WHITE PAPER

The Impact of Open Access Latin American Scholarship:
Digitizing the Backlist of El Colegio de México's Press

John Kiplinger
Valerie Yaw

In 2018, JSTOR received a grant from the Andrew W. Mellon Foundation to support the digitization of out-of-print titles from the Dirección de Publicaciones de El Colegio de México, A.C., as well as the dissemination of those titles on an openly accessible basis. Throughout the year-and-a-half-long project, we worked in deep collaboration with El Colegio de México Press to complete this project. This white paper is intended to document the significance of this work, the process we used to select titles, and what we have learned so far about the usage of these titles on the JSTOR platform. We hope this will help to benefit other initiatives interested in increasing access to out-of-print materials.

The Dirección de Publicaciones de El Colegio de México, A.C. was established in 1938. It offers a catalog of more than 2,400 titles and nine academic journals across the humanities and social sciences.

JSTOR, a service of the not-for-profit organization ITHAKA, collaborates with the academic community to help libraries connect students and faculty to vital content while lowering costs and increasing shelf space; provides independent researchers with free and low-cost access to scholarship; and helps publishers reach new audiences and preserve their content for future generations.

JSTOR gratefully acknowledges the contributions and cooperation of the following:

1. Gabriela Said Reyes, Director, Dirección de Publicaciones de El Colegio de México, A.C.

2. Ninel Salcedo Romero, former Director of Marketing, Dirección de Publicaciones de El Colegio de México, A.C.

3. Micaela Chávez, Director, Biblioteca Daniel Cosío Villegas de El Colegio de México

4. Guadalupe Vega, Coordinator of Data Organization, Biblioteca Daniel Cosío Villegas de El Colegio de México

5. Brian Connaughton, Área de Historia Regional y Comparada, Departamento de Filosofia, Universidad Autónoma Metropolitana

6. Robert Darnton, Carl H. Pforzheimer University Professor and University Librarian, Emeritus, Harvard University

7. Gilbert Joseph, Farnam Professor of History & International Studies, Yale University; Past President, Latin American Studies Association

8. Herbert S. Klein, Gouverneur Morris Professor Emeritus of History, Columbia University; former Director of the Center for Latin American Studies and Professor of History at Stanford University; Research Scholar & Latin American Curator, Hoover Institution, Stanford University

9. Jocelyn Olcott, Associate Professor, History and Gender, Sexuality & Feminist Studies, Duke University

10. William B. Taylor, Muriel McKevitt Sonne Professor of Latin American History, Emeritus, University of California, Berkeley

11. Pardha Karamsetty, President, Content & Media Solutions, Apex CoVantage; CEO, Apex CoVantage India

12. Prabhanjan Mattam, Project Manager, Apex CoVantage

Summary

In 2018, JSTOR received a grant from the Andrew W. Mellon Foundation to support a collaboration with the Dirección de Publicaciones de El Colegio de México, A.C., the press of El Colegio de México, a graduate research institution in Mexico City[1]. This grant enabled JSTOR to digitize nearly 700 books from the press's backlist in the humanities and humanistic social sciences, and make these books freely and openly available on the JSTOR online platform.

The goal of this project was to digitize and make openly accessible scholarship from the backlist of El Colegio de Mexico's Press that would be of significant value to students and researchers in a range of humanities disciplines.

The work on this project proceeded in three phases, including a preparation and selection process, in which JSTOR worked with experts in the field to determine which books would be digitized; a digitization and ingest phase resulting in the books being hosted openly on JSTOR; and an analysis phase, in which JSTOR sought to develop a better understanding of the impact that foreign-language materials can have when hosted on a global platform.

This project brought together Colmex's rich scholarly backlist with JSTOR's experience managing retrospective digitization projects and helping to increase the impact of academic content by making that content easy to find and use online. Colmex and JSTOR have collaborated over the past several years to make Colmex's frontlist books available to readers around the world through JSTOR.org. In this project, we sought to build on that collaboration by making a selection of books from the Press's backlist available in digital form for the first time. In this white paper, we document our process for selection and digitization of books and provide a high-level analysis of usage of the content on the JSTOR platform.

Introduction: History, Context, and Significance of the Collection

The press of El Colegio de México has published a body of important scholarship over the course of the last eight decades.

---

[1] Throughout this paper, we generally refer to Dirección de Publicaciones de El Colegio de México, A.C. simply as El Colegio de México or by its common name "Colmex."

The press was established in 1938 in Mexico City. It attracted a group of pathbreaking scholars in the humanities and social sciences, and Colmex's press—one of the earliest scholarly publishers in Latin America—provided an outlet for their work, which foregrounded some of the ongoing lines of inquiry in Mexican and Latin American studies, including scholarship on migration to and from Mexico, the interplay between church and state in Latin America, and women's rights.

The university's press published its first title in 1938 and continued to publish significant work throughout its history. The list of the press spans disciplines in the humanities and qualitative social sciences, with special emphases on history, sociology, literary criticism, and political science. For the most part, the books focus on Mexican and Latin American contexts.

In addition to a robust books program, the press of El Colegio de Mexico publishes seven journals, including *Historia Mexicana,* arguably the leading journal of Mexican historical studies. Over time, the press has also been an important outlet for making foreign-language writing available in Mexico: as one example, its journal *Diálogos* was the first to publish Milan Kundera's work in Spanish for a Mexican audience.

Since 2013, Colmex's press has published some of its new books in digital form and distributed them through digital scholarly platforms, including JSTOR. Like many established scholarly presses, Colmex licenses access to its frontlist titles to university libraries to help sustain its ongoing publishing program. However, much of Colmex's backlist was out of print and the press had never digitized it due to limited funding. In today's increasingly digital landscape, the lack of electronic copies of this important body of scholarly created, in essence, a barrier to accessing those titles.

This project sought to overcome this barrier and make these books discoverable and accessible for free by a worldwide audience. As noted in the Summary, El Colegio de México and JSTOR have collaborated over the past several years to make Colmex's frontlist books available to readers around the world through JSTOR. In this project, we built on that collaboration, bringing together Colmex's rich scholarly backlist with JSTOR's experience managing retrospective digitization projects and helping to increase the usage of academic content by making that content easy to find and use online. JSTOR has seen high usage and impact for both archival journals and for backlist monographs; in fact, two thirds of ebook usage on JSTOR is for titles published at least three years earlier.

Our Approach: Selection and Digitization

JSTOR digitized nearly 700 titles, or almost 50% of the press's backlist. Significantly, none of Colmex's backlist titles were previously available digitally. For every book made available through this project, each page was scanned and OCR processed, and accompanying book and chapter-level metadata was captured to make the books fully searchable, discoverable, and usable for scholars and teachers.

Selection

We asked a group of scholar-advisors to help us assess the broader significance of Colmex's list in Mexican and Latin American Studies by drawing our attention to books that are noteworthy and that should be highlighted in outreach about the project to scholars, librarians, students, and general readers.

Our scholar-advisors assisted with the selection process mainly in two ways. First, they gave us high-level guidance to inform our strategic sense of the collection's value. One advisor wrote to us that the press's list "[provides] studies of the economic, social, demographic, and political history of Mexico unparalleled by any other publisher."

Several of the scholars also noted the broad discipline coverage of Colmex's list; while we expected that the bulk of the books would be of greatest interest to historians, another advisor wrote to us that "[s]ociologists, economists, demographers, linguists and students of literature, geographers, and historians will all benefit by achieving the digital availability of these works." It is worth noting, as some of our advisors did, that the Press also has a strong list in Asian studies, and the set of titles that we digitized through this project includes books from that area. While the inclusion of these titles may initially seem like an odd fit for a project that focuses for the most part on Mexican and Latin American studies titles, the press's list in Asian studies reflects a critical aspect of the Mexican academy's global engagement. Colmex's Center of Asian and African Studies is, as one adviser noted, "the only functioning center on Asian studies in Latin America," and Colmex's press, picking up on this strength, has become "the major publisher of studies of Asian history in Spanish." To the extent that this digitization project is meant in part to reflect the strengths and disciplinary breadth of Colmex's backlist, it seemed important to include these titles in the project.

Second, while acknowledging the overall value of Colmex's backlist, our advisors also directed us to particular titles that have become classics in their field. For example, some of these titles include Silvio Zavala's multi-volume *El servicio personal de los indios en la Nueva España,* a study of labor and slavery in the 16th to 18th centuries; books and edited volumes by Andrés Lira on Spanish exiles in Latin America after the Spanish Civil

War; and *Los bienes de la Iglesia en México,* a study of the conflict between church and state in the 1800s.

Of particular note among the books we digitized is the *Historia general de Mexico,* a multi-volume work completed in the 1970s and edited by the Colmex historian Daniel Cosío Villegas. This work covers the range of Mexico's history from the dawn of human habitation. As one reviewer in a scholarly journal noted, Cosío Villegas had a longstanding interest in reaching non-academic audiences, and so the scholars who penned essays for the *Historia general* were asked to write such that a general audience could read the work. Thus, one project advisor wrote, the volumes are well suited to "students at the high school and university level as well as to adult readers who give them the time and attention they deserve." Despite the essays being shaped for a non-academic audience, one of our advisors noted that the *Historia general* remains "the standard general history [of Mexico] used by all scholars."

With this guidance in mind, the list of books we digitized resulted from a winnowing process, the stages of which are outlined below[1]:

(1) At the start of this project, Colmex had the necessary permissions to digitize and make freely available in digital form a significant number of titles in their backlist, in many cases because the author was a faculty member at Colmex. Given the sizable expense involved in clearing digital rights, we determined that there was significant value in focusing our efforts on books that did not require painstaking rights research. Of the 1,411 titles in the backlist, Colmex's press has distribution rights for 741.

(2) This list was then refined to exclude a small number of books that were not scholarly in nature (e.g., technical guides from the 1990s). We retained in the list, however, a small number of literary or primary source titles that would be useful for research and teaching.

(3) The list was further refined to exclude titles that did not fit well with the humanities and humanistic social sciences profile[2]. For example, books that focused on environmental policy were considered out of scope for this project.

(4) Finally, based on cost estimates, we initially aimed to reach a final list of approximately 600 titles. Given cost constraints, we made the difficult choices, including moving approximately 40 social science-leaning titles (many in political

----

[1] It is important to note that the winnowing process was undertaken by the project team with guidance from a set of scholar-advisors for the project, given that it was not feasible to ask these advisors, who are also full-time faculty members, to engage in a title-by-title selection process for a list of this size.
[2] This project was funded through the Mellon Foundation's Humanities Open Book Program, which emphasizes out-of-print humanities books.

science) to a B-list. It is important to note that, while these titles were not included in the starting list for digitization, lower-than-anticipated costs allowed us to include these titles in our final output. We acknowledge that this initial selection process was not perfect, but we are pleased with the final outcome since these books hold value for humanities researchers (especially historians).

At the end of our initial selection process, we had an A-list of 611 titles. While the vast majority of the books on this list were in history, literature, or other humanities fields, there were also a number of titles that were exceptions. Some titles on the list leaned more toward the social sciences, including a number of books on public policy. We felt that it would be appropriate to include them because they would be of interest to scholars of Mexican and Latin American history. In addition, a handful of titles on the list (fewer than ten) are literary or primary texts (for example, a Spanish-language translation of Giambattista Vico's *Scienzia nuova)*.

Production

JSTOR's production unit converts over 9 million pages of scholarly journal and book content per year, of which 2 million includes scanning from print sources. We have longstanding relationships with several digitization vendors, and we believed that our experience managing large-scale digitization projects would position us well to accomplish the digitization of Colmex's backlist books quickly, cost-efficiently, and to a high quality.

For books, JSTOR normally receives and processes PDFs from publishers. These PDFs go through automated workflows at JSTOR's end as well as processing by a third-party vendor. This project was different because the source document for each book was a print version[1], and one of the required outputs was an ePub for each book. JSTOR selected one of our current conversion vendors, Apex CoVantage, to handle all vendor processing for the books in the project. This included scanning of the print copy, return shipment of the print copy to Colmex, creation of the PDF from the page images, OCR for creation of searchable full text, metadata capture to JSTOR standard specification for books, and then creation of an ePub. JSTOR negotiated a per page price of $0.83 that covered all these tasks. The project covers 684 books.

Colmex sent nine shipments of print books to Apex CoVantage's production facility in Hyderabad, India. After Colmex's library (Biblioteca Daniel Cosío Villegas) harvested all the metadata needed in the production process, Colmex shipped the first batch in mid-April 2018; the final batch was shipped early-May 2019. Each shipment contained an

---

[1] Although JSTOR has scanned a relatively small number of print books outside this project, the bulk of our print scanning continues to be for journals. However, the same imaging specifications are used regardless of whether they are journal or book pages.

average of 76 print books. Apex conducted non-destructive scanning with each page scanned as 600 dpi bitonal TIFF and grayscale/color content scanned at 300 dpi for RGB TIFF images. We instituted a discrepancy process wherein Apex reported damaged, missing, or other problematic pages to JSTOR. JSTOR assessed these reports and, as needed, worked with Colmex, Harvard University Library, and University of Michigan Library collections to locate and scan replacement pages from other extant copies. The resulting page scans were then used in place of the damaged or otherwise unusable pages or to fill gaps where there were missing pages so that the PDF would represent a complete and intact version of the print original.

Apex submitted the completed PDFs to JSTOR and shipped the print copies back to Colmex. JSTOR's systems then ingested the PDF as well as spreadsheet-based supply chain metadata (SCM) provided separately by Colmex. The PDF and SCM were matched by the system and then were automatically sent to Apex for standard processing, which consists of OCR as well as book- and chapter-level metadata capture.

As Apex completed the standard processing for each book, they then put the books through an ePub creation process that, while very familiar to Apex, was new to JSTOR. The ePubs were created to the EPUB standard version 3.0.1 or higher. Additionally, the processing agreed upon between Apex and JSTOR ensured functionality such as links from footnote anchors in the text block to the footnotes themselves. However, features such as tables were captured as images rather than as HTML. During both the standard processing and ePub creation, Apex occasionally raised metadata capture queries that were reviewed and resolved by JSTOR's metadata librarian team of Karen Aufdemberge, Emily Betwee, and Rachel Ross, thus ensuring a higher and more consistent quality for the metadata.

Apex grouped the ePub, the PDF, and the book- and chapter-level metadata XML files into a zip file for delivery to JSTOR. JSTOR systems then ingested the zip file and ran quality control scripts across the files to ensure that they adhered to our specifications. For the initial batch of books, we also conducted a limited amount of manual quality control reviews of the metadata and of the ePub. To accommodate the ingest of these zip files, however, our content management systems staff had to update the JSTOR software to recognize and accept the different directory structure and files that were present (i.e., the directory containing the ePub as well as the ePub file itself) but had not been present in previous book deliverables from our vendors.

Furthermore, downstream systems for our content delivery platform had to be updated to recognize and appropriately route and make available the ePub file. JSTOR opted to treat the ePub in a manner similar to that of supplementary materials. The ePub is available as a downloadable file via a clickable "Download EPUB" button at the top of the page for each book in the project. Otherwise, the book is treated in a similar manner to any other Open Access title on JSTOR.

Of the 684 books in the project, the first titles became available on the JSTOR site on September 11, 2018. The most recent releases were on July 19, 2019. There are currently four books for which processing cannot be completed because the books do not have ISBN assignments, and the JSTOR systems require an ISBN.

While this project had typical logistical challenges, the challenges that were new to JSTOR were:

1. The need to send the books to one particular vendor instead of dividing them equally between our two vendors, which was addressed earlier in these comments; and

2. The lack of electronic version ISBN (EISBN) assignments for any of the books.

ISBN best practice indicates that an electronic version of a book should have an ISBN that is distinct from its print version counterpart. In fact, different electronic versions (e.g., PDF vs. EPUB) can have their own ISBN assignments. However, JSTOR opted to use a single ISBN assignment to cover both electronic versions of each book. Going into the project, Colmex did not have EISBN assignments for the books, and 102 of the books did not have a print version ISBN (PISBN) either. One problem was that, for Mexican-published works, the ISBN are assigned by a third-party agency, and the turnaround times for the assignments, particularly for large batches of requests, are unpredictable. Therefore, to not to jeopardize the overall timeframe for the project, JSTOR opted to use the PSIBN for any book that had a PISBN assignment. This would allow us to ingest the supply chain metadata into JSTOR systems and to keep individual book processing moving beyond the print-scanning stage.

Meanwhile, Colmex would apply for EISBN assignments attempting to prioritize the assignments for those books that had no ISBN assignment at all. For books that had no ISBN assignment at all, we could move them back into production post-scanning once we had the EISBN assignment. For books that had a PISBN assignment, we plan to do a mass swap of the PISBN for the EISBN once we have all those assignments. At the time this paper was written, 128 books were still awaiting an EISBN assignment, including four books that have no ISBN assignment at all and that therefore cannot proceed beyond the scanning stage.

We are currently planning a project to swap the PISBN for the EISBN for those books where we have the EISBN assignments. We will finish the processing and/or ISBN swaps for the remaining books when the EISBN assignments are available. If we were to do a similar digitization project for backlist books, we would certainly investigate the EISBN situation at the earliest possible stage and work with project partners to secure EISBN assignments as soon as possible.

Usage: What We've Learned So Far

This project represented not only an opportunity to digitize and make available books from the publication run of Colmex's list, but also to measure the usage of these books over time and, ultimately, to understand better the impact that foreign-language materials can have when hosted on a globally accessed platform.

Our objective in measuring usage was to understand how frequently the Colmex books are read online as evidenced by generally accepted metrics such as views and downloads of the chapter files. Additionally, we wanted to understand how this usage compares with the usage of approximately 4,500 openly accessible English-language books hosted on JSTOR.

JSTOR facilitates the discovery of ebook content in a variety of ways. We offer free MARC records to libraries through OCLC, and distribute metadata and full text to discovery services and search engines for indexing. Another important factor in driving usage is co-locating ebook chapters with journal articles on JSTOR's integrated platform, enabling users to cross-search all types of content at once. For many scholars, JSTOR is a starting point for research—in fact, our traffic referral data shows that more than 40% of visits to ebook pages are by users who were already searching and using JSTOR. Faculty and students are incorporating ebooks into their established research workflows on the platform. In addition, we promoted the availability of the Colmex titles via a short animated video in English and Spanish, email campaigns to librarians and faculty in Latin American studies, announcements shared via JSTOR and Colmex's web and social media channels, and promotions to members of the Latin American Studies Association, including advertisements and a presentation at the association's annual conference.
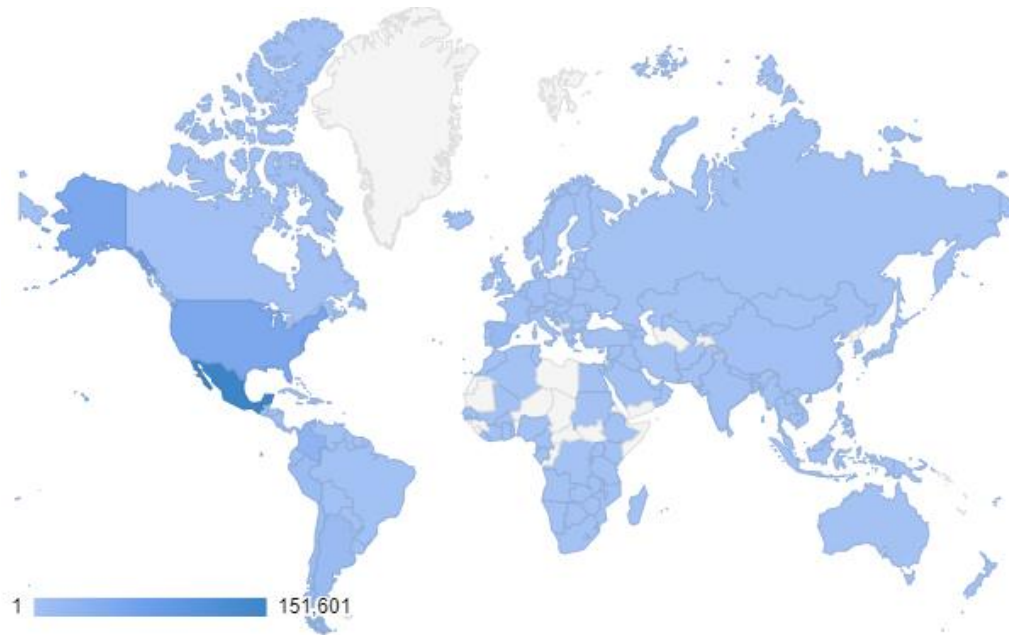
The Colmex titles digitized through this project have been heavily used on JSTOR. The 680 titles made available on JSTOR between September 2018 and July 2019 have been used a total of 502,134 times through October 28, 2019. Every single title has been used. The most-used titles are listed below.

Top ten most-used titles

Title	Copyright year	Usage through 10/28/19

Historia económica general de México: de la colonia a nuestros días	2010	13,251
Historia general de México: volumen I	1994	9,323
Los intelectuales y el poder en México	1991	5,156
De amicitia et doctrina: homenaje a Martha Elena Venier	2007	4,785
La lingüística en México, 1980-1996	1998	4,564
Diccionario del español usual en México	1996	4,503
Introducción a la historia de la vida cotidiana	2006	4,300
Historia de la lectura en México	1997	3,888
Cuestiones de teoría sociológica	2005	3,659
Historia general de México: volumen II	1994	3,595

The data show that there is a broad audience for this scholarship. The titles have been used in 173 countries and territories. While high levels of usage were recorded in Spanish-speaking countries, as we expected, usage also occurred in 161 countries and territories where Spanish is not a national or official language. The map below shows the countries in which we have recorded usage for the Colmex titles, and the table lists the ten countries with the highest usage.

Top ten countries that recorded the most usage

Country	Usage through 10/28/19
Mexico	151,601
United States	54,035
Colombia	29,369
Spain	17,001
Argentina	13,909
Peru	11,408
Chile	9,172

Ecuador	9,143
Costa Rica	4,770
United Kingdom	3,580

Because JSTOR works with thousands of institutions around the world, we can measure the usage of these titles at institutions that participate in our services. We recorded usage of the Colmex titles at 4,285 institutions. This included not only college and universities, but also community colleges, secondary schools, government and not-for-profit organizations, and public libraries.

JSTOR's ebook program had not previously hosted EPUB files; for this project, we added the capability for users to download the full book as an EPUB file from the table of contents page, as well as the standard option to view or download chapter-level PDFs. There were 19,234 downloads of EPUB files for the Colmex titles through the end of October 2019—just 3.8% of the total usage of the titles in that timeframe.

This project also gave us the opportunity to compare the usage of Spanish and English-language titles available on JSTOR. On average, the Colmex ebooks are used 57% as much as the Open Access titles in English on the platform. While there are other variables that may affect the level of usage (such as discipline or copyright year), this figure shows an impressive amount of usage of Spanish-language titles on a primarily English-language scholarly content site.

We've also received positive feedback from librarians and scholars regarding the access to this content. For example, responses to the news on Twitter included praise for the initiative ("Excelente noticia para @elcolmex y el ámbito académico de México y el mundo") and recommendations of specific titles ("Una de las joyas liberadas en acceso abierto [PDF / EPUB] por el Colmex a través de Jstor es /Los intelectuales y el poder en México/ (1991) un nutrido volumen colectivo que contiene muy buenas intervenciones, algunas de ellas referencias obligadas.")

Conclusion

As a result of this project, 680 significant works of scholarship (with four more coming when EISBN assignments are available) that were previously out of print are now

available to anyone who wishes to use them. They are easy to discover and access within researchers' existing digital workflows. The value of these titles is apparent in the strong usage we've seen over the relatively short period they've been available: more than half a million views and downloads across 173 countries. Scholars and students in Latin America and around the world are enriching their research with this content, and we have ensured that it will be available to future generations.

In addition, the Mexican government launched a project earlier this year, the Estrategia Nacional de Lectura, to promote reading and guarantee that books are accessible to the entire population. The 684 digitized titles will be openly available to the Mexican people and promoted as part of this project.

This project also built a foundation for continued work on the Open Access dissemination of Latin American scholarship. JSTOR is currently participating in a pilot led by the Latin American Research Resources Project (LARRP), a consortium of research libraries that is funding the Open Access distribution of 200 titles published in 2018-2019 by the Latin American Council of Social Sciences (CLACSO). This initiative, developed and supported by libraries, will test a framework for the sustainable, long-term stewardship of Open Access scholarly monographs.

We are grateful that the Humanities Open Book program grant funded by The Andrew W. Mellon Foundation provided the opportunity for JSTOR to partner with El Colegio de México to make its important scholarship available for researchers around the world to discover and use. We look forward to continuing to build on what we've achieved together.