

Flujo de partículas de matriz digital abstracta, conexión de datos digitales, concepto de tecnología/ KanawatH/iStock

Introducción

Archivomex¹ es un esfuerzo único en América Latina, y en gran parte del mundo, para poner la colección de datos históricos en las *Estadísticas históricas de México* (de aquí en adelante identificadas como EHM) al alcance de cualquier usuario. Nuestra metodología para brindar acceso amplio a esta fue la creación de una estructura de datos disponible, descargable y consultable en línea: en plataformas abiertas estructuradas de

manera transparente e inteligible para usuarios que vengan del entorno académico o de aquellos interesados en conocer el pasado de su país a través de estadísticas y/o mediante la recuperación de documentos históricos digitalizados.

Esta iniciativa tiene tres objetivos: 1) el desarrollo de una la plataforma que facilita el acceso a las EHM a través de un acervo de bases de datos modernas almacenado en el Banco de Información del Laboratorio Nacional de Políticas Públicas (LNPP) del Centro de Información y Docencia Económicas (CIDE), A. C.; 2) de aplicaciones web para visualizar-

¹ Portal al proyecto Archivomex, 2022, en línea: <https://mx.digital/archivomex/>.

los en línea; y 3) de herramientas digitales para semiautomatizar el procesamiento de tablas de estadísticas mexicanas en formato PDF a archivos Excel.

El rescate de la historia, así como su preservación, protección y distribución, han sido del dominio de museos y archivos, institutos históricos o arqueológicos, bibliotecas y universidades (donde los documentos y objetos se mantienen en estados frágiles y privilegiados). La transformación digital del mundo de la información también cambió a estas instituciones y sus repositorios, pero el acceso a muchas de sus colecciones se ha mantenido en lo físico.

Este artículo se enfrenta al tema del acceso a través de la digitalización de recursos históricos. En las próximas páginas se presentará Archivomex como un ejemplo tanto de historia digital como pública. Vamos a discutir de donde nace la primera y en qué se distingue de la de los recursos documentales impresos y a compartir cómo entre sus participantes se discutió y desarrolló una metodología de digitalización, organización y estructuración para proponer un caso de cómo crear, desarrollar y manejar un proyecto de digitalización exitoso para que el producto final tenga el mayor uso y utilidad para un público en línea.

Primera parte: Archivomex en su contexto histórico

Los inicios de la historia digital

La acidez de la tinta en documentos antiguos siempre ha causado problemas para la preservación de la información histórica.² Aun con el miedo a volverse obsoletos, el temor a que los bienes archivados se esfumen en pedacitos de papel quemado por la tinta siempre ha impulsado la exploración de métodos de respaldo más robustos que un *sobre antiácido*. Los primeros esfuerzos de digitalización eran, principalmente, para proteger los documentos originales —la intención detrás de esta era la

preservación, no el acceso al público—; no obstante, cuando se empezaron a hacer copias en microficha (en la década de los 30 del siglo pasado) y más tarde en discos duros (en la de los 70),³ el acervo histórico continuó manteniendo el acceso a estas, y el usuario tenía que viajar al archivo físico para poder consultarlas en dichos medios. Esto cambió en el siglo XXI, cuando la digitalización, que va viento en popa desde ya décadas atrás, se ha enfrentado a una generación de computadoras y plataformas que permiten un acceso mucho más democrático a la información.

La digitalización de texto en el sector de preservación histórica empezó en 1971 con el *Project Gutenberg*,⁴ que se dedicó a elaborar copias de documentos clásicos. El primero fue la *Declaración de Independencia de los Estados Unidos*, que se subió al servidor central de la Universidad de Illinois. El *Oxford Text Archive*⁵ se creó en Inglaterra en 1976 y lleva desde entonces almacenando imágenes y transcripciones de textos de la época medieval y anterior, e incluso mucho más recientes. El *ARTFL Project*⁶ de 1982 es una colaboración entre la Universidad de Chicago y el gobierno de Francia para brindar acceso a colecciones de documentos en francés que están en las colecciones de esa casa de estudios. Por otra parte, la iniciativa de codificación de texto (*Text Encoding Initiative*)⁷ es un consorcio internacional que se creó en 1987, y que colectivamente se dedica a desarrollar y mantener los estándares para la representación de textos en forma digital. Su tarea principal es establecer las normas para su codificación con el fin de que sean legibles en máquina. Estos son los inicios estructurales de las humanidades y la historia digitales (en la primera era del internet, las estructuras eran internas y académicas, y con el advenimiento de la web 2.0, o como se le llama a la segunda versión de la red, van a cambiar los parámetros de participación).

2 Gwinn, Nancy E. "The Fragility Of Paper: Can Our Historical Record Be Saved?", en: *The Public Historian*. Vol. 13, No. 3. Preservation Technology, Summer 1991 (DE) <https://www.jstor.org/stable/3378551>.

3 BMI Imaging Systems. *The History of Microfilm* (DE) <https://bmiimaging.com/blog/microfilm/history-of-microfilm/>.

4 Hart, Michael. *The History and Philosophy of Project Gutenberg*. Proyecto Gutenberg (DE) https://www.gutenberg.org/about/background/history_and_philosophy.html.

5 Universidad de Oxford. *Oxford Text Archive. Repositorios* (DE) <https://ota.bodleian.ox.ac.uk/repository/xmlui/>.

6 University of Chicago. *Project for American and French Research on the Treasury of the French Language* (DE) <http://artfl-project.uchicago.edu/>.

7 Huma-Num. *Text Encoding Initiative* (DE) <http://www.tei-c.org/index.xml>.

Web 2.0 y la historia digital

Una vez que se lanzó el primer navegador (Mosaic), en 1993, todo cambió. Los exploradores pusieron el internet al alcance de un público muchísimo más amplio y, de ahí en adelante, las interacciones entre humanos y computadoras se volvieron constantes y mucho más tangibles. Esto representa una revolución en el uso de la red, las computadoras personales y de los métodos digitales y computacionales en el análisis histórico.

Este brusco cambio se entiende en el siglo XXI bajo el término de web 2.0, y afecta a cada nivel de la historia y la preservación. En tanto que la primera ola de digitalización se enfocó en preservación y análisis, la segunda se orienta en el acceso y la interacción con estas colecciones digitales, lo que implica una priorización del usuario final en el diseño de los mecanismos de consulta.⁸

El giro digital ha permitido proporcionar acceso a documentos históricos de una manera imprescindible: miles de páginas de periódicos, documentos oficiales, cartas, diarios, libros de cuentas y fotografías se pueden poner al alcance del público que no tiene que pedir permiso para acceder a estas colecciones. Las bibliotecas y archivos tradicionales tienen múltiples formas para proteger sus acervos, y en general son métodos que filtran el acceso a estos. Al construir versiones digitales se dan dos grandes procesos: por un lado, se protegen las colecciones al crear una copia de estas y por el otro, si las réplicas digitales se ponen en línea, no hay límite en el número de consultas que se pueden hacer. El embudo que inevitablemente se genera en la consulta de documentos físicos no existe en la red.

Más allá de los procesos de preservación o acceso, los nuevos métodos digitales también empezaron con la apertura de puertas hacia proyectos innovadores en los que se reinventa, lo cual es un trabajo de historia. Los primeros esfuerzos de esta índole se enfocan no solo en poner archivos a dis-

posición del público, sino también en incluir series historiográficas y cuestionamientos sobre lo que son fuentes históricas y cómo entender la historia de colectividades en el contexto de estudios académicos. El proyecto de Edward Ayers, de la Universidad de Virginia, acerca de dos comunidades durante la Guerra Civil (*The Valley of the Shadow: Two Communities in the American Civil War*⁹) y el *Race and Place: An African American Community in the Jim Crow South*,¹⁰ que fue una colaboración entre el Virginia Center for Digital History y el Carter G. Woodson Institute of African and Afro-American Studies, son excelentes ejemplos de estas nuevas maneras de imaginar cómo poner proyectos de investigación académica al alcance de un público más amplio.

A partir de estos nacieron muchos más que aprovecharon las oportunidades de la web 2.0 para desarrollar un nuevo modelo de investigación histórica. Proyectos como el de la historia de la ciudad de Victoria, Canadá,¹¹ o el *Bracero History Project*¹² presentan una manera de analizar el pasado que no solo se apoya en el archivo histórico, sino que también invita al público a participar como investigador y fuente. En México, la Biblioteca Digital Mexicana¹³ es una iniciativa del Archivo General de la Nación (AGN) y otras instituciones para poner mapas, planos, libros, entre otras fuentes, al alcance de la sociedad en forma digital; además, para lo que no está disponible en línea, el AGN proporciona servicios de digitalización para investigadores.¹⁴ Esta reconceptualización de lo que se entiende como servicio al usuario y de la manera en que lo digital entra dentro de lo histórico es clave para entender cómo el manejo de datos se ha vuelto una necesidad en el trabajo de investigación histórica.

8 Samouelian, Mary. "Embracing Web 2.0: Archives and the Newest Generation of Web Applications", en: *The American Archivist*. Vol. 72, No. 1, Spring-Summer 2009.

9 Ayers, Edward L. *The Valley of the Shadow: Two Communities in the American Civil War*. Valley Archive, 2007 (DE) <https://valley.lib.virginia.edu>.

10 University of Virginia. *Race and Place: An African American Community in the Jim Crow South*. Charlottesville, VA, 2002 (DE) <http://www2.vcdh.virginia.edu/afam/raceandplace>.

11 Lutz, John and Patrick Dunae. *Victoria's Victoria*. University of Victoria, 2002 (DE) <http://web.uvic.ca/vv/index.html>.

12 Center for History and New Media, George Mason University. *Bracero Historic Archive* (DE) <https://braceroarchive.org>.

13 Biblioteca Digital de México. *Acervo histórico* (DE) <http://bdmx.mx/presentacion>.

14 Archivo General de la Nación. *Digitalización de archivos históricos, proceso de trámite*. México (DE) <https://www.gob.mx/tramites/ficha/digitalizacion-de-archivos-historicos-en-el-agn/AGN5757>.

Las bases de datos como fuente del pasado

La disciplina histórica¹⁵ abarca tanto las humanidades como las ciencias sociales y, aunque los bancos de información no se cuentan típicamente entre las fuentes históricas con las que trabaja un historiador, son una gran parte del repositorio con el que se apoya: los censos de población, de agricultura o industria; las cifras económicas; los reportes de gobierno, presupuestales y sindicales; los planes de desarrollo de las compañías de trenes, constructoras de autopistas o de otra infraestructura pública, entre otros, son insumos que se pueden analizar estadísticamente y en agregado.

Para investigadores de la era contemporánea, muchas de estas bases de datos están accesibles en línea y organizadas en archivos descargables a programas y plataformas de análisis estadístico y visualización (como R, Tableau o Stata). Para el historiador, esa información está, en gran parte, contenida en cuadros PDF o en documentos .xls que replican la estructura de las tablas históricas originales. Un sinnúmero de ellos ha trabajado así: transcribiendo tablas impresas en anuarios y copias en PDF y con copias guardadas en *floppy discs*, CD o cualquier otro formato, y muchos otros, sin acceso a otras fuentes, han trabajado con datos disponibles en la página web de *México Máximo*¹⁶ (un esfuerzo loable del ingeniero Manuel Aguirre Botello); sin embargo, al no permitir descargas, el sitio se vuelve poco práctico para el investigador.

Esta falta de acceso a bancos de información en línea resulta en que muchos historiadores trabajan con datos de orígenes diversos, en formatos distintos, sin manera de compartirlos o hacerlos visibles o de transparentar sus métodos de construcción, que además no están al alcance de usuarios en la red o catalogados en una guía en línea.

Una base de datos digital, para poder mantener su integridad en cualquier formato, tiene que estar estructurada de una manera estandarizada, con metadatos que indiquen la metodología, el origen de la información y quién la construyó. Es una labor bastante complicada que requiere múltiples colaboradores con diferentes especialidades (transcripción, manejo de datos y metadatos, infraestructura digital, por contar los básicos). Aquí unos ejemplos de estadísticas históricas en línea:

- Estados Unidos de América (EE. UU.): censo nacional, mantenido desde 1949.¹⁷
- Irlanda: datos estadísticos de 1916, proyecto especial de la oficina de estadísticas del gobierno irlandés.¹⁸
- Argentina: el Instituto Nacional de Estadística y Censos de la República de Argentina incluye series históricas para algunas de sus categorías, pero, en muchos casos, lo que se consideran datos históricos se refieren a los de la década precedente.¹⁹

En general, las series estadísticas que van más allá de 50 años atrás no son fácilmente rastreables en las bases de datos de los órganos estadísticos de cada país que se enfocan en el presente y pasado reciente. En México, los datos históricos, cuando existen, están en tablas PDF y en recónditas esquinas de la página del Instituto Nacional de Estadística y Geografía (INEGI), Banco de México o de cualquier otra organización que ponga material estadístico a disposición de un usuario. Es una lástima, porque debería ser un objetivo de cualquier institución pública que sirva a la sociedad: construir acervos digitales y bases de datos para preservar la información histórica nacional y hacerla accesible y descargable de forma sencilla.

Estamos viviendo en tiempos de grandes retos para la investigación histórica: el cambio climático,

15 Cohen, Daniel and Roy Rosenzweig. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. University of Pennsylvania Press, 2006. // Burg, Steven. *The Future is Here: Public History Education and the rise of Digital History*. National Council on Public History, July 2012 (DE) <https://ncph.org/history-at-work/the-future-is-here-public-history-education-and-the-rise-of-digital-history/>.

16 Aguirre Botello, Manuel. *México Máximo*. 2004 (DE) <http://www.mexicomaximo.org/>.

17 United States Census Bureau. *Historical Statistics, 1789-1945* (DE) https://www2.census.gov/library/publications/1949/compendia/hist_stats_1789-1945/.

18 Central Statistics Office, Republic of Ireland. *Life in 1916 Ireland: Stories from statistics* (DE) <https://www.cso.ie/en/statistics/lifein1916irelandstoriesfromstatistics/>.

19 Instituto Nacional de Estadística y Censos. República Argentina (DE) <https://www.indec.gob.ar/>.

una pandemia global y el constante peligro de recursos limitados son algunas de las variables que exigen de nosotros que imaginemos nuevas maneras de salvaguardar el patrimonio nacional. No es suficiente ni basta con construir nuevos edificios, faltan estructuras y estrategias digitales para almacenar el pasado y darle un futuro. Archivomex es un ejemplo de cómo llevar a cabo dicha estrategia desde los órganos académicos e institucionales que mejor conocen los acervos para promover su uso en colaboración académica y docente en el país, y saber dónde termina el patrimonio nacional relacionado con los datos del pasado.

Segunda parte: construyendo y expandiendo Archivomex

El proyecto se enfocó en transformar una colección de datos históricos clave para cualquier investigador de la historia moderna de México, específicamente relacionados con las EHM, las cuales son una colección de estadísticas que abarca un periodo que empieza a finales del Porfiriato y continúa hasta mediados de 1980. Estas se publicaron en tomos físicos, y fueron escaneados hace algunas décadas, por lo que existen copias en PDF. Algunas tablas también han sido transcritas en Excel. La primera generación de proyectos digitales del INEGI, que fue lanzada al mismo tiempo que inició la institución bajo su primer presidente, Pedro Aspe Armella, en 1983, se enfocó en la colección de datos estadísticos, lo que incluyó digitalizar algunas fuentes originales. Varias fueron subidas a partes de su página web, pero no en ubicaciones dedicadas a la historia, o donde se podrían encontrar fácilmente. Al día de hoy, no existe un acervo aplicado a las EHM que esté en su totalidad en formatos legibles por máquina, descargables y con los metadatos necesarios para satisfacer las necesidades archivísticas del momento, es decir, no tienen su debido lugar en la plataforma del INEGI.

Archivomex integró métodos históricos con ciencia de datos con el fin de generar un nuevo entorno colaborativo para el almacenamiento, la preservación, visualización y difusión de las EHM en

línea. Lo que estamos haciendo en este proyecto las lleva hacia la segunda generación de procesos digitales. En las dos primeras décadas del siglo XXI, el giro digital se ha enfocado en la transformación de las estructuras originales de la información hacia una nueva organización de datos legibles por computadora para facilitar su preservación en diversos formatos de almacenamiento distribuido (a lo que se le llama coloquialmente *la nube*) para que sean visualizados en nuevas aplicaciones web y difundidos por medios sociales en línea. A la fecha, esta evolución en el INEGI se orienta, más que nada, en datos recientemente generados.

En lo que sigue, se explora el inicio y funcionamiento de Archivomex documentando los procesos implementados para poner las EHM en línea y al alcance de todos, así como establecer una nueva plataforma colaborativa e histórica al servicio de historiadores, académicos, maestros y el público en general.

Por ejemplo, el uso de las metodologías de la ciencia de datos para organizar y crear bases de datos modernas fueron clave en el procesamiento de las tablas de las EHM. La siguiente descripción del trabajo de procesamiento de estas viene del segundo reporte técnico del equipo Archivomex entregado al Consejo Nacional de Ciencia y Tecnología (CONACYT):

“En su forma original, cada una de las tablas de la EHM fue organizada como aparecía en la página del texto impreso dificultando cualquier trabajo estadístico. El equipo de captura y visualización de datos tuvo como tarea reorganizar las EHM y, simultáneamente, trabajar en estrategias para simplificar la estructura de los datos y crear metadatos para cada archivo de manera homogénea.

“El proceso paso a paso: Antes de empezar el trabajo, todos los archivos de las EHM se movieron a Dropbox en carpetas representativas de la categoría en la cual existen en las EHM – la nomenclatura de la colección nunca se modificó para mantener la colección en el formato organizacional el más fidedigno posible.

“En simplificado, cada tabla en cada carpeta se revisó y se transformó en base de datos. Un ejemplo ilustrativo de este proceso es como sigue. Primero se tomó uno de los archivos que aparecieron originalmente en las EHM: por ejemplo, en la figura 4 se ve la tabla C1-2-2.xls, de población del estado de Aguascalientes entre los años 1830-1995. En la primera tabla, las celdas copian la estructura de la tabla original. Hay múltiples columnas entre la variable “año” y la variable “población” y hay datos adicionales, pero sin referencias.

“En la nueva base de datos se utiliza una estructura simplificada con dos columnas: la primera variable es el año y la segunda, la población. El nuevo nombre de archivo es el título de la gráfica, que es el tema Aguascalientes_1830-1995.xlsx [ver (figura 1)].

Los metadatos correspondientes detallan toda la información complementaria relacionada con esta tabla, incluidas las notas al pie de página. Esta estructura simplificada permite que la base de datos sea leída y que se modele fácilmente con mínimas modificaciones en otras aplicaciones, como qGIS, R, STATA o Tableau”²⁰

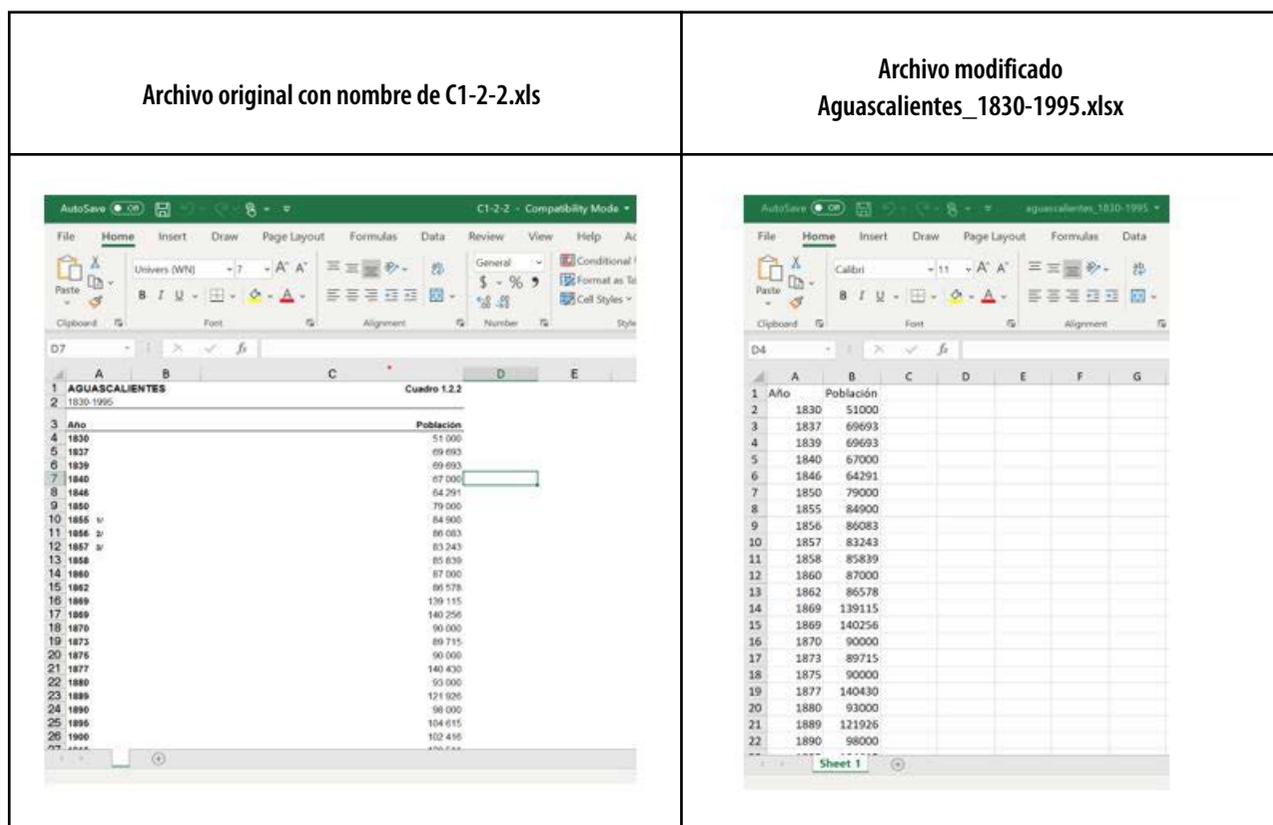
Entorno ambiental: proyectos comparativos

Las universidades de Emory, Michigan State y Columbia, entre otras, han patrocinado importantes iniciativas para la digitalización y preservación de

²⁰ MX.Digital. *Acerca de Archivomex* (DE) <https://mx.digital/4351-2/>.

Figura 1

Comparación de una tabla original de las EHM y la misma información en formato de una base de datos moderna



datos históricos. En esta próxima parte, revisamos algunos antecedentes influyentes para Archivo-mex que destacan por su alcance, metodología y aplicabilidad al trabajo del proyecto.

*Slave Voyages*²¹ es una de las más importantes iniciativas digitales para abordar las trayectorias para construir un programa basado en la preservación y modelado de miles de datos particulares. Su historia y la infraestructura establecida ilustran perfectamente los retos que cualquier proyecto digital histórico va a enfrentar. Sus raíces están basadas en la década de los 60, cuando el historiador Herbert S. Klein, de la Universidad de Columbia, y un equipo de investigadores y académicos acumularon datos no publicados sobre los viajes transatlánticos de esclavos, y decidieron formatearlos en código legible por máquinas. En los siguientes 20 años, ese colectivo había creado 16 distintas bases de datos con 11 mil registros individuales; esto presentó un problema de cimiento: cómo manejar tantos datos a través de tantas plataformas distribuidas. En 1990, David Eltis y Stephen Behrendt, de la Universidad de Emory, propusieron crear una sola base de datos para el tema, que los llevó a la publicación de un CD-ROM de este nuevo archivo digital en 1999. En el 2006, una nueva generación de investigadores en las universidades de Harvard y Emory exploró las posibilidades de presentar la base de datos de *Slave Voyages* en línea. La información del CD-ROM se transfirió a la web y el nuevo archivo dio a luz en el 2008.

Hubo importantes revisiones de código y preservación entre el 2015 y 2018. Es una base de datos enorme y muy importante, y su existencia es un argumento suficiente para seguir trabajando en poner más bancos de información en línea de esa manera. Pero el hecho de ponerlos en la red generó un nuevo proyecto de historia digital. Un equipo de visualización de la revista *Slate* los utilizó para ilustrar visualmente el volumen del transporte de esclavos de África a América sobre un mapa interactivo del océano Atlántico entre 1520 y 1900, dándole al artículo de Jamelle Bouie y Andrew

21 Slave Voyages Consortium. *Slave Voyages v2.2.13*. 2021 (DE) <https://www.slavevoyages.org>.

Kahn un elemento adicional de fuerte impacto que sus palabras solas no podían convalidar.²²

Los proyectos digitales de la División de Historia de la Universidad de Michigan State presentan otros ejemplos de cómo representar y modelar los datos almacenados para hacerlos relevantes a un público en general. Por ejemplo, el *Vietnam Group Archive* es un acervo dinámico de fuentes históricas presentadas en múltiples formatos, desde versiones originales digitalizadas hasta archivos codificados en sistema de información geográfica. El mapa interactivo, escrito en PHP, lleva más de 4 mil visualizaciones de documentos e imágenes, clasificados por año, categoría temática y región de Vietnam. La aplicación ofrece a los usuarios una interfaz fácil de navegar y un camino sencillo para acceder a información histórica sobre un capítulo de la historia que sigue siendo de gran importancia en el presente.

Este ejemplo muestra lo importante que es poner datos no solo en un resguardo adecuado, sino también en un espacio dinámico para su visualización. Este tipo de trabajo requiere que tanto historiadores como científicos de datos y especialistas en su visualización estén en comunicación con programadores y diseñadores de interfaces para crear aplicaciones de web que puedan ser utilizadas en la mayoría de los equipos electrónicos (PC, tabletas, *smartphones*) y que tengan mayor relevancia para un público en general.

Para México, existen iniciativas digitales lanzadas por historiadores que comparten su investigación. Pablo Piccato *et al.*, de la Universidad de Columbia²³ (EE. UU), hicieron una labor excepcional para preservar y digitalizar las estadísticas históricas del crimen en el país. El acervo digital en su sitio web ofrece bases de datos útiles para la investigación, en conjunto con aplicaciones básicas para visualizar esa información. Asimismo, el trabajo de

22 Kayhn, Andrew and Jamelle Bouie. "The Atlantic Slave Trade in Two Minutes. Mapa interactivo", en: *Slate*. 16 de septiembre del 2021 (DE) <https://slate.com/news-and-politics/2021/09/atlantic-slave-trade-history-animated-interactive.html>.

23 Piccato, Pablo, Sara Hidalgo y Andrés Lajous. *Estadísticas del crimen en México: Series históricas, 1926-2008*. Princeton University (DE) <https://ppiccato.shinyapps.io/judiciales/>.

Carlos Marichal, de El Colegio de México,²⁴ acerca de la historia bancaria y financiera nacional y de América Latina es un acervo importante desde la época colonial hasta el siglo XX (ambas colecciones están almacenadas en páginas en Excel y PDF). La existencia de los dos proyectos apunta a la viabilidad de trabajar con datos históricos de México, así como a las amplias fuentes disponibles, pero también señala la necesidad de unificar las formas de estos acervos para que sean más fácilmente accesibles, ubicables y descargables.

Tercera parte: cómo preparar un proyecto histórico de digitalización

Para llevar a cabo uno como Archivomex —que tiene como meta la accesibilidad, seguridad y máxima utilidad de la información que contiene la base de datos— fue necesario elaborar una infraestructura técnica; asimismo, se formaron recursos humanos para fomentar nuevos vínculos entre las disciplinas de historia y ciencia de datos. El acervo digital que se construyó está dedicado a las *Estadísticas históricas de México* y está definido por un marco estandarizado y simplificado para la elaboración individual de bases de datos y sus metadatos. En esta sección se plantea el proceso técnico para la creación de los nuevos bancos de información basados en datos históricos.

Estandarización, captura y visualización de datos

Uno de los principales retos para el estudio de las EHM es que la gran mayoría de tomos disponibles están en el formato PDF u otros anticuados (CD-ROM) que limitan su utilidad. No hay soluciones fáciles para manipular datos en formatos de archivo heredados, se tuvieron que reconstruir estos en nuevos, como CSV, XSL, entre otros. También, se necesitó un significativo trabajo de modificación porque muchas tablas en las EHM no estaban or-

ganizadas de manera estándar y, para uso en aplicaciones de revisión y visualización de datos, se tuvieron que hacer manualmente. No hay manera de hacer que este proceso sea rápido, se tiene que realizar a mano, en formato *long* con un oficio acompañante de metadatos para categorizar cada elemento de la tabla; esto es una seria inversión de tiempo y en gente y, sin duda, es uno de los aspectos más exigentes de la digitalización y modernización de datos históricos.

Para ampliar la captura estandarizada de las estadísticas históricas, hay que implementar métodos claros para historiadores, asistentes de investigación y otros participantes. Una de las tendencias típicas al inicio de la creación de una base de datos es copiar el formato y organización del texto como aparecen en la página impresa. Esto es un error crítico porque limita la capacidad para que esta información se procese en aplicaciones de almacenamiento y visualización.

A lo largo de la historia del estudio de datos históricos, la comunidad profesional ha producido enormes cantidades de archivos que se pueden calificar como idiosincráticos, es decir, archivos que funcionan para el trabajo del investigador individual, pero que, por su organización particular, limitan su difusión y utilidad a otras personas. Por eso, un proceso de orientación y capacitación es clave para enseñar las características de un archivo original y cómo procesarlo a una versión más útil y estandarizada, además de tener un espacio para contestar dudas sobre casos especiales encontrados en los datos.

El proyecto se manejó con tres equipos:

- El primero, para el procesamiento manual de los archivos originales a la nueva estructura de las bases de datos en formato *long* y la elaboración de metadatos. Este trabajo implicó la revisión individual de cientos de archivos para entender su estructura con la necesidad de llevar a cabo ensayo y error para determinar la mejor manera de reorganizar los datos en un formato más sencillo y legible a computadoras.

²⁴ Marichal, Carlos. *Research Projects*. El Colegio de México (DE) <https://carlosmarichal.colmex.mx/en/index-en.html>.

- El segundo trabajó con estos nuevos archivos, revisándolos para errores de dedo, guardándolos en distintos formatos (.txt, .csv, .xlsx, .pdf) con sus metadatos y subiendo todo al acervo histórico del Banco de Información del LNPP. Juntos, estos trabajaron como equipos A y B para asegurar la uniformidad de los datos, compartiendo retroalimentación y eliminando errores.
- El tercero fue de programación, encargado de la creación de aplicaciones web para visualizar los datos y también facilitar su procesamiento. Crearía productos en aplicaciones como Tableau y ArcGIS y programaría aplicaciones web de código abierto en Python, Node.js y Apache, entre otros, para ampliar el acceso a las nuevas bases de datos de las EHM.
- **Segundo:** la presentación del acervo digital al público; si el sitio web es difícil de navegar o encontrar los datos implica que estos no están disponibles para la gran mayoría de posibles usuarios: el diseño de la infraestructura digital para el usuario final es igual de importante que el de la base de datos misma.

La historia de la creación de las EHM corre una trayectoria paralela a las iniciativas mencionadas en la sección anterior. A partir de la década de los 80, una primera generación de historiadores económicos y otros académicos implementaron nuevas recolecciones de datos y lanzaron investigaciones que resultaron en la publicación de los primeros tomos de las EHM. Fue un importante primer paso en la organización y consolidación del conocimiento histórico-económico del país. En la década de los 90, esta información fue procesada en el formato CD-ROM, preservando su estructura original, pero migrándola al entorno digital. Luego, trabajos adicionales continuaron la preservación de las estructuras originales de los tomos publicados en físico, consolidándose en archivos maestros con todos los datos disponibles en PDF o pestañas distintas en Excel. En muchos casos, el acceso a esta información fue restringido debido a los límites intrínsecos de ciertos formatos, como CD-ROM y PDF, o porque los archivos fueron almacenados en sitios web difíciles para navegar y descargar.

Infraestructura digital y usuario final

Uno de los enfoques principales de Archivomex fue elaborar un programa de capacitación y, al mismo tiempo, crear y utilizar nuevas herramientas para simplificar el proceso de captura y procesamiento.

En este contexto, cabe mencionar que hay dos retos principales para la creación de bases de datos digitales:

- **Primero:** se debe asegurar que la infraestructura digital sea adecuada para almacenar los datos a largo plazo (un mínimo de cinco a 10 años) en colaboración con especialistas de ciencia de datos y con la necesidad de tener un plan para la continuación del proyecto después de la etapa inicial; en el caso de los *Slave Voyages*, la investigación pasó por muchas manos institucionales antes de llegar a la infraestructura necesaria para garantizar la preservación de su acervo digital; esta historia influyó en la creación de Archivomex con el fin de colaborar con instituciones en México y el extranjero para preservar los datos en distintos formatos.

El objetivo del proyecto Archivomex fue superar estos límites. Se hizo necesario crecer el entorno discursivo, invitando a los historiadores a colaborar con especialistas de ciencia de datos, sistemas de información geográfica y programación para repensar la organización de las estadísticas históricas y construir las herramientas para ampliar su acceso al público.

Imaginamos la creación de un nuevo acervo digital basado en la elaboración de archivos distintos para cada base de datos, guardados en diversos formatos, apoyándose en los metadatos para facilitar su organización y elaborar nuevas visua-

lizaciones de estos en línea. En conjunto con este trabajo, propusimos la creación de nuevas herramientas abiertas de *machine learning* (aprendizaje de máquina) para facilitar el procesamiento de los datos a la nueva organización estructural. Esta iniciativa particular, dentro de la coyuntura de Archivomex, fue un proyecto piloto para entender las características gráficas de la información numérica en las EHM para semiautomatizar su procesamiento y establecer una prueba para la creación de futuras herramientas para procesar otras fuentes históricas, incluyendo la transformación de los anuarios estadísticos guardados en PDF a nuevos archivos legibles por máquina.

Con apoyo del Fondo Sectorial INEGI-CONACYT, Archivomex inició como proyecto de investigación en el 2019. El trabajo implicó el reclutamiento de profesores de universidades y centros públicos de investigación en México, la creación de equipos de investigadores y asistentes de investigación, trabajando en Ciudad de México y Aguascalientes, y la selección de ciertas tecnologías que iban a respaldar toda su infraestructura técnica.

La transición de una propuesta a un proyecto implementado requirió involucrarse con distintos lenguajes profesionales de especialistas para unir nuestra conceptualización de cómo organizar y presentar las estadísticas históricas. Por ejemplo, la primera discusión principal con los científicos de datos del Laboratorio Nacional de Políticas Públicas del CIDE se enfocó en la necesidad de crear un acervo dedicado a datos históricos. En lugar de distribuir todas las bases de datos de las EHM a lo largo de acervos temáticos ya establecidos en el Banco de Información del Laboratorio, se decidió abrir un acervo histórico independiente con la finalidad de simplificar el acceso a esta información para historiadores.

Este esfuerzo llevó a la elaboración de un nuevo entorno de colaboración para que diversos equipos de especialistas puedan contribuir a modernizar los bancos de información de las EHM para que sean legibles a computadoras y ampliar su acceso con nuevas aplicaciones web diseñadas para inves-

tigadores y el público en general. El resultado ha sido positivo, y consideramos que Archivomex es prueba de que más proyectos históricos son factibles.

El objetivo de largo plazo de Archivomex es invitar a más historiadores y otros especialistas a compartir nuevas fuentes y bases de datos para procesarlas e incorporarlas en la red de colaboración establecida. Las EHM son solo la primera serie de datos en su acervo digital histórico. El proyecto ha abierto un espacio de participación multidisciplinaria que identificó brechas y confluencias en los distintos vocabularios y prioridades profesionales y académicas de los diferentes grupos de especialistas involucrados, y los llevó a reconstruir una base de datos para poner a las EHM en un formato accesible y de uso en las plataformas más comunes en estos inicios del siglo XXI.

Conclusión

Las metas de Archivomex se han cumplido: las EHM están en línea, en una estructura de base de datos que es descargable y manejable en cualquier formato, con documentos de metadatos que acompañan cada base. El proyecto se cumplió bajo circunstancias excepcionales, en gran parte porque estuvo organizado desde su inicio con transparencia y flexibilidad, porque estuvo imaginado como un instrumento de consulta en línea y entre el equipo de trabajo, a cada nivel, se mantuvieron abiertas las vías de comunicación; asimismo, fue muy poco jerarquizado: el líder, los profesores que colaboraron y los asistentes que contribuyeron, todos, eran parte del mismo equipo y, en gran parte, lo que requiere un proyecto digital de esta índole es mucha cooperación, coordinación y comunicación, además de un mínimo de jerarquía; de igual manera, el entorno digital está en flujo constante: todas las plataformas cambian, el código se modifica y las necesidades del sistema, así como del usuario final, también. El aferrarse a un proceso establecido o a funciones específicas dentro del grupo, o a títulos, no iba a coadyuvar en el flujo de trabajo, y este es lo que se privilegió en cada momento.

Nuestra labor en las EHM se ha completado, pero queda mucho por hacer para que las colecciones de datos históricos estén al alcance de la comunidad; hay un sinnúmero de estas que se deben actualizar y llevar por un proceso de digitalización de esta índole, no basta con poner paginas Excel sobre una de web personal, lo que el acervo histórico de México necesita y se merece es un acervo completo, organizado mediante metadatos y con

facilidad de descarga para que los públicos general, docente y de investigación tengan acceso a la historia del país.

Otra fuente

Library of Congress (EE. UU.). *American Memory Project* (DE) <http://memory.loc.gov/ammem/>.