

SDS 2002 Annual Report

Institute for Advanced Technology in the Humanities
© 2003 University of Virginia. All rights reserved.

Table of Contents

1. Overview	1
2. Summary of progress this year	1
3. Staff and Spending	3
3.1. Staff	3
3.2. Presentations	4
3.3. Spending	5
3.4. Training, papers, consulting	5
4. Individual projects	6
4.1. METS	6
4.2. GDMS Editor	9
4.3. Salisbury	11
4.4. Pompeii	12
4.5. Tamino	14
4.5.1. Tibet	16
4.5.2. Whitman	18
4.5.3. Rossetti	19
4.5.4. Blake	19
5. Technical committee report	19
6. Policy Committee	21
7. Observations	21

7.1. Communication	21
7.2. Documentation	22
8. Future Work	22
App. GDMS code snippets	23
A-1. Salisbury	23
A-2. Pompeii	28

Overview

In 2000, the University of Virginia's Institute for Advanced Technology in the Humanities (IATH) and The University of Virginia Library's Digital Library Research and Development group (DLR&D) began a three-year project called "Supporting Digital Scholarship" (SDS) funded by the Andrew W. Mellon Foundation and co-directed by John Unsworth (Director, IATH) and Thornton Staples (Director, DLR&D). The SDS project's goal is to propose guidelines and document methods for second-generation digital libraries. The specific digital library problems under examination are:

1. scholarly use of digital primary resources;
2. library adoption of "born-digital" scholarly research; and
3. co-creation of digital resources by scholars, publishers, and libraries.

Summary of progress this year

Much of the work over the past year has centered around migrating projects from their working space (whether with a publisher or research group such as IATH) to a library-owned repository. This is not a simple matter of moving files and collecting metadata: there is a great need for collaboration between the project staff, resource group, publisher, and library. This will help avoid unnecessary work, unrealistic expectations, and wrong or useless data, all of which can lead to an uncollectable or unpreservable project. Digital libraries' policies can set guidelines that encourage authors to follow community-based standards, but policies are likely to be flexible enough to accommodate interesting and unusual work, leaving authors room to run (creatively) amuck.

We are studying how the collection process can be streamlined and standardized and what impact this has on the project. Figure 1 shows a high-level view of how the collection process might work.

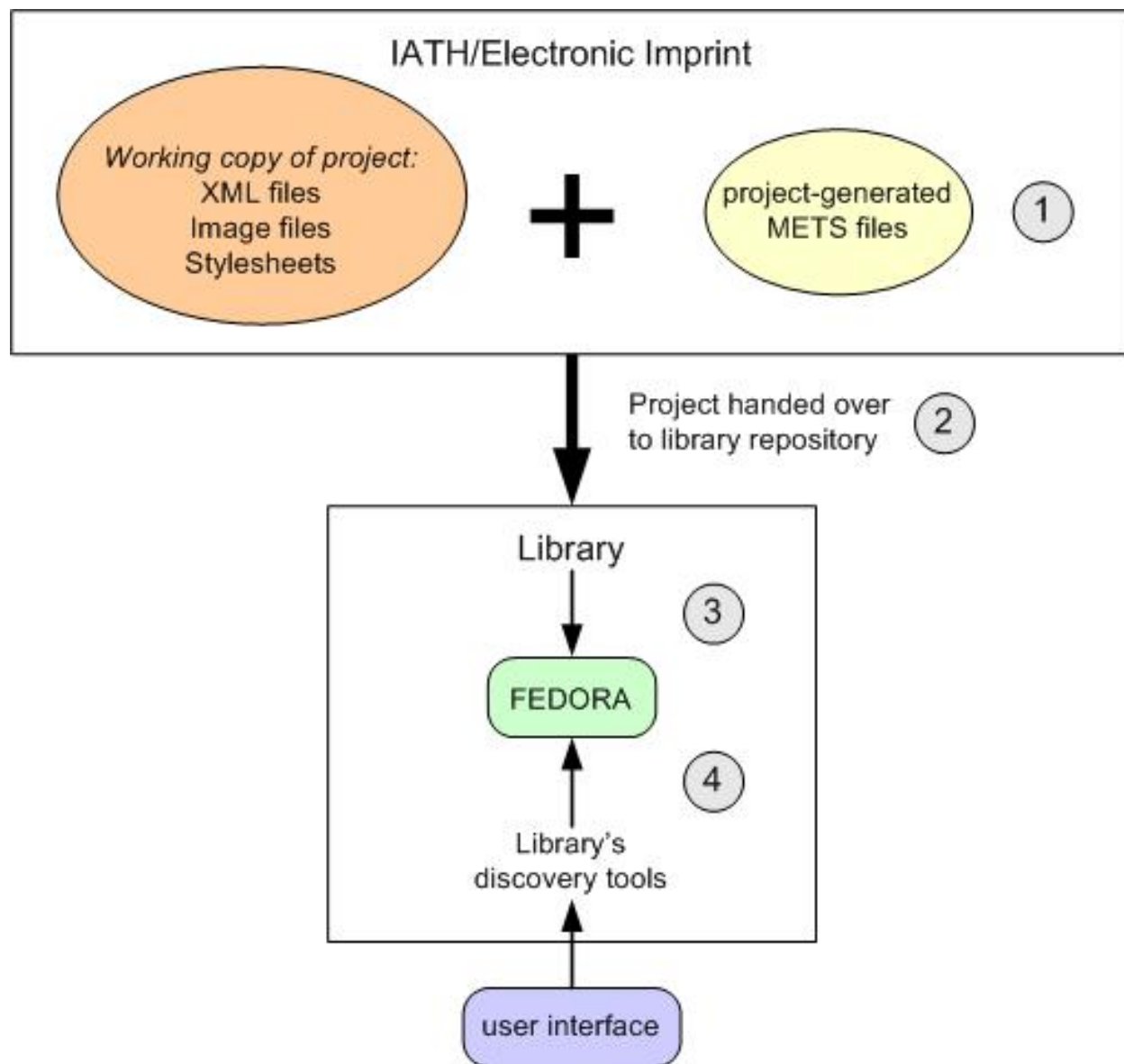


Figure 1. Moving projects into library repositories

Once a project is ready to be collected, the project generates Metadata Encoding and Transmission Standard (METS, section 4.1) files for the library (step 1 in figure 1). METS files are used as wrappers for the project's resources: they tell the library what files are in the project, what metadata, how the project is structured, and what behaviors the project uses. The package of project files and METS wrappers is handed off to the library (step 2). The library can then import the project into FEDORA for deposit and generate its own catalog records (step 3). Once the project is in FEDORA, library users can use it via the library's discovery and dissemination tools (step 4). METS files can be automatically generated by the project (whether by the project staff or a resource group or a publisher). However, the library will need some minimum sets of metadata (e.g., for generating MARC records), so the project authors must be informed of the library's collection requirements before handing the project over for deposit.

We have also been working on making General Descriptive Modelling Scheme (GDMS) easier to use. The GDMS editor (section 4.2) now has template and display features, making it easier to tag a project with GDMS. The template feature lets users reuse code snippets, which helps ensure that resources are properly tagged and that metadata is properly distributed. This can help avoid problems later on when the project is ready to generate METS files. Both the Salisbury and Pompeii projects tested the tool (see sections 4.3 and 4.4). Salisbury has imported almost all of its web site into GDMS and Pompeii has started importing its image catalog.

We are working closely with Software AG's Tamino software development group to import four test projects, so that we can use Tamino as an XML search engine (see section 4.5). We had thought that Tamino would prove to be a useful tool but we have not yet been able to use Tamino on live sites, primarily because of problems with indexing complex projects. Tamino is optimized for business users, but doesn't support all of the features that digital libraries require. Software AG asserts that these and other problems will be solved with the next release of the Tamino software (due in the 2nd or 3rd quarter of 2003). Although we have been trying to overcome our reliance on structured proprietary software, it has proved difficult. We cannot find software that offers all of the features that we require. Tamino offers some support for XML but it has not yet proved able to cope with more demanding projects, such as the Rossetti project. We've been looking into eXist (<http://exist-db.org/>), an open source native XML database, but it doesn't offer full XPath support and may not scale up to handle our larger projects. We would consider using commercial software if we could find an application that fits our purposes. The UVA library has been investigating XPAT (<http://dlxs.org/products/xpat.html>), an XML/SGML-aware search engine developed at the University of Michigan, but XPAT does not support Unicode (which is used as a basis for many common fonts). The library doesn't currently consider XPAT a good long-term candidate but it may be good stop-gap for FEDORA collections. Xindice (<http://xml.apache.org/xindice/>), another open source native XML database, and Ipedo (http://www.ipedo.com/html/products_xml_dat.html), a proprietary XML database, may also be worth investigation. One possible temporary solution to this situation is to reverse engineer queries, using style sheets to translate from X-Query into a format used by a proprietary application.

The Technical and Policy Committees met regularly for much of 2002 and set up subcommittees to tackle specific tasks. The Technical Committee (section 5) is looking at both authoring and user issues, centering around questions of metadata. The Policy Committee (section 6) is developing a draft policy statement that incorporates recommendations from the Research Library Group (RLG, <http://www.rlg.org>) and the Online Computer Library Center (OCLC, <http://www.oclc.org/home/>).

Staff and Spending

In 2002, we hired Cindy Girard to replace Kirk Hastings as an IT specialist. Cindy's position is partly supported by SDS funds.

3.1. Staff

There are two working committees currently directing SDS activities, one for technical issues and one for policy issues.

The committee members are:

Technical Committee:

Rob Cordaro, Library (Digital Library Research and Development Group) *
Chris Jessee, IATH *
Worthy Martin, IATH/Computer Science
Daniel Pitti, IATH
Perry Roland, Library (DLRDG)
Thornton Staples, Library (DLRDG), Co-Chair
John Unsworth, IATH/English, Co-Chair
Ross Wayland, Library (DLRDG)

*staff wholly or partly supported on SDS funding.

Policy Committee:

George Crafts, Library (Humanities Services)
John Dobbins, Art
Bradley Daigle, Library (Special Collections)
Daniel Pitti, IATH, Chair
Thornton Staples, Library (DLRDG)
Melinda Baumann, Library (Digital Library Production Services)
Leslie Johnston, Library (Digital Services Integration)

3.2. Presentations

We invited several outside experts to give public presentations to the SDS and UVA community and to consult on the SDS project. Links to on-line versions of several of these presentations are on the SDS web site (<http://jefferson.village.virginia.edu/sds/>). In 2002 these events were:

- * "Authenticated Electronic Editions Project," presentation about Just In Time Markup, by Paul Eggert and Philip Berry of the Australian Centre for Scholarly Editions. February 2002.
- * "Digital Preservation: Challenges, Services and Costs," a presentation by Stephen Chapman of the Weissman Preservation Center at the Harvard University Library. March 2002.
- * Presentation by Anne Kenney of the Cornell University Department of Preservation and Conservation. April 2002.
- * "Dimensions of Document Trustworthiness," a presentation by Heather MacNeil of the School of Library, Archival and Information Studies at the University of British Columbia. May 2002.
- * "Functional Requirements for Bibliographic Records: What is FRBR and how will it impact our cataloging rules and authority control?," a presentation by Barbara Tillett, Chief, Cataloging Policy and Support Office, of the Library of Congress. November 2002.
- * "Preserving the Digital," a presentation by Clifford Lynch, Director of the

3.3. Spending

SDS spending to date:

	Expenditures Calendar year 2000	Expenditures Calendar year 2001	Expenditures Calendar year 2002	Grand total To date
Salaries	\$98,058.49	\$225,741.17	\$277,243.78	\$601,043.44
Consulting	1,589.62	5,421.13	6,156.57	13,167.32
Travel	6,856.12	17,633.00	10,909.03	35,398.15
Training	4,491.73	17,117.59	6,160.85	27,770.17
Research	4,750.89	31,188.82	5,705.04	41,644.75
Total spent:	\$115,746.85	\$297,101.71	\$306,175.27	\$719,023.83

3.4. Training, papers, consulting

Training

- * Tamino training, Software AG, at IATH. April 2002.
- * Chris Jessee, Mac OS X Server Essentials and Mac OS X Administration Basics, Reston I, Reston, VA. June 2002.

Travel and Papers delivered or proposed on matters relevant to SDS

- * John Unsworth, "Using Digital Primary Resources to Produce Scholarship in Print," delivered as part of "The Future of Literary Studies," a conference of the English Department at the University of Virginia. April 2002.
- * John Unsworth, "Humanities and Computing at IATH," delivered with Worthy Martin as the Design and Technology Lecture at the Franke Institute for the Humanities, University of Chicago. April 2002.
- * John Unsworth, Panelist, "Reinventing Publishing: Digital Libraries," part of Publishing in the 21st Century, Library of Congress, Washington, DC. April 2002.
- * John Unsworth, Presenter, "Session One: New-Model Scholarship and the Creation of Web-Based Documents," part of Preserving Web-Based Documents, a symposium convened by the Council on Library and Information Resources, Washington, DC. April 2002.

- * John Unsworth, Panelist, "The Emergence of Digital Scholarship: New Models for Librarians, Archivists, and Humanists." RBMS Program, American Library Association Annual Meeting, Atlanta, GA. June 2002.
- * Thornton Staples, "The Supporting Digital Scholarship Project," paper delivered to the Digital Library Federation, Seattle, WA. November 2002.
- * John Unsworth, "Cutting a Gordian Knot," delivered as a talk in the Text Studies series at The University of Nebraska, Lincoln, NE. November 2002.
- * John Unsworth, "The Emergence of Digital Scholarship: New Models for Librarians, Scholars, and Publishers," delivered as part of The New Scholarship: Scholarship and Libraries in the 21st Century, Dartmouth College, Hanover, NH. November 2002.
- * John Unsworth, "Using Digital Primary Resources to Produce Scholarship in Print," presented as part of Literary Studies in Cyberspace: Texts, Contexts, and Criticism. Concourse G, Hilton. Modern Language Association Annual Convention. December 2002.

Consultants

- * Jerome McDonough, NYU Digital Library Group, January 2002. Consultation for implementation of FEDORA and use of METS.
- * Paul Eggert and Phil Berry, Australian Centre for Scholarly Editions, March 2002. Public presentation and discussion of Just In Time Markup technology.
- * Stephen Chapman, Weissman Preservation Center, Harvard University, March 2002. Public presentation and discussion of preservation and collection of scholarly digital work at Harvard.
- * Anne Kenney, Cornell University Department of Preservation and Conservation, April 2002. Public presentation and discussion of preservation issues.
- * Heather MacNeil, School of Library, Archival and Information Studies, University of British Columbia, May 2002. Public presentation and discussion of authenticity.
- * Barbara Tillett, Library of Congress, November 2002. Public presentation and discussion of LoC cataloging policy and SDS.
- * Clifford Lynch, Coalition for Networked Information, December 2002. Public presentation and discussion of preservation.
- * Terry Catapano, December 2002. Consultation for tech specs for METS.

Individual projects

4.1. METS

We are working with Terry Catapano, a Librarian and Special Collections Analyst in Columbia University Libraries Digital Library Program and formerly the Electronic Text Manager at the New York Public Library, to develop technical specifications for Metadata Encoding and Transmission Standard (METS),

<http://www.loc.gov/standards/mets/>) files. Terry has extensive consulting experience in XML-based Digital Library project development, and has also consulted for the New York Academy of Medicine, the Pierpont Morgan Library, and the New York Botanic Garden. We will be using METS to prepare projects for migration from their work space (such as the jefferson server at IATH) to a library repository. We are working to develop technical specifications that are generic enough to use with multiple projects. The files should be automatically generated by a publisher, a research group such as IATH, or the project staff before the project is handed off to the collecting library (as shown in figure 1, above). We are using **The Complete Writings and Pictures of Dante Gabriel Rossetti** and **The William Blake Archive** as test cases.

The METS files will function as a roadmap for the library's collecting processes, providing information about a project's structure, its files, and its metadata. They will not detail the relationships and behaviors of the resources, since this can be deduced from the program's existing mark-up, but they will allow the library to quickly and efficiently learn how a project is put together. The library may then want to generate its own METS files, for cataloging or tracking purposes, based on the METS files provided by the project.

There are three basic levels of information, which can be contained in one or more METS files. The outline follows the Functional Requirements for Bibliographic Records model recommended by the International Federation of Library Associations and Institutions (<http://www.ifla.org/VII/s13/frbr/frbr.htm>). The FRBR model differentiates between a work and its expression. A "work" is an abstract entity, an idea that is represented by one or more "expressions." Expressions are realizations of a work. For example, in the Rossetti project, the work *The Blessed Damozel* is expressed as both a poem and a painting. A physical copy of the printed poem (in a book) is called a "manifestation" of the work, but the METS files won't go down to that level of detail.

1. **Project.** This is the starting point for the project. If there are multiple METS files, this will be the starting point and will contain pointers to the other files.
2. **Works and commentaries.** This is information about what works are in the projects and any commentaries about those works. This section may be in one or more separate files.
3. **Expressions.** This lists the resources in the projects and metadata about those resources. This section may be in one or more files.

A METS file has five sections, as shown in figure 2: descriptive metadata, administrative metadata, file groups, a structural map, and a list of behaviors.

<p>Dmd Section</p> <p>dmdSec id="xxx" mdWrap</p>
<p>Amd Section</p> <p>amdSec id="yyy" mdWrap</p>
<p>File Groups</p> <p>filegroup id="abc" file id="aaa" dmdID="xxx" amdID="yyy" file id="bbb" dmdID="xxx" amdID="yyy"</p>
<p>Structure Map</p> <p>div id="struct1" label="chapter1" fileID="aaa" fileID="bbb"</p>
<p>Behaviors</p> <p>behaviorSec id="diss1" structID="struct1" label="make_chapter" interfaceDef label="make_chapter definition" mechanism lable="make_chapter mechanism"</p>

Figure 2. METS file

The Descriptive Metadata section contains descriptive metadata, which the library may use for MARC records, Dublin Core records, etc. The Administrative Metadata section contains metadata related to copyright, digital provenance, source information and technical information. There can be multiple descriptive and administrative metadata sections, each containing a chunk of information and each having its own unique identifier. In both of these sections, the metadata can be included in the METS file in an <mdWrap> element or stored in an external file and referred to in an <mdRef> element.

The File Groups section holds the names and locations of all the project files. These groups may be files of a specific format, files that share a common source, or any other arrangement that seems appropriate. A parent <fileGrp> tag can hold individual file names or one or more child file groups, each with its own id. A file group will associate its files with a particular chunk of metadata by including an pointer back to one or more descriptive metadata sections (via the <dmdID> element) and administrative metadata sections (via the <amdID> element). In figure 2, for example, file group abc includes the metadata in the xxx and yyy metadata sections.

The Structural Map explains the project's structure and how the files fit into that structure. It contains nested groups of <div> elements in a hierarchical structure that reflects the project's organizational or navigational arrangement. The <fileID> elements point to individual files or file groups in the File Groups section. The <div> usually has attributes indicating what its purpose is: in figure 2, struct1 contains the files that make up chapter 1 of a book.

Finally, the Behavior section matches behaviors with the Structure Map <div>s. A behavior is a piece of executable code that implements a particular task, such as displaying an image. The <interfaceDef> element defines the behavior and the <mechanism> element points to the executable code. The diss1 behavior section shown in figure 2 describes a make_chapter behavior, which will assemble the files in files aaa and bbb into a chapter.

4.2. GDMS Editor

A General Descriptive Modelling Scheme (GDMS) editing tool has been under development in the last year, and has been used for two test projects, **The Pompeii Forum Project** and **The Salisbury Project**. Staff from these projects began working with the tool in the second half of 2002 and have produced working GDMS XML documents.

Two major new features have been added: templates and display. The template feature (shown in figure 3) lets users create snippets of XML tags and metadata that can be re-used as necessary. This saves time when tagging a group of resources that have common metadata, such as a group of photos that were taken by the same photographer on the same day. Rather than type the photographer name in over and over again (and risk spelling and consistency errors), the user can use the template. The user can edit the working copy of the template as necessary. Figure 3 shows a template for **The Salisbury Project** template.

Input Template

Add res Node

res:

id:

label:

type:

description:

Content:

Attributes:

agent:

Content:

Attributes:

role:

form:

type:

Figure 3. GDMS template

The display feature lets the user open the document with an XSL style sheet to review the document's contents, see how the resources are presented, and check that the information is correct. A default style sheet is provided, but users can edit it or add other style sheets as necessary. Figure 4 shows a Salisbury document in a display window.

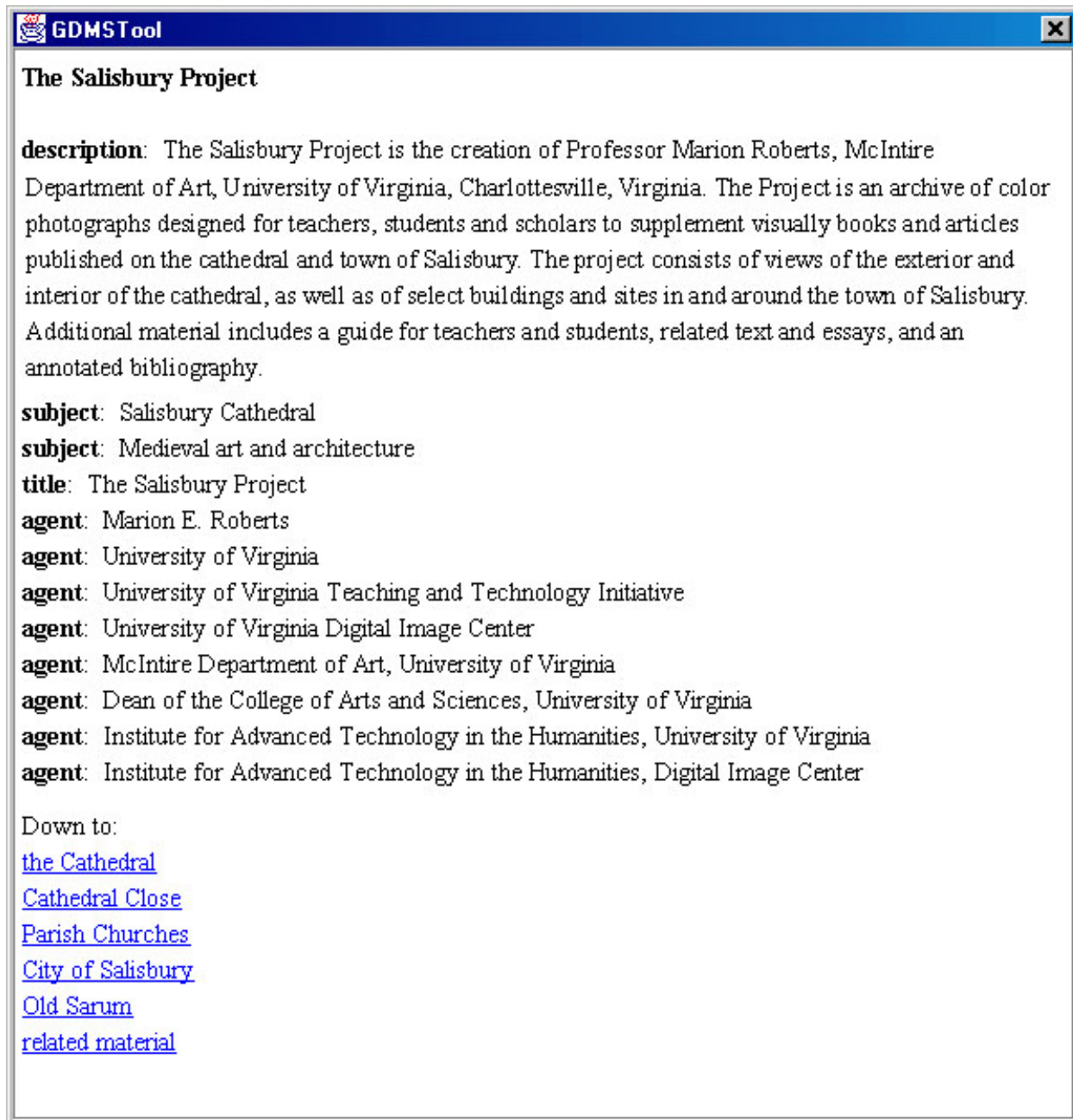


Figure 4. GDMS display

4.3. Salisbury

The Salisbury Project (<http://www.iath.virginia.edu/salisbury/>) was the first project that we worked with in SDS. In our first pass, we moved the cathedral image archive section of the project from EAD into GDMS, which was explicitly designed to be able to handle that kind of application. In 2002, the project manager, Catherine Walden, spent two months in the summer and several weeks at the end of 2002 moving the rest of the project into GDMS, using the GDMS editor. Salisbury is one of the primary SDS test projects and this transition was intended both to test the GDMS tool and to see how

difficult it might be to move a complete project from a web site into GDMS.

In addition to the cathedral image archive, the project originally contained a variety of other materials that were developed as web pages directly in HTML. These included a variety of teaching materials associated with Salisbury, including maps, bibliographies, lists of links and a variety of texts written specifically for the project. In addition, Marion Roberts, the project director, had originally intended to include image archives for other sites around Salisbury that related directly to the cathedral. Catherine and Marion digitized many new images to fill these in, while Catherine captured the other materials and links to the new images in GDMS.

The GDMS editor proved to be a very useful tool, especially since Catherine was not familiar with the GDMS DTD structure. She was able to create a GDMS file that contains the entire project (part of the file is in Appendix 1, section A-1). The project does not yet have style sheets for GDMS documents, so the site has not yet used these files, but that will be done in the current year of the SDS project.

4.4. Pompeii

The SDS Technical Committee has been working with **The Pompeii Forum Project** (PFP, <http://www.iath.virginia.edu/pompeii/>) to build an information structure in GDMS. PFP presents problems associated with collecting, generating, categorizing, and storing metadata that describes information resources (such as pictures of the forum) and higher level concepts (such as intellectual information about the Forum). The site's primary resources are images and analytic information gathered by PFP team members during on-site digging and research. The images are photos taken by PFP team members of the physical site during site digs and research visits. The photos were later digitized and marked with identifying information (where and when they were taken, what is shown in each photo, etc.). The analytic information consists of field notes and observations. The site also has secondary research, educational tools, etc. but these elements have not yet been imported into GDMS.

The resources and the metadata are being imported into GDMS with the GDMS tool, but the project's wealth of information require a carefully thought-out information structure. The general structure is fairly simple: digital image files are stored in an image catalog. Observations and field notes are in an observation catalog. Intellectual information about the Forum itself is kept in <div>s in the physically oriented structure (POS), which uses the Forum as a virtual structural hierarchy for arranging resource metadata. Figure 5 shows a high-level view of this structure.

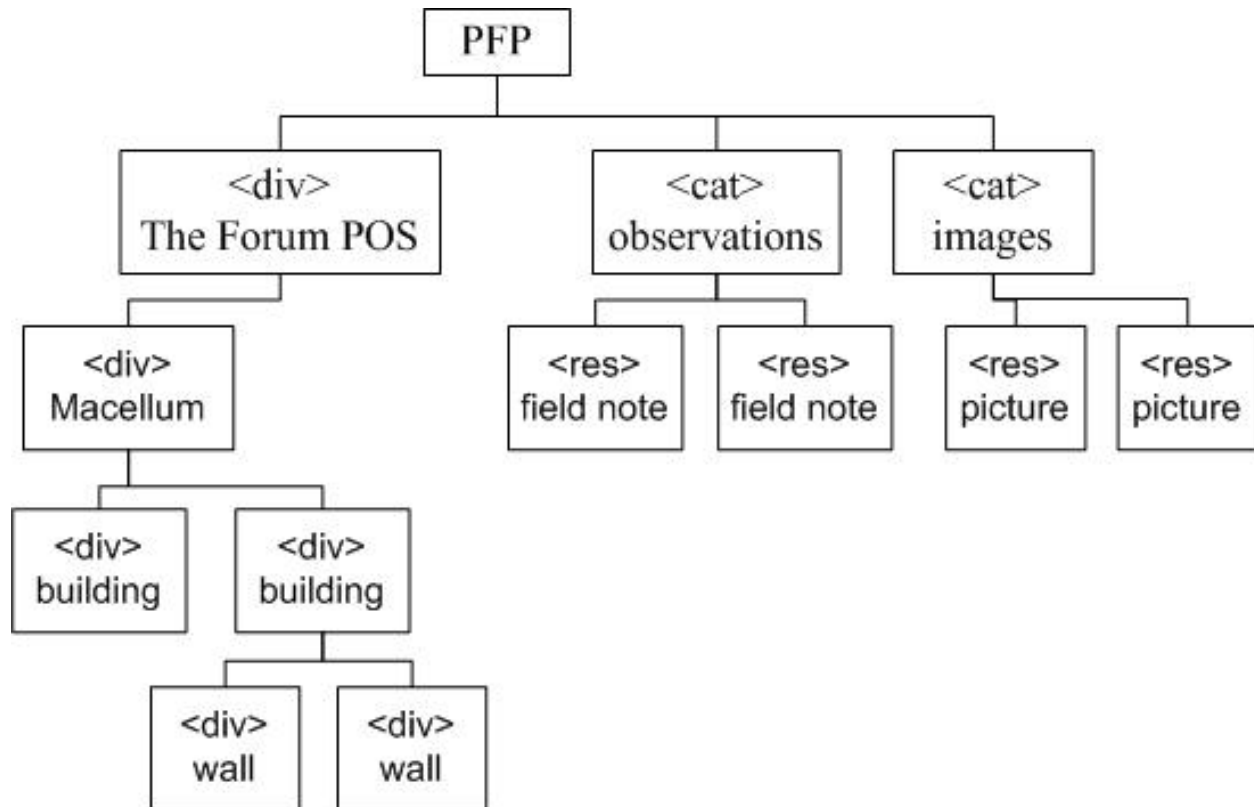


Figure 5. PFP in GDMS

The POS stores details related to the entire forum at the top of the hierarchy and details about individual walls and columns at the bottom. The catalogs can hold metadata related to the image and observations files. This is a convenient and intuitive arrangement for the project staff, but can quickly become inconsistent and controversial if different members of the project staff have different ideas of how to categorize and organize information. It may seem clear that metadata about images should be stored in the image catalog, but what about the description of an image's content? If a photograph shows a building, should that metadata be kept in the image catalog or in the POS <div> that describes the building? What if it shows two different buildings, which are described in two different <div>s? Furthermore, whatever information structure the project decides upon must be robust enough to be collected by the library and imported into FEDORA.

One of the more time-consuming aspects of this problem is identifying and organizing metadata that explains relationships between resources. If a photo shows two walls standing next to each other, how should this relationship be noted? Wall A may stand next to Wall B, but of course Wall B also stands next to Wall A. Relevant information about this relationship may be in an image file that shows the two walls, a journal article that describes the exact location of the walls, field notes with measurements, and in metadata that describes other resources (such as Column C, which is in front of Wall B). If there are dozens or hundreds of objects, however, the user must choose a standard method for recording this information. Figure 6 shows three different ways to record the relationship between the two walls.

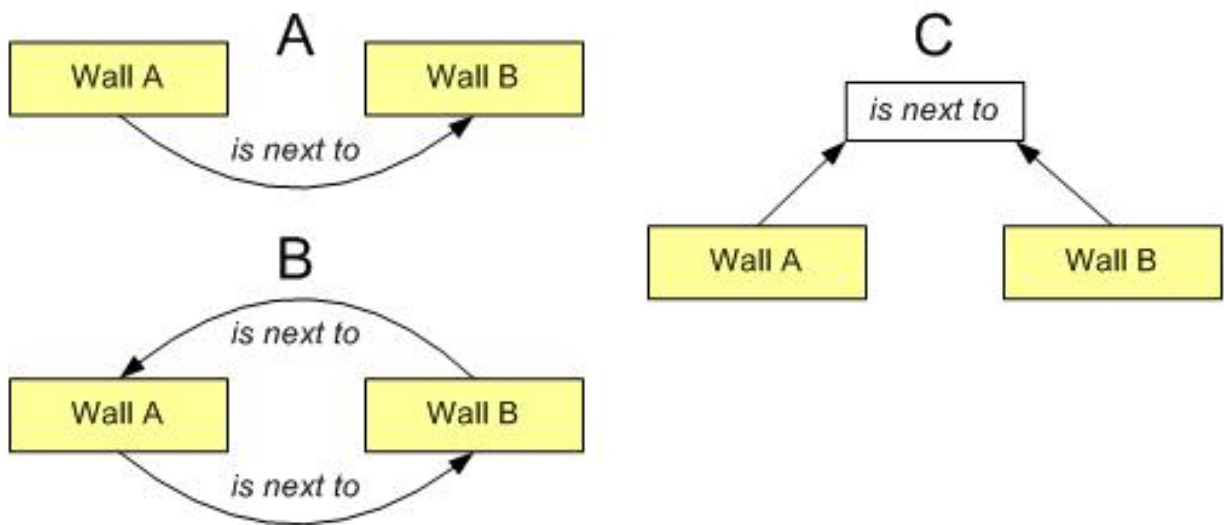


Figure 6. Representing relationships

In 6A, Wall A knows that it has a "is next to" relationship with Wall B. This is called a binary relationship, since it involves two objects, and it can be very useful. However, Wall B doesn't know about Wall A, and if a user wants to know everything about Wall B, he would have to find out about Wall A proximity from another source. In 6B, both walls know that they have an "is next to" relationship with something else. This arrangement becomes very difficult to maintain if there are many objects that dutifully describe their relationship with every other object. An alternative is to use an n-ary relationship, which treats the relationships itself as a quasi-object that can be separated from the objects, as in 6C. This avoids problems with explicit or implicit order and is much easier to maintain.

A PFP staff member is currently working the GDMS tool to build a GDMS version of the site. These files will then be imported into the library's FEDORA system. A sample of the GDMS file, with part of the image and observation catalogs and the POS <div> structure is in Appendix 2 (section A-2).

4.5. Tamino

Tamino (<http://www.softwareag.com/tamino/>) is an XML software package distributed by Software AG, a German software development company. We purchased Tamino last spring, intending to use it as an XML search engine. Tamino allows queries to be expressed in a standards-based query language, X-Query, rather than in a proprietary language. It stores XML documents and associated non-XML documents in a database and then indexes them. The indexed documents are used by a web-based user interface that searches the documents.

We have not been able to get Tamino up and running as quickly as we had expected, unfortunately. The software was successfully installed by mid-summer, and over the fall we spent time converting DTDs and files from SGML (which was used for DynaWeb disseminations) to XML, working on scripts to import documents into Tamino, and preparing style sheets for processing the results of Tamino searches. However, we had

two serious problems to contend with: entities and indexing.

Entities are used in XML documents to represent special characters (such as ampersands and m-dashes), create short-cuts for typing repetitive phrases and names, and to include external documents. Entities can be internal, referring to material within the current document (such as a short-cut for a phrase that is repeatedly used in the document), or external, referring to material outside the current document (such as an image file). Entities are resolved when the XML document is parsed by a web browser or some other processor. When documents are imported into Tamino, Tamino parses and resolves the internal and external entities. This means that if the external entities change (e.g., an image file is updated), the document in Tamino will not reflect that change: you cannot reverse engineer entities. If you were to import the same document with a changed entity, Tamino would treat it as a separate document.

This is not necessarily a flaw, since you can't properly index a document if you don't parse it. However, it can make it difficult to maintain an accurate and current set of project files, since old files may need to be removed by hand when updated files are added.

Indexing has also proved complicated. We intended to use Tamino's search engine capabilities to replace DynaWeb, but our test projects, Rossetti in particular, revealed weaknesses in Tamino's indexing function. The current Tamino release indexes documents only to a certain level of recursion. TEI-based documents have infinite levels of recursion and confound Tamino's indexing tools. The Tamino development team has been very cooperative in trying to work with us and is using the Rossetti project as a test case. They expect to have this problem solved with the next Tamino release, which is expected in March.

An example of how Tamino is going to be used is below:

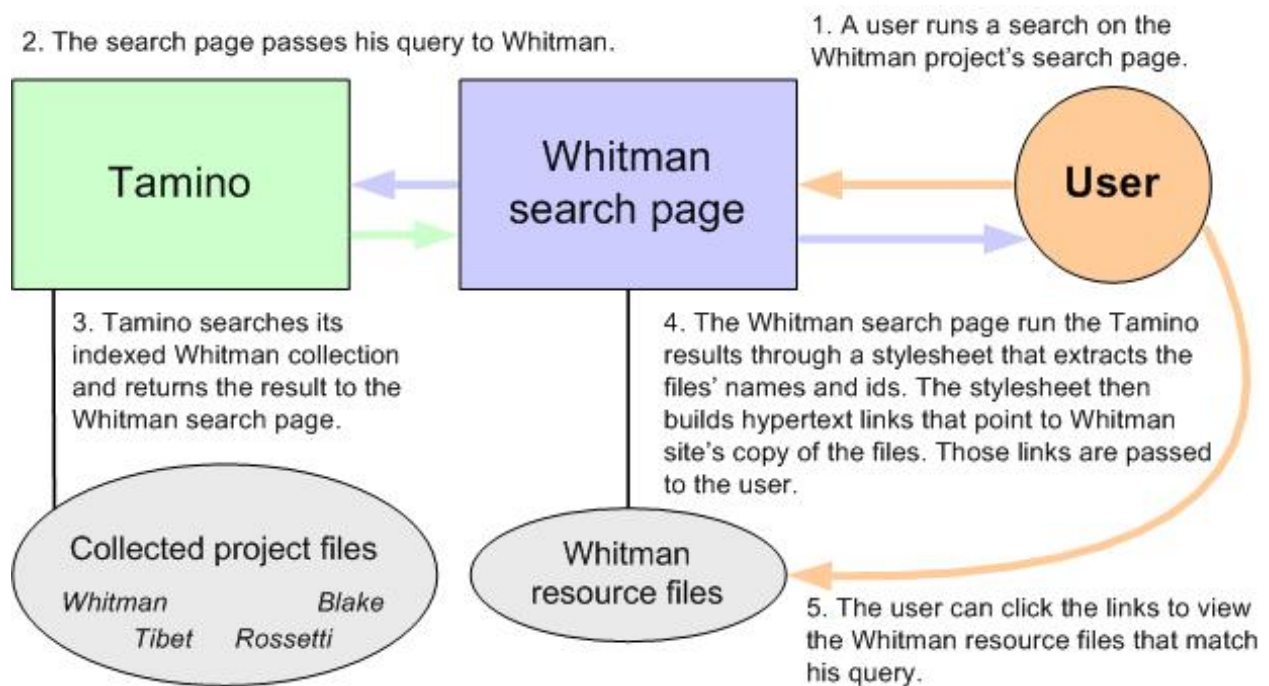


Figure 7. Tamino

We have four test projects for Tamino: **The Samantabhadra Collection of Nyingma Literature**, **The Walt Whitman Archive**, **The Complete Writings and Pictures of Dante Gabriel Rossetti**, and **The William Blake Archive**. Progress in each project is discussed below. Cindy Girard, an IATH IT specialist, and Robert Bingler, an IATH senior programmer/analyst, both received Tamino training and have been working with the Tamino projects.

4.5.1. Tibet

SDS funds helped pay for work done on David Germano's **The Samantabhadra Collection of Nyingma Literature** (<http://iris.lib.virginia.edu/tibet/collections/literature/nyingma.html>), a collection of Tibetan and Himalayan literature, over the summer and fall of 2002. Nathaniel Garson, the project manager, converted the project's DTD and bibliographic records from SGML to XML. This was done in part to migrate out of DynaWeb and in part to prepare files for Tamino. The process is described in detail on-line at <http://jefferson.village.virginia.edu/tibet/tech/sg2xm.html>, but a summary is below.

The project's SGML TIBBIL DTD was converted with Pizza Chef (<http://www.tei-c.org/pizza.html>), a free on-line tool provided by the TEI Consortium for generating XML DTDs to the XML xtibbil DTD. Some tweaking was necessary to accomodate some features in XML, but on the whole it was a successful conversion.

The SGML bibliographic records based on the original TIBBIBL DTD had been imported into the SGML database Astoria, which interfaces with the SGML editor Epic. Astoria served as a safe-repository and a version control mechanism. An Astoria script exported the files.

The exported files needed to be renamed in line with their Astoria-assigned ID (e.g., Tb.381.bib.xml). A perl script handled this task and also commented out local character-entity references used for diacritics. Such character-entities were resolved in SGML via catalog files, which are not allowed in XML, so the script created an XML catalog file listing all titles embedded in the appropriate space within the Doxography/Volume/Text hierarchy represented by <DIV1>/<DIV2>/<DIV3> tags hierarchy. It also created a batch file for running James Clark's SX program on each of the renamed files and created a file list in each volume's folder for uploading those files to Tamino. SX converted the files to XML and modified them through XSLT transformations. Approximately 1600 records were retrieved and converted.

The files could then be uploaded to Tamino. Tamino requires a database and a collection folder for each new project before loading any files. Each collection in Tamino is associated with an XML Schema, and Tamino's schema editor (below) has an import function that automatically converts a DTD into a schema.

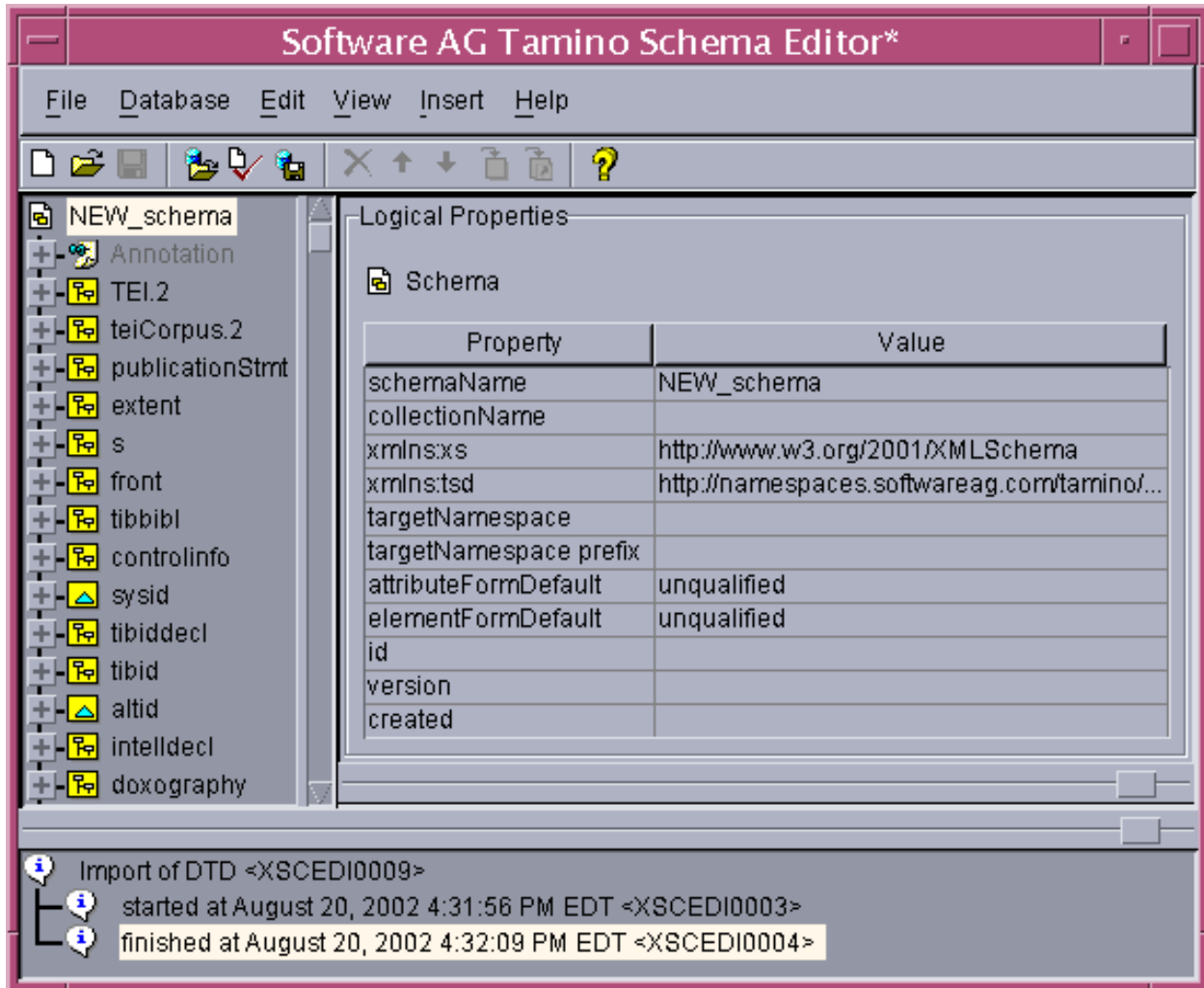


Figure 8. Tamino Schema Editor

Each collection has to have a schema defined for it (figure 8), even though the same schema is used across different collections. Different schemas within a collection can be based on the same DTD but use a different element for their root, when defined.

Once a schema has been defined for a collection, the documents can be uploaded to a Tamino database/collection. This is done in one of two ways, either through the Tamino Interactive Interface (TII) (shown in figure 9), which accessed from the web, or through the TaminoWebLoader perl script. TII provides access to the database through a browser window.

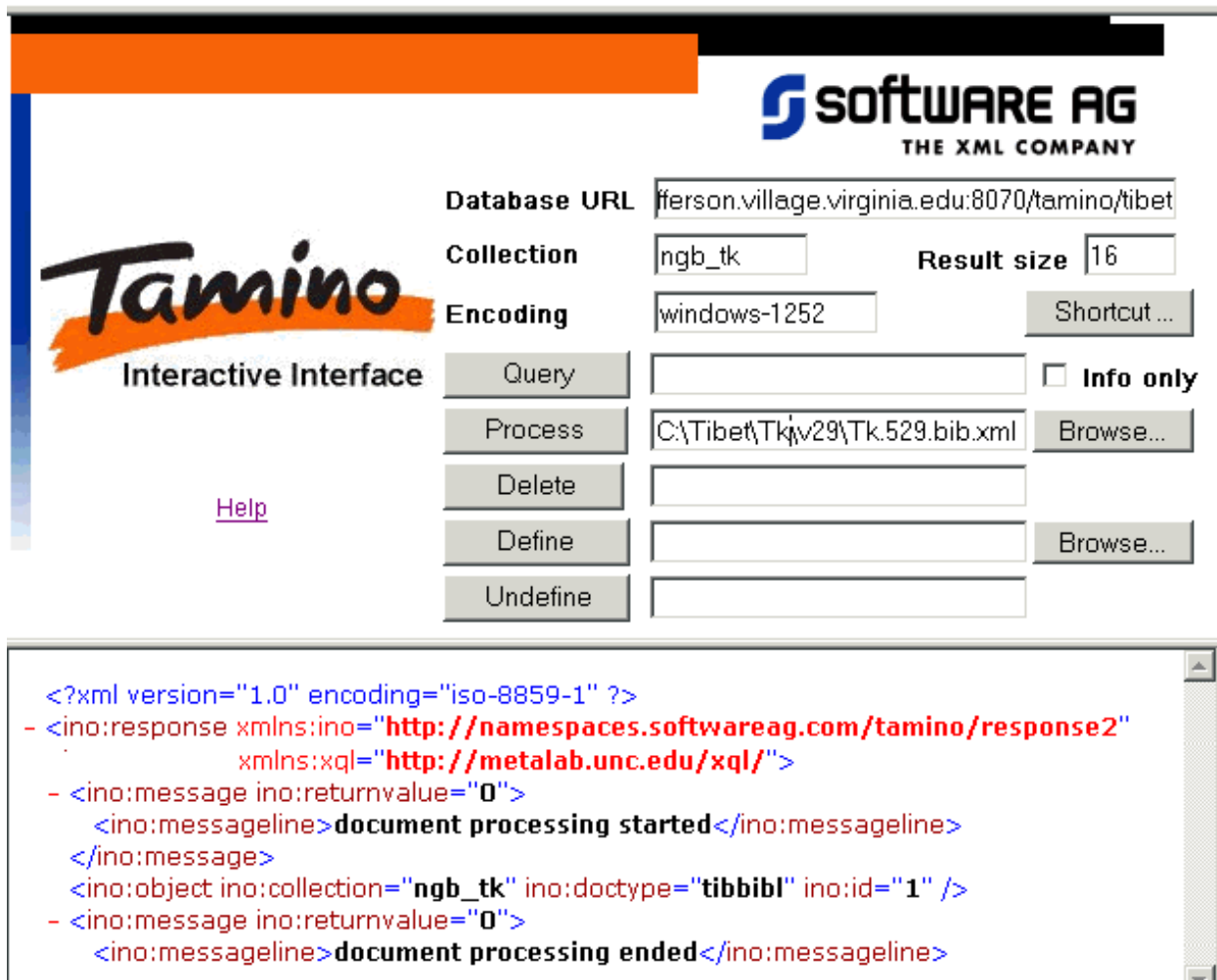


Figure 9. Tamino Interactive Interface

The other way to upload documents to Tamino is in a batch through a perl script provided by Ross Wayland. This was the method used for the Tb edition. Another batch program runs the TaminoWebLoader separately for each volume of the Tb.

The project then needed new style sheets which would mimic the DynaWeb navigational structure. These are finished, with the exception of the search function (which is waiting for the next Tamino release). An example of an XML recreation of the project's catalog with the emulated navigational structure can be seen on-line at [http://dl.lib.virginia.edu/servlet/SaxonServlet?source=http://jefferson.village.virginia.edu:8070/tamino/tibet/ngb?_xql=TEI.2\[@id='Tb.ed'\]&style=http://jefferson.village.virginia.edu/tibet/styles/tamino/tb_full.xsl](http://dl.lib.virginia.edu/servlet/SaxonServlet?source=http://jefferson.village.virginia.edu:8070/tamino/tibet/ngb?_xql=TEI.2[@id='Tb.ed']&style=http://jefferson.village.virginia.edu/tibet/styles/tamino/tb_full.xsl).

4.5.2. Whitman

Cindy Girard used **The Walt Whitman Archive** (<http://www.iath.virginia.edu/whitman/>) as the first test case for moving projects into Tamino. The Whitman project already had an XML DTD and the files were already in XML, so she could start working on developing a schema in the Tamino Schema Editor. She was then able to import the

files into Tamino, first using the Tamino Interactive Interface to test a few files and then using the perl batch script to move everything. The files have been indexed and there is a working test search page that uses Tamino. The test search page includes a basic keyword search and an advanced search using keywords within specific elements. The live version of the project is not yet using Tamino as a search engine, however, since at this point the project does not have enough resource files for there to be a noticeable difference between indexed and non-indexed searches. The project will soon be importing a large resource file that should have a measureable impact on indexed versus non-indexed searches.

4.5.3. Rossetti

The Complete Writings and Pictures of Dante Gabriel Rossetti (<http://jefferson.village.virginia.edu/rossetti/>) used SDS money to prepare, convert, and move files into Tamino. All character entity references were converted to Unicode. The site already had a search page, which had been using a DynaWeb search engine. The search queries therefore were translated into X-Query conformant queries via a style sheet (step 2 in figure 7). Rossetti has an SGML DTD but did not convert it to XML. Tamino uses XML schemas, so the Tamino Schema Editor was used to define a schema for the Rossetti collection, as with the Tibet project. The TaminoWebLoader perl script was then used to load 8,000 files into Tamino.

The Rossetti project is very large, so it hopes Tamino will increase the speed of its search function. However, Rossetti's schema uses a great deal of recursion and Tamino (as discussed above) can not currently handle many levels of recursion. For the moment, Rossetti cannot be indexed by Tamino. The Tamino developers are hoping to solve this problem with in the next product release.

The project has also been working on developing and refining a search engine interface that points to Tamino's search facilities, displaying Tamino's results as links to the project's resource files, and rendering those files.

4.5.4. Blake

The William Blake Archive (<http://www.blakearchive.org/>) will use Tamino as its search engine. The Blake DTD was converted from SGML to XML last summer and the resource files are currently being converted from SGML to XML, with SX and perl scripts. Character entities will be converted to Unicode entities. The project will use an external METS file for image pointers.

Technical committee report

The Technical Committee has been studying the Pompeii project's information structure, as discussed in section 4.4 in this report. The committee has been investigating technical issues associated with transforming digital materials (whether born-digital or converted) into library resources, and an important aspect of this transformation is making a project's resources into a coherent whole. A project's metadata is crucial to identifying and tracking those resources, so the committee has been discussing the issues involved in generating and organizing metadata. There is no generic solution, since while there are some types of information that all projects will produce, each one will have a particular set of information that it must store and track. An archeology

project, for example, might be primarily interested in mapping information against a physical location while a literary or historical project might need to track information against a timeline or intellectual construct. We have tools that allow a great deal of flexibility and creativity in working with metadata, but we are working on recommendations that can help authors develop practical and robust information structures.

Another issue that the committee has been working on is developing a method for members of a library community to create individual collections of references to digital objects in the library's repository. Essentially, we want to find a way for users to "bookmark" parts of collections. This is not a straightforward problem, since digital objects may move or be replaced if a project releases a new edition or changes the released version and not all users will have access to all objects. Worthy Martin, Rob Cordaro, and Chris Jessee have been developing a prototype Digital Object Collection Application (DOCA), which is intended to build individual object collections for University of Virginia users. Figure 10 shows a simple view of how the user might start a DOCA session.

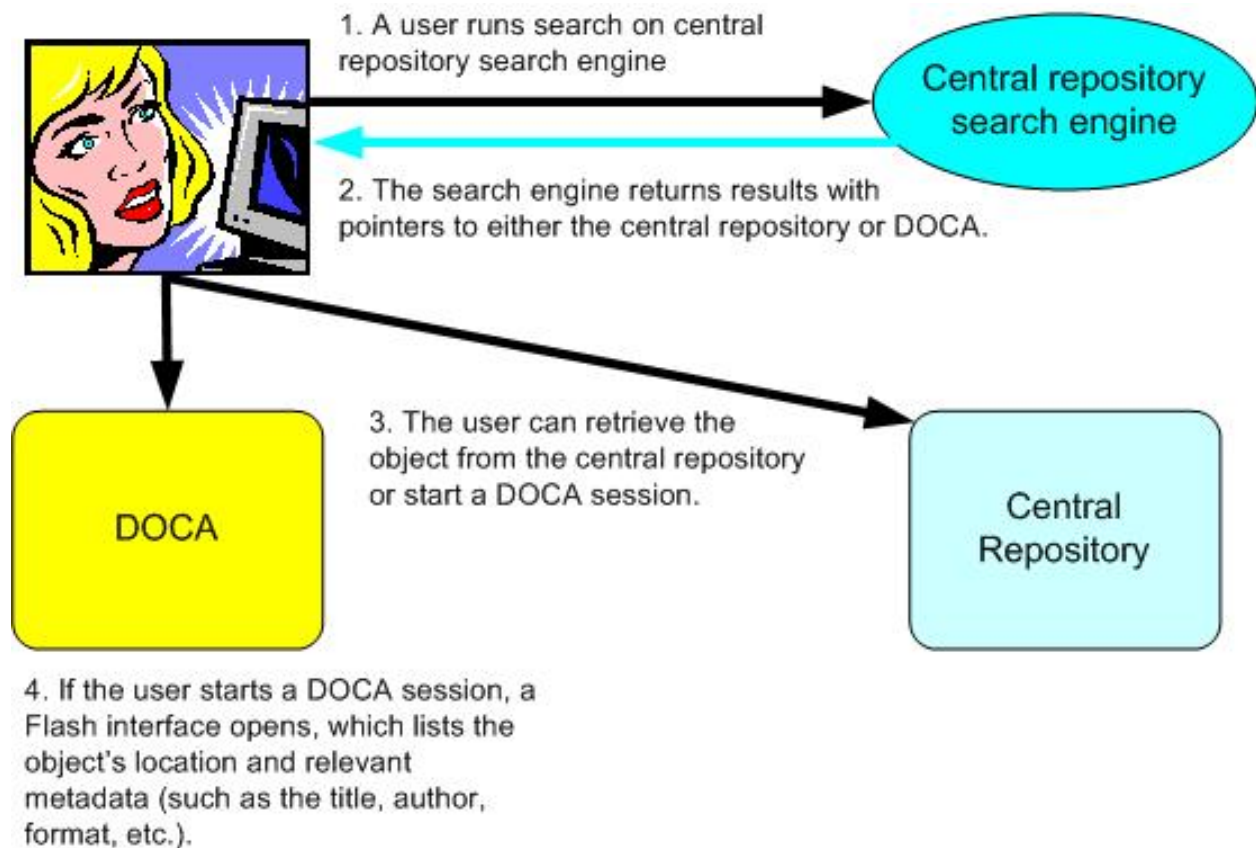


Figure 10. DOCA session

The user runs a search on the library's central repository search engine and gets a list of hits. The search engine offers two pointers for each hit: one directly to an object in the central repository and one to DOCA. If the user clicks the DOCA option for an object, a DOCA session starts up. The user will probably be required to have a user id

and password for DOCA, so she may need to log-on in order to start a session.

The DOCA window will display the object's title and various other pieces of metadata (such as the object's format and creation date). The user will be able to examine the object with an appropriate application (such as iNote) and make notes about it in the DOCA window. She can also run more searches in the central repository and add more objects to her DOCA session. When the user finishes the session, her list of objects, annotations, and personal identification information will be stored in the DOCA database. The next time she starts a DOCA session, she will be able to see the results of her previous searches. Note that the DOCA database will not store object files, but the objects' persistent identifiers.

DOCA still has several large concerns that we have not addressed in detail. It is not clear whether or not users outside the library or university community would be able to use DOCA: the repository may limit access to some objects to privileged users and some projects may limit access to their objects. It is also not clear how DOCA would handle objects that aren't located in the repository. These and other issues will be considered in the current year of the SDS project.

Policy Committee

The Policy Committee met regularly through much of 2002. The committee decided to follow the RLG and OCLC's recommendations for using the Open Archival Information System (OAIS) recommendations (<http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>) as a basis for a conceptual framework of an archival system that can preserve and maintain access to digital information over the long term. The committee has also drawn on "Trusted Digital Repositories: Attributes and Responsibilities," (<http://www.rlg.org/longterm/repositories.pdf>) published by the RLG and OCLC.

The committee decided to use OAIS's categories of submission, archiving, and dissemination for organizing subcommittees that would focus on those categories and how they impact digital library policy. The subcommittees were instructed to identify issues and concerns in these categories and consider possible solutions (and problems that could arise from these solutions). The results are currently being pulled together into a single document that discusses general policy recommendations and issues and recommendations for selection, submission, and collection; archiving, control, maintenance, and preservation; and discovery, delivery, and dissemination. A rough draft of this document can be viewed on-line at http://jefferson.village.virginia.edu/~spw4s/SDS/SDS_policy.new.frame.html .

Observations

7.1. Communication

Among the issues that have risen in the past year, perhaps the most important is the need for clear lines of communication between project authors, research centers (such as IATH), and library or repository staff involved in transferring project resources to the library repository. This is necessary at all stages of the project's development, so that the project staff do not build an uncollectible or unpreservable project. This is especially relevant if the work will be in actively development after collection, since the collecting

repository will expect future versions or editions of the project to mesh easily with previous published versions. If there are multiple parties involved, there should be a collective discussion of key questions, such as what will be collected for deposit, what will constitute a working copy versus an authoritative source for collection, who will maintain the authoritative source, and so on. If the library plans to collect a project that is the co-creation of the author(s) and a research center, the parties should also discuss issues related to software, standards that will be used, server usage, credits, and file storage. These issues can compromise the project if they are not addressed.

7.2. Documentation

Along the same lines, the project must generate correct and detailed documentation about the project's formal structure, contents, and history. The documentation should describe all aspects of the project and should be given to the collecting library along with the project resources. This can dramatically improve a project's lifespan, since delivery mechanisms and user expectations change over time and future digital librarians may need to migrate today's projects to as yet unknown technologies. It will also help project staff generate technically consistent and correct data and metadata, as well as a historical record for institutional memory. It can also be an opportunity for scholarly authors to describe to their current and future peers what the project's academic and technical intentions were, how the project works (or doesn't), and what kind of problems had to be overcome.

One area of documentation that is worth further investigation is databases. There are established tools for documenting table structures, primary keys, and such, but there is no standard system for recording *why* a particular set of tables and keys were used. Syntax information (data types, style sheets, query languages, etc.) is easy to document, especially if the project uses community-based standards. Semantic information -- what the author intended -- cannot be automatically generated and requires time and energy from the project staff. This requires some work from the project staff, but it will benefit the project in the long run: if a repository staff member is asked in the year 2110 (or 2010) to repair or rebuild a database that was designed in 2003 and has documentation that explains why the project chose that particular database structure, the chances are much better that the new database will be faithful to the original design.

Future Work

2002 was the third year of the SDS project, but it received a one-year extension in 2002. For 2003, now the last year of the project, we have set several goals:

1. Prepare as many projects as possible for collection by the library's repository. The library will then collect the projects, place them in the repository, and actively disseminate them via the library's catalog.
2. Prepare final reports from the Technical and Policy Committees, as well as a procedural manual that outlines how an IATH project will migrate to the UVA library repository. The manual should detail the mechanisms for moving projects, who will be responsible for performing various tasks, and when these tasks will be performed.
3. Produce working METS packaging, which has been tested on several projects.

4. Produce documented, working examples of the GDMS editor.

Assuming that we can meet these goals, at the close of this project we will have useful guidelines as well as practical examples for digital libraries, scholars, and publishers. The GDMS editor and DOCA tool may also prove to be helpful tools.

GDMS code snippets

A-1. Salisbury

```
<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE gdms SYSTEM 'http://dl.lib.virginia.edu/bin/dtd/gdms/gdms.dtd'>

<gdms id="image_archive">
  <!-- Minimum Header -->
  <gdmshead>
    <gdmsid>
      <system>salisburyparent.xml</system>
    </gdmsid>
    <filedesc>
      <pubstmt>
        <title>The Salisbury Project</title>
        <agent type="creator" role="author">Marion Roberts</agent>
        <agent type="provider">University of Virginia</agent>
        <imprint>
          <date type="creation" era="ad">1997-2003</date>
        </imprint>
      </pubstmt>
    </filedesc>
    <revisiondesc>
      <change>
        <changedesc>Transferred existing web site content and added images of
          Town, Close, Old Sarum, and parish churches.</changedesc>
        <date type="revision" era="ad">7/02-2/03</date>
      </change>
    </revisiondesc>
  </gdmshead>
  <div id="s1" type="project" label="The Salisbury Project">
    <divdesc>
      <description label="">The Salisbury Project is the creation of Professor
        Marion Roberts, McIntire Department of Art,
        University of Virginia, Charlottesville, Virginia.
        The Project is an archive of color photographs
        designed for teachers, students and scholars to
        supplement visually books and articles published
        on the cathedral and town of Salisbury. The project
        consists of views of the exterior and interior of
        the cathedral, as well as of select buildings and
        sites in and around the town of Salisbury.
        Additional material includes a guide for teachers
        and students, related texts and essays, and an
        annotated bibliography.</description>
      <subject>Salisbury Cathedral</subject>
      <subject>Medieval art and architecture</subject>
      <title>The Salisbury Project</title>
    </divdesc>
  </div>
</gdms>
```

```

<agent role="Project Director" type="creator">Marion E. Roberts</agent>
<agent type="provider">University of Virginia</agent>
<agent role="funding"
  type="contributor">University of Virginia Teaching and
    Technology Initiative</agent>
<agent role="funding"
  type="contributor">University of Virginia Digital Image
    Center</agent>
<agent role="funding"
  type="contributor">McIntire Department of Art, University of
    Virginia</agent>
<agent role="funding"
  type="contributor">Dean of the College of Arts and
    Sciences, University of Virginia</agent>
<agent role="funding"
  type="contributor">Institute for Advanced Technology
    in the Humanities, University of Virginia</agent>
<agent role="technical support"
  type="contributor">Institute for Advanced
    Technology in the Humanities,
    Digital Image Center</agent>
</divdesc>
<div id="uva10262332983371" type="structure" label="the Cathedral">
<divdesc />
<div id="uva10298613011123" label="Exterior tour of Cathedral"
  type="virtual tour">
<resgrp label="tour home">
<res id="uva10401419660672" label="photograph" type="image">
  <agent role="Photographer" form="persname"
    type="creator">Roberts, Marion</agent>
  <mediatype type="image">
    <form>digital</form>
  </mediatype>
  <respgrp>
    <resptr actuate="onRequest"
      href="http://jefferson.village.virginia.edu/salisbury/
        plans/label01.jpg"
      inline="true" show="replace" type="simple"
      xlink="http://www.w3.org/1999/xlink" />
    </respgrp>
  </res>
<res id="uva10401419689423" label="photograph" type="image">
  <agent role="Photographer" form="persname"
    type="creator">Roberts, Marion</agent>
  <mediatype type="image">
    <form>digital</form>
  </mediatype>
  <respgrp>
    <resptr actuate="onRequest"
      href="http://jefferson.village.virginia.edu/salisbury/
        plans/label02.jpg"
      inline="true" show="replace" type="simple"
      xlink="http://www.w3.org/1999/xlink" />
    </respgrp>
  </res>
<res id="uva10401419710044" label="photograph" type="image">
  <agent role="Photographer" form="persname"

```

```

    type="creator">Roberts, Marion</agent>
<mediatype type="image">
  <form>digital</form>
</mediatype>
<respgrp>
  <resptr actuate="onRequest"
    href="http://jefferson.village.virginia.edu/salisbury/
      plans/label03.jpg"
    inline="true" show="replace" type="simple"
    xlink="http://www.w3.org/1999/xlink" />
  </resptrgrp>
</res>
[etc....]
</resgrp>
</div>
<div id="uva10298613261884" label="Interior tour of cathedral"
  type="virtual tour">
  <resgrp label="tour start">
    <res id="uva103100223007014" label="Photograph" type="image">
      <description>View to Trinity Chapel</description>
      <agent role="Photographer" form="persname"
        type="creator">Roberts, Marion</agent>
      <agent role="photographer" form="persname"
        type="creator">Sargent, Robert</agent>
      <mediatype type="image">
        <form>digital</form>
      </mediatype>
      <respgrp>
        <resptr actuate="onRequest"
          href="http://jefferson.village.virginia.edu/salisbury/
            interior/images/entire.jpg"
          inline="true" show="replace" type="simple"
          xlink="http://www.w3.org/1999/xlink" />
        </resptrgrp>
      </res>
      <res id="uva103100227269515" label="plan" type="image">
        <agent role="delineator" form="persname"
          type="creator">Roberts, Marion</agent>
        <agent role="photographer" form="persname"
          type="creator">Sargent, Robert</agent>
        <mediatype type="image">
          <form>digital</form>
        </mediatype>
        <respgrp>
          <resptr actuate="onReq"
            href="http://jefferson.village.virginia.edu/salisbury/
              interior/images/plan.gif"
            inline="true" show="replace" type="simple"
            xlink="http://www.w3.org/1999/xlink" />
          </resptrgrp>
        </res>
      </resgrp>
    </div>
    <div id="uva10267448971318" label="Tour of Computer Model"
      type="virtual tour">
    <div id="uva10312602412501" label="Understanding architectural space"

```

```

type="virtual tour">
<resgrp label="Terminology">
  <res id="uva10312606338285" label="Terminology" type="text">
    <rescon label="">
      <section>
        <head>Terminology</head>
        <list type="ordered">
          <item>Arcade: a range of arches carried on piers or
            columns.</item>
          <item>Buttress: a mass of masonry projecting from or
            built against a wall to give additional strength,
            usually to counteract the lateral thrust of an arch,
            roof or vault.</item>
          <item>Clerestory: the upper stage of the main walls of a
            church above the aisle roofs pierced by
            windows.</item>
          <item>Cross rib: the diagonal arch of a ribbed
            vault.</item>
          <item>Lancet: a slender, pointed arched window.</item>
          <item>Nave: the western limb of a church, west of the
            crossing.</item>
          <item>Pier: a solid masonry support as distinct from a
            column.</item>
          <item>Plinth: the projecting base of a wall or column
            pedestal.</item>
          <item>Quadrant: the quarter of a circle. (Quadrant Arch:
            Flying Buttress Hidden from view beneath side aisle
            roof.) ed.</item>
          <item>Queen posts: a pair of vertical timbers placed
            symmetrically on a tie-beam.</item>
          <item>Rib vault: a framework of diagonal arched
            ribs.</item>
          <item>Set-off: a step-back or set-back in the masonry of
            buttress.</item>
          <item>Shaft: slender column between base and
            capital.</item>
          <item>Spandrel: the surface between two arches in an
            arcade.</item>
          <item>Springer: the point at which an arch springs from
            its supports.</item>
          <item>Tie beam: horizontal, transverse timber.</item>
          <item>Triforium: an arcaded middle level, above the main
            arcade and below the clerestory. In French buildings
            it contains a passage.</item>
          <item>Truss: a number of timbers braced together to bridge
            a space.</item>
        </list>
        <p>Definitions taken from John Fleming, et al, The Penguin
          Dictionary of Architecture, 4th ed., London, 1991.</p>
      </section>
    </rescon>
  </res>
  <res id="uva103126207012514" label="model" type="image">
    <agent role="delineator" form="persname"
      type="creator">Jessee, Chris</agent>
    <mediatype type="image">
      <form>digital</form>

```

```

</mediatype>
<respgrp>
  <resptr actuate="onRequest"
    href="http://jefferson.village.virginia.edu/salisbury/
      model/space/image/start.jpg"
    inline="true" show="replace" type="simple"
    xlink="http://www.w3.org/1999/xlink" />
  </respgrp>
</res>
<res id="uva103126211182815" label="model" type="image">
  <agent role="delineator" form="persname"
    type="creator">Jessee, Chris</agent>
  <mediatype type="image">
    <form>digital</form>
  </mediatype>
  <respgrp>
    <resptr actuate="onRequest"
      href="http://jefferson.village.virginia.edu/salisbury/
        model/space/full/image/start.jpg"
      inline="true" show="replace" type="simple"
      xlink="http://www.w3.org/1999/xlink" />
    </respgrp>
  </res>
[etc....]
</resgrp>
</div>
<div id="uva10312602653432" label="Nave construction stages"
  type="virtual tour">
  <resgrp label="Plan">
    <res id="uva103126339059350" label="Plan" type="text">
      <rescon>
        <section>
          <p>The plan of the cathedral is laid out on the ground with
            stakes and ropes. Foundation trenches are dug under the
            outer walls, buttresses, and pier bases, and courses of
            cut limestone and rubble fill are laid in the
            trenches.</p>
        </section>
      </rescon>
    </res>
    <res id="uva103126388934358" label="model" type="image">
      <agent role="delineator" form="persname"
        type="creator">Jessee, Chris</agent>
      <mediatype type="image">
        <form>digital</form>
      </mediatype>
      <respgrp>
        <resptr actuate="onRequest"
          href="http://jefferson.village.virginia.edu/salisbury/
            model/nave/image/start.jpg"
          inline="true" show="replace" type="simple"
          xlink="http://www.w3.org/1999/xlink" />
        </respgrp>
      </res>
    <res id="uva103126394656262" label="model" type="image">
      <agent role="delineator" form="persname"
        type="creator">Jessee, Chris</agent>
    </res>
  </resgrp>
</div>

```

```

<mediatype type="image">
  <form>digital</form>
</mediatype>
<respgrp>
  <resptr actuate="onRequest"
    href="http://jefferson.village.virginia.edu/salisbury/
      model/nave/full/image/start.jpg"
    inline="true" show="replace" type="simple"
    xlink="http://www.w3.org/1999/xlink" />
  </resptr>
</respgrp>
</res>
<res id="uva103126396765663" label="model" type="image">
  <agent role="delineator" form="persname"
    type="creator">Jessee, Chris</agent>
  <mediatype type="image">
    <form>digital</form>
  </mediatype>
  <respgrp>
    <resptr actuate="onRequest"
      href="http://jefferson.village.virginia.edu/salisbury/
        model/nave/preview/start.jpg"
      inline="true" show="replace" type="simple"
      xlink="http://www.w3.org/1999/xlink" />
    </resptr>
  </respgrp>
</res>
</resgrp>
<resgrp label="Outer Walls">
  <res id="uva10312639973451" label="Outer Walls" type="text">
    <rescon>
      <section>
        <p>Outer walls and buttresses and the inner piers are built
          up. Walls consist of two courses of stone with rubble
          filling the space between.</p>
      </section>
    </rescon>
  </res>
  [etc....]
</resgrp>
</div>

```

A-2. Pompeii

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE gdms SYSTEM 'file:lib/gdms.dtd'>
<gdms version="" id="">
  <gdmshead label="Pompeii Forum Project - PFP">
    <gdmsid>
      <system>URN://</system>
    </gdmsid>
    <filedesc>
      <pubstmt>
        <title>new-project</title>
      </pubstmt>
    </filedesc>
  </gdmshead>
</gdms>

```

```

</gdmshead>
<cat id="uva10445760376050" label="image catalog">
  <cathead label="PFP site images">
    <projectdesc label="PFP">This catalog is the base collection
      of the image files for the
      PFP.</projectdesc>
  </cathead>
  <res id="uva10374803092401" label="pfp-image-c-2001-6-10" type="Image">
    <description id="" type="caption" lang="Color"
      label="After SU 2">After SU 2</description>
    <description label="from east"
      type="orientation">from east</description>
    <mediatype id="" label="jpg" lang="Landscape" type="image">
      <form label="full" lang="37kb">600X400</form>
      <form label="thumb" lang="10kb">100X67</form>
    </mediatype>
    <source id="" type="Color Slide Kodak 35 mm. Kodachrome 64, PCD-2079"
      label="photograph">
      <identifier label="photograph"
        type="photograph">PFP-2001, roll 6, frame 10</identifier>
      <agent role="Photographer" type="creator"
        label="photographer">John J. Dobbins</agent>
      <time id="" label="photograph" type="full-thumb">
        <date era="ad" label="capture" type="capture">30 May 2001</date>
      </time>
      <physdesc label="film" type="medium">Kodachrome 64</physdesc>
      <physdesc label="size" type="extent">35mm</physdesc>
    </source>
    <source label="file" type="digital">
      <identifier label="cd" type="cd">PCD-2079-10</identifier>
      <time label="file" type="file">
        <date era="ad" type="full-thumb"
          label="digitize">2 April 2002</date>
        <date era="ad" type="full" label="modify">2 April 2002</date>
        <date certainty="" era="ad" label="modify"
          type="thumb">02 April 2002</date>
      </time>
    </source>
    <respstrgrp>
      <respstr actuate="onRequest"
        href="file:/C:/My Documents/Pompeii/
          Pompeii Forum Project/pompeii forum project/
          c-2001/c-2001-trench1/c-2001-6-10.jpg"
        inline="true" show="replace" type="simple"
        xlink="http://www.w3.org/1999/xlink"
        role="full" title="full" />
      <respstr actuate="onRequest"
        href="file:/C:/My Documents/Pompeii/
          Pompeii Forum Project/pompeii forum project/
          c-2001/c-2001-trench1/c-2001-6-10t.jpg"
        inline="true" show="replace" type="simple"
        xlink="http://www.w3.org/1999/xlink"
        role="thumb" title="thumb" />
    </respstrgrp>
  </res>
</cat>
<cat id="uva10445764385821" label="observation catalog">

```

```
<res id="uva10445764766672" label="pfp-obs-001">
  <rescon label="obs-001">
    <section label="alignment">
      <p label="" ">Standing to the west of wall 13 and near
        building 5 one can see that wall 13 aligns with
        column 15.</p>
    </section>
  </rescon>
</res>
</cat>
<div id="uva10216689717404"
  label="POS - physically oriented structure"
  type="division">
  <div id="uva10346246147200" label="Observation Catalog"
    type="Catalog" />
  <div id="uva10346246534401" label="Physical Model" />
  <div id="uva10445768698623" label="building 5" type="structure" />
  <div id="uva10445768923644" label="column 15" type="feature" />
  <div id="uva10445769187125" label="building 3" type="structure">
    <div id="uva10445769378506" label="wall 13"
      type="component structure" />
  </div>
</div>
</div>
</gdms>
```