

Linking The Americas

A Project Supported by the Andrew W. Mellon Foundation

Final Report

Stanford University Libraries & Stanford University Press

28 September 2000

Summary

The Linking the Americas project succeeded in converting to digital form and mounting on the Internet a collection of books pertaining to Latin America that had been published conventionally by the Stanford University Press. In addition, some rare and manuscript materials from Stanford's Special Collections have also been converted and mounted. This project has taught a variety of lessons in both policy and production issues surrounding digitization of previously published materials. The Stanford University Libraries are pleased to acknowledge the support and patience of the Andrew W. Mellon Foundation in the evolution of this project, the fruits of which may be examined at <http://lta.stanford.edu/>. A guided tour has been placed online for use in conjunction with this final report: <http://lta.stanford.edu/report.tour.html>.

This version of the report excludes all three appendices.

Review of Project Activities

Conversion of Monographs

The project has converted 109 Stanford University Press monographs related to Latin America via a contractor, as described below. After conversion, it was discovered that an undetermined number of titles are officially out of print, and the Stanford University Press has allowed the copyright to revert to the authors. We are currently exploring the implications of this issue with regard to online sales. While it is not known at present exactly how many titles we can offer for sale, we have moved forward with delivery of access to all of the project's materials for the Stanford community, for our Latin American partners at UC Berkeley and UT Austin, and for a variety of institutions in Mexico, Central America and South America who have expressed strong interest in online access to these materials as well as future additions to the repository as they may become available.

We have also included in the set a scholarly anthology published in Spanish by the Office of the Presidency of Chile and now being published in English by the Stanford University Libraries, *Chile en los noventa / Chile in the Nineties*. This is the first title available through the project in two languages. The Spanish edition was converted like the SUP titles, i.e., from paper to PDF and HTML. However, we were able to convert the English edition directly from the Quark files on which the print edition is based. This method, although not available for most conventionally-published titles, is highly efficient and effective, though there is a learning process or significant cost, depending on whether this

work is done in-house or through a vendor.

The project has focused on obtaining a large volume of converted electronic content through concurrent contracts with several service vendors, a process that provided some perspective and insight into project specification and quality assurance/control.

Conversion of other collections

To richen the mix of documents and methods in the project, we selected over 2,500 pages of material held by the Stanford University Libraries' Special Collections Department and pertinent to Latin American studies. (These materials are listed in Appendix 3.) As most of the materials involved are either unique manuscript documents or extremely rare, it was critical to engage a vendor able to work onsite and in close cooperation with Libraries' staff. We contracted with a local vendor to convert these documents via scanning and conversion of page-based files (as opposed to character-based keyboarding and tagging). Numerous working issues of quality control, workflow, and data flow were successfully resolved over the months of effort. Through scanning documents onsite, Altos staff output high-resolution TIFF files, which they subsequently (and offsite) reprocessed for quality and for creation of PDF files, which have since been successfully loaded both at the LTA website and by eBrary.com.

The PDF files are searchable via Acrobat Reader based on 60-70% success with single-pass, unedited OCR from the source TIFF files. We have had some hesitation about use of proprietary file formats such as PDF, but given the relative ubiquity of PDFs in the online world and the availability (now and for the foreseeable future) of Acrobat Reader at no charge, we felt reasonably confident of the utility and prudence of this format. As a backup, of course, we intend to retain indefinitely the underlying TIFF files for possible future use as source records for generation of better, yet-to-be identified, delivery mechanisms.

Access

The LTA website was made available initially to Stanford's Latin American Studies community and subsequently to several ARL and SALALM institutions during the early Fall of 2000. Incremental increases in access, tuning of access functionality, policies, and models, as well as analysis of methodologies and initial public reception will ensue.

As described in our previous report to the Mellon Foundation, we contracted with Versaware, Inc. to convert printed, bound books to electronic form, using off-shore facilities. These titles have been further converted from the contractor's internal format to HTML and PDF. All titles are available through the project website in HTML, at least for local readership; some, are also available for purchase through Versaware in their proprietary format (readable through their client software). Additional access will be enabled as copyright clearances continue to be obtained.

It is worth emphasis that we insisted the vendor provide all "e-book" functionality and navigation capability via standard web browsers (i.e., Internet Explorer and Netscape Navigator), without proprietary desktop software. We took this position to ensure the broadest possible access to resources regardless of the level or brand of desktop

technology used to “read” these materials (no matter what version of Windows, what flavor or Mac or UNIX desktop, any reader with a standard HTML browser and Internet access will be able to view). We made this decision fully aware that such a "stateless" nature of the web poses real challenges for interface design. As delivered by the contractor, the books in HTML include – and allow navigation among – the books’ complete:

- tables of content
- chapters
- footnotes, references, indexes

Links to and within these texts include:

- Tables of Contents
- From footnote notations within text to footnotes (and back)
- Tables, pictures, etc. to higher resolution versions
- Entries in bibliographic lists linked to searches of Stanford’s union catalog
- Links from Stanford’s union catalog to the content of each book

The website provides adjustable scope of search, beginning with browse and search access to citations, topics, and geographic areas, and including full text search of contents of each title, as well as searching among the contents of multiple titles. The result of multiple-titles searches highlights the hits within a list of titles, allowing convenient examination of pertinent titles and sections within titles. The reader may add / remove titles from the group of titles being searched.

It will also be possible for the reader to establish a “personal” library, allowing the reader to determine, organize, and categorize a subset of the available titles into virtual “bookshelves.”

Versaware’s sales vehicle, its online bookstore, allows individuals to purchase titles for downloading in their proprietary format. An initial set of titles is included in the “for sale” category; once we have further explored rights of ownership of out-of-print titles, we will proceed accordingly. We are curious what market acceptance this format and sales vehicle will develop; there has not been sufficient time and experience yet to gauge these.

Stanford University Libraries staff, working with Stanford MediaWorks’ professional web developers, built a unified project-specific website, to provide several ways of searching and identifying converted materials and to link to kindred sites elsewhere, as well as a place to describe and document the project itself. The development effort required attention to the several classes of material involved, the several formats of document available, the related business issues, and desire to make as flexible and powerful a site as possible. The site provides links to the several players’ web sites: the Mellon Foundation, Stanford University Press (and its online hardcopy sales function, where desired), the Versaware site (for downloading of complete titles in proprietary format),

The project materials are also tightly linked to and from Stanford's OPAC, called Socrates, allowing a user of Socrates to jump directly from a bibliographical record to the full text document (whether in HTML or PDF) and to automatically click a bibliographical citation in the project-generated texts and use it to search Socrates. We think this is a model of sound practice for digital texts in the research library environment, and although the functionality may not be apparent to all readers outside the Stanford domain (for a variety of reasons), this model is valuable as exemplar to other institutions and systems.

At a point we feel the site is robust and well-populated, we will issue a press release (both on paper and via email) to the world at large.

Meanwhile, Stanford has initiated discussion with partner institutions about broadening access to the texts and searching capabilities. The goal is to provide optimal functionality for readers globally, as well as to continue experimenting with business models and partnerships. For example, we have initiated discussion of site licensing of the entire suite to institutions (especially abroad) that might find the texts of particular value. We expect to use this corpus as a quid pro quo for exchange of other scholarly digital texts, preferably pertinent to Latin America. We feel that such arrangements are an appropriate means of sharing resources, motivating others to produce shareable digital properties, and direct attention to the LTA materials.

Linking the Americas will shortly be opened to Latin American partner sites at the University of California, Berkeley and the University of Texas, Austin. Several Latin American institutions (ranging from Mexico's Instituto Tecnológico de Monterrey and El Colegio de Mexico to Argentina's Universidad Nacional de Cordoba, Universidad Torcuato di Tella, Centro de Estudios de Poblacion-CENEP, and Peru's Pontificia Universidad Católica) have expressed interest in obtaining access to the monograph collection. In addition, Latin American Studies librarians from ARL institutions that have tested the project are very enthusiastic:

Wisconsin: "I am extremely happy to know that Stanford is on the vanguard in this area of e-text for the Americas."

Notre Dame: "This project has certainly raised the bar for Latin American Studies electronic projects."

Lessons Learned

This project has yielded insights in several overlapping areas: production, formats, distribution & marketing, and copyright.

Production: We determined that it was inefficient for monographic materials to be manually coded (e.g., for HTML), and that algorithmic approaches (which have served us well for journals) did not create useful efficiencies. Until we had become conversant with the searchable PDF format, quite late in the project, it was demonstrably most efficient to contract out the document conversion services. It proved generally more cost effective to cooperate with a vendor whose business plan involved acquisition of titles for individual

sales (in their own environment) than to rely on vendors whose relation to the material ended on delivery of converted files. This approach coincidentally provided us with a mechanism to allow individual sales of book titles, one of the objectives of the project.

Formats: The market has been changing rapidly for online provision of book-length text information. We have seen the elevation of PDF, in particular, into a de facto standard for text exchange, despite its reliance on proprietary software both for creation and reading. It provides (in its higher-end variants) for both accurate page images and search functions. We did learn in the process that minimal-resolution PDF files, however, were not adequate for the purpose of online publication. Even with PDF files, we found, it is critical to insist on detailed specifications in contracting with vendors. We also encountered extreme variations in the per-page prices for conversion between vendors. To some extent, there was a correlation between price and quality of result, but it is not clear that such correlation is inherent. Libraries are well advised to work with several vendors on a limited scale, concentrate on clear specifications, investigate deliverables and adherence to schedules carefully, and only then determine which vendor(s) to prefer.

Distribution & Marketing: The Stanford Libraries (presumably like other libraries) are not well positioned to sell these titles individually by itself (both for logistical and marketing reasons). It is not evident there is any strong, cost-effective mechanism for marketing of individual titles at isolated points (such as a single university webspace). Our distribution and marketing strategy follows a two-fold path: to incorporate titles into larger bodies of titles (namely Versaware and eBrary) with the expectation that searching and browsing functions will help independent readers find specific titles of interest; and to make the entire set of Latin American monographs available to interested academic institutions for a nominal fee (or reciprocal barter-like access to digital resources). We have solicited interest from a wide variety of institutions across the Americas in such an arrangement. Access is provided to the Stanford LTA site by outside connections on an IP-range basis.

Copyright: As we built up a larger pool of candidate titles to include in the Linking the Americas collection, we increasingly encountered uncertainty and delay regarding copyright. Although the Stanford University Press was the holder of copyright upon publication of all converted monographs, some author agreements stipulated reversion of copyright to the author (and/or to heirs and estates as appropriate) once the title went officially out of print. Thus identifying and contacting the current holder of copyright for some works can be extremely difficult. This process continues for some titles converted, but not yet publicly available, through this project. Concerns were expressed, as we increased the pool of titles during the project, that the online availability of certain currently-popular titles might adversely affect their hardcopy sales. However, it was eventually agreed that such impact would be minimal, and the objection to online posting of those titles removed.

Conclusion

The completion of the project does not, of course, mean an end to the LTA site and Stanford's efforts to disseminate the converted materials to as wide an audience as possible. We will continue to add titles to the available set as copyright arrangements permit. We will welcome addition of new direct-digital Latin American titles from the Stanford University Press, should their new management team find it consistent with their future business plans. We will seek new partners for site or other access mechanisms. In short, we will continue our effort to evolve the LTA site and content with the rapidly-changing digital landscape of academic publishing and library resource sharing.