

# 49

## DIGITALIZACIÓN DE DOCUMENTOS: PROYECTO DIGISAN

Francisco Belmonte Díaz

Diseño e implementación de Sistemas Informáticos. Coordinación de Tareas de Programación  
Servicio de Gestión Informática. Consejería de Sanidad de la Región de Murcia

Salvador de Sampedro González

Funciones de Integración entre Sistemas y Programación. Estudios e implementación de Software  
Servicio de Gestión Informática. Consejería de Sanidad de la Región de Murcia

María Brotóns Román

Coordinación del Área de Programación. Análisis de Proyectos Informáticos  
Servicio de Gestión Informática. Consejería de Sanidad de la Región de Murcia



En el ámbito de cualquier sistema de información que maneja documentos físicos para la toma de decisiones, y especialmente en el de la administración pública, nos encontramos con los siguientes problemas en relación a los archivos necesarios para almacenar tales documentos, a su ubicación y al uso que se hace de ellos:

- **Accesibilidad**, atendiendo a la necesidad de que el archivo físico esté ubicado en un lugar distante del que se hallan los usuarios. Esto es algo normal si el volumen de documentos es elevado y las instalaciones donde se realiza la gestión tiene problemas de espacio, o si la organización tiene instalaciones geográficamente dispersas.
- **Disponibilidad**, el almacenamiento físico de los documentos en archivadores, armarios, etc., a los que inevitablemente tenemos que acudir para su consulta, conlleva un coste en el tiempo de disponibilidad o la dependencia de personas especializadas en tareas de archivo a las que se debe solicitar la información que se desea.
- **Concurrencia**, la posibilidad de que varias personas necesiten acceder a la vez a los mismos documentos.

Para solucionar estos problemas, se concibió Digisan como un sistema de gestión de documentos cuyo objetivo inicial es la incorporación de toda la información en soporte físico que maneja un sistema (papel, radiografías, mamografías, etc.) convirtiéndolos en documentos electrónicos, de manera que:

- Una vez digitalizada la información, el soporte físico se puede almacenar en un sitio especializado a tal efecto, o incluso destruirse (si se cumplen los preceptos legales establecidos). Accederemos a la información digitalizada siempre a través de una aplicación informática, con la que visualizaremos una imagen digital exacta, desapareciendo el problema del espacio y posibilitando el acceso en tiempo real, sin esperas ni problemas de concurrencia.
- Accediendo a la base de datos de documentos digitalizados mediante el uso de las nuevas tecnologías, principalmente Internet, podemos disponer de la información de forma segura prácticamente desde cualquier lugar del mundo.

Pero Digisan es más que un sistema de digitalización de documentos. En primer lugar, destaca por la utilización de una base de datos como soporte de la información, descartando la utilización de sistemas de ficheros y directorios tradicionales. La base de datos facilita la gestión de la seguridad de acceso, garantiza la integridad de la información y optimiza el acceso concurrente, y lo que es más importante, facilita el diseño de un modelo de datos relativo a documentos digitales reutilizable e integrable en cualquier otro modelo de datos de gestión.

Para comprender los conceptos de este modelo de datos, vamos a describir la funcionalidad de Digisan, es decir, las fases por las que pasa un documento físico desde que llega al sistema hasta que queda digitalizado, incorporado a la base de datos y asociado a su expediente.

1. *Purgado y clasificación de documentos manual*. En primer lugar, es preciso preparar los documentos para su escaneo quitando grapas, sobres, etc. e incluso descartando documentos que nada tienen que ver con el sistema. A continuación, se van clasificando en lo que hemos denominado lotes, de manera que:
  - a. Los documentos físicos se agrupan en función de la *unidad de gestión* a la que pertenecen.

- b. Para cada grupo, se crean a su vez *subgrupos* que reúnen o bien documentos del mismo tipo (p.e., acuses de recibo) o bien documentos de un mismo expediente. Esta clasificación se hace para optimizar la fase posterior de *indexación*.
2. *Captura de documentos*. Se escanean los documentos, de manera que para cada cara de las hojas se genera un fichero en formato TIFF al que hemos denominado elemento. Estos elementos se ubican en directorios, que se corresponden con los lotes preestablecidos. Por tanto, al final de esta fase tenemos una estructura de directorios (los lotes) que contienen los ficheros imagen (los elementos).
3. *Carga a la base de datos*. Se seleccionan los lotes escaneados y se cargan a la base de datos, es decir, los ficheros imagen se convierten en registros de tablas de la base de datos con información adicional como fecha de carga, usuario que la realiza, etc.. Se elimina la estructura de directorios una vez cargada.
4. *Indexación*. Se crean los *documentos electrónicos* o *eDocumentos* y se vinculan al expediente correspondiente a partir de los lotes pendientes de indexar, de manera que:
  - a. Se localiza un expediente, mediante un sistema de búsquedas ágil.
  - b. Se crean los *eDocumentos* y se indexan los *elementos* que pertenecen al mismo documento (un elemento corresponde a una cara de una hoja).
  - c. Se introducen los *atributos* previamente definidos en base al tipo de documento. Estos atributos serán reutilizados en la gestión.

Así pues, cualquier sistema de información que use Digisan se encuentra con los siguientes conceptos ya estudiados y diseñados:

- Elemento. Cada una de las caras digitalizadas de un documento físico.
- Lote. Conjunto de elementos digitalizados en un momento dado para una unidad de gestión. Son de un mismo tipo o bien pertenecen a un mismo expediente.
- eDocumento. Equivalente digital de un documento físico, compuesto por los elementos (caras) en el orden correspondiente, junto con los atributos necesarios para su modelado y posterior gestión.
- Expediente. Asociación de una persona (o entidad) en el ámbito de una unidad de gestión, que agrupa el conjunto de trámites que se realizan y que contiene los eDocumentos necesarios para la toma de decisiones.

Desde el punto de vista técnico, la ventaja que aporta el hecho de materializar el modelo de Digisan en una base de datos es que el administrador de la misma va a disponer de un conjunto de objetos ya creados (tablas, vistas, procedimientos, funciones, etc.) con sus restricciones de integridad ya definidas.

Las fases descritas del funcionamiento de Digisan apoyadas en su modelo, se han desarrollado:

- Definiendo la infraestructura necesaria para la fase de escaneado: tipo y modelo de escáner, software de escaneado a usar, estructura de ficheros y directorios para documentos escaneados, etc..
- Mediante un módulo programado en Java para las fases de carga (o preprocesamiento), indexación y consulta de eDocumentos y expedientes.

De esta manera, la implantación de Digisan en cualquier proyecto es inmediata: sólo hay que definir los atributos particulares de cada uno de los tipos de documentos que maneja y los

permisos de acceso de los usuarios a la base de datos, labor que también se ha simplificado al utilizarse un mecanismo basado en roles (basta con asignar el rol correspondiente al usuario).

Hay que resaltar el hecho de que la programación se haya realizado en *servlets* Java por dos motivos fundamentales:

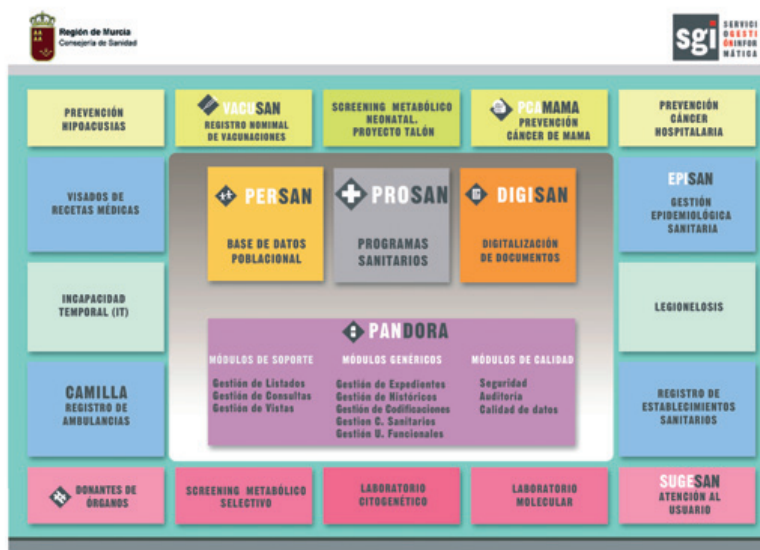
- La independencia de la plataforma cliente: Windows, Linux, etc.
- La accesibilidad a través de un navegador Web desde prácticamente cualquier lugar.

En el aspecto organizativo, se pueden separar las tareas de gestión de un sistema de información que utilice Digisan, de manera que:

- La incorporación de los documentos a la base de datos la realicen profesionales especializados en clasificación de documentos, manejo de escáner de alta tecnología e indexación.
- Los profesionales que realmente trabajan con los expedientes son expertos en su gestión, utilizan la información digitalizada como un elemento de toma de decisiones y no precisan conocer detalles acerca de la digitalización.
- En un sistema de grandes dimensiones, Digisan permite la existencia de varias unidades de digitalización o de una única unidad centralizada, que cree y asocie los eDocumentos a los correspondientes expedientes de cada unidad de gestión. De esta forma, se optimizan los recursos necesarios para invertir en la digitalización, principalmente dado el elevado costo de los escáneres de alto rendimiento necesarios, que serían rápidamente amortizados.
- Para un sistema que decida utilizar Digisan y disponga ya de información histórica previa, permite la creación de un grupo especializado para su digitalización, que conviva con el resto de profesionales que trabajan en el día a día.

## DIGISAN EN LA CONSEJERÍA DE SANIDAD

Para explicar la implantación de Digisan en la Consejería de Sanidad, es necesario comprender la integración entre los sistemas de información existentes.



Todos estos sistemas de información tienen una parte común a la que hemos denominado PROSAN: aspectos relativos al **modelo de datos**, y funcionalidades **comunes** de todo sistema en el ámbito de la Consejería de Sanidad.

En particular, el concepto de Expediente, como identificador único de una persona o entidad, unido al concepto de Documento constituyen los nexos de unión a DIGISAN en el marco de Proyectos integrados en PROSAN.

Dado que en nuestro caso los expedientes son relativos a personas, debemos mencionar otro módulo independiente denominado PERSAN, cuyo nexo de unión con PROSAN es también el expediente. Trata de aspectos relativos a la persona en el ámbito sanitario: datos personales, datos identificativos, medios de comunicación, direcciones, centro de Atención Primaria, Médico de cabecera, etc.

PROSAN, PERSAN y DIGISAN utilizan los conceptos de otro módulo importante en este esquema: PANDORA, que aporta un conjunto de *módulos genéricos* (gestión de información histórica, codificaciones, unidades funcionales o de gestión, etc.), *módulos de soporte* (gestión de consultas, vistas y listados definibles por el usuario, etc.) y *módulos de calidad* (gestión de seguridad de acceso, auditoría de la base de datos, herramientas para la mejora en la calidad de los datos, como eliminación de expedientes y personas duplicadas, etc.).

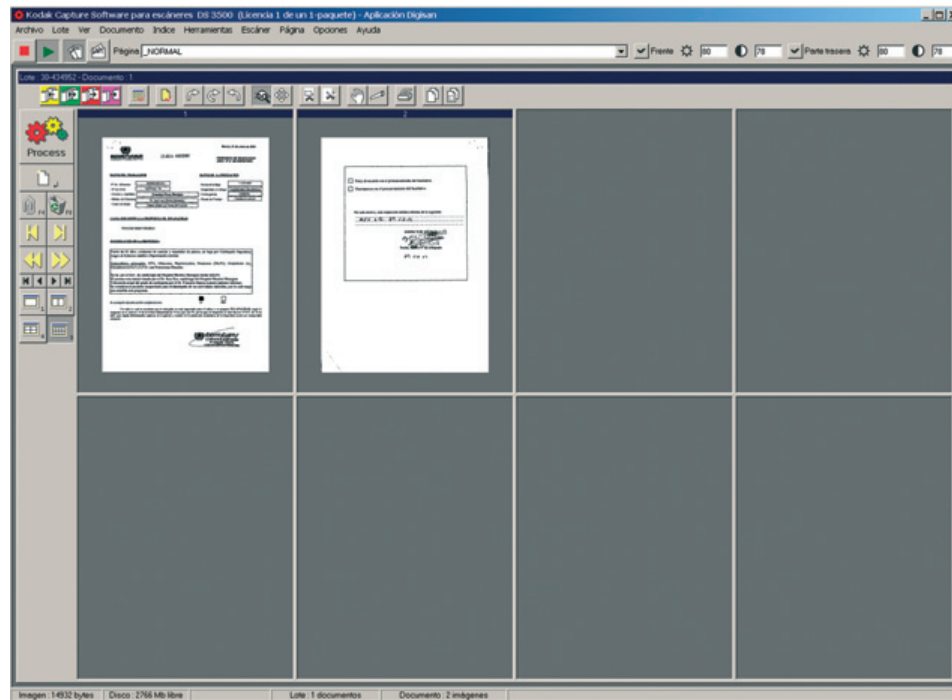
Resumiendo podemos decir que un sistema de información a implantar, como por ejemplo IT, ya nace con las funcionalidades y el modelo de datos de PROSAN, y por tanto, también las de PERSAN, lo que quiere decir que ya tenemos diseñada (e incluso implementada en la parte de la aplicación informática) todo lo relativo a la persona, al expediente, etc. Si además el sistema precisa de digitalización de documentos, reutiliza DIGISAN. Únicamente quedan por desarrollar los atributos particulares de los documentos de IT: partes de baja, notificaciones, etc..

## UN CASO CONCRETO: DIGISAN-IT.

Vamos a describir la aplicación de Digisan a la gestión de expedientes de Incapacidad Temporal (IT).

### ESCANEO DE DOCUMENTOS

En primer lugar, con respecto a la fase de escaneado, se han adquirido escáneres de gran capacidad debido al gran volumen de datos a manejar y en previsión de incorporar al equipo de digitalización otras unidades de gestión del sistema. Se utiliza el software del escáner para la generación de la estructura de archivos y directorios para la carga.



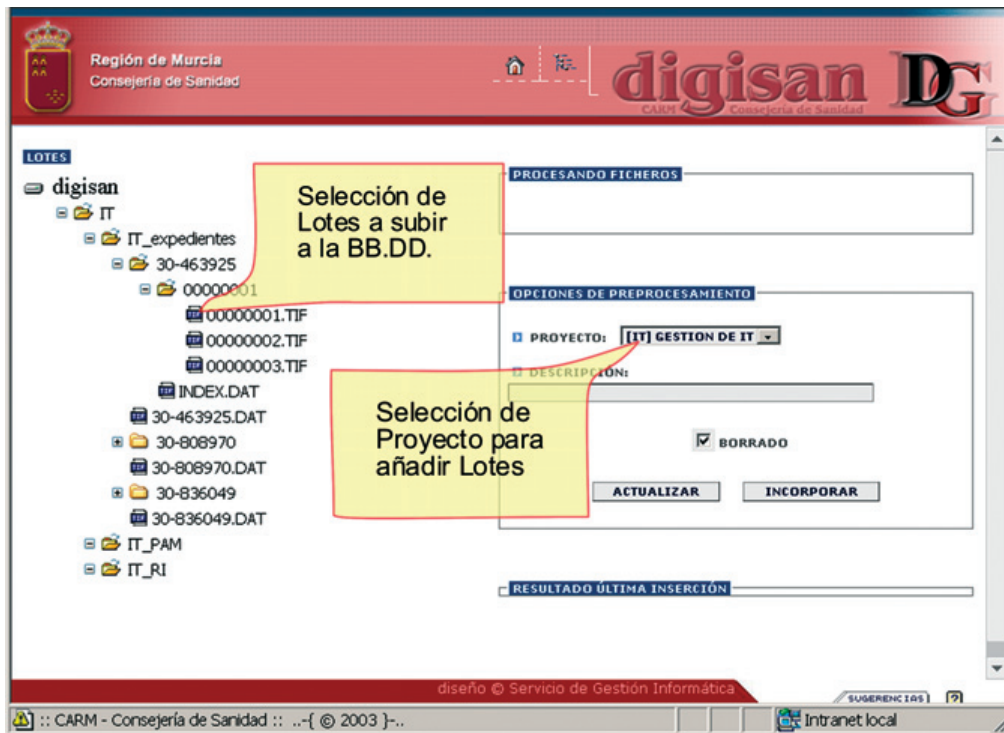
## CARGA O PREINDEXACIÓN

Para la carga de los documentos digitalizados, se procede a la identificación del usuario en la base de datos, el cual debe tener el correspondiente rol de acceso.





A continuación, se seleccionan los lotes del proyecto o unidad de gestión que se quieren cargar: aquellos que están pendientes de indexar. La selección se hace a partir de la estructura de archivos y directorios creados en la fase de escaneo.



## INDEXACIÓN

Una vez cargados, los lotes y sus elementos se visualizan en la ventana de indexación, mostrando detalles de la carga e incluso las imágenes de los elementos.

A continuación, en la zona de búsqueda de expedientes se localiza al expediente que corresponde, utilizando como criterios apellidos, nombre, identificador de persona o identificador de expediente. Se dispone de una búsqueda *avanzada* que nos ofrece más criterios para buscar: fecha de nacimiento, teléfono, etc.

Una vez seleccionado el expediente, se elige el tipo de documento de que se trate, con lo que se mostrarán los atributos asociados que el usuario deberá introducir, los cuales se extraen de la imagen del documento que se visualiza en pantalla. Estos atributos los usarán posteriormente los gestores de la aplicación de IT.

Para terminar, se asocian los elementos en el orden correspondiente al eDocumento que se está creando.



The screenshot displays the DIGISAN web application interface. On the left, there is a search section with fields for 'Nombre', 'Apellido', 'Person', and 'Expediente'. Below this is a table of 'Expedientes seleccionados' with columns for 'ID PERSONA', 'DATOS PERSONALES', 'NSS', and 'FECHA NACIMIENTO'. A 'Gestión de documentos digitalizados del Expediente' section shows a table of documents with columns for 'FECHA DOC.', 'TIPO', and 'NOTAS'. A 'Visualizador de Imágenes TIFF' section displays a scanned document with handwritten notes. Red callout boxes highlight these features: 'Lotes pendientes de indexar.', 'Busqueda de Personas y/o Expedientes', 'Gestión de documentos digitalizados del Expediente.', 'Atributos de los documentos.', and 'Visualizador de Imágenes TIFF'.

## CONSULTA

Se dispone de una ventana de consulta, para que los usuarios puedan consultar los eDocumentos asociados a un expediente, del mismo modo que se hace en la fase de indexación. De esta manera, los inspectores de IT consultan los documentos desde su ordenador (realmente desde cualquier ordenador, pero con sus claves de acceso) sin necesidad de acceder al expediente físico original.

The screenshot displays the DIGISAN web application interface. On the left, there is a search section with fields for 'Nombre', 'Apellido', 'Person', and 'Expediente'. Below this is a table of 'Expedientes seleccionados' with columns for 'ID PERSONA', 'DATOS PERSONALES', 'NSS', and 'FECHA NACIMIENTO'. A 'Gestión de documentos digitalizados del Expediente' section shows a table of documents with columns for 'FECHA DOC.', 'TIPO', and 'NOTAS'. A 'Visualizador de Imágenes TIFF' section displays a scanned document with handwritten notes. Red callout boxes highlight these features: 'Busqueda de Personas y/o Expedientes', 'Gestión de documentos digitalizados del Expediente.', 'Atributos de los documentos.', and 'Visualizador de Imágenes TIFF'.

