

ESCUELA COMPLUTENSE LATINOAMERICANA

*preservación
digital de
documentos*
archivos, bibliotecas y museos

Hitos y desarrollo de una BD: *selección*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Antecedentes

- En la digitalización concurren demandas y requisitos que deben equilibrarse. Cuando se seleccionan materiales originales para su digitalización se presentan dos cuestiones básicas, si los documentos:
 - Deben convertirse
 - Pueden convertirse

Antecedentes

- Las cuestiones implicadas en la selección de los materiales que se van a digitalizar se examinarán desde dos puntos de vista:
 - Principales razones para digitalizar
 - Criterios de selección

Ejercicio

Supuesto

**Hay una colección de 100.000 documentos
fotográficos, y se deben digitalizar 10.000 en 4 meses,
por tanto...**

Ejercicio

1. *Describir las principales razones que animan el proyecto. ¿Se trata de mejorar el acceso, de apoyar la preservación, o de ambas?*
2. *Crear un conjunto de criterios de selección y tomar en cuenta los recursos humanos y materiales, tiempo necesario, velocidad...*
3. *Considerar la creación de un equipo de especialistas y otros investigadores que representen a usuarios potenciales de estos archivos digitales y que ayuden a seleccionar*
4. *Capturar imágenes que se puedan utilizar para satisfacer distintas necesidades y almacenarlas en un soporte barato y seguro (copia maestra)*



Principales razones para la digitalización

- MEJORAR EL ACCESO
- FACILITAR LAS FORMAS DE USO Y ACCESO.
- PRESERVACIÓN

Principales razones para la digitalización

Mejorar el acceso

- La mejora del acceso a un fondo concreto
- La creación de un único punto de acceso a documentos de distintas instituciones
- “Reunificación virtual” de colecciones y fondos de procedencia única, pero que en la actualidad están dispersos

Principales razones para la digitalización

Mejorar el acceso

- Democratización de registros públicos más ampliamente accesibles
- La ampliación de la disponibilidad de los materiales como apoyo a la educación y a otros proyectos de investigación

Principales razones para la digitalización

Facilitar nuevas formas de acceso y uso

- Posibilitar el uso de materiales (manuscritos, impresos antiguos, mapas únicos, objetos museísticos, libros raros, autógrafos originales, etc.):
 - Cuyos originales sólo pueden consultar visitando sus depósitos específicos
 - Que se han deteriorado y necesitan apoyo tecnológico para mostrar su contenido o forma (recuperación de datos)

Principales razones para la digitalización

Facilitar nuevas formas de acceso y uso

- En estos casos el interés debe centrarse principalmente en cómo añadir valor al material original y no en la digitalización como tal

Principales razones para la digitalización

Preservación

- Estas reproducciones deben satisfacer tanto a los usuarios actuales como a los potenciales usuarios del futuro, y por tanto deben ser de una gran calidad y poseer una gran estabilidad física de modo que puedan mantenerse a lo largo del tiempo
- Un método de selección de materiales originales para su preservación es clasificarlos en tres categorías:

Principales razones para la digitalización

Preservación

1. Documentos raros, únicos o frágiles, archivos (como colección documental), otros objetos de valor material que necesiten conservarse en su forma original
2. Materiales originales muy deteriorados de un alto valor intelectual pero de un bajo valor material y con un bajo nivel de uso
3. Materiales originales sin alto valor material pero que se usan mucho

Criterios de selección

- Cuando se planifica un proyecto de digitalización es útil tener en cuenta las políticas establecidas por otras instituciones para sus propios proyectos. Muchos de estos proyectos están actualmente disponibles para su consulta en la Web



Criterios de selección

- Un ejemplo es el de la **Universidad de Columbia** que ha desarrollado un conjunto de criterios de selección para la digitalización de imágenes dividido en seis categorías:
 - *desarrollo de la colección*
 - *valor añadido*
 - *derechos de propiedad intelectual*
 - *preservación*
 - *viabilidad técnica*
 - *control intelectual*

Criterios de selección

- Otro ejemplo es la Biblioteca del Congreso de los EEUU donde la selección para la preservación de la reconversión digital se basa en:
 - *el valor*
 - *el uso*
 - *las características del documento original y la conveniencia de la reproducción digital para el uso*
 - *el acceso*

Hitos y desarrollo de una BD: *digitalización*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Marca y modelo



DigiBook



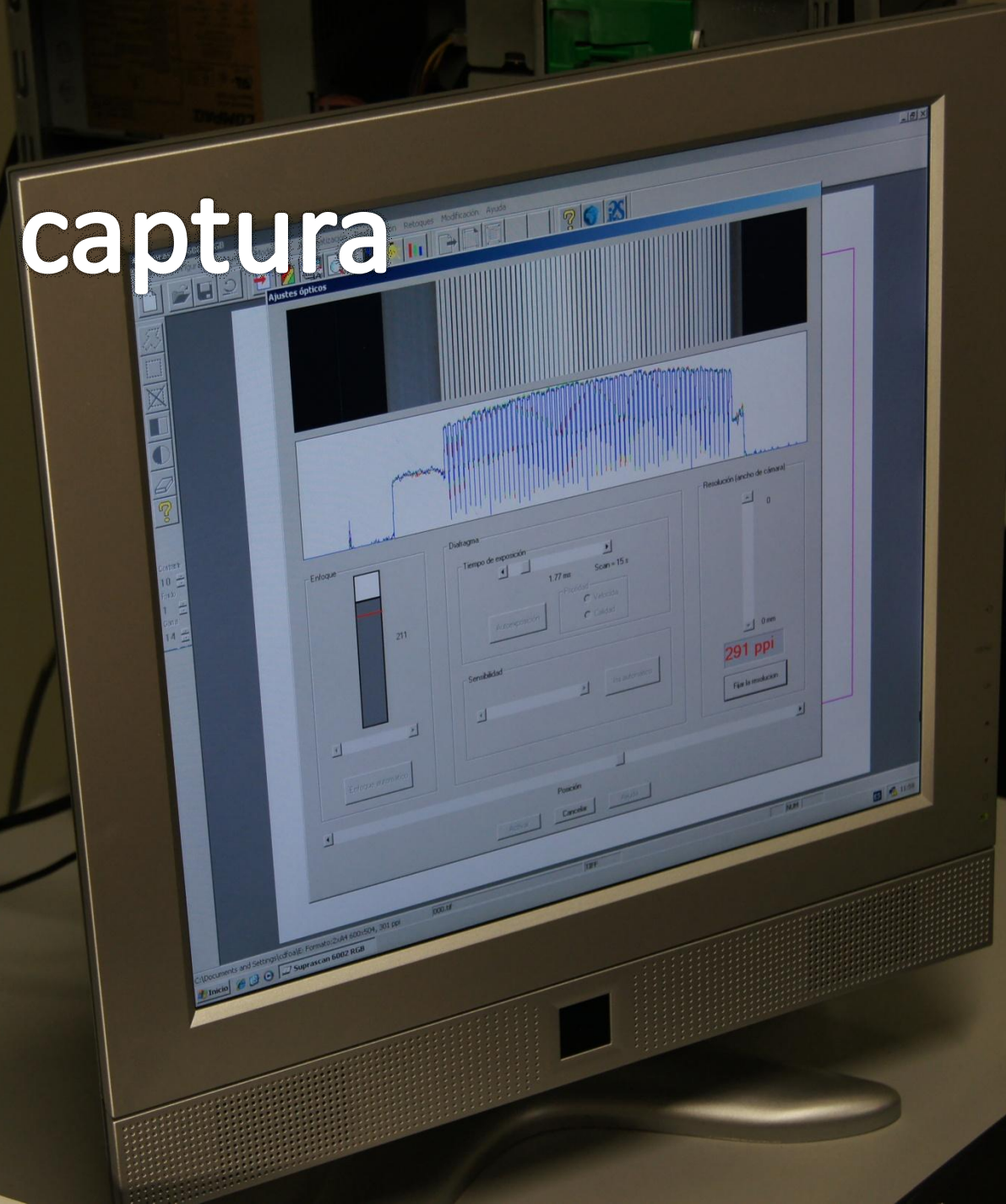
SUPRASCAN

6002 X

Sist. captura



Soft. captura



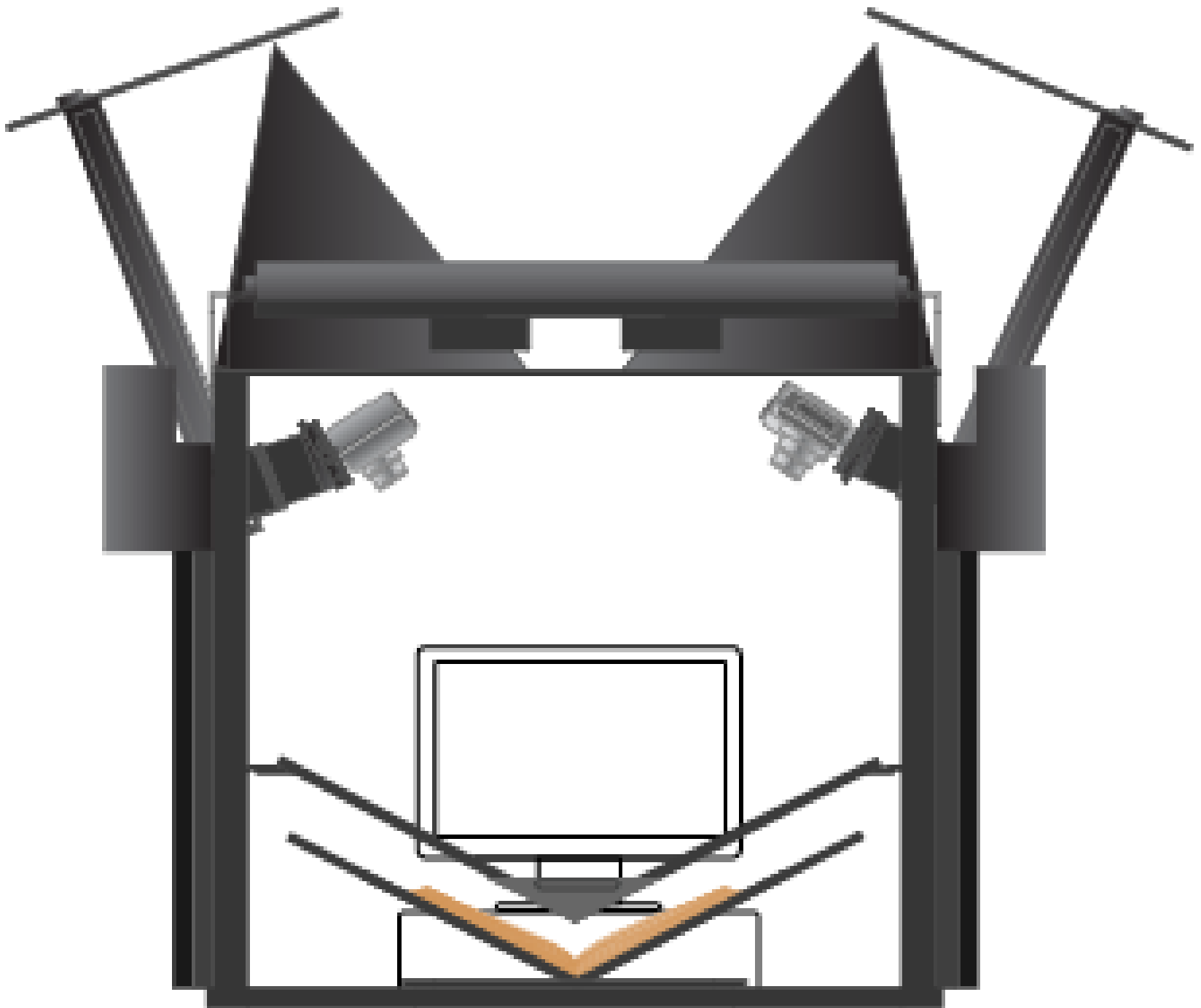


Mesa de trabajo



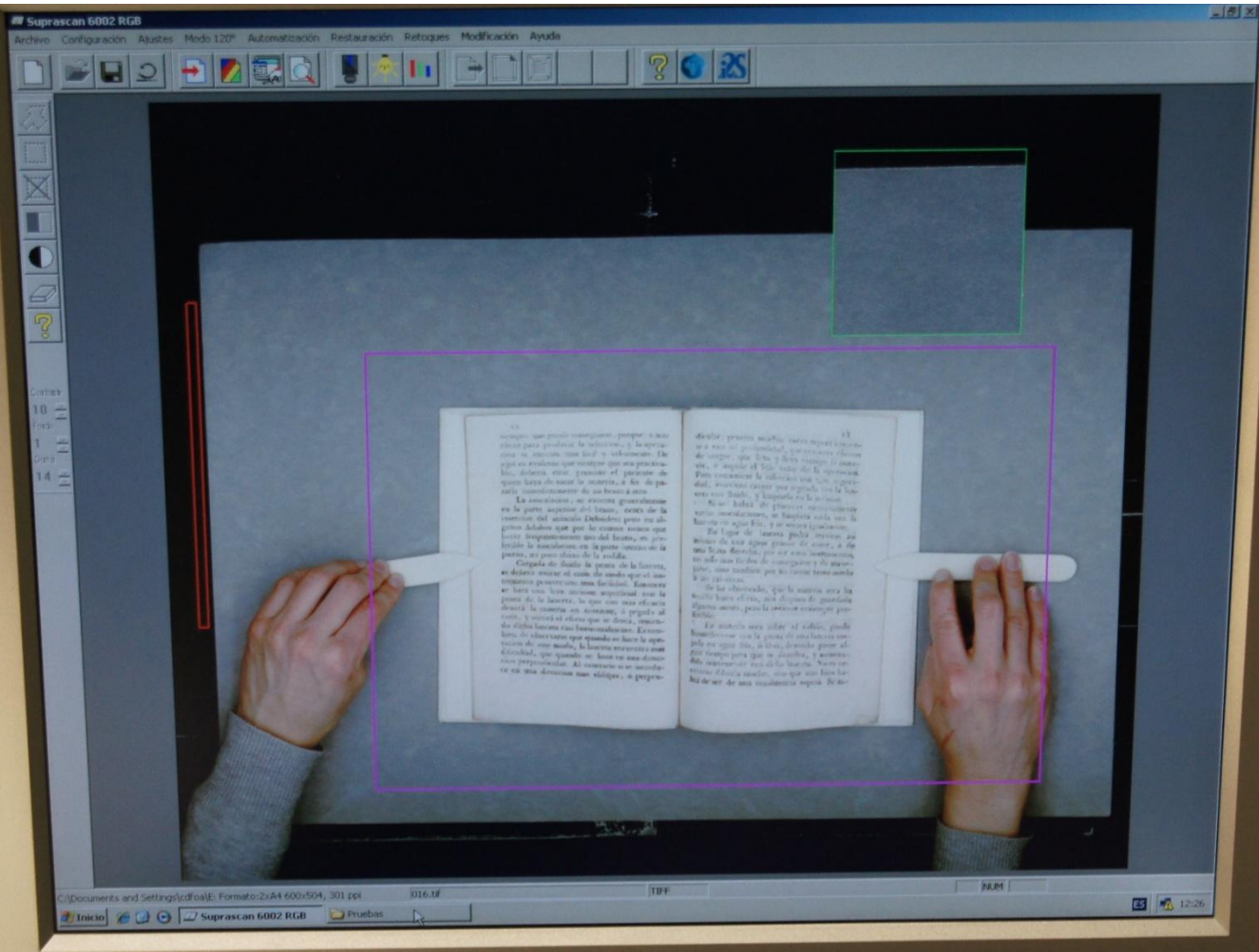
Sist. de iluminación





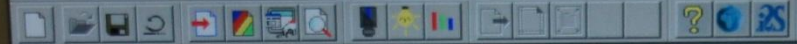






Suprscan 6002 RGB

Archivo Configuración Ajustes Modo 120° Automatización Restauración Retoques Modificación Ayuda



CHERRY
10
113
1
100%
14

C:\Documents and Settings\cdf\MyE... Formato:2x44 600-504, 301 ppi 0116.MJ [PDF] NUR | ES 12:26

Inicio Suprscan 6002 RGB Pruebas

Digitalización

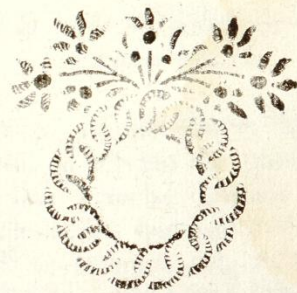
- Se obtienen imágenes en formato TIFF sin compresión a doble página en RGB 24 bits, con resolución óptica de impresión de 300 dpi., en gris o a color, dependiendo del tipo de obra
- Para ejemplares gráficos, no verbales, se fija una resolución óptica mayor, 400 dpi

INSTRUCCIONES

PARA

LA INOCULACION

VACUNA.



BUENOS-AYRES

IMPRESA DE NIÑOS EXPOSITOS

AÑO DE 1813.

Digitalización

- Utilizar carta de color





UNIVERSIDAD COMPLUTENSE



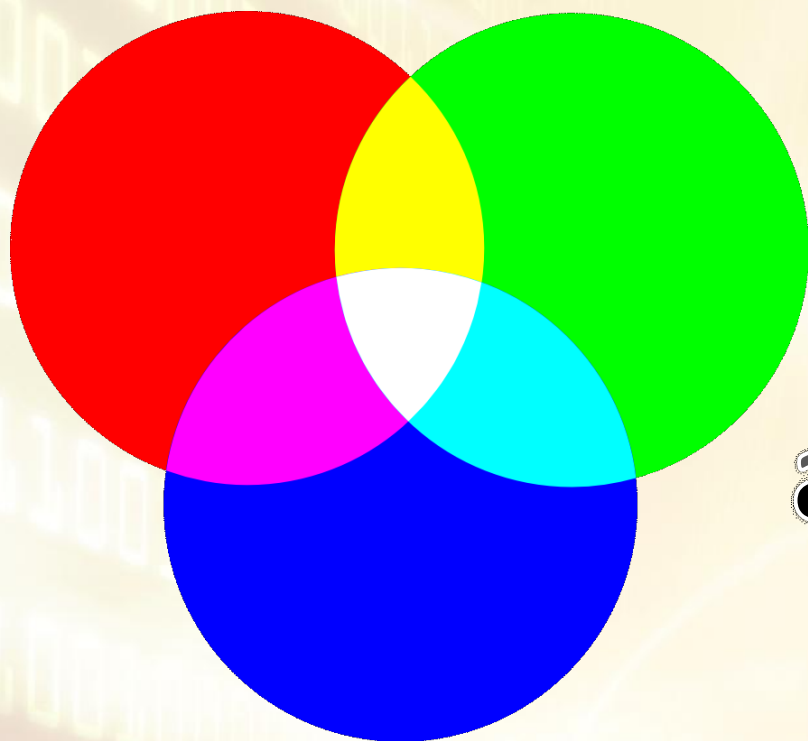
5325290792

b 23699887
i 35220399

UCM
Biblioteca
Histórica
3321
FG

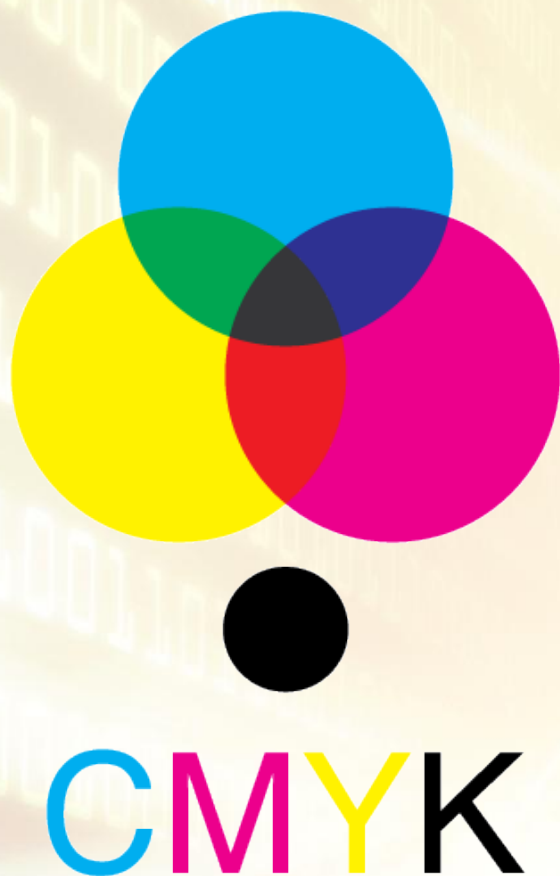
ColorChecker CLASSIC

Digitalización



color
aditivo/síntesis

Digitalización



color
sustractivo

Digitalización

- Utilizar carta de color
- Usar cartulina de color negro/opaco para visualizar de forma correcta los bordes de las obras
- En el caso de haber faltas de material (por ejemplo, orificios por xilófagos) se interpondrán hojas de papel japonés para evitar ver las siguientes páginas
- Almacenar en un servidor de forma simultánea

Digitalización

TIPO DE DOCUMENTO	OBJETIVO	RESOLUCION	PROFUNDIDAD DE COLOR	NOTAS
Texto impreso SIN ilustraciones, prensa, panfletos, páginas mecanografiadas	Imagen del Texto	300 ppp mínimo	Escala de grises 8 bits *	*Color (24 bits) cuando el color sea una característica importante del documento
	Texto con OCR	400 ppi	Escala de grises 8 bits *	
Música: partituras, escalas anotadas, manuscritos de música	Acceso al contenido	300 ppp mínimo	Escala de grises 8 bits*	
	Reconocimiento de sus características materiales	400 ppi	Escala de grises 8 bits*	
Manuscritos: escritos a mano, copias mecanografiadas	Acceso al contenido	300 ppp mínimo	Escala de grises 8 bits*	
	Reconocimiento de sus características materiales	400 ppp	Escala de grises 8 bits*	

Digitalización

Mapas: caracteres impresos color impreso hasta un tamaño 56 cm x 87 cm	Búsqueda	250 ppp mínimo *	24-bit color	*La resolución (ppp) depende del tamaño del mapa, sobre todo en los casos en los que las secciones del mapa tienen que unirse y el tamaño del archivo sobrepase los 500 MB
	Reproducción	400 ppp	24-bit color mínimo	

Digitalización

Fotografías: tono continuo, color	Acceso al contenido	300 ppp mínimo	Escala de grises 8 bits*	*Color (24 bits) cuando el color sea una característica importante del documento
	Reproducción	Máximo soportado	24-bit color mínimo	
Material gráfico	Acceso al contenido	300 ppi mínimo	Escala de grises 8 bits*	*Color (24 bits) cuando el color sea una característica importante del documento
	Reproducción	Máximo soportado	24-bit color	

Hitos y desarrollo de una BD: *control de calidad*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Calidad

- Control de calidad de las imágenes TIFF MÁSTER una a una, para detectar saltos de página, desenfoques...

Transversal, aunque después de la digitalización es cuando se debe prestar más atención...

Hitos y desarrollo de una BD: *gestión de la colección*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Gestión

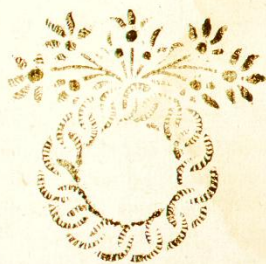
- Enderezar las imágenes con algún programa (LIMB Processing en BNE, por ejemplo)
- Según el material, se digitaliza a una o dos páginas, después se recortan, es decir, se genera un fichero por cada página
- Cada imagen resultante de la digitalización (tanto TIFF MÁSTER como TIFF recortado), debe ser identificado por su signatura

INSTRUCCIONES

PARA

LA INOCULACION

VACUNA.



BUENOS-AYRES

IMPRESA DE NIÑOS EXPOSITOS

AÑO DA 1813.

BH-FG-3321_005.TIFF

Gestión

- Revisar de nuevo las imágenes tras el primer recorte automatizado puede quedar algún marco negro a su alrededor., corregir manualmente



INSTRUCCIONES

PARA

LA INOCULACION

VACUNA.



BUENOS-AYRES

EMPRESA DE NIÑOS EXPOSITOS

AÑO DE 1813.

BH-FG-3321_005.TIFF

Gestión

- En los casos en los que se vaya a obtener OCR y, sólo si se considera que se precisa un tratamiento previo para mejorar su calidad, las imágenes deben ser tratadas con el fin de mejorar el texto sin perder información, desechando manchas y suciedad que han sido originadas por el paso del tiempo, el uso y/o la calidad del papel
- Corregir la inclinación del texto

Gestión

- Centrar el texto manualmente y unificar la medida del tamaño de todas y cada una de las imágenes, utilizando una media sobre la medida del tamaño de las páginas, logrando así que sea lo más fidedigno al documento
- Para mejorar los resultados del OCR, las imágenes se procesan automáticamente para convertirlas a blanco y negro. En el proceso se eliminan la suciedad o manchas que hayan podido quedar después del tratamiento de las imágenes TIFF

Hitos y desarrollo de una BD: *presentación y publicación*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Publicación

- Se generan los archivos de difusión, en formato PDF (125 ppp) o JPEG
- En el caso de que se digitalicen monografías (fondo moderno) se realizará el proceso de OCR
- Se crean los marcadores en los PDF
- Se inserta la marca de agua a través de un archivo de imagen GIF o similar

Hitos y desarrollo de una BD: *mantenimiento a largo plazo*

Lea Plangger

Universidad Complutense de Madrid

leaplanger@edu.ucm.es

Mantenimiento

- Se genera la estructura de metadatos de preservación PREMIS de cada TIFF MÁSTER
- Se relaciona posteriormente cada archivo de difusión PDF/JPEG con su registro/s MARC, generando la estructura METS, MARC, etc.
- Antes de la carga en el SGOD, se realiza el control de calidad del 20% de los títulos de un lote y dentro de esa muestra se revisa un 20% de páginas incluyendo marcadores

Mantenimiento

- Para la BNE la tasa de fiabilidad de la imagen será del 99, 25% y la de los marcadores del 100%

