

IX CONFERÊNCIA INTERNACIONAL SOBRE BIBLIOTECAS E REPOSITÓRIOS DIGITAIS DA AMÉRICA LATINA

BIREDIAL-ISTEC

SÃO PAULO • BRASIL

30 • 31 DE JULHO / 1 • 2 DE AGOSTO 2019

2019

Rumbo a la autoevaluación: la elaboración de un Manual de Procedimientos de Digitalización

Carlos J. Nusch

carlosnusch@prebi.unlp.edu.ar

PREBI-SEDICI

Universidad Nacional de La Plata

CESGI

Comisión de Investigaciones Científicas



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](#).



Preservación en un repositorio

El planeamiento de la preservación dentro de un repositorio:

- ❑ es un elemento clave de la autoevaluación y de la evaluación externa y
- ❑ exige confirmar lo realizado a través de documentos
- ❑ responsabilidades y los responsables delimitados
- ❑ marco sistemático de buenas prácticas de gestión documental.



El Flujo de Trabajo

El desafío del Manual está en describir todo el flujo de trabajo:

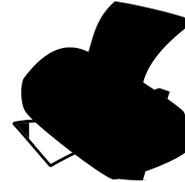
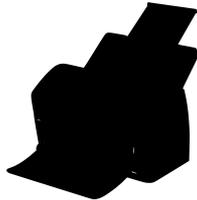
1. El ingreso del material, la prioridad asignada, el depósito de la obra física y el seguimiento de los procesos de digitalización.



El Flujo de Trabajo

El desafío del Manual está en describir todo el flujo de trabajo:

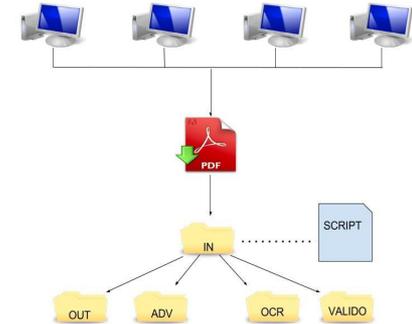
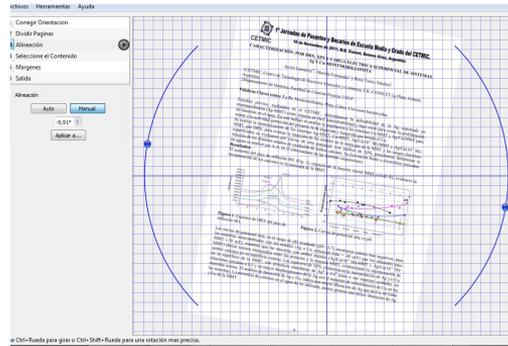
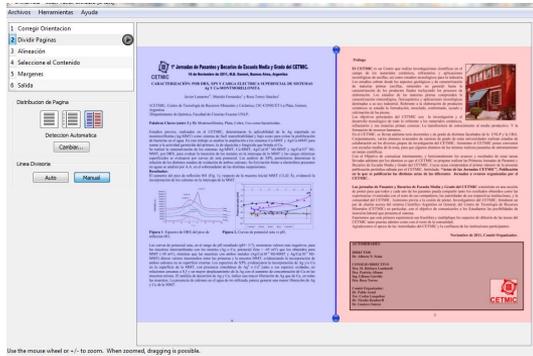
2. Los detalles del proceso de digitalización en sí mismo: software, hardware, cámara y escáner.



El Flujo de Trabajo

El desafío del Manual está en describir todo el flujo de trabajo:

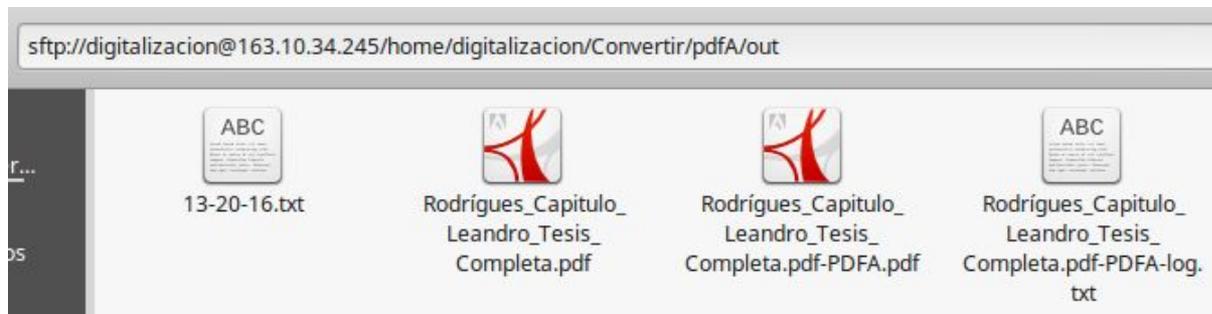
3. Las transformaciones a las que se someten los archivos tras la digitalización, las distintas copias generadas y los diferentes formatos de archivos utilizados.



El Flujo de Trabajo

El desafío del Manual está en describir todo el flujo de trabajo:

4. El modo de trabajo en relación al guardado de los archivos.



Los primeros escáneres

El SEDICI contó siempre con escáneres estándares a los que se les daba un uso esporádico.



HP modelo 5590



HP N8400

Los escáneres usados

El SEDICI contó siempre con escáneres estándares a los que se les daba un uso esporádico.

- ❑ Un escáner HP 7500.
- ❑ Un escáner de alimentación automática (ADF) modelo Fujitsu Fi 6170.
- ❑ Un escáner de grandes dimensiones modelo Contex IQ Quattro para el escaneo de planos y documentos de gran tamaño.



Los escáneres DIY

Se trabajó en el armado de diferentes escáneres de bajo costo para materiales de encuadernación delicada.

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.



derecho a LEER

Derecho a LEER - Otro blog en apoyo al Profesor **Horacio Potel**.

A principios de 2009, a instancias de una editorial francesa, la Camara Argentina del Libro realizó una denuncia judicial que inició una causa criminal contra Potel, su delito: difundir a través de internet textos de célebres filósofos: Nietzsche, Derrida y Heidegger. Espontáneamente se iniciaron campañas y muestras de apoyo a Potel por toda la red... hasta que se consiguió su sobreseimiento!



Donar un libro a una biblioteca para que todo el mundo lo lea.
Subir un libro a internet para que todo el mundo lo lea.

La digitalización masiva de libros, es la mejor respuesta a la [criminalización de los bibliotecarios](#), de la cual [son cómplices, lamentablemente, hasta los secretarios de cultura de los gobiernos progresistas](#)... Sin embargo, los [escáneres de libros](#) están un poco caros, desde U\$S 10.000 en adelante, hasta para instituciones como universidades y bibliotecas **los costos son prohibitivos**... ¡Así que hicimos uno nosotros!

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>



Bienvenido a la comunidad de escáneres de libros DIY

Somos una comunidad de personas que construyen escáneres de libros 'Do It Yourself'. Nos hicimos cargo de la preservación de libros, y la tomamos en nuestras propias manos. Somos el puente faltante entre las estanterías de tu biblioteca y tu e-reader. ¡Únete! Involúcrate probando construir un escáner simple, construyendo un kit, o explorando los límites de las tecnologías de digitalización. Si tus dudas no pueden ser respondidas mediante la lectura, escribinos un correo electrónico a: [diybookscanner \[at\] gmail.com](mailto:diybookscanner@gmail.com).

¿Qué es un escáner de libros?

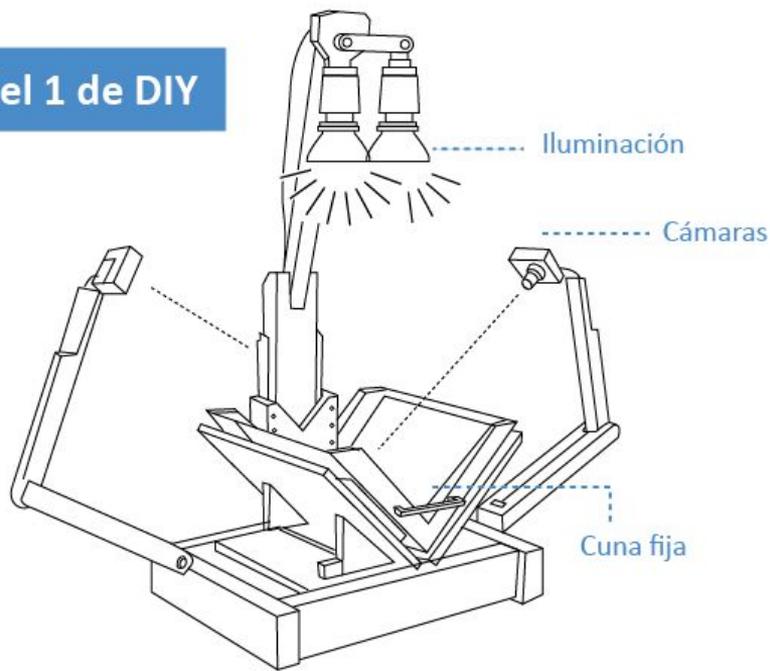
Un escáner de libros es cualquier dispositivo utilizado para digitalizar las páginas de un libro. Cada escáner de libros tiene tres partes. Las cámaras se utilizan para capturar las imágenes de cada página. Un controlador dispara las cámaras y salva las fotografías resultantes en algún lugar conveniente. La plataforma del escáner provee de un ambiente óptimo para la fotografía, mientras sostiene todas las piezas en su lugar.

Diseños de escáneres de libros DIY. (s. f.). Recuperado 30 de julio de 2018, de

<http://diybookscanner.org/es/designs.html>

Escáner basado en el Model 1 de DIY

Model 1 de DIY



Los escáneres **DAL** utilizados para el escaneo de libros en buen estado no pudieron utilizarse porque el giro de las páginas hacía que el papel pudiera quebrarse.

Escáner basado en el Model 1 de DIY



Softwares descartados



- Los softwares utilizados para la captura de imágenes y posprocesamiento debieron adaptarse y transformarse dejando de lado el firmware modificado (**CHDK**) y el **Scan Tailor**.
- El **Scan Tailor** no tenía funciones avanzadas de ajuste de colores y nitidez.
- Firmware modificado CHDK: requería la configuración in situ de cada cámara por separado y el uso de una tarjeta de memoria para mover las imágenes a la computadora.



Escáner Archivist 2014

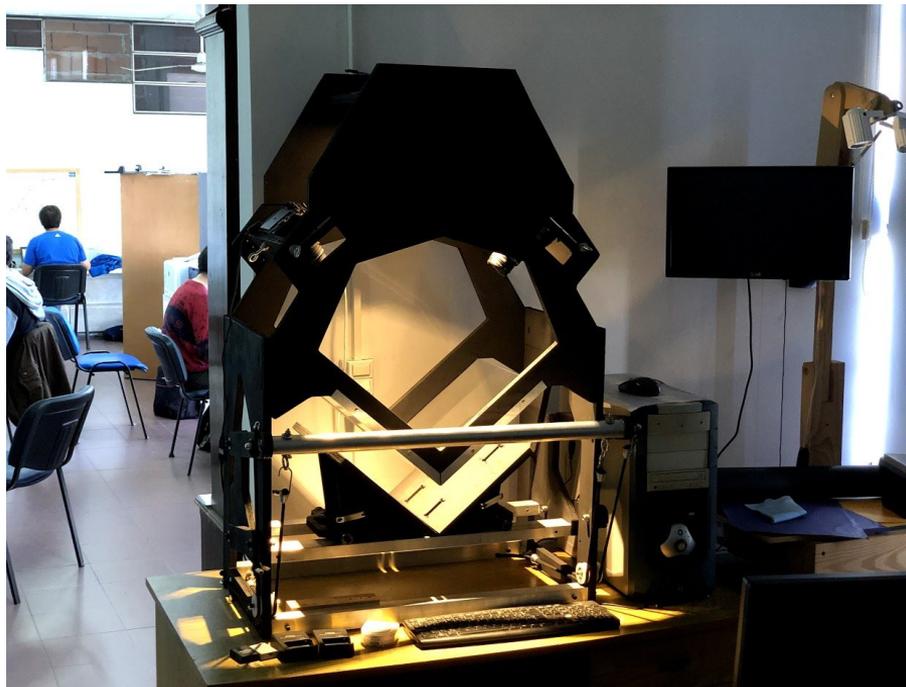
Archivista 2014



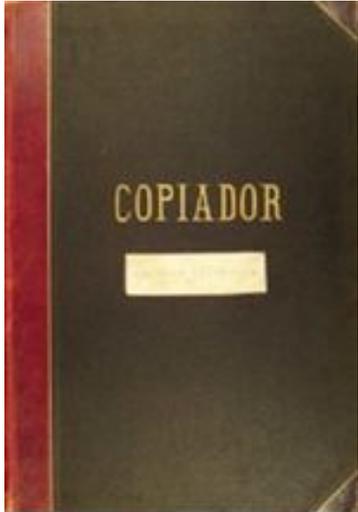
Modelo Archivista 2014

Presentaba muchas mejoras pero no pudo emplearse para el escaneo del libro Copiador por el tamaño y estado de deterioro.

Escáner Archivist 2014



Libro Copiador



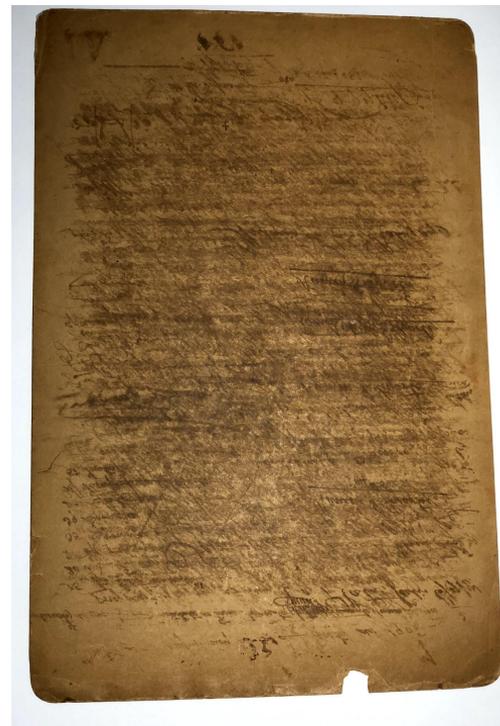
Los libros copiadores constituyen un corpus que evidencia las **relaciones científicas, institucionales, políticas** y aún personales de los miembros de la incipiente comunidad científica de las **épocas fundacionales** y, a diferencia de las publicaciones oficiales, están escritos y firmados por los propios protagonistas de la física de entonces.



¿CÓMO FUNCIONABAN LOS LIBROS COPIADORES?

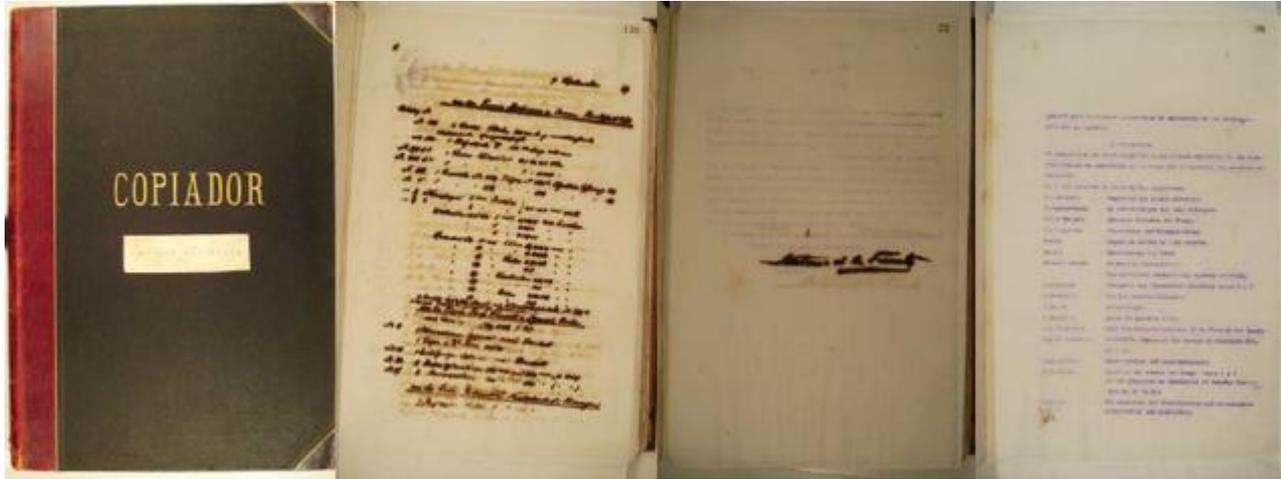


Hojas de material translúcido

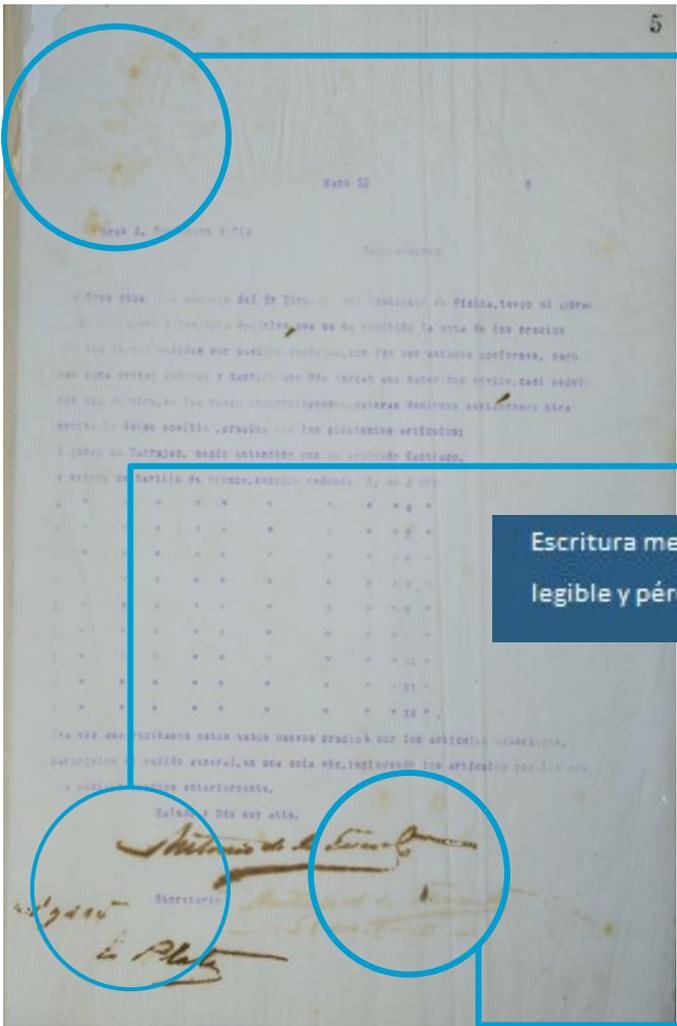


Hojas de transferencia

ESTADO DE CONSERVACIÓN



- Papel amarronado, débil, friable con roturas y desprendimientos.
- Escritura manuscrita con tinta difundida en el papel y transferida a los siguientes, con pérdida de nitidez e imagen doble.
- Escritura mecanográfica poco legible.
- Encuadernación con especiales requerimientos de manipulación.



Escritura mecanografiada poco legible y pérdida de nitidez



Doblesces y desprendimientos



Tinta difundida en el papel y transferida a los consecutivos.

12
3

La Plata Abril 8 de 1808

N. 120

Señor Decano de la Facultad de Ciencias
Físicas y Matemáticas
Ing. Don Benjamín Sal

Congo el honor de decir al Sr. Decano
que la adjunta cantidad de ~~cientos~~ pesos
por gastos efectuados por el Instituto de Física
hasta el 31 de ~~enero~~ ~~de~~ ~~este~~ año, cuya suma ascen-
de a (2.846.00). De los ochocientos noventa
y seis pesos con ~~cinco~~ ~~cientos~~ ~~noventa~~ ~~y~~ ~~seis~~ ~~centésimos~~ ~~de~~ ~~real~~ ~~de~~ ~~los~~ ~~cuales~~
de ~~los~~ ~~cuales~~ ~~se~~ ~~ha~~ ~~entregado~~ ~~la~~ ~~cantidad~~

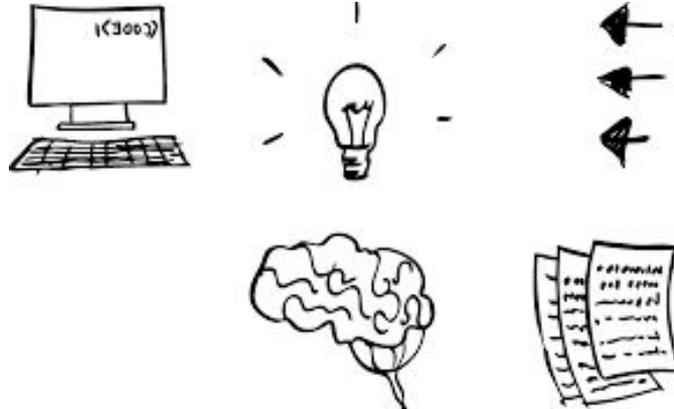
Principales problemas para el escaneo

- ❑ Dimensiones de la base donde se sitúan los libros y la distancia al cristal, que impedían la libertad de movimientos necesaria para un material en estas condiciones.
- ❑ Giro de las páginas durante el proceso de escaneo hubiera provocado el deterioro por su delicado estado y
- ❑ El peso del libro habría podido dañar la encuadernación textil del lomo.



Rediseño del escáner

- ❑ Model 1 de DIY Rediseñado
 - ❑ Cámara cenital
 - ❑ Dos luces LED dicroicas
 - ❑ Consumen hasta un 90 % menos,
 - ❑ No generan calor considerable
 - ❑ Un Índice de reproducción cromática CRI alto. (FADGI, 2016: 17).
- ❑ Sostén del libro a 150°
- ❑ Balance del peso para no forzar la encuadernación.



Nuevas características

- ❑ Cámaras Nikon D5600,
 - ❑ cámara reflex permite el intercambio de lente
 - ❑ un lente del tipo zoom 18-55mm
 - ❑ formato de sensor DX de 24,2 Megapíxeles
 - ❑ permite obtener documentos escaneados a 400dpi



Nuevo Software



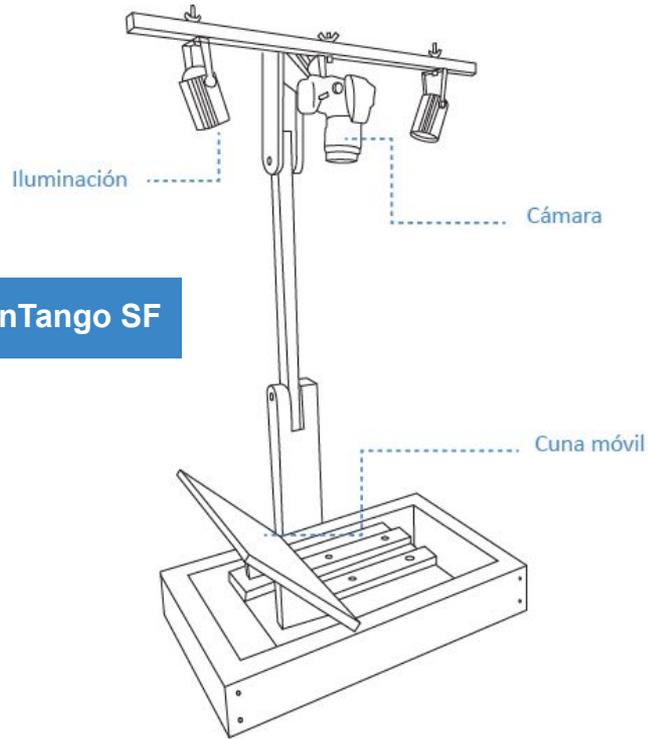
El **digiCamControl 2.0.72.0** fue un programa rápido, confiable y versátil para la captura de imágenes y permitió el manejo de las cámaras directamente desde la computadora.

Nuevo Software



Photoshop CS6 permitió mayor libertad en la manipulación de las imágenes destinadas al reconocimiento de texto. Se aplicaron filtros y se automatizó el procedimiento estándar para todas las imágenes

Escáner Cenital



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 de luz cálida cuya temperatura no daña el material.

Escáner cenital



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 de luz cálida cuya temperatura no daña el material.

Nuevos desafíos: mayor demanda

Aumento del caudal de materiales

- ordenar el proceso
 - la toma de decisiones técnicas
 - actores que intervenían.
- Distribución del material según el tipo de escaneo a realizar

SELECCIÓN DE ESCANER



Selección de Escáner

- Buen estado que no esté encuadernado o
- cuya encuadernación pueda desarmarse
 - escáner de alimentación automática (ADF), cuyo desempeño ronda la captura de 60 páginas por minuto (ppm); en nuestro caso el Fujitsu Fi 6170 o el Hp 7500.



Selección de Escáner

Encuadernación permite un ángulo de 180 grados:

- escáner plano, en nuestro caso el HP 7500.



Selección de Escáner

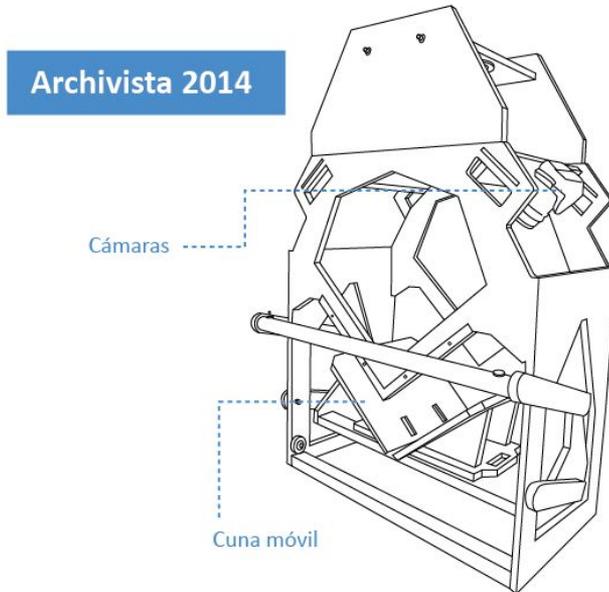
Planos y documentos de gran tamaño, un escáner de grandes dimensiones será necesario para:

- Contex IQ Quattro hasta un ancho máximo de 44 pulgadas o 111,76 centímetros.



Selección de Escáner

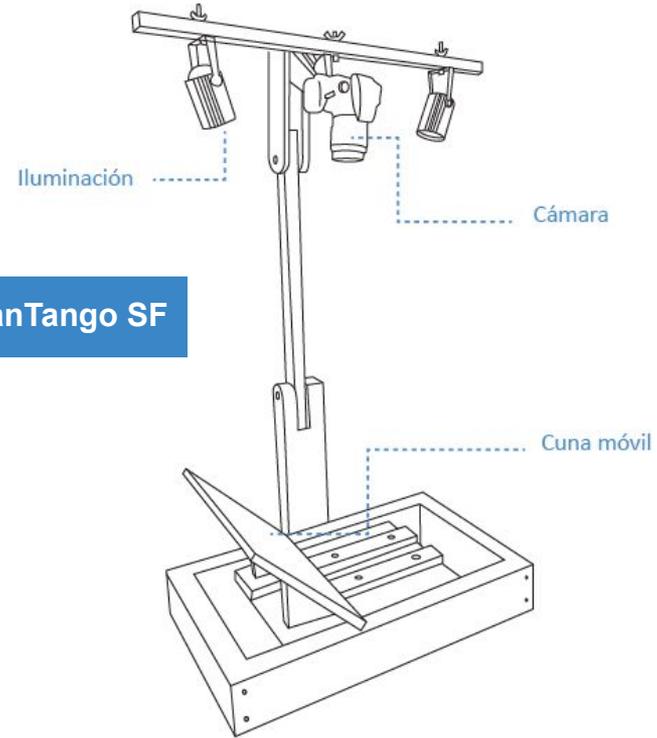
- Materiales cuya encuadernación requiere un ángulo de apertura menor a 130° .
- Tamaño de página inferior a 35 cms de alto.
 - Archivist 2014



Selección de Escáner

- Estado delicado de conservación
 - Escáner cenital

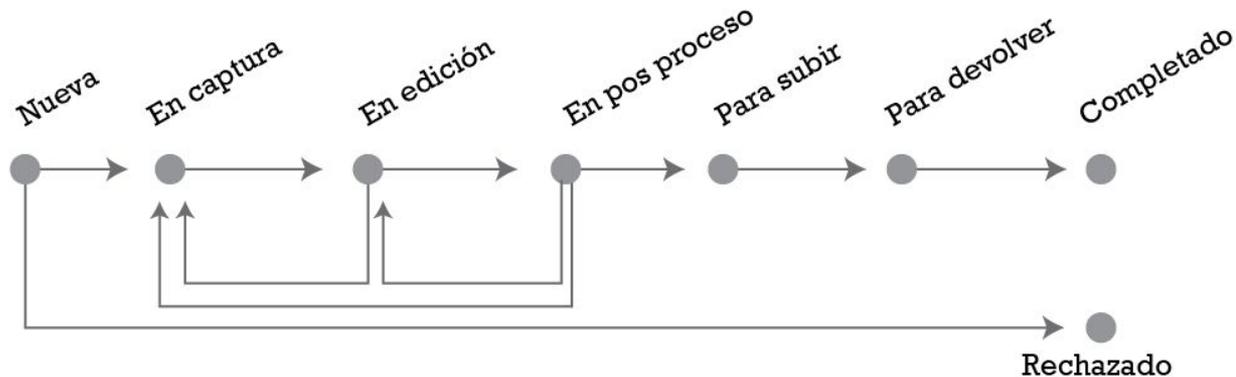
ScanTango SF



Flujo de Trabajo

Se elaboró un flujo de trabajo que atiende a los diferentes tipos de materiales a ser digitalizados y las particularidades de su digitalización así como un registro de todos los procesos llevados a cabo con el material digital resultante.

Este flujo de trabajo consiste en una serie de etapas (ver gráfico), mayormente consecutivas, que organizan y rigen las acciones a realizar sobre cada material.



Tareas del proceso de digitalización en Redmine

Para la gestión de los distintos estados y avances de los materiales ingresados para digitalizar se utiliza Redmine, una herramienta de gestión de tickets. Este sistema permite realizar un seguimiento completo de todas las etapas que atraviesan dichos materiales, los responsables de cada momento y mantiene un registro de las acciones realizadas.

<input type="checkbox"/>	#	Tipo	Estado	Categoría	Prioridad	Asunto	Asignado a	Actualizado	Tipo de recurso	Escáner
■ Nueva 8										
<input type="checkbox"/>	5767	Recurso	Nueva		Normal	Sección II. Labor de los Centros de Estudios. 1940	digitalizadores	24 Mayo 2019 04:42 PM	Documento institucional	DAL
<input type="checkbox"/>	5637	Recurso	Nueva		Normal	Marasso Arturo - Rubén Darío y su creación poética	Pablo Mendez Moura	29 Abril 2019 09:57 AM	Libro	DAL
■ En captura 1										
<input type="checkbox"/>	5771	Recurso	En captura		Normal	Castiñeiras, J.R., Síntesis histórica de las Universidades Argentinas.	Pablo Mendez Moura	31 Mayo 2019 09:09 AM	Documento institucional	DAL
<input type="checkbox"/>	5668	Recurso	Comentarios		Normal	Cremschl y Tauber - Acceso Ferroviario en Trinchera a la Ciudad de La Plata	Analia Pinto	25 Abril 2019 10:31 AM	Otros	Automático
■ En edición 2										
<input type="checkbox"/>	5751	Recurso	En edición		Normal	UNLP. Digesto n°2. 1935	Pablo Mendez Moura	31 Mayo 2019 01:55 PM	Documento institucional	Plano
<input type="checkbox"/>	5770	Recurso	En edición		Normal	Revista de la Universidad Vol 28	Esteban C Fernandez	31 Mayo 2019 09:44 AM	Boletín/Revista	DAL
■ En postproceso 3										
<input type="checkbox"/>	5760	Recurso	En postproceso		Normal	Sección III. Anuario. 1937	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
<input type="checkbox"/>	5761	Recurso	En postproceso		Normal	Sección III. Anuario Bibliográfico. 1937.	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
<input type="checkbox"/>	5762	Recurso	En postproceso		Normal	Sección V. Leyes decretos ordenanzas. 1937	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
■ Para subir 21										
<input type="checkbox"/>	5741	Recurso	Para subir		Normal	Brelles Mariño, Graciela -Interacción planta-bacteria	sediciano	31 Mayo 2019 06:11 PM	Tesis	DAL
<input type="checkbox"/>	5545	Recurso	Para subir		Alta	Boletín de la asociación de Astronomía Nro14	Carlos J. Nusch	31 Mayo 2019 07:02 AM	Boletín/Revista	Automático
<input type="checkbox"/>	5628	Recurso	Para subir		Normal	Revista de la Universidad Vol 2	Analia Pinto	16 Abril 2019 12:47 PM	Boletín/Revista	DAL
■ Para devolver 22										
<input type="checkbox"/>	5756	Recurso	Para devolver		Normal	Gallo, G., Discursos. 1976	digitalizadores	31 Mayo 2019 11:16 AM	Documento institucional	Plano

Etapa 1. Nueva: recepción y ponderación de materiales

Se deberá registrar las particularidades de cada material, a saber:

- si el material puede desarmarse o no
- el escáner apropiado de acuerdo al formato,
- persona responsable de realizar digitalización
- la complejidad percibida luego de un análisis preliminar
- los metadatos propios del material (Autor, Título etc)
- los datos del aportante, a quien se le debe devolver el material

Etapa 2. Captura

Para la determinación de esta etapa se utilizaron como insumo diferentes recomendaciones internacionales:

- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA), Consejo Internacional de Archivos (ICA), y Unesco. 2014. «Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos».
- Junta de Castilla y León. «Recomendaciones para la digitalización de documentos en los archivos».
- Puglia, Steven, Jeffrey Reed, y Erin Rhodes. 2016. «Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files» ed. Federal Agencies Digital Guidelines Initiative (FADGI).

De todos los anteriores la más conveniente fueron las recomendaciones FADGI, en particular para definir resoluciones, aspectos de iluminación y color.

Parámetros de captura

De acuerdo al escáner elegido se utilizará un programa de captura diferente. Por ejemplo, en SEDICI usamos:

- digiCamControl para los escáneres de libros y Cenital:
- PaperStream para el escáner automático Fujitsu FI-7160
- NextImage para el escáner de gran formato Contex IQ Quattro
- Software de escaneo de documentos inteligente de HP para el escáner automático y plano Hp 7500.

Estos programas deberán ser configurados de acuerdo al material en cuestión y a los parámetros de color, resolución y formato definidos. De forma resumida, se presenta a continuación la tabla de resoluciones utilizada para la captura:

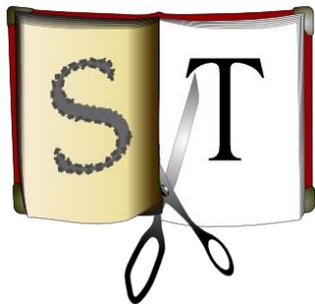
Tipo de material	Profundidad de bit	DPI
Color	24bit	400
Escala de grises	8bit	400
Blanco y negro	1bit	600

Etapa 3. En edición

Para la edición y mejoramiento de imagen se toman los archivos generados en la etapa de captura, y se procesan las imágenes para:

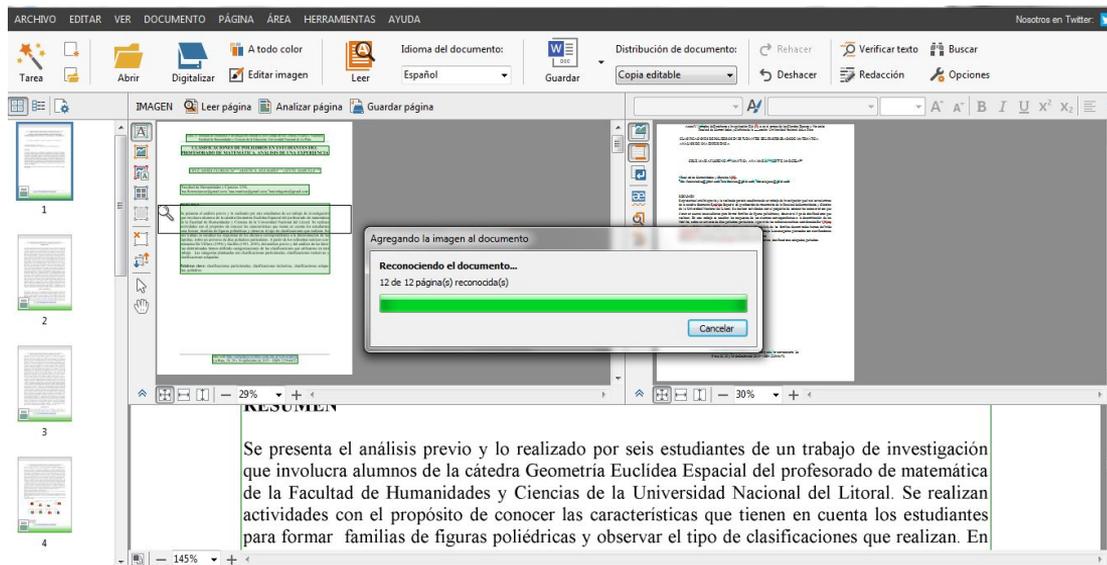
1. Rotar páginas
2. Enderezar las imágenes
3. Ajustar márgenes
4. Eliminar manchas, puntos indeseados
5. Normalizar color
6. Mejorar contraste entre texto y fondo

Para realizar estas acciones se utilizan los productos Scantailor Advanced y Adobe Photoshop CS6.



Etapa 4. de post-proceso

Las imágenes en alta calidad, normalizadas en formato, resolución y color, son incorporadas luego al software de **OCR** para su procesamiento. Se utiliza un método semi-automático de OCR, que consiste en escanear las imágenes con el software Abbyy FineReader y revisar manualmente las elecciones realizadas por el programa. Desde este software se genera el archivo en formato PDF maestro.



Formatos

- Para la preservación digital se guardan los archivos PDF en formato de archivo PDF/A
 - asegura el acceso y la legibilidad a largo plazo de los archivos
 - es un estándar (ISO 19005-1:2005 e ISO 32000-1)
 - está 100% auto-contenido: se puede ver todo su contenido sin hacer uso de fuentes o recursos externos.



PDF/A-1



PDF/A-2



PDF/A-3

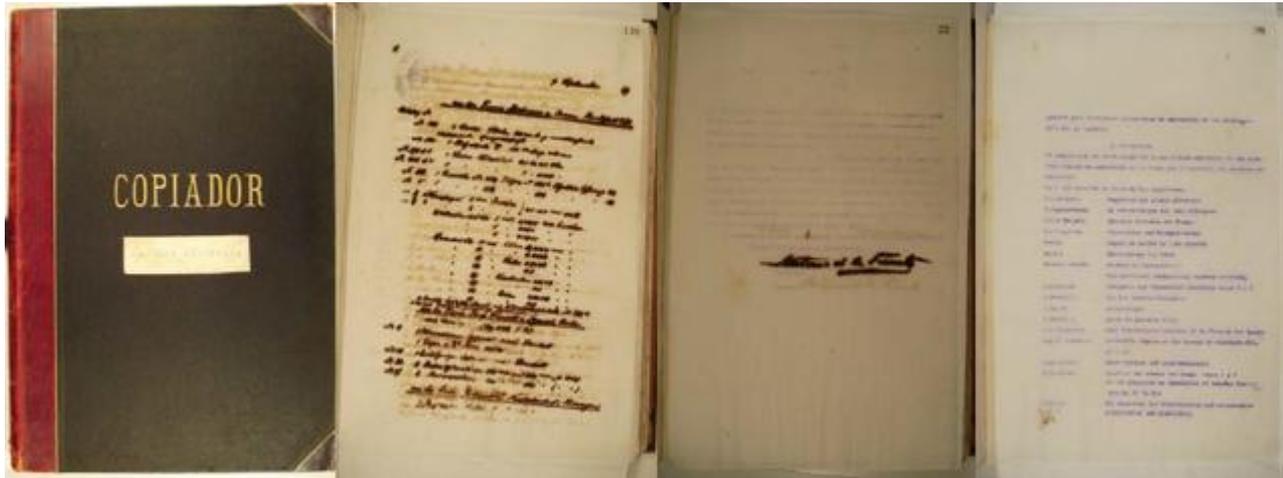
El guardado de los archivos

- El guardado de los archivos se realiza en un espacio de almacenamiento compartido en la nube.
- Se guardan 3 archivos
 - el primero es un archivo TIFF sin comprimir
 - segundo un archivo PDF/A sin compresión
 - un archivo PDF/A con compresión para la difusión
 - ◆ más amigables para el usuario final y para la indexación por parte de crawlers.



Etapas finales del proceso

1. Reunir al material que debe ser subido al repositorio
2. Organizar el material que resta ser devuelto a su aportante original.
 - a. Este estado es extremadamente importante ya que todos los materiales antiguos tienen un elevado valor cultural para la institución y deben ser devueltos lo antes posible a su responsable y/o curador.
3. Una vez completados estas etapas se cierra la tarea y se da por completado el trabajo.



Conclusiones

- Flujo de trabajo ordenado y administrado a través de Redmine
 - mejor aprovechamiento de los recursos
 - mayor certeza a la hora de saber
 - qué hardware utilizar,
 - qué metodología de edición tomar y
 - qué software resulta más conveniente
- El aumento de número de escáneres
- mejoramiento de los escáneres construidos por el personal de SEDICI
 - aumento exponencial en el uso
 - posibilidad de digitalizar un mayor número de documentos en un complejo estado de conservación.
- En los procesos posteriores
 - optimización y ordenamiento,
 - etapa de guardado
 - preservación de los archivos digitales
- La existencia de un Manual de Procedimientos de Digitalización permite al personal contar con un documento al cual recurrir como guía de mejores prácticas y permite, al mismo tiempo, transparentar y dar cuenta de los procesos llevados a cabo con los documentos digitalizados.



Carlos Nusch

carlosnusch@prebi.unlp.edu.ar



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



UNIVERSIDAD
NACIONAL
DE LA PLATA



CESGI
Centro de Servicios en
Gestión de Información
cesgi@cc-poa.gov.ar



PREBI SEDICI
prebi.unlp.edu.ar sedici.unlp.edu.ar



CIC COMISIÓN DE
INVESTIGACIONES CIENTÍFICAS
Ministerio de Ciencia, Tecnología e Innovación