

Restricting the public distribution of NSDL Metadata via OAI-PMH

Dean B. Krafft
July 26, 2007

The Issue

Many of the potential NSDL publisher partners are concerned about having the metadata that they provide us redistributed via the OAI-PMH protocol. By default, all metadata in the central NSDL Data Repository (NDR) is available for harvest via OAI. The community presumption is that metadata harvested using this protocol can be reused and redistributed freely. The publishers wish to clearly restrict the uses of the metadata that they provide NSDL and to prohibit further bulk redistribution.

The Solution

To address this need, the Cornell CI team is implementing a new feature: a *Not available via public OAI* metadata flag. Roughly speaking, this flag could be set for any collection of metadata records that we get from a single source – typically a publisher. When this flag is set, the metadata so flagged, except for title and URL, would not be available for harvest via our publicly available OAI-PMH service. We are reserving the right to make URL and title available through the OAI service, although in the initial implementation they may, or may not, actually be served.

We are currently projecting to have this feature fully implemented and available as part of the NDR by September 15, 2007.

Caveats and Warnings

This flag **only** restricts the bulk redistribution of the metadata through the OAI-PMH service. There are currently a number of other ways that an external user or service could get access to all the metadata from a provider. These include crawling the web-based resource pages for all the resources in the flagged collection; performing a series of queries to the Search REST service to access the items in the collection and their metadata; and directly querying the NDR API for the metadata. In general, using these mechanisms to bulk harvest information from NSDL without permission will probably violate our published terms of service, but there are no technical measures in place to keep this from happening. It would be extremely difficult to try to put such measures in place, and, given the open design of NSDL and the NDR, it would be irresponsible to represent that we could successfully restrict this kind of access.

Impact on Pathways

This new mechanism will be available for any collections of metadata that you or your partners provide for the NDR. So long as everyone understands the cautions and limitations, there may be cases where this would be useful. If you have questions about the details of the implementation of this capability, or if you have suggestions for variations that might be particularly useful for your project, please don't hesitate to contact us.