PROCESO DE DIGITALIZACIÓN EN LA BIBLIOTECA NACIONAL DE ESPAÑA

Actualizado 11/03/2024

Contenido

PROCESO DE DIGITALIZACIÓN EN LA BIBLIOTECA NACIONAL DE ESPAÑA 1

1. INTRODUCCIÓN 3

2. PASOS PREVIOS A LA DIGITALIZACIÓN 4

2.1 Criterios de selección 4
2.2 Extracción del catálogo automatizado 6
2.3 Préstamo de títulos 6
2.4 Revisión y selección de ejemplares 6
2.5 Catalogación y Clasificación 8
2.6 Planificación y lotes de trabajo 8
2.7 Préstamo de obras al digitalizador 9

3. EL PROCESO DE DIGITALIZACIÓN 9

3.1 Los laboratorios de digitalización 12
3.1.1 El material librario 12
3.1.2 El material audiovisual 14

4. FICHEROS MÁSTER Y DERIVADOS 16

4.1 Archivos de preservación 16
4.1.1. Formatos de los archivos de preservación 17
4.2 Archivos de difusión 17
4.2.1. Formato de los archivos de difusión 17
4.2.1.1. PDF de difusión 18

5. METADATOS 18

5.1 Metadatos descriptivos 18
5.1.1 La exportación de metadatos. OAI-PMH 19
5.2 Metadatos de preservación (METS-PREMIS) 19

6. CONTROL DE CALIDAD 21

6.1 Control de calidad previo a la carga en el portal de acceso 21
6.2 Control de calidad posterior a la carga en el portal de acceso 23

7. EL PORTAL DE ACCESO 23

7.1 Sistema de Gestión de Objetos Digitales. 23
7.2 El motor de búsqueda. 23
7.1 La unificación de los portales de acceso. 25

8. EL SISTEMA DE PRESERVACIÓN DIGITAL 25

1. INTRODUCCIÓN

En este documento se detalla el flujo de trabajo para llevar a cabo el proceso de digitalización que permite poner una obra a disposición del usuario en los portales de acceso de la Biblioteca Nacional de España, así como los criterios de selección, técnicos y de calidad que guían todo su desarrollo.

Para conocer mejor los antecedentes de este proceso, tenemos que remontarnos a finales de los años 90, momento en que la digitalización fue paulatinamente sustituyendo a los procesos de microfilmación en la Biblioteca Nacional de España (BNE), convirtiéndose en el único medio de reproducción de las colecciones analógicas. Los primeros impulsos comienzan con la digitalización de la prensa moderna para facilitar la consulta y prevenir el deterioro que el uso continuado provocaba en los originales. Más adelante, se le suma la prensa histórica, dando origen al primer portal digital de la Biblioteca, la Hemeroteca Digital.

Pero es en 2008 cuando la Biblioteca desarrolla una política de digitalización sistemática de sus fondos que da origen a la Biblioteca Digital Hispánica (BDH). Este proyecto recibió un gran estímulo gracias a los convenios de colaboración con entidades privadas (Telefónica 2008/2016) y públicas (Red.es 2015/2018), consolidándose en 2022 con la financiación de los fondos del Plan de Recuperación, Transformación y Resiliencia de la Unión Europea (PRTR).

Este modelo económico de colaboración entre el sector público y el privado se puede considerar en todos los sentidos plenamente beneficioso para la institución, que ha recibido una significativa financiación por parte de entidades externas, manteniendo, sin embargo, toda la autonomía necesaria para organizar el proyecto y para establecer los criterios de calidad que son exigibles a esta institución.

Sin embargo, este modelo no es la única fuente de crecimiento de las colecciones de fondos digitalizados. La Biblioteca cuenta con sus propios laboratorios de fotografía y digitalización y dispone, además, de un equipo de personal especializado con amplios conocimientos en la digitalización del patrimonio documental de la BNE.

La digitalización de colecciones se debe considerar a estas alturas, más que un proyecto, un proceso que afecta a casi todos los Departamentos y Áreas de la BNE. A día de hoy las tareas implicadas desde que se decide qué se va a digitalizar hasta que finalmente está accesible en el portal se han incorporado de modo casi pleno en todas las rutinas de los diferentes Servicios de la Biblioteca.

Aunque el proyecto nace sobre la base de algunas colecciones ya existentes, lo cierto es que el camino recorrido desde su aparición ha sido casi por completo una novedad. Fue necesario fijar los procedimientos de trabajo, establecer los criterios de selección, los controles de calidad, diseñar e implantar un nuevo interfaz de búsqueda, etc. Además de las dificultades propias de esta tarea, hay que tener en cuenta la evolución

tecnológica que no se detiene y que exige permanente revisión y actualización de criterios y procesos.

En definitiva, este proceso hay que valorarlo no sólo por lo que mejora el servicio que la BNE ofrece a los usuarios, sino también por lo que ha supuesto de adaptación al cambio para la institución y para las personas que la componen.

2. PASOS PREVIOS A LA DIGITALIZACIÓN

La creación de colecciones digitales se enmarca dentro de un proceso transversal que, a grandes rasgos, puede resumirse en las siguientes etapas:

- Selección de la colección
- Extracción de registros del catálogo
- Revisión y selección de ejemplares
- Planificación de los lotes de trabajo
- Digitalización
- Control de calidad
- Carga en el portal de acceso y enlace desde catálogo

2.1 Criterios de selección

Para el proceso de selección ha sido necesario ir creando procedimientos y estableciendo criterios que, en ocasiones, han evolucionado en función de las necesidades o de las posibilidades que iban apareciendo.

La selección de colecciones se planteó desde el comienzo teniendo presente un criterio temático que permitierá organizar los documentos de acuerdo a características comunes. Los diferentes tipos de documentos que se han ido incorporando han hecho que el propio concepto de colección haya evolucionado y en cierto modo se haya ampliado en función del aumento de documentos digitalizados y de la incorporación de nuevos materiales.

El proyecto de digitalización de la BNE se centra principalmente en los fondos de dominio público, es decir, aquellos que están libres de derechos de autor. Esto supone referirse a las obras conservadas en la BNE cuyo autor haya fallecido hace 70 u 80 años (dependiendo de su fecha de nacimiento) y, por supuesto, siempre que no se trate de ediciones nuevas también protegidas en la ley de propiedad intelectual actualmente vigente en España.

Al margen de su pertenencia al dominio público hay otra serie de principios generales que guían la selección de documentos:

- La relevancia del contenido. Muchas de las colecciones se han seleccionado mediante un trabajo exhaustivo por parte del Servicio de Información Bibliográfica, con el fin de crear relaciones de documentos especialmente pertinentes para un determinado tema: ocio, viajes, ciencia, independencia

hispanoamericana. En otras ocasiones esta unidad proviene de tratar de ofrecer la producción completa de un autor.

- El interés del material. Desde las Salas de lectura de la BNE se han seleccionado colecciones de documentos cuyas características materiales son interesantes en sí mismas: manuscritos, incunables, dibujos de arquitectura, grabado alemán…

- Otro criterio general que se combina a la hora de seleccionar los documentos es el interés que puede tener para los usuarios. La aplicación de este criterio es posible gracias al conocimiento de la colección y de los usuarios que tienen los bibliotecarios de la BNE.

- El valor patrimonial es otro criterio que inspira la selección de documentos y que persigue que las obras aceptadas tradicionalmente como obras maestras figuren en la BDH.

- Aspectos relativos a la preservación del documento físico que se digitaliza. Dado que un documento digitalizado sufrirá un menor número de consultas, la digitalización se convierte así en una técnica privilegiada de preservación.

- La obsolescencia de los soportes es otro de los criterios de selección a tener en cuenta en estos días. Los fondos audovisuales más antiguos muestran signos de degradación a pesar de haber sido conservados en condiciones ambientales controladas; y cada vez es más complicado encontrar y mantener los equipos necesarios para su reproducción. La digitalización es una manera de conservar el contenido de estos soportes y, además, preserva los originales evitando reproducirlos continuamente lo que podría acelerar su degradación.

- El seguimiento de unos criterios u otros puede suponer en ocasiones tomar decisiones diferentes. Éste es el caso, por ejemplo, la elección de las ediciones que deben ser digitalizadas. En colecciones cuyo fin principal es tratar una materia se optó por seleccionar una única edición. Sin embargo, cuando se persigue recoger la producción integra de un autor, se digitalizan las sucesivas ediciones de una obra al entender que en este caso las variaciones pueden ser de interés para el especialista.

- En el caso del proyecto de digitalización de la prensa histórica, los criterios generales son los mismos, aunque conviene añadir algunas precisiones. El objetivo es cubrir la evolución histórica de la prensa española, desde sus inicios hasta principios del siglo XX, respetando siempre las limitaciones que marca nuestra legislación en temas de propiedad intelectual. El criterio que ha guiado la composición de esta colección ha sido seleccionar periódicos y revistas representativos de su época, que reflejaran la riqueza temática de la edición hemerográfica hispana y de los que se conservaran colecciones completas. Quienes visiten la hemeroteca encontrarán prensa política, satírica, humorística, científica, religiosa, ilustrada, amena, deportiva, artística, literaria, etc.

2.2 Extracción del catálogo automatizado

Los títulos seleccionados se marcan en el catálogo en un campo local del formato MARC, con un código asignado a tal fin, para proceder después a la extracción de los mismos.

Cada vez que un registro se considere idóneo para la colección, deberá comprobarse antes si la obra ha sido ya digitalizada por la BNE.

Finalizada la fase de selección de títulos, se realiza la extracción de un listado con las signaturas marcadas de acuerdo a los criterios transmitidos. Este listado se carga en una base de datos de trabajo interno para comenzar el flujo de tareas que implican la digitalización de una obra.

2.3 Préstamo de títulos

El personal de salas se encarga de realizar el préstamo de los títulos seleccionados, así como de supervisar la colocación de los fondos, la devolución y el traslado de los fondos a los diferentes implicados en el flujo de trabajo (servicios responsables del proceso técnico y/o digitalizadores).

El préstamo de fondos se lleva a cabo en dos procesos dentro del proyecto de digitalización:

- *Selección de fondos a digitalizar*: En este momento tiene lugar la correspondiente entrega de fondos al Departamento de Proceso Técnico (fondo Moderno) o al Departamento especializado en determinado tipo de material (Patrimonio Bibliográfico, Bellas Artes y Cartografía, Música y Audiovisuales), con el fin de que se revisen los documentos originales y se seleccione el ejemplar a digitalizar. Cada sala establece una media de fondos a sacar diariamente, en función del volumen de trabajo asumible acorde con los recursos humanos y materiales de que disponga. Condición *sine qua non* para este movimiento masivo de fondos, es la colocación del código de barras con el IDITEM del ejemplar.

- *Recopilación de fondos que componen el/los lote(s) de digitalización diaria*.

2.4 Revisión y selección de ejemplares

En esta fase, el Departamento de Proceso Técnico (en el caso del fondo moderno) o el Departamento especializado (para los fondos de carácter singular, incluido el antiguo) proceden a la unificación del registro bibliográfico. Esta tarea consiste en la revisión de los títulos de interés, identificación de bibliográficos repetidos y gestión de altas y bajas de registros y/o signaturas que esto origine. De todo esto se deduce fácilmente que la tarea de digitalización está suponiendo una labor de depuración del catalógo bibliográfico.

A continuación, se revisa y selecciona el ejemplar más adecuado para ser digitalizado. En el caso del fondo moderno esta revisión manual la realiza directamente un técnico del Departamento de Preservación y Conservación. En el resto de casos, esta

participación se hace patente siempre que el personal de los departamentos así lo requiera.

2.4.1. Pautas para determinar si un ejemplar es apto para digitalizar:

A continuación, se reflejan los criterios de exclusión de la digitalización masiva y/o aspectos que el operario de digitalización deberá tener en cuenta para la manipulación.

1. **Libros quebradizos**. Si las hojas presentan pérdidas, se parten fácilmente o se desprenden del libro no podrá ser escaneado automáticamente. Se deberá localizar, en caso de que exista, un ejemplar en mejores condiciones u optar, en su defecto, por digitalizarlo en el Laboratorio de Fotografía y Digitalización.

2. **Libros quebradizos en la zona del lomo**: Encuadernaciones a gutapercha o a la americana, rústicas y encuadernaciones de lomo rígido (especialmente entre los siglos XV-XVIII). El exceso de refuerzo en el lomo y de encolado puede provocar la rotura de la encuadernación. Se rechazarán todos aquellos ejemplares que no puedan ser abiertos con comodidad en un ángulo de 135º.

3. **Algunas encuadernaciones presentan encolados ácidos en el lomo.** Se examinará la costura del ejemplar y se rechazarán todos aquellos en los que aparezcan hojas sueltas por rotura en el pliegue. Este problema suele ser más frecuente en las primeras y últimas hojas.

4. **Encuadernaciones rotas** en la zona del cajo –bisagra– de la tapa delantera o trasera.

5. **Ejemplares con ataques graves de microorganismos o de insectos**, en los que aparezcan fragmentos de papel, hojas sueltas o débiles.

6. Ejemplares con graves problemas físicos, como **hojas desgarradas, sueltas o con pérdidas.**

7. Ejemplares con mapas o grabados **desplegables**.

8. Ejemplares gravemente **deformados** por acción del agua o de la mala colocación. En impresos del XVII y XVIII pueden presentar deformaciones importantes en la caja de texto.

9. **Papel rígido**. A pesar de que no presenten problemas de acidez, no se podrán escanear los libros en los que el papel sea grueso y no se flexione con facilidad.

10. Cuando existan varios ejemplares de una misma obra y uno de ellos sea una **edición facticia**, se elegirá otra opción.

11. Cuando se trate de **folletos** –signatura VC– y existan varios ejemplares se elegirá siempre el que esté encuadernado.

12. Cuando existan varios ejemplares de una misma obra y uno de ellos presente los **cuadernillos cerrados *in-tonso*** se escogerá otro ejemplar. Si no es posible, se anotará la incidencia y se enviará al Departamento de Preservación y Conservación.

13. Cuando se trate de un **ejemplar único** debería digitalizarse con especial cuidado. Requerida marcado o identificación especial.

14. En el caso de que no existan otras opciones y deban elegirse ejemplares con contenedores de conservación hay que prestar especial atención, ya que pueden estar protegiendo ejemplares con encuadernaciones valiosas, ejemplares muy deteriorados o ejemplares únicos.

2.5 Catalogación y Clasificación

Una vez se ha seleccionado el ejemplar más adecuado para digitalizar, el personal encargado de la catalogación y clasificación de los registros:

- Corrige y amplía la descripción bibliográfica

- Y asigna una colección temática a la obra según el esquema temático abreviado que se maneja en el portal de acceso y que deriva de la CDU (Clasificación Decimal Universal).

2.6 Planificación y lotes de trabajo

Una vez seleccionados los ejemplares y revisada la catalogación se organizan los lotes de digitalización, en función del ritmo diario que puede asumirse (teniendo en cuenta turnos de digitalización, nº de equipos y tipo de documento).

En concreto la preparación de los lotes de digitalización implica:

- Elaboración de lotes diarios de trabajo para su digitalización, organizados por colección, teniendo en cuenta la creación de colecciones distintas para materiales que se encuentran en la sede de Alcalá y la de Recoletos. Para la preparación de los lotes se manejan dos parámetros básicos:

  - ✓ El límite máximo de páginas o minutos a digitalizar/día según equipo y el tipo de material a digitalizar

  - ✓ El número máximo de préstamos de ejemplares diarios que pueden asumir las salas prestatarias.

- Marcado de los registros en el campo 899 para identificar el ejemplar digitalizado y el proyecto en el que se realiza su digitalización. Se comprueba, además, la presencia/ausencia de 856. De esta forma se minimiza el riesgo de

introducir en la cadena de digitalización obras que ya cuentan con una reproducción de calidad.

- Extracción de los ficheros de metadatos descriptivos en formato MARC 21. Posteriormente, se tratan conforme a las especificaciones requeridas para su carga en el portal de acceso de la BNE.

Una vez elaborados los lotes, a todos los implicados en el proceso se les distribuye un listado con los datos mínimos que permiten identificar los títulos y ejemplares a digitalizar cada día. Dichos implicados son: salas prestatarias y digitalizadores. Los lotes se podrán ajustar según el ritmo real de digitalización y las incidencias que puedan producirse en el día a día.

2.7 Préstamo de obras al digitalizador

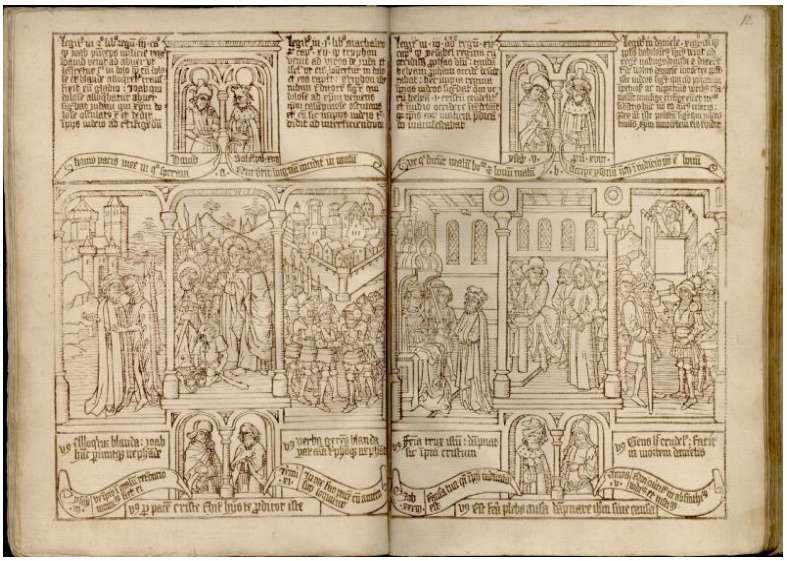
A partir del listado de lotes elaborado se realizan las tareas necesarias que permiten el préstamo al digitalizador a nivel lógico y a nivel físico comenzando así el proceso de escaneado propiamente dicho.

3. EL PROCESO DE DIGITALIZACIÓN

El proceso de digitalización de **material librario** se realiza siguiendo los siguientes criterios:

1. Se generan imágenes máster con las siguientes características:

   - En formato TIFF
   - Sin compresión
   - En color
   - Con profundidad de color de 24 bits
   - Con resolución óptica de 400 píxeles por pulgada
   - En modo RGB
   - Sin marcas de agua

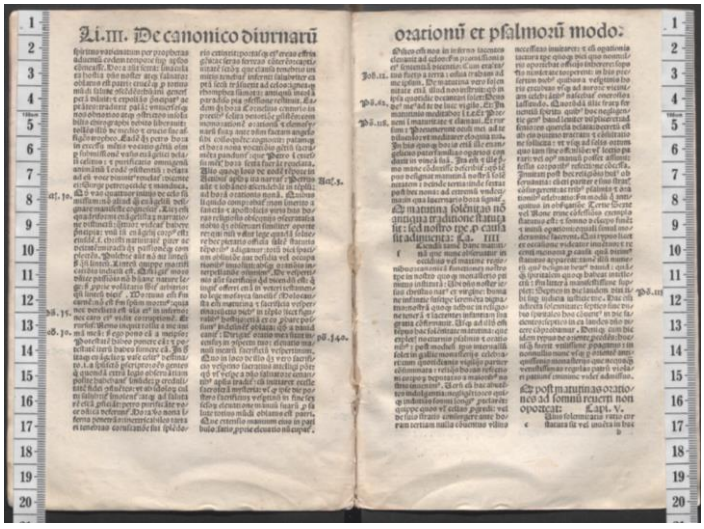


2. En la digitalización de cada ejemplar se utiliza una carta de color en la primera página sobre las 2 hojas del ejemplar.

3. En el proceso de digitalización de las obras se usa una cartulina de color negro debajo de los ejemplares para visualizar de forma correcta los bordes de las obras. En el caso de haber faltas de materia (por ejemplo, orificios por xilófagos) se interpondrán hojas de papel japonés para evitar ver las siguientes páginas.

4. Para el fondo digital manuscritos, monografías, impresos e incunables, las imágenes se digitalizan a doble página. Para el fondo digital Bellas Artes y Cartografía, las imágenes se digitalizan en una a página simple, recto y verso en el caso de que el verso pueda contener algún tipo de información relevante sobre el ejemplar (sello, nº de registro, etc…).

5. En el caso de los incunables, además de la carta de color, se utilizarán dos reglas que van colocadas y pegadas al borde de cada hoja, en el margen, para poder medir el tamaño de los tipos, esencial para identificar el taller de imprenta de origen de cada ejemplar.



6. Cada imagen resultante de la digitalización queda identificada con la signatura del ejemplar.

7. Tras el proceso de digitalización se realiza un primer control de calidad que consiste en revisar que los niveles de la carta de color estén bien, que no haya saltos de páginas, imágenes duplicadas, rotadas, desenfocadas y que todas ellas estén correctamente alineadas.

8. Edición de las imágenes para obtener los ficheros que se utilizarán para la difusión en los portales digitales. En este momento se realizan una serie de procesos que se automatizan con softwares especializados: enderezado automático de páginas, recortes e identificación automática de caracteres para extraer el texto de las páginas. El resultado es un PDF con las siguientes características:

- Versión PDF/A 2u
- Resolución de 150 ppp
- Texto oculto
- En color
- Desprotegidos (sin contraseñas ni otras opciones de seguridad)
- Optimizados y comprimidos para su lectura en web

9. Se procede después a la verificación de nombrado de imágenes, ejecutando un proceso que comprueba que todos los ficheros de cada ejemplar tienen exactamente el mismo nombre y que su número de secuencia comienza por 0000 y continúa sin saltos.

10. De cada objeto digital y su derivado se genera un fichero de metadatos de preservación PREMIS.

11. Finalmente, se realiza el control de calidad de los objetos digitales.

12. Una vez superados los controles de calidad, se procede a la carga de los ficheros de disufisión al portal de acceso.

El proceso de digitalización de **material audiovisual** se realiza siguiendo los siguientes criterios:

1. Se generan los ficheros máster procurando conseguir una reproducción fiel de los originales, sin realizar ningún tipo de interpretación ni aplicar ningún proceso de mejora. Los formatos de los ficheros máster variarán dependiendo del soporte original y atendiendo al contenido.

2. Tanto para las grabaciones sonoras como para los vídeos se han de generar másteres que incluyan la digitalización del soporte analógico junto con los materiales impresos que lo acompañan.

3. Cada ejemplar será digitalizado desde el inicio físico y hasta el final físico del objeto, a pesar de que haya cortes o espacios sin señal. De esta manera se garantiza la reproducción idéntica al ejemplar original.

4. Se digitalizan las pistas de audio tal y como se encuentran en el ejemplar original. Si el original tiene señalmono, se generarán másteres con señal mono.

5. En el caso del material audiovisual, la reproducción digital mantendrá la relación de aspecto (proporción entre el ancho y el alto) de la videograbación original. Además, audio y vídeo deben estar correctamente sincronizados.

6. Partiendo de los másteres se generarán copias de difusión de menor peso. Al igual que en el caso de los másteres, las copias de difusión están formadas por la copia digital de la grabación sonora o del vídeo junto con la copia digital de los materiales impresos que acompañen a estas grabaciones analógicas (carátulas, folletos, etc.).

7. De cada objeto digital y su derivado se genera un fichero de metadatos de preservación PREMIS.

8. Finalmente, se realiza el control de calidad de los objetos digitales.

9. Una vez superados los controles de calidad, se procede a la carga de los ficheros de difusión al portal de acceso.

3.1 Los laboratorios de digitalización

Aunque los proyectos de digitalización masiva se suelen hacer con fondos y empresas externas, la Biblioteca cuenta con sus propios laboratorios de digitalización para realizar algunas líneas internas de digitalización que no encajan en proyectos masivos.

3.1.1 El material librario

Las características técnicas de los escáneres utilizados son las adecuadas para el escaneado de los fondos, y que en modo alguno puedan deteriorar los originales. El modelo de escáner utilizado varía en función de la tipología del documento que se digitaliza, de acuerdo a las especificaciones técnicas de la BNE.

Los elementos que condicionan o intervienen en la elección del sistema de digitalización son:

- **Formatos**:
    - Tamaño: El abanico de formatos comprende, al menos, desde un 8º a mayor de A1, teniendo en cuenta por tanto el porcentaje de los documentos
    - Grosor/peso: El fondo contiene ejemplares de distintos grosores.
    - Desplegables: relacionado con el formato del documento los desplegables suponen un añadido a tener en cuenta, tanto para la manipulación del documento como para las dimensiones del escáner.
- **Características del documento**:
    - Original en color: La mayoría de los manuscritos contienen algún motivo coloreado de interés documental y necesario para su estudio o consulta. Esto hace necesario que el escáner ofrezca garantías de reproducción cromática fiel.
      En relación con el color los códices miniados presentan una dificultad añadida en la reproducción de los dorados. No es aconsejable el modelo copibook para originales donde el color sea una característica esencial.
    - Encuadernación: Las encuadernaciones rígidas no permiten una apertura total del libro en 180º; asimismo las encuadernaciones cerradas pueden tener pérdida de información en la parte central del documento. En ambos casos es aconsejable un escáner que permita realizar la reproducción página a página con objeto de obtener la menor pérdida de información, así como mayor planitud del documento.
    - Material: Un porcentaje importante de manuscritos aparece en soporte pergamino. Las características específicas del soporte: hojas onduladas, pérdida de zonas en el pliego, dureza del mismo etc. determinan una manipulación especial, como inclusión de hojas que aíslen las partes

perdidas, y un escáner que asegure el foco en los distintos planos del documento.

- **Estado de conservación**:
    - Fragilidad del material
    - Falta de información: mutilaciones

- **Diversidad de fondos**: Los materiales específicos de las secciones especiales, presenta diferentes características que motivan la diversidad de sistemas de escaneado:
    - Códices en pergamino
    - Papel ácido y friable
    - Grabados
    - Encuadernaciones históricas
    - Dibujos
    - Fotografías
    - Colección de Ephemera
    - Carteles publicitarios y grandes formatos en general

3.1.1.1 Equipos de captura

La BNE cuenta con diversos modelos de escáneres y cámaras para digitalizar el patrimonio documental y bibliográfico:

- **3 Copibook HD 600– RGB y BW** (tamaño máximo: 428 x 642 mm).

  Este tipo de escáner es un buen referente para la digitalización de documentos encuadernados hasta el formato DIN A2, tiene el más alto nivel de beneficios en la calidad de imagen, productividad, ergonomía, fiabilidad y conservación de documentos.

  Entre sus características destacar las siguientes: Color RGB; Cabezal de cámara: 3 x 35 Mpixels; Resolución óptica de 300 dpi; Auto-foco; Cristal con apertura automática y ajuste de presión sin riesgo; Posición 90° para trabajar sin cristal

  Se emplean fundamentalmente para **la digitalización a color de obras impresas (siglos XVIII a XIX)**

- **1 Digibook Suprascan Quartz A1** (tamaño máximo: 630 x 1030 mm).

  Entre sus ventajas destacar las siguientes: Balanza porta-libros con sensores y cristal con control de presión(opcional); Iluminación móvil no destructiva; el área de escaneo más extensa para A1 110 cm. X 63 cm; Alta calidad de reproducción, con una sola cámara; cámara lineal (CCD lineal móvil); corrección de iluminación en tiempo real; alta productividad; simplicidad de uso.

  Se emplean fundamentalmente para **la digitalización a color de obras manuscritas o impresas encuadernadas, mayoritariamente ilustradas y de fondos gráficos en hojas sueltas (fotografías, carteles, mapas, colecciones de Ephemera, prensa y revistas…)**

- **1 Metis DRS 5070 (tamaño máximo: 500 x 700 mm).**

  Es un escáner cenital de barrido de alta resolución, que comprende un prensa-libros con sensor de presión regulable. Opciones de escaneado en escala de grises de 8 bits y color de 24 bits. Ajuste automático de color, enfoque y exposición.

  Se emplean para **la digitalización a color de obras que exijan una manipulación cuidadosa debido al tipo de soporte (códices, manuscritos miniados, manuscritos con tintas traspasadas, tintas ferrogalicas, encuadernaciones históricas con elementos metálicos…).**

- **5 cámaras fotográficas**, que se utilizan para distintos trabajos de reproducción, ya sea para cartelería, cartografía, dibujos, fotografías, negativos, etc.:

  - **2 cámaras Nikon D700**, con sensor CMOS de 12,1 megapíxeles en formato FX (23,9 x 36 mm). Sensibilidad de ISO de hasta 6400. Disparos continuos de hasta 5 cuadros por segundo. AF de 51 puntos con seguimiento de enfoque tridimensional.

  - **1 cámara Nikon D800**, con sensor CMOS de formato FX (fotograma completo) de 36,3 megapíxeles. ISO 100-6400, que permite alcanzar hasta 25.600 (equivalente) y descender hasta 50 (equivalente). Disparo continuo a 4 fps en el modo recorte FX/5:4. 5 fps en los modos de recorte DX/1,2 aumentos. Motor de procesamiento de imágenes EXPEED 3.

  - **1 cámara Nikon D850**, con sensor CMOS de formato FX de 45.7 megapíxeles de retroiluminación (BSI). ISO 64-25600. Disparo continuo de 7 CPS. Sistema AF de 153 puntos. Rango dinámico ISO hasta 64.

  - **1 cámara Phase One**, modular, que se compone de un cuerpo XF, un visor de cintura, un objetivo de 80 mm, otro de 120 mm y un respaldo IQ3 trichromatic de 100 megapíxeles.

3.1.2 El material audiovisual

El proceso de digitalización del material audiovisual consiste en una conversión analógico / digital, tanto en vídeo como en audio. Para ello, se utilizan interfaces que convierten la información grabada en el soporte analógico en ceros y unos a los que se les asigna un número de muestras y un voltaje determinado. Esos archivos, una vez exportados en los formatos de destino digitales serán los ficheros resultantes.

Los pasos de la digitalización comprenden el correcto conexionado de los equipos, los niveles de grabación, la monitorización y la edición y metaetiquetado de los archivos para su posterior exportación.

3.1.2.1 Manipulación de soportes

La manipulación de los soportes analógicos debe realizarse con guantes de tela especiales. En el caso de los discos no se debe tocar nunca el surco.

En cuanto a la limpieza, se debe eliminar la suciedad o restos que puedan afectar a la reproducción. Los discos de vinilo pueden ser limpiados con agua ionizada y alcohol o

Tergitol, mientras que los de goma laca se deben limpiar con jabón neutro. Los paños de microfibra se deben emplear en el sentido de los surcos (también se debe evitar humedecer las etiquetas). Los discos se deben almacenar en posición vertical sin reclinar y en fundas para evitar dobleces y/o roturas por inclinación o sobrepeso, así como evitar su conservación en álbumes. En cuanto a los soportes de cinta magnética, lo habitual es no manipularlos en exceso y que sea el dispositivo de reproducción el que reciba la limpieza de los cabezales utilizando isopos con alcohol o cintas desmagnetizadoras.

La conservación de estos soportes debe realizarse en condiciones de poca humedad (entre el 30 y el 50%) y a temperaturas comprendidas entre los 4 y 18 grados Celsius.

Para el arreglo de roturas se pueden emplear diferentes técnicas. En el caso de la cinta, un corte tiene fácil solución pegando ambas partes de forma meticulosa y con un adhesivo específico. Las cintas arrugadas serán prácticamente irrecuperables. Los discos de goma laca partidos se pueden reconstruir alrededor del eje del tocadiscos utilizando masilla adhesiva reutilizable (se puede prescindir de ella y no utilizar nada). No se debe utilizar pegamentos. Una vez reconstruido el disco, se debe evitar que la aguja se vea empujada a una superficie más alta para que no sufra daños: debe caer a una superficie más baja para mejorar la transferencia resultante. Las dobleces de los discos se pueden aplanar en un horno de laboratorio con ventilador.

La IASA prioriza la transferencia de discos de acetato a archivos digitales debido a sus características de fabricación. Si están almacenados con alta humedad y/o temperatura, el revestimiento de laca se contrae en su núcleo, mientras que el sustrato de aluminio no. Esto da como resultado tensiones internas que provocan la descamación del revestimiento de laca. Estas escamas se pueden pegar provisionalmente con pequeñas cantidades de vaselina entre la base y la capa de laca. Esto puede permitir la reproducción del disco pero es probable que los efectos a largo plazo de este procedimiento sean perjudiciales, y el método se utiliza solo para intentar reproducir discos que se estimen no reproducibles por cualquier otro medio en la actualidad.

3.1.2.2 Equipos de captura

La BNE cuenta con diversos equipos para la digitalización del patrimonio audivisual que alberga:

- **Reproductores y dispositivos para digitalización de discos de 16, 33$^{1/3}$, 45 y 78 RPM**:
    - Tocadiscos Lenco L75-016
- **Reproductores y dispositivos para digitalización de discos de 33 y 45 RPM**:
    - Technics SL-1200K2 Quartz
    - Technics SL-1210M5G Quartz

  Todos los tocadiscos utilizan el preamplificador Phono TEAC PA-4. La cabina también dispone de las agujas AT-VM95SH/H y AT33EV, además de la Ortofon 235759 colocada en el Lenco.

- **Reproductores y dispositivos para digitalización de cintas de casete**:
    - Doble pletina JVC TD-W 306
- **Reproductores y dispositivos para digitalización de cintas magnéticas abiertas:**
    - Magnetófono Revox B77 MK II
- **Reproductores y dispositivos para digitalización de cintas magnéticas digitales DAT:**
    - Tascam DA-30 MK II
- **Reproductores y dispositivos para digitalización de VHS:**
    - JVC HR-XV3
    - JVC SR-V10
    - JVC SR-VS30 (también MINI-DV)
- **Reproductores y dispositivos para digitalización de BETAMAX**:
    - Sony SL-F30
- **Soportes ópticos**: La información de los soportes ópticos se extrae introduciéndolos en ordenadores con bandejas específicas (bandejas para BluRay, Compact Disc, DVD y HD-DVD).
    - Los CD-A contienen WAV a 44.1 kHz y 16 bits divididos en las pistas que contenga dicho álbum. También se pueden extraer los MP3 a 320 kbps.
    - Los formatos de vídeo como los DVD o los BluRay contienen, por lo general, varios archivos de vídeo: el principal y extras, capítulos, etc. Es obligación del técnico extraer todos los archivos en MP4 a 6500 kbps respetando relación de aspecto, resolución y FPS originales. Además, también se puede construir la IMAGEN ISO a partir de los AUDIO Y VIDEO_TS, que contienen los codecs necesarios para montarla

4. FICHEROS MÁSTER Y DERIVADOS

Como se ha mostrado, tras la digitalización de una obra se generan dos tipos de archivos:

4.1 Archivos de preservación

Se considera archivo máster de preservación a aquel que se ha realizado con la máxima calidad posible para los fines establecidos en cada caso.

Sin alteraciones posteriores a su realización representa la copia de seguridad y se utiliza para producir los ficheros derivados o de difusión.

Las recomendaciones descritas se han definido teniendo en cuenta las directrices de distintas instituciones bibliotecarias que se enfrentan a una casuística análoga a la de la BNE. Tales recomendaciones deben considerarse como valores de referencia mínimos, y pueden variar en función de los avances tecnológicos o necesidades de la institución y de sus usuarios.

4.1.1. Formatos de los archivos de preservación

El cuadro siguiente presenta las características técnicas adecuadas a cada tipo de documento relacionado en la primera columna:

TIPO DE DOCUMENTO	FORMATO	RESOLUCION/ FRECUENCIA DE MUESTREO	PROFUNDIDAD DE COLOR
Material librario (libros, revistas, partituras, manuscritos, material gráfico…)	TIFF, sin compresión	400 ppp	24-bit color
Libros especiales o raros: objetos de gran valor	TIFF sin compresión, RAW	600 ppp	24-bit color
Grabaciones de audio analógico	WAV, MIDI	96 kHz	24-bit color
Grabaciones en vídeo	MOV, MP4	45,88 Mbps	4:2:2

4.2 Archivos de difusión

Los archivos de difusión, son los ficheros resultado de las copias máster, y que se someten a un procesado acorde con las especificaciones técnicas de la difusión en la Biblioteca Digital Hispánica.

Los archivos de difusión se cargan en la Biblioteca Digital Hispánica como dos tipos de objetos:

- Simples: un registro bibliográfico con un único archivo digital
- METS: un registro bibliográfico con varios archivos digitales

4.2.1. Formato de los archivos de difusión

Los archivos de difusión van a tener distintos tipos de formatos según el tipo de material:

TIPO DE DOCUMENTO	ARCHIVO DE DIFUSIÓN
Material librario	PDF y OCR
Material gráfico digitalizado directamente del soporte original (grabados, estampas, dibujos, fotografías, carteles)	JPEG/PDF
Mapas y planos	JPEG/PDF
Audio	MP3
Vídeo	MP4

4.2.1.1. PDF de difusión

La imagen de difusión se generará a color, con el fin de representar fielmente el original físico digitalizado. Dado que han ido variando los distintos criterios en la difusión de ejemplares a lo largo de los años, es posible que algunos de los PDF o JPG presentados en Web puedan verse en blanco y negro, otros con portadas a color y el resto en blanco y negro, o en blanco y negro y páginas internas en color o escala de grises. Estos casos se están sustituyendo progresivamente para mejorar la calidad de lo cargado en el portal de acceso.

5. METADATOS

Los metadatos son el conjunto de informaciones relacionadas con los objetos digitales, cuyo objetivo es facilitar la descripción, búsqueda, uso y gestión de las colecciones digitales.

Los metadatos son las herramientas de las que disponemos para especificar la información contextual asociada a cada documento: su contenido, el historial de las transformaciones sufridas por cada objeto digital, las especificaciones de los equipos físicos necesarias para la construcción de los emuladores, los formatos de cada fichero, los programas que permitirán acceder a cada registro.

Los objetos digitales del portal de acceso se dotan de metadatos descriptivos (consultables en el interfaz público) y de preservación (a través de un perfil METS-PREMIS sólo accesible en los servidores de la BNE).

5.1 Metadatos descriptivos

De cada una de las obras digitalizadas se generan metadatos descriptivos en formato Marc XML para la difusión de los contenidos en los portales de acceso. Los ficheros ".mrc" de las obras a digitalizar se obtienen desde el Catálogo (ISO 2709). Estos ficheros se desglosan en dos partes:

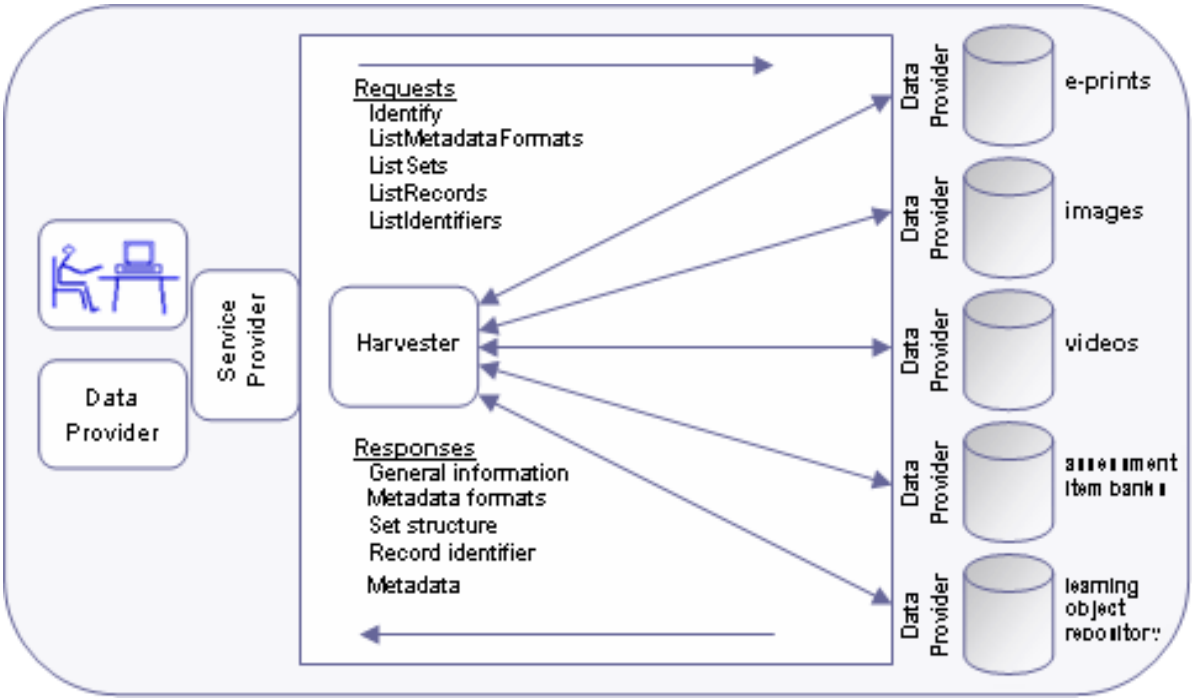
- .mrc que correspondan a objetos simples (documentos compuestos por una sola imagen).

- .mrc que correspondan a objetos complejos (documentos compuestos por varias imágenes). Se generará un fichero XML siguiendo la estructura METS XML por cada objeto complejo que vaya a cargarse en el portal de acceso.

5.1.1 La exportación de metadatos. OAI-PMH

El protocolo OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) se utiliza para la transmisión de metadatos en Internet. Su arquitectura basada en el modelo cliente – servidor pone a disposición del público metadatos en distintos formatos para que puedan ser recuperados. La comunicación se realiza mediante el protocolo http y las respuestas están codificadas en XML.

El portal de acceso cuenta con un servidor OAI (http://bdh-rd.bne.es/oai.vm?verb=ListRecords&metadataPrefix=marcxml), a través del cual se pueden recolectar los metadatos descriptivos en él contenidos. Esto se puede realizar a través de comandos OAI o a través de programas del tipo MEdit, pudiéndose recuperar registros concretos, grupos de registros y los sets de OAI que hay definidos en el portal.



Modelo de exportación de metadatos OAI-PMH

5.2 Metadatos de preservación (METS-PREMIS)

Las obras digitalizadas dentro del proyecto de digitalización sistemática, se conforman en objetos digitales, que se ingestan en el sistema de preservación digital de la BNE.

Cada objeto digital, se recoge en una carpeta contenedora principal, denominada siguiendo unas pautas de normalización y con un identificador único, y en la que se incluyen distintas subcarpetas de:

- **másteres**, con ficheros crudos resultantes de la digitalización (*.tif ,wav, *.mid, *.mov)
- **difusión**, con ficheros que se utilizarán en la web (*.jpg, *.pdf,*.mp3, *.mp4)
- **mets**: que contiene el fichero de metadatos METS-PREMIS, con extensión *.xml

Conviene recordar que los ficheros METS-PREMIS se utilizan en el ámbito de la preservación digital, tanto para documentar (metadatos descriptivos y administrativos) los objetos digitales como para comprobar la integridad de los mismos mediante la consignación de huella digital (MD5), comprobaciones con antivirus y otras validaciones (eventos) en función de las extensiones de los ficheros, y todo ello para garantizar su uso en un futuro.

El perfil del fichero de metadatos METS-PREMIS, se diseñó siguiendo las recomendaciones internacionales en materia de preservación digital, y resulta ser la combinación de dos modelos de metadatos ampliamente extendidos en el mundo bibliotecario: METS (Metadata Encoding Transmission Standard)[1] y PREMIS (Preservation Metadata: Implementation Strategies)[2]. En base a estas se decidió que se utilizarían unidades semánticas PREMIS referidas a dos de las 5 entidades a las que puede referirse un archivo PREMIS: objeto y evento. Dichas unidades semánticas encontrarían ubicación fundamentalmente en la sección de metadatos administrativos de un fichero METS.

De manera muy sintética la estructura o esquema del perfil METS-PREMIS sería la siguiente:

- **Cabecera:** METS header <metsHdr>
  Contiene la lista de los estándares de metadatos utilizados (namespace), la codificación del fichero y datos del creador, editor del fichero….

- **Metadatos descriptivos:** Descriptive metadata <dmdSec>
  Contiene la información bibliográfica que se extrae del catálogo de la BNE, Symphony, mediante una sentencia en el que se especifican los datos de interés como título, autor, fecha edición, notas….

- **Metadatos administrativos:** Administrative metadata <amdSec>

  - Technical metadata <techMD>
    Habrá tantos como ficheros, másteres (M) o derivados (P) formen el objeto digital. Dentro de su unidad semántica <premis:objectCharacteristicsExtension> se especificarán los metadatos técnicos , haciendo uso del modelo de metadatos MIX

  - Rights metadata <rightsMD>
    Habrá tantos como tipos de archivos, en el caso de la BNE serían: máster o derivados, haciendo referencia a los metadatos administrativos mínimos sobre los derechos intelectuales

[1] Disponible en: https://www.loc.gov/standards/mets/
[2] Disponible en: https://www.loc.gov/standards/premis/

asociados al objeto digital o sus partes

- Digital provenance metadata <digiprovMD>
  Tantos como eventos se realicen sobre los ficheros y de los que se guarda información acudiendo a unidades semánticas de la entidad evento de PREMIS

  - Evento Validación (JHOVE)
  - Evento Antivirus
  - Evento Huella digital (Checksum)
  - Evento Validación Derivado (CRAN) (pdf)

- **Inventario de ficheros:** File metadata <fileSec>
  - <fileGrp>
    A modo de "inventario" se agrupan todos los tipos de ficheros (másteres, derivados) descritos, su identificador y todos los metadatos que se han ido asignando a lo largo del fichero METS-PREMIS.
  - <structMap>
    Donde se informa de la estructura y ordenación de todos los ficheros descritos.

En la BNE, los ficheros METS-PREMIS, se generan con la herramienta CRAN MD, que utiliza el lenguaje de programación Python y permite:

- Generación de ficheros METS-PREMIS: *.mets.xml
- Validación de ficheros METS-PREMIS y/o
- Elaboración de informes: *.html

En sus inicios sólo recogía los metadatos propios de material librario (manuscritos, fotografías, grabados, mapas …), pero actualmente se contemplan características y necesidades de otros materiales como prensa o audiovisuales, tanto audio y vídeo, y con perspectivas de ampliarse a materiales digitales con soporte físico, nacidos digitales…

6. CONTROL DE CALIDAD

De los ficheros obtenidos en el proceso de digitalización, se realiza un control de calidad previo consistente en las siguientes tareas:

6.1 Control de calidad previo a la carga en el portal de acceso

- Comprobación y revisión de signaturas que se van a cargar.

- Revisión de xml: De los xml generados se procede a su revisión total de cada uno de los lotes que se va a cargar en Pandora. En la revisión de los xml se centra la atención en las siguientes etiquetas:

  - ✓ Nombre que en el caso de los simples debe de coincidir con el del correspondiente fichero de difusión.

  - ✓ Cabecera: Comprobamos que en la secuencia de las posiciones 21-24 de la cabecera no hay ningún espacio en blanco (ej.: debe de ser 4500 y no 45 0). Y que el tipo de material también está bien codificado.

  - ✓ Etiqueta 008: Comprobamos que las posiciones correspondientes a los datos relativos a fecha de edición se expresan de la forma adecuada (no debe de haber espacios en blanco y los cuatro dígitos han de estar completos) y que el formato de fecha se ajusta a los esquemas del mapeo.

  - ✓ Etiqueta 080: Revisaremos que aparece esta etiqueta con su codificación de materia (CDU) o, en su defecto, que en la etiqueta 899 aparece la codificación simplificada de materia elaborada a tal efecto y basada también en la CDU. Estos códigos y marcas simplificados los podemos consultar en el documento de trabajo elaborado a tal fin.

  - ✓ Etiqueta 300: Se revisa esta etiqueta para comprobar si la obra es o no multivolumen y por tanto que la elección del tipo de carga en Pandora (simples, mets) es la correcta.

  - ✓ Etiqueta 899: Se revisa que esté marcada la signatura digitalizada que se va a cargar. En el caso de que exista otra signatura marcada en el catálogo, añadimos aquella que se va a cargar en el portal de acceso. Además, se verifica que tiene un marcado en el subcampo "x" que indica el responsable de su digitalización.

  - ✓ Etiqueta 856: En los casos en los que ya exista el enlace al portal de acceso, se estudia si se sustituye o se mantiene el que hay. Se analiza también si hay otros marcados 856 atribuibles a enlaces a inventarios como el Catálogo Inventario General de Manuscritos de la BNE, el Catálogo de los códices griegos de la BNE, etc. o a marcados manuales desde descripciones analíticas.

  - ✓ Si se trata de un XML METS, se comprueba el nombrado de archivos del fileSec y el structMap.

- Control de calidad de todos los ficheros PDF y JPG: Comprobación de la existencia de la marca de procedencia de la BNE (en su caso), recorte, calidad de la imagen, OCR (en su caso), etc.

6.2 Control de calidad posterior a la carga en el portal de acceso

La carga de objetos digitalizados se realiza en el servidor de producción de Pandora. Una vez se ha realizado dicha carga, se comprueba su correcta carga en el sistema, prestando especial atención a los siguientes puntos:

- Visualización correcta de los metadatos
- Visualización correcta de los archivos y de su correspondiente marca de agua
- En caso de obras con derechos de autor, se activa el acceso restringido.
- Liberación de las obras, una vez comprobada su carga correcta.
- Actualización del listado de control de cargas, con identificadores y estado de las mismas.
- Actualización del recuento global en BDH.

7. EL PORTAL DE ACCESO

7.1 Sistema de Gestión de Objetos Digitales.

Pandora es el Sistema de Gestión de Objetos Digitales (SGOD) que actualmente utiliza la BNE. Se trata de un repositorio de documentación para recopilar las imágenes digitales de fondos documentales permitiendo a sus usuarios la recuperación de los mismos a partir de avanzados criterios de búsqueda tales como tipo de documento, cabecera, la fecha o rango de fechas y/o términos contenidos en el texto completo de los documentos indexados.

Pandora es un instrumento que permite la recopilación, conservación y difusión de fondos documentales digitalizandos mediante una interfaz de acceso simultáneo y concurrente.

Se trata de un desarrollo que, basándose en el motor de indexación Full Text Lucene (desarrollo de dominio público bajo licencia de software libre de Apache), permite la indexación de archivos digitales, para su inclusión en una base de datos documental.

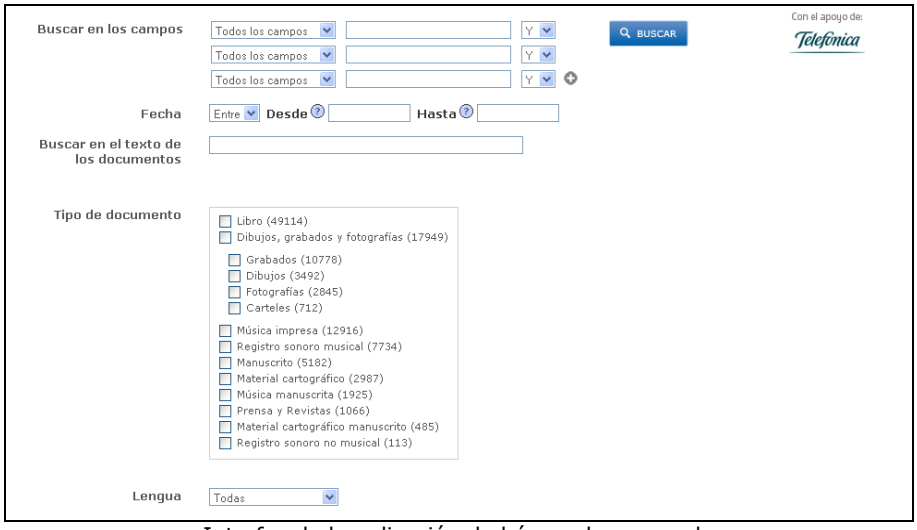
7.2 El motor de búsqueda.

El objetivo principal de un motor de búsqueda en un proyecto de digitalización, es lograr que la localización y navegación sobre volúmenes inmensos de materiales digitalizados sea lo más sencilla, intuitiva y relevante posible.

Actualmente, la Biblioteca Digital Hispánica (BDH) utiliza SOLR como motor de búsqueda, un software de búsqueda de código abierto, que permitirá poder extender y desarrollar funcionalidades propias, al disponer del código fuente.

A través de SOLR se indexan automáticamente los contenidos publicados en Pandora (SGOD), visibles en una interfaz de búsqueda personalizada.

Interfaz de la aplicación de búsqueda sencilla

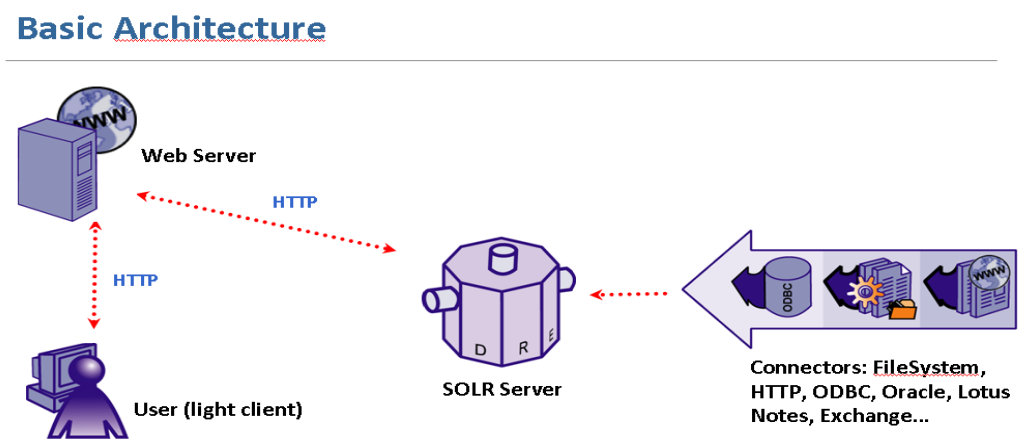


Interfaz de la aplicación de búsqueda avanzada

SOLR indexa a través de OAI, tanto el contenido estructurado (metadatos) como el contenido desestructurado (el propio texto de las obras).

Entre las funcionalidades que ofrece este motor de búsqueda a través de BDH se señalan las siguientes:

- Búsqueda básica y conceptual
- Sugerencias de búsqueda según se escribe (auto-cumplimentación)
- Búsqueda paramétrica (filtros de navegación)
- Hipervinculación (relaciones entre documentos)
- Expansión de consultas
- Resúmenes resaltando las palabras encontradas



Arquitectura básica

7.1 La unificación de los portales de acceso.

Pandora se utiliza en tres entornos diferentes dentro de la Biblioteca Nacional de España: en la Biblioteca Digital Hispánica, para el fondo digitalizado; en la Hemeroteca Digital, que reúne la prensa y publicaciones seriadas digitalizadas; y en el repositorio de publicaciones electrónicas, donde se reúnen los fondos nacidos digitales o los archivos digitales previos a la impresión que los editores deben entregar en virtud de la Ley 8/2022 por la que se modifica la Ley 23/2011 de depósito legal.

Actualmente la Biblioteca está llevando a cabo una migración desde Pandora a una nueva versión del sistema, Biblos, que mejorará las funcionalidades y la infraestructura tecnológica de esta plataforma. La integración y unificación de los tres entornos permitirá establecer un único punto de acceso a todos los contenidos digitales. Además, normalizará los procesos para así optimizar al máximo el aprovechamiento de los recursos.

8. EL SISTEMA DE PRESERVACIÓN DIGITAL

En línea con los fines y funciones de la BNE los ficheros digitalizados se vuelcan en el sistema de preservación digital institucional. Concretamente, y como ya se ha referido se preservan: masters, másters editados (en caso de que éstos existan), archivos de difusión y perfil de metadatos METS-PREMIS. Este sistema asocia al material una determinada política de preservación en la que se definen entre otros parámetros, el número de copias, las políticas de acceso a los ficheros, la realización de comprobaciones periódicas, etc.

El sistema de preservación digital permite chequear la integridad y autenticidad de las colecciones digitales, auditarlas, protegerlas de pérdidas fortuitas y gestionar una migración de formatos de archivos en caso de que estos se quedaran obsoletos.

Consta de un software que permite, entre otras funciones, la recuperación de copias de las imágenes de preservación con el fin de poder atender peticiones de reproducciones de los usuarios u otras instituciones. Se garantiza así la preservación del material analógico (se evita repetir la reproducción digital), del material digital (los ficheros máster originales permanecen en el sistema de preservación), al tiempo que se trabaja a favor de la difusión (atender las peticiones de usuarios y/o instituciones).