

---

**Interim Report**  
**to the Andrew W. Mellon Foundation on the Project:**  
(excerpts)

***AmericanSouth.Org:***

**A Collaborative Project to Improve Access to Digital Resources on  
Southern History and Culture**

*Martin Halbert – Principal Investigator*

*Sandra Nyberg – SOLINET Grants officer*

*John Burger – SOLINET/ASERL Program Manager*

*Kate Nevins – SOLINET Executive Director*

**August 2002**

# Executive Summary

The AmericanSouth.Org project has made good progress on all goals and outcomes as set forth in its project plan. The OAI metadata harvesting and indexing infrastructure has been established and is in operation. Input from scholars, including members of the project's Scholarly Design Team, have provided feedback that has guided the development of the service. OAI providers have been installed at several partner libraries and are now in operation. Additional research collections have been identified for inclusion in the portal. Some scholarly portal functions have been deployed for testing. Finally, a number of project meetings and presentations have been held to coordinate and disseminate project information.

The project has resulted in a number of unexpected findings around the process of metadata aggregation and challenges to collaboration for this purpose. These findings include the need for further fostering adoption of the OAI protocol and the issue of "metadata format collisions". Next steps are identified for the second half of the project.

## **Accomplishments to Date:**

### ***Central Harvesting and Indexing System***

The creation of a reliable and robust central system infrastructure to handle OAI harvesting and indexing functions was perhaps the most critical goal for the first half of the project, and it consequently received a great deal of attention. The central system infrastructure had to be capable of routinely harvesting metadata from the partner sites of the proposed metadata aggregation network that forms the core of the AmericanSouth.Org proposal. The system also had to be capable of providing search and retrieval services for this metadata in order to make it useful to scholars. The software had to be scalable to hundreds of thousands of records. Finally, the software used had to be open source software (OSS).

Two systems emerged as the most promising candidates for collaboration. The first was the Open Digital Libraries (ODL) system under development at the Digital Library Research Laboratory at Virginia Polytechnic Institute (VPI). This software was being developed by the research group of the project's chief technical consultant, Dr. Edward Fox, and was a fully featured solution to the functions of harvesting and indexing metadata through the OAI-PMH. The second possibility seriously considered was the ARC system developed by the Old Dominion University Digital Library Research Group under Dr. Kurt Maly. ARC also represented a comprehensive solution to the question of harvesting and indexing metadata via the OAI-PMH. Both research groups were approached for more information about two areas: 1) making their software systems open source, and 2) the possibility of collaboration with the MetaScholar Initiative. Both research groups indicated their willingness in both areas. An extended evaluation of both products was therefore undertaken to determine which was the best option technically. ARC was selected as the underlying software for the central system infrastructure. Extended testing of the software during 2002 has shown the software to be reliable and adaptable to the purposes of the project.

### ***OAI-PMH Providers at Partner Sites***

A key feature of the proposed AmericanSouth.Org project was the creation of a metadata harvesting network of OAI provider systems set up and maintained at partner research libraries. This challenging goal entails a series of three-way collaborations between the MetaScholar project staff, partner institution staffs, and the project's OAI consultants at Virginia Polytechnic Institute (VPI). Each collaboration requires 1) the development of a working relationship between the three groups, 2) reaching a shared understanding of the local research collections relevant to the subject domain of the project, 3) reaching a shared understanding of the metadata relevant to the research collections under consideration, as well as the local systems infrastructure in which this metadata is maintained, and finally, 4) a customized solution for creating and grafting an OAI provider system on top of the local metadata infrastructure. We have found each collaboration complex, unique, and challenging. The individual collaborations have uniformly been intriguing and rewarding encounters among talented people working toward a common goal.

## ***Discovery of Scholarly Resources***

The project proposal described aggregating metadata for particular scholarly resources at particular research institutions. In addition to these specified research collections, an expectation was articulated that additional relevant collections and material would be opportunistically identified during the course of the project as relevant for inclusion. These opportunities might occur systematically by project staff studying the partner institutions, through discourse among members of the SDT, during site visits to the partner institutions, or in other unplanned ways during project collaboration activities.

These anticipated opportunities have indeed arisen. As project staff members have collaborated with partner site staff, additional research collections have been identified at AmericanSouth partner sites. Some examples include Auburn's Transforming America collection on Black history and civil rights in Alabama (see <http://diglib.lib.auburn.edu/transam>), Emory's Southern Changes Online collection from the Southern Regional Council (see <http://chaucer.library.emory.edu/schanges/southernchanges.html>), and several of the new digital library collections at the University of Tennessee at Knoxville (see <http://diglib.lib.utk.edu/dlc>). Most of these collections did not exist at the time that the grant proposal was submitted, and project staff heard about them while they were under development during meetings with the partner institutions.

As AmericanSouth.Org was conjoined with MetaArchive.Org, many more opportunities for additional research collections arose. The vast majority of the collections identified from the beginning for inclusion in MetaArchive are natural additions to the AmericanSouth subject domain, since they primarily represent archival collections of Southern political figures and religious institutions. More than two dozen collections from various MetaArchive sites concern Southern topics, and will in all likelihood be included in AmericanSouth.Org. As project staff entered into discussions with MetaArchive partner institutions, they also began to uncover previously unknown collections that were of relevance to AmericanSouth subject domains.

All of these opportunistic additions were desirable occurrences, but they had the effect of further elaborating the process of collaboration. This was because collection evaluation activities had to now be factored into the work, as well as increasing the number of custom OAI provider systems that had to be developed for AmericanSouth and MetaArchive sites.

Interesting boundary classification decisions arose in these evaluative activities, in a way similar to collection development activities in libraries. Were particular collections relevant to AmericanSouth? MetaArchive? Both? The project proposal anticipated the development of criteria for selecting resources for AmericanSouth (p. 6), and indeed this was necessary. However, without the formal guidance of the SDT, these criteria decisions were sometimes hard to adjudicate. An example was *the Martyred President Sermons*, an online collection of 57 sermons given on the occasion of the assassination of Abraham Lincoln (see <http://chaucer.library.emory.edu/lincoln/html/>). The sermons are concerned with Lincoln's assassination, a topic of relevance to the Civil War and potentially AmericanSouth, but all the sermons were delivered in northern states. The documents concern a major political figure (a MetaArchive subject domain), but are not from archives of Lincoln's papers. The sermons are religious institutional documents, another MetaArchive subject domain, but very different from other items in the MetaArchive collections. The metadata for the sermons will be harvested for AmericanSouth.Org, but we also intend to query the SDT about the decision, as all such harvesting decisions will be reviewed by the SDT.

Another category of relevant metadata that became obvious in the first months of the project work were research collections for which OAI provider systems had already been established, notably *American Memory*. The complicating issue is that of harvesting only relevant records from a body of metadata as broad as *American Memory*, which does not implement sets in such a way that records associated with the American South can be harvested. A keyword match approach is now being investigated in conjunction with input from the SDT. The project team intends to harvest records from other OAI provider systems such as *American Memory*, once an acceptable filtering process can be developed.