

Estrategias de digitalización de fondos patrimoniales



conservar - preservar - enriquecer - transmitir

Maria Susana Seguin

Universidad Paul-Valéry Montpellier III
IHRIM UMR 5317 CNRS - ENS de Lyon

Institut Universitaire de France

susana.seguin@ens-lyon.fr

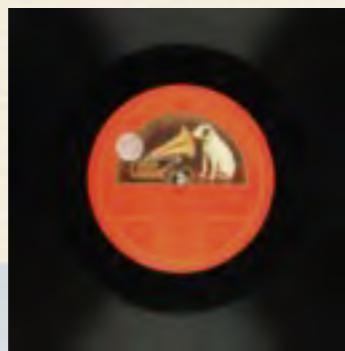
Antes de comenzar...

Qué es el patrimonio digital?

« *El patrimonio digital* está formado por los materiales informáticos de valor perdurable dignos de ser conservados para las generaciones futuras, y que proceden de comunidades, industrias, sectores y regiones diferentes. No todos los materiales digitales poseen valor perdurable, pero los que lo tienen exigen metodologías de conservación activas para mantener la continuidad del patrimonio digital. »



Fondos digitales, fondos digitalizados



EL PAÍS
THE TIMES
Le Monde.fr
LA NACION

Tout le monde parlait du livre. Dans les rues de New York, je ne pouvais plus déambuler en paix, je ne pouvais plus faire mon jogging dans les allées de Central Park sans que des promeneurs me reconnaissent et chuchotent : « Hé, c'est Goldman ! C'est l'écrivain ! » Il arrivait même que certains émittent quelques pas de course pour me suivre et me poser les questions qui les taraudait : « Ce que vous y dites, dans votre bouquin, c'est la vérité ? Harry Québec a vraiment fait ça ? » Dans le café de West Village où j'avais mes habitudes, certains clients n'hésitaient plus à s'asseoir à ma table pour me parler : « Je suis en train de lire votre livre, Monsieur Goldman : je ne peux pas m'arrêter ! Le premier était déjà bon, mais alors celui-là ! On vous a vraiment filé un million de dollars pour l'écrire ? Vous avez quel âge ? Trente ans à peine ? Trente ans ! Et vous avez déjà amassé tellement de pogos ! » Même le portier de mon immeuble, que

Carta sobre la preservación del patrimonio cultural de la Unesco

Artículo 1 – Alcance

El patrimonio digital consiste en recursos únicos que son fruto del saber o la expresión de los seres humanos. Comprende recursos de carácter cultural, educativo, científico o administrativo e información técnica, jurídica, médica y de otras clases, que se generan directamente en formato digital o se convierten a éste a partir de material analógico ya existente. Los productos “de origen digital” no existen en otro formato que el electrónico.

Los objetos digitales pueden ser textos, bases de datos, imágenes fijas o en movimiento, grabaciones sonoras, material gráfico, programas informáticos o páginas Web, entre otros muchos formatos posibles dentro de un vasto repertorio de diversidad creciente. A menudo son efímeros, y su conservación requiere un trabajo específico en este sentido en los procesos de producción, mantenimiento y gestión.

Medidas necesarias

Artículo 6 - Elaborar estrategias y políticas

Es preciso elaborar estrategias y políticas encaminadas a preservar el patrimonio digital, que tengan en cuenta el grado de urgencia, las circunstancias locales, los medios disponibles y las previsiones de futuro. La colaboración de los titulares de derechos de autor y derechos conexos y otras partes interesadas a la hora de definir formatos y compatibilidades comunes, así como el aprovechamiento compartido de recursos, pueden facilitar esa labor.

Artículo 7 - Seleccionar los elementos que deben conservarse

Al igual que ocurre con el conjunto del patrimonio documental, los principios de selección pueden diferir de un país a otro, aun cuando los principales criterios para determinar los elementos digitales dignos de conservación sean su significado y valor duraderos en términos culturales, científicos, testimoniales o de otra índole. Indudablemente, se deberá dar prioridad a los productos “de origen digital”. Los procesos de selección y de eventual revisión subsiguiente han de llevarse a cabo con toda transparencia y basarse en principios, políticas, procedimientos y normas bien definidos.

Artículo 8 - Proteger el patrimonio digital

Los Estados Miembros han de disponer de mecanismos jurídicos e institucionales adecuados para garantizar la protección de su patrimonio digital.

Hacer que la legislación sobre archivos, así como el depósito legal o voluntario en bibliotecas, archivos, museos u otras instituciones públicas de conservación, se aplique al patrimonio digital, ha de ser un elemento esencial de la política nacional de preservación [...]

Artículo 9 - Preservar el patrimonio cultural

Por definición, el patrimonio digital no está sujeto a límites temporales, geográficos, culturales o de formato. Aunque sea específico de una cultura, cualquier persona del mundo es un usuario en potencia. Las minorías pueden dirigirse a las mayorías y los individuos a un público de dimensión mundial.

Hay que preservar y poner a disposición de cualquier persona el patrimonio digital de todas las regiones, naciones y comunidades a fin de propiciar, con el tiempo, una representación de todos los pueblos, naciones, culturas e idiomas.

Confirmado por...

- ❖ Conferencia Unesco « The Memory of the World in the Age of Digitization and Preservation » (Vancouver, 2012)
http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_ubb_vancouver_declaration_en.pdf
* « Digitization, while not in itself a major means of preservation, can protect invaluable documents from handling and further deterioration. For some types of material, it is the only means of ensuring their survival »
- ❖ Comunicación de la comisión europea sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital, 2006.
http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=162

1. Qué se puede digitalizar?

- ❖ Papel: libros, imágenes, periódicos, revistas, notas, etc
- ❖ Objetos: maquetas, estatuas, monedas, etc.
- ❖ Audio / Video: Cassettes, VHS, Disquettes, CR-Roms, etc
 - * Que necesitan instrumentos de lectura
 - * Que ya disponen de versiones digitales
- ❖ Originales escasos? Únicos? Frágiles?
- ❖ La versión digital es fiel a la realidad?

En qué contexto?

- ❖ Para su conservación o para preservar las condiciones de utilización?
- ❖ Documentos bajo derechos de autor o libres?
- ❖ En el marco de una convención con una institución pública? Privada? con el autor?
- ❖ En el marco de un proyecto de investigación?

Recuperación de los primeros documentos digitales

- ❖ CD-Roms, bases de datos:
 - * Todavía disponibles?
 - * Técnicamente accesibles?
 - * Todavía útiles?
 - * Reemplazados por una versión en línea?

Libros y periódicos

- ❖ Comprados o accesibles en línea?
- ❖ Bajo licencia?
- ❖ Catalogados? Indexados?
- ❖ Libres o protegidos?
- ❖ Formato? PDF, HTML, Epub, indefinido?
- ❖ Qué relación a la versión papel? Idéntica o similar a la versión original?
Versión totalmente original?
- ❖ Publicaciones electrónicas nativas?

Documentos digitales para la investigación

- ❖ Contenidos complejos...
 - * Se aplican a un área específica, a veces a un solo investigador
 - * Cantidad importante de datos
- ❖ Pero particularmente ricos y útiles
 - * Posibilidad de hacer avanzar los conocimientos técnicos
 - * « Big data »

El caso de los archivos

- ❖ Archivos públicos
- ❖ Archivos privados, comprados o legados
- ❖ Qué derechos sobre los archivos?
 - * Conservación durante un periodo determinado de utilización?
- ❖ Qué variedad de documentos? (papel, emails, documentos / interfaces electrónicas?)
- ❖ Administrado por una biblioteca? Una universidad? Una institución, o un organismo público?

Internet

- ❖ Posibilidad de prever el depósito legal de publicaciones en línea: exposiciones, documentos pedagógicos, blogs culturales, etc.

Necesidad de conservar de manera perenne

LONGEVIDAD CONTRA DENSIDAD



Obsolescencia, fragilidad, usura...

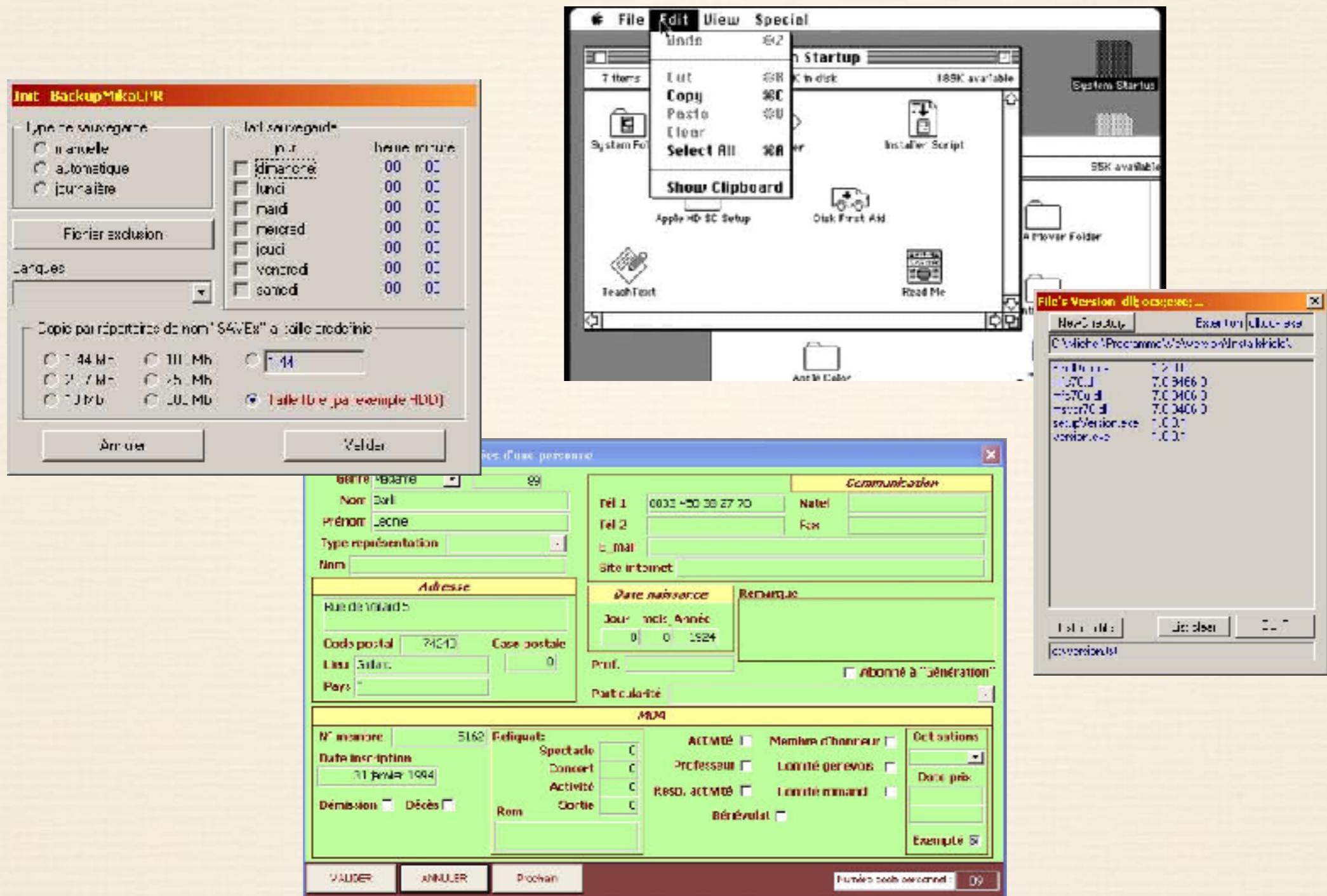
Riesgo de obsolescencia del material informático



Desaparición del material capaz de leer los soportes o de hacer funcionar los programas necesarios a la lectura de datos

Obsolescencia de softwares

que permitan leer los datos en los formatos en que han sido conservados



Obsolescencia de los datos

Necesidad de conservar una descripción mínima de la información

- ❖ Formato
- ❖ Estado de conservación (corrupción?)
- ❖ Identificación
- ❖ Historia de las modificaciones
- ❖ Estructuración
- ❖ Relación con otros documentos
- ❖ Etc.

2. Antes de digitalizar ...

- ❖ Importancia de los medios técnicos elegidos:
 - * En función de los objetivos perseguidos (conservación, preservación, difusión, investigación)
 - * En función de los medios técnicos disponibles (Formato, resolución, color, etc).
 - * Según los fines de la digitalización (modo imagen, texto con OCR, índice, enriquecimiento de tipo REN, edición científica, etc)

Plan de conservación digital

- ❖ Supone varios ejes de reflexión:
 - * Medios económicos
 - * Medios humanos (formación continua)
 - * Gestión del proyecto

Quién hace qué?

- ❖ Los **informáticos** proveen el servicio técnico, no la definición del servicio...
- ❖ Los **bibliotecarios, archivistas, investigadores** (por separado o conjuntamente) definen las metodologías y estrategias del trabajo:
 - * Estrategias preventivas
 - * Planes de conservación
 - * Definición de prioridades en función del valor / fragilidad de las colecciones.
 - * Gestión de riesgos
 - * Definición de utilización de los fondos digitalizados (consulta, explotación, anotación, textometría, etc).
 - * Alertas, test de instrumentos, detección de problemas, etc.
 - * Planes de urgencia

Verdadero trabajo previo de restauración digital

- ❖ Asegurarse del estado de conservación del papel: desacidificación, tintas ferrogálicas, estado de los cuadernillos y del encuadernado, marcas de goma de borrar sobre la celulosa del papel...)
- ❖ Conservar documentos digitales, es preservarlos...
- ❖ Pero digitalizar y guardar, no es archivar ...

3. Cómo digitalizar?

Una estrategia técnica

La cadena de digitalización

- ❖ Digitalización
- ❖ OCR
- ❖ Índices
- ❖ Documentos «nómades»
- ❖ Entrega final

La digitalización

- ❖ Seleccionar los documentos
- ❖ Preparación de los documentos
- ❖ Recuperación y transporte de los documentos (cuidado con las condiciones climáticas y sanitarias!)
- ❖ Digitalización
- ❖ Retorno de los documentos a la institución de conservación y verificación general de los documentos.

El taller de digitalización

- ❖ La digitalización no se hace con un simple scanner familiar...
- ❖ Interna o por prestatario exterior?
- ❖ Si prestación de servicio:
 - * En la institución?
 - * En un taller exterior? Importancia de las condiciones de transporte. Eventualmente contratación de seguros...

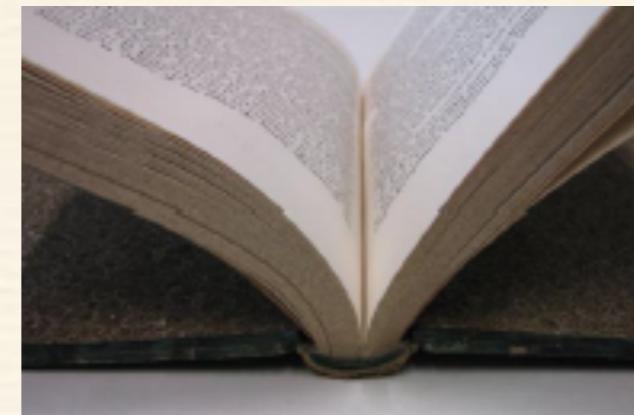
Ejemplos de talleres de digitalización



Como elegir el scanner?

- ❖ En función de los documentos: fragilidad, rareza...
- ❖ Grado de apertura del encuadrado
- ❖ Formato

Grado de apertura y formato de un documento...

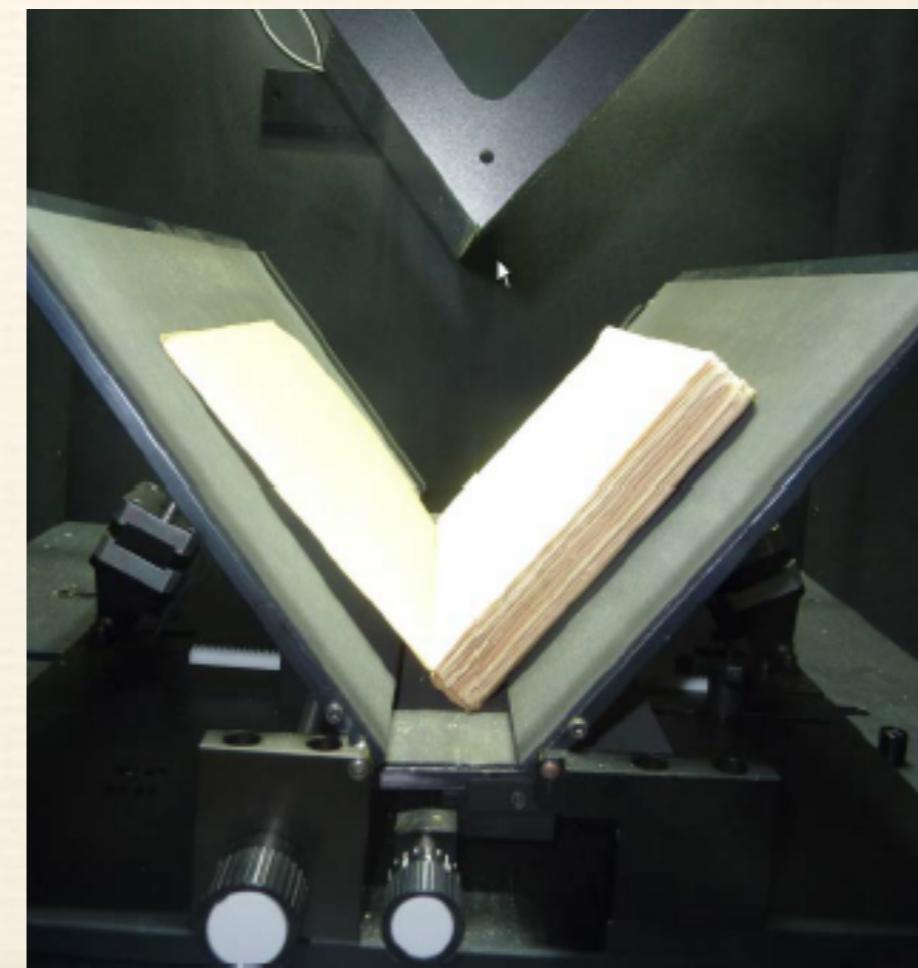
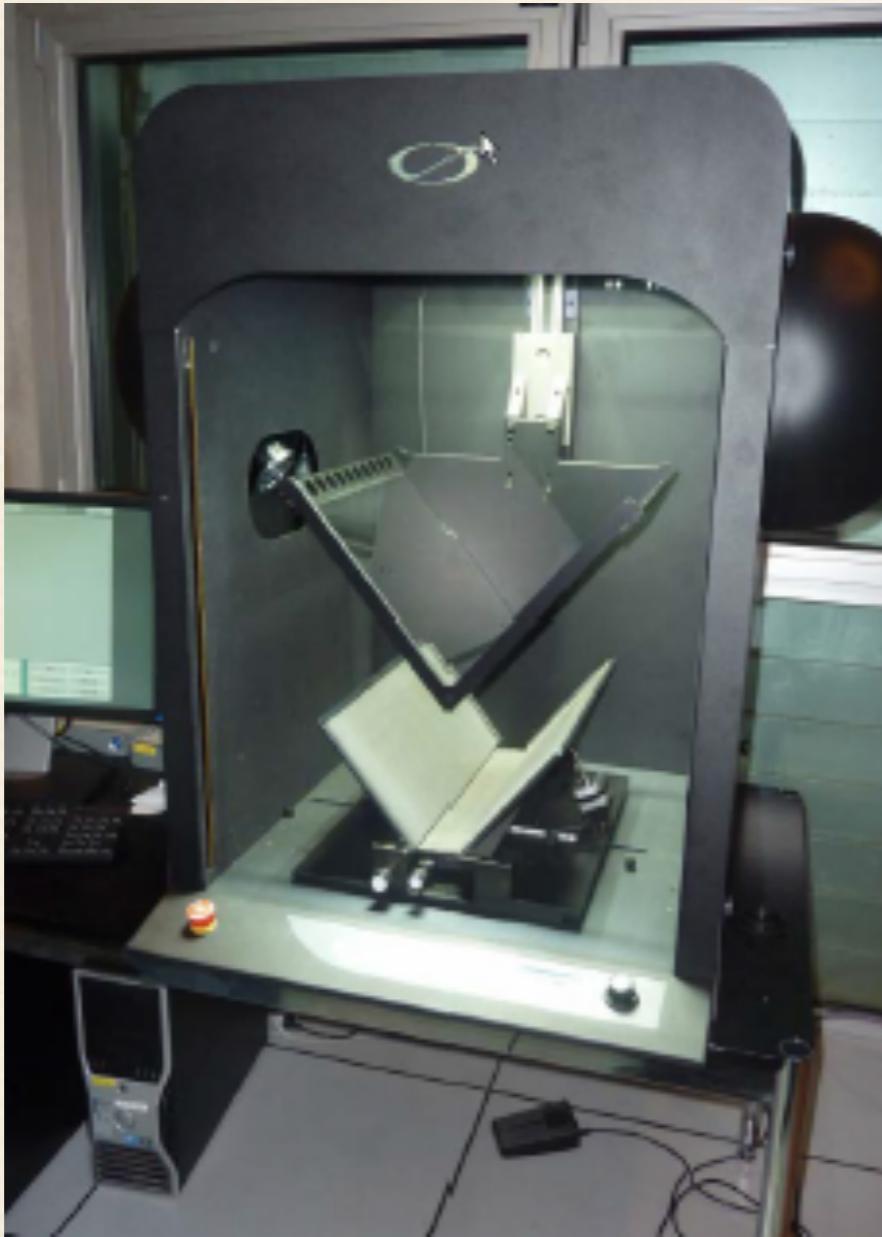


Digitalización plana o apertura tradicional

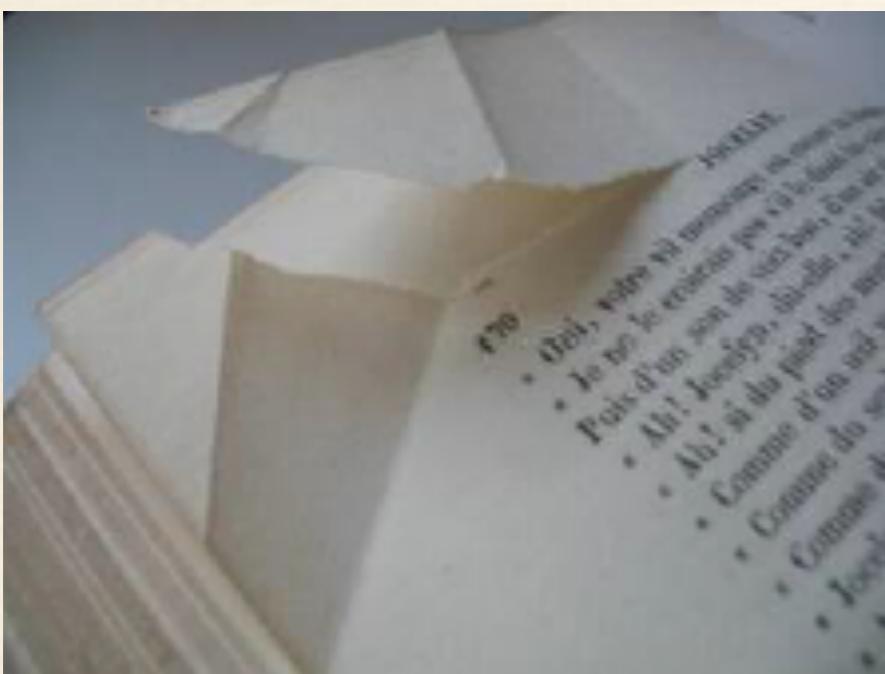
- ❖ [http://www.youtube.com/watch?
v=ck4mpfShLvo&feature=related](http://www.youtube.com/watch?v=ck4mpfShLvo&feature=related)



Quidenus : apertura reducida



Daños frecuentes

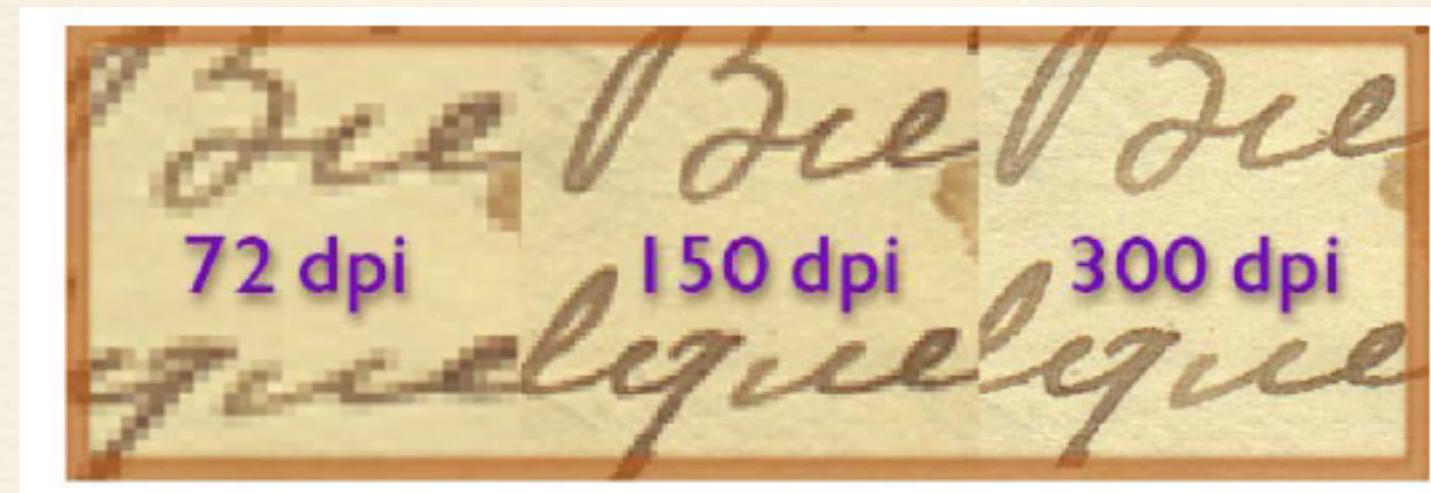


Importancia de la elección del formato

- ❖ **JPEG** (Joint Photographic Experts Group): norma que define el formato en que se graba y el algoritmo que decodifica la representación numérica de una imagen fija.
- ❖ **PDF** (Portable Document Format): Formato de archivo de descripción de páginas de impresión creadas por la sociedad ADOBE. Su objetivo es conservar el formato de origen del documento (estático o dinámico).
- ❖ **TIFF** (Tag(ged) Image File Format): Formato de archivo para imagen digital, que funciona como un « container »: puede encapsular o contener datos de formatos arbitrarios. Puede comprimir la información o conservarla en estado « bruto ».

Importancia de la elección de una buena definición

dpi: número de pixeles por pulgada



- ❖ 72 dpi para la difusión en línea
- ❖ 150 dpi para impresión simple
- ❖ 300 dpi para impresión a escala real
- ❖ 400 dpi para digitalización de un documento original
- ❖ 600 dpi para ver pequeños detalles
- ❖ Equivalencia a definir para los transparentes en función de los documentos para digitalizar

Reproducción de colores

- ❖ Depende de un gran numero de variables:
 - * El nivel de luz en momentos de la captura
 - * La profundidad de bit capturado (dpi)
 - * Las capacidades del sistema de digitalización
 - * La representación de las características colorimétricas.
- ❖ Caracterización: creación de un perfil que describa el método de reproducción de los colores
- ❖ Calibración: definición de un marco colorimétrico de referencia
- ❖ **Es necesario calibrar el periférico (scanner) conforme a normas predefinidas y en función del resultado esperado...**

Comparemos ...



Principios de la digitalización:
B&N - 300 dpi TIFF G4
1 página : 500 Ko



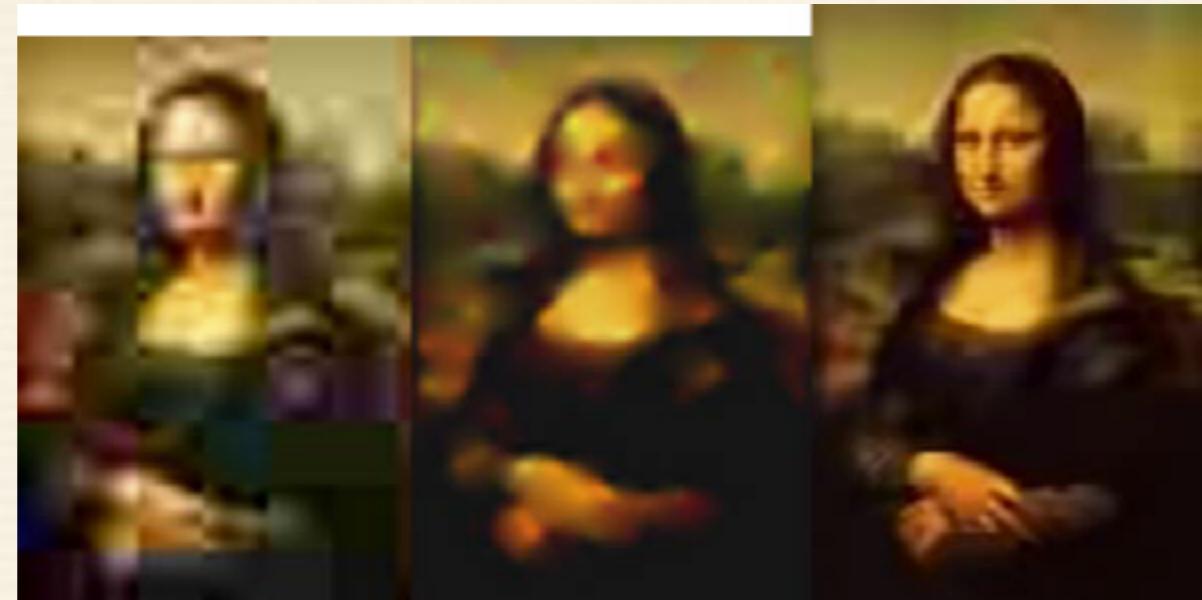
Digitalización actual:
Color - 400 dpi - TIFF no comprimido
1 página: 80 Mo



500 veces más

Compresión de documentos *versus* calidad de digitalización

Con compresión

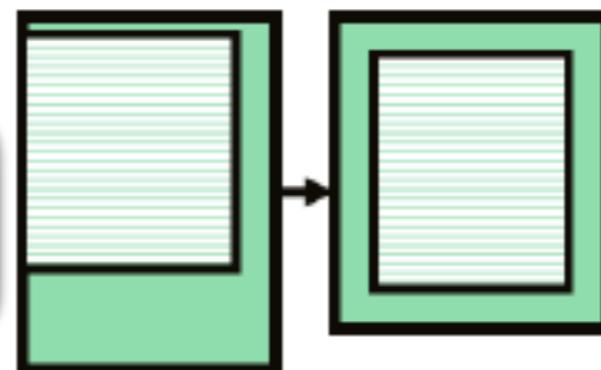


Sin compresión

**Una mayor resolución
supone un mayor número de colores
pero también una imagen más « pesada »
(mayor volumen en octets)
y exigir por consiguiente estructuras
especiales de archivage
(grandes infraestructuras - políticas culturales y científicas)**

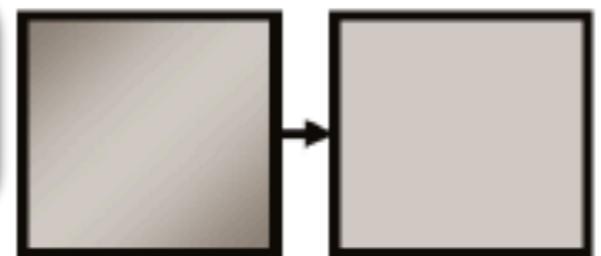
Preparar el documento digitalizado

Recuadrar



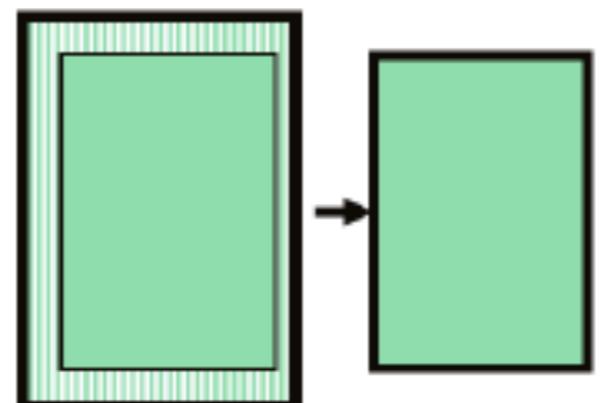
**Enderezar
líneas**

**Corregir
la luz**

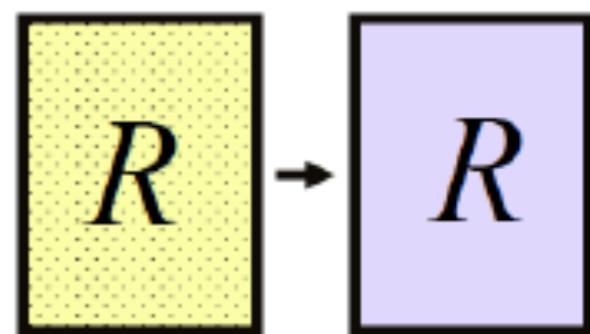
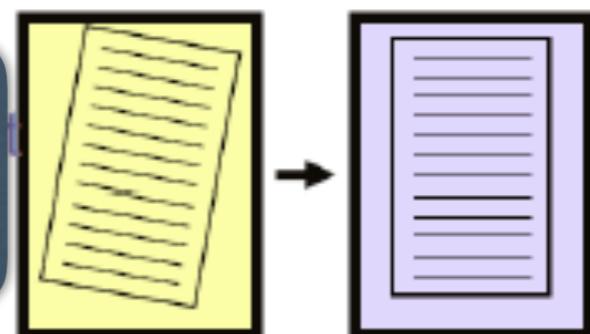
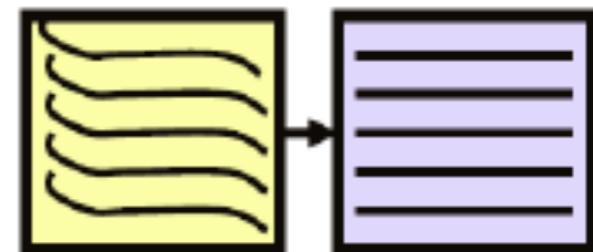


**Enderezar
la imagen**

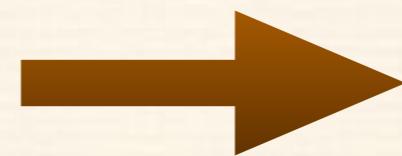
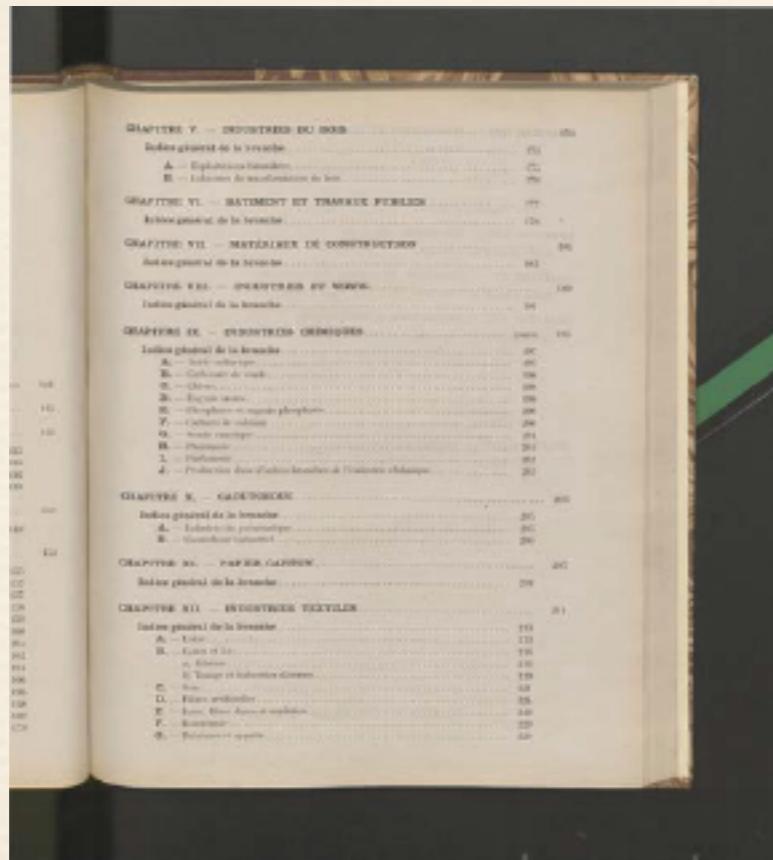
**Suprimir
bordes**



**Limpiar el
fondo**



Resultado de la corrección



CHAPITRE V. — INDUSTRIES DU BOIS		pages
Indice général de la branche		173
A. — Exploitation forestière		173
B. — Industries de transformation du bois		176
CHAPITRE VI. — BATIMENT ET TRAVAUX PUBLICS		177
Indice général de la branche		179
CHAPITRE VII. — MATERIAUX DE CONSTRUCTION		181
Indice général de la branche		183
CHAPITRE VIII. — INDUSTRIES DU VERRE		189
Indice général de la branche		190
CHAPITRE IX. — INDUSTRIES CHEMlQUES		195
Indice général de la branche		197
A. — Acide sulfurique		197
B. — Gélatine de poisson		198
C. — Chaux		198
D. — engrain analysé		199
E. — Phosphates et engrains phosphatés		200
F. — Gélatine de colombe		200
G. — Soude caustique		201
H. — Pharmacie		201
I. — Parfumerie		202
J. — Production divers d'autres branches de l'industrie chimique		202
CHAPITRE X. — CAOUTCHOUC		203
Indice général de la branche		205
A. — Industrie du parapluie		205
B. — Caoutchouc naturel		206
CHAPITRE XI. — PAPIER-CARTON		207
Indice général de la branche		209
CHAPITRE XII. — INDUSTRIES TEXTILES		211
Indice général de la branche		213
A. — Laine		213
B. — Coton et fibres		216
c) Filature		216
d) Tissage et industries diverses		219
E. — Soie		221
D. — Fibres artificielles		223
E. — Jutes, fibres dures et cordères		226
F. — Boussole		227
G. — Teintures et apprêts		229

Standard de archivo para documento digital

- ❖ Archivo origen en TIFF sin comprimir
- ❖ Encabezado por:
 - * Información técnica sobre la imagen
 - * Información de producción
 - * Información administrativa

Tratamiento OCR (Optical Character Recognition)

(BnF Gallica TOUT GALICA ▾ Rechercher... RECHERCHE AVANCÉE | SÉLECTIONS | FR ▾

Accueil > Consultation
Les origines de l'alchimie / par M. Berthelot,....

Recherche dans le document x

sanctuaire

Résultats de recherche
sanctuaire : 15 premières pages trouvées

p.413
Sanctuaires, 118(...) Sanctuaire (livre), 125, 340

p.180
Le prêtre répondit : Je suis le prêtre du sanctuaire et je suis sous le poids de la puissance qui m'accable

p.43
Ce dernier sanctuaire était surtout médical : la parenté étroite qui a toujours existé entre la préparation des médicaments et les études chimiques, nous explique pourquoi les alchimistes regardaient comme leur plus vieux laboratoire(...) un jour quelques indices des pratiques chimiques des Égyptiens, quelques fragments de ces fourneaux que Zosime décrit, d'après ce qu'il a vu lui-même dans le temple de

36 LES ORIGINES DE L'ALCHIMIE grecque ; un autre caractère polyphone, traduit par le *poi* grec, et un troisième polyphonie, aussi très net, le *hōum*, traduit par la *thūta*. Des alphabets magiques analogues existent dans le manuscrit astrologico-alchimique n° 419 de la Bibliothèque nationale, et des alphabets pareils se lisent dans les papyrus thébains de Leide (1). Les noms mêmes des laboratoires où l'on préparait la pierre métallique, c'est-à-dire la pierre philosophale, sont transcrits à la suite d'un traité de Jean l'archiprêtre (2). Les voici : terre de la Thébaïde, Heliopolis, Lycopolis, Aphrodite, Apollinopolis, Elephantine ; ce sont là en effet toutes des villes connues en Egypte et siège de grands sanctuaires. Cette liste semble reproduire du début d'un passage d'Agatharchide, relatif aux exploitations métallurgiques de l'Egypte (3) : peut-être que les lieux où l'on extrayait l'or de ses minerais étaient les mêmes que ceux où l'on préparait le fabriquer. En tous cas la liste est fort ancienne ; car ces noms n'étaient plus guère connus après la conquête musulmane, et il n'y figure aucun lieu étranger à l'Egypte, tels que ceux que nous retrouvons plus tard dans les listes écrites au VIII^e siècle.

Tout ceci nous ramène constamment vers l'Egypte et même vers l'Egypte grecque et hellénisée d'Alexandrie, telle qu'elle existait à l'époque de la domination romaine, aux III^e et IV^e siècles de notre ère.

Cependant dans ces fables, il n'y a, en somme, la preuve

(1) N° 25 ; Bessière, I, 49.
(2) Ms. n° 323, fol. 24v, v.
(3) Voir p. 25. Le manuscrit de saint Marc a donné tout ce passage sous le titre : *Des pierres multiformes* (éd. 158).

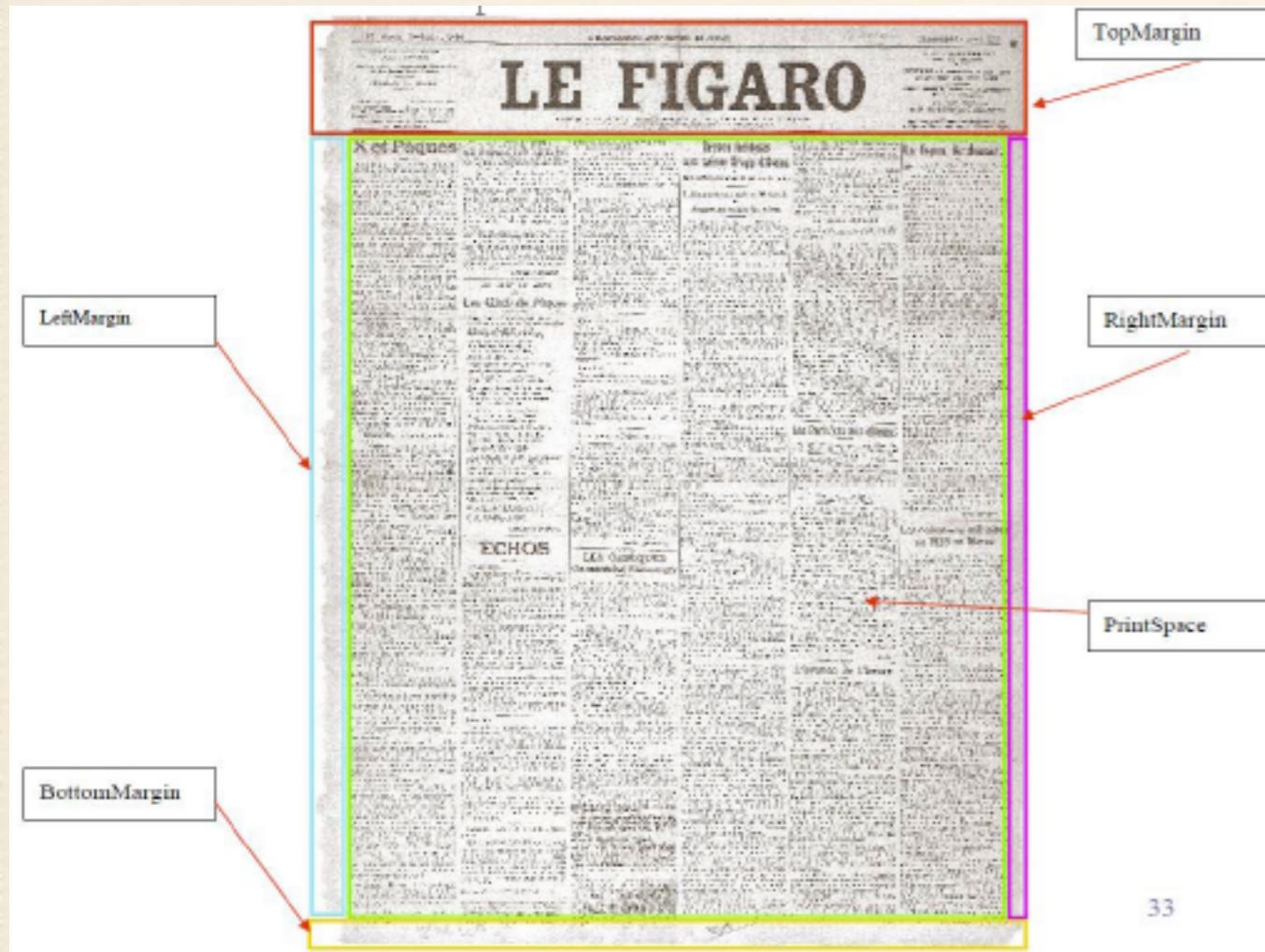
Page | 36

El formato ALTO

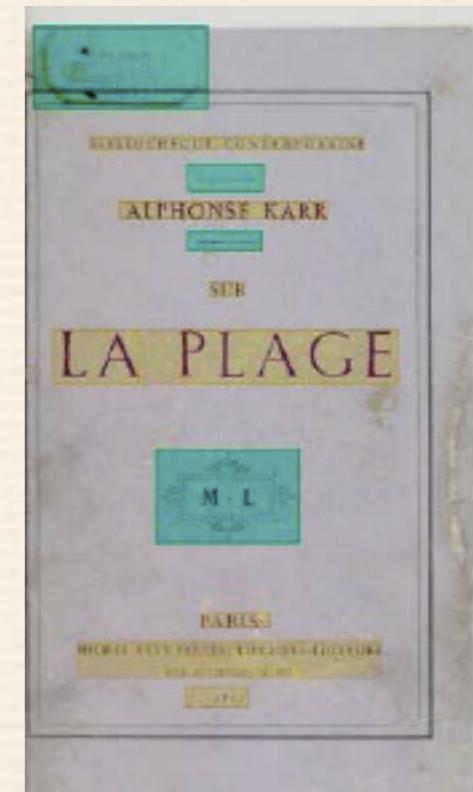
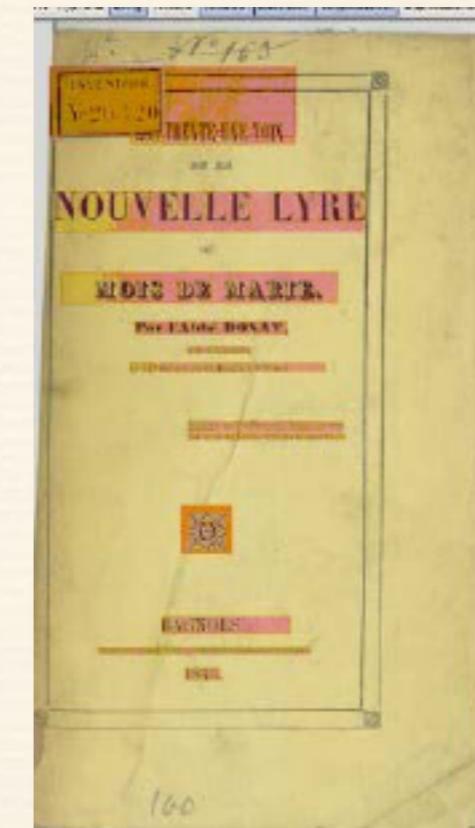
(Analyzed Layout and Text Object)

- ❖ Formato basado en el lenguaje XML, que sirve para codificar el lenguaje, creado por ADOBE pero mantenido por la Biblioteca del Congreso de EEUU.
- ❖ Es el formato más utilizado para convertir el texto contenido en las imágenes de documentos patrimoniales.
- ❖ Conserva las coordenadas geométricas de los contenidos en una imagen (texto, ilustraciones, gráficos) y permite la superposición de la imagen y del texto (color de la palabra buscada en un texto).
- ❖ Permite la segmentación de la página en diferentes sub-elementos (títulos, párrafos, frases, notas de pie de página, etc).

La segmentación del texto

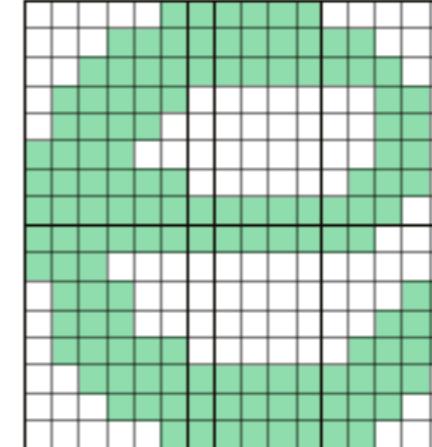
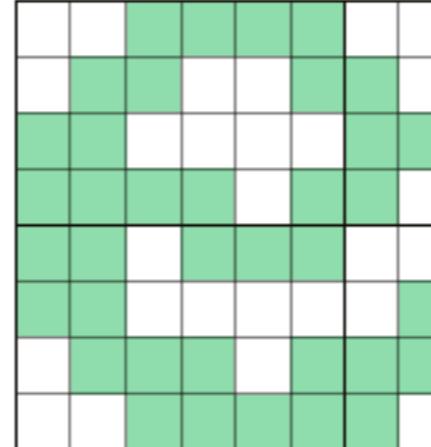
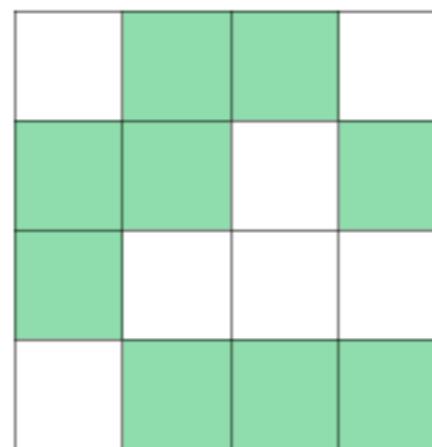
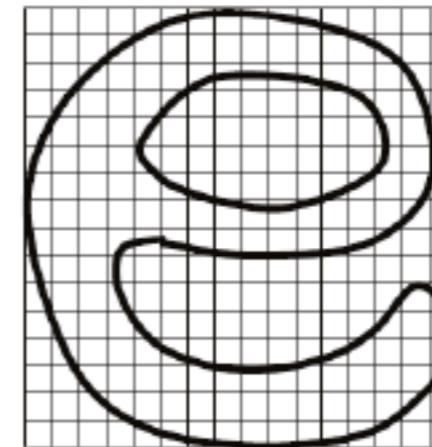
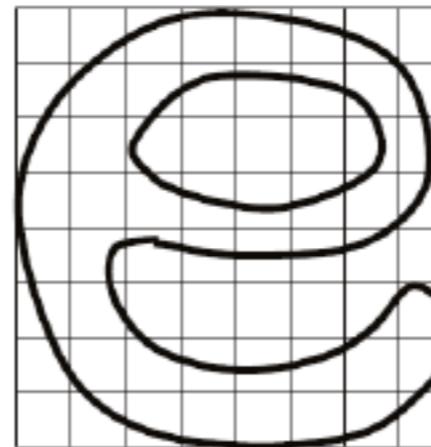
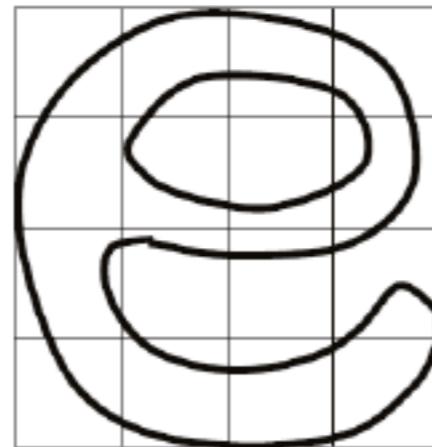


33

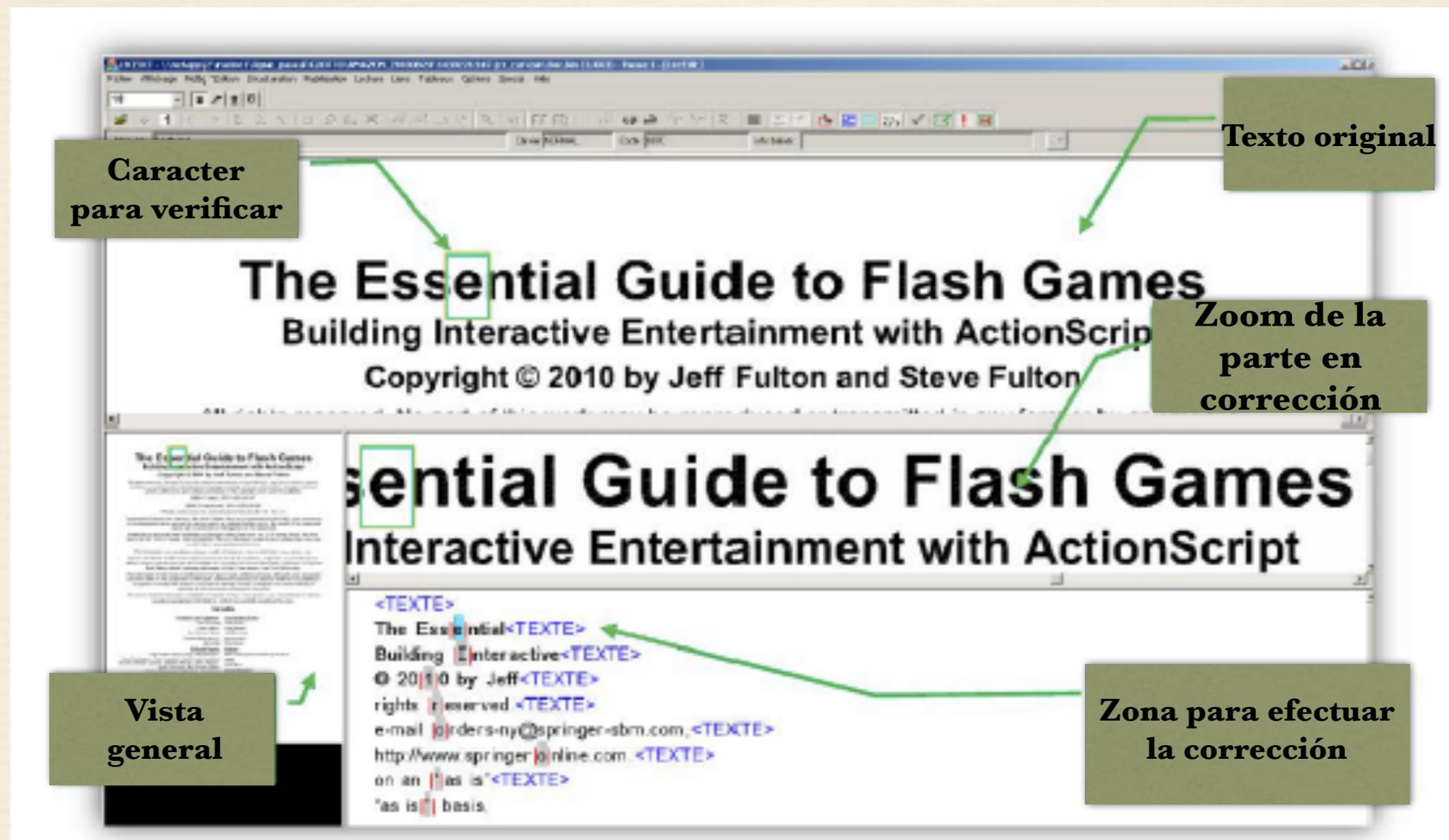


Reconocimiento de caracteres: importancia de la resolución

En función del número de píxeles captados en la digitalización



OCR: necesaria verificación y corrección...



Indices

N°	Titre	Vol.	Pag.
1	FRONTISPICE. — Pavillon du Louvre, place Napoléon III.	I	Iزادat. I
2	PARIS EN 1860. — Vue à vol d'oiseau, prise au-dessus du rond-point des Champs-Élysées.	I	— XII
3	PARIS EN 1860. — Vue à vol d'oiseau, prise au-dessus du quartier de Saint-Gervais.	I	— XII
 C. de Adenet, Dn. II, 30. de Calmont, Dn. II, 91. de Canaples, Dn. II, 34. de Candale, C. II, 3, D. II, 20. de Capres Dn. II, 88. Capital de Buch, II, 2, 3. de Caravas, C. II, 77. de Carency, Dn. I, 20, II, 26.			
 HISTOIRE			
	<i>Etudes sur les mœurs du V^e siècle</i> , par M. DURANT DES AULNOIS, membre titulaire.	1	
	<i>Recherches sur le village natal et la famille du chancelier Gerson</i> , par M. Henri JADART, membre titulaire	17	
 CH. GIVELET.			
Sur les vitraux récemment placés dans la chapelle absidale de la cathédrale de Reims, t. II, p. 62.			
 DELACREAU (Ch.). Bibliographie , III, 219. DELOIXON (Adrien). Du Positivisme dans une jeune école médicale , I, 241. — De quelques théories récentes sur la liberté			
VI, 264. DELTUR (Paul). Sensations d'un infidèle , II, 689.			
HICQUES (Gustave a'). Le Tiers-Etat en 1614 , I, 754. — L'Autriche et le corps germanique pendant la Révolution, III, 618. — Bibliographie , V, 203. — Etudes historiques sur la famille impériale : le Prince Eugène, VI, 339.			
JAMES (Jules). Horace et son temps , V, 268,			

Indice post OCR

(BnF) Gallica TOUT GALlica Rechercher... RECHERCHE AVANCÉE | SÉLECTIONS | FR

Accueil > Consultation
Les origines de l'alchimie / par M. Berthelot,...

Table des matières

Rechercher dans la table des matières

SOURCES	
CHAPITRE I. - Division du livre	7
▪ CHAPITRE II. - Les Origines mystiques	9
▪ CHAPITRE III. - Sources égyptiennes, chaldéennes, juives, gnostiques	21
▪ CHAPITRE VI. - Les témoignages historiques	67
▪ CHAPITRE V. - Les Papyrus de Leide	80
▪ CHAPITRE VI. - Les Manuscrits grecs des Bibliothèques	95
▫ LIVRE SECOND LES PERSONNES	127
CHAPITRE I. - Les	127

LIVRE PREMIER

LES SOURCES

—

CHAPITRE PREMIER.

INTRODUCTION DU LIVRE

 Une science doit être placée dans son cadre historique, si l'on veut en comprendre le véritable caractère et la portée philosophique : ce travail est surtout nécessaire pour une doctrine en partie réelle et en partie mystique, telle que l'alchimie. C'est pourquoi nous allons comparer d'abord les assertions et les textes des premiers alchimistes avec les croyances religieuses et mystiques qui régnent en Orient dans les premiers siècles de notre ère ; ce sera l'objet du second chapitre (sources mystiques) et du troisième chapitre (sources orientales, c'est-à-dire sources égyptiennes, babyloniennes, grecques et juives).

Page | 7

Documentos « nómades » (pueden leerse sin pasar por la interface de consultación de la institución)

- ❖ Amplias funciones: adaptación de tamaño de caracteres, de formato de página, etc.
- ❖ Interface dinámica, interactiva (subrayado, marcapáginas, etc).
- ❖ Lectura sobre tablet, teléfono, etc...
- ❖ Ejemplo tipo: « Epub » (formato tipo de los « eBook »)

0000016.jpg - IdenView (Zoom: 539 x 883)

File Edit Change Options View Help

88/152

Tous les animaux d'étable, sauf par les bœufs, se réveillent pour dans des circonstances aussi curieuses.

Le bœuf du chêne de Dracut est certainement plus amusante que celle de Pekin.

En effet, lors d'un terrible orage qui dévaste le village de Noyers, dans l'Aube, le matin d'un des vendredis, il aperçoit soudain qu'les animaux de son étable entrent toutes dans la grange où les réfugiés de l'orage espéraient qu'ils étaient par les flammes et la fumée, y pensant mal et à force d'aboiements et de mugissement, il fit sortir une première fois un cheval, une vache et quelques moutons. Il revint de nouveau dans la grange, et de la même façon il poussa dehors une vache et encore quelques moutons.

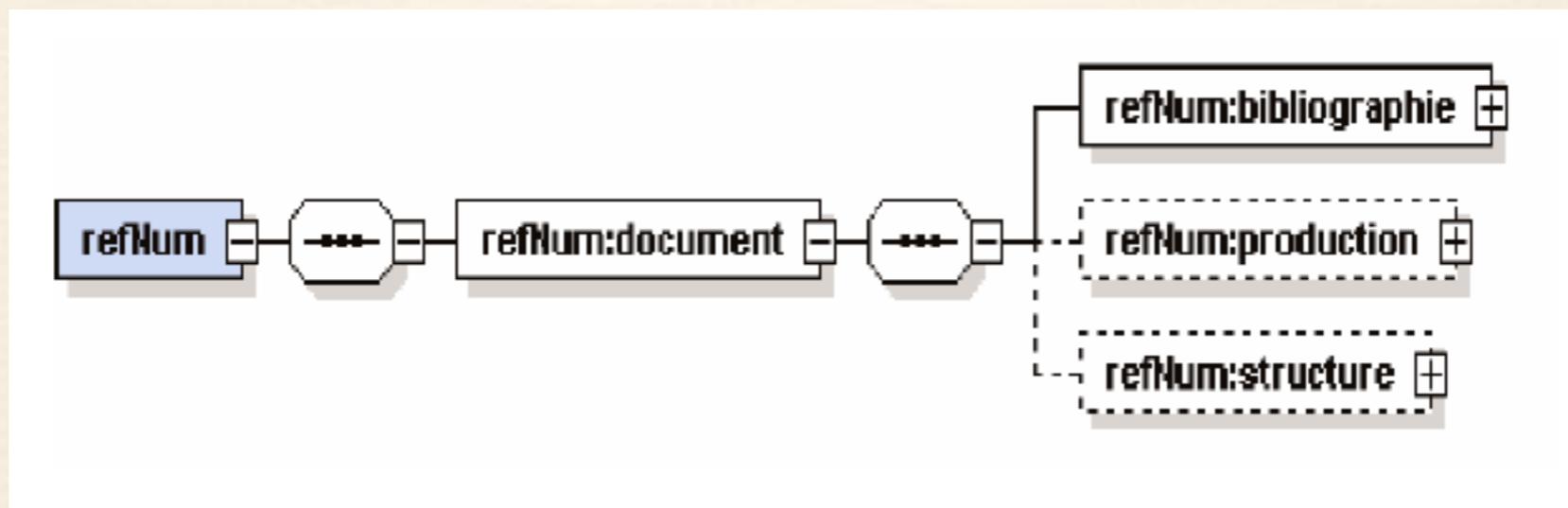
Il voulut rentrer une deuxième fois dans la ferme, mais, dès le seuil, il aperçut que les animaux qui étaient restés dans la grange étaient inactifs. Il revint en arrière et, après, se coucha dans un pré, pensant que la maison allait de bruler.

Devant de gosses pressés d'entendre, les hommes ne pouvaient

Il veillait contre une troisième fois dans la ferme ; mais, dès le seuil, il s'aperçut que les animaux qui étaient restés dans la grange étaient inactifs. Il revint en arrière et, également, se coucha dans un pré, pendant que la maison achevait de brûler.

- ❖ Un **documento digital** es una serie de archivos ligados entre si, identificado por un **referente único**, y por una serie de **metadatos**, es decir, un conjunto estructurado de informaciones que describen un objeto determinado (no necesariamente digital).

RefNum: carta de identidad de un documento digital



Metadatos de un documento digital:

❖ Metadatos descriptivos:

- * descripción bibliográfica profunda y detallada en un formato normalizado que permita el intercambio de información
- * ligar el documento al original o a diferentes versiones de un documento
- * dar acceso a la copia digital.

❖ Metadatos de estructura:

- * ligar los diferentes archivos de un mismo documento entre si
- * reconstruir la estructura del documento: conocer la relación entre los archivos que componen un documento (texto, imagen, etc); conocer la relación física entre los archivos (orden de presentación, archivo principal que da acceso al conjunto, etc.)

❖ Metadatos administrativos:

- * gestión de los derechos: derecho de autor, confidencialidad, derecho de impresión, de reproducción, de modificación, etc.
- * preservar las informaciones técnicas necesarias a la lectura de los archivos
- * garantizar la integridad y el seguimiento de las eventuales modificaciones.

Identificación de un documento digital

METS: (*Metadata Encoding and Transmission Standard*, Biblioteca del Congreso, EEUU): standard que permite codificar los metadatos descriptivos, administrativos y de estructura específicos de los objetos digitales.

- ❖ **Encabezado (metsHdr):** indicaciones sobre el archivo de metadatos en si, sobre todo las fechas de creación y de modificación.
- ❖ **Bibliografía (dmdSec):** datos generales que describen el tipo de documento original (iconográfico, monografía, periódico, etc), título, autor, fecha de edición, número de páginas. No reemplaza a los datos del catálogo.
- ❖ **Gestión (amdSec):** datos sobre las condiciones de digitalización, fecha, entrega, lista de tratamientos, historial, lista de agentes responsables de cada etapa, etc).
- ❖ **Inventario de los archivos (structMap):** expresa modos de navegación propias al documento digital, representadas por mapas de estructura (por ejemplo el índice de una monografía)
- ❖ **Metadatos específicos:** Dublin Core (metadatos de bibliografía), PREMIS (metadatos de procedencia).

Resultado post-digitalización

Nom	Type	Date de modification	Taille	Rati
T	Folder			
X	Folder			
6200003.tif	Image TIF	21/09/2011 12:37:42	4,543,412	61%
T6200003.xml	Document XML	21/09/2011 12:41:02	9,962	79%
K6200003.xml	Document XML	21/09/2011 12:41:04	41,758	92%

Nom	Type	Date de modification	Taille	Rati
T000001.tif	Image TIF	21/09/2011 12:33:54	18,703,982	53%
T000002.tif	Image TIF	21/09/2011 12:33:56	17,601,363	74%
T000003.tif	Image TIF	21/09/2011 12:33:58	17,508,004	78%
T000004.tif	Image TIF	21/09/2011 12:34:00	17,708,794	70%
T000005.tif	Image TIF	21/09/2011 12:34:02	18,106,338	49%
T000006.tif	Image TIF	21/09/2011 12:34:04	17,281,630	72%
T000007.tif	Image TIF	21/09/2011 12:34:04	17,520,374	71%
T000008.tif	Image TIF	21/09/2011 12:34:06	17,896,973	60%
T000009.tif	Image TIF	21/09/2011 12:34:08	17,520,374	74%
T000010.tif	Image TIF	21/09/2011 12:34:10	17,536,661	71%
T000011.tif	Image TIF	21/09/2011 12:34:12	17,520,374	76%
T000012.tif	Image TIF			
T000013.tif	Image TIF			
T000014.tif	Image TIF			

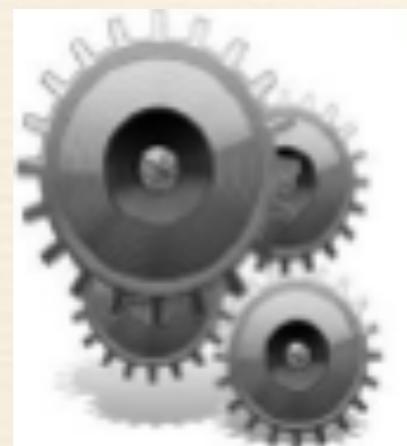
Nom	Type	Date de modification	Taille	Rati
x000001.xml	Document XML	21/09/2011 12:41:02	1,531	59%
x000002.xml	Document XML	21/09/2011 12:41:02	946	49%
x000003.xml	Document XML	21/09/2011 12:41:02	946	49%
x000004.xml	Document XML	21/09/2011 12:41:02	946	49%
x000005.xml	Document XML	21/09/2011 12:41:02	13,367	64%
x000006.xml	Document XML	21/09/2011 12:41:02	946	49%
x000007.xml	Document XML	21/09/2011 12:41:02	13,180	63%
x000008.xml	Document XML	21/09/2011 12:41:02	8,066	81%
x000009.xml	Document XML	21/09/2011 12:41:02	44,240	87%
x000010.xml	Document XML	21/09/2011 12:41:02	47,783	87%
x000011.xml	Document XML	21/09/2011 12:41:02	33,968	86%
x000012.xml	Document XML	21/09/2011 12:41:02	47,214	87%
x000013.xml	Document XML	21/09/2011 12:41:02	42,172	87%
x000014.xml	Document XML	21/09/2011 12:41:02	52,711	87%



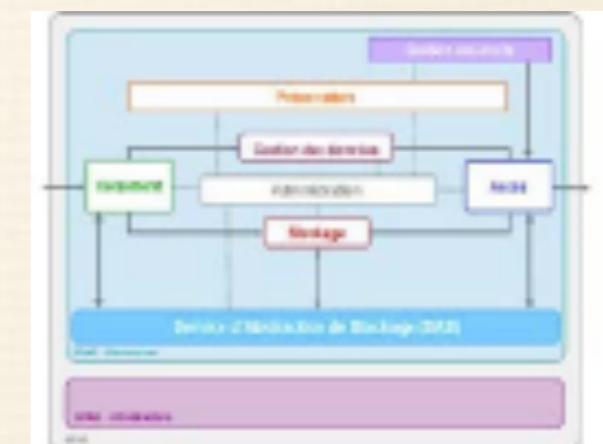
Control de calidad



Consulta pública

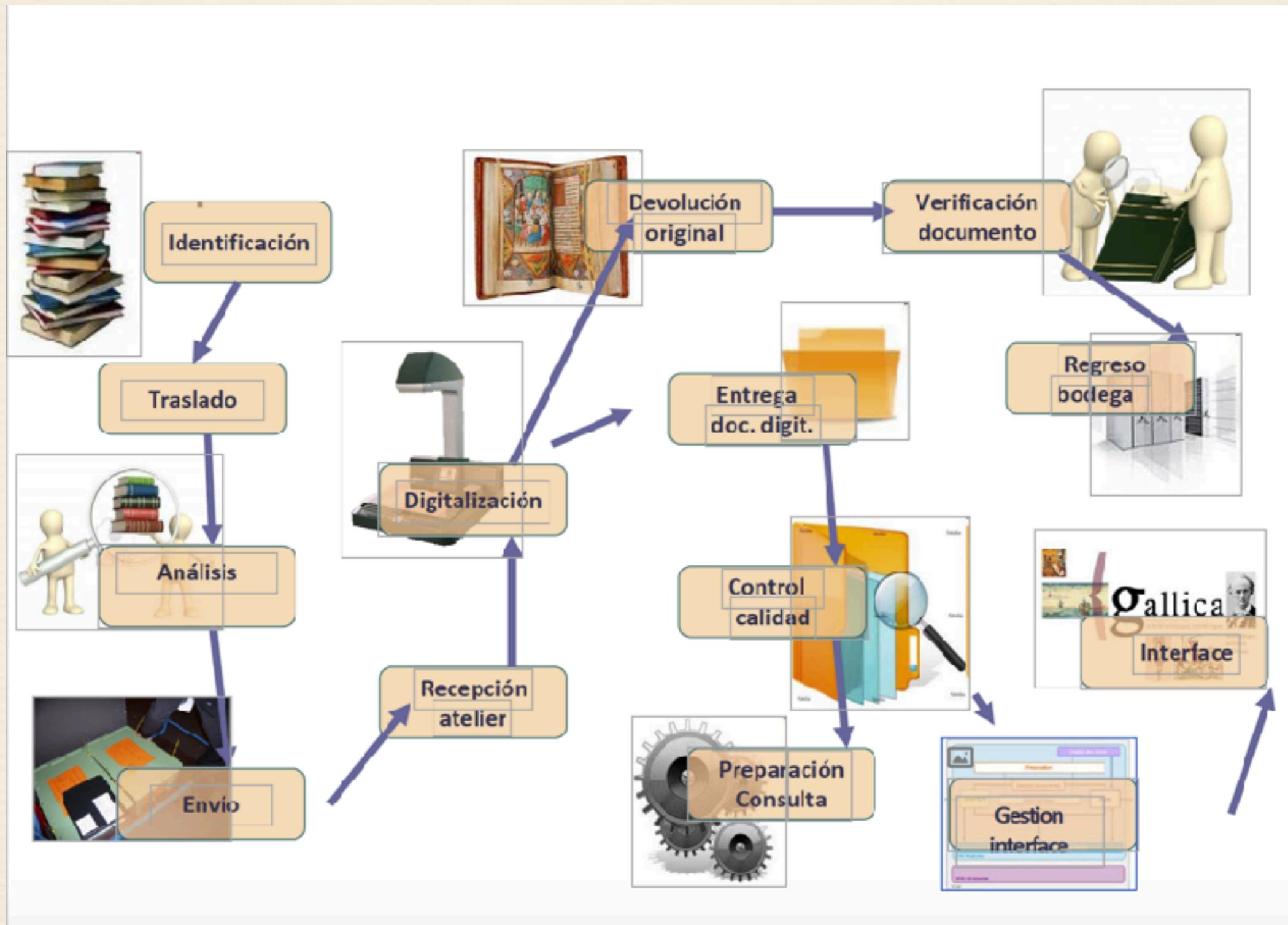


Preparación de la consulta



Gestión de la interface

En síntesis...



* * * * *

La gestión de los documentos digitalizados

El « depósito digital »



- ❖ Un documento digital se conserva en un espacio físico...
- ❖ La infraestructura de stock cuesta caro y ocupa espacio (hay que protegerla y duplicarla)
- ❖ A partir de una cierta cantidad, se necesita una gestión de archivos y de datos por un software adaptado.

Y supone un « catálogo digital »

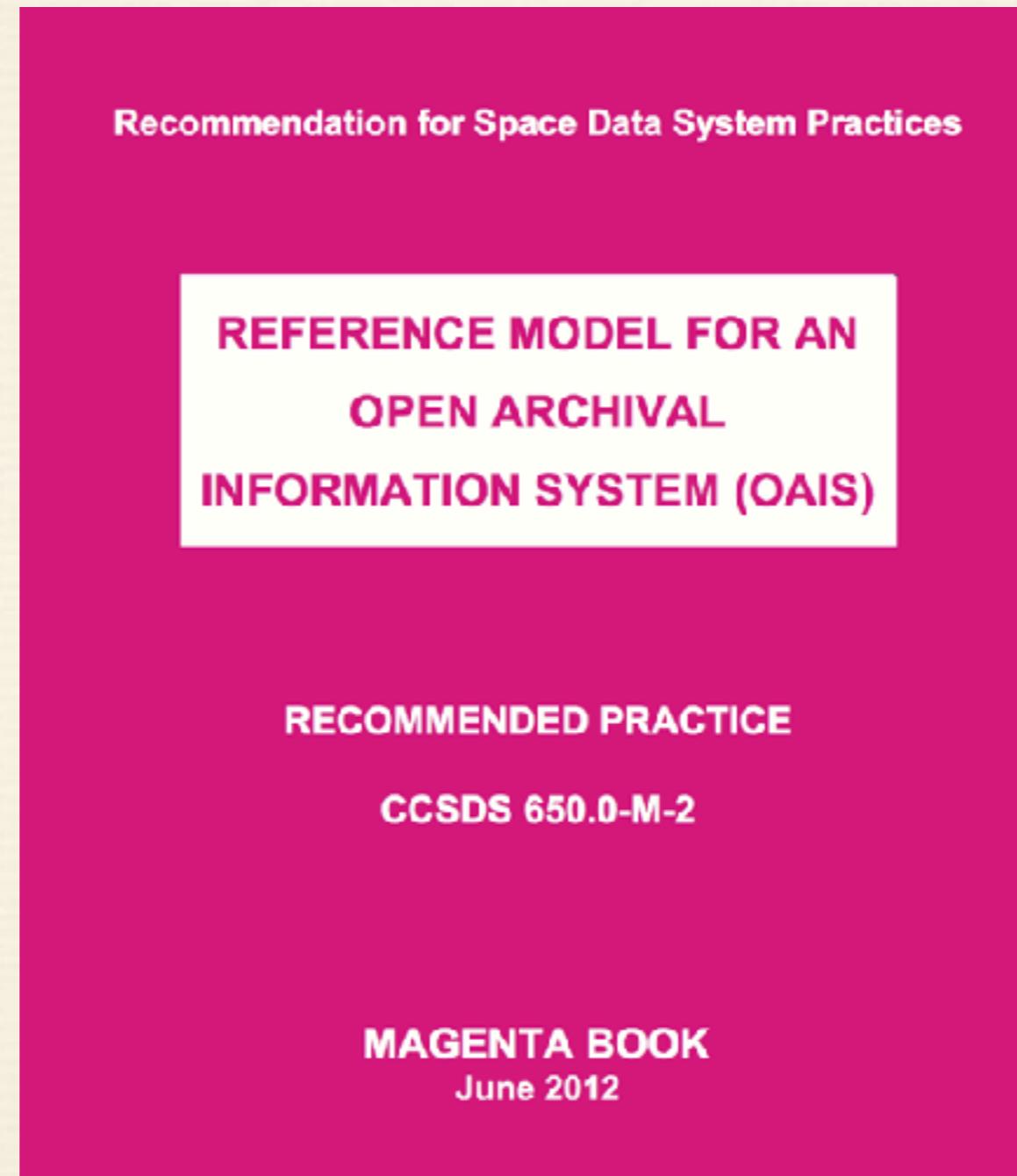
- ❖ La descripción no se limita a la presentación tradicional de los documentos (u objetos) digitalizados: metadatos técnicos y administrativos
- ❖ Felizmente, las informaciones técnicas pueden ser automatizadas (por extracción)
- ❖ Hay que pensar el documento de varios puntos de vista: como fuente de información, pero también como objeto...

```
- <metadata>
- <oai_dc:dc>
  <dc:identifier>http://gallica.bnf.fr/ark:/12148/bpt6k10
- <dc:title>
  Trente jours de voyage en zigzag : de Paris à Venise
</dc:title>
<dc:creator>Abadie, Louis (18..-19.. ; avocat)</dc:cre
<dc:publisher>impr. de A. Lebon (Paris)</dc:publish
<dc:date>1862</dc:date>
<dc:subject xml:lang="fre">France -- Descriptions et
<dc:subject xml:lang="fre">Italie -- Descriptions et
<dc:language>fre</dc:language>
<dc:type xml:lang="eng">text</dc:type>
<dc:type xml:lang="fre">monographie imprimée</dc:
<dc:type xml:lang="eng">printed monograph</dc:ty
<dc:rights xml:lang="fre">domaine public</dc:right
<dc:rights xml:lang="eng">public domain</dc:right
<dc:format>application/pdf</dc:format>
</oai_dc:dc>
</metadata>
</record>
- <record>
- <header>
  <identifier>oaibnffr:gallica/ark:/12148/bpt6k39258m<
  <datestamp>2007-01-19</datestamp>
  <setSpec>gallica</setSpec>
  <setSpec>gallica.9</setSpec>
  <setSpec>gallica.9.944</setSpec>
  <setSpec>gallica:monographies</setSpec>
<header>
- <metadata>
- <oai_dc:dc>
  <dc:identifier>http://gallica.bnf.fr/ark:/12148/bpt6k39
- <dc:title>
  Bâtiments de la Bibliothèque nationale : rapport / M
</dc:title>
- <dc:contributor>
  Barthélémy-Saint-Hilaire, Jules (1805-1895). Rédact
</dc:contributor>
<dc:creator>France. Ministère de l'instruction publiqu
<dc:publisher>Impr. nationale (Paris)</dc:publisher
<dc:date>1879</dc:date>
<dc:subject xml:lang="fre">Bibliothèque nationale (I
<dc:language>fre</dc:language>
```

Preservar es conocer...

- ❖ El punto de partida de todo plan de preservación es el conocimiento de las colecciones:
 - * De un punto de vista documental: prioridad según el valor patrimonial, científico, documental...
 - * De un punto de vista material: prioridades según el estado...
- ❖ Para lanzar el proceso de preservación, hay que considerar:
 - * La cantidad de documentos
 - * Los formatos
 - * Los metadatos disponibles
- ❖ **Anticipar todo lo que sea posible!**

OAIS (ISO 14721)



<http://public.ccsds.org/publications/archive/650x0m2.pdf>

Check list básica:

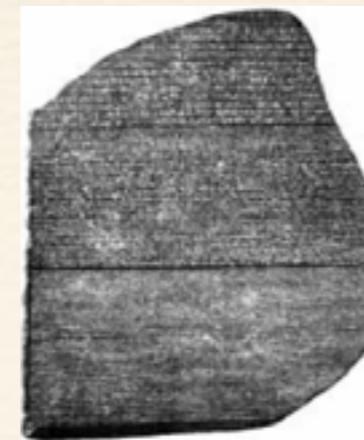
- ❖ **Definir un modelo de información:** Cómo describir mis objetos para asegurar su preservación?
- ❖ **Definir un modelo funcional:** Cómo organizar el trabajo para asegurar esta actividad de preservación?
- ❖ **Definir las estrategias de preservación:** cuales son los métodos para evitar la obsolescencia tecnológica?

Modelo de información

Mundo real



+



+



Mundo digital



+

Información
de representación:
Lengua, formato

+

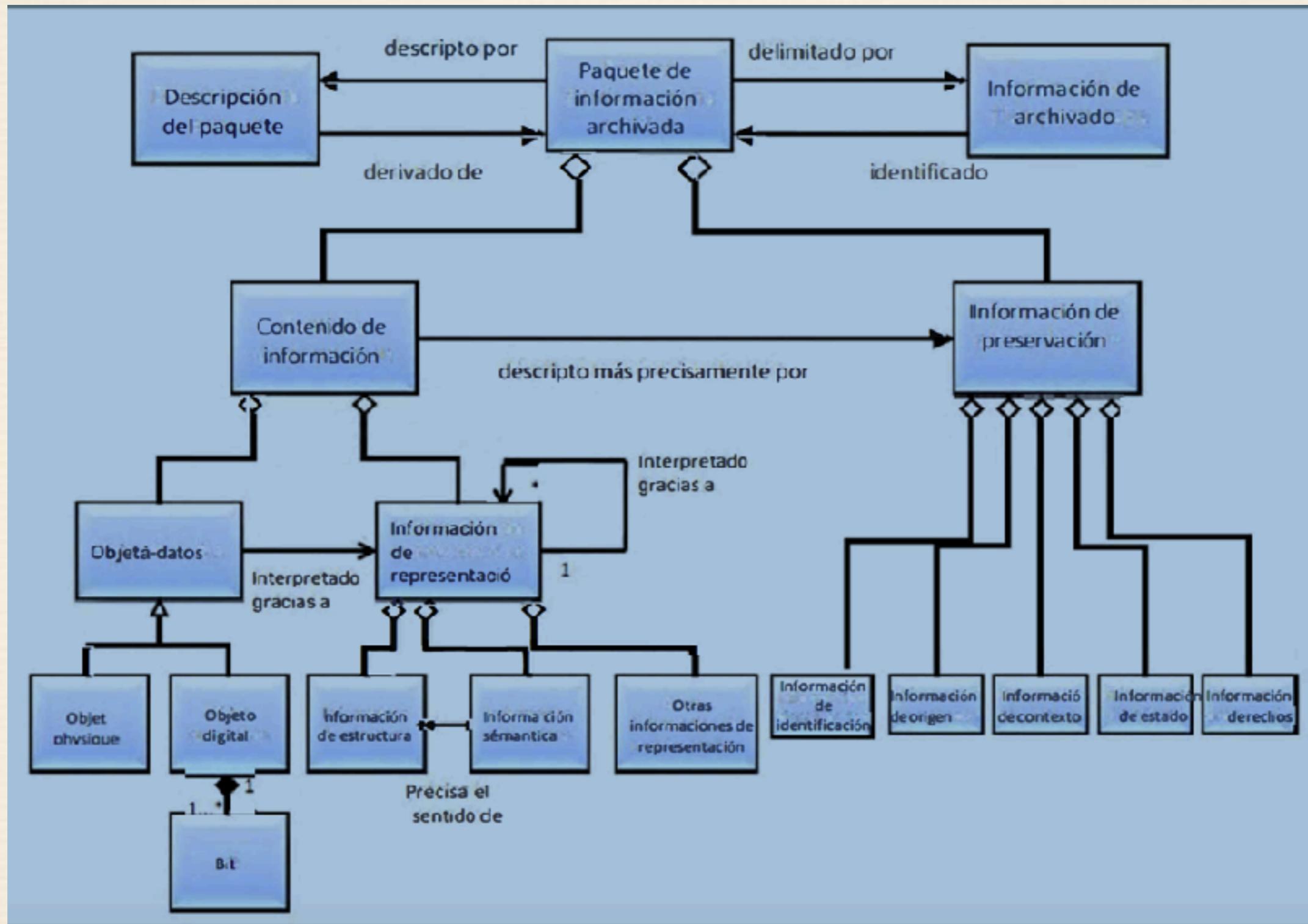
Información de preservación:
Origen, contexto,
estado de conservación...

Paquete de datos

Objeto-datos

Metadatos

Modelo de información estándar



Documentar: contexto y origen

- ❖ Dejar traza de todas las decisiones tomadas para cada tipo de documento
 - * Especificaciones de la digitalización
 - * Evoluciones, modificaciones.
 - * Problemas identificados
 - * Procesos de gestión
 - * Fecha mínima y máxima de los mercados
- ❖ Registrar las elecciones de la estructura del documento:
 - * Digitalización por página o capítulo?
 - * Ocerización de toda la colección o de una parte?
 - * Adoptar una política común para nombrar los documentos.

Documentar es representar

- ❖ Incluir (o hacer incluir) metadatos en el archivo imagen:
 - * Fecha de creación
 - * Software de origen
 - * Calibrage del scanner
 - * Metadatos JFIF (para los JPEG) o TIFF
- ❖ Realizar por lo menos un control manual de los archivos para validación.

Instrumentos de trabajo estandarizados

- ❖ Identificación, validación, caracterización:

<http://description.fcla.edu/>

- ❖ Para el trabajo profesional en serie:

- * El veterano de la identificación: DROID

<http://digitalpreservation.github.com/droid/>

- * Un nuevo instrumento: FIDO

<http://www.openplanetsfoundation.org/software/fido>

- * Identificación, validación, caracterización: JHOVE y JHOVE2

<http://sourceforge.net/projects/jhove/>

<https://bitbucket.org/jhove2/main/wiki/Home>

Para documentos audiovisuales:

<http://mediainfo.sourceforge.net/fr>

Como consignar los metadatos

- ❖ Como con los catálogos papel,
 1. Hay que consignar informaciones esperadas
 2. Hay que elegir un formato para expresarlas
 - * Es importante armonizar desde el primer momento para evitar los problemas mas tarde y facilitar el trabajo a largo plazo y las colaboraciones
- ❖ Existen diccionarios de datos para expresar las informaciones según el área:
 - * PREMIS
 - * MIX, etc
- ❖ Existen formatos para « agrupar » metadatos
 - * METS, XFDU
 - MPEG-21, DIDL, etc.

Preservar implica responsabilidades...

- ❖ Como con la conservación física, hay que definir el rol de cada persona y cada servicio (**trabajo de equipo obligatorio!!!!**):
*Responsables de conservación, encargados de bibliotecas digitales, responsable de sistemas de información, etc.
- ❖ La norma OAIS ofrece guías prácticas sobre la organización de responsabilidades, que deben ser adaptadas a cada contexto particular.

Qué plan de preservación?

- ❖ Idealmente, el mismo plan para colecciones materiales y digitales.
- ❖ Algunas iniciativas interesantes en la definición de niveles de servicio:
 - * Artículo de la National Library of Australia sobre la preservación:
<http://www.dlib.org/dlib/january13/webb/01webb.html>
- ❖ Instrumentos para prevenir la obsolescencia tecnológica:
 - * la Biblioteca del Congreso de EEUU, piloto del National Digital Information Infrastructure and Preservation Program:
<http://www.digitalpreservation.gov/>

Otros instrumentos en open access

❖ Instrumentos disponibles según el tipo de colección:

* Archivos de Internet:

<http://netpreserve.org/about/index.php>

* Audiovisual:

<http://prestocentre.eu/>

❖ Instrumentos libres de los proyectos europeos via Open Planets Foundation:

<http://www.nationalarchives.gov.uk/aboutapps/pronom/>

❖ Registro de formatos e instrumentos de identificación PRONOM y DROID:

<http://www.nationalarchives.gov.uk/aboutapps/pronom/>

* UDFR:

<http://www.udfr.org/>

* JOHVE:

<http://hul.harvard.edu/jhove/>

* JOHVE2:

<https://bitbucket.org/jhove2/main/wiki/Home>

**En síntesis, los objetivos de la planificación
de una estrategia digital son...**

1. Constituir una colección digital de referencia

- * Digitalización de colecciones patrimoniales existentes en formato tradicional
- * Colectar las obras digitales nativas
- * Archivage web

2. Identificar y conservar

- ❖ Garantizar la perennidad de los objetos digitales
- ❖ Hacer visibles los fondos patrimoniales

3. Innovar

- ❖ Creación de interfaces eficaces
- ❖ Considerar los diferentes tipos de públicos
- ❖ Tomar en cuenta las prácticas actuales de lectura y de interacción cultural

4. Trabajar en red

- ❖ Colaboración nacional e internacional
- ❖ Instituciones patrimoniales, universidades, centros de investigación, estructuras nacionales, etc
- ❖ Consorcios internacionales (p.e. Biblioteca virtual Miguel de Cervantes: <http://www.cervantesvirtual.com>)

5.Simplificar el uso de instrumentos y procedimientos

- ❖ Mutualización de instrumentos digitales
- ❖ Simplificación de la colaboración con partenaires
- ❖ Armonización de procedimientos técnicos.
- ❖ Si prestación de servicios externos, definición de criterios y exigencias comunes

6. Proponer una innovación durable

- ❖ Acompañar el desarrollo tecnológico
- ❖ Políticas de colaboración con los sectores de educación terciaria y universitaria (y hasta secundaria)
- ❖ Inscripción en redes y consortium de reflexión digital (p.e., Centro de competencias Impact:
<https://www.digitisation.eu>)

Muchas gracias por su atención.

Preguntas?