

Matias Miguel

Centro de Documentación e Información Educativa,
Dirección General de Cultura y Educación

La digitalización de impresos como garantía de acceso al patrimonio educativo bonaerense.

Proyecto de preservación de la revista *Anales de la Educación Común*

Resumen

Se expone el proyecto de digitalización de la colección de revistas *Anales de la Educación Común* que lleva adelante el Centro de Documentación e Información Educativa de la provincia de Buenos Aires. El objetivo principal es optimizar la calidad de la colección digital con el fin de garantizar su preservación a largo plazo y mejorar la accesibilidad para la comunidad usuaria. Se describen las actividades desplegadas, tales como la investigación de estándares internacionales y el contacto formal con actores institucionales claves en el campo de la digitalización del patrimonio histórico.

Palabras clave

DIGITALIZACIÓN, PATRIMONIO DOCUMENTAL, CONSERVACIÓN Y PRESERVACIÓN DE DOCUMENTOS, PROVINCIA DE BUENOS AIRES.

Introducción

El presente artículo está enmarcado en el proyecto de digitalización y preservación de la colección de revistas *Anales de la Educación Común* que se lleva adelante en el Centro de Documentación e Información Educativa (CENDIE),¹ perteneciente a la Dirección General de Cultura y Educación (DGCyE) de la provincia de Buenos Aires. La edición permanente de esta publicación periódica, es com-

¹ La gestión del proyecto está a cargo del Departamento de Documentación, coordinado por el jefe departamental Javier Peón dentro de diversas acciones y líneas de trabajo propias de la cadena documental del CENDIE.

petencia (delegable) del Director o la Directora General de Cultura y Educación (inciso d del artículo 69, Ley Provincial de Educación de la provincia de Buenos Aires n° 13.688).

No faltan razones para implementar un proyecto de digitalización de la revista. Dentro de las funciones principales que conlleva la digitalización de patrimonio histórico, la International Federation of Library Associations and Institutions (IFLA) menciona incrementar el acceso, mejorar la calidad del servicio para la comunidad usuaria, y reducir la manipulación de los impresos, ya frágiles. Se debe agregar la función de preservación del material digitalizado y prolongación de la vida útil del recurso de información (Pichinini & Pené, 2008).

La revista resulta relevante para quien tenga la intención de investigar la historia de la educación y de la gestión educativa bonaerenses. Según un trabajo de Bracchi y otros, desde su fundación, concretada por Domingo Faustino Sarmiento en 1858, es fuente básica de consulta y “patrimonio documental de gestión oficial en materia de políticas públicas educativas, criterios de gestión, propuestas innovadoras, trabajos de investigación, fuentes estadísticas, designaciones docentes, entre otros” (Bracchi, Vazelle, Deldivedro & Gabbai, 2006, pp. 1-2). Se concluye también en ese trabajo, la importancia de la preservación de las revistas como patrimonio bibliográfico. Bracchi como directora del CENDIE, junto a la plantilla laboral activa de la primera década del milenio, dieron inicio a una reconstrucción histórica y al proyecto de la primera digitalización de la publicación por medio de cámaras digitales.² Esta colección digitalizada en forma completa en formato PDF se distribuyó y difundió por medio de cuatro DVD. Es interesante releer el texto de Bracchi et al. (2006), también para destacar algunos hitos, dentro de las “continuidades y rupturas” que atraviesan la publicación de *Anales de la Educación Común*, así como sus cambios de denominación, estilo y perfil a lo largo de más de 150 años. A su vez, entre sus reflexiones finales se reconoce la necesidad de su permanente edición. Cabe destacar que lleva casi veinte años de publicación desde la última reedición en noviembre de 2005.

A partir de 2020, año complejo atravesado por la pandemia de COVID-19, es posible consultar gran parte de la revista de manera *online*. La colección digitalizada se gestiona por medio de la plataforma *Open Journal System*, software que permite llevar adelante la edición y difusión de los números actuales. A través de esta política, el Centro buscaba facilitar el acceso de la comunidad a la información, en momentos en que era obligatorio el aislamiento, favorecer la visibilidad a través de la indización en sistemas de información y la recuperación mediante buscadores web.³

² Digitalización concretada en la gestión del Director del CENDIE Jorge Levoratti, en 2009.

³ En aquel momento se encontraba como secretario de redacción de la revista Matías Maggio-Ramírez.

Dentro de los principales lineamientos disponibles, se debe mencionar la presencia de la IFLA con las Directrices para planificar la digitalización de colecciones de libros impresos antiguos y manuscritos, del año 2014. Allí se menciona que la digitalización de colecciones antiguas y especiales requiere de consideraciones singulares y distintos modos de actuación. Estas directrices puntualmente se detienen en la planificación conceptual del proceso. En cuanto a la preservación a largo plazo de la colección digital, las bibliotecas deben tener las colecciones digitales en alta resolución en servidores distribuidos, y contar con los procesos y los sistemas necesarios para supervisar la integridad de los archivos digitales a lo largo del tiempo. Asimismo, es recomendable generar copias almacenadas en diferentes lugares geográficos.

En la actualidad, el proyecto de digitalización y preservación tiene entre sus objetivos mejorar la calidad de imagen de los documentos digitalizados, para asegurar la preservación a largo plazo de la colección digital y el acceso global. En ese sentido, se decidió iniciar el proceso de digitalización a partir de los volúmenes en soporte papel que resguarda el CENDIE, e incorporar el reconocimiento óptico de caracteres (OCR). La mejora se da gracias a la dotación de un escáner aéreo y de su incorporación al flujo de trabajo. Este equipo cuenta con un nivel de calidad de imagen intermedio para nuestro propósito de recrear virtualmente, para la comunidad usuaria, la materialidad de la *Revista Anales*. En un paso previo e ineludible, se realiza un proceso de conservación y preservación a partir de la limpieza y desinfección de todos los documentos físicos, y luego de desinsectación de algunos ejemplares afectados por el paso del tiempo, mediante la exposición prolongada a baja temperatura. Finalmente, a futuro, otro objetivo integral del proyecto será su traslado a un repositorio digital permanente, según directrices IFLA ya mencionadas.

Desarrollo. Flujo de trabajo

La primera actividad del proyecto fue observar el contexto en relación a digitalizaciones de colecciones de publicaciones antiguas y/o especiales similares, existentes en otras instituciones. Por un lado, se realizó una lectura de las directrices y guías a nuestro alcance, y por otro, se observó cómo han trabajado y qué resultados obtuvieron otras unidades de información, en proyectos logrados y/o en curso, así como la lectura de las comunicaciones de sus avances.

Además de las directrices IFLA anteriormente citadas, se consultaron las *Technical Guidelines for Digitizing Cultural Heritage Materials* (editadas y revisadas en 2023), que desarrolla y publica la Federal Agencies Digital Guidelines

Initiative (FADGI).⁴ Estos lineamientos, disponibles en idioma inglés, se utilizaron para conocer y orientar los criterios estandarizados de digitalización más actuales en cuanto a resolución de la imagen, perfil de color, y formato de los archivos máster de preservación, así como de pautas para los posteriores documentos derivados de acceso. A su vez, FADGI desarrolló un “sistema estelar”, que se condice con la norma ISO-19264 en la cual se define a la calidad de imagen en tres instancias, A, B y C. Las directrices de FADGI, hacen corresponder esta norma ISO con un sistema en el cual otorga cuatro estrellas para “A o Mejor nivel”, tres para “B o Nivel bueno”, y dos para “C o Nivel aceptable”. Adicionalmente otorga una estrella para imágenes que no se ajustan a la norma ISO-19264.

Se debe agregar que a mediados de 2024, personal del CENDIE mantuvo una reunión con integrantes del Servicio de Digitalización perteneciente al Servicio de Difusión de la Creación Intelectual (SEDICI) de la Universidad Nacional de La Plata (UNLP). Este servicio pionero en el campo del acceso abierto en nuestro contexto, implementó su repositorio en el año 2003, con la finalidad de difundir y preservar la producción intelectual de la universidad (Nusch, Calamante, Fernández, & De Giusti, 2021). A partir de ese encuentro, se sugirió trabajar para la obtención de imágenes de *tres estrellas* en la escala de los lineamientos FADGI en al menos 200 dpi,⁵ y todo ello en escala de colores.

Planificación. Selección

En la composición del acervo bibliográfico del CENDIE, la colección es significativa en tanto es la principal documentación que contiene la memoria histórica del sistema educativo bonaerense desde sus orígenes. Conviene decir que, si bien la colección no está completa, el número de ejemplares existentes es importante. Uno de los primeros pasos que iniciamos a partir de lo observado en las diferentes directrices fue la selección de los documentos originales a digitalizar. En principio, el objetivo planteado es la digitalización de la colección resguardada en el CENDIE. No obstante, será necesario completar la colección con el acceso a los volúmenes faltantes a partir de la colaboración por parte de distintas organizaciones. Cabe mencionar, que para la primera digitalización que se llevó a cabo entre 2007 y 2009, fue necesario previamente, relevar otras instituciones que cedieron en préstamo estos volúmenes.

En cuanto al estado de conservación de la revista, el grado de deterioro que tienen las encuadernaciones, en algunos casos, limita su manipulación y obliga a

⁴ Las directrices de FADGI (2023) fueron diseñadas en consonancia con la International Organization for Standardization (ISO), responsable del desarrollo de la norma ISO 19246 que define la métrica para la digitalización profesional de imágenes del patrimonio cultural en todo el mundo.

⁵ Los “Dots Per Inch” o “Puntos por pulgada” son una medida de resolución de impresión.

pensar consideraciones extra al digitalizarlas. Asimismo, se presentó en ciertos volúmenes un problema adicional, ya que poseen hojas desplegadas en tamaños variados que superan el común de las páginas de las revistas, y que a su vez presentan dobleces y roturas que dificultan su manipulación.

Captura, equipo y parámetros

Luego de la incorporación de un escáner aéreo modelo *SV600 Scansnap* de la marca *Fujitsu*, se decidió digitalizar en una resolución de 200 dpi, y con escala de colores. El formato elegido fue el PDF sin compresión (para no afectar la calidad de la imagen digital), ya que los materiales, en general, no requieren un alto nivel de detalle por parte del usuario, situación que nos llevaba a utilizar un formato TIFF para el máster de preservación.

Conviene aclarar, que si bien con el *Scansnap* es posible escanear libros y publicaciones impresas encuadernadas, este no es un escáner diseñado para digitalizar específicamente ese tipo de material. En ciertas partes del proceso resulta difícil capturar la imagen, tanto por la dificultad de mantener aplanado el material original como por el tiempo que demora en realizarse la captura, razones por las cuales el procedimiento pierde cierta eficacia. Existen en el mercado escáneres diseñados para esos formatos, los cuales resultan muy costosos de acuerdo a los recursos económicos con que se cuenta. Por otro lado, hay iniciativas globales colaborativas que utilizan escáneres caseros realizados con estructuras de madera y cámaras digitales mucho más accesibles y con excelentes resultados, como las del Servicio de Digitalización PREBI-SEDICI en la UNLP de la ciudad de La Plata en la provincia de Buenos Aires, cuyas colecciones digitalizadas se pueden consultar en el repositorio institucional de la universidad. Otra opción existente es la de seleccionar el mejor ejemplar y guillotinar las hojas para digitalizarlas en un escáner de cama plana (Pichinini & Pené, 2008; IFLA, 2014). Se descartó esta modalidad porque en la colección de revistas antiguas contamos con ejemplares únicos, que impiden su realización.

El trabajo de escaneo dio comienzo al desarrollo de ensayos cuyos resultados consideramos que se encuentran algo alejados de los esperados, con imágenes de una o dos estrellas en la escala FADGI. El artefacto ofrece la posibilidad de corregir la inclinación de las líneas de texto y encuadrar las imágenes de manera automática pero no fue posible conseguir escaneos óptimos por esta vía, al menos no con material encuadernado. Ante la imposibilidad de lograr la calidad de imagen esperada las pruebas continuaron y se ensayaron diferentes formas y flujos de trabajo. Progresivamente se consiguieron unos resultados preliminares deseables para nuestro objetivo. Se podría afirmar que se logró adaptar el uso del escáner al material a digitalizar de una forma alternativa a la indicada por el fabricante en su

manual de procedimiento original. De esta forma, en un primer momento se escanean las páginas de cada revista en sentido creciente, para luego girar la revista en su eje sagital y tomar el escaneo de las páginas en sentido decreciente. En la carpeta de destino -previamente creada y nombrada en un directorio de la computadora- queda guardado un documento en formato PDF por cada página con un número con el que se identifican y ordenan las imágenes escaneadas.

Posprocesamiento y presentación de la colección

Luego de escanear cada número y con el objetivo de ofrecer objetos digitales de libre acceso para la comunidad usuaria, la tarea consiste en unir todas las páginas y ordenarlas en un documento PDF multipágina que será visible al público. Se utiliza la aplicación ABBYY *Fine Reader* para tal fin, con su herramienta 'Convertir a PDF' y luego con la pestaña 'Organizar páginas'. Hay una intervención mínima que se realiza sobre las imágenes a las cuales se aplica la opción 'Enderezar las líneas de texto' junto con el mencionado reconocimiento óptico de caracteres. El producto derivado de este trabajo es un documento que permite recrear una experiencia más cercana a la lectura del original e incluso ir más allá, puesto que posibilita buscar y hallar palabras a texto completo. Respecto del ordenamiento de las páginas, este es un paso obligado producto del modo alternativo que se estableció para escanear los documentos, primero en un sentido (páginas en forma creciente) y luego en el otro (páginas en forma decreciente).

El archivo resultante del posprocesamiento es un PDF multipágina más liviano que el conjunto de los archivos máster. Para dar una idea, si un fichero consta de 90 elementos y su tamaño total ronda los 185 megabytes, el documento derivado va a ser comprimido a 100 megabytes aproximadamente. Perseguimos así el objetivo de que aquellas usuarias y aquellos usuarios en línea que acceden a la colección no dependan de la calidad de su conexión a internet para cargar documentos demasiado pesados, que terminarían por entorpecer la consulta.

Registro del progreso y control de calidad

El flujo de trabajo se acompaña de un mecanismo de registro de los procedimientos. Por medio de una planilla de cálculo se registra el trabajo al tiempo que se ejecuta, y así dejar constancia del avance y orientar el proceso. Simultáneamente, se realiza el control de calidad de cada imagen escaneada con el fin de detectar discrepancias que requieran repetir el escaneo de alguna página y de no volver a hacer el proceso de digitalización.

Nombramiento y almacenamiento de archivos

Los nombres de cada imagen incluyen, en parte, una serie de metadatos que se autoreferencian. Se decidió emplear el sufijo “aec” (Anales de la Educación Común) para unificar, dadas las diferentes denominaciones que la revista contó en su historia. Se incluyó también el año de publicación y el número de la revista. El nombre se completa con el orden de la secuencia de escaneos. A modo de ejemplo:

aec-1899-1-0001 *primera página*

aec-1899-1-0002 *segunda página*

aec-1899-1-0003 *tercera página*

A su vez, el mismo esquema se emplea para nombrar los directorios que alojan los objetos digitales, en primer lugar una carpeta por cada número y otra que engloba todos los números de cada año.

En relación al almacenamiento, de toda la documentación relevada y analizada, se observa que no hay directrices, guías o lineamientos que no adviertan sobre la necesidad de preservar las imágenes digitalizadas en distintos soportes –como el disco rígido de una computadora– y sobre la redundancia de copias en diferentes ubicaciones geográficas. Por otra parte, se proyectó usar DVD para grabar copias de las carpetas con las digitalizaciones, grabados a baja velocidad y almacenados posteriormente en lugares de resguardo seguros, frescos y lejos del sol.

Cabe recordar las recomendaciones de la IFLA (2014) en cuanto a la preservación a largo plazo de la colección digital. En ese sentido se estableció una política que incluye distintas estrategias de almacenamiento y difusión. Entre ellas, trasladar las copias máster de preservación y sus derivadas multipáginas al repositorio institucional.

Metadatos

Los metadatos hacen posible varias funciones clave, tales como la identificación, la gestión, el acceso, el uso y la conservación de un recurso digital (FADGI, 2023). Se prevé en un futuro que el repositorio institucional aloje diversas colecciones digitales, además de la revista *Anales de la Educación Común*. Por cada colección se deberá desarrollar y adoptar un perfil de aplicación de metadatos propio y los conjuntos de elementos se adaptarán al requerimiento de cada material. La asignación de metadatos a los objetos digitales se realizará por medios externos incorporados como campos de una base de datos relacional (Marzal García-Quismondo, Calzada Prado & Cuevas Cerveró, 2006). Para el caso puntual de la

colección de revistas, se pensó un perfil de aplicación que combina campos del esquema *Dublin Core* con campos *ad hoc*, obligatorios y no obligatorios, para agilizar la carga de documentos al repositorio, y para dotar de los metadatos suficientes para la recuperación del objeto digital.

Avances preliminares y reflexiones finales

El objetivo fundamental del proyecto es asegurar el acceso permanente a la colección de revistas. En ese sentido, su despliegue se suma al anhelo de gestiones previas por poner al alcance de la comunidad usuaria la revista en línea, en buena calidad de imagen y recuperable a texto completo, y así favorecer más lecturas, más investigaciones de sus páginas, más citas de sus aportes, y de generar más conocimiento a partir de esta fuente de información.

El ejercicio de evaluación comparativa de otras instituciones con el fin de identificar sus mejores prácticas (*benchmarking*) desplegado por el equipo, supuso en un principio la observación de diferentes repositorios y bibliotecas digitales líderes en el universo de la preservación de patrimonio intelectual impreso, con el objetivo de incorporar criterios y técnicas de calidad externas en una búsqueda constante de adaptar y mejorar los procesos y productos internos. Asimismo, salimos a la búsqueda de contactos formales con instituciones clave en el sector, tales como el SEDICI, pero también con el Archivo Histórico “Dr. Ricardo Levene” de la provincia de Buenos Aires. Comparar sus resultados habilitó el descubrimiento de cuáles podrían ser las mejores prácticas para nuestro caso particular.

Finalmente, en cuanto a resultados preliminares, persisten algunos inconvenientes como las páginas desplegadas. A su vez, en algunos máster de preservación se observa distorsión y rotación, a sabiendas de que las directrices determinan que “los archivos de imagen maestra deben estar rectos en relación con el contenido de la imagen”, aunque ahora permiten aplicar la rotación de imágenes (FADGI, 2023). Este problema se resuelve al menos en la etapa de posprocesamiento con *ABBYY Fine Reader* que mejora considerablemente la calidad en las copias derivadas de lectura.

La experiencia del CENDIE en la digitalización de *Anales de la Educación Común* sirve como oportunidad de construir una memoria institucional histórica de la labor realizada. Con este proyecto como antecedente, se puede considerar además, un aprovechamiento por otras instituciones que resguardan el patrimonio histórico y desean ponerlo al alcance de su público, así como instancias de intercambio institucional. Por eso, se presenta el recorrido, con sus aciertos y complejidades propias de la labor. La gestión del proyecto de digitalización

implica la creación de un nuevo producto de información, con renovadas funciones y mejor calidad orientadas a la comunidad. La propuesta busca abonar el terreno de la memoria educativa bonaerense, al hacer más accesibles las colecciones que reflejan su devenir histórico y social.

Referencias bibliográficas

- Bracchi, C., Vazelle, M., Deldivedro, V. & Gabbai, M. (2006). *Revista de Educación: Cuando la historia se hizo revista*. XIV Jornadas Argentinas de Historia de la Educación, 9 al 11 de agosto de 2006, La Plata, Argentina. En Memoria Académica. https://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.13228/ev.13228.pdf
- Federal Agencies Digital Guidelines Initiative. (2023). *Technical Guidelines for Digitizing Cultural Heritage Materials. Third edition*. https://www.digitizationguidelines.gov/guidelines/FADGITechnicalGuidelinesforDigitizingCulturalHeritageMaterials_ThirdEdition_05092023.pdf
- Federal Agencies Digitization Guidelines Initiative. (2009). *Digitalization activities. Project Planning and Management Outline*. <https://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf>
- International Federation of Library Associations and Institutions. (2014). *Directrices para planificar la digitalización de colecciones de libros impresos antiguos y manuscritos*. <https://repository.ifla.org/handle/20.500.14598/455>
- Maggio-Ramírez, M. (2021). Anales de la Educación Común, una nueva plataforma para más lecturas. *Anuario de Bibliotecas, Archivos y Museos Escolares*, 1(1), 38-47. <https://cendie.abc.gob.ar/revistas/index.php/abame/article/view/606/2658>
- Marzal García-Quismondo, M. A., Calzada Prado, J., & Cuevas Cerveró, A. (2006). Desarrollo de un esquema de metadatos para la descripción de recursos educativos: El perfil de aplicación MIMETA. *Revista Española de Documentación Científica*, 29(4), 551-571. <https://doi.org/10.3989/redc.2006.v29.i4.308>
- Nusch, C. J., Fernández, E. C., Calamante, L., & De Giusti, M. R. (2021). ¿Cómo y por qué digitalizar? El Servicio de Digitalización PREBI-SEDICI: Marco institucional y características técnicas. *Publicaciones de la Asociación Argentina de Humanidades Digitales*, 2, e021. <https://doi.org/10.24215/27187470e021>
- Pichinini, M., & Pené, M. G. (2008). Guía para la digitalización de documentos impresos y preservación de las imágenes digitales. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Biblioteca de Humanidades

Profesor Guillermo Obiols. Memoria Académica. <http://www.memoria.fahce.unlp.edu.ar/libros/pm.4571/pm.4571.pdf>

Sistema Nacional de Documentación Histórica. (2019). *Guía general de digitalización de documentos*. https://www.argentina.gob.ar/sites/default/files/guia_general_de_digitalizacion_de_documentos_vf.pdf

Williman, M., & Sanllorenti, A. M. (2017, 2 de noviembre). Una experiencia de digitalización masiva en la Biblioteca Digital de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires. Acceso a la información: uso e impacto. 15ª Jornada sobre la Biblioteca Digital Universitaria, Ciudad Autónoma de Buenos Aires, Argentina. http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.10453/ev.10453.pdf