

NSDL Metadata Formats

All metadata records in the NSDL are xml documents that are xml valid and xml well-formed and use one of the formats described below. Xml principles are described and each metadata format is explained and is linked to a how-to page for more details on creating xml records in that particular format.

Formats covered here:

Resource metadata formats

nsdl_dc

lar

Usage metadata formats

comm_anno

comm_para

XML Valid and XML Well-formedness

The NSDL can only take in metadata records that are well-formed, in the correct metadata format, and that are valid with respects to the format's schema. In order for an xml record to be well-formed, it must follow the guidelines of the W3 Consortium, such as correct capitalization, inclusion of both start- and end-tags, and proper nesting. A good tutorial for learning the basics of well-formed xml can be found at w3schools.com. Additionally, official documentation about the xml standard can be found at w3.org.

Xml records are valid when they comply with the metadata schema for the format they are written in. For example, all valid lar records must comply with the lar schema. The schema stipulates which fields must be present in lar metadata records, whether they can be used multiple times, whether they have a controlled vocabulary, etc. There are many xml editors which will check for valid and well-formed records.

Many times, it takes some back-and-forth communications between a collection builder and NSDL support staff to get the metadata records correct. This is best done early in the process, before the collection builder invests a lot of time and effort creating many (possibly incorrect) records. Luckily, since xml is text-based, it's a fairly easy process to email the information back and forth to get the metadata just right.

Resource Metadata Formats

While the NSDL is a library of educational resources for STEM education, we do not actually store any of the resources themselves. Instead, we store the information about the resources, such as where it resides (the URL), who is the creator, what type of resource it is (i.e., a flash animation or a podcast), etc. Collection builders submit this information to us using resource metadata records. Each resource metadata record is centered around the URL of the resource, since that URL is used as the central identification with which to organize the entire library. There are two main formats used for resource metadata in the NSDL:

nsdl_dc

nsdl_dc is the canonical metadata format in the NSDL. It was developed at the beginning of the NSDL over a decade ago. Currently, about 3/4 of the library's xml records are nsdl_dc format. However, each month, more and more of the library's resources are described using nsdl_dc's new-and-improved successor format, lar. When new collection builders are submitting resource metadata to the NSDL, we require that the format used be lar due to its superiority and ease of use (both for the collection builders as well as the metadata consumer).

lar

The lar metadata format is used for describing resources. It was conceived to implement the concept of learning application readiness for library resources. It is the newest of our metadata formats and was designed specifically to provide users with highly-curated, classroom-ready resources. In relation to its predecessor, nsdl_dc, the lar format uses more controlled vocabularies (which were carefully developed with the input of NSDL partners and community members). It has some new fields which are useful for classroom use, such as reading level, media run time, or degree or standard alignment.

Significantly, though, it is much simpler and easier to implement than nsdl_dc because many confusing and/or superfluous fields have been condensed down. All of these new features allow the NSDL to build highly-attuned user services which search the lar format's more organized, standardized metadata.

Usage Information Metadata Formats

If you have a website with your collection of educational resources, in addition to sharing those resources with the NSDL, you might have user activities about the resources that you would also like to share with the NSDL. The formats for this are comm_anno and comm_para. The NSDL displays this information alongside the resource metadata. We link the information by using the resource URL as the common key. It is very important that both URLs are exactly the same (this can be tricky if one gets updated and the other doesn't), because otherwise the link will be broken and the additional metadata will not be displayed at nsdl.org.

comm_anno

Comm_anno is a metadata format used primarily for collection builders to provide the NSDL with user reviews or comments from their own websites. If a collection builder has their own portal of resources where users have made comments about a resource which the collection builder also shares with the NSDL, they can use the comm_anno format to send along the user comments to appear at nsdl.org as well. It can also be used to align educational standards to a resource. Since both nsdl_dc and lar also provide the ability for standards alignment, the comm_anno format would only be used if the collection builder would like to add on to another organization's resource metadata (since otherwise, you'd be able to simply add it to the resource metadata yourself and wouldn't need an extra metadata record in comm_anno).

comm_para

The comm_para format is designed to capture paradata, or usage data, about a resource. Like the use of comm_anno to submit user comments, comm_para is used when the collection builder has a separate website where users can partake in actions centered around a resource which is also shared with the NSDL. The collection builder can send along information about whether users have voted, viewed, downloaded, favorited. For example, you might share that your educational resource has an average user rating of 4 out of 5, and that 23 people have voted. Since the number of user comments or reviews can also be captured as well as the text of those comments or reviews, there is some overlap with comm_anno, so collection builders should carefully determine what is the most efficient way of sharing their usage data with the NSDL.