Spartan Archive: A Program in Transition

Cynthia Ghering
August 23, 2011
Goals of Spartan Archive

- 3 yr project, **Apr 2009 to Mar 2013**, $251K for IT staff
- **Preservation environment** for born-digital records
- **Automate** ingest, manage descriptive and preservation metadata, archival storage, and public access
- **Scalable**, sustainable, and transferrable
- Structured/un-structured **digital data**
- **Integrate** with automated archival management
- Four records series from **Registrar’s Office**:
  - Academic Programs
  - Course Descriptions
  - Schedule of Courses
  - Student Directory
Analog to Format-Neutral Repository

- Holistic approach to electronic records
- Hybrid (analog and digital) collections
- Service-oriented archives
  - Flexible, adaptable, nimble archivists
- Funding model
  - Who pays for the new “virtual storage”? 
- Workflows for collections management
  - Regardless of material format

*Thanks to Jeanne Kramer-Smyth for the graphic. See her Spellbound Blog to order cool t-shirts!
Workflow for Collections Management

- DOS-based, flat database
- Migrated 4,000 records
- 2,000+ records on cards still
- No name or subject authorities
- Updating physical locations

- Modified AT for records management
- Updating paper based workflows
  - transmittals forms
  - records inventories
  - retentions schedules
Importance of Records Management

- MSU’s new Enterprise Business Systems
  - Finance and HR records now born-digital or imaged
  - Research administration in phase 2 starting soon
- Hybrid systems of record, “paper-sparse”
  - Compliance with federal and state regulations
- Records Authority by Denver University
  - Open-source, web-based software tool for RM
  - Surveying departments, inventorying records
  - Creating, managing, and distributing retention schedules
Spartan Archive – Progress to Date

- Selected software:
  - Fedora, iRODS, BagIt, SIARD, ARK, NOID
  - Islandora, DROID, JHOVE, JHOVE2
- Surveyed Registrar’s Office databases
- Clarified database structures and data fields
- Created XML schemas for each records series
- Develop Fedora collection model for each series
- Drafted ingest procedures for Registrar’s data
## Course Descriptions

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Full Name: Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj_Code</td>
<td>Subject Code: Defines the general subject of the course. Need a list of full names.</td>
</tr>
<tr>
<td>Crse_Code</td>
<td>Course Code: Identifier for the specific course.</td>
</tr>
<tr>
<td>Initial_Action</td>
<td>Appears to contain “Converted” or “Changed” What do these refer to? Are there other values?</td>
</tr>
<tr>
<td>Course_Status</td>
<td>Whether the course has been approved or archived. If archived then this indicates that there has been a change to the course’s attributes. Are there other values?</td>
</tr>
<tr>
<td>Start_Term_Seq_Id</td>
<td>?</td>
</tr>
<tr>
<td>Start_Term_Code</td>
<td>?</td>
</tr>
<tr>
<td>End_Term_Seq_Id</td>
<td>?</td>
</tr>
<tr>
<td>End_Term_Code</td>
<td>?</td>
</tr>
<tr>
<td>Interdprmntl_Flag</td>
<td>Y/N Identifies an interdepartmental course.</td>
</tr>
<tr>
<td>Interdepartmental_With</td>
<td>Identifies the cooperating department. Can there be multiple values? Is the code used the same as the Department Code?</td>
</tr>
<tr>
<td>Administered_By</td>
<td>Identifies the department that actually runs the course. Uses Department Code (?)</td>
</tr>
<tr>
<td>Coll_Code</td>
<td>College Code: Identifier of College that offers the course. Need a list of full names.</td>
</tr>
<tr>
<td>Dept_Code</td>
<td>Department Code: Identifier of Department that offers the course. Need a list of full names.</td>
</tr>
<tr>
<td>Crse_Title</td>
<td>Course Title: Name of the course</td>
</tr>
<tr>
<td>Crse_Description</td>
<td>Course Description: Description of the course contents.</td>
</tr>
<tr>
<td>Semester</td>
<td>Semester in which the course is offered. This is a text field that describes the general rule for when the course will be offered.</td>
</tr>
<tr>
<td>Credit_Type</td>
<td>Values seem to be F, V or Null. Null appears to be associated with archived courses. What do the other codes mean(V means variable?, F fixed?) and are there other values?</td>
</tr>
</tbody>
</table>
Registrar’s Office – Data Sources

The Academic Program

Sessions
Period Covered by Registration
Accreditation
Equal Opportunity
Student Consumer Information
Graduation Rates
Colleges, Schools and Departments
Programs and Areas of Study
  - Degrees
  - Teacher Certification
  - Course Designations
  - University Organization
XML Schemas for Data Sources

```xml
<xs:schema attributeFormDefault="unqualified">
  <xs:element name="publicCoursesSubmission">
    <xs:annotation>
      <xs:documentation>Ex. <publicCoursesSubmission semester="FS11"></xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element name="courseDetail" type="courseType" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="semester" type="termType" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:simpleType name="subjectType">
    <xs:annotation>
      <xs:documentation>Ex. MTH</xs:documentation>
    </xs:annotation>
    <xs:restriction base="xs:string">
      <xs:pattern value="[A-Z]{1,4}"/>
    </xs:restriction>
  </xs:simpleType>
  <xs:simpleType name="termType">
    <xs:annotation>
      <xs:documentation>Ex. FS11</xs:documentation>
    </xs:annotation>
    <xs:restriction base="xs:string">
      <xs:pattern value="[A-F]{0,2}[0-9]{2}"/>
    </xs:restriction>
  </xs:simpleType>
  <xs:simpleType name="creditType">
    <xs:annotation>
      <xs:documentation>Ex. 1 or 1.5</xs:documentation>
    </xs:annotation>
  </xs:simpleType>
</xs:schema>
```
Fedora Architecture

Each Registrar's Office data source is a collection.

Collections

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

Content Model

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
Federated Repositories

- Content models for research data, learning content, digital humanities
  - NSF mandate to preserve data for long-term access
  - Moodle requires a learning content repository for implementation
  - Multiple digital humanities projects underway
  - (beta) MSU Policy Repository

- Leverage complementary needs on campus
- Share progress currently underway at MSU
Project Team:

Rich Burgis, burgisri@msu.edu
Ed Busch, buschedw@msu.edu
Cynthia Ghering, ghering@msu.edu
Lisa Schmidt, Ischmidt@msu.edu