Emory University

General Libraries
540 Asbury Circle
Emory University
Atlanta, GA 30322

Phone 404 727 2204
Fax 404 727 0827

Cyberinfrastructure for Scholars Interim Project Report

Mid-point Report on Project Creating a Sustainable Cyberinfrastructure Program for Scholars based on Focused Interdisciplinary Subject Domain Portals

April 2007

TABLE OF CONTENTS

1.0 Executive Summary and Highlights

The Cyberinfrastructure for Scholars Project, now at its mid-point, has a range of accomplishments to report to the Andrew W. Mellon Foundation. The project is systematically exploring the question of how to create sustainable inter-institutional scholarly portal services. The project Emory University wishes to thank the Mellon Foundation for its generous support of this important work. The following are highlights of the project findings:

- Several alpha systems for the proposed scholarly portal SouthComb have been created and are now being evaluated by small groups of Southern Studies scholars. The latest version of the portal is available at the following URL: *[Insert test system URL]*

- A group of eight subject-domain scholars have been convened as a consulting committee to provide systematic advice concerning the utility of the SouthComb portal for Southern Studies research and teaching purposes. This committee includes an endowed chair, a university provost, and the winner of a Pulitzer Prize. This group has been invaluable in conceptualizing the services to be offered by the scholarly portal.

- Focus groups of faculty, students, and librarians have been convened at the University of Mississippi, the University of South Carolina, and Emory University to gain in depth understanding and feedback regarding the proposed services of SouthComb. These focus groups provided an enormous range of ideas, useful guidance, and advice as we have refined the plans for the SouthComb portal. We have further engaged these focus groups in an ongoing discussion of how to improve networked access to humanities collections in Southern Studies.

- A great deal of programming work has been completed in the first year of project work. These activities have led to both working systems and formalized models for the underlying system components.

- A marketing and web design firm was engaged to work with us on this project. This has led to a compelling and attractive user interface for SouthComb and refinements of the service descriptions.

- A national directory of Southern Studies programs, scholars, publications, and conferences has been created in the course of project market research work. Such a directory has not previously existed and is now searchable in the SouthComb portal.

- To advance the project investigation into sustainability issues, a symposium entitled Sustaining Digital Libraries was held on October 6, 2006. This symposium was well attended, and featured papers presented on a large range of topics related to sustaining digital libraries. These papers are currently being edited and will be released later this year as a specialized monograph on the topic.

2.0 Review of Goals and Challenges Addressed

The Cyberinfrastructure for Scholars project is the culmination of several years of investigation at Emory University into a range of questions concerning the creation of scholarly portal services and systems. This prior research informed and set the stage for the project. The following is a brief review of the goals and challenges addressed by this project, consolidated here for convenience.

2.1 Challenges Addressed

[Intro / recap]

- ***Challenges in Searching Across Heterogeneous Information Realms:*** [Summarize]
- ***Challenges in Engaging Scholars in Innovation:*** [Summarize]
- ***Challenges in Engaging Cultural Heritage Institutions:*** [Summarize]
- ***Challenges in Managing Production Systems:*** [Summarize]
- ***Challenges in Creating a Sustainable Program:*** [Summarize]

2.2 Three Project Goals

[Intro / recap]

A. Build a Sustainable Combined Search Portal Service: [Summarize]

B. Improve Networked Access to Humanities Collections in the South: [Summarize]

C. Explore Sustainable Models for the Advancement of Scholarly Cyberinfrastructure: [Summarize]

3.0 Building a Sustainable Portal System

[Need opening statement]

3.1 SouthComb Planning Overview

SouthComb is a set of tools that allows scholars, researchers, and students to better research, access and understand Southern Studies information from a variety of sources that include the worldwide web, library catalogs, digital archives and user submitted content. The information that is indexed by subject category and genre include web links, metadata records, images, audio/video clips, and maps. The first priority for SouthComb is providing a compelling, high-quality set of services. These services include

- SouthComb Search and Browse
- Southern Studies Directories
- SouthComb Atlas – an interactive mapping service
- Teaching Resources and Media Library
- Today's South news service

Examples of common SouthComb tasks include:

- Search for Southern Studies information by keyword, subject, geography, or time period
- Create a profile and save your search queries for future use
- Display, manipulate, and combine geographic information
- Publish your teaching resources, images, audio/video clips and maps online so others can find them
- Find a Southern Studies graduate program or colleague

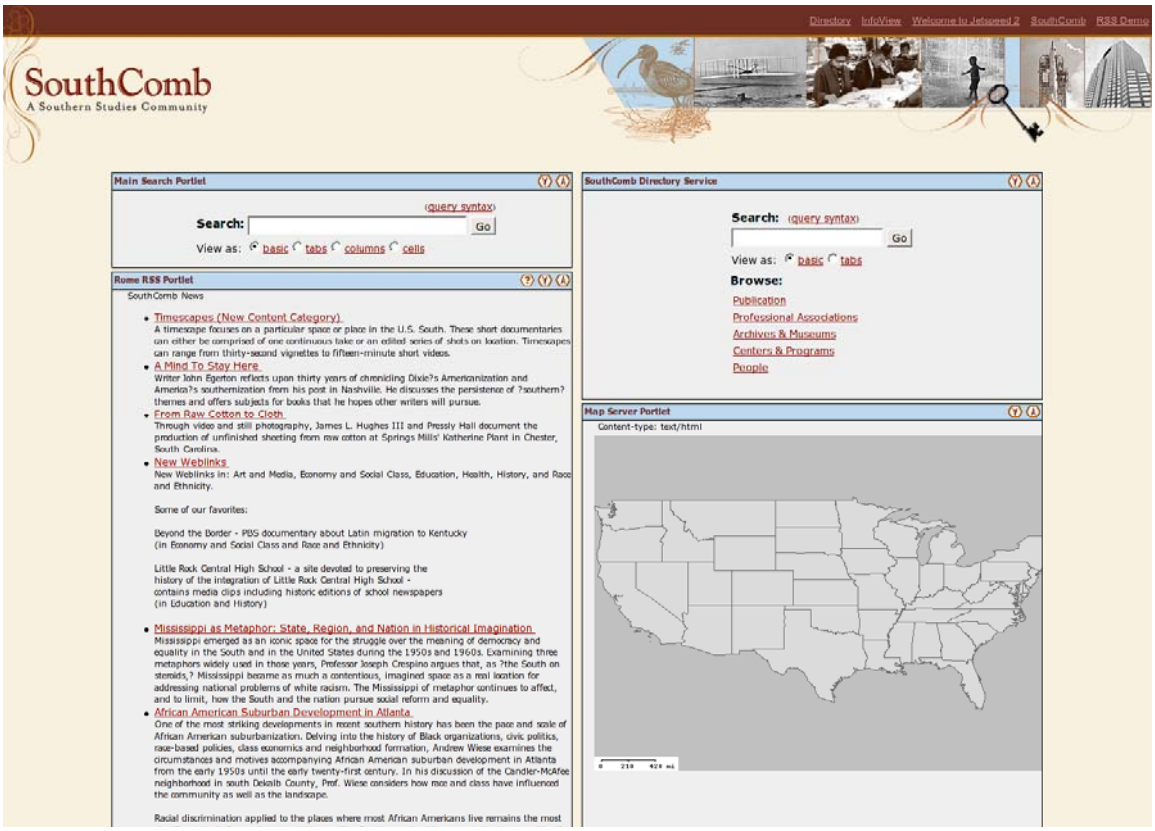
The first release of SouthComb is planned for August 2007. This release will include a complete set of services ready for usability testing and bug fixing.

3.2 Technical Overview

We have made considerable progress in exploring portal technologies for the central aspects of the SouthComb system, as well as “portlets” for the individual SouthComb services. We have working (albeit early) implementations of many of the basic services at this point.

The first prototype system is implemented in the Apache Jetspeed portal environment, with all services extant as JSR-168 portals (though their core implementations may vary in framework). Below is a screenshot of this SouthComb development system, including all the portlets developed to date (except the nascent user contribution portlet), visible on the screen:

Figure 3.2.1 SouthComb development system in JSR-168 portal environment.



Note that high-quality overall styling is in place. However, we have determined that the above “bag of portlets” view should not be the default, public-facing view of SouthComb, thus, we are now sketching out the high-level navigation and presentation of the site's content and services (discussed in the planning section). We intend, however, to make such a “direct-to-portlet” view and functionality available to advanced users who may want to skip directly to their favorite service portlets.

Following the initial prototype, we decided to explore alternatives to the JSR-168 standard, due to issues discussed later. After some exploration, we chose the Ruby on Rails rapid development framework. Reusing much of the original functionality, we were able to quickly deploy a new prototype using Ruby on Rails. Displayed below is the new homepage of the prototype.

Figure 3.2.2 SouthComb homepage prototype developed with Ruby on Rails.

3.2.1 Technologies and Techniques

SouthComb is building on numerous technologies and techniques developed and investigated in previous MetaScholar projects:

- **OAI aggregation (MetaArchive, AmericanSouth)** – These projects began our experiments to test the feasibility of building useful "meta" collections and libraries based on extant, disparate materials, harvested and aggregated using the Open Archives paradigm. The projects were a successul, but also a beginning, indicating that further work was needed to improve the quality of service of these "meta" offerings, due to the heterogeneity of the underlying records.
- **combined searching and browsing (MetaCombine)** – In this project we tested the feasibility of building combined searching and browsing services, using free, open source tools, which would integrate both library and web sources (and potentially others). Our test deployments and experiments showed that this could be done usefully and without too much difficulty, with various success in part related to the quality and heterogeneity of the underlying records and collections.
- **classification and semantic clustering (MetaCombine)** – Also on MetaCombine we pursued a variety of machine learning techniques to help better organize and integrate heterogeneous records. *Semantic clustering* is one such area, which "discovers latent topics" and helps to organize collections (even completely ad hoc collections) lacking any sort of a prior classification scheme or even ontology. (Text) Classification is another family of techniques, which take an

existing classification and records placed under it (the "training set") and uses the word occurrence data therein to place unclassified (or *unlabeled*) records into the same classification scheme. We tested and demonstrated both techniques on MetaCombine, and showed that services based on these methods could indeed be of use to digital library patrons.

- **focused crawling (MetaCombine)** – Focused crawling is a method which performs the "autonomous" discovery of relevant resources on the web. It is a modification of the traditional "web crawler" or "spider" system, which attempts to traverse hyperlinks *to only on-topic pages* rather than any page. This is achieved by integrating a text classifier with a web crawler, and classifying each page retrieved. If a page is judged as on-topic with a high-enough confidence by the classifier, its links will be crawled. What this arrangement achieves is a relatively high-degree of topic-specificity, as well as crawl efficiency (one does not need to traverse or store "the whole web"). We successfully developed an open source focused crawling system (based on freely-available component systems) and performed focused crawling to generate topically-relevant "web collections".
- **searching for scholars (Quality Metrics)** – In the Quality Metrics project, we tackled some special problems and issues in information retrieval for the scholarly setting. This setting entails not only library and digital library information, but also a more sophisticated and specialized user. The name of this project is due to our central objective of better exposing and utilizing measures of quality (both latent and explicit) that would be of interest and use to scholars, and which are often available in library/digital library collections. Along these lines, we developed a working prototype system, QMSearch (built on top of the Lucene open search search engine) which allows for the overlay of a variety of search *profiles* to change how results are presented and organized in a way that takes advantage of value-added information attributes.
- collection development/refinement (MetaCombine, OCKHAM) – On these projects we initiated the development of a framework for enhancing and refining collections in the form of Open Archives Repositories in a standard fashion. This was done in the interest of developing a federated framework of digital library tools and services that would be useful for building and maintaining DLs. It was desired in this work to help minimize the amount of ad hoc programming needed to develop, refine, and manage digital "meta" collections. The generalities of the model (termed "OXF") were worked out and some rough working prototypes were built, but more work is needed to provide a routinely useful framework, involving a good set of tools that will be useful in most digital library collections.

We are integrating all of the above tools and techniques into SouthComb. However, these do not in themselves provide all services and capabilities needed. So in addition, we plan to develop new capabilities in the following areas:

- **community feedback/contribution and moderation** – In the prior projects, we worked exclusively with "library-produced" content. In this project, we are building a real scholarly community (not just a database or library), so will need to support interaction from the user base. This includes record contribution, commenting, and other forms of feedback. Any time an information system has its input opened to users, especially in a context where administration and moderation resources are limited, care must be taken to "do things right." Put more directly, the goal is to find the "maximum" in value-added over top-level work done (and resources consumed) with the design of the community input system. There is no easy solution to this problem, and different arrangements work for different communities. However there are a number of precedents we will examine, and we will adapt them and possibly add new innovations for our situation.
- **"plugin" digital library/portal architecture** – Also key in making SouthComb sustainable is making the technical aspects of the system low-cost to deploy, maintain, and update. Thus we are putting more of our technical efforts into integration aspects of the portal than the specific services (many of which have been pioneered by us on the earlier projects mentioned above, or by others). Little exists in the way of a recombinant set of tools that can be applied to digital libraries and other social information systems. We hope to further the state of that art, thus meeting the goals of our project and helping others as well.

3.2.2 Open Source Software Evaluated & Utilized

We have evaluated a number of systems in the course of selecting technology for the implementation of SouthComb. Some of these systems, in their respective categories, were:

JSR-168 Portal Servers:

- **liferay** – Liferay was the first portal system we utilized, and we began building our prototype system in it. It is produced by a private company but released as open source software; support is available for a fee. The system had a number of attributes that seemed attractive to us, including the support for users and some groups, interface templating, and speed. However, when a bug necessitated an upgrade that did not go smoothly (with no back-port of the bug-fix available), we re-considered our choice of liferay. We decided that we could not go with a system that would make us a "slave" to the upgrade cycle, so we began moving forward with another system.
- **jBoss** – This was the second system we used, put out as a Java Community open source system. It was somewhat slower and less flexible and dependent on the Tomcat server, but seemed to us to be the next-best choice. However, soon after "porting" our old templates and portlets into this system, we discovered a bug that terminally prevented the portlets from working correctly. We reported this bug, but after about two months of no response from the development team, we gave up on jBoss, considering it a bad sign that the JSR-168 standard was not being fully-supported.
- **jetspeed** – This was the final JSR-168 compliant system we used, from the Apache Foundation. We determined its status as Apache software and large community of support would make it at least as suitable as liferay or jBoss, and found that it did fully support the JSR-168 spec, which meant our portlets worked properly. Unfortunately, it suffered from the same flexibility problems as the other JSR-168 portal servers.
- **uPortal** – Early on, we evaluated the uPortal system, but found that it was too inflexible in terms of templating to change the appearance and functionality.

  In general, the JSR-168 portal environments were found to be very heavyweight and inflexible. Although there was a promise of easy integration with portlet services, the details of the integration proved to be much more difficult than expected. View and layout formatting is difficult, as the portal server has its own preferred way of handling layout and display, which can be tricky to override. Inter-portlet communication is unsupported in the current JSR-168, although it is promised in the next release. However, depending on features yet to be implemented is a very risky proposition. In addition, many of the core JSR-168 features, such as moving, resizing, and generally customizing the portlets turned out to be undesirable for SouthComb from a graphical design standpoint.

  Finally, and most importantly, the JSR-168 environment is fragile from a configuration standpoint. Dozens of interrelated configuration options are necessary to successfully deploy the portal server and its associated portlets. If any of these options are not set correctly, the portal and portlets would not work as expected. Unfortunately, debugging these problems is very time-consuming, due to poor feedback from the portal environment.

Application Development Frameworks:

- **Zope** – Zope is a web application platform. We decided that Zope did not provide enough facilities to make it easy to build a portal system, but had the drawback of lots of bulk besides.
- **Ruby on Rails** – Ruby on Rails is language-centric platform for web development. As such, it is lightweight (unlike Zope, which is library-centric). We are currently experimenting with building a "from scratch" portal system in Ruby on Rails, and have a working prototype. The drawback is the de novo development; however the benefits are the lightweight nature, robustness due to no need for java/Tomcat, flexibility, and rapid development.

Storage/repository:

- **Fedora** – Fedora is a digital library object repository system with development centered at the University of Virginia. It is being used in the National Science Digital Library project (NSDL). Fedora handles metadata objects directly, as opposed to more granularly-decomposed data elements (as in a traditional relational database system, or RDBMS). This makes it convenient in the digital library context. Fedora allows for *behaviors* to be defined for these objects (such as *disseminate*), which forms the foundation for library services.

  Although Fedora is a compelling solution, we determined that developing the significant skillset required to use Fedora constituted a major risk to the project schedule. Fedora is comprised of several interwoven components, and each one has its own purpose, interface, and documentation. We determined that due to limited developer resources, it was too risky to base the entire project on an unknown tool. Furthermore, since Fedora would constitute the backend of the entire project, it would be difficult to make any headway in other areas while we struggled with the learning curve of Fedora.

- **Fez** – Fez is a Fedora-based repository system with a front-end for interacting with the collection (both as admins and users). We evaluated Fez, but found it too specific to the institutional repository scenario, and likely too difficult to customize into other scenarios.

- **mysql** – Mysql is a mainstay open source relational database management system (RDBMS). Following our decision to eliminate Fedora, we have moved forward using a standard RDBMS for the backend storage and persistence of SouthComb. Currently, we are using mysql, but we are also exploring PostgreSQL as a possible alternative, due to PostgreSQL's advanced features related to XML.

3.3 Development Detail

3.3.1 Architecture

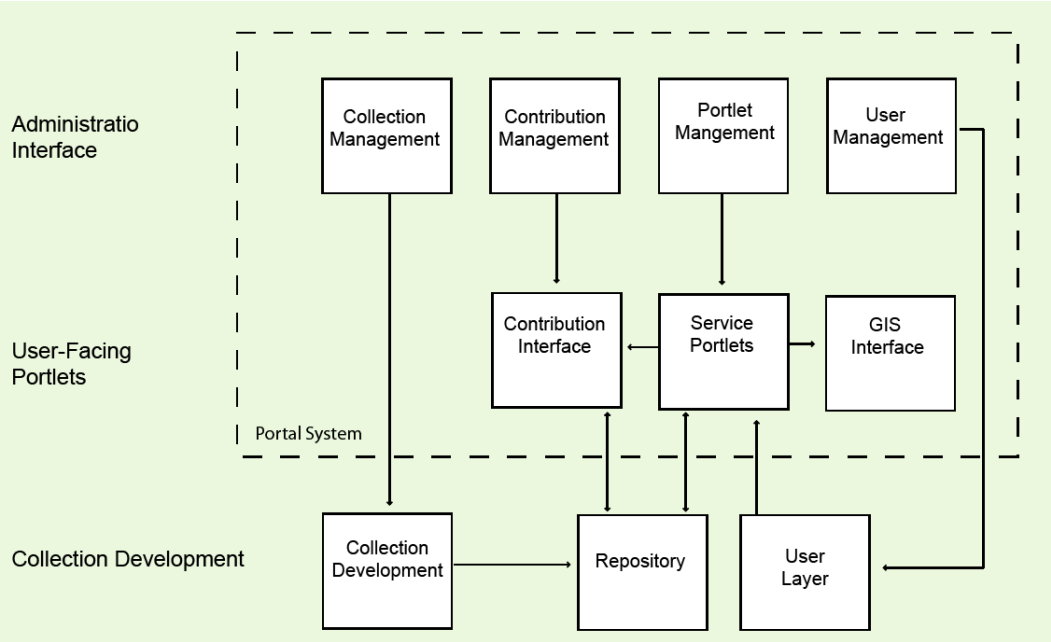
In this section we describe our current vision of the high-level architecture for the SouthComb system.

The system consists of the following main components:

1. repository system
2. user-facing service portlets
3. management portlets
4. user profile layer
5. collection development/management subsystem
6. the portal environment

The portal environment provides a platform to tie together most of the components. The management porlets and user-facing service portlets “plug in” to this central environment These components and their relationships are illustrated in the following figure:

Figure 3.3.1 SouthComb high-level architecture



.

In the following sections, we discuss the other subcomponents in more detail.

3.3.2 Repository System

The SouthComb repository system will serve as the central data store upon which the various other components will draw. The canonical, working version of the collection and other data objects (as discussed in the content model) will be stored here, even though other versions of the objects may exist in other forms in encapsulated data environments (such as search engine indices).

We have decided to utilize a relational database management system (RDBMS) as the data store for the repository objects. These systems are well understood and have been used for data storage for several decades. Utilizing an RDBMS will allow us to leverage the sizeable existing code base developed for RDBMS/SQL. In addition, many new rapid development technologies such as Ruby on Rails are tightly coupled to an RDBMS backend. Leveraging these tools will allow us to quickly deploy working software and iteratively improve it.

It is not without regret that we have chosen to utilize an RDBMS instead of Fedora. In terms of functionality, Fedora seems like an excellent fit for SouthComb. Heterogeneous object storage, dynamic object dissemination, and a host of other features are all very intriguing. Unfortunately, the SouthComb project is constrained by a paucity of developers, and none are proficient with Fedora. In contrast, the entire development team is very familiar with RDBMS/SQL. So, rather than devoting resources to an unknown and unproven tool, we have decided to take a safer route and work with a proven system.

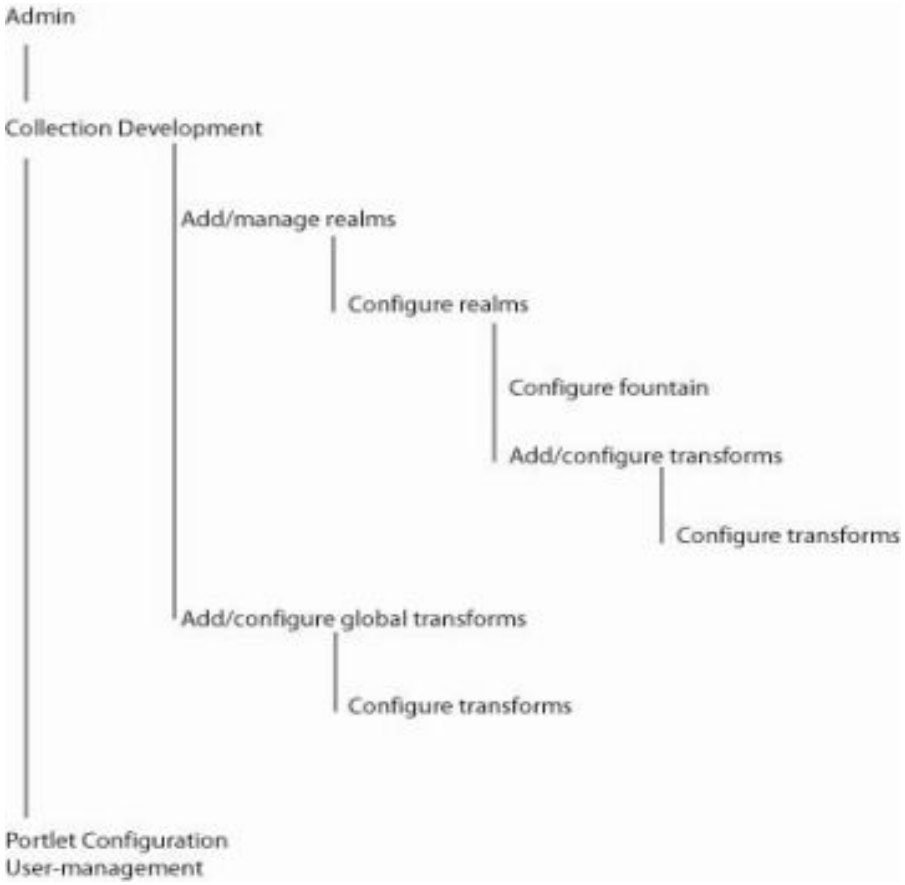
3.3.3 Management Portlets

The management portlets cover all aspects of librarian and administrator-specific interaction with the system. These portlets will be the entry point for:

- managing the collection (connecting with the collection development/management system)
- performing moderation
- managing users (connecting with the user layer)
- configuring portlets

The first two areas are the ones traditionally thought of as “librarian” responsibilities. These are where the “curation” is concentrated. Below is a figure detailing the collection management aspect of this management hierarchy:

Figure 2.1.1 SouthComb Collection management



We have begun extensive discussion and planning on how the moderation process will work. While submissions will not necessary be kept back until “blessed” by a moderator (a labor-intensive model), there will have to be some interface through which a moderator can extert influence by some appropriate conventions.

3.3.4 User Profile Layer

This part of the system will hold the information about user identities, profiles, and preferences, as well as other data specific to and saved by each user. This layer will then have “hooks” into various portlets so that their behavior can vary from user to user. The central portal system will also interact

considerably with this layer besides basic logins, for example, to provide full "breadcrumb" navigation functionality.

3.3.5 Collection Development/Management System

Considerable planning has gone into the collection development and management system. The design for this system continues prior work done on the MetaCombine project, where web services functionality was added to machine learning tools to allow them to function as modular Open Archivfes collection refinement "transformations". We call this model OXF, and describe below how it will be the core of the SouthComb collection preparation subsystem.

3.3.5.1 Introduction

Building on our work in the OCKHAM and MetaCombine projects, we propose to utilize OAI repository transforms as the building locks for a comprehensive, easy-to-use collection management system. Such a system will form the core back-end of the SouthComb site. The componentized, standardized model of the OAI transform framework will be key to lowering the management costs of the SouthComb digital library, thus allowing the overall project to sustain easier.

In the next section, we briefly explain this model. After this, we will sketch out how a collection development system could be built based on it.

3.3.5.2 OAI/OCKHAM Transformations Framework

The OAI/OCKHAM Transformations Framework (henceforth, OXF) has its conceptual underpinnings in three projects:

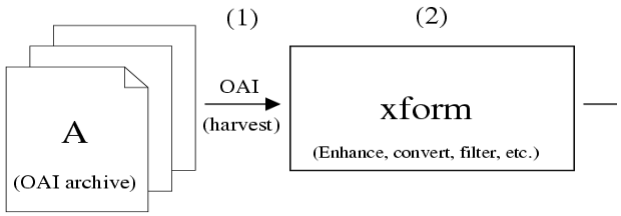
- **The Open Archives Initiative** - This project contributed the Open Archives Protocol for Metadata Harvesting (OAI-PMH) and the notion of a collection or repository as an Open Archive.
- **The OCKHAM Project** - This project (with Emory a participant) developed a generic framework for digital library services, emphasizing discovery of such services and a common language for describing their interfaces (Web Services/WSDL).
- **The MetaCombine project** - This project (at Emory) explored the notion of using machine learning methods to normalize and enhance harvested, heterogeneous collection metadata, as well as federated frameworks for doing this enhancement.

The federated framework efforts of MetaCombine became, for us, something of a capstone to the general development of OAI in the digital library world plus our work on OCKHAM. In our opinion, the way all these notions come together most naturally in the area of metadata improvement is as web services which perform automated metadata alternations on OAI repositories (with a WSDL interface, as in typical OCKHAM services). Such services can then, optionally, be advertised on an OCKHAM network, to be made available for serendipitous discovery by third parties (however, this last bit not necessary for our present purposes).

The atomic piece of this framework is the service which transforms an OAI archive. In detail, such a service inputs an existing OAI archive (via a base URL) and outputs an altered version of that archive at a new base URL. We call this an OAI/OCKHAM transformation service, or OXS.

This basic building block is sketched in the following figure. Such a model immediately allows a useful service to be shared with third parties. But perhaps more powerfully, by having standardized inputs and outputs, OXF allows *pipelined transformations* (this and more are detailed further in [1]).

Much like unix commands, pipelining allows value to be created by the ad hoc arrangement of simple building blocks that do not need to be re-developed each time (and in the case of web services, not necessarily even re-deployed each time).



- Illustration 1: Schematic of an OAI/OCKHAM transform service (OXS); the building block of OXF. Such a service inputs an OAI archive, updates its metadata in some way, then outputs a new archive with the result. Thus, an entire archive is treated like a fungible "collection object."

Since each point in the network (both inputs and outputs) are simply OAI repositories, there is automatic transparency in the workflow. Outputs can be queried in an ad hoc fashion, validated or browsed with tools such as the OAI Repository Explorer[1], copied with OAICopy[2], or harvested with any existing OAI tool. Third parties or third party software written with no knowledge of OXF can therefore still interface with an OXF pipeline.

In the next section we discuss how this model can form the basis of an easy-to-manage collection development system.

3.3.5.3 Collection Development With OXF

OXSes allow us to move towards large scale integration of collection streams by encapsulating the problem of particular kinds of metadata enhancement. This isn't to suggest that all metadata enhancement tasks are “solved,” easy, or even *possible* at very high levels of quality. However, OXF encourages us to break down enhancement tasks into separate, typically independent services, in a “divide and conquer” approach. Further, the beneficiary of the constituent OXS services does not need to know how they work, and can simply “drop in” improved versions of services when they become available.

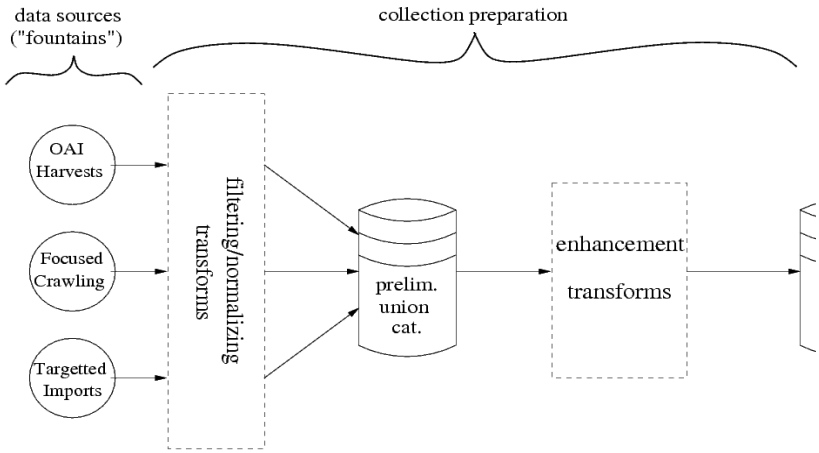
Once it is accepted that we will have OXSes to do the “heavy lifting” of preparing our metadata, we can envision a management system at a higher level of abstraction. Such a management system becomes accessible to normal digital librarians, who do not need to know the nitty-gritty of machine learning or arcane systems programming. This specialized knowledge (and the resources to fund it) remains a problem in the widespread adoption of metadata enhancement techniques that are *already proven*.[3]

1 http://re.cs.uct.ac.za/

2 http://metacombine.org/software/oaicopy-latest.tar.gz

3 Such as clustering, classification, date normalization, and many more.

In the following figure, we sketch the highest-level view of the collection management system founded on OXF. All of the arrows in this sketch represent "flowing" OAI repositories moving between portions of the system.



- Illustration 2: The overall collection management system. Data sources (which could have more instances than shown here) perform specialized work to fetch or create collections of metadata. These sources output OAI repositories. These repositories then feed into first-stage OXF enhancement pipelines, which are tailored to each data source. These separate streams then feed into a common, first-cut union catalog. Universal enhancement (also through OXF) is then performed on this catalog until a final union catalog is produced. This catalog is then suitable for building the desired end-user digital library services (each of which is, optimally, based on harvesting the final union catalog and exposing it in some custom way through a standardized web front-end).

The complexity hidden in this view is chiefly in the dashed boxes. These boxes represent OXF service pipelines, which would be constructed based on the particular data sources that had been fed into the system, and the needs of the end-user digital library services. The service pipelines in the first box are dedicated to the specific data sources, and go into producing a "first-cut" union repository which can be generally enhanced.

Some OXF services we think would be likely at the data source-specific stage would be:

- Subset filtering (selection).
- Editing set hierarchy.
- Metadata extraction (such as dates or places).
- Source-specific metadata normalization.
- Adding collection-level metadata.

Typical OXF services at the common stage, going into producing the final union catalog, might be:

- Clustering.
- Classification.
- Keyword/keyphrase extraction.

- Date normalization.
- Geocoding.
- Adding thumbnails.
- Deduping.
- Citation link analysis.
- Adding in full text (or portions of it).

Despite listing the above OXF service ideas, there are doubtless many more great ideas we (or others) have not even thought of yet. As long a such services can be cast as the production of a new OAI repository from an old one, they can be interfaced with this system with only a "thin" OXF wrapper.

3.3.5.4 OXF Collection Management "Plugins"

The above speaks only to the data infrastructure aspects of digital library collection management. However, even with the above OXF services available, there is much administrative (and even programming) work required to establish the necessary data sources and pipelines for the typical DL. A major goal of the SouthComb project is to "tidy up" all of this back-end work, which we can actually hope to do once it all becomes based on the OXF standard.

To provide a user-friendly management system, we need a couple new ingredients:

- A way to specify the "topology" of the network; which data sources there are, what transforms are to be applied to each of them, and what transforms are to be applied at the common level. This gives us the elements (nodes) and connections (edges) in the system's graph.
- A way to configure each component (both data sources and transformation services).
- A way to "execute" the system (and determine general parameters for how it is executed).
- Error reporting and debugging.

With these provisions, we can establish a plugin system that allows a non-programmer librarian to set up a data fountains-based digital library without touching a command prompt or writing a line of code, armed only with source targets (OAI base URLs and web site URLs) and subject-specific information.

As it turns out, the web services model already provides much of the necessary foundations:

- The inputs and outputs of a service are defined formally by an XML schema description (XSD).
- Execution and completion of a service are part of the standard workflow, with web services libraries already encapsulating all aspects of running.
- Errors are captured and returned in a standard fashion.

In fact, we propose that even the data source plugins can be handled through the web services methodology, even if they are to be hosted locally. This allows these services to automatically take advantage of all of the above standardized aspects. The only difference between a transform service and a service that *originates* data (in the form of an OAI repository) is that the latter has no repository as *input*; only as output.

3.3.5.5 Pulling it all Together

Even with this web services-plumbing, for SouthComb we must develop (1) the management front-end for such services, and (2) an interface for the overall assembly and composition of services and global configuration for the system.

For the first element---the task of configuring a particular system component---we can take advantage of the formal description of parameters to create a universal schema-driven configurator. In fact, we have already begun this work in the form of the "profile editor" for Quality Metrics. This web editor system inputs an XML schema and provides an editor to create or modify instances of that schema.

In QMSearch, these instances become input for each search to specify how the outputs are to be organized. However, there is no reason this notion cannot be extended to other instances of schema-driven XML instance editing, which in the case of OXF components we can apply to configuration. While we may not be able to use the QMSearch profile editor wholesale, we should make sure the current work on this is leveraged for the SouthComb back-end management system.

The second item---the general configuration interface---likely needs to be built from scratch. While an extremely advanced version of this could actually be based on a GUI with a drag-and-drop model for specifying and connecting components, intuitive equivalents could be created that use a more conventional tabular web display. Inside this interface would be callbacks to enter configuration for the individual components.

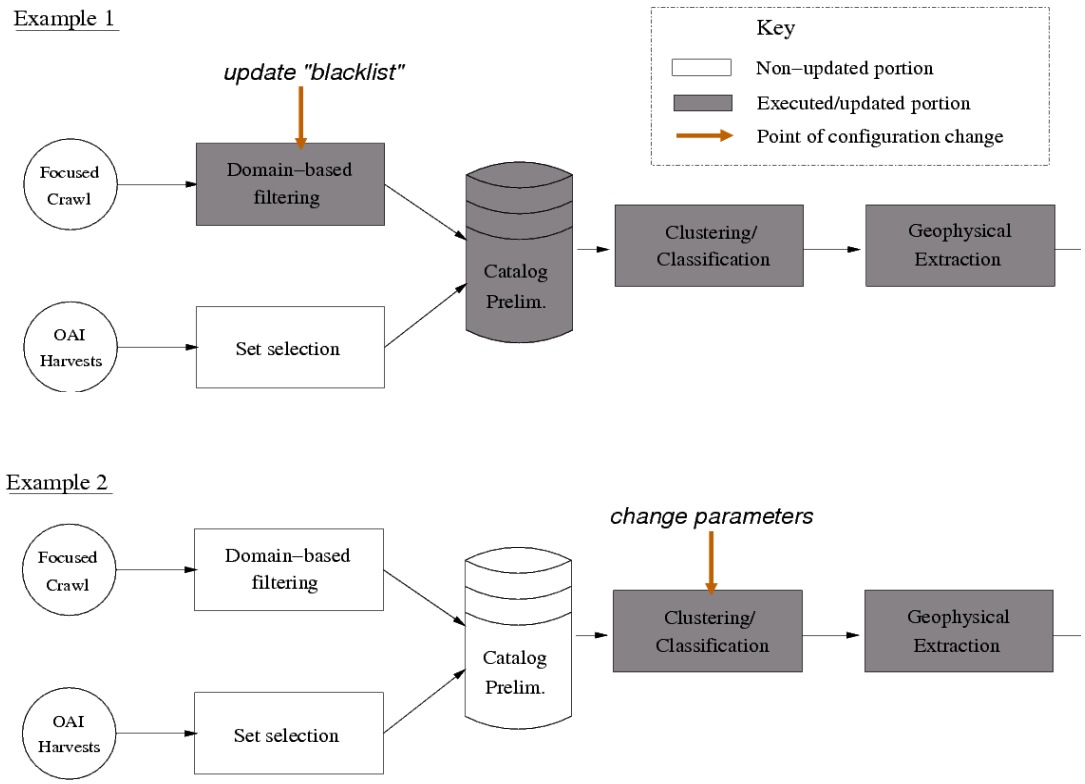
3.3.5.6 Iterative Refinement

The naturally iterative nature of collection refinement can be handled elegantly in this framework. The general workflow is:

- Configure all data sources and transforms that are part of the pipeline.
- Execute the pipeline.
- Examine the output (through the front-end services or "probing" tools, such as the Repository Explorer).
- If the output is deficient, alter the configuration of the responsible data source nodes or transform (proceed to step 2).

For example, if the digital librarian has a pipeline including a focused crawl record discovery and one or more refinement pipelines, he or she may find too many off-topic records are present in the output. In this case, the librarian would go back to the configuration of the focused crawling data source, perhaps altering the positive or negative data or keyword sets, and then would re-run the process.

This process could be conducted optimally by just re-running stages including and after modified stages. Such a "downstream" progressive updating scheme is illustrated in the examples in the figure below.



- Illustration 3: Iterative updates in the OXF collection management framework. OXF pipeline components can be reconfigured all or in part at any time, and the system can then be "re-run" in an update mode which only processes points "downstream" from the change. Here two examples are given, with two different hypothetical, simple OXF pipelines and corresponding refinement edits.

The justification for this workflow for iterative refinement is manyfold. Of course, it is helpful that this workflow eliminates the need for a separate refinement interface. However, more importantly, the plan relies upon viewing the output *in the context of the actual digital library services* in order to evaluate the level of success of the collection preparation. While it is possible to make estimates of collection and refinement quality by other (e.g., analytical) means, such estimations are always limited in accuracy.

For example, browsing a collection in the OAI Repository Explorer does not always give a full sense of how uniform a collection will look through the end-user interface; one must observe the success or failure of services to optimally present records that may be based on data which is in some way heterogeneous or otherwise unexpected.

As a consequence of this refinement methodology, production services should be deployed in concert with a development interface for evaluation of preliminary versions of the collection that have the benefit of a fully-featured---but not live---digital library interface. However, this requirement does not go beyond simply good systems development and deployment practice.

[1] Aaron Krowne, “A Draft Standard for an OCKHAM-based OAI Transformation Service Framework (OCKHAM-xform)”, August, 2005, http://br.endernet.org/~akrowne/ockham/oai_xform_spec/ockham-xform.pdf

3.4 Content Model

The *content model* is the schematic for the kinds of content in an information system, and the interrelationships of the various objects and categories of content to each other. Developing a content model, either explicitly or implicitly, is necessary to build any information system. The structure of the information in the system is the fundamental determinant kinds of information services available in that system: the content model can either be an enabler or hindrance to building the kinds of services which are necessary. Due to the scope and formal nature of SouthComb, we are beginning to explicitly develop our content model at this stage. Below we sketch out the SouthComb content model as an ontological hierarchy, reminiscent of an OOP class hierarchy. Also given at "leaf" nodes are key metadata and referential elements:

- message objects
  - MD: subject
  - MD: message body
  - MD: commentary or editorial
  - REF: other object (any type)
- quality judgment objects
  - ratings
    - MD: score
    - REF: other object
  - reviews
    - MD: review prose
    - REF: other object
  - recommendations
    - MD: positive or negative
    - REF: (target) other object
    - REF: linked field of study (e.g. relevant category or curricular element)
  - relevance link
    - REF: object A
    - REF: object B (inc. ontological object)
- collection ontology objects
  - categories
    - MD: certainty (confidence level – esp. for automatic classifications)
    - REFs: contained objects
  - subcollections
    - REFs: contained objects
  - curricular elements
    - REFs: contained/relevant objects
  - lists/sequences (by users, experts)
    - REFs: contained objects
- collection objects
  - records
    - MD: user-contributed?
      - MD: title
      - MD: contributor
    - MD: autonomously-discovered?
    - MD: moderator-approved?
    - MD: media type
    - MD: genre
    - MD/REF: subject
    - MD ...

A major point of debate will likely be which ontological aspects should be implicit or encoded only in "leaf" node metadata, and which should also be represented as fully-qualified ontology objects, complete with links to other objects in the system. Some such ontologies are (not limited to):

- language

- genre (end-use or applicability-centric)
- realm (where derived from – web, OAI/DL, library catalogy, contributions, etc.)
- subject area ("category")
- format/media type (image, movie, text, presentation, map ...)
- file type/format (word, pdf, jpeg, etc. ...)
- ... (?)

Here is an example of an object that might be a part of the SouthComb system, in terms of its attributes and relationships to other elements of the SouthComb content model:

User-contributed GIS map : Kitty Hawk/Wright Brothers Flights

- MD: title = Kitty Hawk/Wright Brothers Flights
- MD: user-contributed = yes
- MD: contributor = user1234
- MD: moderator-approved = yes (has been seen and verified)
- MD: format/media type = map
- MD: genre = archive resource, teaching resource
- MD: file type = jpeg
- MD: version = 1
- ...
- REF: categories = science & technology, history
- REF: review = { "A useful visual; great teaching resource!", Katherine Skinner }
- REF: rating = { 5/5, "great!", Aaron Krowne }
- REF: rating = { 4/5, "pretty good – needs better labelling", Stacey Martin }
- REF: message = { editorial, "Some labels are hard to read, I suggest ..." , Stacey Martin }
- ...

The content model will be translated from the high-level description into a combination of low-level implementation specifications. Such specifications will include XML schemas, database schemas, and vocabulary sets defining allowed values for the various metadata fields.

3.5 Portal Development

The *portal* is the central portion of the SouthComb system. It integrates all the various user-services, the data, and the management aspects of the system. The portal will be implemented as a Ruby on Rails application that allows easy integration of the portlets described below.

3.6 Portlet Development

Portlets are service components of a portal, typically user-facing (but not necessarily). Considerable development progress has been made on portlets for many of the core planned services of SouthComb. In the following subsections we give an overview of this activity.

3.6.1 Combined Search

At this stage we have a fully functional search service which searches over records derived from a variety of "realms," including:

- OAI (“native” digital library)
- Library (south-related record “mined” from catalogs)
- Web (gleaned from focused crawling)

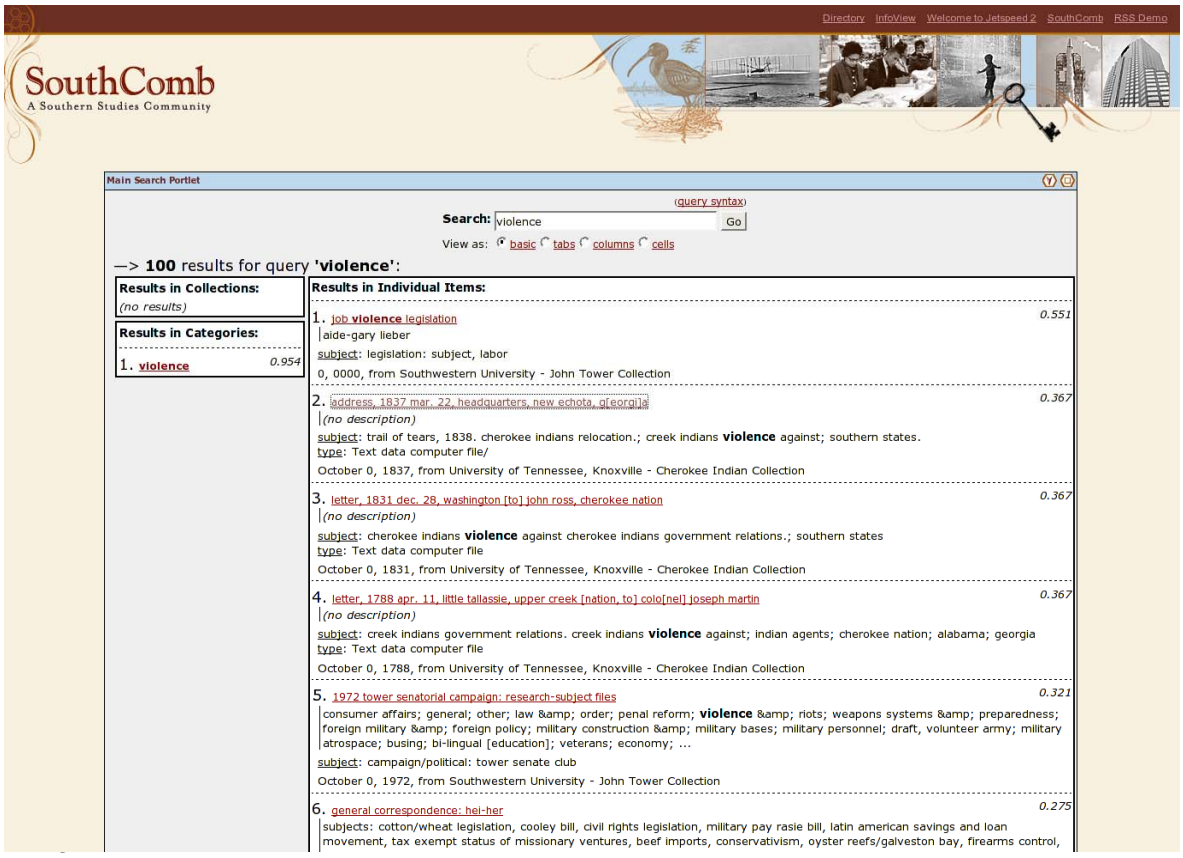
We did not develop this search service from scratch: we have continued and extended the QMSearch metasearch system we developed on the separate “Quality Metrics” project (funded by IMLS), which is in its final evaluation stages. We have successfully applied this system to SouthComb in its development capacity and anticipate the successful application of the system in the production environment (with concomitant polishing and tuning).

To support integration with the core SouthComb portal system, the Combined Search service is deployed as a stand-alone web service. Completely independent of the core SouthComb portal system, the Combined Search service listens for XML-encoded requests from clients, passed via the well-defined SOAP interface. XML-encoded responses are then returned to the client.

In this model, SouthComb is a front-end client for the Combined Search web service. When users enter query text, it is passed from the SouthComb portal system to the Combined Search web service. The XML returned from the service is then transformed into HTML and displayed to the user in their browser. From the user's perspective, it seems as though Combined Search is fully integrated with SouthComb. However, from a development standpoint, the web services model allows for reuse of the Combined Search functionality across multiple applications.

Below we provide a screen shot of this system in action:

Figure 3.6.1 SouthComb Search



the boxes on the left for supplementary results for categories and collections. These are provided by adding to the collection surrogate records representing objects of this sort, specially marked as the appropriate ontological type.

We are still tuning the other three profiles (tabbed, columns, and cells) for display in the portal environment.

Next steps for the combined search service include the further tuning and polishing of default search profiles, and more influentially, the development of a schema-driven editing system so users can create their own ad hoc search profiles. This latter provision will subsume the functionality of what is traditionally called “advanced search”; except much more powerful due to the novel capabilities of the QMSearch system. For example, users will be able construct arbitrary ranking functions by drawing on whichever underlying metadata and quality attributes they wish.

Ultimately, we plan for the combined search system to power much of the browsing and searching of the core digital library collection of SouthComb through the extensive use of selective *filters*. Filters allow narrowing results sets to arbitrary subsets of a collection. For example, a list of all syllabi in a the collection would simply be a null query with results restricted to records of type “syllabus”.

We will also explore more “agile” user interface provisions, such as showing less summary meta-information for each returned result, but providing some Javascript/AJAX controls to “open up” a result on the screen to get an intermediate level of detail. If the user is interested, they could then click on the record's title to get its full rendering, as is done currently.

3.6.2 Southern Studies Directory

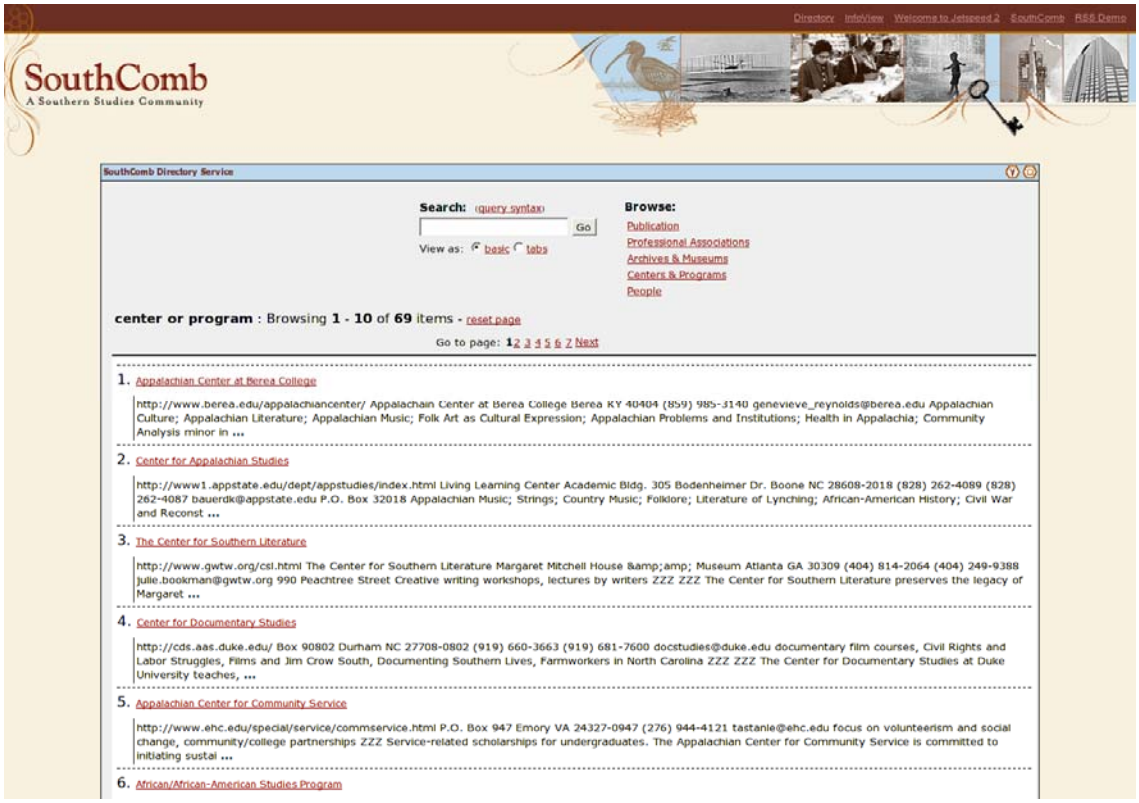
The Southern Studies Directory is one of the simplest, yet most value-added services of SouthComb. Simply put, this service is a browseable database of “entities” in Southern Studies (such as people, publications, and various sorts of institutions). No such specific database exists for the field; all of this information is scattered and ad hoc; and much of it is not online.

We have compiled a modestly-sized database of these records with manual effort already, and placed them in a number of spreadsheet files. These files will grow with time, and in the future we plan to deprecate them in favor of “online” editing of the database, with user community input. These will be described later. Immediately below, we discuss how the current incarnation of the service “1.0” works.

The current Directory service allows for three sub-services: browsing, searching, and viewing a record. Navigating directly between records can also be considered a service. The underlying data is stored in a relational database system (mysql). The search and browse back-end is QMSearch (a browse is simply a search restricted to a category, with a “null” query). There is a script to “ingest” the data from the entity database to the Directory QMSearch module, which has its own Lucene index.

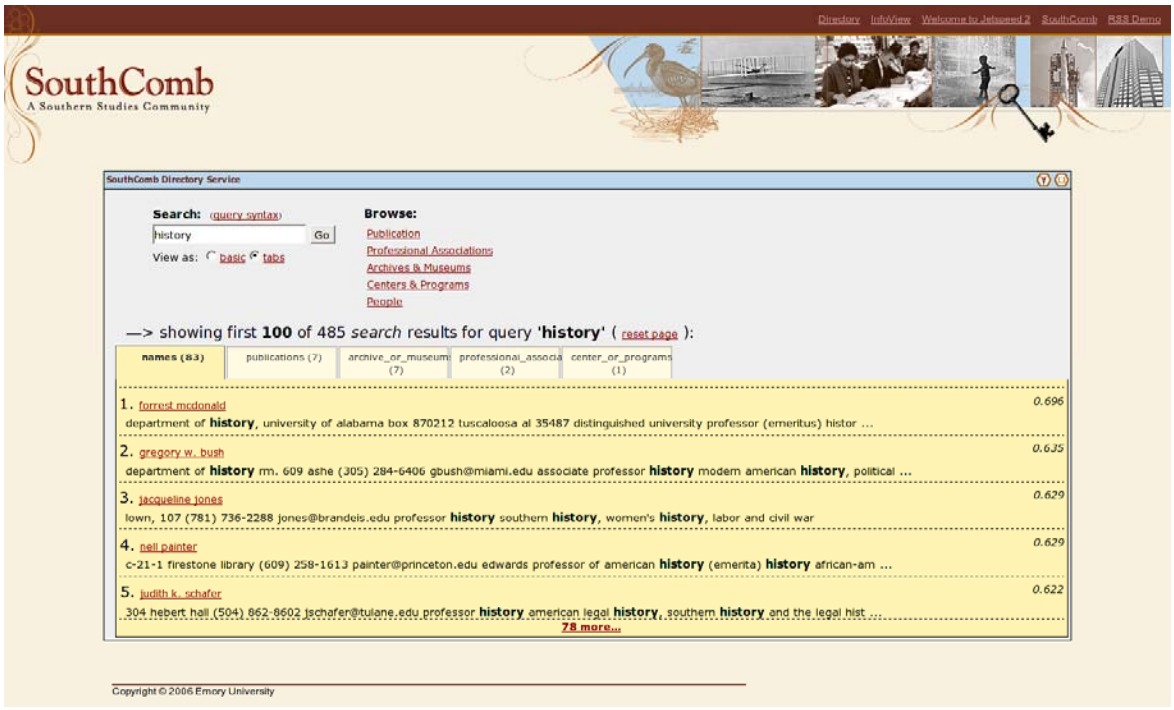
Below is a screenshot of a Directory browse in the current system, for “center or program” entities:

Figure 3.6.2 SouthComb Directory



Next we give a screenshot of a search in tabbed display mode, which separates results out by entity type:

Figure 3.6.2b SouthComb Directory



The search can also be switched to “basic” mode, which gives results in a linear list similar to the browse list shown previously.

As mentioned above, the input to the service is a set of spreadsheet files, encoded using a number of working conventions we have developed that provide data uniformity. These spreadsheets are input periodically into an ingest-and-translation script system (which happens to be written in PHP; though this is not a necessary characteristic of the model).

This script's main task is to take the “flat” data model of the spreadsheets and produce an entity database which encodes all of the otherwise latent inter-linkages between the constituent entities. It does this by running in two stages. The first stage creates an corresponding “flat” entities database (without inter-linkages). The second stage analyzes the relationship data provided in the spreadsheets for each entity (which is specially marked-up) and searches for the corresponding target entities in the database. If they exist, a link is created in a “links” table.

At record display time, each entity's metadata is displayed, as are any links to other entities. Links to other entities are displayed as hyperlinks, which allow navigation to the related target entity. Thus, the Southern Studies Directory is actually a navigable “semantic network” of entities in the professional field. Not only does such an advanced (yet intuitive) service not exist in Southern Studies, but we are have never seen one for any other field either!

Below we show the display of a “Professional Association” record in the current system. In the top box is the entity metadata; in the bottom one are links to related entities:

Figure 3.3 SouthComb display of individual record

Type	Professional Association
Title	Appalachian Studies Association
URL	http://www.appalachianstudies.org/
Address 1	Marshall University
Address 2	400 Hal Greer Blvd.
City	Huntington
State	WV
Zip Code	25755
Phone Number	(304) 696-2904
Email	mthomas@marshall.edu
Scope	general organization for Appalachian Studies

Publisher	Journal of Appalachian Studies
President	Chad Berry
Office Manager	Mary K. Thomas

Directory Home

Next we give the person entity record connected to the above as the second link (note that the second person is printed but not linked; this is because there is no entity record for them in the database):

Figure 2.4.5 SouthComb display if individual record connected as link

Type	Name
Title	Chad Berry
URL	http://www.berea.edu/appalachiancenter/people/chadberry.asp
Address 1	Bruce, Room 128, CPO 2166
Address 2	Berea College
City	Berea
State	KY
Zip Code	40404
Phone Number	(859) 985-3727
Email	chad_berry@berea.edu
Title	Goode Professor of Appalachian Studies
Department	Appalachian Studies
Additional Title	Associate Professor
Additional Department	History
Second Additional Title	Director
Second Additional Institutional Affiliation	Appalachian Center
Area Of Specialization	Appalachian history, Southern migrants, National Barn Dance""

President	Appalachian Studies Association

Directory Home

Note the the link back to the corresponding professional association visible in the link box above. Also note that each link has a link label which explains the type of relationship; in this case, “President”, which is clearly a person-role.

The above record printouts are very austere and meant just to provide a functional start. We are working on more production-caliber designs. Below we give a record mockup (based on real data from the Database) which has a more pleasant and readable appearance:

Figure 2.4.6 SouthComb Individual Record graphical mockup

Note the sections in the above:

- Title, type, and link are displayed first, in a standard way.
- The address and contact information is given second, and printed in a standard way.
- Entity-specific metadata, without special handling, will be printed as a list of fields and values under an "other information" heading.
- Links are printed last, but are clearly marked as links, and are separated out by the *type of the entity*. This adds additional information which is useful for comprehension. On top of the type is given the link label as before, which makes a little bit more sense as a facet underneath target entity type.

Future Versions

Currently, the Directory service is largely its own self-contained universe, complete with its own database and QMSearch module. However, the service is lacking many desirable features, such as searchability with the rest of the collection, hooks into user feedback, and user contribution. To provide these features eventually, it is clear the Southern Studies Directory needs to be integrated with the global SouthComb repository and content model.

This would not necessarily be too difficult to bring about. Using the OXF model, the "output" of the ingest script could, instead of a database, be an OAI repository of the directory records, with inferred links between records encoded as "Relation" metadata tags. These records could then be ingested into the SouthComb acquisition pipeline as OAI records from a "directory" realm, and furthermore

simply treated like normal records (selected or deselected via filtering when they are specifically-desired, or not desired).
In this model, user contributions would simply take the form of users creating a record of the appropriate metadata and "directory" realm type, for inclusion in the collection. This could then go into the usual moderation system, or "directory" records could be routed for special moderation. However, little is needed on the user-facing side of a Directory service other than browse and search functions that are narrowed to records of the "directory" realm type.

3.6.3 Today's South (RSS)

The "Today's South" service is a value-added "news feed" which will be focused on Southern studies-relevant content. While South-related mainstream news qualifies, this also includes third-party commentary, and even popular articles and features from "academic" sources as they are published.

RSS, which has no standardized definition but means something like "really simple syndication," is an XML format to communicate basic "news" items between information sources, in a client-server (or provisioner-consumer) model. More technically, RSS is a family of formats, and competes with another format called Atom, but these details have largely (and thankfully) been abstracted by major RSS-handling software libraries.

RSS provides for a set of standards and conventions to convey basic streams of chronologically-sequenced news items in the form of basic useful metadata. The semantics of these streams are generally "what's new," and can run the gamut from mainstream news to announcements of new records in libraries.

We are employing this standard and the surrounding paradigm for SouthComb to provide a unique news-like service which does not exist anywhere else: an academic-tilted Southern Studies news stream. Most mainstream news outlets are not specifically-geared towards the South, nor do they do a good job of separating out South-related articles. Further, of digital sources that are Southern-specific, most do not provide data feeds in any sort of format, let alone an RSS variant. Thus, such a service is not even remotely available at present, so we believe "Today's South" will be a very low-cost, yet high-value-added service for our patrons and members.

This leaves the question of how we are to solve the problem of aggregating South-specific content. Our solution is two-pronged:

1. Search for South-specific RSS feeds and directly integrate them.
2. Find general news (or other relevant resource) feeds which includes items about the South on a regular basis, and extract just those items for inclusion in "Today's South".

For #1, we have already undertaken a few rounds of search and found a handful of high-quality, dedicated South-specific resources. For example, there are a few blogs that focus on the South in valid interest areas such as justice and rights, law, culture, and the environment. These are prime candidates for direct inclusion. We have also begun to push other sources (especially those to which we have professional ties) to begin providing RSS streams to their content.

Part #2 of our approach to the service is to filter out non-South-related content from general news feeds. For this, we are building on text classification technologies and methods utilized in the MetaCombine project, and have constructed a working version of the service.

The filtering approach works as follows:

- RSS streams are "ingested" by our program, polled on a regular basis (every few hours). The news items are aggregated in a local database.
  - Items from South-only sources are marked as "relevant" in the database.
  - Items which are from mixed sources are stored as "unknown".

- A classifier is run on metadata extracted from each recent, uncategorized item. The training set for this classifier is the same as the one used on MetaCombine to perform classification tasks (including focused crawling) and includes records from the American South collection, articles from the Encyclopedia of Southern Culture, and various other harvested relevant items.
- Items that pass a certain threshold are considered “South-related” and marked as such. The other items are marked as “unrelated.”
- The final-cut South-related items are passed on to a generated, finalized RSS stream.
- This finalized RSS stream is read in by the “Today's South” portlet service and the records are displayed to users.

We plan to expose “Today's South” a variety of ways:

- As a special floating box on the front page to SouthComb.
- As a direct-navigation (full-screen) service, with more records and more details (as well as browsing of past news items).
- As a link from a navigation “sidebar.”

An early version of this service, displaying South-specific records, is shown in the all-portlets screenshot at the beginning of this section.

Future Versions

Despite the sophistication of the above, we consider this version of the service only the first of three natural phases.

A second phase would replace the RSS-specific handling with a bridge layer that draws in RSS and creates an ad hoc OAI repository with records corresponding to the ingested RSS records. These records would be given persistent, unique identifiers, with provenance information recorded as well.

This “synthetic” repository (which would serve just as an initial aggregator) would then be pulled into the general SouthComb acquisition stream as a new “realm” -- “news,” in this case; parallel with “web” and “OAI” -- and put through classification and other sorts of normalization filters (as described in the section on the OXF framework and collection management).

After this point, the records would be aggregated with all SouthComb records, though with a distinct “realm” tag, and processed with global refinements (such as date normalization, geotagging and geocoding). Then all records would be placed in the final-cut SouthComb repository (“the union catalog”), and made available for services.

This more sophisticated but standardized architecture would allow “Today's South” records to be treated as normal collection records, without re-producing the necessary development in the context of the news service. For example, we would like users to be able to search among past “Today's South” items, browse for them in whatever subject hierarchies are available, comment on them, help to moderate them, and so forth. If we are to provide these capabilities, we cannot afford to do so by recording them for “Today's South,” and potentially numerous other services.

3.6.4 GIS/Mapping

GIS and mapping in SouthComb actually breaks down into three distinct kinds of services:

Finding maps which have already been created and saved in the system Creating new maps interactively Mapping for other kinds of objects in the collection

Finding extant maps is perhaps the simplest: these will simply be collection objects which have map images and potentially other sorts of map data files associated with them. They will be searchable and browseable with the rest of the collection, with appropriate metadata categorizing them as maps.

The interactive creation of new maps, however, is definitely a distinct service. There are many aspects to this service, including the interactive interface, the underlying data which is provided, and the "hooks" to allow users to share and contribute their creations back to the collection. Initially, we installed the open source MapServer system and a portlet that acted as a bridge between Mapserver and our Jetspeed portal. The output of this early implementation is visible in the previous figure of the entire portal system. Following the switch to Ruby, we reconfigured MapServer to provide data using interfaces defined by the Open Geospatial Consortium. Basing development on widely-implemented standards gives us the flexibility to choose from a variety of tools and increases the likelihood that our work can be reused in other projects. Considerable design and development remain to fully develop the mapping application and integrate it thoroughly within SouthComb.

Bringing GIS and mapping to other objects in the system is also a sophisticated and involved, yet important objective. We wish to apply GIS to SouthComb to the fullest extent possible. For example, search results, even non-map results which pertain to some specific geographical location should be easy to view within the context of a map. As another example, library locations for library catalog-derived records should be easy to view. Finally, almost all Southern Study directory objects pertain to some physical location, so these should certainly be browsable via GIS.

A major component of connecting the map functionality to the collection objects in the system will likely be geotagging and geocoding. We plan to construct OXF transforms to extract likely geographic labels from metadata records, then generate geocodes (longitude and latitude specifiers) based on these. This metadata will then be fed into a mapping API to allow any objects to be displayed in a variety of ways.

4.0 Improving Networked Access to Collections

4.1 Exposing high-quality collections

A major goal of this project is to reach out to humanities collections in the South that are not easily accessible via the web. Currently, the University of South Carolina will provide a liaison for routing relevant information and links to SouthComb that includes current and archival collections of Southern.

4.2 Collaborative Efforts

There are a range of mechanisms that SouthComb is engaging in collaborative efforts. First, Emory University has visited with interested faculty, students, and librarians at Emory University, the University of Mississippi in Oxford, MS and the University of South Carolina in Columbia, SC to hear about what research and teaching tools they currently use and to gather feedback on the SouthComb system.

4.2.1 Focus Groups

Members of the SouthComb team visited the University of Mississippi in Oxford, MS on August 29, 2006 to discuss potential services with librarians, faculty, and graduate students. A similar series of sessions was conducted at Emory University in Atlanta, GA on October 4 and 16, 2006 and March 29th, 2007 at the University of South Carolina, Columbia, SC. An overview of the project was enhanced with a visual presentation of SouthComb's site design. Participants shared current tools used for research and teaching and were encouraged to share ideas of tools that though could improve their own workflow.

Faculty and Students
Faculty and students alike expressed enthusiasm for the idea of a personalizable portal experience, including the option to save searches and search results and bookmark specific links. They saw value in two specific kinds of searches: serendipitous and object-specific. An object-specific search would be a more typical search procedure, allowing the user to narrow a search by media type, origin, source, etc. On the other hand, a serendipitous search would allow a user to reproduce the "next book on the shelf" experience of browsing a real library, in which he or she might discover an interesting and only somewhat related item near the original item of interest.

One suggestion that emerged from both faculty and student sessions was the idea of a media repository. While image repositories do exist online and some are open to the public for searching, there are a number of limitations to these available facilities. Media repositories do not often specialize in materials dealing with the South, for example. In addition, a number of respondents expressed concern about questions of copyright in the use of many of these materials.

It became clear that both faculty and students would greatly appreciate a reliable repository of media dealing with the South that was understood to be available for reasonable use, such as media under the Creative Commons copyright agreement or an agreement like to that used by Flikr and other similar sites. ["Creative Commons is a non-profit that offers an alternative to full copyright" (creativecommons.org). "For example, a user has the option to allow users to copy, distribute, display, and perform your copyrighted work -- and derivative works based upon it -- but only if given credit, or used for noncommercial purposes only. Users can copy, distribute, display, and perform only verbatim copies of your work, not derivative works based upon it or allow users to distribute derivative works only under a license identical to the license that governs your work." (flikr.com)] Such a repository would also provide added value if it allowed contributors to submit brief summaries and descriptions of their submissions, providing fruitful context for other SouthComb users.

Students

While students shared many concerns with faculty, they also expressed issues unique to themselves. Since so few institutions offer specifically Southern Studies programs, students of the South are often scattered throughout various departments and unable to connect. On the other hand, at institutions with specific programs for Southern Studies, students in those programs are often neglected by other departments when they might benefit from an interaction, thereby missing opportunities such as calls for papers or announcements of funding. In short, students universally expressed a desire to use SouthComb to facilitate networking and to share information. They also expressed enthusiasm for potential socialization services, such as discussions and commenting on extant objects.

Students also presented a unique set of concerns in relation to general searches of the SouthComb collection. They believed it would be best to be able to search across a wide variety of media in a single search and to have the results method be customizable by each user. A single search term could, therefore, return as results articles on the subject, syllabi mentioning the subject, directory information about professors studying the subject, and images, videos, or audio clips pertaining to the subject.

Similarly, students were interested in a geographical dimension to search results. If a result object was in a given student's institutional collection, that information ought to be noted, but if the object was *not* in the local collection, knowing where the nearest copy of it was would facilitate interlibrary loan or trips to visit a different collection.

Faculty

Like students, faculty had a few unique suggestions about SouthComb usage. Several faculty wanted to be able to preview materials which are not available online. By way of example, these faculty praised the indices, tables of contents, and sample pages made available for many books at Amazon.com.

While faculty were enthusiastic about the creation of a multimedia repository, they did feel that overcoming the collective action dilemma and soliciting contributions might be a problem. To some extent, they felt that this problem would be alleviated by the ability of contributors to annotate their contributions and to receive credit for their contributions permanently in the associated metadata. Although such contribution is not quite academic publishing as it is recognized in most institutions, the faculty did feel that scholars who are inclined to share their materials would be enticed by this extra dimension to the SouthComb repository.

Lastly, faculty agreed that they would enjoy using SouthComb as a means to share teaching tips, resources, and syllabi.

Librarians

At Emory University, the librarians were curious about the SouthComb atlas, as they say they receive many requests for assistance with the acquisition and construction of maps from students and faculty. They felt that a tool for creating, archiving, and sharing maps about the South would be extremely useful. They cautioned, however, that the technological expertise among many of their clients is less than optimal, and that the submission process in particular ought to be kept as simple as possible.

At both Emory and the University of Mississippi, librarians stressed the frequency with which they address "Info-literacy" issues. Instructors at both institutions struggle with explaining to their students how to do research and how to be critical of sources online, and often turn to library staff for help. SouthComb could be used to provide either a list of "vetted" sources on the South or a component introducing such info-literacy concepts, or both.

Librarians also expressed interest in finding aids across institutional collections, especially newspaper archives. A better system of archiving collections, such as the one SouthComb aims to provide, appealed to them greatly.

SEE APPENDIX for Focus Session Agenda

4.3 Additional Connections

Additionally, SouthComb's services are reliant are existing and emerging technologies built by Emory as well as other institutions. To expand SouthComb's services, Emory University hosted a forum to share information between the Middlebury College Semantic Indexing Project (www.knowledgesearch.org) and interested scholars and librarians. Both SouthComb and the Semantic Indexing Project are sponsored by the Andrew W. Mellon Foundation to conduct applied research into semantic indexing and clustering systems. In addition to sharing information, the meeting produced an opportunity for both parties to begin work together. Several options for collaboration were discussed and included interest in digital preservation and the clustering and classification of spatial and temporal information.

4.4 Subject Scholar Consultants (KS)

We are pleased to report that at the beginning of the project, all six of our MetaCombine Project Subject Scholar Consultants, Allen Tullos, Will Thomas, Charles Reagan Wilson, Tom Rankin, Barbara Ellen Smith, and Natasha Trethewey, unanimously elected to continue working with our project team as we moved to the SouthComb Project. Their continued participation has allowed the SouthComb project to quickly integrate the expert advice and suggestions of an unusually well-informed advisory board.

We are also pleased to report the addition of two accomplished Subject Scholar Consultants to our project team: Provost Earl Lewis of Emory University and Prof. Patricia Yaeger of University of Michigan. Both are highly regarded within the field of Southern Studies and beyond, and they bring to our group both a broadened subject focus and fresh perspectives on portal features and development.

The Subject Scholar Consultants on this project are as follows:

Dr. Allen Tullos is Associate Professor of American Studies at Emory University and Senior Editor of *Southern Spaces*. From 1982 until 2004 he was editor of the journal *Southern Changes*. He has worked as producer, co-producer, and sound recordist for numerous documentary films and has served as the website coordinator for americanroutes.com. Tullos has published numerous articles and book chapters on popular music, southern film and visual culture, the politics of space, and contemporary southern politics, including the Sydnor Award winning book *Habits of Industry*.

Dr. Charles Reagan-Wilson is Professor of History and Southern Studies and Director of the Center for the Study of Southern Culture at the University of Mississippi, where he has taught since 1981. He is the coeditor of the *Encyclopedia of Southern Culture*, author of several books, and general editor of a new book series, "New Directions in Southern Studies," published by the University of North Carolina Press. He has directed numerous symposia on topics ranging from the Caribbean and the South to Religion and the American Civil War.

Dr. William G. Thomas, III is the John and Catherine Angle Professor in the Humanities for the Department of History at the University of Nebraska-Lincoln. He formerly served as the Director of the Virginia Center for Digital History and Associate Professor of History in the Corcoran Department of History at the University of Virginia. He is the author of *Lawyering for the Railroad: Business, Law, and Power in the New South*, published in 1999 by Louisiana State University Press. He is the co-author and assistant producer of a history of Virginia series for public television, called "The Ground Beneath Our Feet: Virginia's History Since the Civil War." Episode Three, "Massive Resistance," was an Emmy Nominee for 2000 from the Washington, D.C. Chapter of the National Academy of Television Arts and Sciences.

Pulitzer Prize-winning poet Natasha Trethewey is the author of *Domestic Work* (Graywolf, 2000), *Bellocq's Ophelia* (Graywolf, 2002), and *Native Guard* (Houghton Mifflin, 2006), for which she was awarded the Pulitzer Prize in 2007. She is the recipient of fellowships from the Guggenheim Foundation, the Rockefeller Foundation Bellagio Study Center, the National Endowment for the Arts, and the Bunting Fellowship Program of the Radcliffe Institute for Advanced Study at Harvard. Her poems have appeared in such journals and anthologies

as *American Poetry Review*, *Callaloo*, *Kenyon Review*, *The Southern Review*, *New England Review*, *Gettysburg Review*, *Virginia Quarterly Review*, and *The Best American Poetry 2000* and *2003*. Currently, she is Associate Professor of English and Creative Writing at Emory University.

Dr. Tom Rankin is Director of the Center for Documentary Studies and Associate Professor of the Practice of Art and Documentary Studies at Duke University. A photographer, filmmaker, and folklorist, Tom Rankin has been documenting and interpreting American culture for nearly twenty years. His books include *Sacred Space: Photographs from the Mississippi Delta* (1993), which received the Mississippi Institute of Arts and Letters Award for Photography, *'Deaf Maggie Lee Sayre': Photographs of a River Life* (1995), *Faulkner's World: The Photographs of Martin J. Dain* (1997), and *Local Heroes Changing America: Indivisible* (2000).

Dr. Barbara Ellen Smith is Director of Women's Studies and professor of interdisciplinary studies at Virginia Polytechnic Institute and State University. For the past thirty years, she has been an activist-scholar in Appalachia and the U.S. South. She is the author or editor of three books and numerous articles, including *Digging Our Own Graves: Coal Miners and the Struggle over Black Lung Disease* (1987) and *Neither Separate Nor Equal: Women, Race and Class in the South* (1999). Smith has recently completed a community-based research and education project, carried out in collaboration with the Highlander Research and Education Center and the Southern Regional Council, on Latino immigration to the U.S. South. Additional research projects include an analysis of the interplay between flexible labor practices and globalization in the Memphis logistics sector.

Dr. Patricia Yaeger is Henry Simmons Frieze Collegiate Professor of English and Women's Studies at the University of Michigan. Her research interests include twentieth-century American literature and visual arts, southern fiction, feminist theory, literary theory, social geography, and trash in modern/postmodern ethnic American literature. Prof. Yaeger is the author of *Dirt and Desire: Reconstructing Southern Women's Writing: 1930-1990* (University of Chicago Press, 2000); editor of *The Geography of Identity* (University of Michigan Press, 1996); co-editor of *Nationalisms & Sexualities* (Routledge, 1991); and author of *Honey-Mad Women: Emancipatory Strategies in Women's Writing* (Columbia University Press, 1988). Among her current projects is "Luminous Trash: America in an Age of Conspicuous Destruction."

Dr. Earl Lewis is Provost and Executive Vice President for Academic Affairs and the Asa Griggs Candler Professor of History and African American Studies. He is Emory University's first African American provost and the highest ranking African American administrator in university history. Before joining the Emory faculty in July 2004, Lewis served as dean of the Horace H. Rackham School of Graduate Studies and vice provost for academic affairs/graduate studies at the University of Michigan. He was director of the Center for Afro-American and African Studies and also the Elsa Barkley Brown and Robin D.G. 101 Kelley Collegiate Professor of History and African American and African Studies. From 1984 to 1989 he was on the faculty in the department of African American Studies at the University of California, Berkeley. Lewis, who holds degrees in history and psychology, is author and coeditor of seven books, among them *In Their Own Interests: Race, Class and Power in 20th Century Norfolk* (University of California Press, 1993) and the award-winning *To Make Our World Anew: A History of African Americans* (Oxford University Press, 2000). Lewis's research and projects have been funded by the Rockefeller, Ford, Mellon, and National Science foundations. In 1999, Lewis was a recipient of Michigan's Harold R. Johnson Diversity Service Award.

4.4.1. June 2007 on-site meeting

In June of 2007, we convened the Subject Scholar Consultants for their first SouthComb annual meeting. Prior to the meeting, all SSCs were familiarized with the three main SouthComb project goals: 1) creating a portal service for Southernists; 2) conducting an outreach program to improve networked access to southern humanities collections and digital scholarship productions; and 3) working through the challenges of creating a sustainable digital library service.

The purpose of this meeting was to gain a better understanding of what tools and services will make the SouthComb service appealing to a broad Southern Studies constituency. To this end, we engaged the Subject Scholar Consultants in brainstorming discussions and asked them to vet and prioritize our proposed core features for inclusion in the SouthComb portal. Specifically, we discussed four main areas of the portal service

and the utility of each: 1) MetaSearching; 2) A Southern Studies Database; 3) GIS mapping; and 4) Pedagogical Tools.

4.4.1.1. MetaSearching.

We asked the Subject Scholar Consultants to help us identify the types of collections they would benefit most from having access to through a portal service, including publications that have not yet been indexed by other services and other resources, such as historical newspapers, music, literature, scholarship, or other pieces that are out of copyright. We revisited the MetaCombine model and toolkit and discussed what indexing capabilities and displays would most appeal to this group and their colleagues.

The Subject Scholar Consultants talked about the importance of properly vetting web-based content and prominently announcing in the portal service that these are peer-reviewed materials. They also shared with us their desire to see southern materials, including digitized historical newspapers and digitized literature, collected and served out through a central location.

Among the most important topics brought up around MetaSearching was the need for a service that collected RSS feeds from contemporary sources, particularly news feeds, that were publishing news about the south. This service, termed by Will Thomas "Today's South," emerged as a new service that we were asked to prioritize in the portal development.

4.4.1.2. Southern Studies Database.

We asked the Subject Scholar Consultants to respond to one of the features we had recently begun drafting for the portal, the southern Studies database.

The Consultants affirmed that there is no such service elsewhere, and that such a service would be quite useful to them. They agreed that the service should provide contact information and specialty information for scholars (primarily professors and independent researchers, but also graduate students on a self-selecting basis); information on southern studies programs and support structures; and ideally, a list of journals and other publications that focus on southern studies' topics.

4.4.1.3. GIS Mapping.

We asked the Subject Scholar Consultants to share with us what maps would be most important to develop for scholars in the field generally and what sort of GIS service would be most helpful to them.

The Consultants found GIS Mapping services to be among the most highly desirable features for the portal service. They suggested that we develop a series of maps that would be good, general use maps and offered to help us to determine what those maps should be. They also suggested that we create a way for users to deposit their own maps into the system in order to allow the system to grow beyond our team's limited capacity for production. Finally, they lauded the idea of having a GIS person employed on the portal who could respond to user queries for maps by creating maps for users. Funding possibilities for this included institutional subscriptions for x number of maps a year for professors of an institution at a standard rate or a per-map development charge to users. All agreed that such a response system should be a future endeavor, not part of the initial release.

4.4.1.4. Pedagogical Tools.

We asked the Subject Scholar Consultants for their ideas on helpful pedagogical tools that they would use in the classroom or in their advising work.

The Consultants suggested a syllabi service, to which college and university teachers could post their own syllabi; a reading-list service, to which graduate students could post their examination reading lists;

We also discussed sustainability issues, particularly regarding funding models for the portal service. The Consultants were divided in their approach. Some believed their institutions would gladly pay an annual fee in order to allow their professors and students to access the system, and that professors and students could approach their libraries for these subscriptions and meet with success. Others felt their institution would not support such a model. All believed that, if SouthComb became a pay-for service, it should provide some portion of its collection freely in an Open Access manner. All also felt strongly that enabling independent researchers and researchers at institutions that did not purchase a subscription should be able to pay by item or to buy an individual subscription (or both). All also believed that the system should not be expected to be supported by fees alone, but instead with a mixed-funding model that included institutional resources at Emory, sponsored funding, and possibly an endowment.

At the meeting, we also raised the question of what Southern Studies programs we should collaborate with via focus groups and interviews on this project. Charles Reagan Wilson volunteered the University of Mississippi; others suggested that we work with Emory, University of South Carolina, Vanderbilt, and University of North Carolina.

5.0 Investigating Sustainability

** This section needs to talk about the Sustaining Digital Libraries symposium and monograph—I'll work on this next week

5.1 Sustaining Digital Libraries Symposium and Monograph

In 2006, the project team brought together digital library leaders and other interested professionals from around the nation to explore the key issues involved in sustaining digital libraries. The symposium was structured purposefully to elicit a dialogue between the presenters and the symposium attendees on this increasingly important topic.

The symposium used a mix of individual speakers and panels to feature a variety of leadership perspectives on preserving information and making it accessible to widely varying communities. Funding agencies such as the new NSF Office of Cyberinfrastructure articulated the outlook of sponsoring groups. Panels included representatives from major programs such as the National Digital Information Infrastructure and Preservation Program (NDIIPP), the National Science Digital Library (NSDL), and the Digital Library Federation. Representatives from groups that have been successful in similar digital library functions, but which are not always included in such discussions, such as Amazon.com and the Inter-university Consortium for Political and Social Research (ICPSR), also presented at this event.

The keynote speaker, Dr. Gregory Crane (Professor of Classics and Winnick Family Chair in Technology and Entrepreneurship at Tufts University, and Editor-in-Chief of the Perseus Digital Library), opened the event with "Beyond Incanabular Digital Libraries," a discussion of the importance of institutional support and faculty involvement for the long-term success of digital library initiatives. Paul Berkman (NSDL), David Ackerman (NYU), and Tyler Walters (Ga Tech) gave presentations on defining, creating, and sustaining digital libraries. Panel discussions included Chuck Henry (Rice), Chris Greer (NSF), Kaye Howe (NSDL), and Martha Anderson (NDIIPP) talking about sustaining emerging "cyberinfrastructure" in the humanities, social sciences, and sciences, and Darrell Donakowski (ICPSR), David Seaman (DLF), Vicky Reich (LOCKSS), and Robin Asbury (Booksurge/Amazon) sharing their insights on sustainable organizational models.

During the symposium, a call for papers was distributed both to guests and to the larger digital library community for a monograph on this topic entitled *Strategies for Sustaining Digital Libraries*. The co-PIs of this project are serving as the collection's editors. The forthcoming publication will provide resources for digital library stakeholders who seek to better understand how to effectively evolve such efforts from short-term projects to long-term sustainable programs. The monograph will include contributions from leaders in major digital libraries that have made such transitions or which are systematically considering the question of programmatic sustainability, including representatives from groups such as the National Digital Infrastructure and Information Preservation Program (NDIIPP), the National Science Digital Library (NSDL), and the University Corporation for Atmospheric Research.

Contributions cover such topical areas as:

- Business models that have proven effective for sustaining digital library efforts, especially revenue-stream models that are potentially generalizable to a range of other endeavors;
- How to implement best practices for mobilizing collaborative efforts in sustainable ways;
- Exemplars of nonprofit incorporation for digital libraries created in research settings; and
- Taxonomies of sustainable organizations which clarify the options available to digital libraries that are trying to transition into sustainable configurations.

5.2 Operational Models Considered

Many of SouthComb's services will be free; however, the project team is considering making several features available either exclusively or in full-feature mode to SouthComb members. Since SouthComb's user base will vary, the implementation of this payment process will be complex. In all likelihood, SouthComb will offer a range of membership options. Several of these options follow.

Association Model
As SouthComb is being assembled, advisors and consultants have noted a glaring absence of a coherent Southern Studies association of scholars. Unlike most other academic fields, scholars of the South cannot turn to an association to provide services like a directory of other scholars, a calendar of subject-specific events, or an annual conference that is as interdisciplinary as the field itself. Instead, such groups as the Southern Historical Association and The Society for the Study of Southern Literature address southern topics within disciplinary frameworks, serving only small portions of the larger southernist community

Because SouthComb aims to provide many of the valuable services traditionally tackled by subject associations in academia, creating a Southern Studies association parallel to SouthComb could be a viable option. Under this membership model, scholars of the South would join the academic association, and receive full SouthComb membership.

However, the establishment of an academic association may exceed the scope of this phase of the SouthComb project.

Institutional Model
Perhaps the simplest membership model proposed for SouthComb, the Institutional Model would imitate the membership processes already available for a variety of value-added services such as Factiva (www.factiva.com) and Lexis-Nexis (www.lexisnexis.com),[4] Universities, library systems, state school boards, or consortia could purchase various types of institutional memberships, which would feature either unlimited usage from certain IP addresses or a definite number of individual accounts tied to the institution or consortium.

The Institutional Model does have some obvious drawbacks, including the initial cost to SouthComb of convincing institutions that it is a valuable investment. It will be easier to acquire members of this type after SouthComb is an established portal. To this end, many of the SouthComb services may initially be provided at no cost at all, and only later have some features limited to paid members. The self-building community will be SouthComb's best argument for institutional purchasers.

Individual Model
One of the pervasive problems faced by southernists is the lack of coherent definition of the field. In fact, many scholars work at institutions without a program explicitly dedicated to Southern Studies or any similar subject area. These institutions may not be eager to purchase an institutional membership to be used by only a few faculty and students. It is therefore imperative to provide these isolated scholars with a way to become members of SouthComb on their own.

Individual memberships can be priced variously according to a number of factors: students, faculty, and amateur researchers may each require their own pricing scale. Feature availability could be tailored to each group, allowing all faculty to have some services, while a different menu of services is available to students. Alternatively, feature availability could be offered *à la carte*, with each member entitled to a given number of features. This would allow scholars to maximize the utility of their SouthComb experiences while minimizing expense. It would also provide valuable feedback to the SouthComb developers about which services are most used by which types of members.

[4] Please see Bates, Mary Ellen, "Free, Fee-Based and Value-Added Information Services." (Factiva 2002 White Paper Series and 2004 White Paper Series) http://factiva.com/collateral/files/whitepaper_feevsfree_032002.pdf; http://www.factiva.com/collateral/files/whitepaper_feevsfree_0504.pdf.

6.0 Next Steps

Appendices

Appendix 1. Focus Group Agenda

Cyberinfrastructure for Scholars Project
University of Mississippi Focus Groups Visit

Date / Location

Tuesday, August 29, 2006, from 10:30 AM to 3:30 PM. All sessions at Ole Miss will be in the Barnard Observatory conference room.

Purpose of the Focus Groups

These focus groups are being conducted to gather information from potential users to guide the development of the SouthComb Learning Environment, a new academic portal for Southern Studies. The visiting team from Emory University will ask questions during these sessions to better understand the varying needs and interest of faculty, students, and librarians in a learning environment for this broad interdisciplinary area. A secondary aim of the visit is to explore other ways that Ole Miss and Emory might work together on scholarly communication activities (for example: metadata best practices, digital preservation).

Three Goals of the Cyberinfrastructure Project

A. Build a Sustainable Combined Search Portal Service:

Building on several years of digital library research, we will create an easily manageable and reusable software suite for the creation and maintenance of humanities-oriented search portals that implements all of the experimental techniques that we have developed to date in the MetaCombine project for harvesting, automatically classifying, and metasearching information resources combined from multiple sources (Web, OAI, and other sources). We will use this software to implement a scholarly subject portal focused on Southern cultures and history that will index and organize sources reviewed and selected by an advisory panel of scholars. We tentatively intend to call the service SouthComb, a name meant to be evocative in several senses: "Comb" is a root word simultaneously associated with a tool for organizing unruly tangles of hair, an ordered cache of honey created by industrious bees, and an agricultural vehicle that harvests masses of ripe grain. By combining the roots "South" and "Comb" we hope to capture a variety of connotations for such an information harvesting and organization system.

B. Improve Networked Access to Humanities Collections in the South:

There is a great need to mobilize collaborative efforts to improve networked access to humanities research collections at cultural repositories. Our previous work has led us to several effective approaches to addressing this need that we will deploy. Assisting smaller institutions in deploying mechanisms like the OAI-PMH leads to greatly improved exposure of "hidden" collections through metadata harvesting services. Subject portals, especially those with metasearch capabilities, have been shown to have great utility for both scholarship and teaching. New models for digital peer-reviewed publications such as Southern Spaces have provided scholars with ways of producing research that uses such collections and makes them more accessible. We will pursue all of these approaches in the course of this project, and will make a special effort to establish long term mechanisms for engaging in such collaborative efforts.

First, we will work with the faculty of several of the largest Southern studies programs in the country, including the programs based at the university systems of Mississippi, South Carolina, Virginia, North Carolina, Kentucky, Georgia, and Alabama. By working with an expanded group of Southern studies

programs and archives to deploy and refine the technologies we have developed, we intend to make the SouthComb service an integral part of how materials are used by scholars investigating Southern cultures and histories, and how archives provide access to such materials. We will specifically engage in collaborative efforts with these academic programs to tailor sub-portals of the SouthComb service to their research needs, conducting an analysis of their research and teaching needs and then associated usability studies to make sure the service is accomplishing its aims. Several different kinds of contextualization services will be incorporated in this system for purposes of research and pedagogy, potentially including some GIS and recommender capabilities.

Second, in the course of these interactions we will further widen the involvement of scholars in producing new digital publications in Southern Spaces contextualizing and analyzing collections held in cultural repositories throughout the region. We have developed methods for engaging scholars in such digital productions in the course of cultivating the Southern Spaces internet journal and forum. This approach directly connects scholars with the process of improving access to collections.

Finally, we will work with a number of smaller, hidden "treasure-trove" archives in these states to improve access to their holdings. This would be done by advancing our model for regional collaboration developed in the IMLS-funded Music of Social Change project (http://www.metascholar.org/MOSC/) for exposing digitized collections of materials related to various aspects of Southern cultures and histories, including the civil rights movement, Southern folk life, and early Black pamphlets. We will advise institutions in the use of OAI-PMH tools such as the Metadata Migrator (http://www.metascholar.org/sw/mm/) which we recently developed as part of the same IMLS project. We will also work with the partner institutions of the MetaArchive Cooperative preservation network (http://MetaArchive.org) led by Emory University, which has been developed as part of the National Digital Information and Infrastructure Preservation Program (NDIIPP) of the Library of Congress. This collaborative effort will seek to provide better access to the public materials (as opposed to embargoed or private materials) preserved in the MetaArchive network.

Our general goal in all of these interactions would be to seek ways of better exposing high-quality collections for scholarly communication purposes. We are particularly interested in analyzing such interactions with a mind to long term sustainability of such efforts. This relates to our third and final project goal of sustainability for our program.

C. Explore Sustainable Models for the Advancement of Scholarly Cyberinfrastructure

A third goal of this project is to examine models for creating sustainable long-term services for scholars. This will be accomplished through consultation with a variety of experts to explore options for such service programs at Emory University. A careful market analysis will be conducted by consultants in the first year of the project for the services of this sustainable entity. A compelling user interface will be developed, making use of web design consultants. Operational models for the service will be evaluated in consultation with business service consultants. Finally, several meetings will be conducted with experts to explore models for sustainability of digital library services.

University of Mississippi Agenda

10:30 AM-11:45 AM -- Librarian Session

Session Structure:

10:30 – 11:00 Welcome, introductions, coffee, overview of session, and brief intro to Cyberinfrastructure Project

11:00 – 11:30 Discussion w/ Mississippi librarians, ask them to describe their interaction with Southern Studies faculty, database selection process, ways for us to contact them for feedback

11:30 – 11:45 Summary of major points, identify future steps

Desired Outcomes:

1. Share information with librarians about the project
2. Gain a better understanding of the library perspective on the SouthComb service
3. Identify ways of working with them in the future to get feedback

12:00 PM-2:00 PM -- Faculty Session

Session Structure:

12:00 – 12:20 Welcome, introductions, get lunch, session structure

12:20 – 12:30 Overview of SouthComb and context of Cyberinfrastructure Project

12:30 – 12:50 Faculty share teaching and research styles

12:50 – 01:30 Focus Group questions

01:30 – 01:55 Introduce portal and portlets

01:55 – 02:00 Conclusions

Desired Outcomes:

1. Share information with faculty about the project
2. Obtain information about faculty practice in research and teaching, especially as relates to online tools
3. Gain a better understanding of faculty perspective on the SouthComb service
4. Identify ways of working with them in the future to get feedback

2:15 PM - 3:30 PM -- Student Session

Session Structure:

02:15 – 02:30 Welcome, introductions, snacks, session structure

02:30 – 02:45 SouthComb overview, introduce portal and portlets

02:45 – 03:00 Students share learning and research styles

03:00 – 03:30 Focus Group questions

Desired Outcomes:

1. Share information with students about the project
2. Obtain information about student practice in research and learning, especially as relates to online tools
3. Understand their response and perspective on the SouthComb service.

Appendix 2. User Personas

Name: Eva L.

Age: 63

Primary role: Staff Assistant, Mary Buie Museum, University of Mississippi

Education level: M.F.A.

Residence: Oxford, Mississippi

Hometown: Vicksburg, Mississippi

Marital Status: Married 42 years

Children: 3; 7 grandchildren

Hobbies: Junior League, volunteering through her church

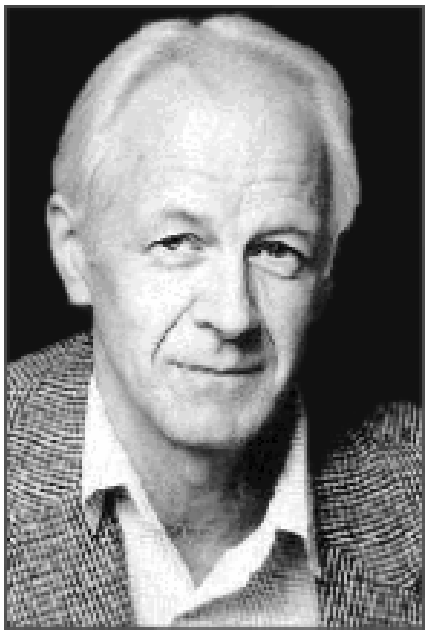
Computing skills: Has a PC at home which she uses to check her e-mail. Uses Internet Explorer to browse the internet. Never installs or modifies software; takes a long time to feel comfortable with new things her children do to the computer when they visit.

Primary software used: Microsoft Word, Internet Explorer, and Image Viewer

Primary goals: Would like to use the internet to spread the word about her idol, Theora Hamblett, and Ole Miss's Hamblett collection (some of which is online on the University's pages). Isn't sure about the legality of posting Hamblett images elsewhere on the internet, but can easily provide an item-level catalog of the museum's holdings in Word format to other interested scholars. Would also love to view images of art similar to Hamblett's.

Primary features: Catalog contribution, browsing images

Frustrations/pet peeves: Doesn't want to be overwhelmed by options. Wants to access the features she needs and wants without having to figure out what everything else does.



Name: Allan W.

Age: 57

Primary role: Professor, Department of American Studies, Vanderbilt University

Education level: Ph.D. (Cultural Migration in the American South)

Residence: Nashville, TN

Hometown: Boston, MA

Marital Status: Partnered

Children: none

Hobbies: Film discussion group, reading, travel

Computing skills: Doesn't seek out new technologies, but isn't afraid to learn to use them, either. Uses word processing and presentation software programs, and browses the internet recreationally. He even has a MySpace account, created to see what his students were talking about all the time. He has recently finished scanning every photograph he has ever taken, as well as every photograph and slide his father ever took, so he has some basic knowledge of photo-editing software. Discovered Flikr while looking for a way to share his photographs after a trip to Europe last summer and has been an avid fan ever since.

Primary software used: Microsoft Word, Picasa Photo software, PowerPoint, Excel

Primary goals: With fabulous teaching evaluations, Professor W. is known for his trendy, pop-culture-laden lectures and of-the-moment insights. He dedicates himself to his teaching and spends a great deal of time looking for interesting, engaging ways to discuss sociological aspects of the American South. He would like a bank of syllabi for inspiration, and he would be grateful for easy access to primary sources, video, and images useful for his lectures and his students' classwork. But balancing a fulfilling personal life and a committed teaching life is difficult, and Professor W. considers chat/bulletin board sites to be pretty much a waste of his time.

Primary features: image repository, syllabus repository, other primary-source access, search feature, teaching tools

Frustrations/pet peeves: Confusion, lack of clarity, or extraneous features that make his research harder to do.

Name: James D.

Age: 20

Primary role: Sophomore, Georgia Tech

Education level: Some college

Residence: On-campus housing

Hometown: Savannah, GA

Marital Status: single

Children: none

Hobbies: Gaming

Computing skills: A true child of Generation Y, James has never been without a computer. He wrote his first simple program at age 7 and has never looked back. Despite his programming competency, he is considering biomedical engineering as a career. Computers are, therefore, tools – omnipresent and generally capable of doing what he needs, when he needs it – but not things about which James feels a passion, in and of themselves.

Primary software used: IM software, word processing, relevant scientific software, really cool widgets he picked up here and there

Primary goals: After a summer conversation with his grandfather, a peanut farmer, James has decided to try taking a limited number of humanities classes while at Georgia Tech, and is currently enrolled in a special topics course on Southern literature. He is enjoying the course, but certainly doesn't have the free time to devote much more than the minimum amount of energy to its assignments. James primarily wants to be able to find sources for his coursework.. He will browse interesting-looking sites, however, and might be interested in sharing his grandfather's collection of Depression-era photographs in a vital photo-sharing community, should he stumble upon it.

Primary features: search feature, image repository

Frustrations/pet peeves: Glitches annoy James to no end. He wants websites to work quickly and easily, and avoids troll-ridden sites at all costs.

Name: Alison B.

Age: 23

Primary role: Student, CSAS, UNC

Education level: M.A. student

Residence: Chapel Hill, NC

Hometown: Miami, FL

Marital Status: Single

Children: A terrier mutt

Hobbies: Movies, music, cooking

Computing skills: Doesn't find computers particularly interesting, but can do everything she needs to do with them – e-mail, keeping in touch with friends, researching, and writing papers.

Primary software used: Web browser, IM software, word processing

Primary goals: Alison lives pretty far from campus and doesn't have to drive in every day, so she tries to do as much of her research online as she possibly can. This living situation, and the essential nature of Southern Studies as a field, can lead to Alison's feeling pretty isolated, however. She'd love a way to meet other people interested in traditional foodways. She'd also love to be notified of the academic opportunities (like calls for papers or conferences) in the English and History departments, from which she is often excluded accidentally.

Primary features: Search, primary source repository, IM/chat, calendar, RSS feed

Frustrations/pet peeves: Really wants a community of like-minded individuals; that is, scholars or those with scholarly inclination. Isn't interested in unmoderated forums or places where the virtual community seems too large. Doesn't want to feel pressured to visit a site every day and back-read threads of conversation.

Name: Amado G.

Age: 29

Primary role: Librarian and subject liaison (History), University of Maryland

Education level: M.A. (History), M.L.I.S.

Residence: College Park, MD

Hometown: Atlanta, GA

Marital Status: married

Children: none

Hobbies: Outdoor sports, photography, music

Computing skills: Very good. Amado is competent at the normal range of academic tasks, but also has eagerly embraced his new digital camera and can edit photos and video handily using basic software. He even leads occasional tech sessions when the regular librarians are busy.

Primary software used: Microsoft Office, Pinnacle Studio, Photoshop, Firefox

Primary goals: Receives many requests from faculty for sources relevant to the Civil War and Civil Rights eras, especially newspapers. Would love to make fulfilling these requests easier, especially by teaching faculty how to find that stuff themselves, if there were an easy source for such things.

Primary features: Search, image repository, video repository

Frustrations/pet peeves: Doesn't really love leading tech sessions for particularly obtuse faculty, and doesn't want to have to answer questions about how to utilize a website.

Name: Bobby J.

Age: 36

Primary role: Professor, Dept. of African American Studies, UC Berkeley

Education level: Ph.D.

Residence: Berkeley, CA

Hometown: Philadelphia. PA

Marital Status: Married

Children: 2

Hobbies: Barbershop quartet, mystery novels

Computing skills: Pretty decent. Having grown up prior to the computer revolution, he isn't quite sure how everything works, but is more than aware of how important computers are. Has worked hard to make sure his computing skills are at least passable, if not quite competitive.

Primary software used: Microsoft Word, PowerPoint, Internet Explorer

Primary goals: Because he has had to struggle to learn how to use some aspects of academic technology, he wants to make very sure his students know how to find reliable academic sources online. Moreover, he appreciates the ease of online citations and the entertainment factor added by well-designed digital presentations, which he requires of his students. And he would love to connect with other academics like himself to swap assignment suggestions.

Primary features: Search, image/video repositories, bulletin board, syllabi, teaching tools

Frustrations/pet peeves: Doesn't want to worry about the credibility of the sources his students use, and wants them to be able to use the sites to which he sends them easily and well.

Name: Patricia S.

Age: 39

Primary role: Anchor, WBRC-6 (FOX)

Education level: M.S.

Residence: Birmingham, AL

Hometown: Decatur, AL

Marital Status: Married

Children: 3

Hobbies: Crafts, golf

Computing skills: Uses computers mainly for communication. Can e-mail and chat easily, and can use a browser to find information she wants.

Primary software used: Microsoft Outlook, AOL-IM, Internet Explorer

Primary goals: Is the secondary anchor on the local 5-o'clock news. Has a weekly segment called "Patsy's Picks" which highlights events, happenings, and history locally and regionally. Would love to be able to scan headlines about local history, local events (especially cultural/artistic sorts of events), etc. to lend a more intellectual note to the segment, which sometimes features cheerleaders washing cars for lack of a better subject.

Primary features: RSS feed, directory (for contact information)

Frustrations/pet peeves:

Name: Liugen X.

Age: 61

Primary role: Professor Emeritus (Faulker)

Education level: Ph.D. (American Literature)

Residence: Beijing, China

Hometown: Tianjin, China

Marital Status: Married

Children: 1

Hobbies: Fishing, birds, martial arts

Computing skills: After two sabbaticals in the United States, bought himself a top-of-the-line PC in Hong Kong. Can use most standard software, although is not familiar with image/video editing, and hasn't figured out how to upload his grandson's baby pictures yet (grandson is now 2 years old).

Primary software used: Internet Explorer, Microsoft Word

Primary goals: To keep abreast of research and other academic happenings in the Faulkner/Southern Literature community. Communicates daily with colleagues in Denmark. Is planning a trip to Oxford, MS for the next Faulkner convention, and is co-authoring a new paper on Faulkner's sexuality with two other (American) colleagues. Relevant primary sources, networking tools, and online access to digital objects will be appreciated.

Primary features: Repositories, directory, bulletin board

Frustrations/pet peeves:

Name: Karin S.

Age: 41

Primary role: Administrative assistant, Tulsa Community College

Education level: M.A. (Art History)

Residence: Tulsa, OK

Hometown: Tulsa, OK

Marital Status: Married

Children: 2

Hobbies: Reading, home renovation, art

Computing skills: Basic. Uses proprietary software at work, so isn't exactly scared of computers, but had kids before she had one in her home so she's been too busy to learn much.

Primary software used: Microsoft Word, Internet Explorer

Primary goals: In her spare time, is co-authoring a book (with her sister) on Swedish and Norwegian art in the American South and Southwest. Has found that much of the primary-source documentation they require is a) in Swedish or Norwegian and b) hasn't been digitalized yet. She and her sister were inspired by discovering a stash of their great-grandmother's paintings and embroideries in a family attic, which they have documented digitally. Would be willing to share these images in a reliable environment in which Great-Grandma would still get credit for her work.

Primary features: Directory of archives, Archive locator (geographically), networking tools (to meet people who may live nearby or who can translate), search feature, possibly image repository

Frustrations/pet peeves: