

INNOVACIÓN EN PROYECTOS DE DIGITALIZACIÓN

Marcos Lloret Romero
Fujitsu España Services

INTRODUCCIÓN

Vivimos momentos de transformación social de enorme trascendencia y de alcance mundial. Las tecnologías de la información y las comunicaciones, simbolizadas por el fenómeno de internet, aceleran el cambio desde la sociedad industrial hasta la sociedad del conocimiento. Estos cambios, provocados por la inquietud del ser humano, generan múltiples escenarios de trabajo vinculados a la cultura y a los *concellos*.

Durante los últimos años se han producido diversas acciones impulsadas desde la Administración y desde organismos privados que tienen como finalidad introducir plenamente la sociedad de la información en el ámbito cultural, contribuyendo así a la incorporación de la sociedad española al uso y aprovechamiento de las tecnologías de la información y de las comunicaciones (TIC).

Estas acciones están implicando desde el primer momento, y mayoritariamente, la implantación o actualización de infraestructuras, como la incorporación de cableado de banda ancha o la dotación de ordenadores, entre otros. No cabe duda que este es el primer paso para introducir el uso de las TIC en los procesos cotidianos, aunque los cambios actuales deben proporcionar servicios útiles para los ciudadanos.

En este sentido, es necesario construir sobre los esfuerzos actuales y potenciarlos al máximo, aprovechando para cambiar los paradigmas de acceso a la información y su utilización; conceptos que representan la temporalidad del acceso y la ubicación. Ya no es necesario hacer colas ni desplazarse al ayuntamiento para solicitar el certificado de empadronamiento, sino que es posible hacerlo desde casa en cualquier momento.



Argumentos tan importantes como la localización geográfica y la disponibilidad horaria han dejado paso al acceso 24x7¹ y desde cualquier ubicación. Esta nueva época se caracteriza por la capacidad de acceder a la información instantáneamente desde cualquier lugar y en la forma deseada.

Bajo este mismo principio, el valor cultural no está en las tecnologías en sí mismas, sino en el uso que hacemos de ellas. Y es precisamente en este uso donde se aúnan esfuerzos para acercar la cultura de los pueblos a sus gentes. En este sentido, representa la aproximación del conocimiento universal a todos. Para provocar dicho acercamiento no es necesario grandes proyectos ni grandes desembolsos de dinero, sino ideas y proyectos concretos que aporten un nuevo punto de vista.

La tecnología existe, es alcanzable y, la mayor parte del tiempo, operativa. Dejemos de tener fondos de conocimiento (libros, fotografías, revistas, etc.) ubicados en espacios cerrados donde son sólo unos pocos quienes tienen acceso a los mismos. Es

1. Acceso 24 horas al día durante los 7 días de la semana los 365 días al año.

necesario que existan iniciativas de la Administración, unidas a iniciativas privadas, que actúen en tal dirección.

Los proyectos de digitalización tienen como objetivo último aprovechar al máximo el potencial de las nuevas tecnologías para desarrollar un espacio de información y colaboración donde tengan cabida y participación todos los ciudadanos y las fuentes de conocimiento.

CONTEXTO DE LA DIGITALIZACIÓN

Todas las iniciativas y proyectos encaminados a cambios en los procesos de captura de información, y especialmente aquellos basados en la digitalización, buscan alguno de los siguientes objetivos generales:

- Ampliar el acceso.
 - Acceso universal y a cualquier hora. Acercar la cultura de los pueblos a sus gentes mediante internet.
 - Promover la participación activa de los ciudadanos a través del uso de las TIC.
- Tener una colección basada en contenido.
 - Catalogación.
 - Investigación.
- Preservar.

Ótros objetivos más concretos que perseguiremos en un proyecto de digitalización en el ámbito de la Administración son:

- Mejorar la accesibilidad, seguridad y fiabilidad de los datos.
- Distribuir la documentación a cualquier puesto de trabajo mediante un navegador Web.
- Facilitar la catalogación, búsqueda y reutilización de documentos.
- Ahorrar en coste y espacio.
 - Convertir documentos en papel a documentos en imagen.
 - Eliminar el papel y ahorrar espacio y coste de almacenamiento de los archivos físicos.
 - Evitar el desplazamiento de los originales.

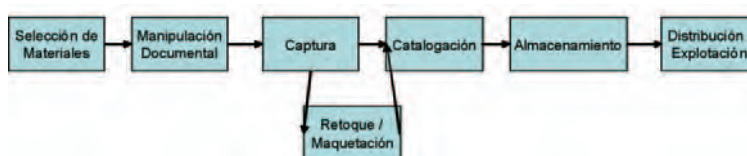
Estos objetivos, que debemos valorar a la hora de afrontar cualquier proyecto de digitalización, nos proporcionarán implícitamente una serie de requisitos tecnológicos y funcionales que muchas veces nos condicionarán el proceso de digitalización.

Como ejemplo, si nuestro objetivo en un proyecto fuera preservar los originales será fundamental establecer un mecanismo de captura que permita aproximar al máximo la copia al original mostrando menor consideración aspectos como el tamaño de almacenamiento, la distribución o la explotación de la información. En otro caso, si lo que perseguimos es ampliar el acceso a la información priorizaremos aquellos mecanismos que nos permiten proporcionar la mayor información a la mayor cantidad de usuarios en el menor tiempo.

Concellos y cultura generalmente persiguen objetivos opuestos, aún cuando las líneas de actuación sean similares. En los proyectos de digitalización en *concellos*, habitualmente se busca el reducir costo y espacio, así como el acceso a los documentos desde cualquier puesto sin necesidad de *viajes* de documentación; se busca el convertir los documentos en papel a documentos en imagen. En proyectos en cultura normalmente se aborda como principal objetivo el preservar los originales y el disponer de una colección basada en contenido que sea útil tanto para procesos de catalogación como para investigación.

PROCESOS DE DIGITALIZACIÓN

A la hora de afrontar cualquier proyecto de digitalización, debemos tratarlo en su conjunto y realizar un análisis cuyo resultado nos marcará las pautas a seguir en el proceso. Si representamos de forma gráfica las fases de un proyecto de digitalización obtendremos el siguiente esquema:



Para abordar todas estas fases es imprescindible contar con un equipo de personas multidisciplinar con especialistas en procesos de catalogación y en sistemas de digitalización.

FASES DE UN PROYECTO DE DIGITALIZACIÓN

Selección de materiales. Nuestra primera fase será inspeccionar los materiales que componen el archivo y conocerlo en profundidad para poder decidir la selección de materiales que digitalizar y la ordenación de la misma. Existen diferentes formas de ordenar los materiales para acceder al proceso de captura, como puede ser por fecha de publicación, estado de la obra, autor, etc. Es importante, en esta fase, seleccionar aquellos materiales que queremos difundir y aquellos que queremos preservar. Es también en esta fase donde procederemos a la evaluación física de los materiales, que seguramente nos condicionará la selección definitiva.

Captura. El proceso de captura está asociado directamente al tipo de material (libros, fotografías, documentos multimedia, cartografía, etc.), al estado de conservación de los materiales o a su antigüedad. Para seleccionar el mecanismo de captura que vamos a utilizar, como norma general, deberemos escoger entre aquéllos que representen fielmente los originales y que no alteren el estado de conservación de los mismos.

Cuando abordamos el proceso de captura escogeremos, a modo de ejemplo, entre algunos de los siguientes mecanismos:

Proceso de digitalización	OCR ² . Captura. Imagen
Selección del escaner	Escáner escala de grises, color, zenitales, etc. Cámaras digitales
Quién realiza el proceso	Personal especializado en manipulación y digitalización
Selección de formatos para preservación y para difusión	TIFF ³ , JPEG ⁴ , PNG ⁵

2. OCR (Optical Character Recognition).

Antes de iniciar el proceso de captura es necesario analizar todas las fases del proyecto para poder definir objetivamente todos aquellos requisitos que tendrá.

Catalogación. Una vez definido el mecanismo de captura y digitalización es necesario que estudiemos qué información queremos obtener de los materiales (fecha de publicación, autor, descriptores, CDU⁶, lugares de edición y publicación, etc.). Para ello definiremos una serie de atributos descriptivos diferenciando los obligatorios de los opcionales. Si por el contrario decidimos utilizar un formato de catalogación basado en un estándar de facto (MARC⁷, ISBD⁸, ISAD(G)⁹, etc.), estos atributos descriptivos estarán establecidos a priori por él. Es conveniente definir también el formato mediante el cual se generará el número de registro único de cada material dentro del catálogo. Este código de registro podría

3. El TIFF (Tagged Image File Format) es un formato padronizado de archivo que permite el almacenamiento e intercambio de informaciones gráficas en formato raster provenientes de «scanners», *frame grabbers* y de programas de retoque y edición de fotos.
4. JPEG (Joint Photographic Experts Group) es un formato de imagen en el Web que fue creado para reducir la talla de las fotografías al comprimirlas. JPEG se refiere al formato de los ficheros gráficos que utiliza este estándar de compresión.
5. PNG son las siglas de Portable Network Graphics, un formato de compresión de imágenes aprobado por el World Wide Web Consortium (W3C) como sustituto del formato GIF. Los archivo tipo GIF utilizan un algoritmo de compresión de datos que está patentado, mientras que el formato PNG no está patentado y no necesita licencia para su utilización.
6. CDU (Clasificación Decimal Universal). La CDU es un sistema de clasificación temática de los libros que permite asignar a cada tema un código específico. La CDU se llama decimal porque las cifras usadas para catalogar los libros se consideran como decimales, no como números enteros. Cada decimal, a su vez, puede subdividirse en un nuevo número decimal, sin que el orden de los números se altere.
7. MARC (Machine Readable Cataloging) define los mecanismos de intercambio, uso e interpretación de la información bibliotecaria.
8. ISBD (A). Descripción bibliográfica internacional normalizada para publicaciones monográficas antiguas.
ISBD (CF). Descripción bibliográfica internacional normalizada para archivos de ordenador.

contener información sobre la ubicación física del material en el fondo así como información numérica secuencial asociada al mismo en el catálogo.

En este proceso de catalogación, una vez definido el formato, deberemos seleccionar el mecanismo mediante el cual catalogadores, investigadores, bibliotecarios y, en general, cualquier persona que pueda aportar información al catálogo tendrán acceso al mismo. Si diferenciamos en cuanto al tipo de acceso las herramientas de catalogación podríamos tener:

- Herramientas cliente / servidor. Útiles si las personas que van a catalogar están ubicadas físicamente en el mismo lugar (la misma red de datos).
- Herramientas Web. En este caso se aprovecha un canal como internet, donde cualquier persona con permiso de acceso puede acceder al catálogo y a su catalogación.

Retoque / Maquetación. Solo en el caso de que los originales se encontraran deteriorados será conveniente analizar el mecanismo de retoque que abordar. En los proyectos que busquen la digitalización *fiel* de los originales no parece necesario realizar modificaciones sobre los mismos. En aquellos proyectos donde la información del documento digitalizado sea importante (*i. e.* una firma en un documento, un sello...) deberemos aplicar técnicas de retoque sobre los originales destacando la zona o zonas donde se

ISBD (CM). Descripción bibliográfica internacional normalizada para material cartográfico.

ISBD (G). Descripción bibliográfica internacional normalizada general.

ISBD (M). Descripción bibliográfica internacional normalizada para publicaciones monográficas.

ISBD (NBM). Descripción bibliográfica internacional normalizada para materiales no librarios.

ISBD (PM). Descripción bibliográfica internacional normalizada para música impresa.

ISBD (S). Descripción bibliográfica internacional normalizada para publicaciones seriadas.

9. ISAD (G). Norma para la descripción de las imágenes fotográficas.

ubique la información que deseamos obtener. Siempre que realicemos cualquier modificación del original deberemos identificarlo como una copia del mismo. De esta forma, siempre tendremos localizados aquellos cambios realizados sobre los originales.

Almacenamiento. Es importante medir y definir el mecanismo de almacenamiento físico de la información digitalizada.

Imaginemos que vamos a digitalizar 2000 documentos tamaño DIN A4¹⁰ y hemos definido que utilizaremos el formato de imagen TIFFxi (300 dpi), cada fichero ocupará aproximadamente 100 KB, ocupando un espacio total en disco de 200 MB. En cambio si lo que digitalizamos son 2000 fotografías y hemos definido que utilizaremos el formato de imagen JPEG extendido (300 dpi), cada fichero ocupará aproximadamente 900 KB, ocupando un espacio total en disco de casi 2 GB.

Dispositivos de almacenamiento	Disco duro/Cinta/DVD/CD-ROM
Características del almacenamiento	Almacenamiento redundante Copias de seguridad Réplicas de disco

Distribución. En esta fase ya debemos disponer de los documentos digitalizados y catalogados, con lo cual decidiremos la información que es pública y cuál no. Se puede llegar al acuerdo de publicar solo un tipo de material, o referido a un autor, o por un periodo de fechas, un tema, etc. Además de definir qué es importante hay que definir como (acceso desde internet, acceso desde los ordenadores del museo...).

En la distribución de la información deberemos tener en cuenta la legislación vigente en materia de derechos de autor puesto que internet es una enorme puerta de entrada de usuarios a

10. DIN A4 es un formato de medidas (297 x 210 mm) similar al tamaño de un folio y corresponde al papel de uso más corriente.

nuestra información pero, por desgracia, también lo es para la copia ilegal. Es por esto por lo que deberemos establecer mecanismos de protección de los originales que velen por la autenticidad de la información.

Explotación. La última fase del proyecto, pero quizá una de las más importantes, es la explotación o utilización del archivo digital con el fin de dar a conocer la información a los usuarios o, incluso, como posible fuente de obtención de recursos. Entre algunos ejemplos de uso en proyectos de digitalización destacamos:

- Consultas al catálogo digital mediante múltiples criterios.
- Envío de una postal a un amigo.
- Descarga de una imagen como fondo de escritorio.
- Participación activa de los ciudadanos en los procesos de catalogación. Imaginemos que mostramos fotografías digitalizadas de la década de los 30 a los ciudadanos teniendo como utilidad la posibilidad de añadir comentarios sobre las fotografías.
- Solicitud de copias.
- Puntos de información.

Una vez descritas todas las fases en un proyecto de digitalización tan solo queda comenzar nuestro primer proyecto de digitalización.