

Digitalizar ¿Para qué?

Principios y criterios a tener en cuenta en los proyectos de digitalización del patrimonio documental

Margarita Vázquez de Parga¹

Ante todo agradezco a los organizadores del Congreso a ¿Archivar para qué?, y muy especialmente a Joseph Dager Alba y a Roisida Aguilar, por haberme invitado a participar en la conmemoración de los 150 años del Archivo General de la Nación de Perú.

Ello me ha permitido el reencuentro con este Archivo, por el que siento un gran afecto y con el que he tenido la suerte de colaborar estrechamente durante los años en que estuve al frente de la Dirección de los Archivos Estatales de España.

Gracias a esa colaboración tuve el privilegio de trabajar con D. Guillermo Durand Flórez, a quien me atrevo a definir como el padre del Sistema de Archivos del Perú, con D. Guillermo Lohmann, César Gutiérrez Muñoz, Luis Enrique Tord, Pedro Casas, Aída Luz Mendoza; todos ellos figuras esenciales en la historia del desarrollo de la archivística en Perú. Y no puedo ocultar la satisfacción que me produce el encuentro con tantos archiveros al frente de importantes archivos, que fueron alumnos de las distintas escuelas y cursos para archiveros iberoamericanos que se celebraron en España.

¿QUÉ SIGNIFICA DIGITALIZAR?

De las distintas definiciones que encontramos de la palabra digitalización nos vamos a quedar con la que define la digitalización como “el proceso de convertir información analógica en información digital”, esto es convertir

¹ Cuerpo Facultativo de Archiveros del Estado. Consultora. Informática El Corte Inglés.

cualquier señal de entrada continua (analógica) en una serie de valores numéricos por medio de escáneres.

Esta definición de digitalización deja clara la diferencia entre copia digital, o imagen digital, de un documento analógico, resultante del proceso de digitalización, y documento electrónico, esto es, el documento que se ha producido por medios electrónicos, “born electronic”, o que ha nacido electrónicamente.

Cuando nos referimos a “proyectos” de digitalización de documentos debemos considerar la digitalización como “un proceso en el que hay que realizar una serie de operaciones que pueden agruparse en cuatro grandes apartados:

- **Planificación del proyecto**
 - Definición del objetivo, el alcance del proyecto y plazo de ejecución,
 - Selección de los fondos documentales sobre los que se va a trabajar,
 - Recursos económicos, técnicos y humanos disponibles,
 - Forma de ejecución del proyecto
- **Preparación e indexación de la documentación**
 - Preparación de los documentos, ordenación, paginación, pequeñas labores de conservación, como es la eliminación de dobles, e incluso en ocasiones la restauración de los documentos,
 - Indexación y extracción e incorporación de metadatos,
- **Digitalización**
 - Digitalización propiamente dicha,
- **Trabajo post digitalización**
 - Control de calidad,
 - Almacenamiento de las imágenes digitales para su conservación a largo plazo,
- **Gestión del proyecto**

Un aspecto fundamental que no debe descuidarse es la gestión del proyecto a lo largo de todo el tiempo de ejecución del mismo, para asegurar que se cumplen los objetivos marcados, que se mantiene la calidad y que no hay desvíos de presupuesto ni en el plazo de ejecución².

² Federal Agencies Digitization Initiative (FAGDI) –Still Image Working Group. Technical Guidelines; 2009 ; www.digitizationguidelines.gov

DIGITALIZAR, ¿PARA QUÉ?

Al igual que el lema del Congreso se pregunta Archivar, ¿para qué?, la primera pregunta que nos debemos hacer al acometer un proyecto de digitalización es para qué queremos digitalizar, qué objetivo perseguimos, y qué alcance va a tener el proyecto, ya que en función de estas premisas se deberán definir los criterios a seguir en su ejecución.

Se pueden distinguir distintos tipos de proyectos en función del objetivo que persiguen. Los objetivos perseguidos pueden ser:

- **Facilitar la consulta en los procesos de negocio:**

Para lo que se plantea la digitalización de documentos de uso frecuente como apoyo a la gestión de las administraciones, con el objetivo de evitar la acumulación de documentos en las oficinas y facilitar y agilizar su consulta.

En este caso los documentos originales en soporte físico que tengan valor administrativo, jurídico y fiscal deben conservarse, aunque puede hacerse fuera de los locales de las oficinas. En general se trata de documentos administrativos en formato estándar que pueden digitalizarse con escáneres de alimentación automática. Este tipo de proyectos de digitalización presentan una serie de beneficios entre los que cabe destacar los siguientes:

- o La posibilidad de que varias personas accedan al mismo tiempo y desde múltiples lugares a una misma imagen de un documento,
- o Permite una mayor integración de la transmisión de imágenes con sistemas de flujo de trabajo, lo que incide en una mejora de la productividad tanto en el ámbito de la administración como en la gestión de los negocios,
- o Permite acabar con uno de los problemas que se producen actualmente, que es la existencia de expedientes mixtos, esto es expedientes en los que unos documentos están en soporte papel y otros en soporte electrónico,

- o Se reduce la necesidad de espacio de depósitos documentales para la conservación de los documentos físicos, al poder archivar los documentos originales en depósitos documentales externos.

Sin embargo, frente a estas ventajas hay que prever y tener en cuenta el coste que supone la gestión, mantenimiento y conservación a largo plazo de las imágenes producidas³.

SUSTITUCIÓN DE DOCUMENTOS FÍSICOS

Digitalización de documentos con el objetivo de sustituir los documentos originales en soporte papel por sus copias digitales. Este tipo de proyectos se hace cada vez más frecuente al amparo de la implantación de la administración electrónica.

En España la normativa que regula la implantación de la administración electrónica⁴ está orientada a eliminar el papel en la administración, en cuyo caso las copias digitales de los documentos deben mantener el valor administrativo, jurídico y fiscal de los documentos originales.

En este caso la digitalización tiene que ser certificada, para garantizar que el proceso de digitalización se ha realizado de forma que no se haya modificado el contenido del documento original, se mantenga su integridad y esté disponible y accesible siempre que sea necesario consultarlo. Para ello durante el proceso de digitalización se debe incorporar firma electrónica y sello de tiempo una vez capturada la imagen, almacenar dicha imagen en un formato estándar, incorporar los metadatos de descripción, técnicos y administrativos a la imagen del documento, y su almacenamiento en un gestor documental, todo ello sin más intervención humana que en el proceso de captura de la imagen.

³ Technical Report ISO/TR 13028:2010. Information and Documentation- Implementation guidelines for digitization of records.

⁴ España. Ley 11/2007 de acceso electrónico de los ciudadanos a los servicios públicos. B.O.E de 23 de junio 2007; España. Real Decreto 4/2010, de 8 de enero por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica, B.O.E. de 29 de enero de 2010; Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Digitalización de Documentos, B.O.E. 30 de junio 2011.

En España este tipo de digitalización está regulado por la Agencia Tributaria⁵ para la presentación de las facturas, para lo cual este organismo ha homologado una serie de softwares que cumplen los requisitos de firma electrónica, sellado de tiempo y el plan de calidad establecido por esta entidad.

PRESERVACIÓN Y DIFUSIÓN

Digitalización para facilitar la conservación, consulta y difusión del patrimonio bibliográfico y documental que constituye la memoria histórica de los pueblos.

Es en este último aspecto en el que vamos a centrarnos a lo largo de esta comunicación. Como hemos dicho anteriormente, lo primero que se debe hacer cuando se plantea un proyecto de digitalización masiva de documentos es definir con precisión el objetivo que se persigue alcanzar, que puede ser:

- Preservación de los documentos originales, sustituyendo su consulta por la de su copia digital, de forma que se evite su manipulación y uso frecuente, que es uno de los factores que más inciden en su deterioro. De esta forma se evita su pérdida, que puede ser por extravío al devolver el documento a su unidad de instalación y ubicación en los depósitos documentales, o por sustracción, problema que lamentablemente se da con relativa frecuencia en los Archivos y que da lugar a que se instalen en éstos sofisticadas medidas de seguridad, que a pesar de su elevado coste no evitan al 100% que se produzcan estos casos.
- Reconstrucción del patrimonio documental creando fondos virtuales mediante la digitalización de documentos de instituciones que compartan fondos documentales sobre un mismo tema o período histórico, lo que favorece, así mismo, el desarrollo de proyectos de colaboración entre instituciones.

⁵ www.aeat.es/AEAT...Agencia_Tributaria/.../Digitalizacion_certificada.../Digitalizacion_certificada_de_facturas.shtml

- Difusión del patrimonio documental en internet, en cuyo caso deberán realizarse copias digitales master y derivadas con bajo peso para facilitar su transmisión y acceso en la red. En este caso es necesario que la institución cuente con la infraestructura de redes de comunicaciones necesaria para la transmisión de las imágenes.
- Edición de facsímiles, para lo que la digitalización deberá realizarse con unos parámetros superiores a los necesarios en los casos anteriores con el objetivo de obtener la mejor calidad de las imágenes.

Digitalización bajo demanda, esto es, cuando los investigadores o usuarios del archivo solicitan la reproducción de documentos, sustituir la fotocopia o microfilmación de los documentos por su digitalización, de forma que se vaya creando un archivo de imágenes. En este caso se deberán digitalizar legajos completos, únicamente se digitalizarán documentos individuales en el caso de documentos de relevancia y valor excepcional.

DEFINICIÓN DEL ALCANCE DEL PROYECTO

Una vez determinado el objetivo que se persigue con el proyecto, es fundamental dejar claro desde el inicio el alcance que va a tener, identificando los fondos documentales que se van a digitalizar, las series documentales que incluyen, el volumen de documentos que comporta, así como sus características físicas y estado de conservación,

Al hacer la selección de documentos sobre los que se va a trabajar de acuerdo con los siguientes criterios básicos:

- Deben digitalizarse series completas de documentos, o al menos legajos completos, nunca documentos sueltos, a menos que su importancia y singularidad lo justifique,
- Se seleccionarán las series documentales más consultadas por los usuarios e investigadores, lo que tendrá como resultado ofrecer un mejor servicio a los usuarios al simplificar y facilitar el acceso

a los documentos, al tiempo que evitará el riesgo de su deterioro o extravío, como se ha indicado anteriormente, y al tiempo repercutirá positivamente en la organización del archivo, al poder dedicar a otras labores más productivas al personal dedicado a servir los documentos a quienes los soliciten.

- Se analizará la importancia que tienen los documentos como memoria histórica, en función de lo cual se debe garantizar su conservación por su valor de patrimonio histórico y memoria colectiva,
- Se determinará el estado de conservación en que se encuentran los documentos, y el riesgo potencial de desaparición y pérdida por su fragilidad, e, incluso, en caso necesario se determinarán los procesos de restauración que deban llevarse a cabo antes de su digitalización,

Se analizará el grado de descripción en que se encuentran, y el esfuerzo que sería necesario realizar para su descripción y extracción de los metadatos que se requieren.

Partiendo de estos criterios básicos se seleccionarán los fondos y series documentales objeto del proyecto, en función del objetivo que se pretende alcanzar, de la urgencia de contar con las copias digitales de los documentos, como puede ser para eliminar su consulta directa y evitar su destrucción por su fragilidad o para participar en una exposición o para una edición facsímil, y de los beneficios que se espera obtener, que puede ser la conservación de la memoria histórica, mejorar el servicio a la comunidad científica y/o el mero disfrute del patrimonio documental por la ciudadanía en general facilitando su acceso en internet.

Dos aspectos adicionales a tener en cuenta al acometer este tipo de proyectos y al hacer la selección de los documentos sobre los que se va a trabajar, son la visibilidad y el prestigio que el proyecto va a proyectar sobre los archivos y sobre el colectivo de profesionales de la archivística, así como el rédito político que se obtendrá, ambos aspectos importantes ya que redundarán muy positivamente en el desarrollo de los archivos.

DEFINICIÓN DE LOS ASPECTOS TÉCNICOS DEL PROYECTO

Una vez que se tiene claro el alcance y objetivo del proyecto, se deberán definir los aspectos técnicos bajo los que se realizará el proyecto, como son:

- El tipo o tipos de escáneres a utilizar en función de las características físicas de los documentos,
- El tipo de digitalización a realizar, bitonal, en niveles de gris, o en color,
- Los parámetros de resolución de las imágenes, que se definirán en función de las características de los documentos y del objetivo del proyecto,
- El formato de los ficheros de almacenamiento de las imágenes,
- El esquema de metadatos a incorporar para las imágenes y en la indexación de los documentos,

Otro aspecto fundamental a tener en cuenta es el coste del proyecto y los recursos económicos y humanos disponibles, para lo que hay que valorar el coste de los siguientes elementos:

- De los escáneres, que estará en función del tipo y número de escáneres, necesarios para el tipo de documentos a digitalizar, de su tamaño, de si se trata de folios sueltos o están encuadernados y en este caso del grosor y facilidad de apertura del cuadernillo o libro, y de la capacidad de resolución que alcanzan, y el número de escáneres necesarios en función del volumen de documentos a digitalizar y el plazo de ejecución del proyecto,
- Del proceso de digitalización o captura de las imágenes, en función del tiempo necesario para digitalizar el volumen de documentos determinado, incluyendo el coste de preparación de los documentos previo a su captura, y que estará también en función de si se digitaliza en bitonal, en niveles de gris o en color y de los parámetros de resolución determinados, ya que el tiempo necesario para digitalizar y la capacidad de almacenamiento a 600 dpi, es muy superior al necesario para digitalizar a 400 o a 300 o a 250 dpi,

- De almacenamiento de imágenes, en función de la capacidad de almacenamiento necesaria y del tipo de sistema seleccionado,
- Del número de copias necesarias: de seguridad frente a posibles desastres, para garantizar la conservación; y derivadas, para su uso en local y con menor peso para acceso por la red; del sistema o aplicación para la gestión de la información y de las imágenes: licencias de software base y de aplicación, posibles adaptaciones necesarias, actualizaciones, instalación, mantenimiento del sistema,
- De la descripción o adquisición de metadatos, que estará en función del nivel de descripción que tengan los documentos: si están descritos a nivel de inventario, o a nivel de catálogo, los tipos de metadatos que incluye la descripción y de si hay que complementarla, o el coste de descripción si los documentos están sin describir,
- Del coste de la migración, y en su caso, de grabación de descripciones ya existentes, la forma de acometer el proyecto: con medios propios o contratando el servicio con una empresa especializada,

ADOPCIÓN DE DECISIONES

Las decisiones finales deberán tomarse teniendo en cuenta las variables mencionadas anteriormente y especialmente en función de:

- Los objetivos marcados,
- Las características, volumen y estado de descripción de los fondos documentales seleccionados,
- Los aspectos técnicos definidos,
- Los recursos humanos y económicos disponibles,
- El plazo establecido para obtener resultados,

ELECCIÓN DE ESCÁNERES

La primera decisión que se deberá adoptar es la elección de los escáneres que se van a utilizar, que, como ya se ha dicho, deberá tener en cuenta el tipo de documentos que se van a digitalizar, sus características físicas y su estado de conservación:

- En el caso de que los documentos estén empastados o cosidos, es fundamental utilizar escáneres cenitales con la profundidad de foco adecuada al grosor de los documentos, y con un nivel de resolución acorde con los parámetros de digitalización adoptados. Así mismo es recomendable que el escáner esté dotado de balanza portalibros, de forma que no sufran los documentos durante el proceso de digitalización.
- También se pueden utilizar cámaras digitales de alta resolución instaladas en un soporte y con los elementos de iluminación adecuados, que permiten la captura de las imágenes de forma cenital, aunque el nivel de resolución no llega al de los escáneres cenitales.
- En el caso de que los documentos están sueltos, instalados en legajos o en cajas, se recomienda la utilización de escáneres planos de cama, cuyo coste es menor, y su nivel de resolución es alto.
- Cuando se trata de digitalizar documentos de carácter histórico en ningún caso se deben utilizar escáneres de alimentación automática por el riesgo que hay de ruptura o deterioro de los documentos. Por otra parte, el grosor del papel y sus características puede dar lugar a que los documentos se atasquen en el momento de su ingreso al escáner.
- Si se trata de digitalizar documentos de gran formato, como mapas, planos, etc., hay que utilizar escáneres adaptados para esos tipos de formatos.

PARÁMETROS DE DIGITALIZACIÓN

Una de las decisiones más importantes y que tiene un mayor impacto en el coste y el tiempo de ejecución del proyecto, es el establecimiento de los parámetros de digitalización que se van a adoptar. Hoy día, desde los organismos internacionales, hay una tendencia a recomendar aplicar unos altísimos parámetros de digitalización para garantizar contar con copias digitales de máxima calidad, que permitan su utilización en proyectos que exijan una imagen excelente, solución que únicamente es viable si se dispone de medios económicos y tiempo ilimitados para la realización del proyecto;

pero poco realista porque la necesidad de copias de altísima calidad sólo se da, de forma excepcional, para documentos y proyectos puntuales, lo que no justifica que en una digitalización masiva de documentos se aplique ese criterio de forma generalizada.

En el caso de que fuera necesario disponer de esas copias de calidad se haría una copia TIFF de muy alta calidad que se guardaría como master y copias derivadas en JPEG para el uso en local y en internet.

En realidad esta calidad máxima de la imagen no es perceptible para el ojo humano, por lo que únicamente está justificada en proyectos en los que sea necesaria una reproducción de altísima calidad y a gran tamaño.

Pero en proyectos de digitalización masiva de documentos en los que el objetivo es la preservación de los documentos originales, sustituyendo la consulta de los documentos originales por sus imágenes y su difusión por internet, esta solución no es viable, ya que tanto los costes de digitalización, de almacenamiento y de copias de seguridad sería muy elevado, y la recuperación de las imágenes en ambos casos sería muy lenta, por lo que consideramos que se deben adoptar soluciones realistas, determinando el nivel de resolución suficiente para capturar todos los detalles significativos del documento original y garantizar que sea legible con buena calidad en la pantalla de ordenador y pueda imprimirse con calidad de lectura⁶.

Para este objetivo y para el caso de documentación histórica los parámetros de digitalización que recomendamos, (en base a la experiencia de más de 100 millones de páginas de documentos históricos digitalizados) son los siguientes:

- Digitalización en 250 niveles de gris o 300 en color verdadero. La elección entre digitalizar en niveles de gris o en color dependerá de los siguientes aspectos:

⁶ Este criterio para la digitalización masiva de documentos se propone, igualmente, en el documento publicado por el Ministerio de Cultura de España: DIRECTRICES PARA PROYECTOS DE DIGITALIZACIÓN de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos, Apéndices A, B y C a la traducción española actualizados a 15 de abril de 2008, Federación Internacional de Asociaciones de Bibliotecas, Consejo Internacional de Archivos, UNESCO, 2011.

- Las imágenes en color reflejan con mayor fidelidad los documentos originales, son más agradables a la vista y para su lectura, y su apariencia es de mayor modernidad, sin embargo su peso en bytes es mayor,
- Las imágenes resultantes de la digitalización en niveles de gris tienen menor peso en bytes, por lo que el proceso de digitalización es más rápido, el coste de almacenamiento es menor, las copias de seguridad se hacen en un menor plazo de tiempo y su transmisión por la red, sea intranet o internet, es más rápida.

Esta resolución de captura asegura su lectura, impresión y velocidad de transmisión con alta calidad. Realmente el ojo humano no es capaz de diferenciar entre un documento digitalizado a 300 o a 400 puntos por pulgada (ppp.), siendo necesario para ello hacer una ampliación del 400%; sin embargo, el peso en bytes de una imagen a 400 ppp. puede duplicar el de la misma imagen a 300 ppp.

Con respecto al formato de las imágenes, como hemos mencionado, el ideal sería guardar la imagen master en TIFF y una derivada comprimida en JPEG, pero dado el elevado coste que supone esta solución, tanto en tiempo de captura como en coste de almacenamiento, en los casos de digitalización masiva de documentos se recomienda guardar las imágenes en formato JPEG comprimido.

En las imágenes siguientes se presenta el mismo documento guardado, la primera en formato TIFF y la segunda en JPEG, la diferencia en MB y peso en bytes es muy alto, en tanto que a simple vista no se aprecia diferencia en la calidad de la imagen.

Con respecto a la documentación moderna se recomienda digitalizar a 300 ppp. en binario, y comprimir las imágenes en formato TIFF G-IV, que da imágenes de bajo peso en bytes y permite su proceso posterior mediante OCR.

En definitiva, los parámetros de digitalización deben definirse en función del objetivo perseguido, de los medios económicos disponibles y de los plazos recomendados para la ejecución del proyecto, teniendo siempre en cuenta adoptar una solución realista adoptando el nivel de resolución suficiente para capturar todos los detalles significativos del

documento original y garantizar que sea legible, con buena calidad en la pantalla de ordenador y pueda imprimirse con calidad de lectura. En el caso de documentos de excepcional importancia o en proyectos específicos que exijan una calidad excepcional se podrán adoptar parámetros de digitalización y formatos superiores.

En el caso de documentos deteriorados, fundamentalmente afectados por bibliófagos, se colocará debajo una hoja de papel de color neutro para resaltar las zonas en las que falta el soporte, tanto si se trata de pequeños agujeros como de zonas de mayor tamaño de la página, comprobando que el tono del papel no favorezca el transparentado de las tintas del reverso.

TÉCNICAS DE RESTAURACIÓN ELECTRÓNICA DE LAS IMÁGENES Y SU APLICACIÓN

Uno de los problemas que plantea la documentación histórica es su estado de conservación, ya que con frecuencia los documentos presentan manchas de humedad, problemas de tintas transparentadas debidas a los componentes férricos de las tintas, o que debido al paso del tiempo la tinta está desvaída, por lo que para su lectura es necesario aplicar técnicas de restauración electrónica de las imágenes. La mayor parte de las aplicaciones de visualización de imágenes incorporan algunas técnicas de restauración que se limitan a ampliar la imagen, a intensificar el color, aumentar el brillo y limpiar algo la imagen. En este aspecto es excepcional el sistema de restauración de las imágenes desarrollado para el sistema informático del Archivo General de Indias, inaugurado en el año 1992, y con el que se llevó a cabo la primera experiencia de digitalización masiva de documentos históricos⁷.

7 Bescós, Julián "Image Processing Algorithms for Readability Enhancement of Old Manuscripts", International Electronic Exposition & Conference, Pasadena, CA, 10-13 April 1989, Electronic Imaging 89,1, 392-397 (1989).

Bescós, Julián, Francisco Jaque and Luis Montoro, "Reflectance and Optical Contrast of Old Manuscripts: Wavelength Dependence", SPIE Proceedings (International Society for Optics and Photonics) 1028, 252-262, (1988).

Bescós, Julián, Juan Navarro y Carlos Ramón, "Mejora de Legibilidad de Documentos Antiguos mediante Tratamiento Digital de Imágenes", Comunicaciones del IV Simposium de la Sociedad de Reconocimiento de Formas y Análisis de Imágenes, 51-58, (1990).

Bescós, Julián, Juan Pedro Secilla, and Juan Navarro, "Filtering and Compression of Old Manuscripts by Adaptive Processing Techniques", Proceedings of the Society for Information Display

Para ello, el equipo experto en tratamiento de imágenes desarrolló una serie de algoritmos que permiten restaurar las imágenes eliminando las manchas de humedad, las tintas transparentadas e intensificando las tintas desvaídas, de forma que se facilita extraordinariamente la lectura de documentos prácticamente ilegibles en su estado original.

Para respetar el documento en su estado de conservación original la imagen del documento se captura y guarda tal y como se presenta con los problemas de conservación que tenga, aplicándose los algoritmos de restauración en el momento de la visualización de las imágenes, nunca sobre la copia master almacenada.

DEFINICIÓN DE METADATOS

En un proyecto de digitalización masiva de documentos hay que tener en cuenta que el coste de la extracción de metadatos durante el proceso de indización es enormemente alto; se calcula que supone el 60% del coste total del proyecto⁸, por lo que hay que incorporar los metadatos de descripción, estructurales, administrativos y técnicos imprescindibles para garantizar la conservación a largo plazo de las imágenes y permitir su recuperación de forma precisa y completa y el uso de acuerdo con las normas legales de acceso a los documentos.

Al igual que ocurre con los parámetros de digitalización, los organismos internacionales tienden a recomendar la extracción de un alto número de metadatos, lo que sería ideal en el caso de contar con recursos y tiempo ilimitados, lo que lamentablemente no ocurre frecuentemente, por lo que no es una solución realista.

CONTROL DE CALIDAD

El control de calidad de las imágenes digitales es un aspecto fundamental en los procesos de digitalización de documentación histórica, con el fin de proporcionar el servicio requerido con las copias digitales y evitar tener

International Symposium 1990, 384-387.

⁸ España, Ministerio de Cultura. DIRECTRICES PARA PROYECTOS DE DIGITALIZACIÓN de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos, Apéndices A, B y C a la traducción española actualizados a 15 de abril de 2008, Federación Internacional de Asociaciones de Bibliotecas, Consejo Internacional de Archivos, UNESCO, 2011

que digitalizar de nuevo los documentos digitalizados defectuosamente o dar acceso a consultar los documentos originales. Para ello se sigue la siguiente metodología:

- Control automático por tamaño, comprobando los ficheros digitalizados y que el tamaño de las imágenes está comprendido en un rango estimado, visualizando para su comprobación las imágenes que estuvieran fuera de dicho rango.
- Control por inspección visual de forma global, visualizando un número determinado de imágenes en miniatura y ajustando el tamaño de éstas.
- Verificación de imágenes por inspección visual individual, atendiendo a los siguientes aspectos: calidad y enderezado de la imagen, recorte, imagen “movida” o desenfocada, presencia de objetos ajenos al documento (por ejemplo, dedos del digitalizador), presencia de bandas (rayas), control de la reproducción del color, resolución (en el caso de documentos con letra muy pequeña), etc.
- Verificación mediante inspección visual de que las imágenes se visualizan correctamente.
- Comprobación de que la indexación de las imágenes digitalizadas es correcta y contienen todos los metadatos definidos para el proyecto que se está realizando.

Como resultado del control de calidad se procedería a lo siguiente:

- Corrección de errores detectados, repitiendo, en su caso, el proceso de digitalización, y/o corrección o ampliación de la indexación.
- Se definirá y cumplimentará una Ficha de Control de Calidad, en la que se indicará la unidad documental y páginas sobre los que se ha hecho el control y los identificadores establecidos. Mediante esta ficha se controla, así mismo, el volumen de producción en la digitalización y la calidad del trabajo de cada digitalizador.

- Revisión de los documentos digitalizados por cada digitalizador en ese día.

EL SISTEMA DE GESTIÓN DE LA INFORMACIÓN

Un componente imprescindible del proyecto es la aplicación mediante la cual se van a gestionar las imágenes y su descripción, y que es la que nos va a permitir y facilitar la organización de las descripciones y de las imágenes, así como la consulta y visualización de las imágenes.

El sistema debe permitir organizar y clasificar los documentos en un cuadro de clasificación que permita organizar los fondos documentales en niveles jerárquicos de acuerdo con su estructura de dependencia: fondo, órgano productor, serie documental, expediente, unidad documental simple. La descripción deberá ser conforme a las normas y estándares internacionales, en el caso de los documentos de archivo, deberán seguir las Normas Internacionales de Descripción de Archivos ISAD (G), de descripción de autoridades, ISAAR-CPF, y deberá permitir la consulta y búsqueda en la base de datos por distintos métodos: mediante navegación por el cuadro de clasificación, búsqueda por descriptores, búsqueda sencilla y experta, texto libre, entre otros.

En cuanto a la visualización de las imágenes y la posibilidad de aplicar técnicas de restauración electrónica a las imágenes que presentan problemas de lectura por su estado de conservación, descrito anteriormente, es un valor añadido que incorpora la aplicación desarrollada por Informática El Corte Inglés para el sistema informático del Archivo General de Indias y que se incorporó a la aplicación ArchiDOC-ArchiGÉS, evolución del sistema del Archivo General de Indias.

Así mismo debe contar con un potente sistema de gestión de usuarios que permita asignar funciones y roles a los distintos perfiles de usuarios internos del archivo y controle y gestione los permisos de acceso a la información y a las imágenes en función de las posibles restricciones de acceso. Por supuesto que deberá tener un módulo de digitalización que permita tanto la digitalización mediante integración de distintos tipos de escáneres, como la incorporación de imágenes producidas desde sistemas externos.

El repositorio de las imágenes deberá estar organizado y gestionado de forma que garantice la conservación y gestión de las imágenes a largo plazo, aspecto no trivial ni por los requerimientos del sistema ni por su coste, por lo que en este tipo de proyectos es recomendable compartir esfuerzos y recursos, creando un sistema centralizado de información en el que se puedan almacenar y difundir en la red la información y las imágenes de un conjunto de archivos, que o bien formen parte de un mismo sistema archivístico o que compartan fondos documentales complementarios sobre un tema o período determinado.

EXPERIENCIA DE DIGITALIZACIÓN DE INFORMÁTICA EL CORTE INGLÉS

A lo largo de los últimos 25 años se ha participado en múltiples proyectos en los que la digitalización ha tenido un peso específico, y en los que se han digitalizado muy distintos tipos de documentos, que cubren períodos cronológicos que van desde el siglo X hasta el año actual, en distintos soportes y diversos tipos y formatos, documentos textuales, documentos gráficos, fotografías en distintos soportes: placas de cristal, negativos, positivos; mapas y planos, microfilmes; libros, códices, entre otros, habiendo digitalizado en torno a 100 millones de páginas de documentos.

El inicio de esta experiencia tuvo lugar con el proyecto, focalizado en el Archivo General de Indias⁹, que supuso el diseño y desarrollo de una aplicación específica para la gestión integral de un archivo histórico, en el que se incorporó la digitalización de documentos antiguos, incluyendo un aspecto absolutamente novedoso, como era la corrección de los problemas de legibilidad que pudieran presentar los documentos, debido a problemas de conservación, lo que suponía una experiencia inédita a nivel mundial.

⁹ Algunas referencias de interés para un mayor conocimiento del proyecto: Vázquez de Parga, Margarita. Archivos transparentes en la era de la información. La experiencia en los Archivos Estatales de España. Miscelánea in honores Carola Kecskemeéti. Número especial 54, Archives et bibliothèques de Belgique, 1998, pag. 517-540. Rütimann, Hans y Stuart Lynn, Computerization Project of the Archivo General de Indias Seville, Spain". Comisión on Preservation and Access, March 1992, Washington. González García, Pedro. Informatización del Archivo General de Indias. Estrategias y resultados. ANABAD, Madrid, 1999

El proyecto, que se desarrolló durante los años 1986-1994, estuvo patrocinado por la Fundación Ramón Areces e IBM-España, para conmemorar el V Centenario del Encuentro de Dos Mundos, por lo que debía estar operativo en el año 1992.

El sistema informático supuso el diseño y desarrollo de una aplicación informática que contaba con una base de datos en la que se incorporó, con un modelo de descripción normalizado, toda la información elaborada en el archivo durante sus doscientos años de vida, y se digitalizaron nueve millones de páginas de documentos, lo que permitiría sustituir la consulta de los documentos originales por sus imágenes digitales.

Contaba con módulos para la gestión de los investigadores desde el momento en que accedían al archivo, la consulta de las bases de datos, la visualización en pantalla de los documentos ya digitalizados, o la solicitud de consulta de documentos originales en el caso de los que todavía no lo estuvieran, identificación y control de los depósitos documentales, gestión de la solicitud de reproducción de documentos, incluida la gestión económica que ésta conlleva.

La digitalización de los documentos se realizó en niveles de gris a 200 ppp. y como gran novedad, no superada aún hoy día, se desarrollaron una serie de algoritmos que permiten la restauración electrónica de las imágenes durante la visualización de los documentos en pantallas de alta resolución, lo que permite eliminar manchas y tintas transparentadas, tan frecuentes en los documentos de esas fechas, debido a la composición férrica de las tintas y la intensificación de las tintas desvaídas. De no menor importancia fue la resolución de la logística para la digitalización masiva de documentos, una experiencia nueva e insólita en aquellas fechas.

Al sistema informático se incorporaron cuatro millones de páginas de documentos relativos a Indias, conservados en el Archivo General de Simancas y en el Archivo Histórico Nacional y sus descripciones correspondientes, realizándose el primer proyecto de centralizar en un único sistema la información y las imágenes de documentos conservados en distintos archivos, creándose así el primer archivo virtual, al tiempo que se realizaron las primeras experiencias de facilitar el acceso a la información y a las imágenes mediante la tecnología de la comunicación.

En 1992 el sistema informático del AGI contaba con una base de datos textual integrada por 350.000 registros y un archivo de imágenes con 9 millones de páginas digitalizadas, ordenadores en los puestos de trabajo de los archiveros, un parque de 20 escáneres para la digitalización de los documentos, la sala de investigación contaba con 40 puestos de trabajo informatizados, y los documentos originales que ya habían sido digitalizados sólo se consultaban a través del sistema informático. Todo ello supuso una auténtica revolución en la forma de trabajo de archiveros e investigadores, que se adaptaron sin problema a la nueva situación.

A partir de la inauguración del sistema se continuó avanzando en la aplicación de la tecnología digital, se introdujeron mejoras en el software de digitalización, se continuó la digitalización de documentos, incrementando el archivo de imágenes hasta once millones de páginas, se depuró y enriqueció la base de datos textual, se editaron conjuntos de fondos en CD-Rom y se continuó explorando la forma de facilitar el acceso a la información a través de las redes de comunicación.

DEL SISTEMA INFORMÁTICO DEL ARCHIVO GENERAL DE INDIAS A ARCHIDOC-ARCHIGÉS

El sistema informático del AGI nació con voluntad de permanecer en el tiempo y de convertirse en una plataforma de gestión archivística de amplia implantación, manteniendo su evolución continuamente para adaptarlo a las nuevas posibilidades que iban ofreciendo las tecnologías de la información y la comunicación, se evolucionó a una nueva versión basada en Windows NT, ArchiDOC-ArchiGÉS, que en sus sucesivas versiones ha ido incorporando los estándares y normas a medida que se han ido desarrollando, así como nuevas funcionalidades, mejorando la integración entre sus distintos módulos, incorporando la digitalización en color y adoptando Oracle como gestor de base de datos.

Entre las nuevas funcionalidades desarrolladas cabe destacar nuevas opciones de búsqueda en la base de datos, el multilingüismo, tanto del interfaz como de los datos, el tesoro multilingüe, la gestión de la reprografía incluida la gestión económica que conlleva, aspecto fundamental para los archivos; más flexibilidad en el diseño de fichas de descripción adaptadas a distintos tipos de fondos, la posibilidad de exportar e importar datos en

XML, nuevos formatos para la producción de publicaciones, la adaptación de la gestión de usuarios a la Ley Orgánica de Protección de Datos, así como el desarrollo de un módulo web para facilitar el acceso a la información por internet.

Se desarrollaron funcionalidades específicas para la gestión de las fototecas y para la gestión completa de los archivos administrativos, incluyendo la posibilidad de su integración o interoperabilidad con las aplicaciones de tramitación administrativa, lo que facilita la transferencia de documentos e información de forma automatizada desde las oficinas al archivo.

Últimamente se ha desarrollado una nueva versión, ArchiDOC Web, basada en JAVA y orientada a la gestión de sistemas de archivos que deben comunicarse entre sí, que incorpora las funcionalidades necesarias para dar solución al problema más actual de la archivística, como es la gestión de los documentos electrónicos resultantes de la administración electrónica y la interoperatividad entre aplicaciones.

El sistema de almacenamiento de las imágenes evolucionó desde los discos ópticos utilizados en los comienzos del proyecto a los CD-Rom, y finalmente a discos magnéticos redundantes; los dispositivos de captura de las imágenes también evolucionaron, desde los escáneres planos iniciales, pasando por cámaras digitales de alta resolución, hasta llegar a los escáneres cenitales de alta resolución utilizados actualmente, siempre siguiendo el criterio de utilizar la tecnología más avanzada y adecuada al tipo de documentos a digitalizar.

Así pues, a lo largo de los veinticinco años transcurridos desde el inicio de la aplicación de la digitalización a los documentos de los archivos, hemos informatizado y digitalizado los fondos de más de cien instituciones públicas y privadas de España, Europa y América, formados por muy distintos tipos de objetos, con soportes documentales muy diferentes, y aplicando distintos criterios de digitalización en función del objetivo del proyecto y de las características de los documentos, unas veces en niveles de gris y otras, cada vez más frecuentemente, en color. Entre los muchos proyectos realizados destacaré algunos de los más significativos, como son:

SISTEMA INFORMÁTICO DE LOS DOCUMENTOS DEL KOMINTERN (MOSCÚ)

En el año 1996, el Consejo de Europa, el Consejo Internacional de Archivos y el Servicio Estatal de los Archivos Rusos firmaron un acuerdo marco para la realización de un Proyecto internacional de informatización de los Archivos del Komintern, con el objetivo de preservar sus fondos dado su interés para la historia de Europa y del mundo en el siglo XX.

La realización del proyecto se llevó a cabo bajo la supervisión de un comité internacional (INCOMKA), formado por representantes del Consejo de Europa, Consejo Internacional de Archivos, Archivo del Estado Ruso para la Historia Social y Política (RGASPI) y las ocho organizaciones que patrocinaron el proyecto: Archivos Nacionales de Francia, Archivos Federales de Alemania, Archivos Estatales de Italia, Subdirección de Archivos Estatales del Ministerio de Cultura de España, Archivos Federales de Suiza, Archivos Nacionales de Suecia, Biblioteca del Congreso de Estados Unidos y la “Open Archives Society” de Budapest.

El interés de este fondo documental radica en que se trata de los documentos producidos por el Comité Ejecutivo de la III Internacional Comunista, fundado por Lenin en 1919, con representación de 35 partidos comunistas, con el objetivo de extender el modelo comunista en el mundo, y fundamentales para el conocimiento de la historia de Europa durante ese período.

Tras la disolución del Komintern, en mayo de 1945 los archivos se transfirieron al Partido Comunista de la Unión de Repúblicas Socialistas Soviéticas (URSS), estando custodiados en la actualidad en el Archivo del Estado Ruso para la Historia Social y Política (RGASPI).

El proyecto consistió en crear una base de datos completa de los Archivos del Komintern, digitalizar un millón de páginas de los archivos y facilitar el acceso a la base de datos e imágenes a través del RAGSPI y de las instituciones patrocinadoras.

Para su ejecución, el Comité Internacional seleccionó el software ArchiDOC-ArchiGÉS, de Informática El Corte Inglés, que fue adaptado para el proyecto: se tradujo al ruso, permitiendo procesar caracteres latinos y cirílicos, realizar búsquedas por texto libre en ruso y en inglés, así como incorporar un amplio rango de descriptores de cada documento para facilitar su identificación, recuperación y consulta. Se incorporó a la base de datos todas las referencias descriptivas, revisadas y ampliadas. La Biblioteca del Congreso de Estados Unidos llevó a cabo la transliteración de 173.000 nombres de personas a caracteres latinos y tradujo al inglés los descriptores institucionales y de materias.

Se digitalizaron un millón de páginas de documentos cuyo papel presentaba problemas de manipulación al ser con frecuencia papel cebolla, segundas copias, papel autocopiativo, recortes.

El proyecto se inauguró en junio de 2003, siendo accesibles la base de datos y las imágenes en línea, desde 17 ordenadores de la sala de consulta del RGASPI. La base de datos contenía 210.000 registros, 184.000 descriptores y un total de 1.062.000 imágenes digitales.

El sistema instalado integraba asimismo estaciones para los módulos de Descripción Archivística, Digitalización, Gestión de Usuarios, Gestión de la Sala de Consulta, Control Económico de las Copias de Reprografía y Generación de Informes, Estadísticas y Memorias, quedando formado el personal del RGASPI para la utilización del Sistema ArchiDOC-ArchiGÉS, así como para la administración diaria del mismo.

Tras la implantación del sistema, se dieron de alta las firmas de los otros tres fondos que gestiona el RGASPI: Movimientos Revolucionarios de Europa Occidental; PCUS; y Colecciones Personales de Marx, Engels, Lenin, Stalin, de forma que el acceso a estos fondos, todavía no digitalizados, pudiera gestionarse también con la aplicación informática.

Finalmente, se instaló una réplica de la base de datos e imágenes en estaciones monopuesto e intranet en las diferentes sedes de las organizaciones patrocinadoras: Archivos Nacionales de Italia en Roma;

Open Society Archives at Central European University, Fundación Soros, en Budapest; Archivos Federales de Suiza en Berna; Archivos Nacionales de Suecia en Estocolmo; Biblioteca del Congreso de Estados Unidos en Washington; Archivos Federales de Alemania en Berlín; Archivos Estatales de España en Salamanca y Madrid y Archivos Nacionales de Francia en Dijon, Paris y Fontenebleau.

SISTEMA INFORMÁTICO DEL ARCHIVO GENERAL DE LA NACIÓN DE COLOMBIA

En el año 1997, el Archivo adquirió la aplicación ArchiDOC-ArchiGÉS, lo que constituyó un gran aporte tanto para la difusión y consulta del acervo documental como para la gestión de los usuarios.

Su aplicación en los procesos de descripción, concordante con las normas internacionales y útil para garantizar la compatibilidad y el intercambio de información, es de gran utilidad para la organización y normalización documental, y es fundamental para la restauración de imágenes de documentos históricos, garantizando un acceso amplio, a la vez que protege los documentos originales de la nociva manipulación directa.

Se migraron las bases de datos existentes, (cerca de 45 fondos de la Sección “Colonia”) y se digitalizaron los documentos correspondientes con dos escáneres, y una cámara digital de alta resolución. En 2007 se actualizó con una versión más avanzada y en 2009 se adquirió el módulo de consulta web de ArchiDOC, lo que permite consultar en internet cerca de 74 fondos documentales, con cuatro millones de imágenes y sus descripciones, con la ventaja de que el usuario puede efectuar tratamiento digital a las imágenes, provenientes de cualquier lugar del mundo, desde la pantalla de su computador, siendo innumerables los accesos que se efectúan a través de la web por investigadores de muy diversas nacionalidades.

ArchiDOC es una herramienta fundamental para la puesta en valor de los fondos del archivo y del enorme trabajo realizado por sus profesionales por décadas. El archivo continúa su modernización, siempre ha estado y estará atento a la demanda de información de los usuarios y la sociedad a

través de tecnologías de avanzada como el programa ArchiDOC.

SISTEMA INFORMÁTICO DEL ARCHIVO GENERAL DE LA NACIÓN DE MÉXICO

En el año 2010 el AGN decidió la implementación de un nuevo sistema informático por el que se renovó totalmente su sistema de información, dotándolo de nuevos equipos de cómputo, implantando la aplicación ArchiDOC-ArchiGÉS, así como equipos de digitalización adecuados para el tratamiento de los documentos históricos. Además el proyecto, realizado por Informática El Corte Inglés, ha incluido la migración de 2.300.000 referencias del fondo documental y cerca de 72.000 del bibliográfico, previa adaptación del modelo de descripción al formato MARC 21, y la validación de 15 millones de imágenes ya existentes en el AGN.

Dicho sistema permitirá al archivo tener un control de todo lo que resguarda y a la vez permitirá al investigador tener acceso de manera más rápida a sus acervos. El proyecto de digitalización también permitirá que los acervos más consultados se conserven de manera más adecuada y así poder garantizar que futuras generaciones también tengan acceso a ellos.