional Resources list at the end of this document (though it may not be updated).

2. Users of the *Guidelines*: New and continuing metadata providers / collection builders

After a brief overview of metadata in general, the *Guidelines* outline the requirements and options for planning and performing the creation and sharing of NSDL_DC-compliant metadata. The document also offers guidance for continuing partners on how to enhance existing NSDL_DC metadata by taking advantage of NSDL's controlled vocabularies, best-practice counsel, and, as needed, the latest version of our in-house cataloging tool (NSDL Collection System, or NCS).

3. How to use and navigate the *Guidelines*

For those new to metadata or looking for a thorough refresher, the *Guidelines* are best consumed in a linear fashion: front to back, from the general to the NSDL-specific. For metadata veterans and established Library contributors looking for particular NSDL specifications, the *Contents* page affords quick-reference access to each section of the *Guidelines*.

II. GENERAL METADATA OVERVIEW

1. What is metadata?

The most common definition for metadata is the pat and circular "data about data," or "information about information."

But as with any tautology, "data about data", while tidy, succinct, and etymologically accurate, is not a very elucidating interpretation. A more useful way to capture the concept of metadata may be to describe what it *does*:

Metadata is structured information used to describe something (anything) whether it's a Website or electronic journal, a car, or even a pair of shoes.

The more structured and detailed the information, the more precise our searches can be.

Metadata is, in fact, a very old concept. Humans have been constructing and employing metadata, and wrestling with its attendant rules and requirements, for centuries. In brick-and-mortar libraries, for instance, metadata has long persisted as card catalogs, union catalogs, and online catalogs. Catalogers have been creating metadata in one form or another for hundreds, if not thousands of years.

In contemporary practice, metadata takes the form of tags, markers, or fields that help identify all manner of descriptive, administrative, and technical information about a given object. A metadata record consists of a standardized set of attributes, or *elements*, that provide a necessary structure for describing a resource or collection of resources.

For NSDL's purposes as a large-scale aggregator of educational material in digital format, key *elements* include title, identifier (URL), subject, description, education level, and resource type.

2. Why use metadata? (And why standardized metadata in particular?)

Metadata allows for the precise description of resources (and the sharing of such descriptions) in relatively small and discrete packages of information (metadata records), without the necessity of involving the resources themselves in the transaction. Think of an NSDL metadata record as a highly distilled stand-in, or surrogate, for the information resource it describes.

Standardized metadata is an investment in current and future interoperability, as it enhances the ability of projects or communities to work together effectively over the long term, regardless of changes in computers, networks, operating systems, and applications.

If all content providers in the NSDL community provide metadata that adheres to a common standard, users can discover NSDL content much more effectively and efficiently. In this sense, metadata functions as a kind of built-in, self-perpetuating marketing tool, increasing site traffic (at contributors' local sites as well as NSDL.org) and improving awareness of available resources. Aggregating standardized metadata in a high-profile and well-supported repository like NSDL is a particularly strong opportunity to "market" one's educational/informational wares.

3. What is NSDL_DC metadata? (The short answer)

The particular metadata schema with which this document is concerned is a variant on the Dublin Core (DC) standard. It is called NSDL_DC. Its design grew out of the idea of *normalized* metadata, or a *common denominator* approach to metadata in an environment where individual projects may use diverse metadata schema based on their own specific needs and adapt it as necessary to satisfy an aggregator's standard. For this reason, Dublin Core was chosen as the preferred format. Dublin Core comes in two flavors: "Simple" (only 15 elements) and "Qualified" (with element refinements and encoding schemes).

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is one of NSDL's primary avenues for acquiring metadata, and requires simple Dublin Core as a minimum. However, Simple DC's 15 basic elements do not provide a sufficient platform for describing NSDL's education- and education research-focused holdings in the science, technology, engineering, and mathematics disciplines. Therefore, NSDL_DC has been developed as a customized application of Qualified Dublin Core. It also includes elements from the IEEE-LOM Learning Object Metadata standard and features NSDL-specific controlled vocabularies for the following elements: Education Level, Audience, Type, and Access Rights. The most important elements in the NSDL_DC schema are: *Title, Identifier (URL/URI), Description, Subject* (and/or keywords), *Education Level*, and *Type*.

Simple Dublin Core is the low-bar metadata threshold that must be met for resources to be accessioned by NSDL. These *Guidelines* are not focused on achieving the bare minimum. We've custom-built a version of qualified Dublin Core in order to enable the most precise description and exposure of STEM-centric resources in a federated, broad-spectrum environment.

(Detailed guidelines on how to populate and express each of the element fields in an NSDL_DC metadata record appear in section IV. Creating NSDL_DC Metadata.)

4. How is metadata used in NSDL?

The chief responsibility of a digital library is to satisfy searchers' needs. To do this, we must optimize accessibility (search, discovery, retrieval). The journey to robust accessibility begins with a solid, standardized, and interoperable metadata framework. Among other benefits, this metadata framework allows for:

1. Description and classification of all manner of information objects, including learning resources, datasets, large websites, and in some cases, continuing resources like online journals
2. 'Smart' search and discovery of collections, and of discrete resources within these collections
3. Sharing of metadata with other repositories, libraries, and aggregators who may in turn provide additional services
4. Identification of similar content so that logical links and other relationships can be made
5. Seamless interaction with NSDL's other Web tools and services to optimize the administration, exposure, and publicizing of NSDL's holdings

When used in concert with other technologies that support resource discovery (including Web and text-crawling) standardized metadata forms a bedrock upon which a digital library may expand and evolve.

5. NSDL_DC metadata version

The NSDL_DC metadata framework is currently at version 1.02.020. Either NSDL_DC version 1.02 or 1.02.02 metadata is acceptable. NSDL provides version 1.02.020 through its web services, OAI, search and other services.

III. METADATA PLANNING

1. Overarching considerations and issues

- Consider the size of your project's staff and establish a budget and timeframe for creation and sharing
- Appoint a metadata manager or custodian for your project who has responsibility for ensuring the quality and currency of the project's metadata
- Review the Collection Development Strategies as a helpful guide in what makes a good collection
- Based on an initial reading of this document, create a comprehensive list of the tools, resources, and expertise that will be necessary for your particular project to comply with NSDL_DC metadata (whether you are generating metadata from scratch or adapting/crosswalking your existing local metadata)
- Plan ahead: establish at the start a detailed plan for ongoing metadata maintenance and upkeep
- Ensure that NSDL_DC metadata records exist for all resources to be shared with the NSDL Repository
- Ensure that each of your NSDL_DC records includes all of the required and strongly recommended metadata elements, and as many of the recommended and optional elements as is appropriate for your content
- Use NSDL's controlled vocabularies for the appropriate fields, following best-practice guidelines
- Pay particular attention to the Title, Subject, Identifier, Type, and Education Level elements to ensure that they contain good quality metadata that is appropriate to the resource being described
- Consider which of the optional NSDL_DC elements should be used to enhance the online visibility and accessibility of your particular resources

The NSDL Repository works with NSDL_DC metadata but collection builders are encourage to use the new new NSDL metadata format, LAR (Learning Applications Ready). It is the responsibility of a collection builder to ensure their metadata is compliant with NSDL systems and services. NSDL can consult and advise you to help make this happen.

2. Background knowledge

Good management and quality-control of metadata resources (human and technical) comes down to indispensable requirements: first, that responsibility for the task be assigned explicitly; second, that there be a plan for ongoing maintenance. Depending on the size and complexity of a collection, responsibility may fall to staff with related responsibilities.

Ideally, the project's metadata team should feature members possessing:

- Familiarity with Dublin Core (DC) metadata as a descriptive and classification mechanism for objects. That is, knowledge of DC metadata fields and their definitions and usage patterns helps in understanding the NSDL_DC metadata framework
- Familiarity with good quality control processes that ensure metadata integrity and viability
- Familiarity in expressing metadata in XML using XML schemas
- Familiarity with how the metadata records are produced (either cataloging tools or other programmatic methods)
- Technical ability to share created metadata via either an OAI provider (Open Archives Initiative) or the NSDL Repository API
- Strong project-management and liaising skills

Planning for metadata creation, sharing, and maintenance should be an integral part of a project's overall collection/information-management planning. The list above is designed to be a starting point for planning the budgetary requirements necessary. All metadata-related choices should be made with an understanding of both local and NSDL community needs.

For instance, you might very well determine that the project requires two 'views' of your resources; the first view is expressed in a local, micro-focused set of metadata records (think your own metadata format), while the second view is expressed via the aggregate-friendly, interoperable NSDL_DC format. If this course is taken, budgets must be adjusted accordingly.

It is generally better to confront the human resource needs for metadata management at the time of initial budgeting for the project, rather than at the later stages. As with any project, good planning and documentation at the outset can sometimes overcome the problem of scarce resources during the lifetime of a project. Unfortunately, leaving metadata decision making for the end of a project rarely works: by that time investments in time and effort have already been made, oftentimes in ways that cannot be fixed, undone, or re-done.

Maintaining interoperable, high-quality metadata is a significant investment for a project. However, high-quality metadata can be exposed through NSDL to further metadata distribution and it can provide the basis for effective local management of resource content.

3. Deciding what to describe

Metadata can be used to describe documents, services, collections or subcollections of data, or offline resources of many varieties.

The operative principles when determining what should be described should be based on the perceived needs of the user community. It is not necessary to create metadata for every single document or individual resource on a Website or in a database. Metadata may or may not relate to Web "pages," depending on the structure and purpose of the information being conveyed.

More metadata is not always the best solution, particularly if in the aggregate it creates large retrieval sets of records with few distinguishing details. Because the data repository does not also contain the content being described, the metadata should efficiently deliver the searcher to a point, often on a project website, from which they can easily find the specific resource they are seeking.

4. Levels of metadata granularity (collection versus resource)

When building a collection, parts of the decision process include making decisions about the levels of granularity for content in the collection (e.g. a collection of portal websites, modules, activities, datasets, individual web pages or some mixture). The same holds true for metadata. You need to decide to what level of granularity the collection will be described.

It is not always necessary to describe everything in a collection. For instance, when the items in a collection are very similar to each other, (a series of photos of Mount Rushmore at different times of the day), it is not necessary to provide metadata for each individual photo. It would be better to describe these very similar resources with just '*collection-level*' metadata that describes the objects as a group.

When the items in a collection differ in topic, description, content structure, location on a website or some other significant difference, then it is important to describe each object with '*item-level*' metadata that highlights these differences. In general, the goals of descriptive metadata for discovery and retrieval are:

* to distinguish one resource from another
* to aggregate resources that are related by topic, format, institution, audience or some criteria

A collection of highly distinguishable resources is best described at the item level. This decision (collection or item) inevitably involves tradeoffs. With collection-level descriptions, the burden shifts to the library user when the eventual target of the search is individual items. With item-level descriptions, the burden shifts to the collection holder in terms of updating and maintaining the resources (if applicable) and metadata for the collection.

These granularity issues result in two primary types of NSDL metadata records, **collection** and **item**. Both records are in the NSDL_DC metadata format but serve different purposes.

Collection records describe the entirety of a collection using subject and other descriptors that pertain to the collection as a whole. They exist conceptually at the top level of the repository, and are expected to be linked to "item" records. Each collection in the repository must have a collection record. All collections are subject to the NSDL Collection Policy. NSDL staff often use descriptive information from a collection's web site, or about pages to create collection-level metadata. Collection builders may contribute this information themselves by using the Submit a Collection form to register a collection with NSDL.

Item records describe individual or discrete resources within a collection. Since different collections can have very different levels of content granularity, a collection of activities may describe a single web page on a site, a .pdf document, or a video clip while another collection focusing on museums that have weather exhibits may catalog only the homepage of museum websites or exhibit sites. Item records may be part of an NSDL collection or be stand-alone items within the library. An item record may become part of a collection at a later time. When the item record is part of a collection, it is normally provided to NSDL by the collection builder who is responsible for the metadata and sometimes the resource content as well. Item records may be associated to more than one collection.

Links between a collection record and its item records allow some information to be inherited from the collection record and therefore gives library users more information about the context of an item. In most cases, library users who discover individual items during search can see both item and collection information.

All collections have additional administrative metadata that includes, but is not limited too, contact information and metadata creator information. This administrative metadata is used to manage the collection within NSDL. It is not visible to library users nor exposed for metadata harvesting by others.

IV. CREATING NSDL_DC METADATA

1. NSDL_DC metadata fields

The table below provides a list of all metadata fields within the NSDL_DC metadata framework. Some fields are hyperlinked to a separate web page that may include the following information:

* the definition of the field
* NSDL usage of the field
* controlled vocabulary terms and definition (if applicable)
* cataloging best practices
* resources use and examples
* XML tips and examples

Collection builders are also encouraged to review Dublin Core documents for additional information:

* Dublin Core Metadata Terms - lists all Dublin Core terms and their definitions
* Using Dublin Core - The Elements - provides guidelines and examples for generating content for simple Dublin Core metadata fields
* Using Dublin Core - Dublin Core Qualifiers - provides guidelines and examples for generating content for qualified Dublin Core metadata fields

Element	Recommended Usage	Simple definition / Notes	Sample XML tags
Title	Required	The name by which the resource or collection of resources is formally known.	<dc:title>...</dc:title>
Alternative Title	Recommended if applicable	A refinement of the Title element used to express varying form(s) of a title [e.g., *Journal of polymer science* (title); *Polymer symposia* (Alternative Title)].	<dct:alternative>...</dct:alternative>
Identifier	Required	URL to the resource	<dc:identifier>...</dc:identifier>
Subject	Strongly recommended	Populate each Subject field with only one subject term (or phrase) that describes the topics, concepts or content of the resource; repeat as needed.	<dc:subject>...</dc:subject>
Education Level	Strongly recommended	Use to describe the appropriate learning level or range associated with a resource. A refinement of the audience element. NSDL controlled vocabulary available.	<dct:educationLevel>...</dct:educationLevel>
Audience	Recommended	A broad category that best describes the recipient or user for whom the resource is primarily intended. NSDL controlled vocabulary available.	<dct:audience>...</dct:audience>
Mediator	Optional	A class of entity that mediates access to the resource and for whom the resource is intended or useful.	<dct:mediator>...</dct:mediator>
Description	Strongly recommended	A free-text account of a resource. May include abstracts or table of contents. Used as primary search field and display field.	<dc:description>...</dc:description>
Type	Strongly recommended	The nature, function or typical use of a resource. NSDL controlled vocabulary and DCMI type list available. To describe the file format, physical medium, or dimensions of the resource, use Format element.	<dc:type>...</dc:type>
Rights	Recommended	Rights information typically includes a free-text statement about various property rights associated with the resource, including intellectual property rights. May be populated with a URL that links to specific rights language in the resource.	<dc:rights>...<dc:rights>
Access Rights	Optional	Information describing conditions or requirements for viewing and/or downloading NSDL material. NSDL controlled vocabulary available; a refinement of the Rights element.	<dct:accessRights>...<dct:accessRights>
License	Optional	A legal document giving official permission to do something with the resource. A refinement of the Rights element.	<dct:license>...</dct:license>
Contributor	Recommended	Entity responsible for making contributions to the resource. Populate each Contributor field with only one contributor term; repeat as needed.	<dc:contributor>...</dc:contributor>
Creator	Recommended	Entity primarily responsible for making the resource.	<dc:creator>...</dc:creator>
Publisher	Recommended	Entity responsible for making the resource available.	<dc:publisher>...</dc:publisher>
Language	Recommended	Primary language of the resource. NSDL_DC recommends use of LOC's *ISO 639-2* controlled vocabulary.	<dc:language>...</dc:language>
Coverage	Optional	Statement of resource's spatial/geographic and/or temporal coverage. Named places (countries, cities, etc.) or time periods (epochs, date ranges, etc.) are typical Coverage values.	<dc:coverage>...</dc:coverage>
Spatial	Optional	Spatial characteristics of the intellectual content of the resource.	<dct:spatial>...</dct:spatial>
Temporal	Optional	Temporal characteristics of the intellectual content of the resource.	<dct:temporal>...</dct:temporal>
Date	Recommended	A point or period of time associated with an event in the lifecycle of the resource. Employ W3CDTF encoding scheme that looks like YYYY-MM-DD.	<dc:date>...</dc:date>
Created	Recommended	A refinement of the Date element	<dct:created>...</dct:created>
Available	Optional	A refinement of the Date element	<dct:available>...</dct:available>
dateAccepted	Optional	A refinement of the Date element	<dct:dateAccepted>...</dct:dateAccepted>
dateCopyrighted	Optional	A refinement of the Date element	<dct:dateCopyrighted>...</dct:dateCopyrighted>
dateSubmitted	Optional	A refinement of the Date element	<dct:dateSubmitted>...</dct:dateSubmitted>
Issued	Optional	A refinement of the Date element	<dct:issued>...</dct:issued>
Modified	Optional	A refinement of the Date element	<dct:modified>...</dct:modified>
Valid	Optional	A refinement of the Date element	<dct:valid>...</dct:valid>
Interactivity Type	Recommended if applicable	The type of interactions supported by a resource (active, expositive, mixed, undefined)	<ieee:interactivityType>...</ieee:interactivityType>
Interactivity Level	Recommended if applicable	The level of interaction between a resource and end user; that is the degree to which the learner can influence the behavior of the resource (very high, high, medium, low, very low)	<ieee:interactivityLevel>...</ieee:interactivityLevel>
Typical Learning Time	Optional	The typical amount of time for a particular education level to interact with the resource.	<ieee:typicalLearningTime>...</ieee:typicalLearningTime>
Format	Optional	Physical medium and/or file/MIME format	<dc:format>...</dc:format>
Extent	Optional	The size or duration of the resource.	<dct:extent>...</dct:extent>
Medium	Optional	The material or physical carrier of the resource.	<dct:medium>...</dct:medium>
Relation	Recommended if applicable	A related resource. Best practice to express relationships to related resources and the item being cataloged is to employ the applicable refinements below. Enter either the title and/or URL of the related resource.	<dc:relation>...</dc:relation>
• conformsTo		A refinement of the Relation element. Also used to provide educational standard via a URI (e.g. as ASN URIs).	<dct:conformsTo>...</dct:conformsTo> <dct:conformsTo xsi:type="dct:URI">http://purl.org/ASN/resources/S1014D95</dct:conformsTo>
• isFormatOf		A refinement of the Relation element	<dct:isFormatOf>...</dct:isFormatOf>

• hasFor mat		A refinement of the Relation element	<dct:hasFormat>...</dct:hasFormat>
• isPart Of		A refinement of the Relation element	<dct:isPartOf>...</dct:isPartOf>
• hasPa rt		A refinement of the Relation element	<dct:hasPart>...</dct:hasPart>
• isRefe renced By		A refinement of the Relation element	<dct:isReferencedBy>...</dct: isReferencedBy>
• refere nces		A refinement of the Relation element	<dct:References>...</dct:References>
• isRepl acedBy		A refinement of the Relation element	<dct:isReplacedBy>...</dct:isReplacedBy>
• replac es		A refinement of the Relation element	<dct:replaces>...</dct:replaces>
• isRequ iredBy		A refinement of the Relation element	<dct:isRequiredBy>...</dct:isRequiredBy>
• requir es		A refinement of the Relation element	<dct:requires>...</dct:requires>
• isVersi onOf		A refinement of the Relation element	<dct:isVersionOf>...</dct:isVersionOf>
• hasVe rsion		A refinement of the Relation element	<dct:hasVersion>...</dct:hasVersion>
Abstract	Optional	A summary of the content of the resource. A refinement of the Description element	<dct:abstract>...</dct:abstract>
Table of Contents	Optional	A list of subunits of the content of the resource. A refinement of the Description element	<dct:tableOfContents>...</dct: tableOfContents>
Bibliographic citation	Optional	A bibliographic reference for the resource. A refinement of the Identifier element	<dct:bibliographicCitation>...</dct: bibliographicCitation>
Instructional method	Optional	Describes process by which knowledge, attitudes, and/or skills are instilled.	<dct:instructionalMethod>...</dct: instructionalMethod>
Provenance	Optional	Statement of ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.	<dct:provenance>...</dct:provenance>
Accrual method	Optional	Method by which items are added to a collection; rarely used in NSDL.	<dct:accrualMethod>...</dct: accrualMethod>
Accrual periodicity	Optional	Frequency with which items are added to a collection; rarely used in NSDL.	<dct:accrualPeriodicity>...</dct: accrualPeriodicity>
Accrual policy	Optional	Policy governing the addition of items to a collection.	<dct:accrualPolicy>...</dct:accrualPolicy>

2. Controlled vocabularies

The NSDL_DC metadata framework uses controlled vocabularies. Some of the controlled vocabularies are Dublin Core supported and therefore an inherent part of NSDL_DC. Others are supported by other institutions, like IEEE-LOM, for the IEEE-LOM fields that are incorporated into NSDL_DC. That is, IEEE-LOM developed the terms and NSDL adopted them into NSDL_DC. And finally, there are some NSDL_DC specific vocabularies that were developed by and for NSDL constituents. The following NSDL_DC metadata fields currently have controlled vocabularies.

- educationLevel
- accessRights
- interactivityType
- interactivityLevel
- audience
- format
- subject* - The GEM subject vocabulary is still allowed in the XML of a metadata record but the vocabulary is not being promoted as a subject vocabulary for NSDL.

The use of these controlled vocabularies is not required, but collection builders are strongly encouraged to use them in addition to any locally used controlled vocabularies. The reference table above provides links to information that describes the controlled vocabularies.

The vocabularies for the above metadata fields are "enforced" in the XML of a metadata record when the XML is written a certain way. For example, if the attribute, xsi:type=" nsdl_dc:NSDLAudience", is included in the audience field of the XML metadata, then the NSDL_DC metadata framework **requires** the value for the field to be part of the audience controlled vocabulary. In the example below, **Professional/Practitioner** is part of the controlled vocabulary. Therefore, the XML below is considered "valid" XML metadata.

<dct:audience xsi:type=" nsdl_dc:NSDLAudience">**Professional/Practitioner**</dct:audience>

If the attribute, xsi:type=" nsdl_dc:NSDLAudience", is not included in the audience field of the XML metadata, then the NSDL_DC metadata framework **does not require** the value for the field to be part of the audience controlled vocabulary. It can be a local (collection builder-defined) controlled vocabulary term or free text. The XML below is also considered "valid" XML metadata. Note the differences with the XML snippet above.

<dct:audience>**Hearing impaired student**</dct:audience>

Please note that if you are using a cataloging tool that generates the XML of the metadata record for you or some other mechanism that does, you may not need to be concerned with the XML details described above. It is included for those developers who need this information. The same goes for other XML information in these *Guidelines*.

3. XML and NSDL_DC

XML is an acronym for the Extensible Markup Language and it is designed to encode documents for the sharing of information. XML is designed to be relatively human-readable and allows users to define the structure of the information that is being shared. A simple XML example is:

<title>NSDL_DC Metadata Guidelines</title>
<description>To describe the NSDL_DC metadata framework.</description>
<date>2007-12-12</date>

NSDL uses XML to provide structure to the NSDL_DC metadata framework. NSDL_DC XML for the above XML snippet would look like:

<?xml version="1.0" encoding="UTF-8" ?>
<nsdl_dc:nsdl_dc schemaVersion="1.02.020" xmlns:nsdl_dc="http://ns.nsdl.org/nsdl_dc_v1.02/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dct=" http://purl.org/dc/terms/" xmlns:ieee="http://www.ieee.org/xsd/LOMv1p0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation=" http://ns.nsdl.org/nsdl_dc_v1.02/ http://ns.nsdl.org/schemas/nsdl_dc/nsdl_dc_v1.02.xsd">
<dc:title>NSDL_DC Metadata Guidelines</dc:title>
<dc:description>To describe the NSDL_DC metadata framework.</dc:description>
<dc:date>2007-12-12</dc:date>
</nsdl_dc:nsdl_dc>

You can see how the NSDL_DC XML gives a bit more information. This extra information is controlling the enforcement of the metadata fields that are allowed (the field "title" versus a field name of "resourceTitle") and enforcing controlled vocabularies if they are used as described above.

The NSDL_DC metadata framework is currently at version 1.02.020. However, NSDL_DC version 1.02.000 metadata is still acceptable. The following link shows the XML for almost all fields.

View XML sample record for NSDL_DC 1.02.020

To determine the version of NSDL_DC metadata that you are using, look at the XML of the metadata record. If there is an attribute like schemaVersion="**1.02.020**" then you know that your metadata is NSDL_DC 1.02.020.

4. Metadata problems (pitfalls to avoid)

Data in the wrong element

This means when a metadata record has good data but it is in the wrong metadata field. This often occurs for the following metadata field pairs: *Type* and *Format*, *Relation* and *Source*, and *Title* and *Description*. The goal is to make sure your metadata makes sense. A title that is 250 words long is not appropriate. That is a description.

Useless data

Occasionally, metadata contributed to NSDL contains various defaulted strings signifying "no information available:" generally something like *unknown* (sometimes abbreviated or misspelled), or values comprised solely of stray characters such as dashes or hyphens. Some examples:

<dc:description>unknown</dc:description>
<dc:description> -- </dc:description>
<dc:description> â€¦ </dc:description>
<dc:description>No abstract available. </dc:description>
<dc:source>No source: created in machine-readable format</dc:source>

These values are of no use in the NSDL context, and should be removed before the metadata is contributed.

Confusing and ineffective data

In the *Subject*, *Creator*, *Publisher*, and *Contributor* elements, values sometimes consist of strings of terms, keywords, names, or organizations, sometimes inconsistently ordered or ambiguously separated (commas used to separate surname from forename and also to separate names). Some examples of this kind of confusing data include:

Confusing:
<dc:subject>Biology, frogs, amphibians, ecology</dc:subject>
<dc:creator>Smith, John, George Jackson, Humphrey Little and Stanley Black</dc:creator>
<dc:contributor>Sanders, G.S., T.R. Brice, V.L. DeSantis, and C.C. Ryder.</dc:contributor>

In order to eliminate confusion, NSDL recommends that collection builders separate names and subjects into separate elements and be consistent in how it is done.

Good:
<dc:subject>Biology</subject>
<dc:subject>frogs</subject>
<dc:subject>Amphibians</subject>

<dc:creator>John Smith</dc:creator>
<dc:creator>George Jackson</dc:creator>

Embedded HTML

Some data contributors populate their *Description* element with information that is also used in a website (or is cut-and-pasted from one) and may include embedded HTML used to specify a particular presentation. Much of this tagging is not only invalid in XML, but may also behave in unexpected ways within the NSDL user interface. In general, do not use embedded HTML in metadata provided to the NSDL. If we encounter it, the metadata record may or may not be accessioned and there is no guarantee that the display of information through the NSDL user interface will be as the collection builder intends.

In most cases, HTML tagging will display in NSDL.org search results as plain text, with tags clearly evident. Since this is often not what is intended, please strip out embedded HTML before providing metadata to the NSDL.

5. Item-level metadata checklist

As a checklist to see if your item-level metadata records in a collection are ready to be contributed to the NSDL, consider the following:

- Metadata meets the cataloging best practices
- For metadata fields with controlled vocabularies (e.g. type), the vocabulary is used correctly
- Metadata does not use placeholder vocabulary values (e.g. 'To be supplied')
- Metadata content meets the intended use of the field
- The URL being cataloged has been checked to verify it is operational
- Acronyms and abbreviations are spelled out, particularly in descriptions and titles
- The general description of the object being cataloged uses complete and well-written sentences that are understandable to a library user
- The metadata content avoids redundant information. For example, resource creator and technical information should not be repeated in the description field
- Correct spelling is used throughout the metadata record
- Metadata descriptions are objective and free of evaluative information or testimonials about the resource
- Records are in XML (eXtensible Markup Language) in the NSDL_DC metadata format or at a minimum, the simple Dublin Core format
- Records have the appropriate required metadata like *Identifier*
- Records have appropriate recommended metadata like *Title*, *Description*, *Subject*, *Education Level*
- Records have characters that are properly XML-encoded in UTF-8 (Unicode Transformation Format)
- Records are valid XML
- Records are provided through OAI or some other programmatic method that communicates directly with the NDR API

_____

V. TOOLS/METHODS FOR CREATING NSDL_DC Metadata

The following tools/methods are only suggestions for creating NSDL_DC metadata. Collection builders are encouraged to use whatever method suits their needs. NSDL can provide consultation information as to what might be the best method.

1. Collection Workflow Integration System (CWIS)

The Collection Workflow Integration System (CWIS) (pronounced see-wis) is a software package to to assemble, organize, and share collections of data about resources and has functionality to customize your own site around its portal capabilities. CWIS has built-in OAI provider capabilities. CWIS is produced by the Internet Scout Project. The requirement for CWIS are:

- A web server that supports PHP
- A web server that supports MySQL
- Apache
- A LAMP (Linux / Apache / MySQL / PHP) platform (but CWIS is being run in the field on OS X, Solaris, and FreeBSD)

The CWIS website has more detailed installation and download instructions. CWIS is ideal for those collections that want portal capabilities in addition to collection management capabilities. Collection builders install the software package locally. NSDL is not available for technical help.

2. NSDL Collection System (NCS)

The NSDL Collection System (NCS) is a web application for creating, editing, managing and developing collections. The NCS supports several built-in metadata formats and can even support user-defined metadata formats if supplied an appropriate XML schema. The NCS has the built-in capability of being able to communicate directly with the NDR-API.

The NCS is a result of a collaboration between NSDL and Digital Learning Sciences at the University Corporation for Atmospheric Research (UCAR).

The NCS may be used locally or collection builders may request to create a collection using a hosted NCS at UCAR.

Local install

The requirements for a local install of NCS are:

- A web server container - Tomcat 5.5
- Java 1.5
- A JVM (works well with the Sun JVM that uses a JDK of 1.5.0_09)

The NCS is available on Sourceforge which includes installation instructions. Collection builders install the web application package locally. NSDL is not available for technical help.

Hosted collection

To create a collection in the NCS hosted by NSDL, please contact NSDL and include the following information in the request:

- Your name
- Institution, project or group name
- Name and purpose of the collection
- A URL (if available) to a sample item in the collection or to the collection homepage

3. Programmatic crosswalk

This method refers to a specific NSDL scenario. If a collection builder provides simple Dublin Core metadata to the NSDL, then the NSDL can automatically crosswalk this metadata to NSDL_DC. This method is for those collection builders who produce simple Dublin Core and are not able to produce NSDL_DC.

VI. SHARING METADATA WITH NSDL

Once you have metadata and reviewed it against the metadata checklist described earlier, you are ready to share this metadata with the NSDL. NSDL can harvest or hold "native" metadata (e.g., whatever metadata format is used by a project or collection as the basis for their management of their resources) so long as it is XML-valid and uses one of the sharing methods below. However, such metadata is not available to library search or other NSDL systems and services. Only the collection builder or their designees have access to such native metadata in the NSDL repository. Therefore, in order to have a collection searchable in NSDL, one must share their metadata, either in simple Dublin Core but preferably NSDL_DC. There are several possible methods for sending metadata to the NSDL.

1. OAI method

The Open Archives Initiative (OAI) specifies a Protocol for Metadata Harvesting (PMH). Currently, NSDL supports OAI-PMH 2.0 and your OAI repository needs to be compliant with the OAI-PMH 2.0 specification. More specifically, your repository must pass validation in terms of OAI validation and XML schema validation:

OAI validation

OAI validation means your OAI repository correctly implements the OAI-PMH. It responds correctly to all OAI-PMH requests, including requests with arguments and various error conditions. OAI-PMH requires that every OAI response be XML schema valid, but XML schema validity is not sufficient to check OAI validation. There are requirements in OAI-PMH that cannot be expressed via XML schemas.

We recommend the following approach to OAI validation:

1. Go through the OAI Server Checklist.
2. Check each verb with as many conditions as you can think of using the OAI Repository Explorer [http://re.cs.uct.ac.za/]
   This is an interactive suite of HTML forms -- you will need to fill in the blanks for every OAI-PMH request you test. Remember to use different combinations of arguments to check each verb. There is an option for XML schema validation, which is set by default.
3. Validate your server using the validator supplied by the Open Archives Initiative. The OAI provided validator is sort of hidden at the bottom of the Registering as a Data Provider page [http://www.openarchives.org/data/registerasprovider.html]. You can validate your OAI server without registering it by clicking in the checkbox that says "only validate and do not register (you may then register later)."

XML schema validation

OAI-PMH requires that all metadata served must be valid according to an XML schema. In other words, every metadata record must be schema valid, and you must indicate the URL of the XML schema, both in ListMetadataFormats responses and in the served records. Note that there is also an XML schema provided for all OAI-PMH responses -- so the OAI wrapper around the served metadata must also be schema valid.Schema validation tools are available to help you with this.

When your OAI server passes all of the above validations, it is ready for NSDL harvesting.

There are many OAI software solutions and packages available. The Metadata Frquently Asked Questions (FAQ) can help with XML questions and OAI questions.

2. Tools that communicate directly with NSDL

Metadata or other information/content can be shared directly with NSDL repository when using the NSDL Collection System (NCS). This tool already includes the programmatic means to communicate with the NSDL repository. For more information about the NCS see the tools/methods for creating NSDL_DC metadata above.

VII. ADDITIONAL RESOURCES

Metadata Frequently Asked Questions (FAQ)

Best Practices for OAI Provider Implementations and Shareable Metadata
(A joint initiative between the Digital Library Federation and the National Science Digital Library)

Metadata Made Simpler, by Gail Hodge