

La experiencia del repositorio institucional SEDICI (UNLP) en la digitalización de obras antiguas: un ejemplo sobre tomos de la colección Cervantina de la biblioteca Pública de la UNLP

Autores: Esteban C. Fernández¹, Lorenzo Calamante¹, Carlos Nusch^{1,2,3}, Marisa De Giusti^{1,2}, Martín Calabrese³, Gloria Chicote³

1 Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Argentina

2 Centro de Servicios en Gestión de Información, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina

3 Instituto de Investigaciones en Humanidades y Ciencias Sociales, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata-Conicet, Argentina

Área Temática: Estrategias y herramientas de preservación digital, planes de preservación

Resumen:

En el presente trabajo se detallan los procedimientos llevados a cabo por el personal del Servicio de Difusión de la Creación Intelectual para la preservación en formato digital de algunos volúmenes pertenecientes a la Colección Cervantina custodiada por la Biblioteca Pública Dardo Rocha. La Colección Cervantina ha sido declarada patrimonio documental por la UNESCO e incluye ediciones del siglo XVII, 75 de los cuales ya forman parte del Programa Memorias del Mundo.

Para la digitalización se siguieron los estándares y directrices internacionales para la digitalización (IFLA, NARA, FADGI, etc.). Sin embargo, muchas particularidades de esta colección no estaban contempladas en la bibliografía y fue necesario aplicar técnicas y procedimientos específicos.

Se digitalizaron inicialmente 9 ediciones (27 tomos), cuyo estado general de conservación era bueno: presentaban algunas manchas propias de la calidad de la tinta y el papel utilizados entre los siglos XVII y XIX. En dos casos, la ilustración exigió modificaciones en la disposición de las luces y se presentaron algunas roturas en las tapas y lomos. La apertura de los ejemplares también generó inconvenientes por su encuadernación apretada y márgenes reducidos. Las anteriores características hicieron necesario el uso de un escáner con cámara cenital y una base adaptada para el sostenimiento de los libros que había sido construido anteriormente para el escaneo de materiales delicados. Para la captura de imágenes se utilizó el software digiCamControl 2.0.72.0,

para la edición ScanTailor Advanced y para el reconocimiento de caracteres y guardado final en PDF/A ABBY Fine Reader 12 Professional. El resultado fue la obtención de documentos de imágenes de alta calidad, destinados a la preservación digital y tras versiones de los mismo documentos comprimidos destinados a la difusión en el repositorio, ambos en formato PDF/A de modo que posibilitaran la búsqueda, el copiado y la transformación del texto por el usuario final y la extracción e indexación.

Palabras Clave: Colección Cervantina; Patrimonio documental UNESCO; Universidad Nacional de La Plata; Preservación Digital; Escáneres DIY;

La colección Cervantina

La Biblioteca Pública Dardo Rocha custodia una Colección Cervantina declarada patrimonio documental por la UNESCO; se trata de una colección de diferentes obras que incluye ediciones del siglo XVII, 75 de las cuales ya forman parte del Programa Memorias del Mundo («Memory of the World», 2017). Varios especialistas en el tema han estudiado y considerado a estas obras como joyas (Mangiaterra & Chicote, 2005; Megías, 2005). Así mismo, alrededor de 55 ediciones del *Quijote* forman parte del Banco de Imágenes del *Quijote* («QBI - Banco de imágenes del Quijote», s. f.), un portal web que se ha encargado, bajo la dirección de Lucía Megías, de digitalizar la iconografía de las ediciones conservadas en diferentes bibliotecas a nivel global. La importancia de la obra de Cervantes ha hecho que las ediciones de sus textos acompañen y recorran prácticamente todas las variantes y tendencias de la industria editorial desde el siglo XVII, dando cuenta, en los ejemplares conservados, de ese desarrollo editorial y de sus características en diferentes espacios y momentos de la historia.

El Catálogo, Aventuras del Quijote en la UNLP da cuenta de la peculiaridades de la conformación de la colección, que se configuró como un proyecto institucional para nutrir a la Biblioteca de un valioso acervo bibliográfico. El 18 enero de 1887 se crea la Biblioteca Pública Provincial, para que puedan albergarse allí diversos archivos que posibiliten a la comunidad científica realizar investigaciones de todo tipo. A partir de ese año, comienza a adquirir ejemplares de libros de variadas índoles, incluyendo obras raras y antiguas. En 1905, por convenio entre la nación y la provincia, la biblioteca pasa a formar parte de la recientemente creada Universidad Nacional de La Plata. En el acto inaugural, Joaquín V. González, Ministro de Instrucción Pública, sienta las bases de la nueva universidad fijando su horizonte en la investigación y el desarrollo académico. Una gran parte de la colección se debe a la figura de Luis Ricardo Fors, director de la biblioteca entre 1898 y 1906, intelectual y bibliófilo español, que fue encargado de nutrir la Colección Cervantina hasta alcanzar en 1906 los 255 ejemplares, de gran valor e incluso piezas únicas. El objetivo del directorio de Fors era llegar al tercer centenario del *Quijote* con una

colección considerable, una edición local de la obra y una serie de actos celebratorios.

Intelectuales de la talla de Amado Alonso han dado cuenta en diversas oportunidades de la importancia que a nivel internacional tiene esta colección, que ha permitido a una gran cantidad de investigadores realizar sus estudios en condiciones bastante similares a las que otorgan importantes bibliotecas de España y Europa (Alonso, 1934).

Entre las obras se destacan por su antigüedad, como la *Primera Parte del Ingenioso Hidalgo don Quexote de la Mancha*, publicada en Bruselas por Huberto Antonio en 1617, adquirida en 1901 o *Vida y hechos del Ingenioso Cavallero don Quixote de la Mancha*, “corregida e ilustrada con diferentes estampas muy donosas y apropiadas a la materia”, publicada también en Bruselas por Juan Mommaerte en el año 1662. Se destaca también la edición financiada por Real Academia Española de 1780, a cargo de Joaquín Ibarra, impresa en Madrid y la de la Imprenta Real, editada por Andrés Ponce de Quiñones en 1797 (primera parte) y 1798 (segunda parte).

La colección cuenta con 30 ediciones en castellano del Quijote, 4 en italiano, 16 en francés, 7 en inglés, 5 en alemán, 1 en catalán, 1 en holandés, 1 en ruso y 1 en serbio. Todas pertenecientes a diferentes períodos y con diferencias observables en cuanto a calidad de impresión.

Se encuentran también en la colección otras obras de Cervantes como la edición parisina de la *Galatea* de 1611, publicada en París; las *Comedias y Entremeses* impresa por Antonio Marín en Madrid en 1749; ediciones de las *Novelas Ejemplares*, en castellano, francés e inglés; la *Historia de los trabajos de Persiles y Sigismunda*, publicada en 1763 y por último el *Viaje al Parnaso*, publicado en 1784.

La importancia de preservar la colección Cervantina

Los repositorios institucionales por su propia constitución deben asegurar la preservación y el acceso a largo plazo, si bien la mayor parte de los contenidos con los que trabaja un repositorio institucional como SEDICI o CIC Digital ya han nacido en formato digital, existen otros materiales que requieren un proceso de digitalización de calidad y siguiendo un procedimiento de trabajo normalizado (Nusch, De Giusti, Fernández, & Lira, 2019). La Universidad Nacional de La Plata cuenta con 17 unidades académicas con sus bibliotecas, la biblioteca pública y una red de museos que aloja una ingente cantidad de obras físicas, en papel e incluso objetos, de gran valor patrimonial y pasibles de ser digitalizados.

En el programa Memoria del Mundo (Edmondson, 2002) de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) se remarca la importancia del patrimonio documental como memoria colectiva y

documentada de los pueblos, y la necesidad de que éste sea preservado y protegido. La sola preservación del material en formato digital no es suficiente, sino también su accesibilidad equitativa y su distribución, como la de sus productos derivados sea amplia y sirva para generar conciencia acerca de la importancia del patrimonio documental (Foster, Russell, Lyall, & Marshall, 1995, p. 5). Del mismo modo que realizar copias de seguridad de los archivos no es hacer preservación digital (Pérez Pineda, 2017, p. 3) el mero hecho de escanear un documento no asegura la correcta preservación y difusión a largo plazo. El proceso de digitalización del patrimonio documental de una institución como la Universidad Nacional de La Plata implica procedimientos complejos que involucran a personal de diferentes áreas y unidades académicas trabajando en pos del cuidado de materiales delicados cuyo daño o pérdida serían enormemente lamentables.

En el ámbito de los archivos y bibliotecas, la conservación implica todas aquellas medidas destinadas a proteger adecuadamente los documentos, con el fin de prolongar su utilización en condiciones óptimas durante el mayor tiempo posible (Pérez Pineda, 2017, p. 7). En el caso del patrimonio documental de una universidad o institución educativa o de investigación, el proceso de digitalización debe contemplar la conservación del soporte de la obra debido a su importancia como objeto y documento histórico (un libro antiguo, un incunable, etcétera) además de lo concerniente a la preservación del documento digital.

Particularidades de los volúmenes a ser escaneados y de las herramientas

Se procedió en principio con la digitalización de 9 ediciones que comprendían en total 27 tomos. En la digitalización de estas obras, el escáner **Modelo Archivista 2014** no pudo utilizarse, porque estas obras tienen un tamaño que excede las dimensiones de la base o cuna (35cm x 25cm). Se resolvió entonces utilizar el escáner cenital, construido anteriormente para el escaneo de materiales delicados (Nusch et al., 2018a, 2018b) al que se le hicieron modificaciones para conseguir libertad de manipulación, confeccionando un nuevo soporte que permitiera una apertura de 150° aproximadamente, y balancear el peso de ambos lados de modo que la encuadernación no sea forzada.

Se dotó al escáner de cámaras Nikon D5600: este tipo de cámara reflex permite el intercambio de lente. En la actualidad se cuenta con un lente del tipo zoom 18-55mm. También cuenta con formato de sensor DX de 24,2 Megapíxeles que permite obtener documentos escaneados a 400dpi y respetar los estándares de digitalización internacional como los de la Library of Congress (Library of Congress, s. f.).

El uso del escáner cenital exigió que se tomara de a una página por vez, por lo que primero fueron capturadas las páginas impares en orden creciente y luego las pares, en orden decreciente. El software utilizado para la captura de imágenes fue digiCamControl 2.0.72.0, el cual permite controlar las cámaras de manera remota en tiempo real desde la PC (foco, distancia, velocidad de obturación, luz, color).

La ausencia de cristal en la cuna del escáner permitió una mejor manipulación y un mejor control sobre el movimiento del papel, sin embargo, al no tener cómo que presionar y alisar las páginas, fue necesario sostenerlas por los bordes y hubo entonces que recortar la imagen a un tamaño menor que el original.

El escáner cuenta con dos luces LED dicróicas de luz cálida, cuya temperatura no daña el material dispuestas hacia el objeto que se captura. Estas luces tienen varias ventajas con respecto a las luces halógenas e incandescentes utilizadas en el pasado. Consumen hasta un 90 % menos, no generan calor considerable y tienen un Índice de reproducción cromática [CRI](#)¹ alto (FADGI, 2016: 17). Sin embargo, en algunos casos, sobre todo en tapas, contratapas y algunas ediciones completas la tinta y el papel provocaron un reflejo que obligó a modificar la posición de las luces para corregir su ángulo de incidencia. Por último, se presentaron algunos problemas particulares cuando se trabajó con cada uno de los ejemplares:

1.- Muchos de los libros presentaban los lomos deteriorados y en el caso de la edición de Londres de 1742 (Lf-154), por esta razón se decidió directamente no incluirlo (Cervantes Saavedra, Tonson, Tonson, & Dodsley, 1742).

2.- La edición de Edimburgo de 1879 (Lf-190) permanecía intonsa, por lo que hubo que recurrir al área de restauración de la biblioteca antes de proceder con la digitalización. A su vez, las hojas tenían tamaños disímiles por lo que hubo que recortar la imagen un tanto más que el tamaño original (Cervantes Saavedra, 1879).

3.- Las ediciones de Barcelona de 1880 (Lf-165) y de las láminas del *Quijote del Centenario* (Lf-281) presentaban una tinta y un papel en las que se reflejaba la luz del escáner. Para resolverlo, se utilizó otras lámparas, que se colocaban a los costados, de modo que la luz no incidiera directamente sobre el libro (Cervantes Saavedra, 1905).

4.- Las láminas de la edición de Barcelona de 1880 (Lf-165) tenían unas hojas protectoras que -por causas que no se han determinado- en algunos casos se pegaron a la imagen, se tuvo entonces que esperar a que en el área de

¹ El índice de reproducción cromática (IRC) es la medida utilizada en relación a una fuente de luz para medir su capacidad de mostrar los colores de un objeto de manera "real". Esto tomando como referencia la iluminación natural.

restauración se retirara con metilcelulosa el papel pegado para proceder con la digitalización.

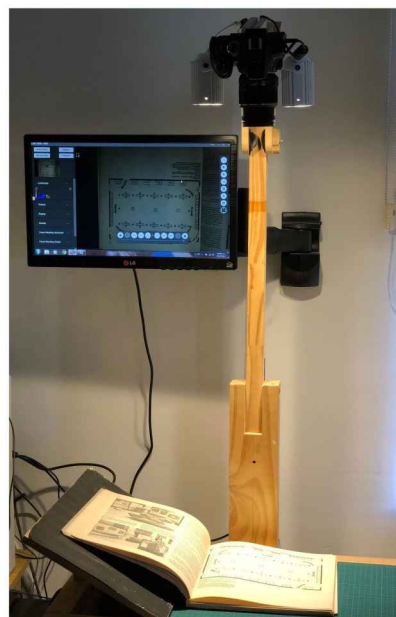
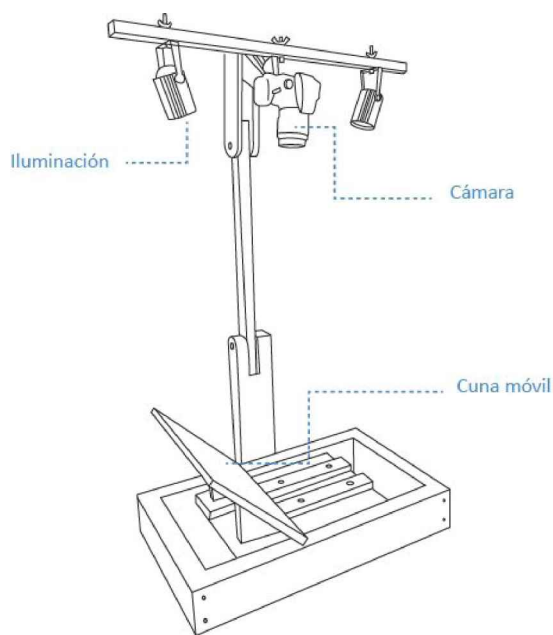


Figura 1: Modelo de escáner con cámara cenital

Procesos realizados para el tratamiento de la Colección Cervantina

Las capturas se realizaron en la sala museo de la Biblioteca pública y se debió prestar especial atención a que estuvieran capturadas todas las páginas, puesto que la edición se realizó posteriormente en la oficina de SEDICI: inmediatamente después de la captura del libro se controló primero las páginas impares, luego las pares y se ordenaron y renombraron con el software de gestión de archivos **Total Commander**, para finalizar se verificó nuevamente que estuvieran todas las páginas.

Posteriormente, en la oficina de SEDICI, se procedió con la edición de las imágenes en **ScanTailor Advanced**, manteniendo su color original y recortándolas sin agregarle márgenes. Este programa procesa las imágenes, allí se definen los DPI (puntos por pulgada), se desinclinan las imágenes se corrigen los márgenes.

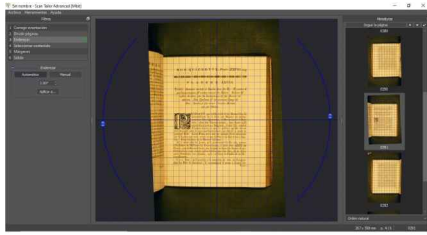


Figura 2: Herramienta “Enderezar” de ScanTailor Advanced.



Figura 3: Herramienta “Seleccionar contenido” de ScanTailor Advanced.

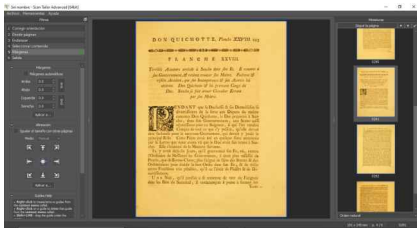


Figura 4: Herramienta “Márgenes” de ScanTailor Advanced.

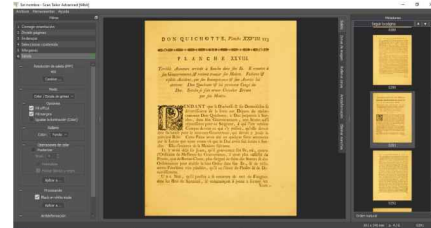


Figura 5: Visualización de la imagen de salida en ScanTailor Advanced.

Para el OCR se utilizó **ABBY FineReader 12**: el mejoramiento de las imágenes y los caracteres tipográficos estuvo enfocado en proporcionar la mejor calidad posible para el proceso de reconocimiento de texto. El material debió dividirse en dos tandas; se seleccionaron las primeras 250 páginas, que se trataron como un solo conjunto en un PDF.

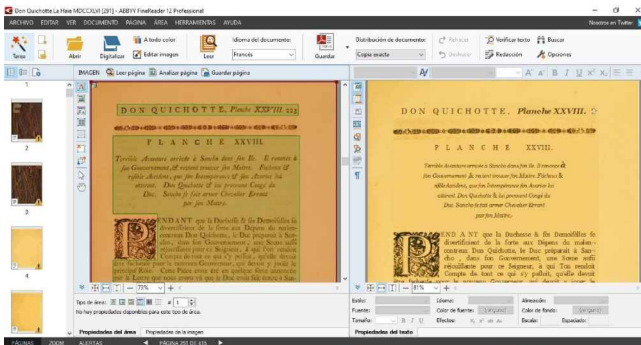


Figura 6: Proyecto de OCR de ABBY Fine Reader 12.

Conclusiones:

Si bien se consultaron los estándares internacionales para la digitalización, las pruebas realizadas con las Joyas Cervantinas demostraron que con el procedimiento estándar no se obtenían los mejores resultados. Se hizo necesario adaptar los procesos de captura y tratamiento de las imágenes por las particularidades del material a digitalizar. El escáner con cámara cenital permitió una manipulación cuidadosa de las páginas y garantizó la protección de la encuadernación. También se realizaron diferentes pruebas de filtros y enfoque digital de imágenes para obtener la mejor calidad posible de imágenes, de acuerdo a criterios de lectura, visibilización, difusión y preservación digital. Para el procesamiento del OCR realizado con ABBY Fine Reader 12 Professional los requerimientos de recursos y memoria de la PC también fueron importantes.

Se generaron dos archivos .pdf de cada uno de los tomos: Uno de alta calidad de imagen (de aproximadamente 500 Mb), que se destinó a la preservación digital y otro comprimido de 200 Mb, que se utilizó para la difusión en el repositorio. En ambos casos, el OCR permitió la búsqueda, el copiado y la transformación del texto por el usuario final y la extracción e indexación.

Bibliografía

- Alonso, A. (1934). La sección cervantina de la biblioteca. *Boletín de la Universidad Nacional de La Plata*, XVIII(4), 196-205.
- Edmondson, R. (2002). *Memoria del Mundo: Directrices para la salvaguardia del patrimonio documental*. UNESCO.
- Foster, S., Russell, R., Lyall, J., & Marshall, D. (1995). *Directrices para la salvaguardia del patrimonio documental*. Recuperado de https://unesdoc.unesco.org/ark:/48223/pf0000105132_spa
- International Organization for Standardization (ISO). (2015). *UNE-ISO 14721:2015 Sistemas de transferencia de datos e información espaciales. Sistema abierto de información de archivo (OAIS). Modelo de referencia*. Recuperado de <https://www.aenor.com/normas-y-libros/buscador-de-normas/iso/?c=062542>
- International Organization for Standardization (ISO). (2017). *UNE-ISO 16363:2017 Sistemas de transferencia de información y datos espaciales. Auditoría y certificación de repositorios digitales de confianza*. Recuperado de <https://www.aenor.com/normas-y-libros/buscador-de-normas/iso/?c=062542>
- Library of Congress. (s. f.). Technical Information. Recuperado 2 de octubre de 2018, de Technical Information, Library of Congress website: <https://memory.loc.gov/ammem/about/techIn.html>
- Mangiaterra, N. E., & Chicote, G. (2005). *Aventuras del Quijote en la UNLP*. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/61575>
- Megías, J. M. L. (2005). Un paseo por las joyas cervantinas de la Biblioteca Pública de la Universidad Nacional de La Plata. En N. E. Mangiaterra & G. Chicote, *Aventuras del Quijote en la UNLP* (pp. 19-32). Recuperado de

- <http://sedici.unlp.edu.ar/handle/10915/61575>
- Memory of the World. (2017, marzo 16). Recuperado 18 de octubre de 2019, de UNESCO website: <https://en.unesco.org/programme/mow>
- Nusch, C., De Giusti, M. R., Fernández, E. C., & Lira, A. J. (2019, julio 30). *Rumbo a la autoevaluación: La elaboración de un Manual de Procedimientos de Digitalización*. Presentado en IX Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'19) (São Paulo, 2019). Recuperado de <http://sedici.unlp.edu.ar/handle/10915/78722>
- Nusch, C., Fernández, E. C., Méndez Moura, P., Reichenbach, M. C. von, Cozzuol, A., Santamaría, M., & De Giusti, M. R. (2018a, octubre 12). *Digitalización y Preservación Digital del patrimonio museológico de la Universidad Nacional de La Plata: Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)*. Presentado en XXVIII Asamblea General del ISTEAC 2018 - Congreso Internacional “La influencia de la tecnología en las comunidades del conocimiento” (La Paz, Bolivia, 2018). Recuperado de <http://sedici.unlp.edu.ar/handle/10915/69958>
- Nusch, C., Fernández, E., Méndez Moura, P., Reichenbach, M. C. von, Cozzuol, A., Santamaría, M., & De Giusti, M. R. (2018b, octubre 25). *Particularidades y desafíos en la digitalización del libro Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925): Un caso de uso de un repositorio institucional para la preservación de Memoria Institucional*. Presentado en VIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales BIREDIAL-ISTEC (Lima, 2018). Recuperado de <http://sedici.unlp.edu.ar/handle/10915/70355>
- Pérez Pineda, I. (2017). *Plan de preservación digital del patrimonio documental digitalizado de la Biblioteca Lafragua*. Recuperado de <http://diposit.ub.edu/dspace/handle/2445/117724>
- QBI - Banco de imágenes del Quijote. (s. f.). Recuperado 19 de octubre de 2019, de <http://qbi2005.windows.cervantesvirtual.com/>