

Fedora Objects and Data Streams

Overview

Fedora is used to store all metadata for the Spartan Archive. This includes Premis and Dublin Core information for each collection and instance, along with data definitions and organizing information. Each object is identified by a permanent ID or PID. All Spartan Archive PIDs use the prefix--msu-uahc:.

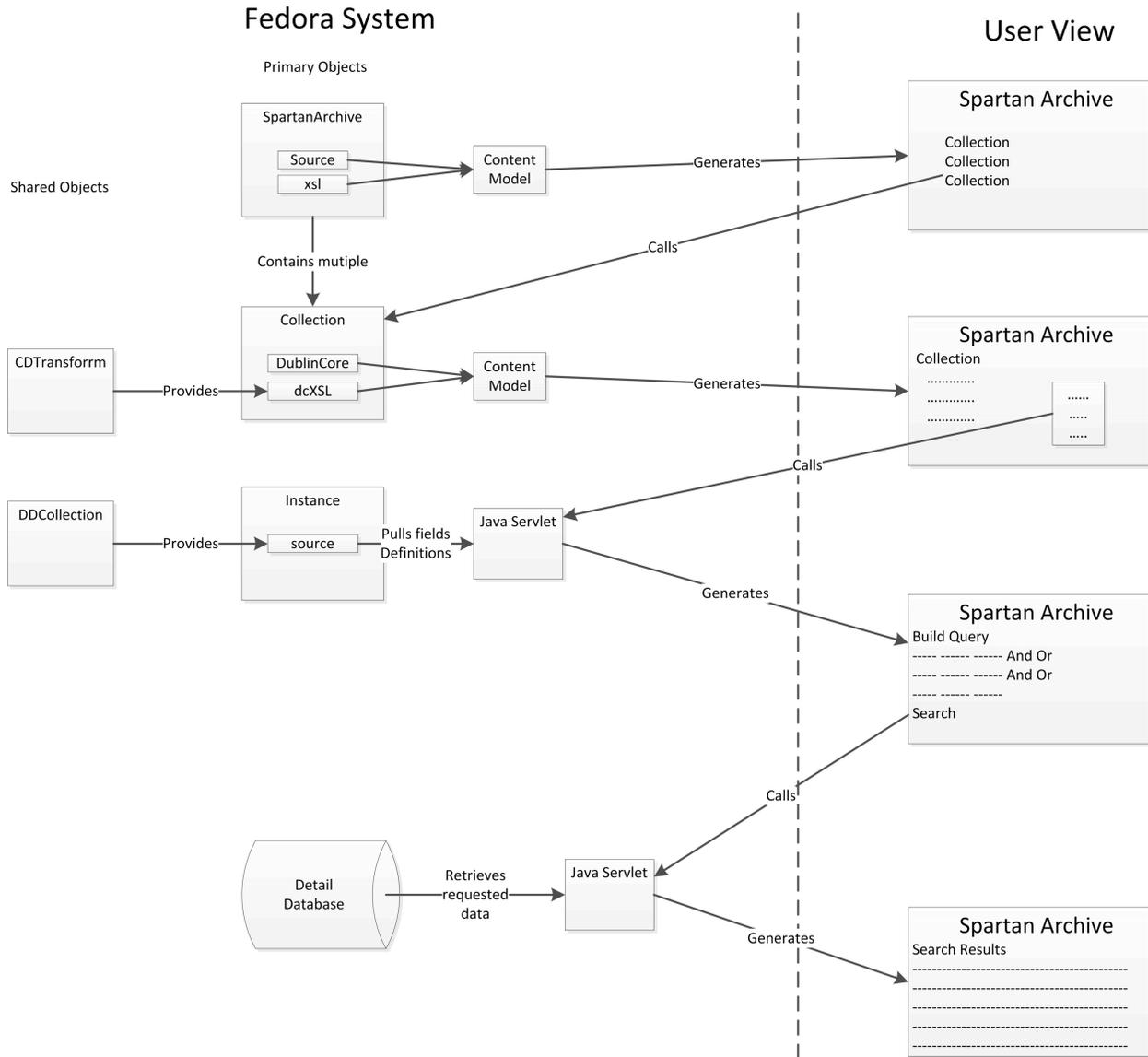
Each object contains one or more data streams. Data streams are Fedora's mechanism for storing data and metadata. Each object is required to contain a 'dc' data stream which is a limited-functionality Dublin Core metadata file containing identifier information for Fedora. Because it is limited in its functionality, many of the data objects in the Spartan Archives implementation of Fedora also contain a 'Dublin Core' data stream which has a richer set of metadata objects. Similarly objects representing instances of collections also contain a 'Premis' data stream which contains premis metadata for the object. Finally any Spartan Archive object which is expected to provide access to its data streams will contain one or more xsl stylesheet data streams which instruct the browser on how to display the data. (For the most part these style sheets are common to multiple objects and are stored in a single object in Fedora. They are then referenced in the individual Spartan Archives objects.)

Another data stream used in most Spartan Archive objects is the 'RELS-EXT' which is intended to document relationships between objects. At this time the only use for these data streams is to associate objects with their content model and related display methods.

Finally Fedora maintains a set of objects used for providing specialized access to data and metadata. These are the content models and the associated display methods. The Spartan Archive currently uses a single content model which supports xslt transformations. In the future additional content models will be added to support access to other types of digital objects.

The following diagram displays the relationship between the objects in the Spartan Archive Fedora system.

Spartan Archives: Fedora and Users



Fedora Object Relationships

Spartan Archive Objects

The entry point to the Spartan Archives is the object `msu-uahc:SpartanArchive`. It contains 3 data streams: 'source', 'xsl', and 'defaultTypes'. The 'source' data stream contains information about each collection in the archive. This includes basic metadata, schedule information and some processing aids. (See the data stream listings for more detail.) The 'xsl' data stream contains the stylesheet needed to present the first page in the Spartan Archive access system (including links to the specific collections). The 'defaultTypes' data stream contains all of definitions of configurable items in the archive system. The actual configuration are maintained in each separate collection.

The content models “GetContent” method generates the following page.

The Spartan Archives Michigan State University's Home for Digital Preservation

Collections

[Academic Programs](#)

Creator: MSU Office of the Registrar

Last Updated: July 16, 2011

Period Covered: 2011-2012

This collection consists of each semester's student directory which consists of local and permanent contact information for MSU students. A new set is collected every semester.

[Course Descriptions](#)

Creator: MSU Office of the Registrar

Last Updated: August 26, 2011

Period Covered: 1992 to 2012

This collection consists of each year's course descriptions. A new set is collected every year.

[Schedule of Courses](#)

Creator: MSU Office of the Registrar

Last Updated: August 26, 2011

Period Covered: Fall 2003 to Fall 2011

This collection consists of each semester's schedule of courses which consists of information about each class offered. A new set is collected every semester.

[Student Directory](#)

Creator: MSU Office of the Registrar

Last Updated: August 26, 2011

Period Covered: Spring 1994 to Fall 2011

This collection consists of each semester's student directory which consists of local and permanent contact information for MSU students. A new set is collected every semester.

Collection Objects:

Each Collection within the Archive has its own object identified by a PID with the form:

msu-uahc:RG-CollectionName

where RG is the record group. So for example the PID for the Academic Programs collection is msu-uahc:UA.6.7-AcademicPrograms.

Each collection object has 6 data streams (along with the default 'dc' data stream). Examples can be found in the Data Stream section. These are:

Data Stream	Purpose
RELS-EXT	Connects the object with the content model for displaying the metadata
source	These are the data definitions for the collection
xsl	The stylesheet for formatting the data definitions.
Dublin Core	The Dublin Core metadata for the collection
dcXSL	The stylesheet for formatting the Dublin Core. This stylesheet is used for all collections and is referenced here from: http://fedora.ats.msu.edu:8080/fedora/get/msu-uahc:CDTransform/dcXSL This is used to display the page for each collection including creating the links to the specific query pages. See examples below:
defaults	These contain the configuration items specific to this collection

Each collection page is generated using the 'GetContent' method of its content model, using the DublinCore xml data stream transformed according to its dcXSL stylesheet to produce a page like the one below.

The Spartan Archives

Michigan State University's Home for Digital Preservation

Current Collection: Schedule of Courses

[Search the complete collection.](#) or choose a period below.

Creator: Michigan State University. Office of the Registrar

Last Updated: 2011-08-15

Period Covered: Spring Semester 1994 - Spring Semester 2012

Schedule listing all courses, locations, instructors and special instructions for a given semester.

[Spring 2012](#)



Schedule of Courses Collection Page

Instance Objects

Each Instance within the Archive has its own object identified by a pid with the form:

msu-uahc:RG-ATAccessionNumber-XXX

where RG is the record group. ATAccessionNumber is the number assigned in Archivist's toolkit for this submission and XXX is a sequential number beginning a 001. This number is used when more than one instance is accessioned at one time. As an example of that the twenty years of Schedule of Courses submissions were accessioned at once. As so each instance is differentiated by the suffix. Thus a sample instance object pid might be msu-uahc:UA.6.7-A.2012.0021-001.

Each instance object has 5 data streams (along with the default 'dc' data stream). Examples can be found in the Data Stream section. These are

Data Stream	Purpose
RELS-EXT	Connects the object with the content model for displaying the metadata
Premis	The premis metadata for the instance
Dublin Core	The Dublin Core metadata for the collection

Data Stream	Purpose
source	The data definitions for the collection. These not actually stored in each instance object. Rather they exist as data streams within a specific data definition object, and are referenced within the instance object by their Fedora url. For example: http://spartanarchive.msu.edu:8080/fedora/objects/msu-uahc:DDAcademicPrograms/datastreams/source/content
xsl	This is the stylesheet for the data definitions. It is referenced from http://fedora.ats.msu.edu:8080/fedora/get/msu-uahc:DDTransform/xsl

Shared Objects

There a number of shared objects, that is secondary objects that support multiple primary objects. These can be classified into three groups:

- Content Models and Services
- Data Definitions
- Stylesheets.

Content Models and Services

Currently only one content model is used in Spartan Archive. Because the only current content in the Spartan Archive is the records based data from the Office of the Registrar, the content model only needs to support displaying the metadata for the collections and their instances. All other data is provided to researchers via the Java backend. These objects are:

Object	Purpose
msu-uahc:ex3CModel	Content model that supports applying a stylesheet against metadata (in xml form) to display a web page. The CModel object has a RELS-EXT data stream that ties it to the service definitions that it supports. In this case those are msu-uahc:ex3CSDef and demo:ex3CSDef.

Object	Purpose
msu-uahc:ex3CSDef	<p>Service definition object that supports displaying the 'source' data streams that define the data definitions via the xsl stylesheet data source.</p> <p>SDef objects contain a MethodMap data stream that names the services provided via this definition. In this case it is getContent.</p>
msu-uahc:ex3CSDep	<p>Service deployment object that maps the data sources (in this case the 'source' and 'xml' data streams) to a web service that actually performs the conversion.</p> <p>SDep objects have a RELS-EXT data stream that ties the deployment back to both a SDef definition, and the content models that use the service.</p> <p>In addition they have a MethodMap data stream that does the actual mapping between the inputs and the web service, a DSINPUTSPEC data stream that defines the types of the input and a WSDL data stream that is required by the web service.</p>

Data Definitions

Each collection has a data definition object that defines each data item in the collection. The object contains a single data stream 'source' which is an xml representation of the data dictionary. This data stream is referenced in each instance object. (See the table in Instance Objects above for an example.) These definitions are used for two purposes in the system. First they provide the the list of fields available for query and second they are used to popup definitions for each of those fields.

Data Definition objects have PIDs of the form msu-uahc:DDCollectionName. So the Academic Programs collection data definition object would have a PID of msu-uahc:DDAcademicPrograms.

Stylesheets

In order to have a consistent presentation for collections, stylesheets are shared across the Spartan Archive system.

The collection definition stylesheet is contained in msu-uahc:CDTransform. This has a single data stream called 'dcXSL.' This data stream transforms a collections Dublin Core metadata into the collection webpage. This data stream is referenced from each collection object.