Spartan Archive Architecture
Each Registrar's Office data source is a collection.

Fedora

Collections

Each Collection has a Content Model that Links the collection To its access methods

Each Collection links to multiple Instances and provides access methods For the instances

Collection Instances

Each ingest forms a collection instance

Fedora

Each instance points at The location of its detail data

Detail Database

Current design has each instance's data in the same database. The primary key includes an instance identifier to allow them to be handled individually

Web Service For accessing data
Spartan Archive Architecture

The Spartan Archives is composed of 4 components:

**Ingest**: The Registrar’s Office will send two files for each data source. The first file contains the actual records of interest. The second file contains a checksum for the first. This component reads that data, verifies it for format and completeness, then modifies it to fit the needs of the Archive. Once that is done the data is split into two parts: Metadata (consisting of information about the data that is of interest to researchers or of utility in finding desired data or maintaining it) and Detail Data (consisting of the specific records from the Registrar.) Detail Data is stored in the Detail Database. Metadata, including the location of the Detail Data, is stored in Fedora.

Records that contain links to external, MSU, web sites will be converted to point to the saved equivalent in the MSU Web Archive.

**Fedora**: Fedora stores the metadata and provides facilities for managing the Archive and accessing its information (including that held in the Detail Database.) Each class of Registrar’s data comprises a collection. So there will be a collection for Academic Programs, Student Directories, etc. Each instance of that collection (the 2011-2012 Academic Programs for example) is a Collection Instance. Each of these (Collections and Collection Instances) is stored as a Content Model in Fedora. These models contain the metadata and have links to the facilities for accessing the data.

In addition Fedora is used to store traditional digital objects such as photos, videos and texts. These will be ingested through a mechanism imported from a separate project. There are a number of existing systems that handle ingest into Fedora for these objects. One of those will be used.

Fedora serves as the primary vehicle to access the archive. Researchers will access Fedora via the web. Fedora will determine the data requested and call appropriate web services to respond to the requests. These web services are defined in each collection’s Content Model.

**Detail Database**: This contains all of the specific records from the Registrar’s Office. Fedora’s Content Models will point to the specific data needed and it will be accessed through the Content Models accessor methods. Data in the database will be stored in as granular a means as possible to allow easy access to unanticipated questions.

**Permanent Archive**: Both the Fedora data and the database are exposed and depend strongly on current technology. In order to ensure that the Archive is available on a long term basis, it must be stored in as technology neutral and secure a way as possible. Thus data from Fedora will be extracted and stored in BagIt format, while database data will be extracted, converted to SIARD format and stored. As each new ingest is performed, the permanent archive data will be read, verified, and moved to duplicate repositories along with the newly ingested data. This process will be managed via iRODS (the integrated Rule Oriented Data System).

The primary storage will be tape. A mechanism will be developed to restore an archive to disk as needed.