

# **Digitalización y Preservación Digital del patrimonio museológico de la Universidad Nacional de La Plata: Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)<sup>1</sup>**

Carlos J. Nusch<sup>1, 2</sup> Esteban C. Fernández<sup>1</sup> Pablo Méndez Moura<sup>1</sup> Cecilia von Reichenbach<sup>3, 4</sup> Ana Elisa Cozzuol<sup>3</sup> Mariana Santamaría<sup>3</sup> Marisa R. De Giusti<sup>1, 2</sup>

1. Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual (SEDICI - PREBI) de la Universidad Nacional de La Plata
2. Centro de Servicios en Gestión de Información (CESGI) de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)
3. Museo de Física del Departamento de Física de la Facultad de Ciencias Exactas Universidad Nacional de La Plata
4. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

## **Resumen**

En el presente trabajo se detallan los procedimientos llevados a cabo por el personal del Servicio de Difusión de la Creación Intelectual y del Museo de Física de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata para la preservación en formato digital de un documento archivístico: el libro *Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)*. Se siguieron los estándares internacionales para la digitalización (IFLA, NARA, FADGI, etc.) sin embargo muchas de las dificultades que presentó el material no estaban contempladas en la bibliografía. El volumen escaneado, con especiales condiciones materiales requirió el diseño de técnicas y procedimientos específicos. El estado de conservación presentaba las siguientes particularidades: papel amarronado, débil, friable con roturas y desprendimientos, tinta difundida en el papel y transferida a los vecinos, pérdida de nitidez en la escritura e imagen como vista a través de un espejo, escritura mecanografiada poco legible. El estado de la encuadernación, con especiales requerimientos de manipulación hizo necesario el diseño *ad hoc* de un escáner con cámara cenital y con brazos especiales para sostener el libro. El procedimiento estándar utilizado para la captura de imágenes y pos procesamiento debió modificarse dejando de lado el firmware modificado (CHDK) y el Scan Tailor. En su lugar se utilizó el digiCamControl 2.0.72.0 para la captura de imágenes, el Photoshop CS6 para el procesamiento de las imágenes capturadas y el ABBY Fine Reader 12 Professional para el OCR. Finalmente, se obtuvieron dos documentos: uno de alta calidad de imágenes destinado a la preservación digital y otro comprimido destinado a la difusión en el repositorio, ambos en formato pdf de modo que posibilitarán la búsqueda, el copiado y la transformación del texto

---

<sup>1</sup> Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

por el usuario final y la extracción e indexación por parte de crawlers o repositorios.

## **Introducción:**

La plasticidad y potencialidad de los repositorios institucionales los ha convertido en un valioso recurso a la hora de difundir y preservar no solo la producción intelectual sino también el patrimonio de las instituciones, ya sean universidades, bibliotecas o archivos. El motivo de este trabajo es la descripción de un caso de preservación en formato digital de un documento archivístico. Se trata del libro *Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)* cuyo procesamiento requirió la intervención del personal del repositorio central de la Universidad Nacional de La Plata (SEDICI) y del personal del Museo de Física de la Facultad de Ciencias Exactas<sup>2</sup> de la misma casa de estudios. Los libros copiadores son el único registro que se conserva de las notas administrativas, de gestión, docencia e investigación enviadas desde el Instituto de Física<sup>3</sup>. Los mismos complementan las publicaciones oficiales (Memorias y Anuarios) y constituyen un reservorio de datos documentales para el estudio de la historia de la Física en La Plata. Al cumplirse 100 años de la Reforma Universitaria de 1918, el Museo de Física consideró oportuno recuperar la historia del Instituto de Física y la Facultad de Ciencias Físicas, Matemáticas y Astronómicas, a través del estudio del libro *Copiador* que abarca los años 1918-1925. La digitalización del libro con especiales condiciones materiales requirió el diseño de técnicas y procedimientos específicos así como de un escáner hecho a medida, todo lo cual, como se explica más adelante, constituye la originalidad de este trabajo.

## **El libro que se quiere preservar: *Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)***

Los libros copiadores constituyen un corpus que evidencia las relaciones científicas, institucionales, políticas y aún personales de los miembros de la incipiente comunidad científica de las épocas fundacionales y, a diferencia de las publicaciones oficiales, están escritos y firmados por los propios protagonistas de la física de entonces. En sus páginas se incluyen conceptos sobre el proceso de institucionalización de la ciencia en Argentina, las dificultades propias de la labor en esta actividad pionera (relacionadas con la falta de insumos para experimentación, problemas de importación de instrumentos, dificultades para realizar las tareas docentes y de investigación, etc.), las relaciones existentes entre los físicos locales y figuras destacadas en la escena internacional (como las cartas de invitación al premio Nobel Walter

---

<sup>2</sup> Anteriormente Facultad de Ciencias Físicas, Matemáticas y Astronómicas.

<sup>3</sup> El Instituto de Física fue fundado, conjuntamente con la U.N.L.P., en 1905, integrándose a la misma bajo la dirección del Observatorio Astronómico. Fue el primero de su tipo en Latinoamérica, respondiendo a la idea positivista de fines del siglo XIX y a la concepción de Universidad esbozada por Joaquín V. González. Es el antecesor del Departamento de Física actual.

Nernst<sup>4</sup> en 1914), dificultades internas (notas de denuncias por problemas entre docentes, que devinieron en problemas institucionales), evidencias de los primeros vínculos científico tecnológicos en física (correspondencia entre Emil Bose<sup>5</sup> y Jorge Newbery<sup>6</sup>), por citar algunas.

Se trata de ejemplares ordenados cronológicamente que contienen copias de documentos cuyos originales no se conservan, con las firmas de los directores y secretarios del Departamento de Física entre los años 1908 y 1948. Las copias se obtuvieron por transferencia de parte de la tinta del documento original al libro *Copiador*.

El interés en esta colección, en particular por los libros comprendidos entre los años 1908 á 1933, no sólo atañe a los investigadores en la historia de la física en Argentina por su importancia, sino también a investigadores locales e internacionales dedicados a la historia de la ciencia en América Latina, que por medio de ellos logran acceder a nuevas fuentes de información no reveladas hasta ahora.

### **Tareas y estado de conservación**

El Museo de Física conserva un total de diecinueve libros copiadores del Departamento de Física. Su encuadernación es rígida y formada por bloques de cuadernillos cosidos, separados en dos secciones, índice y copias, cuyos papeles son de diferentes calidades y gramajes; las hojas están foliadas; presentan escrituras manuscrita y mecanográfica y las tintas utilizadas son ferrogalotánica<sup>7</sup> y de cinta mecanográfica copiativa.

El detalle de los más antiguos es el siguiente:

- I.- “Copiador - Instituto de Física - Años 1908 a 1913”; consta de 499 hojas de copiador, 485 de ellas con registro de escritura manuscrita color marrón; idiomas Español y Alemán.
- II.- “Copiador - Sección Geofísica – Años 1911 a 1912”; consta de 495 hojas de copiador, 58 de ellas con registro de escritura manuscrita de color marrón; idiomas Español y Alemán.
- III.- “Copiador - Instituto de Física – Años 1912 a 1918”; consta de 1000 hojas de copiador, 865 de ellas con registro de escritura manuscrita de color marrón, 102 de escritura mecanografiada de colores celeste y violeta, firmas manuscritas de color marrón; idioma Español.

---

<sup>4</sup> El físico y químico alemán Walther Hermann Nernst Görbitz (Briesen, Prusia, 25 de junio de 1864 - Ober-Zibelle, Alemania, 18 de noviembre de 1941), fue premio Nobel de Química en 1920 por sus teorías sobre el cálculo de la afinidad química como parte de la tercera ley de la termodinámica.

<sup>5</sup> El físico alemán Emil Hermann Bose (Bremen el 20 de octubre de 1874 - La Plata en mayo de 1911) fue el primer Director del Departamento de Ingeniería Eléctrica de la Universidad Nacional de La Plata.

<sup>6</sup> El ingeniero Jorge Alejandro Newbery (Buenos Aires, 27 de mayo de 1875 - Mendoza, 1 de marzo de 1914) fue un aviador, deportista, funcionario y hombre de ciencia argentino que fundó la Aeronáutica Militar Argentina.

<sup>7</sup> La tinta ferrogálica/ferrogalotánica es de reacción química; su trazo, de color negro, se observa luego de un tiempo de efectuado por reacción química de los componentes.

**IV.-** “Copiador - Facultad de Ciencias Físicas, Matemáticas y Astronómicas – Años 1918 a 1925”; consta de 500 hojas; 8 de ellas con registro de escritura manuscrita de color azul y marrón en un mismo trazo y 474 restantes de escritura mecanografiada de colores azul y violeta, firmas manuscritas de color azul y marrón en un mismo trazo; idioma Español.

**V.-** “Copiador - Años 1925 a 1928”; consta de 496 hojas con registro de escritura mecanografiada de colores azul y violeta, 4 hojas con registro de escritura manuscrita de colores azul y marrón o negra en un mismo trazo, firmas manuscritas; idioma Español.

**VI.-** “Copiador - Instituto de Física – Años 1928 (1927) á 1930 (1933)”; consta de 500 hojas con registro de escritura mecanografiada de colores verde y violeta y firmas manuscritas de color marrón; idioma Español.

El deterioro en los libros anteriores se debe principalmente a la calidad de los materiales que los componen (tales como el papel ácido de las secciones de índices y la tinta ferrogálica de la escritura manuscrita) y probablemente al proceso deficiente de transferencia de las tintas de los originales al copiador.

Consecuentemente el papel de las páginas de los índices se observa oscurecido, manchado y debilitado, mientras que el papel de las copias con tinta ferrogálica presenta roturas, friabilidad y pequeños faltantes. Ambas escrituras manifiestan pérdida de nitidez; en el caso de la ferrogálica por difusión y transferencia de la tinta en el papel además del cambio característico de color al marrón; mientras que en la mecanográfica se debe al desvanecimiento del color de la tinta o a la transferencia insuficiente de la misma.

En el año 2006 se observó que los tres primeros libros, con registros manuscritos, se encontraban en un estado de conservación muy delicado. Desde entonces, para ralentizar los procesos de deterioro y favorecer sus permanencias, se evitó manipularlos y se mantuvieron protegidos en condiciones ambientales y de guarda adecuadas y controladas<sup>8</sup>. Simultáneamente se inició la duplicación de ellos para permitir el acceso a la información que contienen mediante fotografía digital, tarea que continúa el SEDICI.

---

<sup>8</sup> Se contienen en una caja construida con materiales de calidad conservación, en condiciones de temperatura y humedad estables, que se controlan periódicamente.

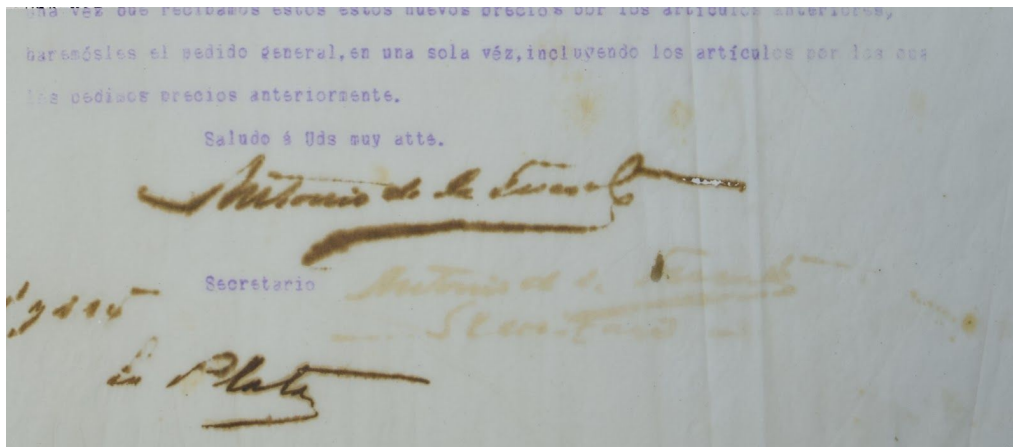


Figura 1. Se observa el papel amarronado y con dobleces. La tinta ferrogálica de la firma se ha oxidado, difundido en el papel carcomiendo una parte por reacción química. También se observa transferencia de texto de la hoja anterior. La tinta azul mecanografiada presenta desvanecimiento y pérdida de nitidez.

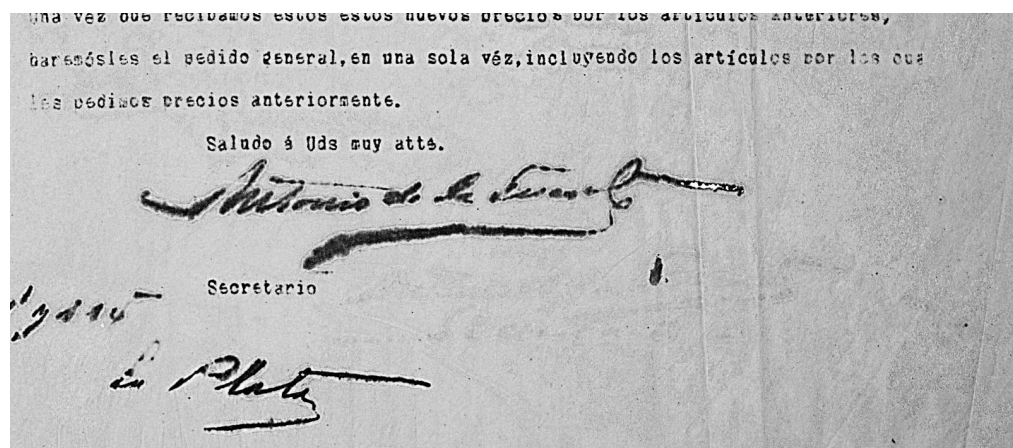


Figura 2. El postproceso y la aplicación de filtros permite una mayor claridad para el reconocimiento de texto.

### ¿Por dónde empezar?

El libro *Copiador* que abarca los años 1918-1925 es el único documento de archivo de esos años que se conserva en el Museo y es fuente de primera mano de los movimientos administrativos del Instituto de Física.

Como parte de la Universidad Nacional de La Plata que mantiene y defiende las conquistas de los reformistas, el Museo quiso hacer su aporte en esta fecha para celebrar nuestra Universidad pública, gratuita y extensionista. Con motivo del Centenario de la Reforma Universitaria se realizó una campaña virtual por redes sociales, que contó a través de una serie de publicaciones diarias: qué fue la Reforma, qué pasaba en La Plata y qué impacto tuvo en la Facultad de Ciencias Físicas, Matemáticas y Astronómicas a través de sus principales protagonistas. Para esto se requirió la participación interdisciplinaria de varios integrantes del Museo, especialistas en las siguientes disciplinas: Física, Historia, Conservación y Museología. Pero las condiciones del libro requerían una manipulación delicada y un tratamiento especial para poder digitalizarlo, con dos objetivos: su preservación física y su preservación digital.

En esta instancia se recurrió al SEDICI, con quienes el Museo trabaja colaborativamente desde 2014.

### **El repositorio institucional SEDICI y las colecciones de preservación**

El Servicio de Difusión de la Creación Intelectual (SEDICI) es el repositorio institucional central de la Universidad Nacional de La Plata y su misión es albergar, preservar, difundir y dar visibilidad a toda la producción científica e intelectual de la universidad con el objeto de aumentar la visibilidad y por lo tanto el impacto institucional.

SEDICI se encuentra organizado en diferentes comunidades, subcomunidades y colecciones, de acuerdo a las unidades académicas de la universidad y el tipo de materiales albergados, además, el repositorio cuenta con subcomunidades y colecciones especiales destinadas a destacar y agrupar trabajos e información relevante para la Universidad y la región ([Red de Universidades con Carreras en Informática \(RedUNCI\)](#), [Academia Nacional de Agronomía y Veterinaria \(ANAV\)](#), [Emergencia Hídrica](#), [Recursos Educativos Abiertos](#), [Radio Universidad](#) entre otras). Una subcomunidad especial dedicada a la preservación del patrimonio es la correspondiente a la [Red de Museos de la UNLP](#) que agrupa documentos de suma importancia tales como los catálogos Colección Max Kohl o textos del físico Tebaldo Ricaldoni procedentes del [Museo de Física](#), así como diferentes libros y registros de instrumentos musicales que incluyen grabaciones de sonido del [Museo de Instrumentos Musicales "Dr. Emilio Azzarini"](#), archivos pertenecientes a la Colección Sala "Alejandro Korn" y a la Colección Sala "Joaquín V. González" pertenecientes a las [Salas Museo de la Biblioteca Pública](#). El libro *Copiador* de la Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925) fue alojado en la subcomunidad Libros y publicaciones de la subcomunidad correspondiente al Museo de Física.

### **Los modelos de escáneres usados para la Digitalización en SEDICI**

En Agosto de 2012 se firmó un Acuerdo entre Wikimedia Argentina y SEDICI con el objeto de realizar un programa de asistencia y colaboración recíprocas para la ejecución de tareas orientadas a la transferencia de tecnologías de digitalización, capacitación en digitalización y disponibilización virtual de contenidos. En el marco de dicho acuerdo se recibieron capacitaciones para el manejo del escáner DAL (Derecho A Leer), así como al proceso de trabajo de digitalización de obras y contenidos y temas referidos a la compatibilidad de la propiedad intelectual de los contenidos con las licencias Wikimedia Commons. El escáner DAL fue desarrollado por el grupo Derecho a Leer para la Facultad de Filosofía y Letras en julio de 2011. Inspirados en la experiencia de *Do It Yourself* desarrollaron un escáner para libros de forma artesanal y muy económica, mediante madera y cámaras digitales estándares para ofrecer soluciones de digitalización de acervos bibliográficos a bajo costo. El primer escáner DAL armado en 2012 para SEDICI se basó en el Model 1 de DIY y en el primer scanner DAL de la Universidad de Buenos Aires. En 2017 fue armado segundo tipo basado en el modelo Archivista 2014 de DIY que

presentaba varias ventajas: mejora en el soporte del libro, mucho más firme y con la cama móvil, mejora en el soporte de las cámaras, menos flexible que el modelo anterior pero más estable, la iluminación con lámparas LED y mejor fidelidad en la captura de colores, cámaras reflex de mayor calidad con lentes intercambiables.



Figura 3. Primer DAL de SEDICI basado en el *Model 1* de DIY.

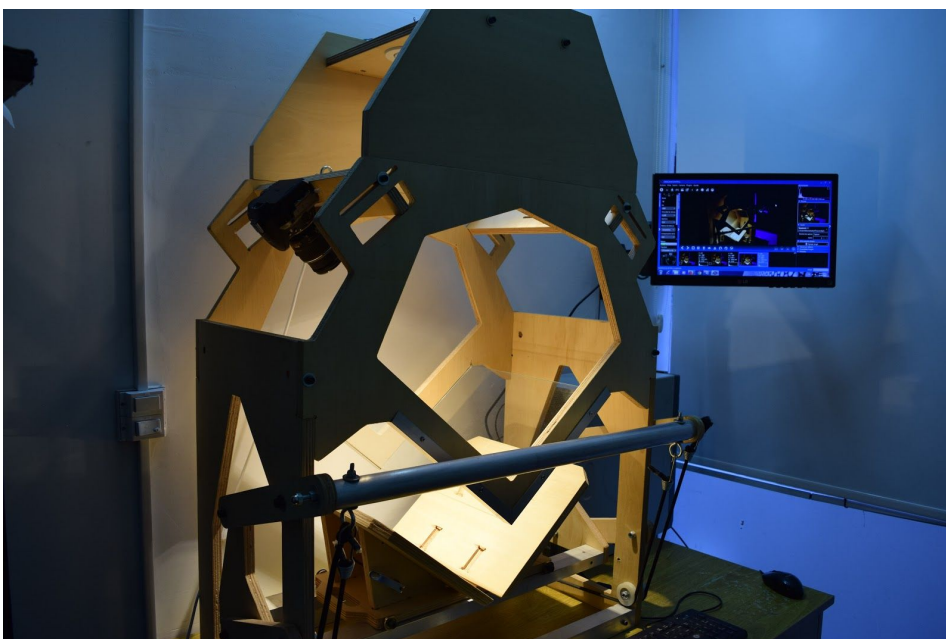


Figura 4. Modelo de escáner *Archivist 2014* usado para la digitalización corriente. Por su forma (36,5 de alto, 25,2 cm de ancho y 3,5 cm de grosor), el libro *Copiador* excedía el tamaño de atril y la forma de sujeción hubiese deteriorado el material.

## El diseño del nuevo escáner para la Digitalización del libro *Copiador*

A pesar de contar con un nuevo escáner *Archivista 2014*, el estado de la encuadernación con especiales requerimientos de manipulación hizo necesario el diseño *ad hoc* de un tercer escáner ya que las dimensiones de la base donde se sitúan los libros y la distancia al cristal impedían la libertad de movimientos necesaria para un material en estas condiciones. De no hacerlo el giro de las páginas durante el proceso de escaneo las hubiera deteriorado por su delicado estado y el peso del libro habría podido dañar la encuadernación textil del lomo.

Teniendo en cuenta las precauciones necesarias para el manejo del material se optó por un sistema de escaneo rediseñado a partir del *Model 1* de DIY, con una cámara cenital apuntando hacia el libro, junto con dos luces LED dicroicas, las luces led 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](#)<sup>9</sup> alto. (FADGI, 2016: 17). Uno de los brazos del atril del DAL se adaptó para poder sostener el libro siempre abierto a 150° aproximadamente, y balancear el peso de ambos lados de modo que no forzara la encuadernación.

Se dotó al Scanner de cámaras Nikon D5300, este tipo de cámara *reflex* permite a diferencia de las cámaras del primer Dal Canon sx130, el intercambio de lente. En la actualidad contamos con un lente del tipo zoom 18-55mm, dicho objetivo es muy versátil y se adaptó bien a esta labor, aunque un lente de focal fija de mayor luminosidad puede dar aún mejores resultados. 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 Library of Congress.

Es importante notar que la distancia focal en este lente es utilizada sobre los 35mm que es donde ya no presenta problemas de distorsión de la imagen. 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). En las versiones anteriores de DAL se utilizaba un firmware modificado ([CHDK](#)) que hacía el proceso de captura pero las imágenes se exportaban de la cámara una tarjeta de memoria. Si ocurría algún error en la captura no era posible detectarlo ni modificarlo en el momento. Con el digiCamControl las fotos pasan directamente a ser almacenadas en la computadora.

---

<sup>9</sup> 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.





Figura 5. Modelo de escáner con cámara cenital armado especialmente para la digitalización del Copiador del Museo de Física. Se utilizó a modo de atril, para sostener la otra sección del libro sin ejercer mucho peso de ninguno de los lados y tener espacio para girar las páginas sin doblarlas.

### **Copiador Museo de Física**

Para la digitalización del *Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas- Años 1918-1925*, se procedió a partir de los estándares internacionales con los que se acostumbra a trabajar, tanto de la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA), como National Archives and Records Administration (NARA) o de la Junta de Castilla y León. En este caso específico se utilizó la Guía técnica para digitalización de materiales del patrimonio cultural de la Federal Agencies Digital Guidelines Initiative en especial para los pasos previos a la digitalización (FADGI, 2016: 13-15) y el uso de la iluminación (2016: 19). Sin embargo, muchas de las dificultades que presentaba el material tan delicado con el que se debía trabajar no estaban contempladas en este documento: tal es el caso del tratamiento de las hojas con papel friable o el escaneo de papel translúcido.

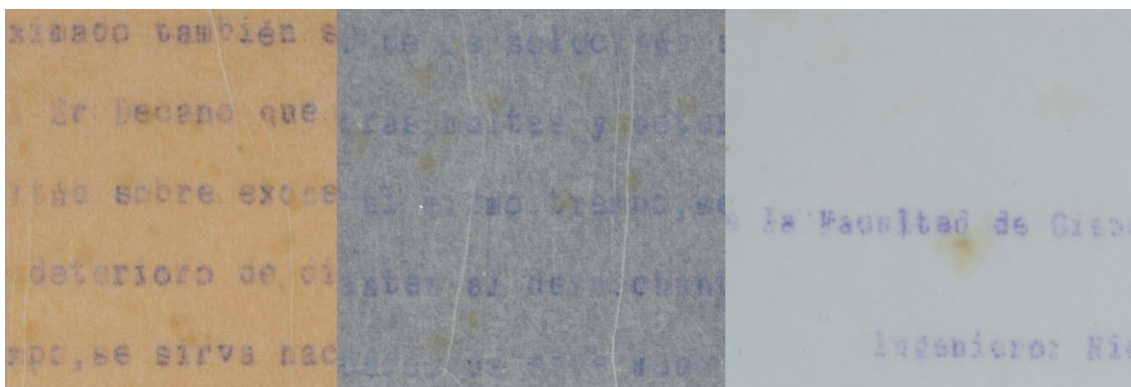
Los procesos previos al tratamiento del libro fueron los siguientes: 1. Se organizó, sistematizó y armó el nuevo escáner y se realizaron pruebas y acondicionamientos. 2. Se repasaron todas las normas y estándares para tratar material antiguo. 3. Se realizaron pruebas de imagen, color y luz y pruebas del proceso de escaneo. El escaneo final del documento llevó 3 días de trabajo.

El paso siguiente fue la revisión de todo el material digitalizado: 1. Se ordenaron las páginas por números y se renombró las imágenes de acuerdo al número de las mismas. 2. Se revisó y rehizo el escaneo de páginas cuyo archivo digital resultante no presentaba la calidad deseada.

## Pruebas previas al escaneo

Entre las pruebas previas al escaneo del material, se utilizó una superficie rígida (separador) para colocar entre hoja y hoja en el momento del escaneo; el objeto era evitar la aparición de textos superpuestos en la imagen recuperada en el caso de hojas translúcidas. Se utilizaron varios colores de separadores, el primero de ellos fue el amarillo, puesto que las manchas tenían estas tonalidades, se trató de visualizarlas mejor para quitarlas luego en el post proceso. Como los textos en imprenta se encontraban en color azul y los de letra manuscrita en negro, el color amarillo podría destacarlos y hacerlos más legibles. Sin embargo, la cantidad de manchas y detalles recuperados con esta metodología fueron excesivas e hizo más difícil la limpieza del texto.

El segundo color utilizado como separador fue el azul con el propósito de resaltar más los textos y despegarlos del fondo. El resultado fue un fondo muy oscuro con contraste insuficiente y pliegues de las hojas demasiado definidos que dificultaba la legibilidad. Por último, se utilizó un separador de color blanco que mejoró el contraste con la tipografía y que solucionó el inconveniente de las transparencias del papel.



*Figura 6. Se muestra el resultado en la captura con los separadores de color amarillo, azul y color blanco.*

Para el procesamiento de imágenes se utilizaron los softwares Photoshop CS6 y ABBY Fine Reader 12 Professional, empleado para la normalización de imágenes, edición de tratamiento de color, ruido y distorsiones, OCR, giro de imágenes recuperadas con cámaras en posición cenital, ajuste digital de foco, desaturación por color y redimensionamiento.

## Proceso corriente con escáner DAL

Para el escaneo de un material de más de 30 años es imperioso prever ciertos detalles: si el material presenta dificultades de manipulación y problemas de deterioro del papel o roturas en el lomo. En condiciones normales el escáner DAL es fácil de maniobrar y lleva aproximadamente un día laboral de proceso de digitalización (5 a 6 horas) el escaneo de 300 páginas.

La captura de las fotos con cada cámara y la demora de disparo, y el seteo de ambas, el seguimiento de cada una de ellas para conseguir resultados similares implican cierto tiempo. Luego de tomar las fotografías se deben mover los archivos digitalizados a la computadora donde se procesan y editan las imágenes con el software Scan Tailor. Este programa procesa las imágenes, allí se definen los DPI (puntos por pulgada) y corrige el sesgado. El paso siguiente es comenzar con la edición de las imágenes para que sea posible visualizarlas con la forma correspondiente a una página de un libro y no solo una foto del mismo. A continuación se realiza el reconocimiento óptico de caracteres (OCR) con ABBYY.

### **Procesos realizados para el tratamiento del libro**

En el caso de las imágenes capturadas con el escáner con cámara cenital se prefirió utilizar Photoshop CS6 y no Scan Tailor por la potencia de procesamiento del primero de los programas. Si bien, Scan Tailor se utiliza normalmente para modificar imágenes y automatizar muchos procesos, el Photoshop tiene más opciones de procesamiento de imágenes, especialmente en cuanto al ajuste de colores y nitidez. Las funciones automatizadas de Scan Tailor no funcionaban bien en los casos en los que las imágenes capturadas tenían la tinta difundida en el papel y la detección automática de renglones, orientación, etc. no funcionaba de manera eficiente.

Se utilizó la función **'actions'** de Photoshop que permite automatizar órdenes y procesos de manipulación de imágenes de manera secuencial con el fin de aplicar diferentes filtros a todas las imágenes capturadas. La gama de posibilidades que presentaba Photoshop frente a Scan Tailor era mucho mayor y permitía la revisión en detalle y el mejoramiento profundo del material digitalizado. Las funciones secuenciadas y automatizadas fueron las siguientes:

1. Girar imagen: el escáner con cámara cenital captura una sola imagen a la vez, y una de las dos páginas enfrentadas siempre se captura invertida. Debieron ser giradas a su posición de lectura original utilizando este filtro.
2. Enfocar (smart sharpen): este filtro sirvió para acentuar el borde de la tipografía en la imagen y mejorar el contraste con el fondo.
3. Desaturación por color (black and white filter): este filtro desatura los colores por separado. Esto permite seleccionar las tonalidades que representan manchas, suciedades y atenuarlos hasta que la superficie se vea homogénea. El papel translúcido con manchas en amarillo y tipografías en azul presentaba varios desafíos. La desaturación por color permitió eliminar las manchas amarillas desaturando la gama de rojo y amarillo y se resaltó la tipografía al saturar la gama azul. De esta manera se logra un apariencia limpia en los contornos de las tipografías mejorando la legibilidad.

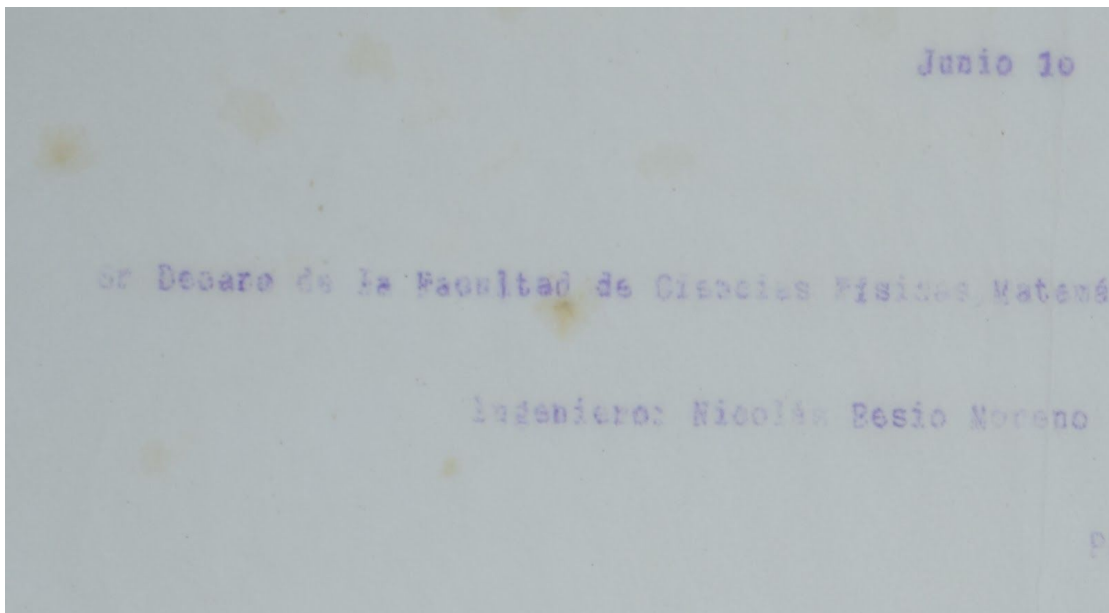


Figura 7. Imagen original sin retoques.

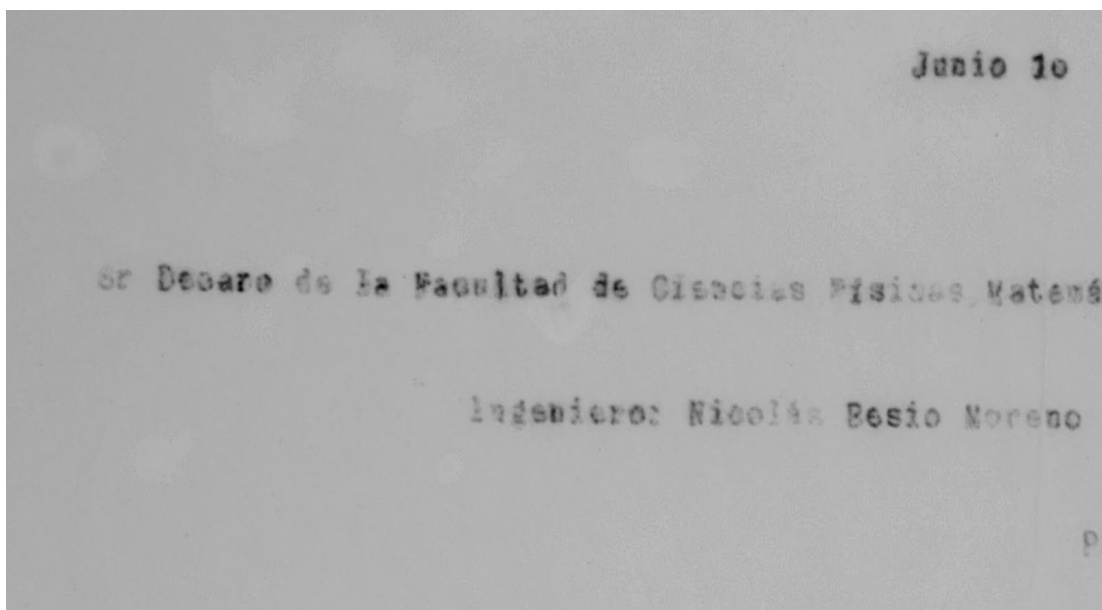
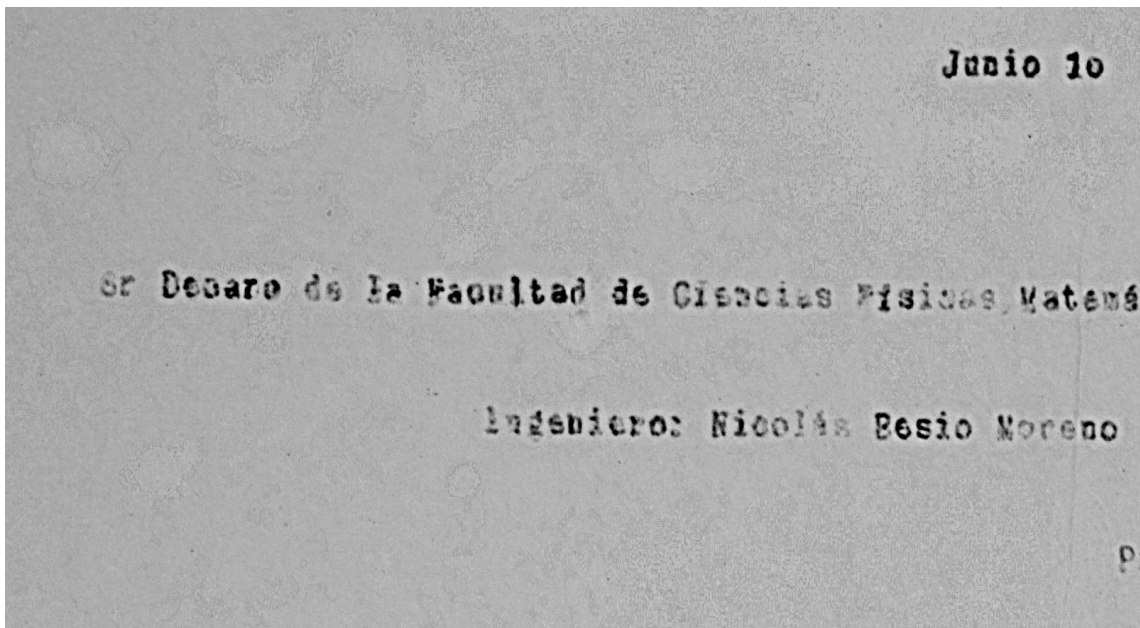
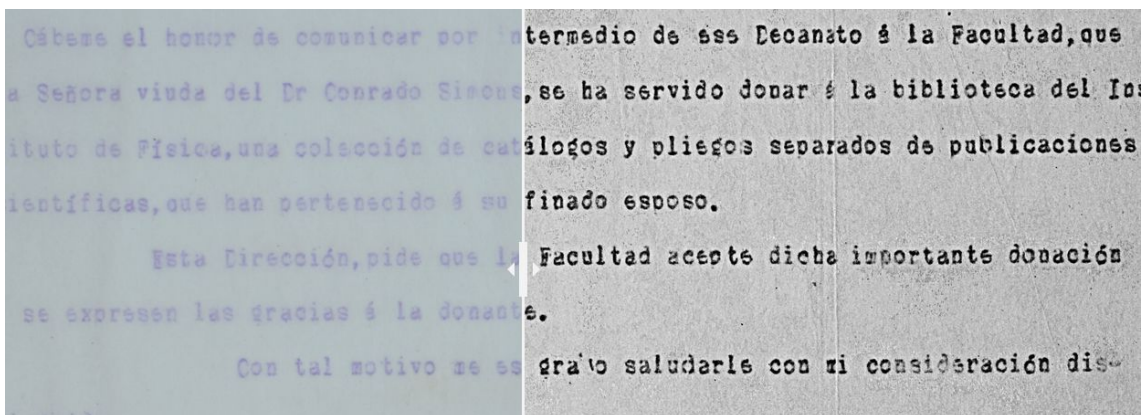


Figura 8. Imagen luego de aplicar el filtro blanco y negro previo al filtro de enfoque..



*Figura 9.* Imagen con el filtro de enfoque inteligente. Photoshop posee varias opciones, algunas automáticas (Enfoque, enfoque de bordes y más enfoque) y dos opciones que permiten modificar las variables (enfoque inteligente y máscara de enfoque). De las dos opciones el enfoque inteligente es el que permite controlar más variables y ajustar de forma más precisa el contraste entre la tipografía y el fondo.



*Figura 10.* Se aprecia la diferencia de nitidez en el material procesado.

4. Tamaño de lienzo: se predeterminó un tamaño de lienzo común a todas las imágenes acorde al tamaño del documento para recuperar las condiciones visuales más similares a la manipulación del documento físico.

5. Droplet: esta es una utilidad de automatización que permite crear un archivo ejecutable al cual se pueden enviar las páginas a procesar. Esta función se encargará de abrir los archivos, aplicar la "acción" (con las mejoras del documento), luego guardar los mismos en la carpeta seleccionada y por último cerrar cada documento.

Para el OCR se utilizó Abby FineReader, 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.

### **Resultados y discusión:**

Si bien se consultaron los estándares internacionales para la digitalización de la IFLA, NARA y de la Junta de Castilla y León y la Guía técnica para digitalización de materiales del patrimonio cultural de la FADGI fue necesario adaptar tanto el escáner como los procesos de captura y tratamiento de las imágenes por las particularidades del material a digitalizar. Por ello y luego de consultar a la especialista en restauración del Museo de Física, se diseñó un nuevo escáner con cámara cenital, ya que su tamaño superaba el del atril del escáner original DAL. 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.

El procedimiento de captura de imágenes se modificó utilizando el software digiCamControl 2.0.72.0 que probó ser mucho más eficiente y cómodo para la captura y el traslado directo de las imágenes al CPU. Para el post proceso y mejoramiento de las imágenes se utilizó el Photoshop CS6 cuyos filtros y funcionalidades de automatización de operaciones probó ser más eficiente que el Scan Tailor, para este caso en particular, en el que el libro original presentaba varias dificultades. Una de las principales dificultades en el uso del Photoshop fue el Espacio en disco. Se trataba de procesos que requerían muchos recursos de memoria y requerían mucho espacio de almacenamiento, tanto de los archivos temporales como de los archivos finales.

Para el procesamiento del OCR realizado con ABBY Fine Reader 12 Professional los requerimientos de recursos y memoria de la PC también fueron importantes. El archivo final pdf tenía un peso de 1GB, por la calidad de las imágenes requeridas para la preservación digital del documento. Un archivo comprimido y de menor tamaño se utilizó para la difusión del material en el repositorio.

### **Conclusiones:**

El trabajo de digitalización y las pruebas realizadas con el el libro *Copiator – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)* probó que con el procedimiento estándar de escaneo, utilizado para libros en estado óptimo no se obtienen los mejores resultados en casos de libros como éste, cuyo estado de conservación es delicado. Tanto el software utilizado para la captura de imágenes y pos procesamiento debieron adaptarse y modificarse dejando de lado el firmware modificado (CHDK) y el Scan Tailor. El digiCamControl 2.0.72.0 fue un programa rápido, confiable y versátil para la captura de imágenes y el Photoshop CS6 permitió mayor libertad en la manipulación de las imágenes destinadas al reconocimiento de texto.

El escáner con cámara cenital permitió un giro cuidadoso de las páginas y la protección de la encuadernación, al no ejercer peso extra. Para el

escaneado de hojas translúcidas se utilizaron separadores que ayudaron a resaltar la tipografía y destacar las manchas del papel que debían ser eliminadas. El tamaño de las imágenes capturadas se cambió a un tamaño de lienzo común para que fuera acorde con el tamaño original del documento y que las condiciones visuales semejaran la manipulación del documento físico. Un archivo de alta calidad de imágenes se destinó a la preservación digital y otro comprimido 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 por parte de crawlers o repositorios.



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

## Bibliografía

- A 100 Años de la Reforma Universitaria. (2018). Recuperado 17 de agosto de 2018, de <http://museo.fisica.unlp.edu.ar/exhibiciones/a-100-anos-de-la-reforma-universitaria-10167>
- ABBYY. (2013). (Versión 12.0.101.264). Recuperado de <https://www.abbyy.com/es-la/?content=global>
- Adobe Photoshop CS6. (2012). (Versión 13.0). Recuperado de <https://www.adobe.com/la/products/photoshop.html>
- CHDK Wiki. (2012). (Versión 1.4.1). Recuperado de <http://chdk.wikia.com/wiki/CHDK>
- CHDK Wiki. (s. f.). Recuperado 7 de agosto de 2018, de <http://chdk.wikia.com/wiki/CHDK>
- Clifford A. Lynch. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age, 226, 1-7.
- Derecho a Leer (DAL). (s. f.). Recuperado 30 de julio de 2018, de <http://derechoaleer.org/blog/acerca-de.html>
- digiCamControl. (2018). (Versión 2.1.0). Recuperado de <http://digicamcontrol.com/>
- Diseños de escáneres de libros DIY. (s. f.). Recuperado 30 de julio de 2018, de <http://diybookscanner.org/es/designs.html>
- DIY Book Scanner. (s. f.). Recuperado 30 de julio de 2018, de <http://diybookscanner.org/es/index.html>
- Edmondson, R. (s. f.). Memoria del Mundo: directrices para la salvaguardia del patrimonio documental; 2002. UNESCO.
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA), Consejo Internacional de Archivos (ICA), & Unesco. (2014, junio 13). Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos. Recuperado de <http://travesia.mcu.es/portalnb/jspui/handle/10421/3342>
- Garro Acón, M. (2015). Red centroamericana de repositorios de patrimonio cultural. Recuperado de <http://www.kerwa.ucr.ac.cr/handle/10669/15333>
- Junta de Castilla y León. (s. f.). Recomendaciones para la digitalización de documentos en los archivos. Recuperado de <https://archivoscastillayleon.jcyl.es/web/jcyl/ArchivosCastillaYLeon/es/Plantilla100Detalle/1253861401355/1253861401355/1284166105811/Comunicacion>
- Library of Congress. (s. f.). Technical Information. Recuperado 2 de octubre de 2018, de <https://memory.loc.gov/ammem/about/techIn.html>
- Muñoz, W. C., & Restrepo, M. C. (2016). Los repositorios como herramienta para la recuperación del patrimonio bibliográfico: el caso de seis bibliotecas públicas municipales del departamento de Antioquia. *Revista Interamericana de Bibliotecología*, 39(1), 57-68. <https://doi.org/10.17533/udea.rib.v39n1a08>
- Nuestro Do It Yourself Book Scanner. (s. f.). Recuperado 30 de julio de 2018, de



<http://derechoaleer.org/blog/2011/08/nuestro-do-it-yourself-book-scanner.html>

- Plata, I. de F. L. (1918). *Libro Copiador de la Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)*. Recuperado de <http://hdl.handle.net/10915/67640>
- Puglia, S., Reed, J., & Rhodes, E. (2004). Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images.
- Puglia, S., Reed, J., & Rhodes, E. (2016). Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files. (Federal Agencies Digital Guidelines Initiative (FADGI), Ed.).
- Quintero, J. A. D., Núñez, L. A., Saldivia, B. E. S., & Heeren, R. T. (2006). Los repositorios institucionales y la preservación del patrimonio intelectual académico. *Interciencia*, 31(1), 22-28.
- Rodríguez Terán, Á. (2016). Repositorios Institucionales: Repositorio del Instituto del Patrimonio Cultural de España. Recuperado de <https://repositorio.unican.es/xmlui/handle/10902/9583>
- ScanTailor. (2012). (Versión 0.9.11.1). Recuperado de <http://scantailor.org/>
- Web: El derecho a leer. (s. f.). Recuperado 4 de octubre de 2018, de <http://www.lapulseada.com.ar/web-el-derecho-a-leer/>

## **Autores: resúmenes biográficos**

**Carlos J. Nusch** es Profesor de Letras por la Universidad Nacional de La Plata, Argentina. Trabaja en el Servicio de Difusión de la Creación Intelectual (SEDICI) de la UNLP, en el Proyecto de Enlace de Bibliotecas (PREBI) desde 2001 y en el repositorio CIC-Digital desde 2014. Es miembro del Comité Asesor del el Centro de Servicios en Gestión de Información (CESGI) desde 2016 y parte del personal del Observatorio Medioambiental La Plata. También trabaja en Estudios Clásicos y Literatura Medieval en el Centro de Estudios latinos (CEL) y en el Instituto de Investigaciones en Humanidades y Ciencias Sociales (IdIHCS). Ha publicado varios artículos sobre trabajo académico colaborativo, repositorios digitales, análisis del discurso político y literatura.

**Esteban C. Fernández** es Productor y Director para Radio y Televisión por el Instituto Superior de Enseñanza Radiofónica (ISER). Trabaja en el Servicio de Difusión de la Creación Intelectual (SEDICI) de la UNLP y en el Proyecto de Enlace de Bibliotecas (PREBI) desde 2004. También cumple tareas en el área de Digitalización y Preservación Digital de SEDICI.

**Pablo Méndez Moura** es estudiante de la Licenciatura en Bibliotecología y Ciencia de la Información de la Facultad de Humanidades y Ciencias de la Educación (FaHCE), Universidad Nacional de La Plata (UNLP). Trabajó en la Cátedra de Montaje y Edición de la Licenciatura en Artes Audiovisuales de la Facultad de Bellas Artes de la UNLP, es parte del área de Digitalización y Preservación Digital del Servicio de Difusión de la Creación Intelectual (SEDICI) desde 2017.

**Cecilia von Reichenbach** realizó la Licenciatura y el Doctorado en Física en la Universidad Nacional de La Plata. Luego de diez años de trabajo en temas de teoría cuántica de campos, y a raíz de su dedicación a la gestión del Museo de Física del Departamento de Física de la Facultad de Ciencias Exactas, cambió la orientación de sus investigaciones (en la UNLP y el CONICET) hacia la Historia de la Física en Argentina, específicamente en la UNLP en las primeras décadas del siglo 20. Es profesora de la Facultad de Ciencias Exactas en una cátedra mixta en la Facultad de Humanidades, y Directora del Museo de Física.

**Ana Elisa Cozzuol** es Técnica restauradora egresada del Instituto Universitario Nacional de Arte (IUNA). Técnica química, egresada de la “Escuela Nacional de Educación Técnica N° 1 de La Plata Albert Thomas”. Es responsable del Área de Conservación y Taller de Restauración de Papel del Museo de Física de la Facultad de Ciencias

Exactas de la Universidad Nacional de La Plata. Se especializa en trabajos de conservación y restauración sobre papel. Ejerció la docencia en el Departamento de Artes Visuales, Universidad Nacional de las Artes (UNA) en la Cátedra “Prácticas de Conservación y Restauración de Bienes Culturales” y “Conservación Restauración de Documentos Gráficos y Papel”. Desde año 2003 hasta 2016.

**Mariana Santamaría** es Licenciada en Museología y Repositorios culturales y naturales de la Universidad Nacional de Avellaneda. Se ha capacitado en gestión cultural, comunicación y archivística. Es museóloga del Museo de Física (2010) y la Biblioteca Pública (2011) de la Universidad Nacional de La Plata. Es vicepresidente de la Fundación Centro de Artes Visuales de La Plata y miembro del grupo de investigación y exhibición del Centro de Arte Experimental Vigo. También se desempeña como coordinadora del Museo y Archivo histórico del Club de Gimnasia y Esgrima La Plata.

**Marisa R. De Giusti** es Doctora en Informática por la por la Universidad Nacional de La Plata, Argentina. También es Ingeniera en Telecomunicaciones y Profesora de Letras. Ha dirigido el Servicio de Difusión de la Creación Intelectual (SEDICI) de la UNLP desde 2003, y el repositorio CIC-Digital desde 2014, y el Centro de Servicios en Gestión de Información (CESGI) desde 2016. También coordina el Observatorio Medioambiental La Plata. Es profesora en la Facultad de Informática de la UNLP y sus intereses de investigación incluyen el campo de repositorios digitales, preservación digital y acceso abierto.