

OAI Provider Configuration Reference

The following configuration values can be set by modifying the WEB-INF/classes/proai.properties file and restarting the webapp.

Contents

- [Contents](#)
- [Front-End OAI-PMH Behavior](#)
- [Back-End Update Behavior](#)
- [File Storage Locations](#)
- [Primary Database Configuration](#)
- [Advanced Database Configuration](#)
- [Fedora Driver: General Settings](#)
- [Fedora Driver: Metadata Format Configuration](#)
- [Fedora Driver: Set Membership Configuration](#)
- [Fedora Driver: MPTStore Configuration](#)
- [Fedora Driver: Advanced Configuration](#)

Front-End OAI-PMH Behavior

- **proai.secondsBetweenRequests** For multi-part OAI responses, this specifies the maximum time to wait between requests before cleaning up the session.
- **proai.incompleteRecordListSize** The maximum number of records to include in each part of a ListRecords response.
- **proai.incompleteIdentifierListSize** The maximum number of record headers to include in each part of a ListIdentifiers response.
- **proai.incompleteSetListSize** The maximum number of sets to include in each part of a ListSets response.

Back-End Update Behavior

- **proai.driverPollSeconds** How often Proai should poll the driver for updates.
- **proai.driverPollingEnabled** Set this to false to start the service without polling enabled. Note: You can also change polling behavior while the service is running by executing the following SQL against the database: `UPDATE rcAdmin SET pollingEnabled = 0` (or 1 to enable it). The value configured via this property will always be put into effect when the service is started.
- **proai.maxWorkers** The maximum number of worker threads to use while retrieving new/updated OAI records from the driver.
- **proai.maxWorkBatchSize** How many requests each worker thread should make before returning the batch to the database to be committed.
- **proai.maxFailedRetries** If a request of the driver for a specific record fails for any reason, this is the maximum number of times that record should be automatically retried before giving up on that record. If this is exceeded, the record information will remain in the rcFailure table until it is dealt with manually.
- **proai.maxCommitQueueSize** The maximum size the in-memory record commit queue is allowed to reach. If this threshold is reached, it means that worker threads are returning results faster than they can be committed to the database by the commit thread. If this occurs, workers will have to wait until more items have been committed before they can add any more to the queue.
- **proai.maxRecordsPerTransaction** The maximum number of record updates (from the in-memory commit queue) that should be committed to the database at one time.
- **proai.validateUpdates** Whether to run schema validation during the update process. This defaults to true.

File Storage Locations

- **proai.cacheBaseDir** The directory where cache files should be stored. This will be created if it doesn't exist.
IMPORTANT: You should change this from /tmp to a permanent directory of your choosing. The default value is just used for testing.
- **proai.sessionBaseDir** Where OAI-PMH session data will be stored when needed. This directory will be created at startup if it doesn't already exist. Each session consists of a series of small data files in a uniquely-named subdirectory beneath this one. They are automatically created and removed as needed.
- **proai.schemaDir** Where to store schema files if validation is enabled.

Primary Database Configuration

- **proai.db.url** The JDBC connection URL for the database that will be used by the cache.
Embedded McKoi example:
`jdbc:mckoi:local://build/test/mckoi/mckoi.conf?create_or_boot=true`
Local McKoi example:
`jdbc:mckoi://localhost/`
Local MySQL example:
`jdbc:mysql://localhost/proai?useUnicode=true&characterEncoding=UTF-8&autoReconnect=true`
Local Oracle example:
`jdbc:oracle:thin:@localhost:1521:proai`
Local Postgres example:
`jdbc:postgresql://localhost/proai`
- **proai.db.driverClassName** The class name of the JDBC driver appropriate for use with the connection url. The appropriate library must be in the classpath.
McKoi example:
`proai.db.driverClassName = com.mckoi.JDBCdriver`
MySQL example:

proai.db.driverClassName = com.mysql.jdbc.Driver

Oracle example:

proai.db.driverClassName = oracle.jdbc.driver.OracleDriver

Postgres example:

proai.db.driverClassName = org.postgresql.Driver

- **proai.db.mySQLResultTrickling** If using MySQL, this indicates whether to stream query results one-at-a-time from the database to conserve memory. If unspecified, this defaults to false.

Background: For large repositories, MySQL's default behavior of reading all results into memory can cause OutOfMemory errors in Proai. If this value is set to true, memory requirements for Proai will be lower, but query performance will suffer. For more detail, please see the [Connector-J Implementation Notes](#).

- **proai.db.username** The database user. This user should already exist in the database and must have permission to create, modify, and query tables.
- **proai.db.password** The password for the database user.

Advanced Database Configuration

- **proai.db.maxActive** The maximum number of connections from the pool that may be active (leased) at once. If unspecified, this defaults to 8.
- **proai.db.maxWait** The maximum number of milliseconds to wait for a connection from the pool if none are available. If unspecified, this defaults to -1, which means "indefinite".
- **proai.db.*** Any other properties defined by the Apache commons-DBCP project can be specified. See the full list (and documentation) at <http://jakarta.apache.org/commons/dbcp/configuration.html>. To use standard DBCP properties, just prefix them with "proai.db." here.
- **proai.db.connection.*** To use JDBC driver-specific properties (in DBCP, these are called "connectionProperties"), prefix them with "proai.db.connection.".
- ***.ddlConverter** The Java class responsible for creating the database tables required by Proai's record cache. The name of the property should be the driverClassName of the JDBC driver you're using, plus ".ddlConverter".
- ***.backslashEscape** Whether the database interprets backslashes in SQL strings as escape characters. This is needed so that Proai can formulate proper queries against the database. The name of the property should be the driverClassName of the JDBC driver you're using plus ".backslashEscape". If unspecified, backslashEscape will be assumed true.

Fedora Driver: General Settings

- **proai.driverClassName** The class name of the proai.driver.OAIDriver implementation to use. This value should never need changing -- it just informs Proai that we're using the Fedora driver.
- **driver.fedora.queryFactory** The class name of the fedora.services.oaiprotocol.QueryFactory implementation to use. This allows the OAI Provider to work with Fedora's Resource Index whether it's backed by Kowari or MPTStore. If the RI uses Kowari, the value here should be fedora.services.oaiprotocol.ITQLQueryFactory. If it uses MPTStore, the value should be fedora.services.oaiprotocol.MPTQueryFactory.
- **driver.fedora.baseURL** The base URL of the Fedora repository to query and get records from.
- **driver.fedora.user** Identifies a Fedora user that has privileges to do a resource index query via the research interface, and can get all needed disseminations that represent OAI records in the Fedora repository.
- **driver.fedora.pass** Password for the Fedora user.
- **driver.fedora.queryConnectionTimeout** When querying the resource index, the maximum number of seconds to wait for the http connection to be established before giving up.
- **driver.fedora.querySocketTimeout** When querying the resource index, the maximum number of seconds of socket inactivity to allow before giving up.
- **driver.fedora.disseminationConnectionTimeout** When getting xml data from Fedora, the maximum number of seconds to wait for the http connection to be established before giving up.
- **driver.fedora.disseminationSocketTimeout** When getting xml data from Fedora, the maximum number of seconds of socket inactivity to allow before giving up.
- **driver.fedora.identify** The location of the XML that should be used to respond to OAI "Identify" requests. This can be any URL, but typically it points to a manually-created datastream within the same Fedora repository that hosts the OAI records.
- **driver.fedora.itemID** The name of the RELS-EXT property whose value (a plain literal) contains the OAI item ID for records that are provided by an object. A value for this property *must* exist in RELS-EXT before any of an object's disseminations will be considered to be OAI records.

Fedora Driver: Metadata Format Configuration

- **driver.fedora.md.formats** Metadata formats to make available. This is a space-delimited list of all formats provided, identified by OAI metadata prefix.
- ***.loc** The location of the W3C schema for each format. Example property name: driver.fedora.md.format.your_format.loc.
- ***.uri** The namespace URI for each format. Example property name: driver.fedora.md.format.your_format.uri.
- ***.dissType** The Fedora dissemination type for each format.
A Fedora dissemination type is a URI starting with `info:fedora//` and ending with a datastream ID (such as "DC"), a Behavior Definition PID followed by `"/methodName"`, or a Behavior Definition PID followed by `"/methodName?name=value"`. The `dissType` is the key to mapping an OAI metadata format to a kind of Fedora dissemination. Here are a few examples: `info:fedora//DC` ; identifies the "DC" datastream. `info:fedora//demo:1/getDC` ; identifies the "getDC" method of the `demo:1` behavior definition. `info:fedora//demo:1/getMD?fmt=dc` ; identifies the "getMD" method of the `demo:1` behavior definition, when invoked with the required "fmt" parameter. When the OAI provider queries Fedora for records in your_format, it uses this special value to constrain the query to only those disseminations that are in the expected format. Thus, all records that the OAI provider considers to be in your_format must have this dissemination type. Example property name: driver.fedora.md.format.your_format.dissType.
- ***.about.dissType** The Fedora dissemination type for each format. This optional property identifies the OAI "about" dissemination type for your_format. If specified for your_format, then the OAI provider will attempt to find disseminations of this type for each object that has a matching your_format.dissType. If such a dissemination is found, for that particular object, the information therein will be used as the "about" metadata for the record. Example property name: driver.fedora.md.format.your_format.about.dissType.

Fedora Driver: Set Membership Configuration

- **driver.fedora.setSpec** The name of the RELS-EXT property that indicates the unique setSpec string of a Fedora object representing an OAI set. This property must exist on any Fedora object that should represent an OAI set.
- **driver.fedora.setSpec.name** The name of the RELS-EXT property that indicates the name of a Fedora object representing an OAI set. In addition to setSpec, this property must also exist on any Fedora object that should represent an OAI set.
- **driver.fedora.setSpec.desc.dissType** The dissemination type for the XML that describes a set. If a dissemination of this type is available for a Fedora object representing an OAI set, the content of the dissemination will be used in ListSets responses.
- **driver.fedora.itemSetSpecPath** The "triple pattern" identifying the path from objects to the setsSpecs (short names) of sets they are members of.
The path must contain \$item and \$setSpec, with the \$item variable representing a Fedora object in the resource index, and the \$setSpec representing the setSpec name.

Fedora Driver: MPTStore Configuration

NOTE The values below only need to be configured if using MPTQueryFactory (Fedora's Resource Index is backed by MPTStore).

- **driver.fedora.mpt.db.driverClassName** The JDBC driver class for connecting to the MPTStore database.
- **driver.fedora.mpt.db.ddlGenerator** The database-specific DDLGenerator to use for MPTStore.
- **driver.fedora.mpt.jdbc.url** The JDBC URL for connecting to the MPTStore database.
- **driver.fedora.mpt.jdbc.user** The user to connect to the MPTStore database as.
- **driver.fedora.mpt.jdbc.password** The MPTStore database user's password.
- **driver.fedora.mpt.db.map** This configuration value is necessary to initialize the MPTStore library, and should not need to be changed.
- **driver.fedora.mpt.db.prefix** This configuration value is necessary to initialize the MPTStore library, and should not need to be changed.

Fedora Driver: Advanced Configuration

- **driver.fedora.deleted** This is the name of the RELS-EXT predicate that indicates the state of all records within the object. The normal behavior is that the Fedora-controlled "state" property of the dissemination (or datastream) itself indicates the oai deleted status. This can be used to specify the status via any object property, if desired.