

# Gestión Documental, Tecnologías ECM y OpenProdoc

U.C.M. - Facultad de Ciencias de la Documentación

19 / 04 /2017

Joaquín Hierro



#### Índice

- Introducción
- Gestión documental hoy
  - Datos comparativos
  - Ejemplos de proyectos
  - Documentos como núcleo de los procesos
- Tecnologías ECM
  - Мара ЕСМ
  - Gestores Documentales
  - Clasificación automática
- OpenProdoc
  - Construcción
  - Funciones
  - Funciones adicionales (tesauros, referencias)
    Seguridad
  - Demostración



#### Introducción

- Este documento repasa de forma muy resumida el estado actual de la tecnología y proyectos de gestión documental, revisa las principales funciones y productos del ámbito ECM y muestra las características principales del gestor documental OpenProdoc.
- OpenProdoc es una herramienta de ECM, un gestor documental desarrollado bajo licencia de Código Abierto (Open Source).
- Un gestor documental es una herramienta muy sofisticada, por lo que esta presentación es solo una breve introducción a algunos aspectos de los gestores documentales y de OpenProdoc en particular.



### Índice

- Introducción
- Gestión documental hoy
  - Datos comparativos
  - Ejemplos de proyectos
  - Documentos como núcleo de los procesos
- Tecnologías ECM
  - Мара ЕСМ
  - Gestores Documentales
  - Clasificación automática
- OpenProdoc
  - Construcción
  - Funciones
  - Funciones adicionales (tesauros, referencias)
  - Seguridad
  - Demostración



## Gestión documental hoy

- Para aclarar el ámbito del documento, hay que resaltar que se habla exclusivamente de "Gestión Documental" referida a "Gestión Electrónica de Documentos" (GED), excluyendo otras tecnologías como S.I.G.B. (Sistemas Integrados de Gestión Bibliotecaria) o Sistemas de Archivos.
- Para evaluar qué se está haciendo hoy en día en países como España desde un punto de vista documental es importante fijarse en las cifras y la complejidad de los proyectos, ya que esto determina el uso de ciertas herramientas y sobre todo condiciona la necesidad de automatización.



## Gestión documental hoy: Datos comparativos (I)

- Existen varias formas de comparar entre la G.E.D. y otros ámbitos como S.I.G.B. o S.A.
- Un primer enfoque puede ser desde el punto de vista del número de instalaciones

Según el I.N.E.:

Bibliotecas: 6.800

Empresas: 3.200.000

\* Autónomos: 3.100.000

 Cualquier empresa o autónomo necesita gestionar sus facturas, contratos, proyectos, nóminas, etc.

X 1000

 Adicionalmente, las empresas, a partir de un tamaño, pueden requerir varias instalaciones.



## Gestión documental hoy: Datos comparativos (II)

- Puede compararse también por número de documentos manejados:
  - B.N.E.: 28 millones de ejemplares, 4 Millones catálogo automatizado.
  - Número de líneas de móvil: 50 Millones → 50 M expedientes, 50\*12 = 600M facturas/año \* 6 años = 3.600 M facturas almacenadas (sin contar los cambios de operadora)
  - Hipotecas firmadas/mes = ± 33.000 \* 12 = ± 400.000
    Hip./año con una duración de unos 25 años → 10 M expedientes vivos con unos 200 M de documentos.
  - Contratos de trabajo anuales = ± 6 M → 50 M docs.



## Gestión documental hoy: Datos comparativos (III)

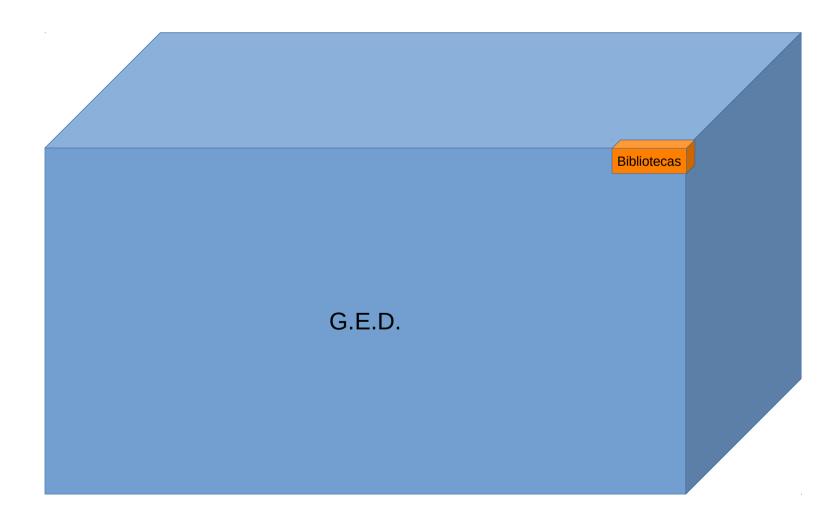
Finalmente podría compararse por complejidad y diversidad:

| Bibliotecas, Archivos<br>Tradicionales               | G.E.D.  |
|--|---|
| Metadatos normalizados (IBSD, MARC, ISAD, METS, NTI) | Metadatos no normalizados                     |
| Independencia de otros procesos                      | Integración con otros procesos                |
| Sistemas tecnológicamente aislados                   | Sistemas tecnológicamente integrados          |
| Seguridad y perfiles limitados                       | Seguridad y perfiles complejos                |
| Usuarios Limitados                                   | Todos los ciudadanos en múltiples sistemas    |
| Acceso controlado                                    | Autoservicio de usuarios.                     |
| Descripción a nivel de libro o expediente            | Descripción a nivel de expediente Y documento |



## Gestión documental hoy: Datos comparativos (IV)

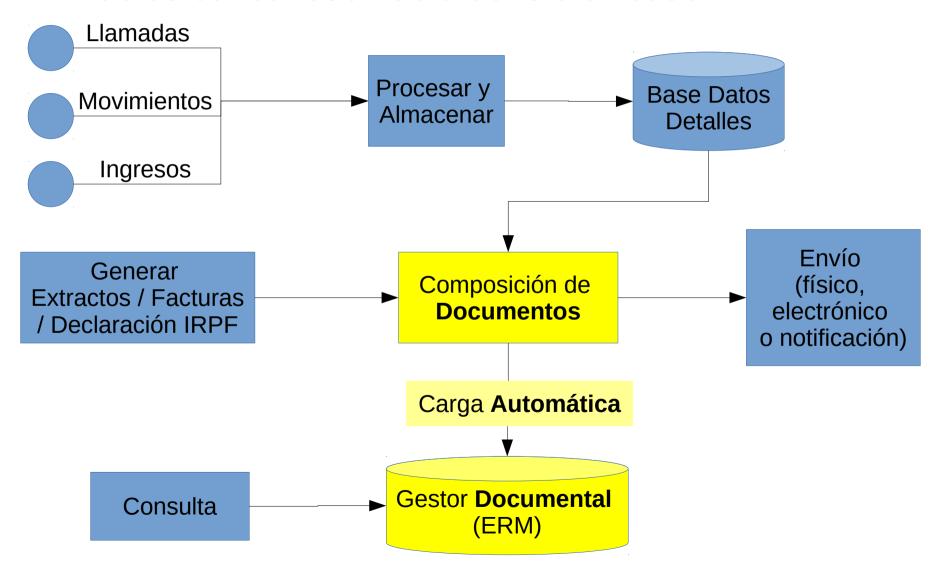
 Si combinamos datos cuantitativos (instalaciones y volúmenes de documentos y usuarios) y complejidad, el orden de magnitud podría expresarse como:





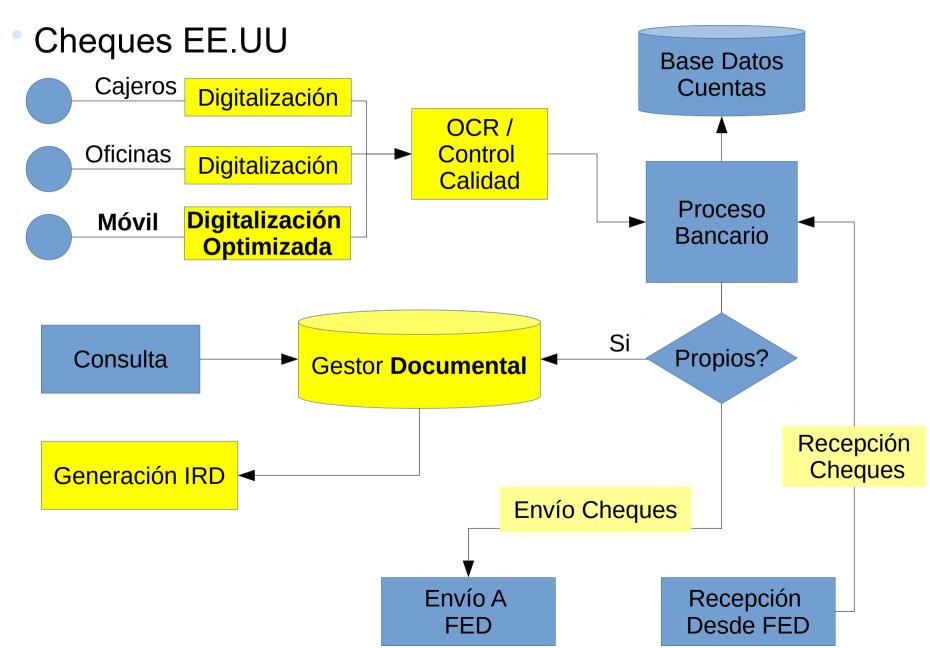
## Gestión documental hoy. Ejemplos de proyectos (I)

Extractos bancarios / Facturas telefónicas / IRPF:



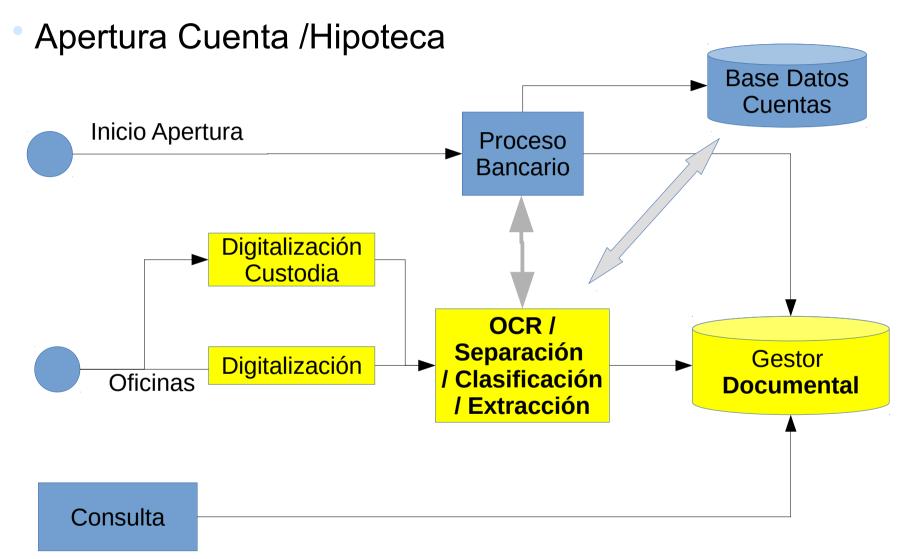


# Gestión documental hoy. Ejemplos de proyectos (II)





### Gestión documental hoy. Ejemplos de proyectos (III)





## Gestión documental hoy: Documentos como núcleo

- Como puede verse, los documentos son el núcleo o detonante de la mayoría de los procesos.
- De hecho puede considerarse que lo han sido siempre pero cada vez, debido a la normalización, legislación y sobre todo a la necesidad de acelerar procesos, la importancia es mayor.
- Dada la cantidad documentos y variedad de formatos (Ej.: la mayor parte de los contratos con operadoras de telefonía son un documento de audio) se hace imprescindible sistemas para automatizar la clasificación y extracción de la información recibida, así como para su almacenamiento y consulta.
- A continuación se repasa algunas tecnologías.



### Índice

- Introducción
- Gestión documental hoy
  - Datos comparativos
  - Ejemplos de proyectos
  - Documentos como núcleo de los procesos

#### Tecnologías ECM

- Мара ЕСМ
- Gestores Documentales
- Clasificación automática

#### OpenProdoc

- Construcción
- Funciones
- Funciones adicionales (tesauros, referencias)
  Seguridad
- Demostración

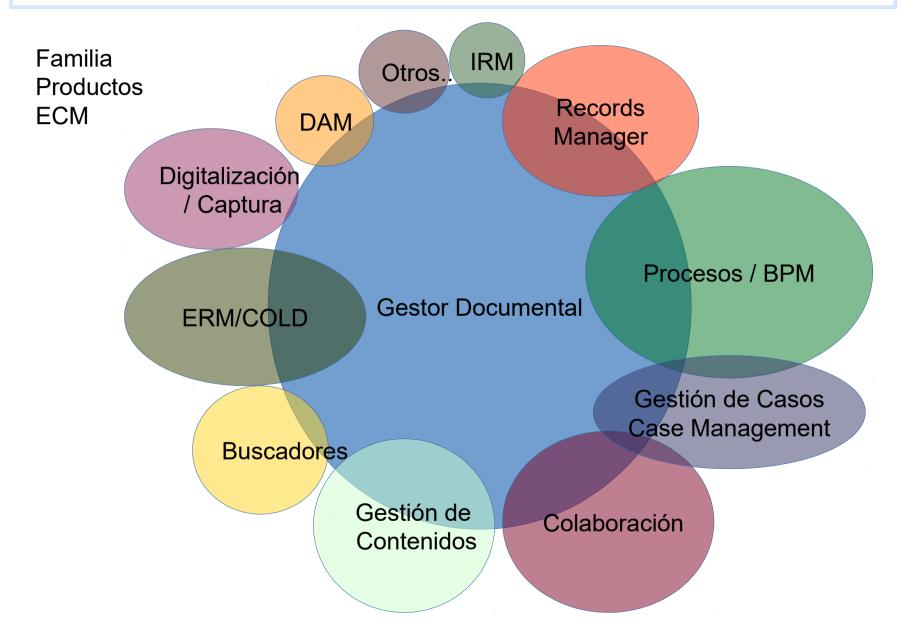


# Tecnologías ECM

- Según AIIM (y Wikipedia):
  - Enterprise Content Management (ECM) son las estrategias, métodos y herramientas utilizadas para capturar, gestionar, almacenar, preservar y entregar contenido y documentos relacionados con los procesos organizativos. Herramientas y estrategias de ECM permiten la gestión de la información no estructurada de una organización, donde quiera que exista esa información.
- ECM cubriría herramientas y tecnologías como:
  - Gestores documentales
  - Gestión de contenidos
  - Herramientas de Colaboración (Wiki, blog, mensajería)
  - Gestión de Registros (Record Management)
  - Captura y digitalización
  - Gestión de Procesos (WorkFlow / BPM)
  - Búsqueda
  - DAM (Digital Assets Management)



## Tecnologías ECM: Mapa ECM



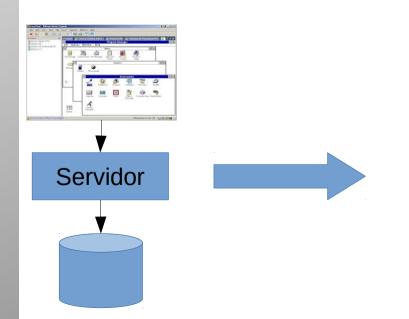


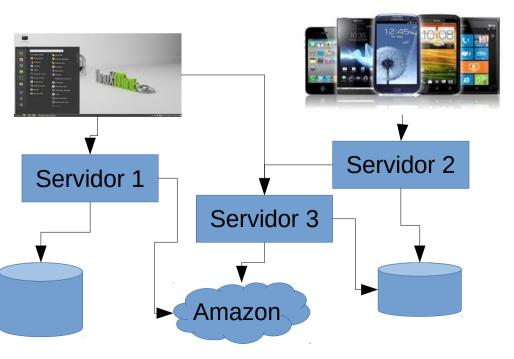
# Tecnologías ECM

Arquitectura









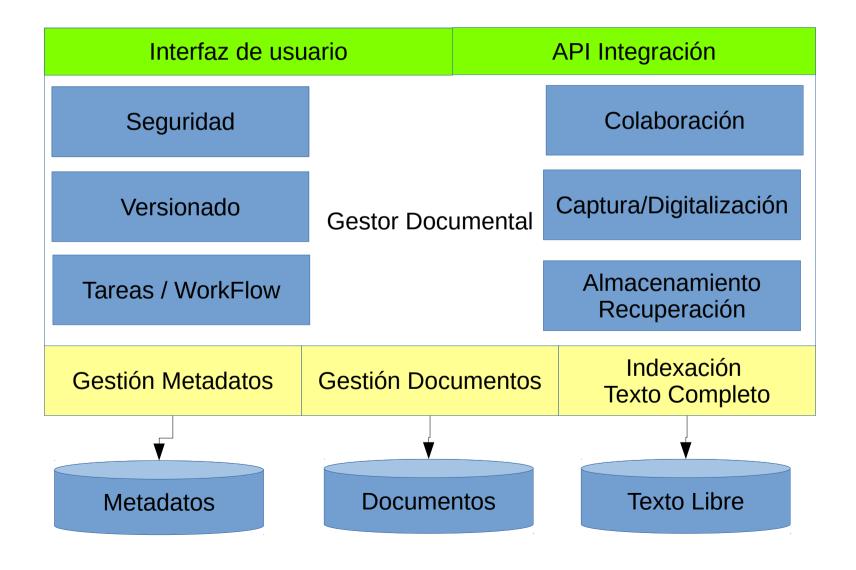


### Tecnologías ECM: Gestores Documentales (I)

- Un Gestor Documental (o D.M.S.) es un programa informático orientado al manejo de documentos de forma compartida entre múltiples usuarios.
- Por manejo debe entenderse el almacenar, buscar y recuperar documentos, asignarles metadatos (fichas descriptivas) o tipologías para poder identificarlos, agruparlos en expedientes o estructuras jerárquicas, asignarles seguridad para poder limitar qué puede hacer cada usuario o grupo de usuarios y definir un ciclo de vida para que, por ejemplo, los documentos se borren automáticamente al cabo de los años.
- Generalmente debe parametrizarse para poder utilizarlo, definiendo las tipologías documentales, los tipos de expedientes, el modelo de seguridad y los procesos por los que pasaran los distintos elementos.



#### Tecnologías ECM: Gestores Documentales (II)





### Tecnologías ECM: Gestores Documentales (III)

- Dada su función como núcleo de muchos procesos, tiene diversos requerimientos básicos:
  - 1) Flexibilidad: Poder ajustar su modelo documental a las necesidades, definiendo tipologías de documentos y expedientes, imponiendo restricciones a los metadatos y creando estructuras jerárquicas de información
  - 2) Seguridad: Ajustar la seguridad a las necesidades del proceso (tanto en cuanto a los roles de usuarios, a los permisos de acceso a cada objeto y al acceso no autorizado)
  - 3) Escalabilidad: poder crecer cuando aumenta la carga de trabajo para mantener la velocidad y servicio, esto puede hacerse tanto horizontal (aumentando servidores) como verticalmente (aumentando memoria o CPU).
  - 4) Integración: Integrarse con otros sistemas, por medio de APIs (Application Program Interfaces) y servicios de forma que otras aplicaciones pueda invocar funciones.



### Tecnologías ECM: Clasificación automática (I)

- Dados los enormes volúmenes de documentos que se manejan es imprescindible contar con sistemas que separen los documentos, los clasifiquen y extraigan los metadatos adecuados, todo ello de forma automática.
- Si además los documentos son imágenes, deberá antes realizarse un reconocimiento óptico de caracteres (OCR).
- Esto se hace no solo con documentos estructurados (formularios) sino con documentos semi-estructurados (nominas, facturas) y documentos no estructurados (poderes, documentos notariales, correspondencia).
- El reconocimiento puede hacerse sobre caracteres impresos, manuscrito, marcas y códigos de barras 1D y 2D.

Ana Domínguez















UPC-A





#### Tecnologías ECM: Clasificación automática (II)

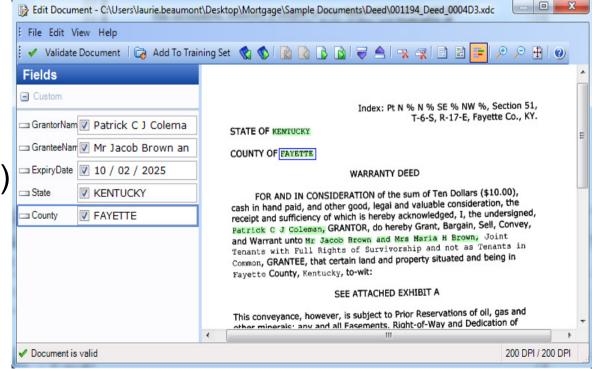
 El entrenamiento en muchos casos se realiza por medio de muestras y ejemplos y el sistema aprende "solo" aunque suele ser necesario ademas complementar con parametrización y desarrollo.

El sistema devuelve siempre el tipo documental obtenido tras el análisis y los metadatos extraídos, siempre con un sixulado en figura en la managa de farma en construcción de la farma en la farma en

nivel de confianza en la respuesta, de forma que según el

documento y proceso puede decidirse aceptarlo o pasar a revisión manual (o incluso dos revisiones)

 Puede obtenerse automatismos del 80%





### Índice

- Introducción
- Gestión documental hoy
  - Datos comparativos
  - Ejemplos de proyectos
  - Documentos como núcleo de los procesos
- Tecnologías ECM
  - Мара ЕСМ
  - Gestores Documentales
  - Clasificación automática

#### OpenProdoc

- Construcción
- Funciones
- Funciones adicionales (tesauros, referencias)
  Seguridad
- Demostración



## Construcción (Orígenes)

- A partir de la experiencia utilizando y administrando los principales productos del mercado, así como realizando procesos de análisis y selección de soft. documental, me planteé crear un gestor documental con las características que consideré más interesantes de ellos.
- Hace unos años no había apenas gestores Open Source ni tenían demasiada calidad.
- El crear un gestor permite aprender mucho sobre el funcionamiento interno y los posibles problemas (de diseño, portabilidad, rendimiento,..), lo que permite analizar y utilizar otros productos con mucha mayor base.
- Por ultimo, como reto personal, ya que crear un producto de este tipo con la calidad, rendimiento y funciones adecuadas es un proyecto muy complejo.

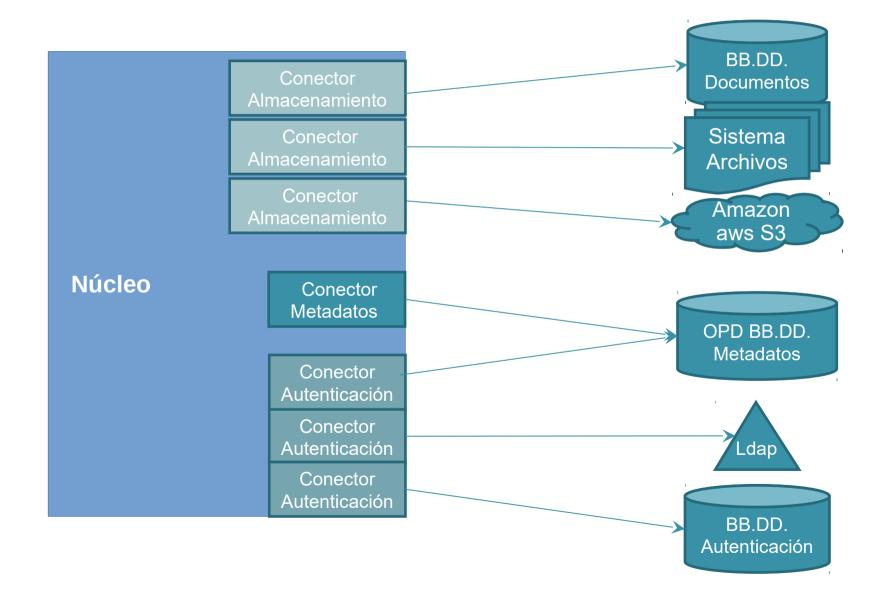


## Construcción (Diseño)

- Para su desarrollo era condición de partida utilizar un lenguaje orientado a objetos por la facilidad de evolución y mantenimiento que aporta. Se ha utilizado Java debido a las ventajas de portabilidad y estandarización de la arquitectura J2EE.
- Otro requisito importante era la independencia de plataforma, de forma que pudiera utilizarse con cualquier base de datos, servidor de aplicaciones y sistema operativo, para ello se ha cuidado el desarrollo y probado en diversos entornos.
- El modelado interno de los metadatos en BBDD ha sido uno de los puntos en los que he empleado más tiempo para asegurar flexibilidad y rendimiento.
- OpenProdoc se basa en conectores para acceder a otros sistemas, de forma que pueda expandirse.

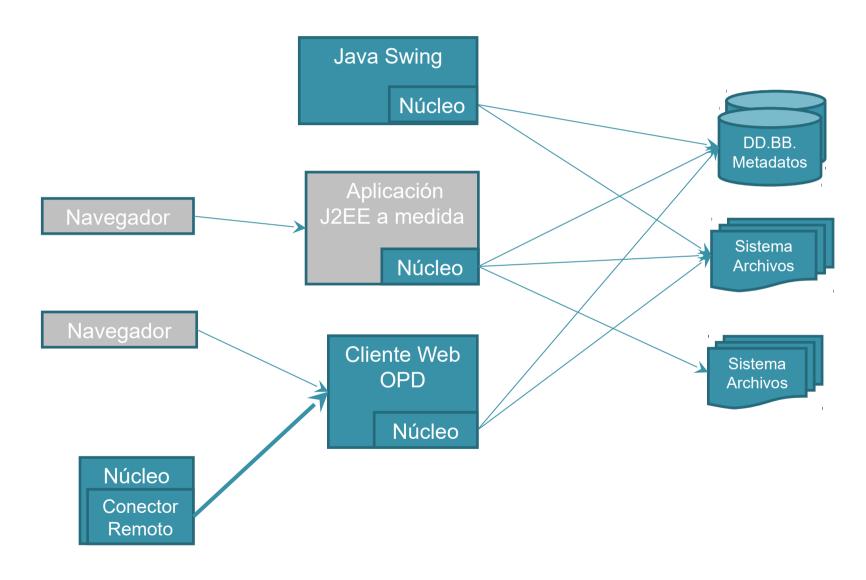


### Construcción (Arquitectura)





## Construcción (Arquitectura)





## Funciones (I)

- Openprodoc incluye las funciones habituales de un gestor documental:
  - Definición de tipos documentales y de contenedores (carpetas/expedientes).
  - Organización de la información en estructuras jerárquicas.
  - Almacenamiento de los metadatos y archivos en contenedores.
  - Gestión de versiones
  - Gestión de usuarios y grupos
  - Seguridad basada en ACL (Listas de Control de Acceso)
  - Búsquedas por los metadatos
  - Gestión del ciclo de vida
  - Posibilidades de integración con otros sistemas.
- A ello añade:
  - Gestión de múltiples tesauros multilingües
  - Gestión de referencias
  - Funciones (limitadas) de DAM y DSI.
  - Búsqueda por texto libre.



## Funciones (II)

- OpenProdoc permite la definición de tipos documentales y contenedores orientada a objetos.
- Cada tipo documental se define como subtipo de uno ya existente. Por defecto existe un tipo básico de carpeta y un tipo básico de documento con un mínimo de metadatos. Todos los tipos documentales creados serán subtipos del tipo básico o de otro subtipo.
- Los metadatos y comportamiento se heredan; es decir un tipo documental tiene la suma de todos los atributos de cada uno de sus antecesores.
- A cada tipo documental puede añadírsele metadatos, indicando el nombre y tipo (cadena, fecha, entero, booleano, controlado por un tesauro), si es obligatorio, si es multivaluado y si es un valor único.



## Funciones (III)

- La orientación a objetos tiene varias ventajas:
  - Permite una evolución gradual del modelo. Es muy complejo definir una estructura de tipos documentales para una organización grande con diversos departamentos. Esperar a tener un cuadro de clasificación completo retrasaría el arranque mucho tiempo. Adicionalmente es habitual una evolución del modelo por causas organizativas o legales.
  - Es más potente, ya que puede buscarse por un tipo documental o por un tipo y sus subtipos (Ej. Buscar por unos criterios en "Informes" o en "Informes" y todos sus subtipos). Esto aplica a también a otras operaciones definidas.
  - Es más sencilla y evita tener que definir una "estructura plana", se realiza un análisis gradual que genera una estructura jerárquica. Los metadatos o propiedades definidas para un tipo se heredan para los subtipos.
  - Es una representación más "real", ya que las definiciones y tipologías documentales no son totalmente independientes.



### Funciones adicionales (Tesauros I)

- Puede crearse múltiples tesauros multilingües estructurados de acuerdo a las necesidades.
- Cada "tesauro" puede utilizarse como un tesauro completo (NT, BT, RT, UF) o simplemente como una lista de materias o una lista de términos, que puede ser jerárquica.
- Los metadatos de los tipos documentales pueden asociarse a tesauros concretos, de forma que se elija el valor del metadato dentro de los disponibles en el tesauro y se controle la integridad. Por ejemplo puede así definirse un metadato "País" que se valide contra una lista de países o un metadato "Tema del informe" que se valide contra un tesauro de materias.



### Funciones adicionales (Tesauros II)

- Para importar y exportar se utiliza el estándar SKOS-RDF.
- OpenProdoc no necesita crear microtesauros. Las búsquedas pueden acotarse a partir de un término a todos los términos específicos del mismo, restringiendo la búsqueda a una rama del árbol.
- OPD no maneja tesauros poli-jerárquicos, de forma que la importación de un tesauro que incluya relaciones de ese tipo generará un aviso y guardará una sola relación.



# Funciones adicionales (Referencias Bibliográficas)

- OpenProdoc puede manejar referencias bibliográficas de forma sencilla.
- No es un gestor especializado y tan completo como Zotero o Mendeley pero permite incluir las referencias combinadas con expedientes y documentos almacenados en el repositorio.
- Admite el estándar de referencias RIS, tanto de entrada como de salida.
- Estas referencias pueden generalizarse y manejarse como cualquier tipo documental, definiendo los metadatos necesarios (que no tiene que ser RIS) y referenciando páginas web o cualquier contenido remoto.



### Seguridad (Perfiles)

- La seguridad en OpenProdoc se cubre de varias formas.
- La primera y básica son los perfiles. Los perfiles indican los tipos de operaciones que puede realizar un usuario (por ejemplo, administrar usuarios, dar de alta documentos, administrar definiciones de documentos, administrar grupos, administrar los tesauros,..)
- Puede crearse perfiles con la combinaciones de permisos que se desee, sin necesidad de un único administrador "todopoderoso".
- Puede crearse distintos tipos/perfiles de administrador, por ejemplo Administrador de Seguridad (encargado de usuarios y grupos), Administrador Documental (Encargado de definiciones de documentos y tesauros), Administrador tecnológico,...



### Seguridad (ACL I)

- En segundo lugar, y más importante está la seguridad aplicable a cada objeto (carpetas/expedientes o documentos).
- En OpenProdoc se utiliza seguridad basada en ACL (Listas de Control de Acceso) como es habitual en los principales gestores.
- Un ACL contiene un nombre y una lista de usuarios y grupos con los permisos asignados a ellos.
- Ej:
  - ACL: DocEconomicos
    - GrupoContabilidad: Lectura-Escritura-Borrado
    - GrupoDireccion: Lectura
- Si un grupo no aparece no tiene permisos, y no podrá acceder a los elementos con ese ACL.



### Seguridad (ACL II)

- ¿Porqué un modelo de ACLs y no un sistema más elemental?
- Por varios motivos.
  - Los sistemas basados en permisos de lectura-escritura o niveles de acceso (0-9) no reflejan la complejidad de un gestor ni una empresa. Un usuario puede poder escribir en un expediente y solo leer en otro dependiendo del área o tipo de expediente.
  - Las listas de usuarios asignados a cada elemento son inmanejables y si hay cambios obligan a cambiar miles de documentos o expedientes. Esos cambios, o deben hacerse manualmente o de forma jerárquica, lo que obliga a una estructura irreal.
  - Ademas el documento o expediente puede sufrir cambios durante su vida y tramitación

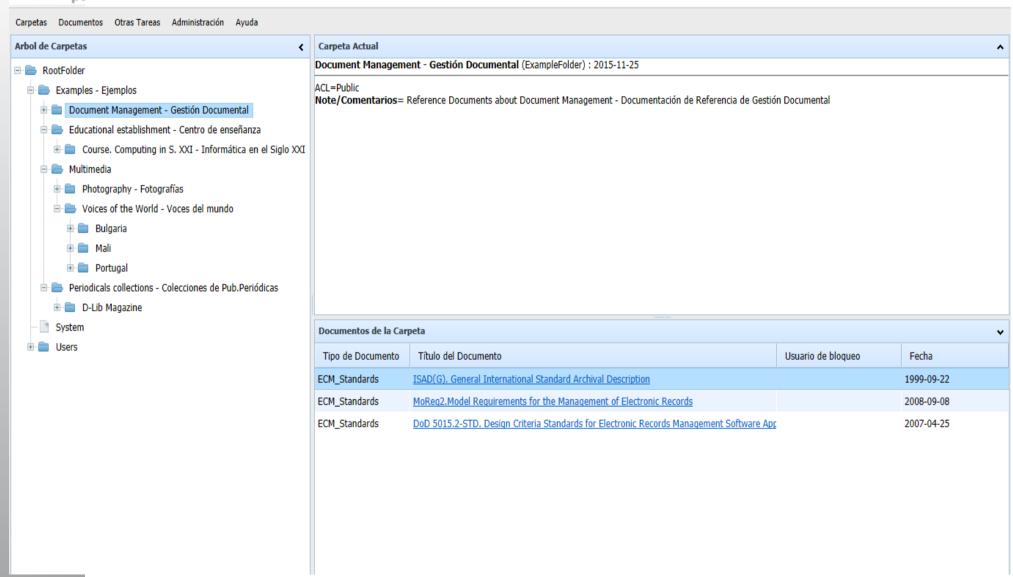


## Seguridad (ACL III)

- Con los ACL, podemos cambiar la definición de seguridad y al momento millones de documentos y expedientes cambian sus permisos.
- Los ACL reflejan la seguridad del elemento y muestran el sentido mejor que una larga lista de nombres. Ej:
  - "DocumentosConfidenciales",
  - "ExpedientesTramitados",
  - "PendientesRevisionDireccion"
- Ademas puede cambiarse en los distintos estados de la vida. Ej
  - "Cuando un expediente se recibe su ACL será
    'RegEntradaDepartamentoX' tras su revisión se asignará a los departamentos A o B con ACL 'ExpedientesA' o 'ExpedientesB' "

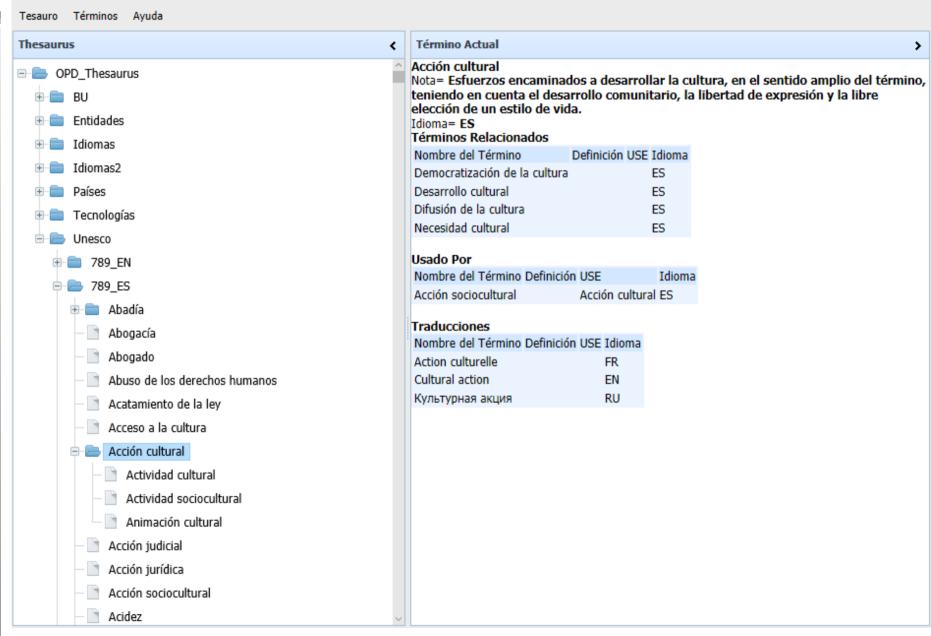


#### Demostración

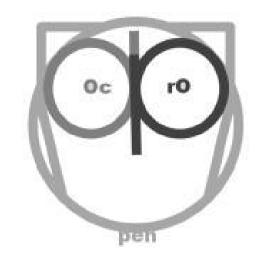




#### Demostración







#### **OpenProdoc**

Joaquín Hierro openprodoc@gmail.com

http://jhierrot.github.io/openprodoc/index\_ES.html

http://www.dokumentalistas.com/articulos/openprodoc-creando-un-gestor-documental-en-solitario/

http://www.dokumentalistas.com/articulos/openprodoc-superando-los-silos-de-informacion/

http://nosturi.es/2017/02/14/disena-modelo-documental-openprodoc/

http://www.biblogtecarios.es/author/joaquinhierro/