Proposal to CI and the Pathways Regarding Metadata
Prepared by Susan Jesuroga
10-3-2006

Overview

Key priorities for the Pathways include expanding their collections (including resources discovered through NDR), and better understanding and meeting their user needs. This proposal is not about metadata for metadata's sake, but focuses on the types of information that appear necessary for Pathways (and other NSDL users) to do their work of building collections and providing user services.

Areas for Discussion and Decisions

To organize the issues, I looked at several sources to provide direction, including a review of the search and browse features on each site (summarized in the Appendix), the recent nsdl.org user survey by Mick Khoo, and a handful of recent digital library articles. The following recommendations are meant to be a starting point for discussions at the October 18 meeting.

Although we all know the cost of creating metadata, and the limitations of relying on metadata (such as mismatches between user and cataloger vocabularies), there is key information that I propose this group find consensus on. This information will help the Pathways and CI build a solid body of information on which more comprehensive user services can be built.

1) Agreement that the Pathways supply audience metadata and the gradeLevel qualifier. This issue started at the August meeting and it seems reasonable to finalize the gradeLevel vocabulary under discussion. Almost all sites use gradeLevel in search services, including returning it in results, or using it to limit and refine searches. Not all web sites use the audience field, but I also recommend requiring audience, because this information will be critical for those searching the NDR to expand their collections and to help users narrow their search results. Once basic vocabularies are agreed upon, CI can publish and advertise these vocabularies as required information (either explicitly supplied or crosswalked) for incoming metadata from other sources.

Joint work required:
- Agree on gradeLevel vocabulary (at Annual Meeting).
- Assign the audience vocabulary problem to a working group of interested parties.
- Work with SDSC to backfill gradeLevel metadata missing in existing records.
- Pathways make this metadata available to the NDR.

2) Agree on other minimum metadata to be provided by the Pathways to help them make gains towards their goals. And I mean minimum. Mick's search survey suggests that users would like to refine searches by resource type and grade level. Studies by Mimi

Recker and others also suggest that these are important fields. Pathways search results suggest that in addition to gradeLevel and audience, the following fields would be widely used and should be required:

Title (which is linked to the URL)
Description
Resource Type
Subject

Title is used by all sites. Description, resource type, and subject are used by most sites, in both presenting results and in supporting advanced searches.

Joint work required:
- Resource type will need a working group to determine an appropriate vocabulary. Since current vocabularies are quite different, the initial work needs to clarify the intent and scope of the resource type field.
- Subject will also need vocabulary agreements. Common vocabularies include GEM or LCSH for subjects, or adaptations of each, so it seems this issue might be more straightforward. CI should also decide if iVia might also help with subject classification.
- There might be alternatives for the description field, including indexing resource text and providing snippets (particularly useful for discovery). Most sites are not indexing full text, and this area could be ripe for prototyping and development (perhaps with assistance from CI or through provision of central services).
- Pathways make this metadata available to the NDR.

3) K-12 teachers are very interested in finding resources aligned to standards (underscored in several articles including those by Recker and DLESE). For collections providing education standards information, NSDL should accept only standards identifiers based on the ASN. Agreeing to a standard ultimately makes it easier to discover and share this information, and ASN's persistant identifier helps in managing resource assignments over a long period of time.

Joint work required:
- This decision has implications about the persistance and maintenance of the ASN database. CI needs to clarify NSDL's long-term access to ASN.
- Pathways adopt ASN identifiers and make them available to the NDR.

4) CI needs a strategy for managing disparate vocabularies. Crosswalks can help normalize various approaches, but a strategy to link vocabularies to collections and items via the NDR would be helpful to track and manage changes over time.

Joint work required:
- CI and Pathways need to determine the key information to be managed about vocabularies and agree on how that information would be related to collections and/or

items in the NDR, along with the types of services required to access and use that information.

5) As Pathways build their collections, finding resources in the NDR that can be reused is key. This goes beyond simple user actions related to viewing resources, to users services that allow for the actual download, embedding, or alteration of that resource. NSDL needs to adopt a common licensing framework like Creative Commons, which is my recommendation.

Joint work required:
- For any collection in the NDR lacking such information, CI should take the responsibility for adding, at minimum, simple contact information for each collection. My own opinion is that the number of collections is still small, and this information should be straightforward to add (perhaps using student help).
- For sites in the NDR for which this information cannot be ascertained, I recommend these collections should be deaccessioned.
- Going forward, we agree to only add sites with some sort of explicit copyright statement, or at minimum, a contact email. (This could be formalized in the Collection Development Policy.)
- CI can look into methods of inheriting this information from collection records for individual resources.

I have not detailed all the specific work that comes along with these recommendations including development of crosswalks or other metadata work that might be required by the Pathways. There is, no doubt, other information that could help Pathways in building collections and providing user services, but the information described above is critical to existing services and the work seems completely doable with reasonable amount of resources when considering the overall benefits to all involved.

APPENDIX
Quick review of Pathway search and browse features.

AMSER
http://amser.org/SPT--Home.php

Results show title, link, description, link to more info, and "your" rating.
- more info includes title, URL, publisher, description, LoC classification, GEM subject, resource type, audience, language, source, cumulative rating.

Advanced fielded search on description, title, subject, creator, with limits on source, resourve type, language, audience, format, creator role.

Browse by GEM or LCC classification.

Search on metadata only. Considering full-text indexing in future.

BEN
http://www.biosciednet.org/portal/

Results show title, link, description, author, publisher, collection, last updated, keywords, resource type (**match to DC field**), context (grade level), file format, size.

Advanced search on title, keywords, discipline, pedagogical use (assess, learn, research, etc), resource type (animation, assessment, audio, book, video, etc), ed level, author, authors institution.

Browse by resource type or subject.

Search on metadata only.

ComPADRE
http://www.compadre.org/portal/

Results show title, link, description.

Advanced search on general subject (controlled vocabulary), specific subject (controlled vocabulary), cost, resource type (such as collection, curriculum, learner activity, etc), target level (gradeLevel), target role.

Browse collections or refine collection list by grade level and role (such as K-12, faculty, research, etc).

Search on metadata only. Planning for full-text indexing by fall 2007.

CSERD
http://www.shodor.org/refdesk/Catalog/

Results show title, link, description.

Advanced search on description, title, subject, creator, language, format.

Browse by fields subject, keyword, audience with limits on: education level, resource type.

Search on metadata with some full-text indexing (depends on access to resource).

Journal of Chem Ed DLib
http://jchemed.chem.wisc.edu/JCEDLib/index.html

Browse by collections and topical indexes.

MatDL
http://www.matdl.org/

Results show collection name, date of issue, title, link, authors. Search can be refined by institution.

Browse by institution and collection.

Search on metadata only. Considering full-text indexing in future.

Math Gateway
http://mathgateway.maa.org/do/Home

Results show title and link, collection and link, author, resource type, subject and first few words of description
- refine by collection.

Search on metadata only.

MSP
http://msteacher.org/

Search returns title and link, description, grade level, audience, resource type, link to collection, and link to more info.

Browse by subject in math and science.

Search on metadata only. Considering full-text indexing in future.

TeachEngineering
http://teachengineering.org/index.php

Results show title and link sorted by subject area, curricular units, lessons, activities.

Advanced search by education standards
- refined by source, subject, topic, standards number, grade level, and correlation mapping scheme and correlation strength

Advanced search in fields including title, summary, engineering connection, keywords, grade level, and time required. Modifiers for activities include group size and cost.

Browse by subject area, curricular units, lesson, and activities.

Searches full-text indexing.

NEEDS
http://www.needs.org/

Search returns title and link, author, other contributors, publisher, download size, keywords, search score, possible uses, collection link, and icons for platform, cost, and reviews.

Advanced search includes by keyword, resource type, discipline, grade, title, author, series (controlled vocabulary), collection, publication, and peer-review.

Teachers' Domain
http://www.teachersdomain.org/tdhome.html

Search returns thumbnail and link to full resource info page, title and link to full resource info page, description, grade level, media type and view option, link to standards.
- refine search by media type, grade level, and subject.

Browse by grade level and topic.

Full-text index search including PDFs and video captions.