



Self-Hosted Partner Repository Information

Getting Started

While the South Carolina Digital Library provides three regional hubs for participation in the project, we also welcome partnerships with stand alone digital repositories. The following information will help you down the path of setting up your institution's repository for harvesting into SCDL. Before any repository can be harvested into SCDL, the provider OAI feed will be tested and analyzed for metadata compliance. When your institution is ready for it's first test harvesting, or if at any point you have questions about the process and the preparations necessary, please contact your regional coordinator.

If your institution does not have it's own repository, you can still participate in the South Carolina Digital Library by contacting your regional coordinator. You can find your regional coordinator here: <https://scmemory.org/about/scdl-staff/>

Harvesting your repository into SCDL

We need to know two things to get started:

1. The base URL of the repository's OAI provider, including the port, if specified. Here is roughly what a typical OAI base looks like:

http://your.repository.edu/oai/oai.php?verb=ListRecords&metadataPrefix=oai_qdc

2. The metadata format of the records. SCDL prefers Qualified Dublin Core and the "oai_qdc" prefix. A complete list of our prefix mappings is available here: <https://bit.ly/42TckFP>.

Please let us know which mapping you will use and map your metadata accordingly.

Once we know those things, SCDL can begin reviewing your metadata feed for interoperability and compliance with our existing standards and guidelines. Once your feed has passed the review process, it will be added to the SCDL harvest schedule.

How we use OAI to harvest your records

- SCDL is built using Blacklight and Apache Solr. We use a Blacklight OAI Harvester plugin to harvest our partners' records and then index that data to Apache Solr
- Our OAI Harvester plugin then sends a "ListRecords" request with "from" and "until" parameters to obtain the first batch of metadata records from the repository.



Self-Hosted Partner Repository Information

- Additional “ListRecords” requests with appropriate “resumptionToken” parameters are sent as needed to get the full listing of records.
- We run a number of normalization routines on the harvested records to transform the raw Qualified Dublin Core metadata into SCDL-compliant, normalized Qualified Dublin Core metadata.
- Typically, we do a full initial harvest only once and then do incremental harvests twice/year, prior to DPLA’s scheduled ingest, using the “from” and “until” parameters on the monthly “ListRecords” request. We keep an open schedule of our OAI harvests available here: <https://goo.gl/kTkvWB>

Providing Metadata Using OAI (Open Archives Initiative) Protocol

SCDL can harvest any OAI-compliant stream. The requirements for the OAI protocol for metadata harvesting are available here: <http://openarchives.org>.

SCDL’s metadata schema and guidelines use Qualified Dublin Core and are available here: <http://scmemory.org/for-contributors/guidelines-resources/>. Any participating repository must be compliant with the minimum required in these guidelines.

Some things to consider:

- We want to harvest your metadata at the object level only. This means that we do not want to harvest any page level metadata from your compound objects. For example, if using CONTENTdm be sure that the “Enable OAI” setting is set to “Yes” but that the “Enable compound object pages” setting is set to “No.”
- We recommend that you implement deleted record status. This is not required of OAI repositories but, without it, we have no way to remove from our harvester the records that you delete locally, except by a full delete-and-reload of your entire repository, which we prefer not to do (very often).
- Our default metadata format for harvest is Qualified Dublin Core but we can accept simple Dublin Core if need be, although it sacrifices a lot of metadata complexity.

SCDL and the Digital Public Library of America

All SCDL partners have their content harvested into and fully discoverable within DPLA. Using a Blacklight OAI Provider plugin, we supply our complete SCDL Qualified Dublin Core OAI feed (compiled from all participating repositories) out to the Digital Public Library of America for harvesting into their system. DPLA currently harvests our feed twice/year.