

Software de fuentes abiertas en la Administración electrónica

Archivo electrónico de las Administraciones Públicas

Elaborado por:

Centro Nacional de Referencia de Aplicación de las TIC basadas en fuentes abiertas (CENATIC).

Atos Consulting

Con la colaboración de:

Ministerio de Industria, Turismo y Comercio, Ministerio de la Presidencia, Ministerio de Economía y Hacienda, Ministerio de Defensa, Red.es, Dirección General de Tráfico, Generalitat de Catalunya, Gobierno de Aragón, Gobierno de Cantabria, Junta de Extremadura, Junta de Andalucía, Principado de Asturias, Diputación de Cáceres, Diputación de Teruel, Ayuntamiento de Almendralejo, Ayuntamiento de Mérida, Ayuntamiento de Zaragoza

Edita:

© 2009 CENATIC

(Centro Nacional de Referencia de Aplicación de las TIC basadas en fuentes abiertas).

C/ Vistahermosa, 1-3ª planta
06200 Almendralejo (Badajoz)

Deposito legal: BA-322-09

Diseño y producción:

Comfot, s.l.u.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio, siempre que se mantenga el reconocimiento de sus autores y no se realice ninguna modificación de las mismas.

Este informe está disponible en el sitio web de Cenatic
<http://www.cenatic.es>

Las opiniones expresadas en esta publicación pertenecen a los distintos autores de los textos incluidos y no reflejan necesariamente la opinión de Cenatic, o de las Administraciones Públicas.

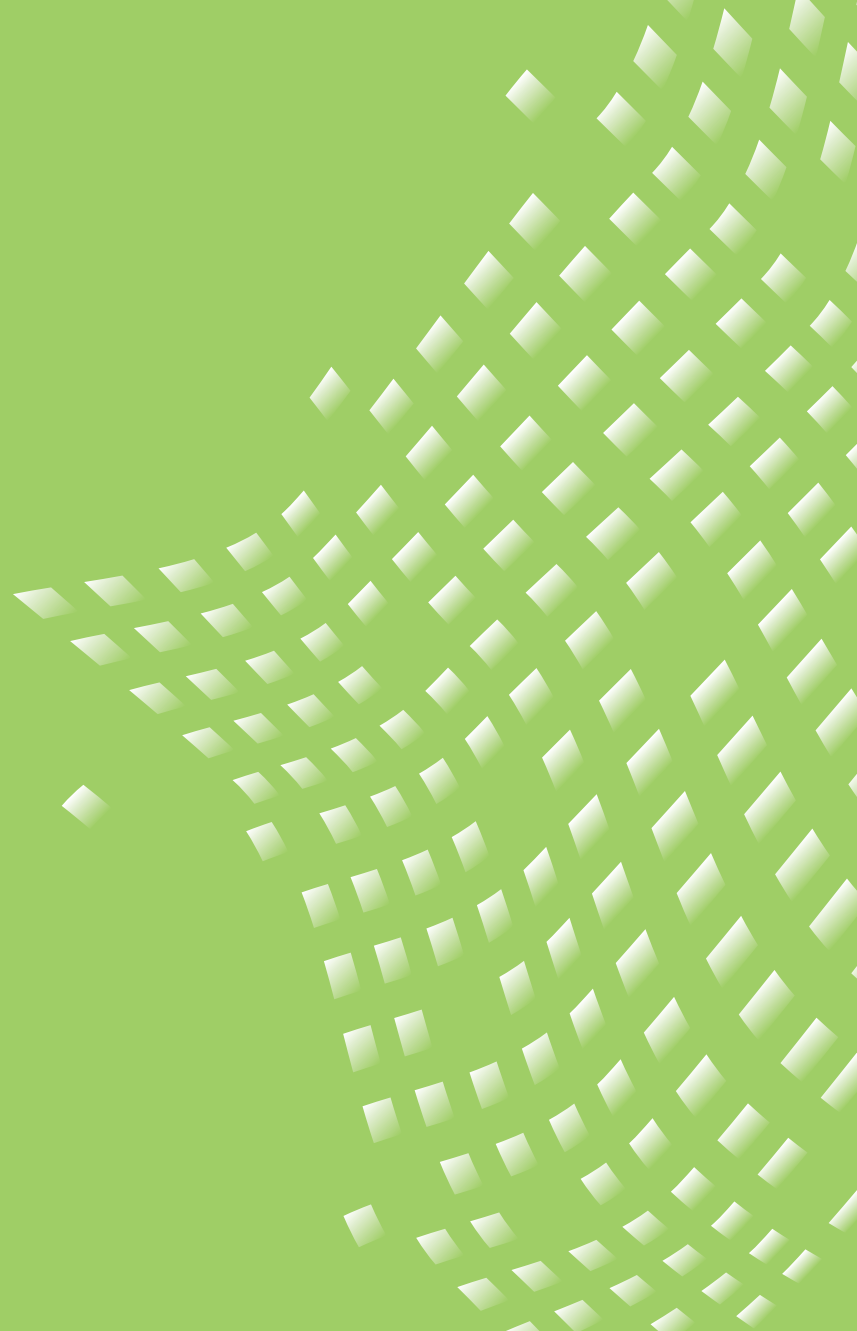
Índice

1. Resumen ejecutivo	5
<hr/>	
2. Introducción	9
2.1. Archivo de documentos electrónicos	11
2.2. Problemática relativa al Archivo Electrónico	12
2.3. Contexto actual del Archivo Electrónico	14
2.4. Expediente electrónico	16
2.5. Valor de los documentos	17
2.6. Documentación susceptible de ser archivada	18
2.7. Metadatos	19
2.8. Diferencias entre un Gestor Documental y un Archivo Electrónico	20
2.9. Ciclo de vida de los documentos	22
<hr/>	
3. Modelo conceptual del Archivo Electrónico	23
3.1. Descripción del Modelo Conceptual	25
3.2. Aspectos clave del modelo	33
<hr/>	

4. Aplicación del modelo conceptual a las AAPP	39
4.1. Flexibilidad del modelo de Archivo Electrónico propuesto	41
4.2. Escenario 1	41
4.3. Escenario 2	43
4.4. Normas internacionales	44
5. Definición funcional del Archivo Electrónico	47
5.1. El modelo OAIS	49
5.2. Adecuación del modelo OAIS	51
6. Implementación tecnológica del Modelo Conceptual	55
6.1. Alcance de la implementación	57
6.2. Contexto tecnológico	57
6.3. Soluciones de repositorio existentes basadas en software open-source	58
6.4. Diseño de la implantación piloto	71
7. Mapa de componentes	87
7.1. Identificación de categorías de componentes	89
7.2. Mapeo funcionalidad - componente	98
8. Glosario	107

01.

Resumen ejecutivo



1. Resumen ejecutivo

El documento se enmarca dentro de la iniciativa impulsada por el **Centro Nacional de Referencia de Aplicación de las TIC basadas en fuentes abiertas (CENATIC)** para analizar la problemática actual del archivo de documentos electrónicos en las Administraciones Públicas.

En el mismo se presentan las conclusiones de las actividades realizadas durante el proyecto, que se han orientado hacia dos líneas de trabajo:

- Definir un **modelo conceptual de Archivo Electrónico** adecuado a las necesidades de las Administraciones Públicas.
- Analizar la **viabilidad de una implantación tecnológica del modelo propuesto** en el ámbito del software de fuentes abiertas.

La definición del modelo conceptual se apoya en un análisis previo de la normativa existente, que ha permitido determinar las directrices que deberán regir la operativa del Archivo Electrónico. Sobre ellas, se ha diseñado una propuesta **válida y suficientemente flexible para dar respuesta a los requerimientos de**

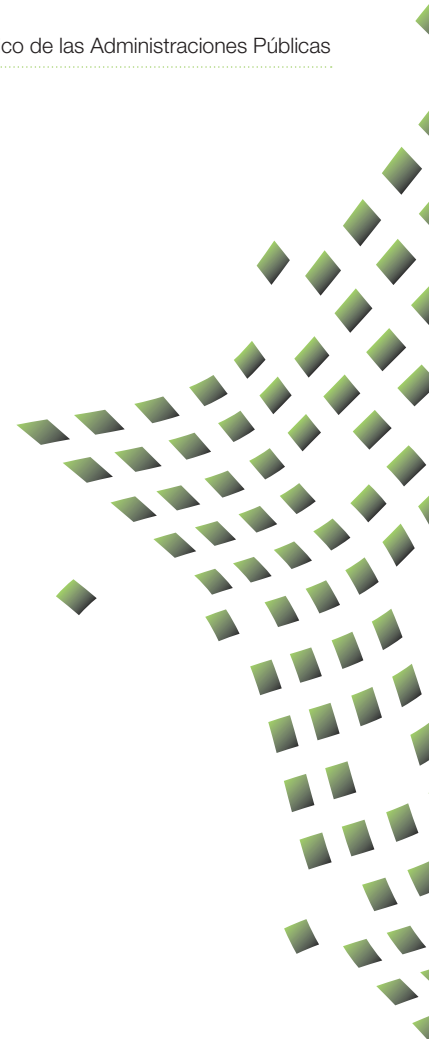
los distintos tipos de Administraciones Públicas existentes en España.

Este modelo segmenta el Archivo Electrónico en distintos niveles para favorecer una centralización de la información y facilitar su alineación con la organización existente.

Cada uno de estos niveles actuará como un sistema independiente. Para la definición funcional de los niveles se ha adoptado una división modular, en la que se ha tomado como referencia el modelo OAIS de Archivo Electrónico, aceptado a nivel mundial.

El modelo propuesto ha sido contrastado con representantes de los distintos ámbitos de la Administración Pública, lo que ha permitido perfeccionar la aproximación inicial, ajustándola al entorno y a las necesidades reales de los potenciales usuarios.

Complementariamente, se ha realizado un estudio para **comprobar la viabilidad de la implantación del modelo propuesto** sobre una plataforma tecnológica



formada exclusivamente por componentes de software de fuentes abiertas.

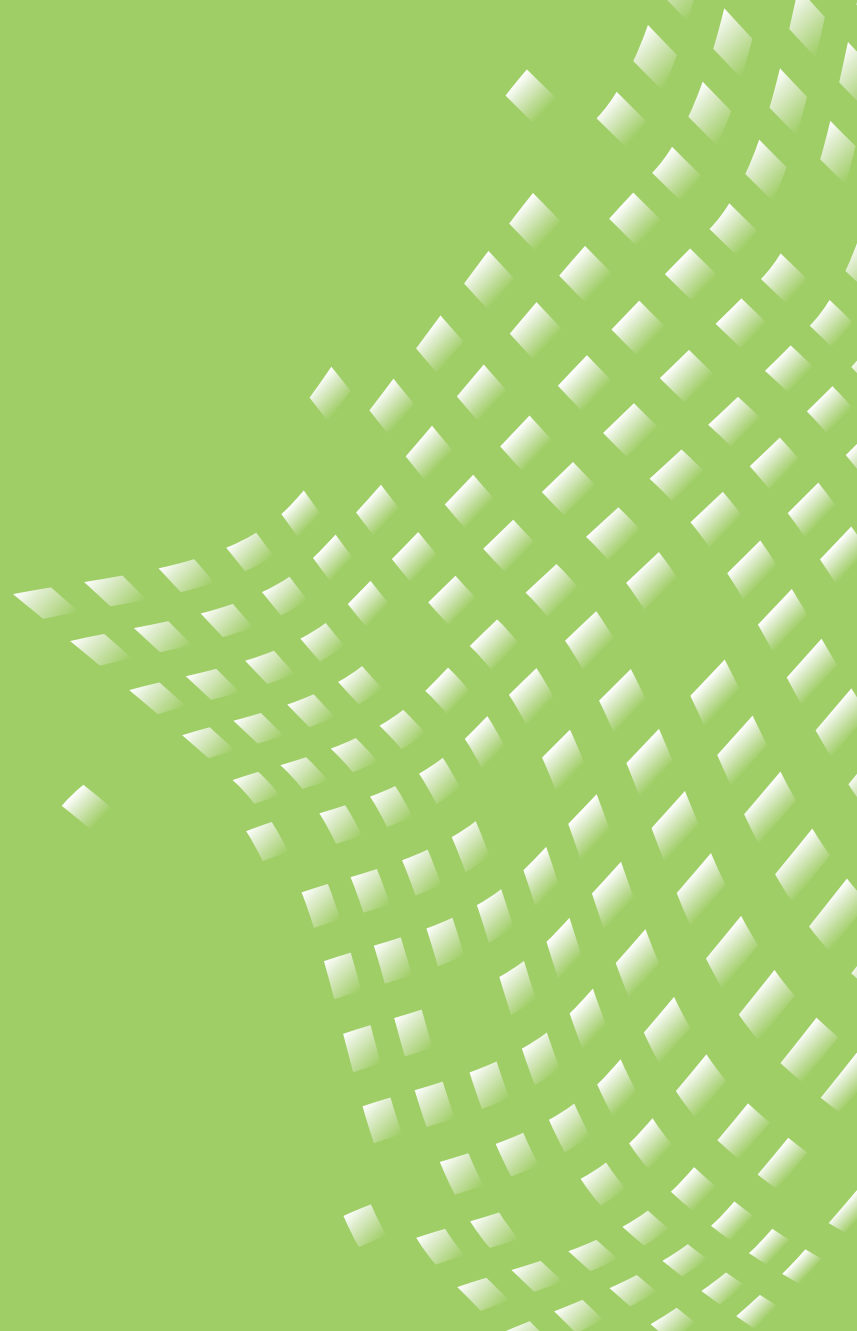
Se han evaluado distintas aplicaciones open-source de repositorio digital, confrontándolas con el modelo conceptual diseñado. Fruto de esta comparación, se han identificado qué aspectos funcionales del modelo no quedarían cubiertos por una aplicación tipo de repositorio digital.

Para solucionar estas carencias, se han buscado aplicaciones y componentes de software de fuentes libres complementarios que puedan ser integrados para conformar una implantación del Archivo Electrónico completa.

Los productos identificados se han recopilado en un **catálogo o mapa de componentes de software de fuentes abiertas.**

02.

Introducción



2. Introducción

2.1. Archivo de documentos electrónicos

El Archivo Electrónico se ocupa del almacenamiento y conservación de documentos generados por medios electrónicos.

Con la progresiva modernización de la Administración Pública, cada vez es más habitual el empleo de sistemas informáticos para la creación y distribución de documentación como parte de la operativa diaria de las organizaciones. Esta informatización está originando una sustitución del soporte papel por otros formatos digitales, lo que supone un cambio en la forma de trabajar de los empleados públicos.

En paralelo se está avanzando en la implantación de mecanismos de relación entre los ciudadanos y las administraciones a través de medios electrónicos, especialmente desde la aprobación de la Ley 11/2007, de Acceso Electrónico de los Ciudadanos a los Servicios Públicos (LAECSP).

Como resultado, durante los próximos años se prevé un importante aumento en el volumen de información

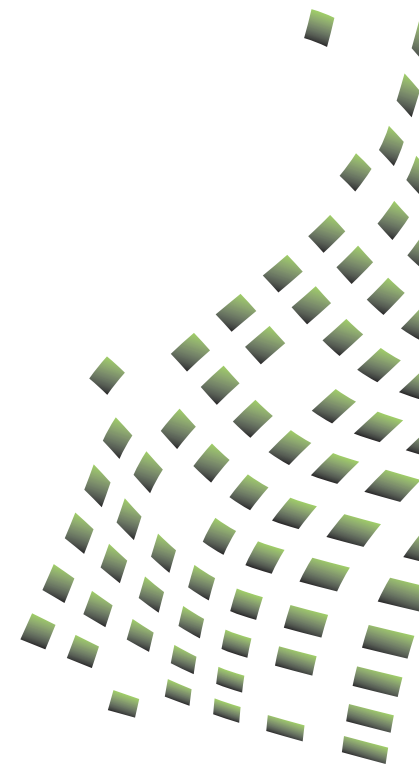
digital gestionada por los organismos públicos, los cuales deberán disponer de sistemas adecuados que permitan almacenarla de una forma ordenada y eficaz.

El archivo de documentación electrónica, como otros aspectos de la Administración Electrónica, viene regulado por la LAECSP, a través de su artículo 31, que indica que:

“Los medios o soportes en que se almacenen documentos, deberán contar con medidas de seguridad que garanticen la integridad, autenticidad, confidencialidad, calidad, protección y conservación de los documentos almacenados. En particular, asegurarán la identificación de los usuarios y el control de accesos, así como el cumplimiento de las garantías previstas en la legislación de protección de datos”.

La LAECSP también define qué se entiende por documento electrónico de archivo, como:

“La información de cualquier naturaleza en forma electrónica, archivada en un soporte electrónico según un formato determinado y susceptible de identificación y tratamiento diferenciado”.



2.2. Problemática relativa al Archivo Electrónico

Aunque el manejo de documentos en formato electrónico tiene ventajas en comparación a sus equivalentes en papel, también es cierto que presenta nuevas problemáticas relativas a su conservación que antes no existían.

2.2.1. Obsolescencia tecnológica

La problemática principal es que el uso de ficheros electrónicos implica la dependencia respecto a una infraestructura tecnológica. Esta infraestructura está formada por las aplicaciones y programas que son capaces de procesarlos y presentar los contenidos a los usuarios, así como por los componentes de hardware sobre los que se ejecutan

Es algo asumido que la tecnología evoluciona con rapidez, lo que da lugar a la renovación constante de productos y aplicaciones por otros más modernos y completos. En la práctica esto obliga a las organizaciones a una continua renovación de sus sistemas informáticos para mantener su infraestructura actualizada.

La obsolescencia tecnológica aparece cuando, fruto de esta evolución, versiones antiguas de los ficheros y

aplicaciones dejan de ser compatibles con la plataforma tecnológica actual.

Estas actualizaciones de la tecnología, aunque frecuentes, no suelen implicar un riesgo para la operativa cotidiana, donde los documentos electrónicos raramente tienen una vigencia superior a un par de años, y los cambios tienden a hacerse progresivamente.

Sin embargo, cuando consideramos la conservación de documentos electrónicos durante largos periodos de tiempo, esto puede convertirse en un problema. Así, puede darse el caso, que el documento que archivamos hoy en un soporte y formato de fichero determinado, no pueda ser recuperado dentro de 10 años, debido a causas como que las aplicaciones ofimáticas no sean capaces de reconocer el formato empleado o simplemente porque el hardware no sea compatible con el soporte empleado.

A modo de ejemplo, basta imaginar un fichero almacenado un disco de 5 ¼, una tecnología muy empleada durante los años 80. En la actualidad, es prácticamente imposible encontrar un ordenador moderno que los admita, por lo que la información que contiene puede darse por perdida.

Además hay que tener en cuenta que los ficheros informáticos no son más que una secuencia de bits (ceros y unos), que el software trata de acuerdo a unas especificaciones concretas, dando lugar a la representación esperada de los contenidos.

Esto implica un riesgo al emplear formatos de fichero propietarios, cuyas especificaciones no hayan sido publicadas por sus creadores, para el archivado a largo plazo. La recuperación de estos ficheros viene determinada por la existencia de soporte por parte de sus creadores. Si la empresa cierra o abandona su desarrollo, nadie más será capaz de desarrollar software que pueda procesar y recuperar estos ficheros.

En el mejor de los casos, es posible que sí exista software compatible en el futuro pero que debamos adquirirlo bajo las condiciones impuestas por el productor, limitando nuestra libertad para elegir con qué herramientas queremos trabajar.

Por ello, la adopción de formatos basados en estándares abiertos y conocidos es un requisito imprescindible a la hora de abordar la implantación de un Archivo Electrónico, ya que el conjunto de especificaciones

que rigen la representación de los contenidos son públicas y accesibles para cualquiera, permitiendo el desarrollo de software adecuado para su manejo en cualquier momento.

2.2.2. Clasificación y recuperación de la información

A medida que el volumen de información almacenada aumenta, la identificación y recuperación de los contenidos se hace más complicada. Esto es común tanto a la documentación en soporte electrónico como a la generada en soportes físicos.

Durante años, los archiveros han establecido mecanismos para el etiquetado y clasificación de la información, dando lugar a los archivos actuales existentes en las Administraciones Públicas. La existencia de cuadros de clasificación, catálogos e índices hace que sea posible encontrar un documento en fondos documentales con miles de volúmenes.

Los sistemas informáticos ofrecen mecanismos útiles para el usuario para la búsqueda y recuperación de ficheros en entornos de almacenamiento. Esto podría llevar a la conclusión de que el uso de métodos archivísticos no sigue siendo necesario.

Sin embargo, estos mecanismos de búsqueda pierden efectividad y requieren de una mayor cantidad de procesamiento a medida que crece la cantidad de ficheros almacenados, por lo que sigue siendo imprescindible la catalogación de la documentación generada.

Los usuarios deben ser conscientes de ello y adoptar una actitud proactiva al respecto, describiendo y etiquetando la información que generan, ya que no resulta viable dejar toda esta responsabilidad en manos de los archiveros.

Por ello, hay que evitar la percepción del Archivo Electrónico como algo relacionado únicamente con la gestión de Sistemas Informáticos y favorecer la participación de los archiveros en el desarrollo e implantación de procesos de conservación de la documentación digital.

2.2.3. Seguridad de la información

La seguridad es otra de las problemáticas respecto a la conservación de documentos electrónicos. A diferencia del papel, un documento electrónico es fácilmente modificable, pudiéndose añadir o eliminar partes del documento sin que se aprecien evidencias de los cambios.

Hay que tener en cuenta que la incorporación de un documento a un procedimiento administrativo le confiere un valor legal, por lo que es necesario garantizar la autenticidad y procedencia del documento. Por ello, es imprescindible establecer los mecanismos adecuados que permitan garantizar la integridad de la información archivada, así como auditar todas las acciones que se han realizado sobre la misma y quién las ha llevado a cabo.

Por otro lado, el uso de la tecnología persigue facilitar la compartición y el acceso a la información por parte de los usuarios. Dado el carácter confidencial de parte de los documentos incluidos en la tramitación administrativa, es imperioso establecer métodos de autenticación y autorización que limite su difusión únicamente a los usuarios afectados por el procedimiento.

2.3. Contexto actual del Archivo Electrónico

La LAECSP ha impulsado la implantación real de la Administración Electrónica en las Administraciones Públicas españolas. La complejidad de estas iniciativas y la limitación de recursos disponibles está favoreciendo que esta implantación se lleve a cabo de forma

progresiva por fases, dándose prioridad a aquellos servicios más visibles por el ciudadano, como el registro electrónico o el inicio de trámites por vía telemática, así como a aquellos sistemas que tienen un mayor impacto en la gestión interna.

En este contexto, el Archivo Electrónico no se está percibiendo como una prioridad a corto plazo, ya que la tecnología actual permite adoptar soluciones de contingencia a bajo coste. Así, es habitual que las Administraciones Públicas habiliten sistemas de almacenamiento donde mantener la información digital generada sin abordar el desarrollo de mecanismos que garanticen su conservación a largo plazo.

Aunque esta aproximación puede ser válida a corto plazo, no es asumible posponer la resolución de este problema indefinidamente, ya que, como se ha comentado anteriormente, es inevitable la aparición de problemas relacionados con la obsolescencia de la tecnología en el futuro, comprometiendo el acceso a los contenidos archivados.

Esta percepción como problema a resolver en el futuro, está retrasando un análisis exhaustivo en las Administraciones Públicas sobre su aplicación. Se ha

constatado que, en general, existe una sensación de indefinición alrededor del Archivo Electrónico, propiciado en parte por una falta de regulación en algunos de los aspectos claves que le afectan, como:

- Criterios a seguir para la conservación de los contenidos.
- Qué información es obligatorio conservar.
- Definición de responsabilidades sobre la preservación de la información electrónica.
- Definición de mecanismos para la interoperabilidad con otros sistemas.

Esto se refleja en una práctica inexistencia de soluciones comerciales orientadas a dar respuestas eficaces a esta problemática. De este modo, las Administraciones Públicas carecen de referencias sobre la dirección a seguir.

Como se ha podido comprobar durante el contraste del modelo, existen principalmente tres tendencias para abordar el archivo de documentos electrónicos:

- Implantar **soluciones orientadas a automatizar la gestión de documentación en papel**, que aunque permiten el almacenamiento de información digital, no entran a resolver su problemática. Es-

tas soluciones suelen encontrarse en plataformas integrales de Administración Electrónica, como SIGEM o W@nda, ampliamente adoptadas por Entidades Locales.

- **Ampliar la funcionalidad de los sistemas de Gestión Documental existentes** para habilitar ciertos mecanismos de preservación. Se ha constatado que esta es la alternativa preferida por muchas Administraciones Regionales que ya disponen de sistemas corporativos como Documentum o Alfresco.
- **Habilitar espacios de almacenamiento de documentos digitales**, dejando la implantación de políticas de preservación para más adelante. Esta opción se da en Administraciones que generan volúmenes de información reducidos.

En general, estas iniciativas tienden a impulsarse a través de departamentos de informática y sistemas, siendo poco habitual la participación de personal relacionado con el Archivo de la entidad en el proceso de diseño y definición.

Como excepción cabe mencionar el proyecto **iArxiu** emprendido por la Generalitat de Catalunya, a través de la Agència Catalana de Certificació (CATCERT), para la definición e implementación de un sistema de

Archivo Electrónico orientado a la conservación a largo plazo y a la implantación práctica de políticas de preservación.

2.4. Expediente electrónico

El Expediente Electrónico es un concepto clave a la hora de abordar el estudio del Archivo Electrónico. Así, puede verse como la representación en el entorno digital del **expediente administrativo**, que no es sino el conjunto de documentos, en los que la Administración basa sus decisiones y actuaciones sobre un tema en concreto.

Según la LAECSP, en su artículo 32, el expediente electrónico se define como

“El conjunto de documentos electrónicos correspondientes a un procedimiento administrativo, cualquiera que sea el tipo de información que contengan”.

La LAECSP también indica cómo debe estructurarse:

“El foliado de los expedientes electrónicos se llevará a cabo mediante un índice electrónico, firmado por la Administración, órgano o entidad actuante,

según proceda. Este índice garantizará la integridad del expediente electrónico y permitirá su recuperación siempre que sea preciso, siendo admisible que un mismo documento forme parte de distintos expedientes electrónicos.”

En otras palabras, al hablar de expediente electrónico estamos refiriendo a un conjunto de documentos electrónicos que dan soporte a un procedimiento administrativo junto con un índice que los referencia y sirve como elemento integrador.

Esta aproximación nos permite, entre otras cosas, que un mismo documento pueda formar parte de distintos expedientes administrativos.

Entre los contenidos que pueden formar parte del expediente electrónico se encuentran:

- Documentos creados electrónicamente por las Administraciones Públicas
- Digitalizaciones de documentos creados en formato papel
- Digitalizaciones de documentación aportada por los ciudadanos
- Documentos aportados por los ciudadanos en formato electrónico

- Otros contenidos como registros de Bases de Datos, justificantes de transacciones electrónicas, etc.

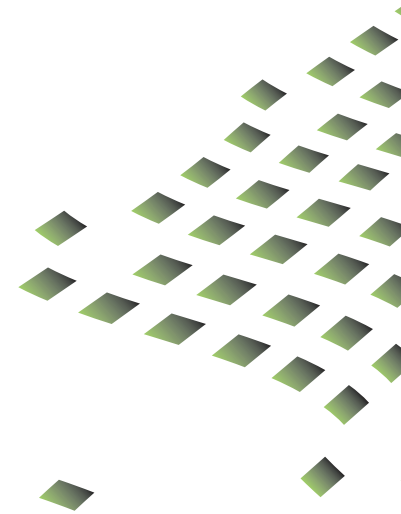
Hay que destacar que la digitalización de documentación en papel tendrá consideración de copia auténtica, pudiendo procederse a la destrucción de los originales, siempre que se cumplan los requisitos y actuaciones previstas en el artículo 46 de la Ley 30/1992, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común.

También es importante indicar que, aunque durante la tramitación administrativa, el expediente electrónico es un elemento “vivo”, al que pueden añadirse documentación y editar la existente, cuando llega al Archivo Electrónico, tanto el índice como los documentos que contiene no podrán modificarse.

2.5. Valor de los documentos

Los documentos, según la finalidad para la que fueron producidos, podrán tener los siguientes valores:

- **Valor primario:** aquel que va unido a la finalidad inmediata por la cual el documento se ha producido. Este valor tiene asociado un plazo preestablecido,



denominado **vigencia administrativa**, durante el cual el documento genera derechos y obligaciones, y después del cual el documento perderá ese valor. Dentro de este apartado, se distinguen los siguientes valores:

- **Valor administrativo:** cuya finalidad es dejar constancia de la gestión de un proceso administrativo.
 - **Valor fiscal o contable:** valor asociado al cumplimiento de obligaciones tributarias o de justificación de operaciones relacionadas con la contabilidad.
 - **Valor legal o jurídico:** valor asociado a derechos u obligaciones legales.
 - **Valor informativo:** valor inherente a cualquier documento relacionado con sus contenidos. Su finalidad es servir como documentación de apoyo a la gestión administrativa.
- **Valor secundario:** aquel que obedece a otras motivaciones que no son la propia finalidad del documento, y que se va adquiriendo a lo largo del tiempo:
- **Valor histórico:** cuya finalidad es servir como fuente para la investigación histórica y cultural.

2.6. Documentación susceptible de ser archivada

Las Administraciones Públicas tienen la obligación de conservar los expedientes administrativos mientras no expire su periodo de vigencia, durante el cual los documentos aportados mantienen un valor probatorio.

Sin embargo, ésta no es la única documentación que puede ser objeto de archivado, pudiendo considerarse beneficioso la conservación de otros tipos de documentación que aporten cierto valor informativo o se espere que adquieran un valor histórico en el futuro.

A continuación se indica qué tipo de documentación suelen conservar las Administraciones Públicas:

- Los **documentos que forman parte de los expedientes**, que materializan actuaciones secuenciales y procedimentales de la Administración encaminados a la resolución de un asunto, y que se deberán clasificar según una serie de normas.
- Los libros o recibos de **registros** para la recepción y remisión de solicitudes, escritos y comunicaciones.
- La **correspondencia o comunicaciones electrónicas**, que no deben confundirse con los documentos que forman parte de los expedientes.

- La **documentación de apoyo a la gestión administrativa**, como textos legales, boletines oficiales, publicaciones, folletos comerciales, fotocopias de legislación, etc. que ayuden en la gestión de los procesos administrativos.

2.7. Metadatos

Los metadatos son un concepto básico asociado a la archivística y pueden definirse como “la información referente a la información”.

En la práctica es información complementaria, independiente del documento, que permite categorizarlo y situar los contenidos dentro de un contexto que facilite su comprensión.

Dado que el Archivo Electrónico persigue la conservación de la información durante largos periodos de tiempo, la capacidad de recuperar el contexto en el que se creó el documento en el futuro es un requisito importante.

Existen distintos tipos de metadatos, dependiendo de su función informativa:

- **Descriptivos:** Aportan información sobre aspectos genéricos del documento como el autor, el título, la fecha de creación, la fecha de publicación, etc. También pueden proporcionar un breve resumen de los contenidos.
- **Técnicos:** Informan de las características de los soportes electrónicos, proporcionando datos sobre aspectos tecnológicos para el procesado del fichero. Pueden variar en función de la naturaleza de los contenidos. A modo de ejemplo se mencionan algunos de los más comunes:
 - ❖ **Imágenes:** formato empleado, tamaño en píxeles de la imagen, mecanismo de compresión, resolución, esquema de color.
 - ❖ **Audio:** formato empleado, tasa de compresión, duración.
 - ❖ **Texto:** fuentes empleadas, características de presentación
- **Preservación:** Se orientan a garantizar la integridad y accesibilidad de los contenidos, haciendo referencia a aspectos como firma de los documentos o pistas de auditoría.

Los documentos en formato papel almacenan tanto los contenidos de información como la forma en que se presentan al usuario. En el mundo digital, los con-



tenidos se codifican en ficheros electrónicos como secuencias de unos y ceros, que son procesados por programas y aplicaciones para generar una representación de estos contenidos. Distintas interpretaciones de las secuencias de unos y ceros darán lugar a representaciones diferentes de los contenidos. Por ello, es tan importante preservar la información como documentar las propiedades y características que permitan reconstruir la representación original.

Esta información de representación se registra a través de los metadatos, principalmente de carácter técnico, y debería estar siempre disponible junto a los contenidos del documento.

2.8. Diferencias entre un Gestor Documental y un Archivo Electrónico

Otra puntualización importante es establecer la distinción entre los **Sistemas de Gestión de Documentos Electrónicos (SGDE)**, también llamados **Gestores Documentales**, y los **Sistemas de Gestión de Documentos Electrónicos de Archivo (SGDEA)**, o Archivos Electrónicos.

Los Gestores Documentales son herramientas que facilitan la creación, edición y compartición de documentos electrónicos generados como resultado de la operativa de la unidad organizativa, permitiendo a los usuarios modificarlos y borrarlos. En otras palabras, manejan documentos “vivos” que pueden cambiar como resultado del trabajo diario.

En un Gestor Documental pueden conservarse distintas versiones de un mismo documento, mantener duplicados, así como almacenar cualquier documentación de apoyo que se considere oportuno para la tramitación.

Sólo cuando el documento electrónico alcanza su versión definitiva y no se permiten nuevas modificacio-

CARACTERÍSTICAS	GESTOR DOCUMENTAL	ARCHIVO ELECTRÓNICO
Tipo de documentos	Documentos en producción y trámite	Documentos de archivo
Modificación de documentos	Permitido, pudiendo existir varias versiones de un mismo documento	Prohibido, sólo se podrá incluir la versión final del documento
Versiones de documentos	Pueden conservarse varias versiones de un mismo documento	Sólo se conserva la versión final, que no podrá ser modificada
Eliminación de documentos	Permitido	Prohibido (excepto en transferencias de un Archivo a otro y casos excepcionales)
Sistemas de clasificación y almacenamiento	Opcional (gestionado por los usuarios del sistema)	Obligatorio, necesidad de un Cuadro de clasificación, gestionado por archiveros
Políticas de conservación	Opcionales	Obligatorias, teniendo en cuenta que los documentos se guardan a largo plazo
Objetivo	Facilitar la gestión de documentos en el trabajo diario	Proporcionar un repositorio seguro para la conservación de los documentos

nes, se ingresa en el sistema de Archivo Electrónico, que se ocupará de su custodia y conservación. Una vez en el Archivo Electrónico, el documento podrá consultarse pero no editarse ni borrarse.

El Archivo Electrónico deberá asegurar la accesibilidad, disponibilidad, integridad y autenticidad de los documentos electrónicos en el futuro, independientemente del medio físico de almacenamiento y del formato del fichero.

Para favorecer su preservación, pueden imponerse algunos requisitos a los documentos previamente a su ingreso en el Archivo Electrónico:

- **formato:** será necesario que los documentos electrónicos de archivo tengan unos formatos definidos para su conservación y consulta, recomendándose que estos formatos estén basados en estándares abiertos.
- **firma electrónica:** se podrá exigir una serie de requisitos a los documentos firmados electrónica-



mente para poder verificar su autenticidad e integridad.

- **digitalización:** los documentos de papel que se digitalicen, deberán cumplir unos requerimientos normativos para que se puedan considerar copias auténticas del documento original.

Para facilitar el proceso, se recomienda que las normas, requisitos y políticas de conservación se se tengan en consideración desde el mismo instante en el que se genera el documento, y no sólo en el momento de su ingreso en el Archivo Electrónico.

El Archivo Electrónico deberá asegurar la accesibilidad, disponibilidad, integridad y autenticidad de los documentos electrónicos en el futuro, independientemente del medio físico de almacenamiento y del formato del fichero.



¹ El valor administrativo se entiende en este contexto como el valor que tienen los documentos por formar parte de un procedimiento administrativo, incluidos su valor legal, fiscal contable.

2.9. Ciclo de vida de los documentos

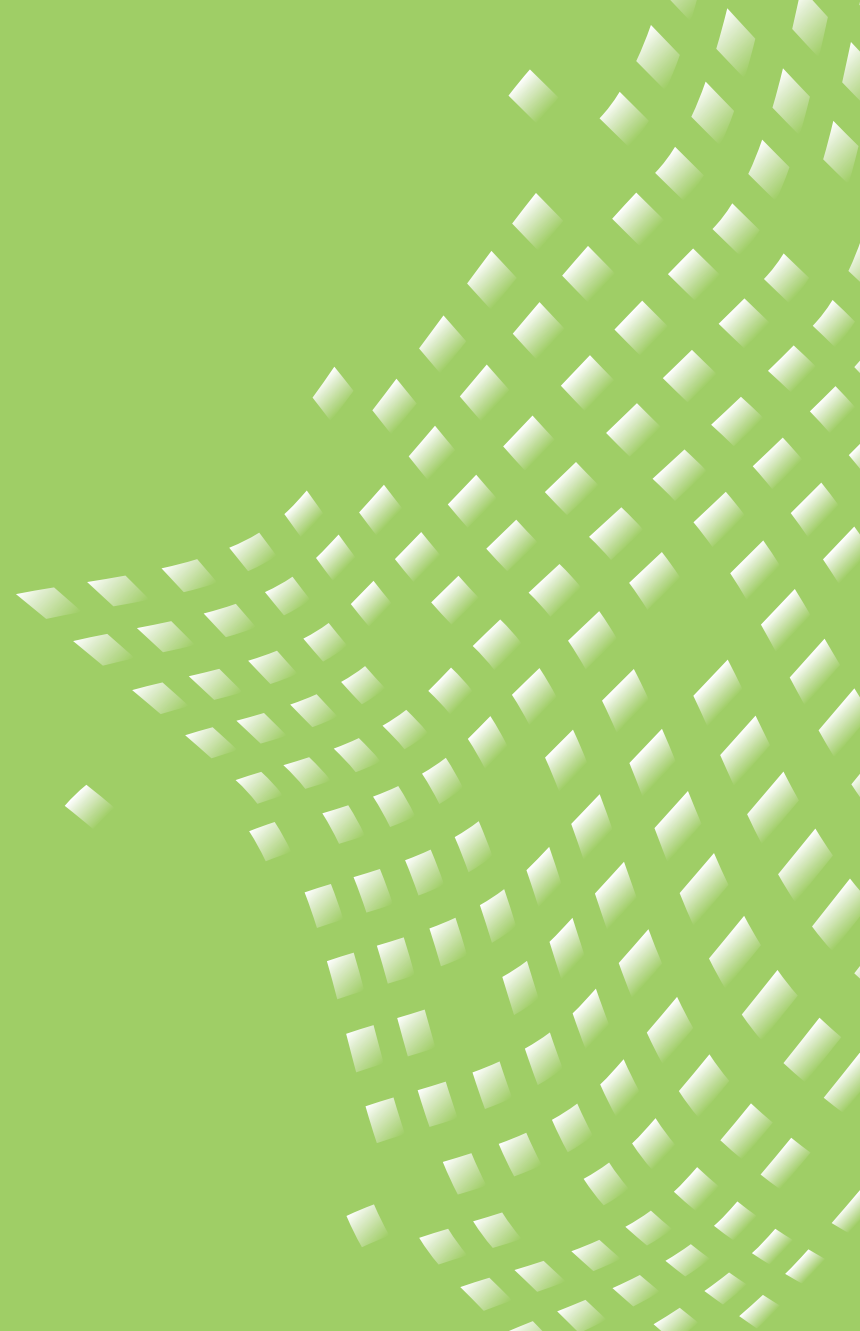
Los documentos administrativos evolucionan de acuerdo a un ciclo de vida, que comprende desde que su producción o recepción en las unidades administrativas hasta su eliminación o conservación por su valor histórico.

Las fases del ciclo de vida de los documentos son las siguientes:

- **Documentación activa:** es la etapa en la que los documentos son producidos o recibidos, su valor administrativo¹ está vigente y su consulta suele ser frecuente.
- **Documentación semi-activa:** es la etapa en la que el trámite administrativo ha finalizado pero su valor administrativo se mantiene vigente. Su consulta es esporádica.
- **Documentación inactiva:** es la etapa en la que su valor administrativo ha desaparecido, pero a su vez el documento ha adquirido un valor secundario, que puede ser fuente de investigación histórica.

03.

Modelo conceptual del Archivo Electrónico



3. Modelo conceptual del Archivo Electrónico

3.1. Descripción del Modelo Conceptual

Aunque las tecnologías de almacenamiento actuales permitirían centralizar toda la información generada en un sistema único de Archivo Electrónico, se ha optado por proponer una estructura jerarquizada de Archivos. Esta aproximación sigue el modelo archivístico tradicional, adoptado por la mayoría de las Administraciones Públicas para la gestión de sus documentos en formato papel.

Esta división se fundamenta en las fases del ciclo de vida de los documentos vistas en el apartado anterior, de modo que cada tipo de archivo se encargará de la conservación de documentos con requerimientos similares.

Los niveles de la estructura se identifican como:

- Archivo de Gestión o de Oficina
- Archivo Central / Intermedio
- Archivo Histórico

Como resultado de esta división, se favorece el establecimiento y asignación de responsabilidades con-

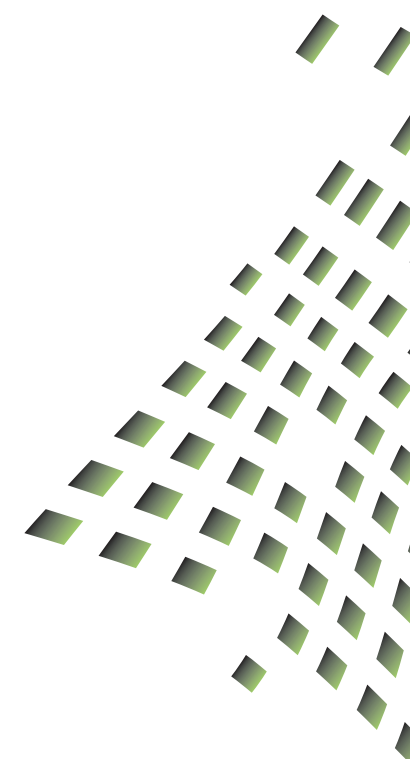
cretas sobre la conservación de la información a los distintos órganos de la entidad.

Este modelo es una simplificación del modelo clásico archivístico estructurado en cuatro niveles de Archivos:

- Archivo de Gestión o de Oficina
- Archivo Central
- Archivo Intermedio
- Archivo Histórico

Esta organización se ha ido desarrollando para facilitar el archivado de grandes cantidades de documentación en papel, condicionado por la disponibilidad de espacio físico en las dependencias de los Archivos. Con la capacidad de almacenamiento que dan las tecnologías actuales, esta limitación queda superada, por lo que se considera beneficioso simplificar el modelo y agrupar las funciones de Archivo Central y Archivo Intermedio en un único nivel de la estructura.

Hay que indicar que existen administraciones, como el Gobierno de Cantabria, que están adoptando esta

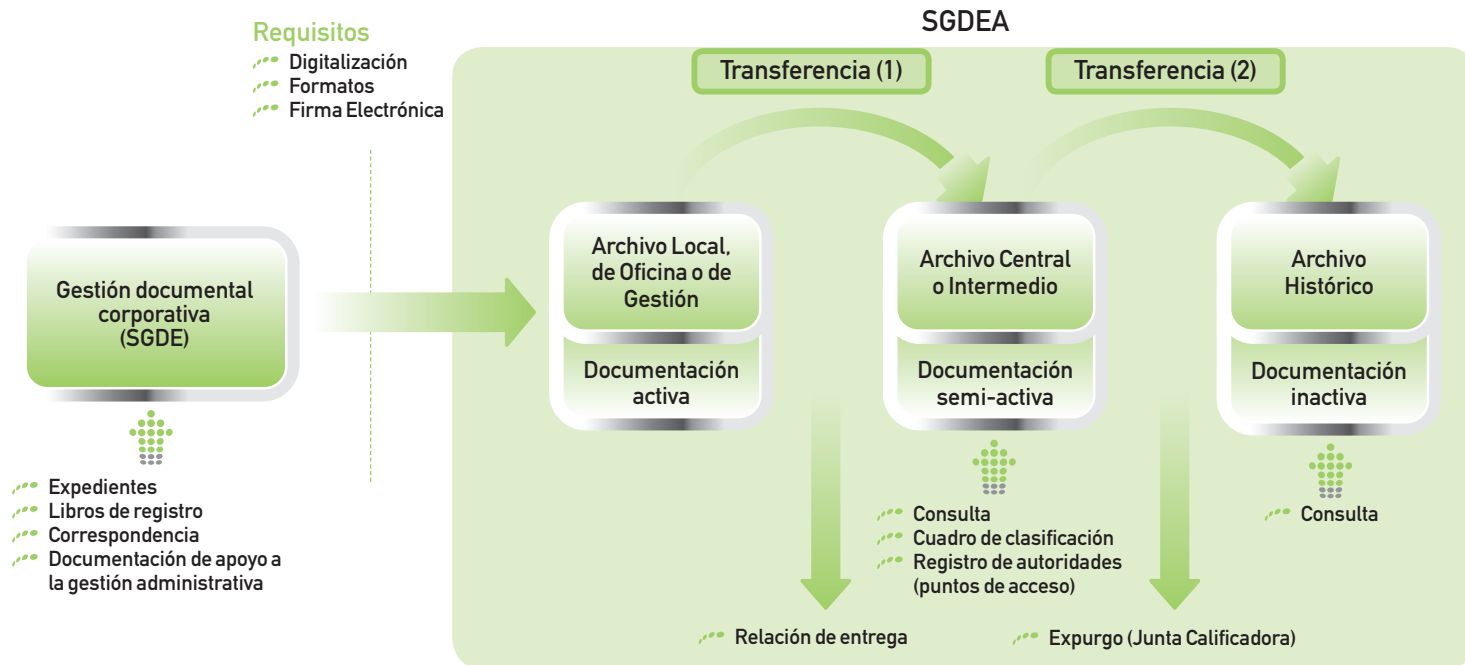


división clásica para sus sistemas electrónicos de Archivo, siendo igualmente válido a efectos prácticos.

A continuación, se incluye una representación esquemática del modelo, donde se puede ver los tres niveles de Archivos y las interacciones entre ellos.

En el esquema también se establece la **relación entre el Gestor Documental y el Archivo Electrónico**. El primero será el encargado de la gestión de la documentación durante el periodo de tramitación, ingresando posteriormente las versiones definitivas en el Archivo Electrónico para su conservación.

Modelo conceptual del Archivo Electrónico de las Administraciones Públicas.



Como se verá posteriormente, se ha contrastado que esta división no siempre se establece claramente, **siendo habitual que el Archivo de Oficina llegue a estar integrado dentro del propio sistema de Gestión Documental.**

El traslado de documentación de un archivo a otro viene determinado por unos actos formales denominados **transferencias**, para los cuales se establece un procedimiento definido.

Este evento no tiene porque implicar el movimiento de información de un sistema a otro, sino que representa más bien el traspaso de las responsabilidades sobre la conservación de esa documentación.

En los siguientes apartados se presentan los distintos niveles del modelo y se explican los mecanismos de transferencia entre niveles.

3.1.1. Junta Calificadora de documentos administrativos

La **Junta/Comisión Calificadora de documentos administrativos** de cada órgano se responsabiliza del estudio y dictamen de las cuestiones relativas a la calificación y utilización de los documentos administrativos, así como a su integración en los Archivos y al régimen de acceso e inutilidad administrativa de tales documentos.

Las funciones de la Junta/Comisión, cuya composición, funcionamiento y competencias específicas se establecen por vía reglamentaria, incluyen:

- **Identificación:** análisis de la estructura de la organización y de sus funciones, para poder definir las series documentales que se producen en cada uno de los órganos.
- **Valoración:** definición de los valores primarios y secundarios de los documentos electrónicos que integran las series.
- **Consulta:** establecerá los criterios de accesibilidad y utilización de los documentos y series documentales.
- **Transferencia:** definición de las características y requisitos de los documentos y series documentales a transferir de un Archivo a otro, así como los plazos de permanencia en el Archivo Central.

- **Expurgo:** eliminación de los documentos o series documentales que han superado su periodo de vigencia administrativa y carecen de valor histórico.

3.1.2. Archivo Local, de Gestión o de Oficina

Función: Es el Archivo que mantiene la documentación mientras dura su tramitación y siempre que su consulta sea constante. Es decir, se encarga de la gestión de los documentos en fase activa.

Este es el nivel de Archivo más próximo a la unidad productora y donde los documentos son más fácilmente accesibles.

Responsabilidad: La conservación y la custodia de los documentos existentes en los Archivos de Gestión es responsabilidad de los encargados de las unidades organizativas de las que formen parte, quienes pueden delegar la gestión, clasificación y ordenación de los documentos en uno o más responsables.

Acceso: Se podrá acceder por distintos medios, siendo necesario siempre la identificación del usuario. Se establecerán perfiles de usuario, dependiendo del grado de acceso a los documentos requeridos.

Estos usuarios tendrán posibilidad de añadir o consultar documentos según los permisos pertinentes (perfiles de usuario).

Valor de los documentos: informativo, administrativo, legal o fiscal (valor primario).

Modificación / eliminación: no se podrán modificar o eliminar documentos. Los documentos que forman parte del Archivo de Oficina deberán ser siempre versiones finales de los documentos generados.

Permanencia: Entre 2 y 5 años. Se recomienda no conservar los documentos electrónicos por un periodo mayor, ya que raramente se implementan mecanismos de preservación a este nivel, siendo preferible transferirlos al Archivo Central.

3.1.3. Transferencia del Archivo de Gestión al Archivo Central

Se llama transferencia al proceso archivístico por el que, transcurrido el periodo de permanencia en las unidades productoras, se traspasan las series documentales al Archivo de nivel superior. Implica el traspaso de las responsabilidades sobre el tratamiento y custodia de dicha documentación.

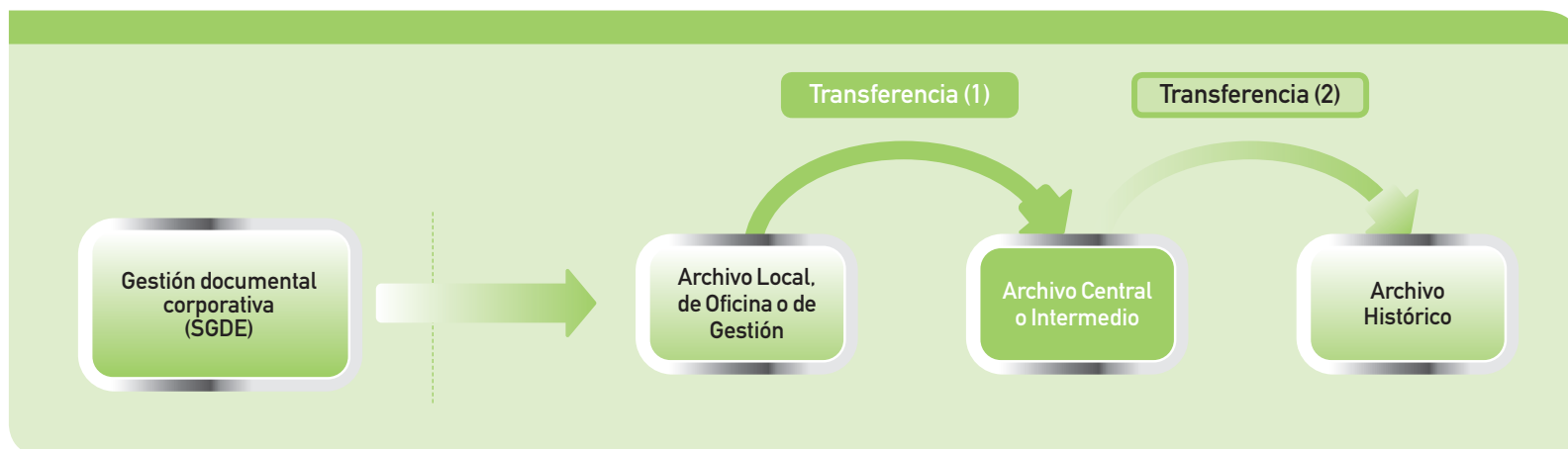
Una **serie documental** es el conjunto de los documentos, producidos o gestionados por una unidad orgánica en cumplimiento de sus funciones, y que tienen el mismo tipo documental o se refieren a un mismo asunto. Las series documentales suelen ser archivadas, conservadas, utilizadas, valoradas, transferidas o eliminadas como una única unidad.

Son candidatos al traslado al Archivo Central todos aquellos expedientes administrativos cuya tramitación a concluido, así como cualquier otro tipo de documentación para la que se determine que su consulta no es habitual.

Los documentos se clasificarán de acuerdo a un **Cuadro de Clasificación**, esquema donde se encuentran



Transferencia el Archivo de Gestión al Archivo Central.



estructuradas de forma jerárquica y lógica las series documentales que conforman las secciones de un **Fondo Documental** (conjunto de documentos reconocido dentro de los Archivos).

Sólo se transferirán los documentos cuyo valor sea administrativo, legal o fiscal, eliminándose previamente cualquier copia o duplicado, así como de la documentación de apoyo a la gestión administrativa.

Responsable de la transferencia: Jefe de la Unidad, quién debe firmar electrónicamente la relación de entrega de los documentos a transferir.

Esta **Relación de Entrega** es un documento en el que se detallan:

- listado de documentos que se van a transferir
- persona encargada de realizar la transferencia
- serie documental y expediente al que pertenecen los documentos,
- periodo de vigencia administrativa

Receptor de la transferencia: será la persona encargada del Archivo Central, que deberá verificar la documentación entregada y validar si es correcta y se corresponde con lo indicado en la relación de entrega.

3.1.4. Archivo Central o Intermedio

Función: Es el Archivo que reúne la documentación que se debe conservar por un periodo de tiempo definido, debido a obligaciones administrativas, legales o fiscales, pero que se consulta rara vez. Es decir, se encarga de la gestión de la documentación semi-activa.

Responsabilidad: Unidad de Archivo dentro del órgano administrativo, el cual se encarga de la conservación y custodia de los documentos.

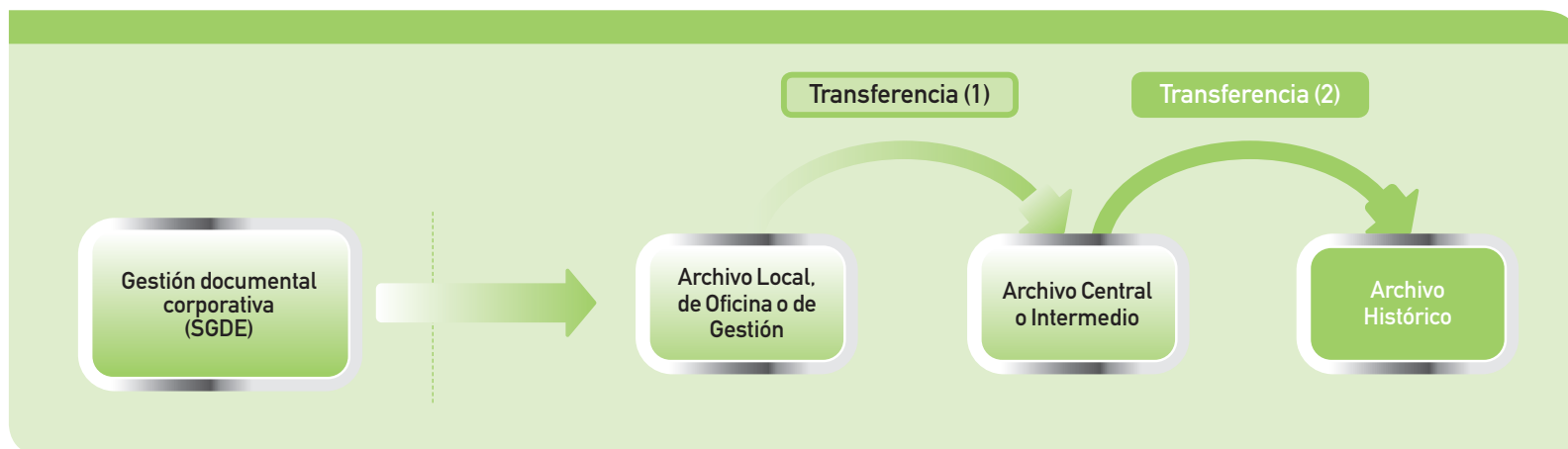
Acceso: La consulta de los documentos archivados requerirá la solicitud a la Unidad de Archivo, quienes deberán autorizar al solicitante. La Junta/Comisión Calificadora de documentos administrativos fijará una serie de normas para el acceso a los documentos dentro del Archivo Central.

Dependiendo del tipo de documento requerido, se podrán habilitar distintos medios, no siendo posible en ningún caso la alteración del documento.

Para añadir nuevos documentos a un expediente, **en circunstancias muy excepcionales**, será necesario autorización del responsable de la Unidad de Archivo.



Transferencia el Archivo de Gestión al Archivo Histórico.



Valor de los documentos: administrativo, legal o fiscal (valor primario).

3.1.5. Transferencia del Archivo Central al Archivo Histórico

Cuando la **vigencia administrativa** de un documento o serie documental haya finalizado, se procederá a su eliminación del Archivo.

Como excepción, aquellos documentos con interés histórico o que puedan ser objeto de investigación científica, se transferirán al Archivo Histórico donde se conservarán permanentemente.

En este paso, la Junta/Comisión Calificadora de documentos administrativos realizará un **expurgo**, que es el proceso de evaluación crítica de documentación

con vistas a su posterior conservación o eliminación, en función de la existencia de valor histórico. Este proceso tendrá como resultado:

- Eliminación de los documentos o series documentales cuya vigencia administrativa haya finalizado y carezcan de un valor documental.
- Conservación y transferencia al Archivo Histórico de los documentos que hayan adquirido un valor histórico a lo largo de su ciclo de vida.

Responsable de la transferencia: Junta/Comisión Calificadora de documentos administrativos, que será la encargada de decidir los documentos que se deben transferir al Archivo Histórico y los que se deben eliminar.

Receptor de la transferencia: Unidad de Archivo Histórico, encargado de los documentos históricos.

3.1.6. Archivo Histórico

Función: Reúne los expedientes y documentos que deben conservarse indefinidamente por su valor histórico. Se asocia a la última fase del ciclo de vida de los documentos, cuando la documentación se considera inactiva.

Responsabilidad: Unidad de Archivo Histórico, la cual se encargará de su gestión y conservación.

Acceso: mediante solicitud a la Unidad de Archivo Histórico, quién deberá autorizar al solicitante.

Valor de los documentos: histórico, debido a que ha finalizado la vigencia administrativa de los documentos, perdiendo su valor primario (administrativo, fiscal o legal).

3.2. Aspectos clave del modelo

A continuación se presentan la aproximación hecha a algunos de los aspectos más determinantes del Archivo Electrónico en el ámbito de la Administración Pública.

Son los temas que han generado un mayor interés durante las entrevistas de contraste, en las cuales se han generado interesantes aportaciones que han permitido centrar y completar el estudio inicial.

3.2.1. Políticas de conservación de los contenidos

La mayor dificultad que se presenta al abordar el Archivo Electrónico es cómo mantener los contenidos accesibles para los usuarios durante largos periodos de tiempo, evitando que los soportes queden obsoletos.

La manera de solventar este problema es planificando y poniendo en práctica políticas de preservación de los documentos almacenados en el Archivo Electrónico. Esta tarea se aborda a través de las siguientes aproximaciones:

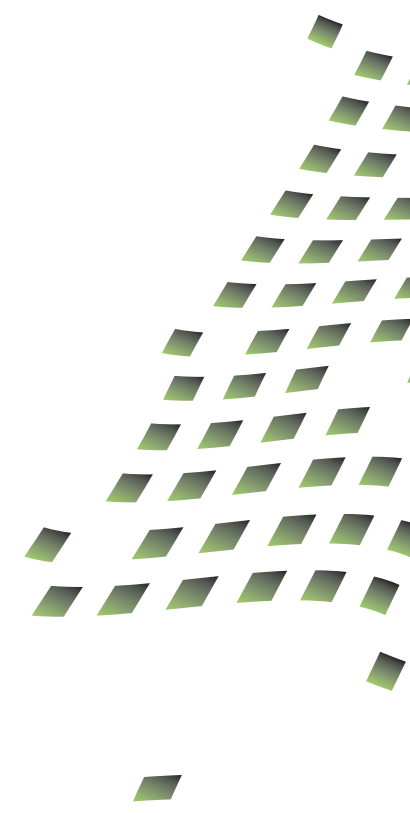
- **Migración:** Implica una conversión del documento a un formato más actual, por lo que implica un pro-

ceso de transformación en el que pueden perderse algunas de las características de presentación del formato original.

- **Emulación:** Es la construcción de aplicaciones basadas en la tecnología del momento para que sean capaces de manejar el formato original y mostrar una representación de sus contenidos.
- **Mantenimiento de la infraestructura original:** Supone conservar los sistemas, tanto hardware como software, existentes en el momento de creación del documento. Es una solución compleja y costosa ya que implica el mantenimiento de múltiples plataformas.

Durante el contraste del modelo se ha verificado que la migración de formato es la alternativa preferida por la mayoría de las Administraciones Públicas consultadas, pudiendo complementarse con la implantación de emuladores en algunos casos.

Hay que destacar que, para acometer políticas de migración y especialmente de emulación, es necesario conocer las especificaciones técnicas de los formatos empleados, por lo que existe coincidencia en considerar el uso de formatos basados en estándares libres como un requisito fundamental.



3.2.2. Formatos de fichero admitidos por el Archivo

Con objeto de homogenizar la gestión de los contenidos archivados, la mayoría de las entidades entrevistadas coinciden en la necesidad de restringir los tipos de formatos admitidos por el Archivo a un conjunto predeterminado y consensuado con los productores.

Esta aproximación permitiría centralizar los recursos disponibles en el desarrollo de políticas de migración y emulación únicamente para un número reducido de formatos distintos, favoreciendo su implantación y reutilización.

Estos formatos deberán presentar unas características apropiadas que favorezcan la preservación de los documentos durante largos periodos de tiempo y su presentación en diversas aplicaciones y entornos tecnológicos.

Como norma general, se favorecerá la adopción de formatos basados en estándares abiertos y reconocidos. En todo caso, se podrían aceptar formatos propietarios siempre que sus especificaciones sean públicas.

3.2.3. Preservación de documentos firmados electrónicamente

El valor legal de los documentos archivados debe mantenerse durante todo el tiempo que establece la legislación, para lo cual deberá garantizarse la autenticidad e integridad del documento. Eso se realiza en el escenario digital a través de la firma electrónica.

Establecer la validez de estas firmas durante periodos prolongados de tiempo es uno de los aspectos a garantizar por el Archivo Electrónico, ya que, debido a la caducidad de los certificados digitales usados para la firma, no es posible asegurar su accesibilidad en el futuro.

En este sentido, una solución consistiría en recopilar, junto con la firma digital, todos los datos necesarios para su validación y registrarlos en el repositorio junto con el documento firmado. Entre los datos a preservar se indicarían el mecanismo de firma usado, cadenas de certificados que prueben la validez de la CA, listas de revocación, etc.

Esta información deberá estar disponible en el momento del ingreso en el Archivo. Así el sistema será

capaz de validar la autenticidad del documento en ese momento y “dar fe” de la misma a través de una sello temporal electrónico.

Esta aproximación se ha contrastado como adecuada por la práctica totalidad de las Administraciones Públicas entrevistadas.

La firma electrónica garantiza el contenido del fichero original, pero no es extensible a cambios futuros en el fichero, por ejemplo producto de una migración de formato. Surge por tanto el problema de cómo garantizar que la nueva versión es una representación veraz del documento firmado.

La recomendación en este caso es mantener tanto el fichero original como la nueva versión en el archivo, documentando el proceso realizado. Sobre estos datos se realizará un nuevo sello temporal como garantía de la preservación de los contenidos.

Estos sellos temporales tienen también validez por tiempo limitado, debido a la caducidad de los certificados digitales empleados. Por ello deberán establecerse políticas de resellado que garanticen la autenticidad a lo largo del tiempo de archivado.

3.2.4. Clasificación de la documentación

Los volúmenes cada día mayores de información archivada hacen necesario la adopción de mecanismos de clasificación desde el mismo momento de generación de los documentos electrónicos como la única manera de catalogarlos y facilitar mecanismos para su consulta en el futuro.

Para ello deberá favorecerse la adopción de criterios archivísticos en la gestión de la documentación, así como la participación del colectivo de archiveros de las organizaciones en los procesos de diseño e implantación del Archivo Electrónico.

Uno de los mecanismos principales para facilitar esta clasificación es la incorporación de metadatos a la gestión de la documentación.

3.2.5. Aplicación del concepto de expediente electrónico

La mayoría de las aproximaciones al Archivo Electrónico usan el documento como unidad de gestión. Sin embargo, la tramitación administrativa actualmente se fundamenta en el concepto de expediente administrativo, el cual agrupa diversos documentos.

En este sentido, la representación que da la LAECSP del Expediente Electrónico, como un índice electrónico firmado por la Administración que referencie los documentos electrónicos que lo componen, permite superar esta diferencia y adecuarse a las soluciones implementadas por las Administraciones entrevistadas.

Se considera necesario que, independientemente de las soluciones implantadas, se adopte esta interpretación, para lo cual deberá habilitarse mecanismos para el establecimiento de relaciones y referencias entre los documentos archivados.

Por otro lado, la LAECSP permite que un mismo documento forme parte de distintos expedientes. Esto plantea problemas derivados de los distintos tiempos de vigencia de los documentos y puede ocasionar que expedientes caducados deban ser preservados porque contienen información referenciada por otros expedientes aún vigentes.

En este sentido, habrá que validar cada expediente previamente a su eliminación, asegurando la conservación de aquellos documentos que formen parte de otros expedientes activos.

3.2.6. Soporte para la documentación en formato papel

El modelo de Archivo propuesto, aunque se orienta fundamentalmente a la gestión de documentos en formato electrónico, proporciona ciertas facilidades para la gestión de documentación en papel.

Como es obvio, los contenidos de un documento en papel no pueden incorporarse al Archivo Electrónico. Sin embargo, lo que sí puede incluirse en el Archivo es una reseña electrónica que describa la naturaleza y procedencia del documento así como su ubicación física. Esta reseña electrónica puede incorporarse al expediente electrónico, que recopilará un listado con la totalidad de documentos que lo forman, independientemente de su formato de origen.

El Archivo Electrónico se convierte así en un medio de centralizar las consultas de los usuarios, que no necesitarán conocer el soporte en que se generó el documento. En caso de requerir un documento en formato electrónico, el usuario podrá acceder a su contenido, mientras que si está en formato papel, recibirá la información sobre dónde solicitarlo.

De todos modos, la tendencia actual se encamina a la homogenización de los soportes a través de procesos

de digitalización de la documentación en papel. Así, los documentos se escanean y se firman electrónicamente para dar prueba de su autenticidad e integridad, lo que permite incorporarlos al Archivo Electrónico.

3.2.7. Seguridad de la información

Dado que el contenido de los expedientes administrativo puede contener información personal y/o confidencial, es necesario establecer el alto nivel de control tanto sobre los resultados devueltos por las consultas al sistema como sobre el acceso a los propios documentos.

Así, debería garantizarse que la información confidencial sólo sea presentada a:

- Los funcionarios con competencias atribuidas sobre la materia que trata el expediente.
- Los propios usuarios interesados en el procedimiento.

En este aspecto, la normativa de referencia es la LOPD, por lo que el sistema tendrá que ser auditado para verificar que cumple con los parámetros establecidos en la Ley.

En cualquier caso, se considera menos perjudicial restringir el acceso e implementar mecanismos al-

ternativos para realizar esta consultas (solicitud expresa a la unidad productora, procesos de autorización).

3.2.8. Organización

Se ha detectado que existen organizaciones, principalmente en el ámbito local, que no han establecido aún procedimientos adecuados para el archivo de la documentación en formato electrónico. Esto lleva a la falta de una definición formal sobre la responsabilidad sobre la conservación de los documentos a lo largo de su ciclo de vida.

En esta situación, se favorece que las unidades productoras conserven la información durante largos periodo de tiempo, limitando la visibilidad de la misma para otras partes de la organización.

Además, mucha de esta información no llega a archivar-se nunca, manteniéndose localmente en los puestos de los usuarios, con pocas garantías de conservarse y se accesible a medio plazo.

En este contexto, implantar políticas de seguridad eficaces que abarquen todo el ciclo de vida de los documentos es costoso y de difícil mantenimiento.

La centralización de la información dentro de la unidad administrativa facilitaría la asignación de responsabilidades sobre su conservación, así como la homogenización de su gestión por parte de los distintos usuarios.

Para ello, será necesario promover buenas prácticas relacionadas con la gestión de la documentación, como:

- Establecimiento de criterios comunes para el etiquetado y catalogación de los documentos
- Definición de procedimientos de generación, edición y revisión de la documentación dentro de la organización.

3.2.9. Entorno normativo

Durante las entrevistas realizadas, las Administraciones encuestadas han manifestado su preocupación relativa a la existencia de un vacío normativo que regule la gestión adecuada de la documentación electrónica, principalmente en aspectos como:

- Definición de plazos de conservación de la documentación en cada uno de los tipos de archivo.
- Estandarización de metadatos como medio de favorecer la interoperabilidad entre organismos públicos.

- Regulación de los formatos de archivo requeridos a los ciudadanos y en las comunicaciones entre Administraciones.
- Definición de procedimientos para la transferencia de ficheros digitales entre los distintos tipos de fichero.
- Regulación de procedimientos para la consulta y acceso a la información.
- Definición de criterios de validez que rijan los procesos de transformación de ficheros digitales, especialmente de aquellos firmados electrónicamente.

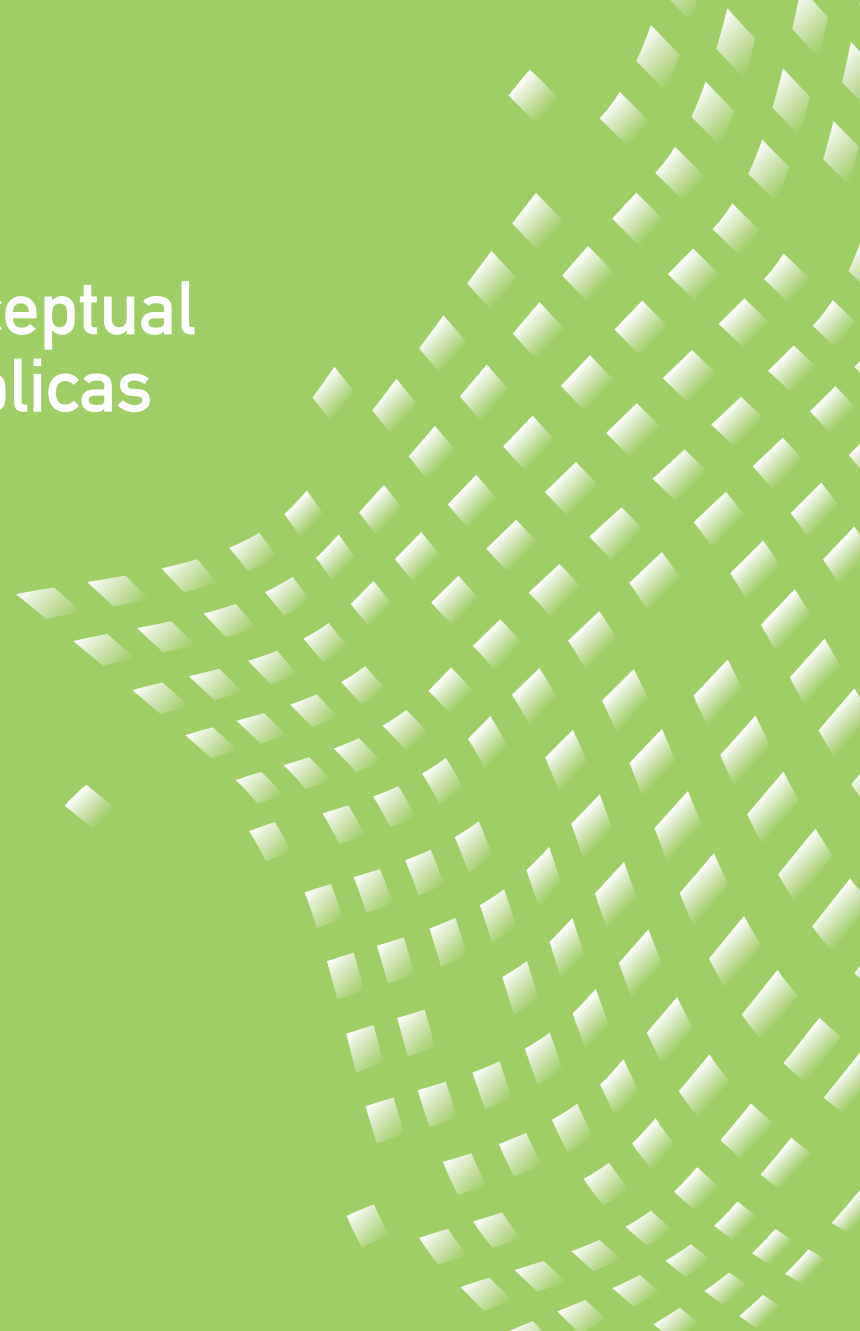
Esta indefinición está provocando el retraso en la implantación de medidas concretas por parte de las Administraciones Públicas y, como consecuencia, muchas de ellas están optando por adoptar criterios “de contingencia” de manera particular, soportados por normativa particular.

En este sentido, se espera con interés la aprobación de nueva regulación de ámbito nacional que clarifique algunos de estos aspectos anteriores, en concreto:

- Reglamento de aplicación de la Ley 11/2007, de Acceso Electrónico de los Ciudadanos a los Servicios Públicos (LAECSP).
- Esquema Nacional de Interoperabilidad.
- Esquema Nacional de Seguridad.

04.

Aplicación del modelo conceptual a las Administraciones Públicas



4. Aplicación del modelo conceptual a las AAPP

4.1. Flexibilidad del modelo de Archivo Electrónico propuesto

La división del modelo en diferentes niveles de Archivo facilita la asignación de responsabilidades sobre la conservación del documento en función del punto de su ciclo de vida en que se encuentre.

Este modelo intenta dar respuesta a la problemática desde una perspectiva organizativa, tratando en lo posible de independizarlo del entorno tecnológico existente en cada organismo y dotándolo de suficiente flexibilidad para que pueda ser adaptado a los requerimientos de los distintos tipos de Administraciones Públicas.

En este sentido, durante las reuniones de contraste, se han identificado dos escenarios particulares, determinados por la infraestructura tecnológica existente en la organización, y que podría pensarse que quedan fuera del modelo propuesto, aunque como se verá a continuación no es así.

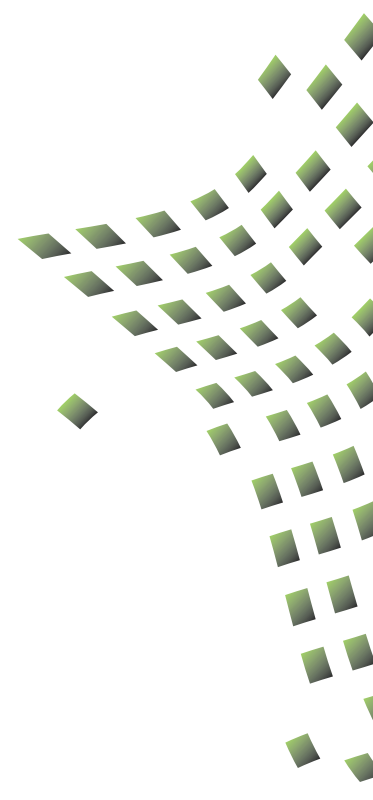
4.2. Escenario 1

Este escenario se produce en unidades administrativas con un **volumen reducido de documentación a gestionar** y que, hasta la fecha, no han tenido la necesidad de implantar un sistema informático de Gestión Documental.

En este contexto, la información habitualmente se encuentra descentralizada, diseminada en distintos sistemas informáticos, y no existen los mecanismos necesarios para garantizar su seguridad e integridad.

En este escenario, la consulta de información suele ser ineficaz y el riesgo de pérdida de información muy alto.

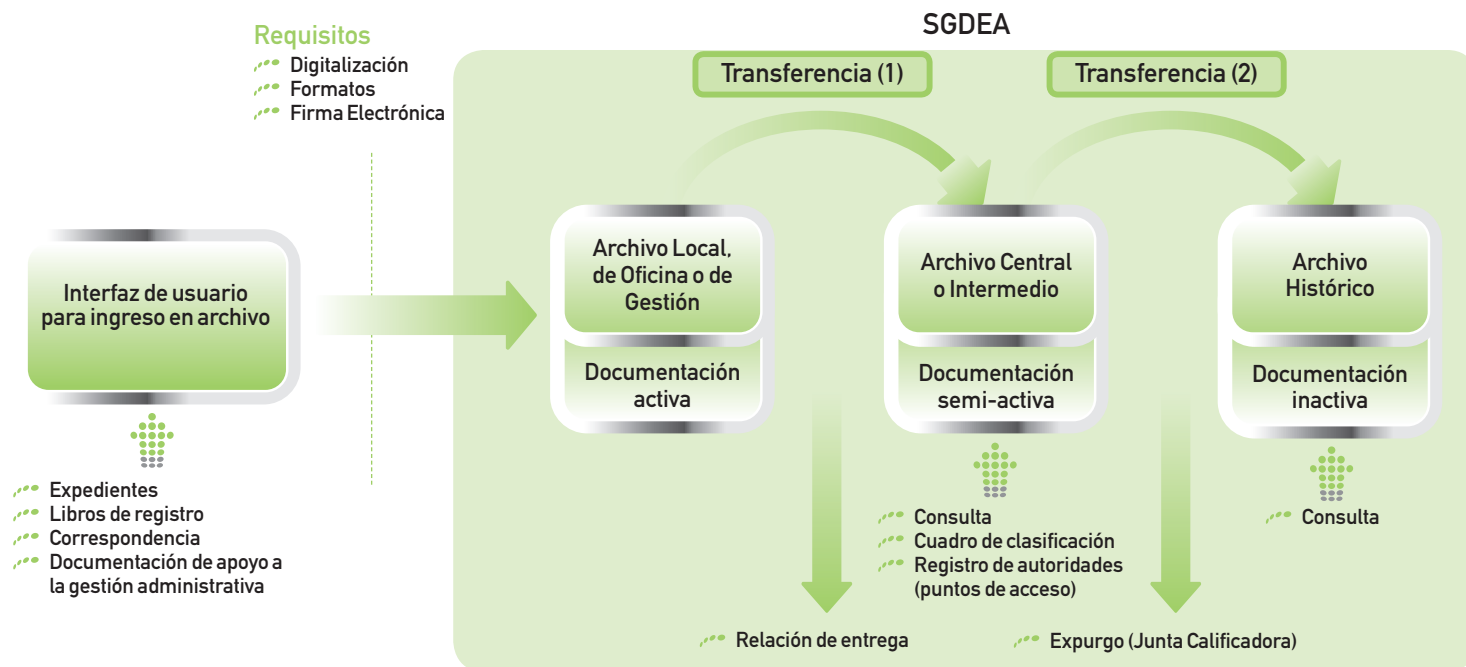
La implantación de un sistema de Archivo Electrónico proporciona un repositorio único en la unidad organizativa donde almacenar la información a conservar, lo que permite centralizar las consultas tanto internas como externas. Así mismo, facilita la implantación de procedimientos de control de accesos y salvaguarda de la información.




Adaptación del Modelo de Archivo al Escenario 1.

Como puede comprobarse en el esquema siguiente, el modelo sigue siendo válido implantando mecanismos que permitan el ingreso de los documentos en el sistema de Archivo por parte de los usuarios.

Este escenario tiende a darse en Entidades Locales de tamaño reducido con una baja implantación de la Administración Electrónica.

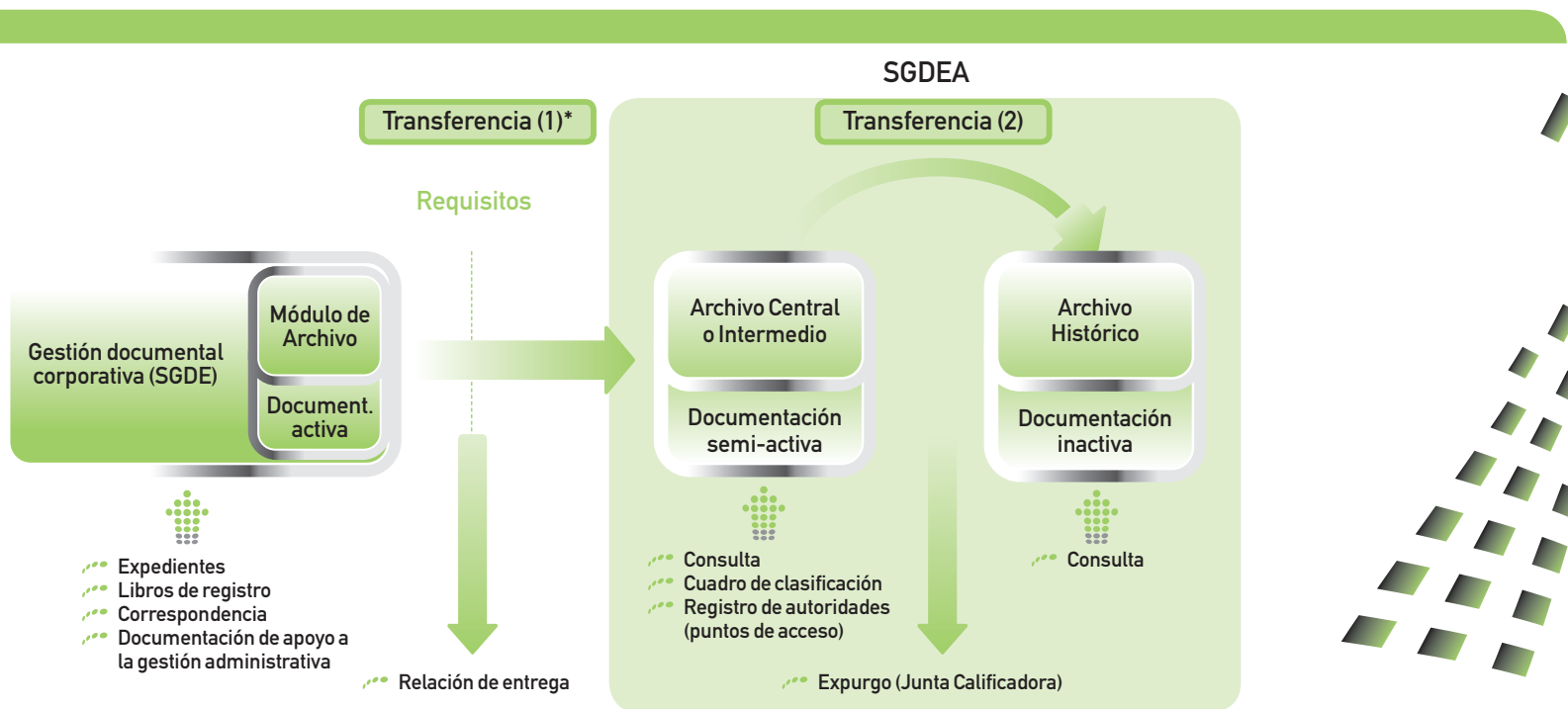


4.3. Escenario 2

Esta situación se corresponde con unidades administrativas que actualmente están asumiendo la res-

ponsabilidad sobre la preservación de la información electrónica que generan durante periodos de tiempo extensos, ya que no disponen de un Archivo Central corporativo donde derivar sus fondos documentales.

 Adaptación del Modelo de Archivo al Escenario 2.



* Los ficheros a transferir deberán cumplir los criterios de formato, firma, etc impuestos por el Archivo Central.

Este tipo de organizaciones tienden a implementar **Gestores Documentales complejos**, que implementan funcionalidad adicional para el archivo y preservación de documentos electrónicos. En la práctica se puede ver como que el Archivo de Gestión se ha integrado dentro del propio Gestor Documental. Dado que la responsabilidad de la gestión sigue asignándose a la unidad administrativa, la no existencia de sistemas separados para Gestión Documental y Archivo Electrónico, no invalida el modelo propuesto. En estos casos es importante identificar el organismo que debe realizar las funciones de Archivo Central o Intermedio (como el Archivo General de la Administración para la AGE) y establecer los procedimientos para la transferencia de documentación (formatos, criterios de validación, etc.) con el mismo.

Esta aproximación se está adoptando en distintas Administraciones Autonómicas (como Gobierno de Aragón y Junta de Extremadura) y en organismos dependientes de la AGE, como la Agencia Tributaria.

4.4. Normas internacionales

A continuación se muestra una relación de normas y modelos desarrollados para facilitar el diseño y gestión de sistemas de Archivo y sistemas de Gestión Documental.

Ninguna de ellas es de obligado cumplimiento y su propósito es establecer recomendaciones para el desarrollo y gestión de sistemas que trabajen con documentos electrónicos.

4.4.1. MoReq2 - Model Requirements for the management of electronic records

Esta especificación describe un Modelo de Requisitos funcionales para la gestión de documentos electrónicos de archivo, destinados a facilitar el diseño y la evaluación de un SGDEA.

4.4.2. ISAAR (CPF) - International Standard Archival Authority Records for Corporate Bodies, Persons and Families

Esta norma sirve de guía para elaborar registros de autoridad de archivos que proporcionan descripciones de entidades (instituciones, personas y familias) asociadas a la producción y a la gestión de archivos.

4.4.3. ISAD (G) - General International Standard Archival Description

Esta norma constituye una guía general para la elaboración de descripciones archivísticas, identificando y explicando el contexto y el contenido de los documentos de archivo con el fin de hacerlos accesibles, e integrando esta información contextual, definida en la norma complementaria ISAAR (CPF) y en las normas nacionales existentes.

4.4.4. NEDA - Normas Españolas de Descripción Archivística

Normas profesionales desarrolladas por CNEDA (Comisión de Normas Españolas de Descripción Archivística, creada por Orden del Ministerio de Cultura de 25 de mayo de 2007 (Orden CUL/1524/2007)), y orientadas a favorecer la realización de una prácticas adecuadas y homogéneas en los procesos técnicos de identificación y descripción de fondos/ colecciones de archivo, así como en la introducción y salida de datos en los SGDEA, facilitando la búsqueda y localización de los recursos archivísticos y el intercambio/comunicación de la información descriptiva.

4.4.5. ISO 14721:2003 - Open archival information system (OAIS)

Norma que precisa un modelo de referencia para sistemas de información de archivo, ya sean físicos o electrónicos, estableciendo un esquema de organización con responsabilidades concretas en la preservación de la información y facilitando su acceso a una determinada comunidad de usuarios.

4.4.6. UNE-ISO 15489-1 - Información y documentación - Gestión de documentos - Parte 1: Generalidades

Esta parte de la Norma ISO 15489 regula la gestión de documentos, en todos los formatos o soportes, que producen o reciben las organizaciones, ya sean públicas o privadas, en el ejercicio de sus actividades.

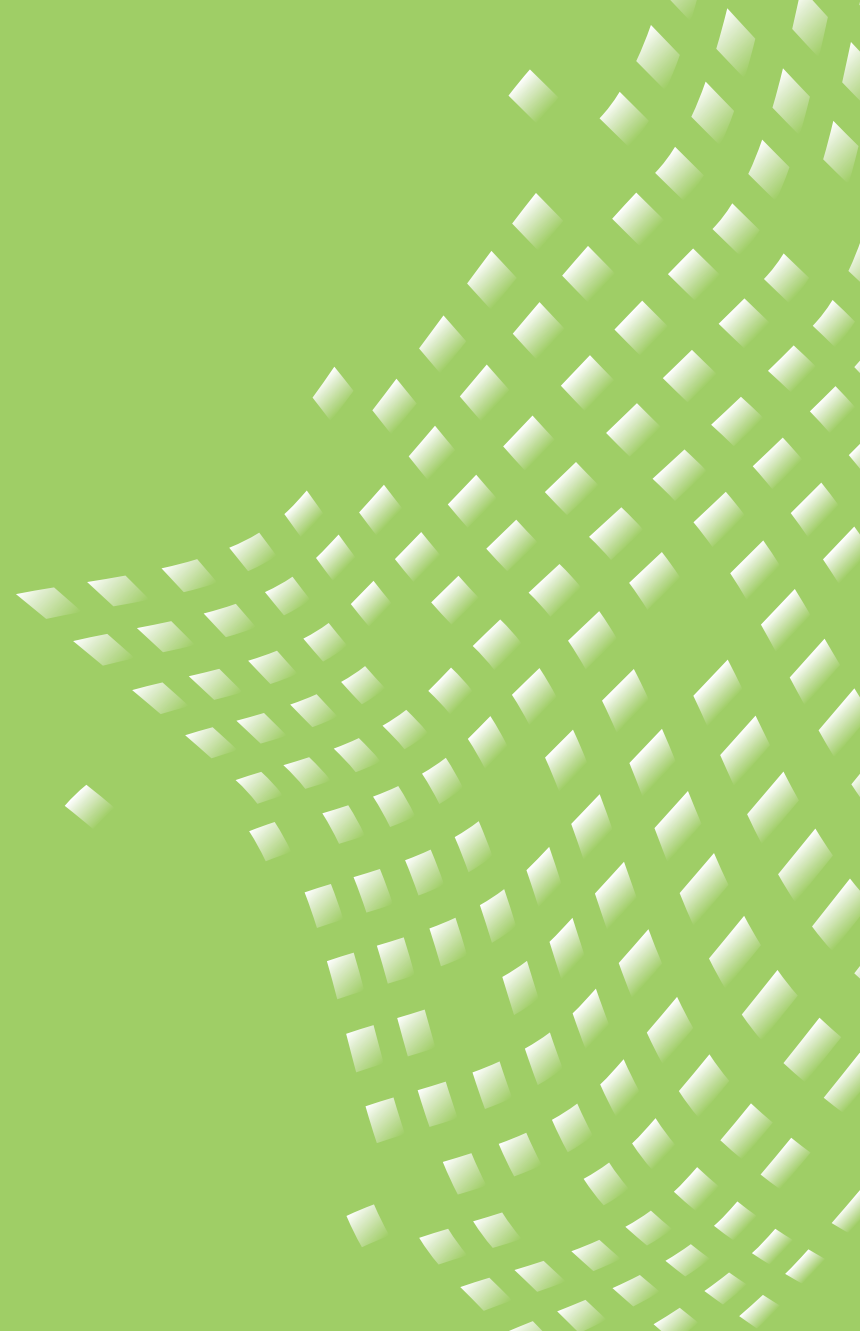
4.4.7. UNE-ISO/TR 15489-2 - Información y documentación – Gestión de documentos - Parte 2: Directrices

Esta parte de la Norma ISO 15489 es una guía de implementación de la Parte 1 para su uso por parte de los profesionales de la gestión de documentos, proporcionando una visión general de los procesos y factores que se deben tener en cuenta en el cumplimiento de la Norma.



05.

Definición funcional del Archivo Electrónico



5. Definición funcional del Archivo Electrónico

5.1. El modelo OAIS

OAIS (Open Archive Information System) es un modelo de referencia para la implantación de un sistema de repositorio digital desarrollado por el Consultative Committee for Space Data Systems (CCSDS) y estandarizado a través de la ISO 14721:2003.

Este modelo se ha concebido como un marco de referencia frente al que evaluar los sistemas de preservación de la información digital. Su orientación hacia la preservación a largo plazo lo distingue de otros modelos y, en la actualidad, es la referencia a la hora de abordar el diseño de un sistema para el archivo prolongado de información digital.

OAIS no especifica una implantación tecnológica del modelo, ni menciona sistemas o tecnologías particulares.

El modelo divide el sistema de Archivo Electrónico en un conjunto de módulos funcionales que abarcan el conjunto de servicios provistos por este tipo de sis-

temas. Para cada uno de estos módulos se definen unos requisitos funcionales así como la forma en que interactúan con otros módulos.

A continuación se describen los distintos módulos de una manera muy concisa. En apartados posteriores se ampliará el propósito de cada uno de ellos.

5.1.1. Módulo de ingreso (Ingest)

Habilita los procesos para la admisión de los objetos digitales procedentes de los productores.

5.1.2. Módulo de almacenamiento (Storage)

Establece mecanismos para la conservación de los objetos digitales en el repositorio.

5.1.3. Módulo de gestión de datos (Data management)

Mantiene y gestiona un registro de metadatos descriptivos de los objetos digitales almacenados.



5.1.4. Módulo de consulta y recuperación (Access)

Proporciona servicios a los usuarios para la realización de búsquedas en el repositorio y la visualización y obtención de copias de la información.

5.1.5. Módulo de administración y configuración (Administration)

Facilita la gestión operativa del sistema.

5.1.6. Módulo de preservación (Preservation Planning)

Monitoriza y hace un seguimiento de la evolución tecnológica para definir políticas de conservación de la información almacenada.



Esquema del Modelo OAIS.



5.2. Adecuación del modelo OAIS

El modelo OAIS ha servido de referencia a la hora de definir la funcionalidad del modelo propuesto de Archivo Electrónico de las Administraciones Públicas.

Sin embargo, no hay que olvidar que OAIS aborda el diseño de repositorios digitales genéricos, por lo que ha sido necesario ampliar la definición funcional incluyendo modificaciones para adecuarlo a los requerimientos propios de las Administraciones Públicas.

5.2.1. Módulo de Pre-ingreso

Éste no es un módulo dentro del sistema de Archivo Electrónico, sino un **complemento al modelo** que actúa como interfaz para facilitar la entrada de datos.

Este módulo, como se detallará más adelante en el documento, permite estructurar los contenidos a almacenar junto con la información de contexto (metadatos) proporcionada por el proveedor dentro de un encapsulamiento único, facilitando el proceso de ingreso en el Archivo Electrónico.

En principio, este módulo sólo se considera sólo para la implantación del Archivo Local, como un modo de facilitar la comunicación entre los sistemas productores (sistemas de Gestión Documental) y el Archivo Electrónico.

Se ha considera adecuado que, a la hora de definir las transferencias, el Archivo origen entregue la documentación en un encapsulamiento adecuado para su ingreso en el Archivo destino, evitando la necesidad de una transformación intermedia de los datos.

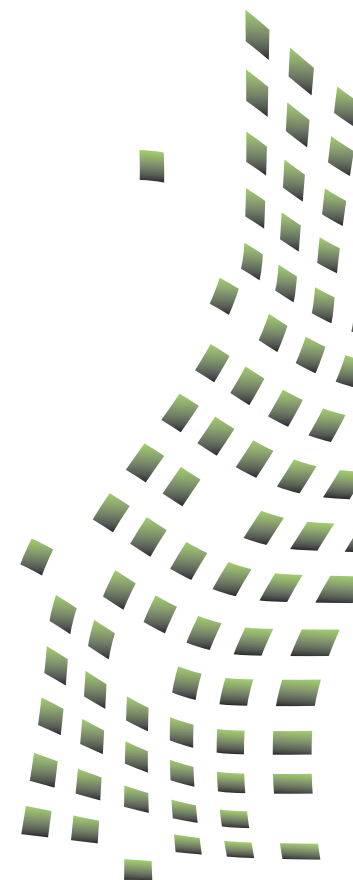
Si esto no fuese así, debería contemplarse la participación de este módulo en el proceso de transferencias.

5.2.2. Módulo de Gestión de Transferencias

El modelo OAIS considera el repositorio digital como un sistema único. Sin embargo el modelo propuesto establece una estructura jerárquica por niveles que se implementará sobre distintos sistemas.

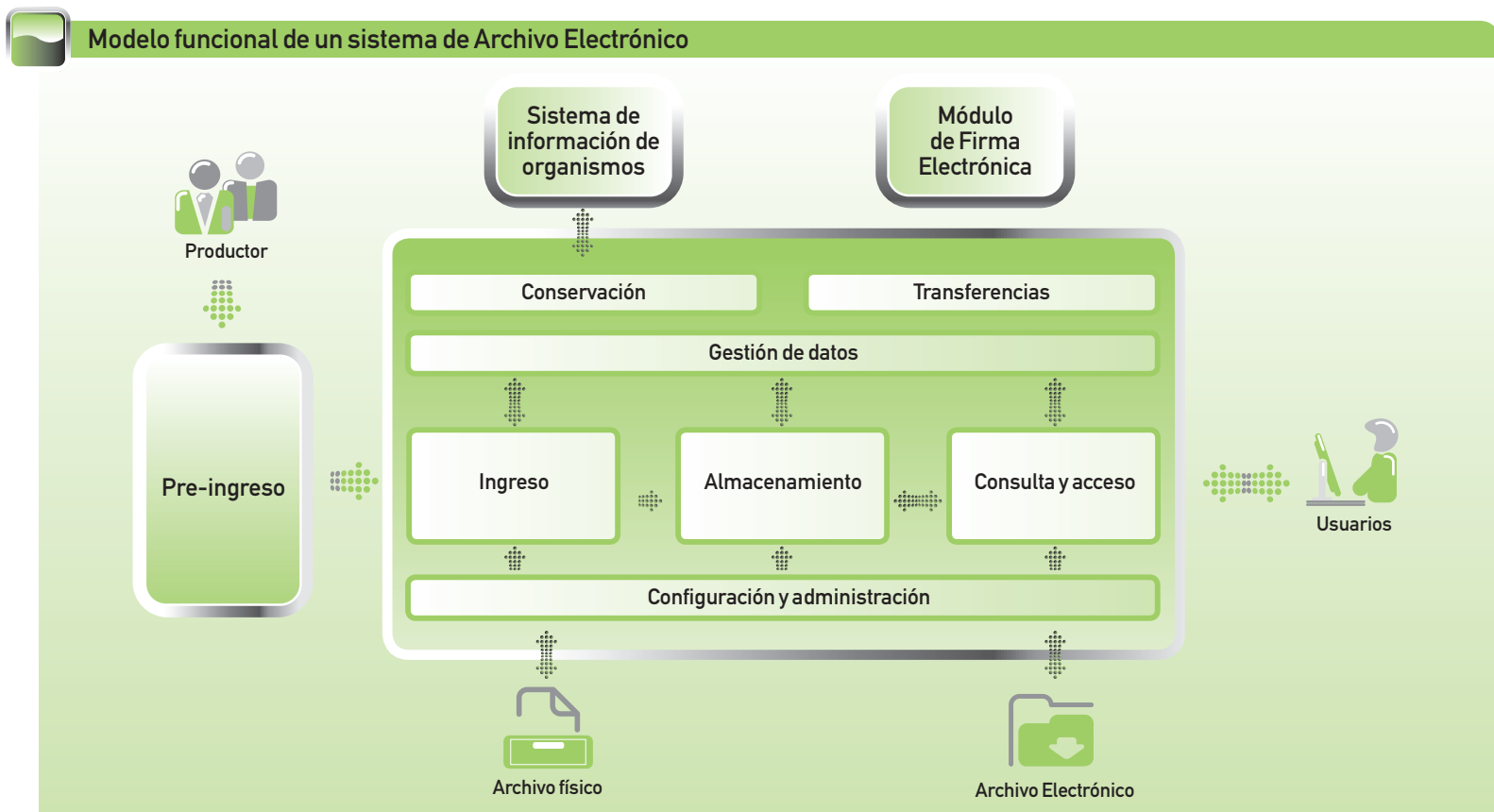
Para facilitar el traspaso de información entre los distintos Archivos que componen esta jerarquía, se ha extendido el modelo definido por OAIS con un nuevo módulo que agrupará la funcionalidad relativa a la propuesta de transferencias de documentación, su aceptación por parte del Archivo destino y la transferencia de la información.

Hay que destacar que esta funcionalidad podría haberse incorporado a otros módulos existentes en OAIS, pero se ha considerado que tiene una importancia suficiente como para poder se tratada independientemente.



5.2.3. Modelo funcional propuesto

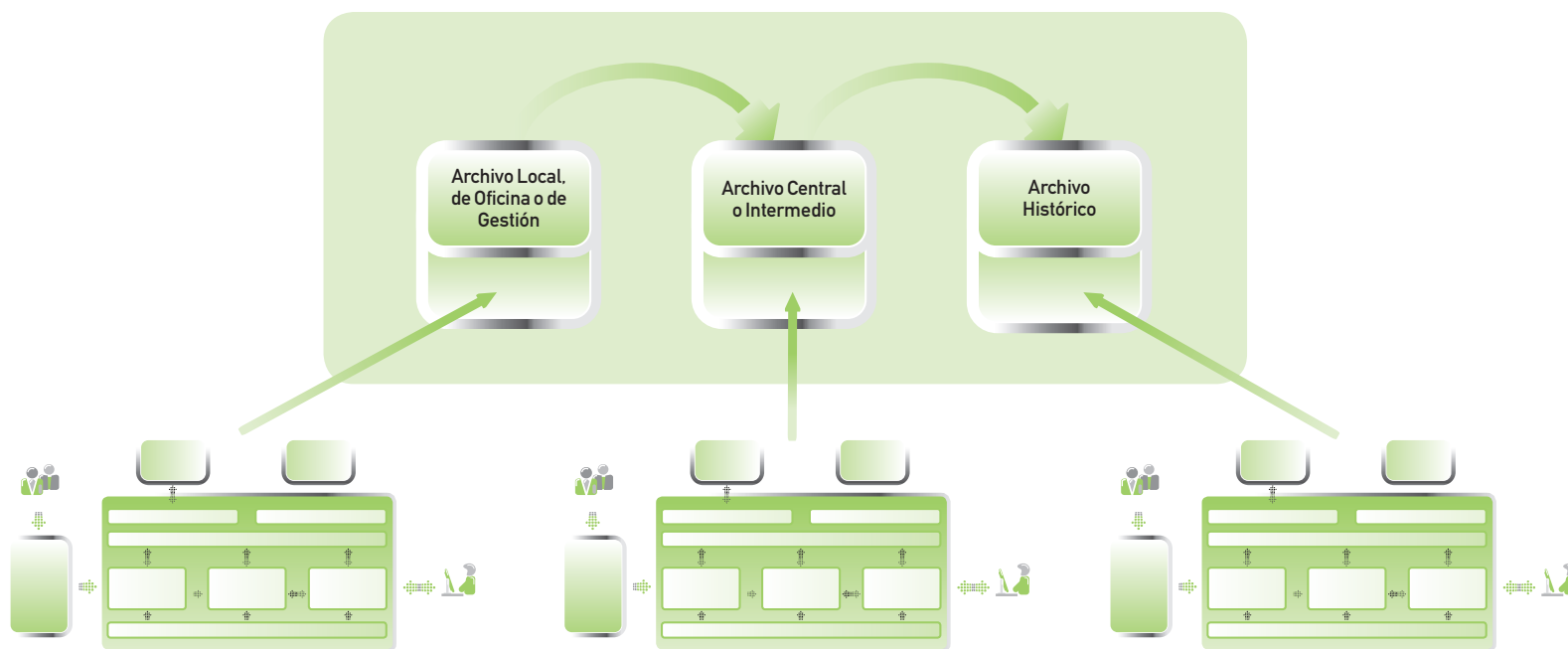
Modelo funcional de un sistema de Archivo Electrónico.



Hay que destacar que este esquema funcional se corresponde con la implantación únicamente de uno de los Archivos que componen la estructura jerárquica (Archivo de Oficina, Archivo Central / Intermedio, Archivo Histórico). El siguiente gráfico ofrece una visión de conjunto del modelo.



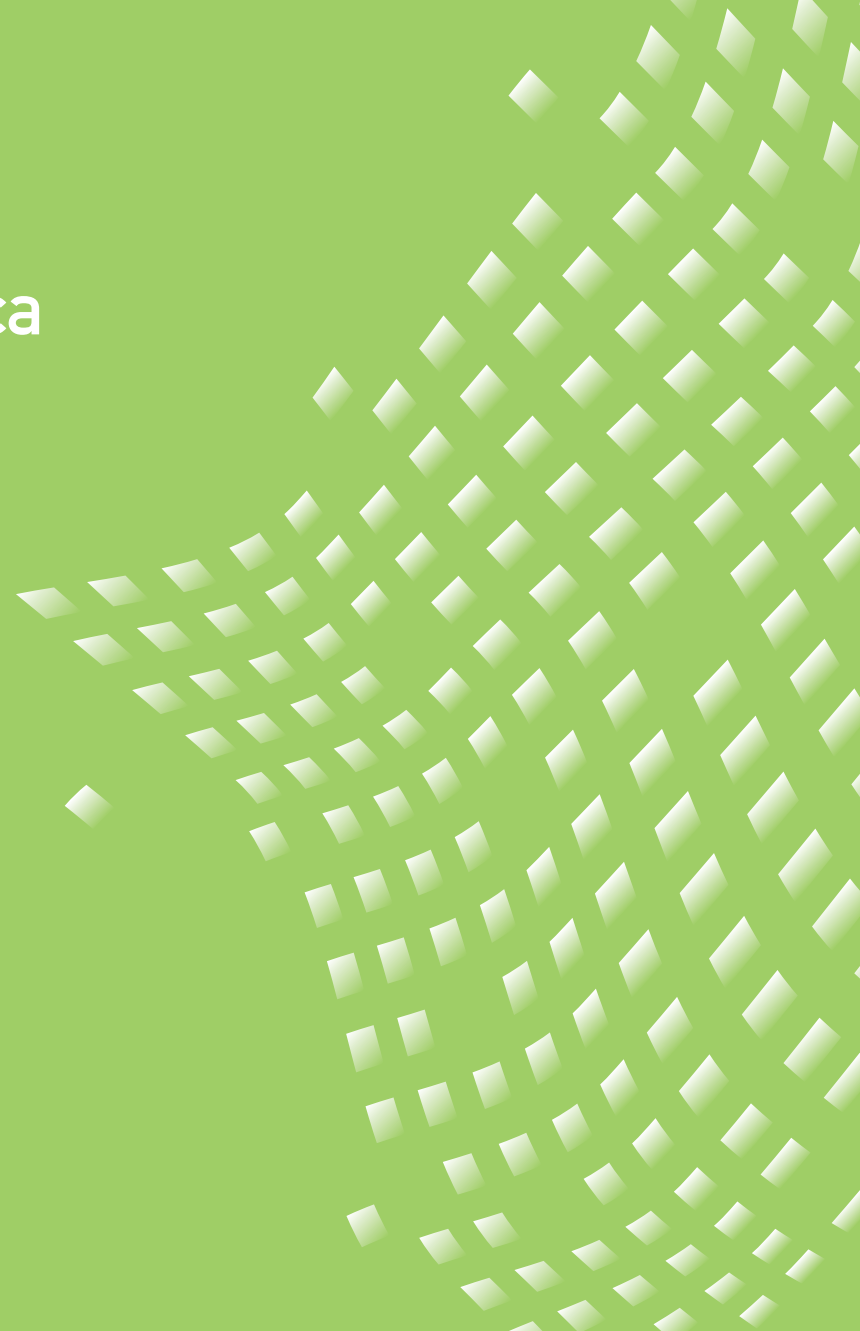
Relación del Modelo Conceptual con el Modelo Funcional.





06.

Implementación tecnológica del Modelo Conceptual



6. Implementación tecnológica del Modelo Conceptual

6.1. Alcance de la implementación

Una vez definido el Modelo Conceptual, los trabajos abordados se han dirigido a determinar la **viabilidad técnica de una implantación real** empleando únicamente componentes de software de fuentes libres.

Así, se ha realizado un análisis del mercado orientado a comprobar la existencia de soluciones open-source consolidadas, que proporcionasen las funcionalidades requeridas en el modelo propuesto y contrastado con las Administraciones Públicas.

Como resultado, se ha generado un inventario de componentes comerciales, en base al cual se ha elaborado el diseño de un posible piloto de Archivo Electrónico.

Es importante indicar que **este diseño no debe verse como una propuesta cerrada**, sino como una validación teórica de que realmente es factible acometer la construcción de un Archivo Electrónico plenamente operativo basado exclusivamente en software open-source.

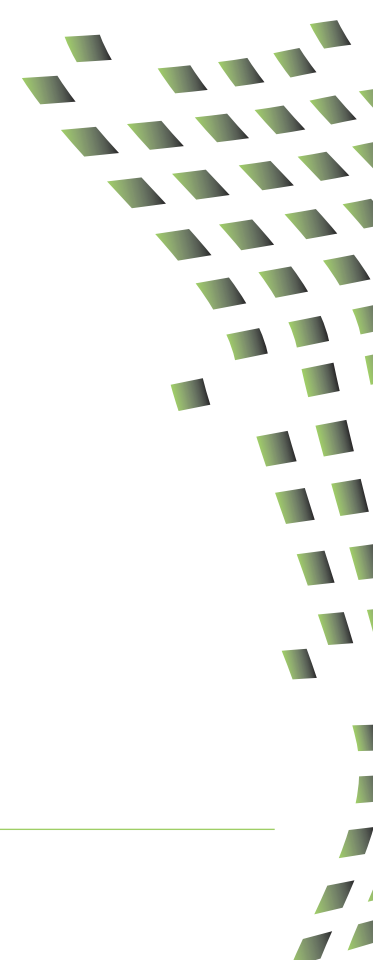
Por ello, durante el diseño, se ha optado por primar la adopción de estándares y modelos reconocidos sobre la implantación de aplicaciones concretas. De este modo, se persigue dotar al diseño de una mayor flexibilidad y favorecer una mayor independencia respecto a tecnologías y proveedores concretos, facilitando la generación de otros diseños a partir de componentes alternativos.

6.2. Contexto tecnológico

A diferencia de otros componentes de la Administración Electrónica, como el Registro Electrónico o la firma digital, el Archivo Electrónico es un campo donde se han hecho pocos avances. Así, en la actualidad, **apenas existen aplicaciones productivas de Archivo orientadas a la tramitación administrativa.**

Como ejemplos cabe mencionar:

- **@rchivA**, impulsada por la Junta de Andalucía.
- **SIGEM**, impulsada por el MITyC.



En general, estos productos implementan mecanismos derivados del archivo tradicional, basado en soportes en papel, adaptados para el almacenamiento de ficheros electrónicos.

Esta aproximación, aunque puede solventar la problemática actual, no proporciona la suficiente flexibilidad y capacidad de expansión que el nuevo escenario vinculado a la Administración Electrónica necesitará en los próximos años.

Sin embargo, a diferencia de en el sector público, el uso de repositorios de archivos digitales sí está muy extendido en el ámbito académico y científico, dónde se utilizan habitualmente para la publicación y compartición on-line de documentación técnica.

Muchas de estas aplicaciones, impulsadas principalmente desde universidades, han sido liberadas como software de fuentes abiertas, lo que ha favorecido la formación de comunidades activas de usuarios que colaboran para mejorarlas y desarrollar nuevas funcionalidades.

Gracias a este trabajo, estos sistemas han evolucionado enormemente durante la última década, y pueden considerarse suficientemente maduros y evoluciona-

dos para considerar su adopción en el entorno de la Administración Pública.

Hay que tener en cuenta que las diferencias entre estos dos ámbitos, impuestas principalmente por el marco normativo, van a suponer que estas herramientas no incorporen toda la funcionalidad requerida por la Administración, por lo que se hará necesario complementarlos con otros componentes que amplíen su funcionalidad para producir una solución de archivo consistente y adecuada a la legalidad vigente.

6.3. Soluciones de repositorio existentes basadas en software open-source

Como parte del trabajo acometido se han seleccionado y analizado diversas aplicaciones de repositorio digital basadas en software de fuentes abiertas. Estos repositorios se orientan a la conservación de cualquier tipo de documento y a su publicación a una comunidad de usuarios.

Dado que el objetivo es la conservación de documentos en su versión definitiva, el estudio se ha centrado en productos orientados a la preservación de la infor-

mación y a facilitar su búsqueda y consulta por parte de distintos usuarios. De este modo, se han primado aplicaciones que:

- Permitan la incorporación de metadatos para clasificar y contextualizar los contenidos archivados.
- Implementen mecanismos de búsqueda eficaces sobre la información archivada.
- Dispongan de un modelo de datos que agrupe contenidos y metadatos como una única entidad.
- Faciliten la presentación de la información a los usuarios, evitando su modificación.
- Posibiliten la implantación de mecanismos de preservación y actualización de la información.

Hay que indicar que aunque en la actualidad se están empleado soluciones de Gestión Documental para el archivo de documentos electrónicos, se ha optado por centrar el estudio en aquellas diseñadas específicamente como repositorio.

En concreto, se han seleccionado las cuatro aplicaciones open-source de repositorio predominantes a nivel mundial. Estas aplicaciones, impulsadas a través de diversas universidades y centros de investigación, cuentan con una amplia aceptación y han sido implantadas en numerosos organismos, empresas e insti-

tuciones. Gracias a ello, existe una gran cantidad de experiencias documentadas que permiten contrastar su rendimiento y funcionalidad.



Hay que indicar que existen otros repositorios con funcionalidad similar pero con una difusión mucho menor, por lo que se consideran productos minoritarios y menos adecuados.

Junto con las cuatro aplicaciones anteriores, se ha estudiado la validez de tres herramientas comerciales cuyo código fuente ha sido liberado por las empresas desarrolladoras.



A posteriori, el estudio se ha completado con dos soluciones más (iArxiu y Alfresco) identificadas durante las reuniones de contraste con las Administraciones Públicas.

A partir de este estudio comparativo, se han identificado qué soluciones ofrecen mejores condiciones para conformar el núcleo funcional del diseño del piloto, valorando, entre otros, los siguientes aspectos:

- Funcionalidad proporcionada y necesidad de incorporar componentes complementarios.
- Capacidad de integración con otros sistemas.
- Facilidad para el desarrollo de nuevas funcionalidades.
- Existencia de comunidades establecidas.
- Documentación y soporte técnico disponible.
- Experiencias de implantaciones previas.

Hay que destacar que, aunque sólo DSpace, Fedora Commons y ePrints cumplen con el modelo OAIS, todas las herramientas proporcionan, en mayor o menor medida, la funcionalidad básica requerida por este modelo, en concreto:

- Un **productor** de información será capaz de:
 - ❖ Añadir nuevos objetos digitales al repositorio para su almacenamiento.
 - ❖ Informar de los metadatos asociados al objeto.

- Los **usuarios** del sistema podrán:
 - ❖ Realizar consultas en función de ciertos criterios y generar listados de resultados.
 - ❖ Recuperar una copia del objeto digital para su visualización.
- Finalmente **un administrador del sistema** podrá:
 - ❖ Añadir, modificar y eliminar metadatos asociados a un objeto digital.
 - ❖ Establecer permisos de acceso a los objetos digitales archivados.
 - ❖ Generar informes de operación del sistema.

6.3.1. DSpace

Es una herramienta desarrollada por el Instituto Tecnológico de Massachusetts (MIT), en colaboración con HP, para la implantación de su propia librería digital. La primera versión del código fue liberada en 2002.

Ha sido ampliamente adoptado por distintos organismos en todo el mundo (se calculan unos 350 repositorios operativos), siendo la más extendida de las aplicaciones consideradas. En la actualidad, cuenta con el soporte de una notable comunidad de usuarios y desarrolladores, muy ligada al ámbito académico.

El paquete DSpace se presenta como una solución completa, proporcionando toda la funcionalidad necesaria de un repositorio digital, incluyendo interfaces para el ingreso de ficheros, administración del sistema, consultas sobre el repositorio, etc.

Antes de su ingreso, el sistema encapsula cada fichero electrónico junto con sus metadatos en un paquete único. Este paquete, que tiene un identificador único, es la unidad de almacenamiento del sistema.

El hecho de que la herramienta haya estado orientada desde el principio a la publicación de documentos está limitando su adaptación para otros propósitos.

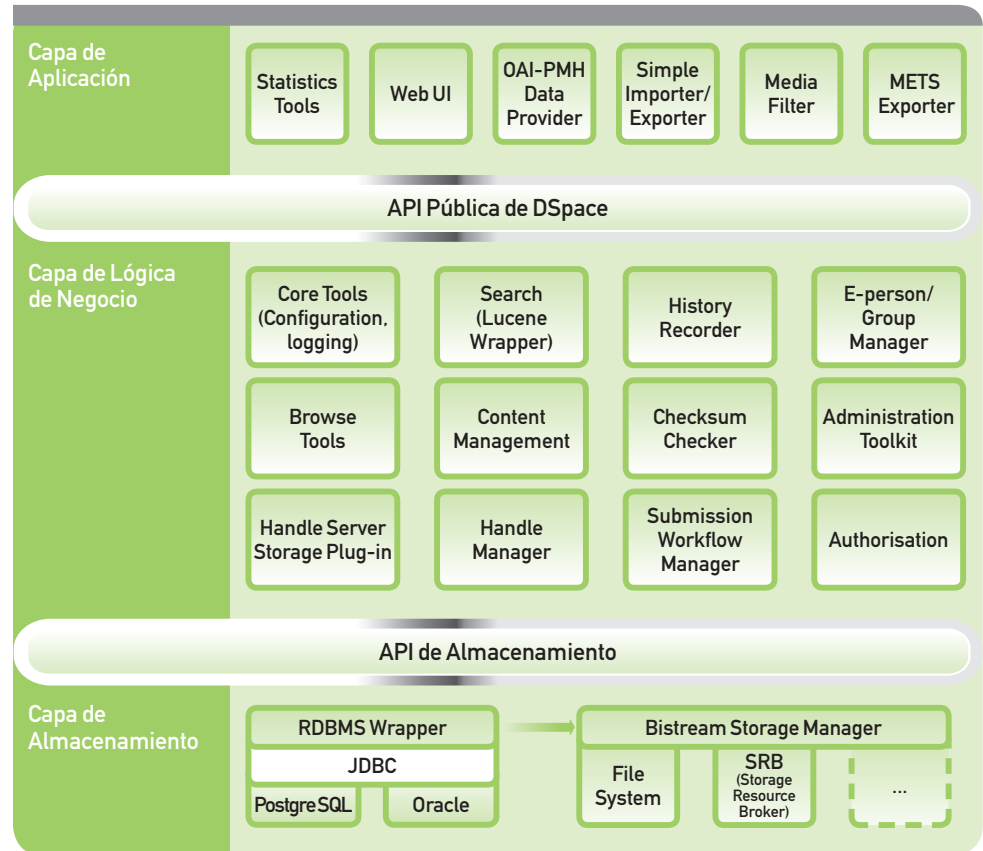
La aplicación se ha construido sobre una arquitectura de tres capas, formadas por módulos y componentes independientes, lo que favorece un encapsulamiento de la funcionalidad.

El core de la aplicación está formado por los módulos ubicados en la capa de negocio. La inclusión de nueva funcionalidad en el core se hace a través del desarrollo de plug-ins.

La inclusión de nuevas funciones en la capa de aplicación, se hace a través de add-ons. Para ello, incor-



Arquitectura modular de DSpace.



para un framework de desarrollo, llamado Manakin, que proporciona a los desarrolladores una serie de interfaces para acceder a la funcionalidad del core del repositorio.

La comunicación entre los distintos componentes de la aplicación se realiza a través de API's.

Como aspecto negativo indicar que, hasta la fecha, no se ha potenciado el desarrollo de mecanismos de interoperabilidad basados en Web Services, lo que dificulta su integración con otros sistemas informáticos.

Actualmente, no dispone de componentes para la implantación de políticas de preservación sobre la información.

6.3.2. Fedora Commons

Se trata de un proyecto iniciado por la Universidad Cornell y la Universidad de Virginia.

Su implantación no está tan extendida como otras aplicaciones (DSpace o ePrints) pero cuenta con una importante comunidad detrás, a la que da soporte una estructura organizativa consolidada y estable.

Dentro de la filosofía de Fedora Commons está proporcionar una plataforma escalable y suficientemente flexible para adecuarse a los requisitos de distintos tipos de organizaciones.

A partir de la versión 3.0, los desarrolladores de Fedora Commons han hecho un importante esfuerzo para adecuar el modelo de objeto digital soportado por el sistema, de modo que sea posible gestionar casi cualquier tipo de contenido.

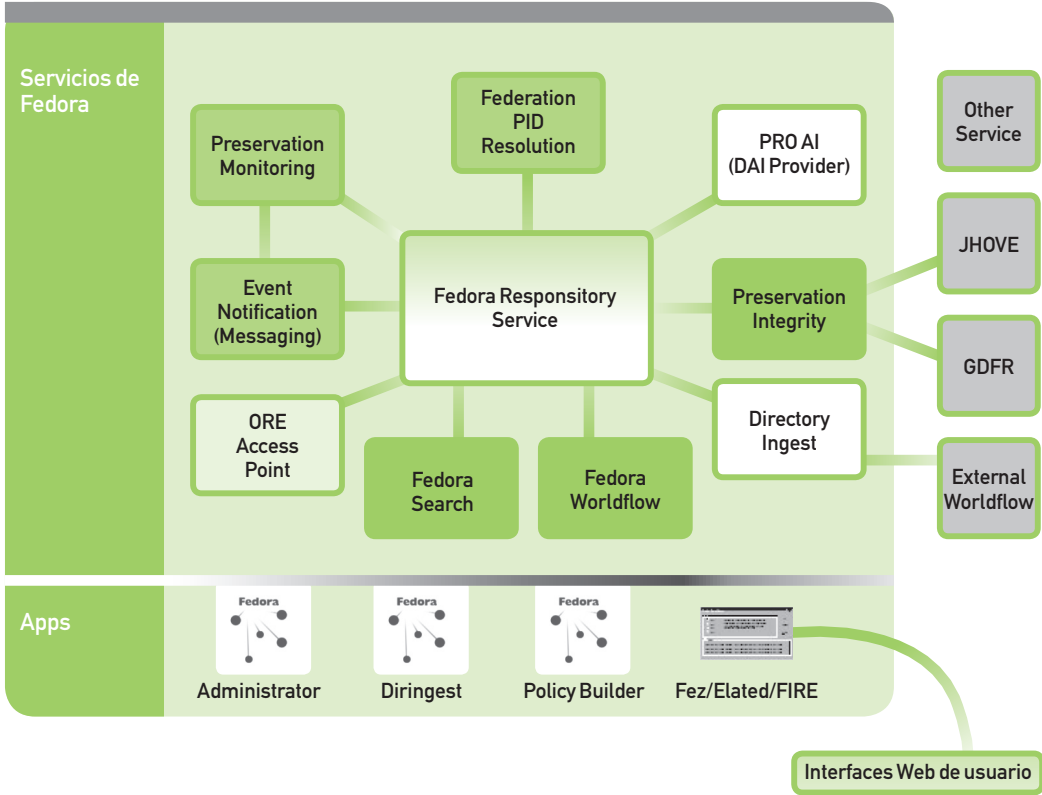
La evolución de la aplicación se estructura a través de dos vías:

- Constitución de grupos de trabajo especializados, orientados a la solución de problemáticas concretas del sistema (desarrollo de mecanismos de preservación de la información, gestión de metadatos, etc.).
- Colaboración con otros proyectos e iniciativas externas con objeto de generar sinergias y favorecer la evolución tecnológica de la aplicación.

El proyecto Fedora Commons se orienta a proporcionar una arquitectura modular fundamentada en un core o núcleo (Repository service) que proporciona la funcionalidad básica del repositorio digital (ingreso, almacenamiento, consulta, etc.).



Arquitectura modular de Fedora Commons.



El objetivo es, una vez conseguido un core estable, las nuevas funcionalidades requeridas por los usuarios se implementan a través de módulos adicionales que hacen uso de los servicios ofrecidos por el core. Para facilitar esto, se ha establecido un framework de desarrollo.

Los servicios proporcionados por el core son accesibles a través de distintos interfaces implementados como Web Services:

- **API-A (SOAP / REST)** – proporciona servicios de acceso y recuperación de objetos digitales.
- **API-M (SOAP / REST)** – proporciona servicios de gestión y administración del repositorio como ingreso de objetos digitales, inserción de metadatos, etc.
- **Basic Search (REST)** – servicios de búsqueda por texto completo sobre los contenidos almacenados.
- **RI Search (REST)** – servicios de búsqueda sobre el Índice de Recursos del sistema (basado en metadatos).
- **OAI-PMH provider** – servicios para la publicación de metadatos a otros repositorios.

Esta aproximación se adecua bien al modelo de Archivo Electrónico propuesto, dando solución a gran parte

de la funcionalidad requerida de los módulos de Ingreso, Gestión de Datos, Almacenamiento y Consulta.

Esta funcionalidad se completaría mediante desarrollos particulares que permitirían incorporar otros procesos complementarios como validaciones, extracción de metadatos, generación de representaciones de los contenidos, etc.

Fedora Commons no implementa una interfaz de usuario integrada en la aplicación. En la actualidad hay diversas iniciativas que han desarrollado sus propios interfaces gráficas (Fez, Muradora, entre otras) y los han liberado como software open-source. Estos interfaces implementan la funcionalidad requerida para el tratamiento de los objetos digitales previo a su ingreso.

Como en el caso de DSpace, los objetos digitales se encapsulan junto con sus metadatos, generando una estructura única que actúa como unidad de información en el sistema.

6.3.3. ePrints

Es un proyecto desarrollado inicialmente por la School of Electronics and Computer Science de la Universidad de Southampton (UK), y con una amplia difusión

principalmente en el entorno académico. Se estima que actualmente existen unos 270 repositorios que lo implementan.

Sus ventajas son una gran facilidad de uso y su rapidez de implantación, lo que lo han hecho muy popular. Además, una misma instancia de ePrints puede ejecutar distintos repositorios concurrentemente.

ePrints permite la definición de estructuras de directorios en los repositorios para facilitar la catalogación de la información de acuerdo a las necesidades de la organización.

La aplicación está muy orientada a su uso como repositorio de documentación, función que realiza eficazmente, pero apenas ofrece facilidades para su adaptación a otros usos. Como ejemplo, se puede decir que, si bien facilita tareas como el ingreso de objetos digitales en el repositorio, no permite validaciones previas como la comprobación de firma electrónica.

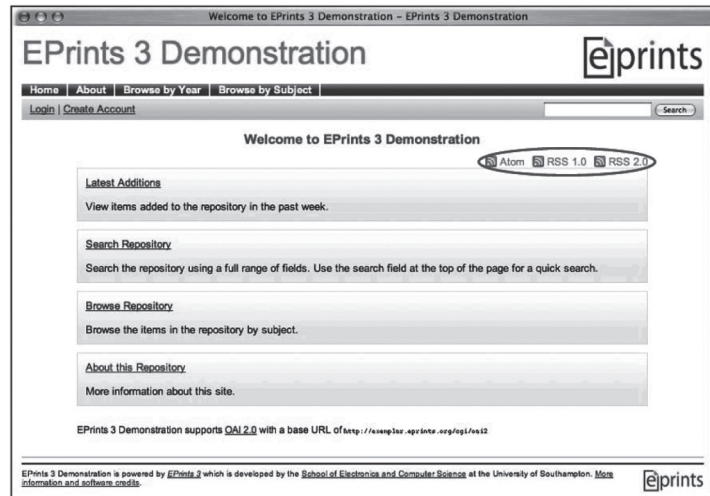
La configuración del sistema se realiza editando los ficheros de configuración, lo que lo hace más complejo. Actualmente, se está trabajando en el desarrollo de una interfaz gráfica.

A nivel de usuario, dispone de una interfaz gráfica web y una interfaz por línea de comandos que ofrece servicios de ingreso y recuperación de ficheros. La interfaz web permite su personalización para adecuarla al look & feel de la organización.

La evolución de la aplicación está basada en el desarrollo de plug-ins, no implementando una interfaz basada en Web Services. Esto supone una importante limitación a la hora de integrar el repositorio en una plataforma tecnológica existente.



Ejemplo del interfaz de usuario en ePrints.



Permite exportar los resultados de las búsquedas en distintos formatos a través de múltiples medios de sindicación (RSS, Atom).

La aplicación no proporciona mecanismos para facilitar la ejecución de políticas de preservación.

6.3.4. Greenstone Digital Library

Greenstone surgió como proyecto a partir de una iniciativa de la National Library of New Zealand (NLNZ) con sede en la Universidad de Waigo, a mediados de los 90. Es una de las iniciativas pioneras en este tipo de soluciones y, de todas las aplicaciones consideradas, es la que tiene un mayor recorrido temporal.

La comunidad de Greenstone colabora con la ONG Human Info y con la UNESCO para extender el uso de repositorios digitales en países en vías de desarrollo.

Su implantación está menos difundida que las otras soluciones (unos 63 repositorios según su web), con gran influencia en África y Asia.

Cuenta con una comunidad de usuarios significativa, principalmente archiveros, siendo mucho más limitado el número de desarrolladores que participan en el

proyecto. De este modo, las evoluciones del software se llevan a cabo por un pequeño grupo de personas que han participado en el proyecto desde sus inicios.

La colaboración con la UNESCO ha favorecido la traducción del paquete completo, no sólo de la interfaz gráfica, a múltiples idiomas, entre ellos el castellano. Además se ha favorecido su compatibilidad con múltiples entornos tecnológicos (Linux, Windows, Solaris, Mac OS-X, FreeBSD).

Aunque implementa las funcionalidades básicas recogidas en los módulos de Ingreso, Almacenamiento y Consulta del modelo OAIS, su diseño no está basado en esta referencia.

En la última versión (v 3.0), se ha rediseñado la aplicación para favorecer la modularidad y la comunicación entre componentes vía XML. Los objetos digitales se empaquetan y almacenan como Greenstone Archival Format, que es un encapsulamiento XML que permite marcar secciones del documento e incluir metadatos tanto a nivel de documento como de sección.

El sistema utiliza plug-ins que implementan las funciones específicas que son invocadas por los procesos

principales. De este modo es posible añadir nueva funcionalidad al sistema, por ejemplo, la rutina de procesamiento de un nuevo tipo de formato.

6.3.5. OpenEditDam / Razuna / ResourceSpace

Del estudio independiente de estas tres aplicaciones corporativas, se puede concluir que existe una gran similitud entre ellas, ya que proporcionan una funcionalidad parecida y tienen ventajas y limitaciones similares.

Por ello, se ha optado por presentarlas de manera conjunta, destacando las diferencias encontradas en durante la comparativa.

Las tres herramientas han sido desarrolladas por empresas comerciales que han decidido liberar el código y proporcionarlos de manera gratuita, centrandos su modelo de negocio en servicios de soporte especializado y hosting. Todas ellas se orientan hacia la compartición y publicación web de ficheros de una manera sencilla para organizaciones de tamaño medio.

Su desarrollo como aplicaciones monolíticas favorece instalaciones sencillas, que no necesitan de avanzados conocimientos técnicos y permiten una rápida puesta en marcha del repositorio.

Por el contrario, esta aproximación no facilita la adaptación del código para incorporar nuevas funcionalidades. En este sentido, sería preferible un diseño orientado a módulos que agrupasen funciones específicas y la definición de interfaces entre ellos.

Sólo ResourceSpace implementa un mecanismo de ampliación, basado en plug-ins que permite incorporar modificaciones sencillas.

Ninguna de las tres aplicaciones está fundamentada en el modelo de referencia OAIS.

Aunque todas ellas permiten ingresar múltiples tipos de ficheros, su orientación principal es hacia la publicación de ficheros multimedia (imágenes, vídeo, audio). De este modo, proporcionan facilidades para generar versiones de baja resolución a modo de previstas, para la transformación de formatos y distintos reproductores dentro de la propia aplicación.

Por el contrario, el conjunto de metadatos es limitado y muy orientado hacia este tipo de contenidos multimedia. En concreto, las tres aplicaciones tienen soporte para metadatos XMP / IPTC y sólo ResourceSpace permite además campos Exif.

Hay que mencionar que este tipo de metadatos se escribe en el propio fichero del objeto digital, así que la modificación de cualquier metadato supone la alteración del objeto digital, lo que puede complicar la gestión de documentos firmados digitalmente.

Aunque el esquema permite la personalización incluyendo nuevos campos de metadatos de forma manual, esta aproximación no se considera operativa cuando se manejan grandes cantidades de objetos. Esto supone una grave limitación a la hora de incorporar información sobre preservación o aspectos técnicos.

Los tres presentan una interfaz web intuitivo y sencillo que favorece un aprendizaje rápido. Sin embargo,

esto limita toda la interacción con el sistema al interfaz gráfico, lo que dificulta la interoperabilidad con otros sistemas y aplicaciones.

Finalmente, mencionar que, dado su orientación como gestores de contenido, no permiten implementar políticas complejas de autorización y acceso a los objetos digitales.

Tampoco permiten establecer relaciones entre distintos objetos almacenados en el repositorio.

La siguiente tabla recoge diversas características particulares de cada una de las aplicaciones comentadas:



OPENEDITDAM	RAZUNA	RESOURCESPACE
<ul style="list-style-type: none"> /// Desarrollado por OpenEdit. Inc. /// Distribuido bajo licencia GNU Lesser General Public Licence. /// Desarrollo Java sobre OpenEdit Framework. /// Opciones de configuración y gestión del sistema muy limitadas. /// Contempla la integración con LDAP. 	<ul style="list-style-type: none"> /// Desarrollado por SixSigns. /// Distribuido bajo GPL Open Source Licence. /// Desarrollado en Java. /// Implementa un API básica. /// Permite conversiones de formato on-the-fly. /// Equipo de desarrollo reducido. /// Documentación técnica escasa y poco explicativa. 	<ul style="list-style-type: none"> /// Desarrollado por Montala. /// Distribuido bajo licencia BSD. /// Desarrollo PHP / MySQL. /// Admite plug-in para incluir pequeñas modificaciones de la funcionalidad. /// Permite "archivar" ficheros (en la práctica los elimina de los resultados de las búsquedas) /// No permite configurar permisos a nivel de usuario.

6.3.6. iArxiu

iArxiu es una aplicación impulsada por la Generalitat de Catalunya a través de la Agència Catalana de Certificació (CATCERT) y que se orienta a proporcionar servicios de Archivo Intermedio y Archivo Histórico para documentos electrónicos.

Desde hace un tiempo CATCERT ofrece la primera versión de la solución (que ofrece una funcionalidad de archivo de documentos electrónicos básica) a las Entidades Locales de Cataluña para su uso, a través de acuerdos de colaboración.

En paralelo, el software a sido completamente revisado y rediseñado, dando lugar a una nueva versión (iArxiu v.2.0), que acaba de salir a producción. Esta nueva versión ha sido definida tomando como referencia OAIS e implementa gran parte de la funcionalidad definida en el modelo.

Esta aplicación es una de las pocas soluciones existentes de Archivo Electrónico planificadas y diseñadas para cubrir los requerimientos reales de la Administración Pública, dando respuesta a la problemática asociada a la preservación de la información digital a largo plazo. Las características principales de la nueva aplicación

fueron avanzadas durante la reunión de contraste con el personal de la Generalitat de Catalunya, varias semanas antes de su puesta en producción. Pendientes aún de un estudio más exhaustivo, a priori parece viable su integración como parte de un futuro piloto de Archivo Electrónico.

Como inconveniente, indicar sólo que la aplicación **no está disponible aún como software de fuentes abiertas**, por lo que no se ha incluido en el estudio realizado. CATCERT ya ha manifestado su interés por liberarla a corto plazo y facilitar la constitución de una forja que de soporte a su evolución.

6.3.7. Alfresco

Fruto de las entrevistas de contraste realizadas, se ha constatado que distintas Administraciones Públicas están adoptando Alfresco como una solución de Archivo Electrónico a corto y medio plazo, por lo que se ha considerado interesante incluirlo en el estudio realizado.

Alfresco es una aplicación de Gestión Documental open-source diseñada para facilitar la gestión colaborativa de documentos electrónicos. Sus objetivos son:

- Favorecer la creación y versionado de documentos en entornos compartidos.
- Favorecer la clasificación de los documentos electrónicos archivados.
- Facilitar el acceso a estos documentos electrónicos.

En este aspecto, no puede considerarse un repositorio digital en los términos que se han empleado hasta ahora.

Sin embargo, es posible extender su funcionalidad básica añadiendo ciertos mecanismos para la conservación de documentos electrónicos. Estos mecanismos se orientan a mejorar la catalogación y monitorización de la información, en ningún caso abordan la problemática de la obsolescencia de formatos y soportes. En vista de esto, Alfresco sólo resulta viable como implantación del Archivo Local o de Oficina.

La interacción del usuario con Alfresco se realiza a través de una interfaz web, por lo que no es necesario la instalación de ningún aplicativo cliente. Además es posible ingresar, mover o borrar ficheros a través de interfaces FTP, WebDAV y CIFS.

Los documentos electrónicos se almacenan en el sistema dentro de contenedores (**spaces**) que permiten la definición de **reglas**. Estas reglas permiten la ejecución de ciertas acciones en respuesta a un evento.

Alfresco define los documentos electrónicos de archivo (**records**) como los documentos incluidos en unos spaces especiales llamados **filesplans**, que añaden la siguiente funcionalidad extra:

- Permiten la definición de periodos de conservación para los records almacenados. Pasado este periodo, el record se marca como obsoleto.
- Incorporan nuevos metadatos para ampliar la información relacionada con el record.
- Añade nuevos tipos de reglas relacionadas con la gestión de archivos (eliminación de records cuando quedan obsoletos, extracción automática de metadatos en el ingreso).

Alfresco gestiona un conjunto predefinido de metadatos (propiedades). Estos metadatos pueden utilizarse como parámetros de búsqueda avanzada sobre el repositorio. Es posible configurar nuevos campos de metadatos editando uno de los ficheros de configuración de la aplicación. Estas modificaciones son comunes a todos los records, no permitiendo por este método la

definición de metadatos particulares según el tipo de contenido almacenado.

Sólo cuando el ingreso del documento electrónico se hace vía interfaz de usuario es posible informar todos los metadatos asociados. En el resto de opciones (FTP, WebDAV, CIFS), sólo podrán incorporarse aquellos metadatos que puedan extraerse automáticamente.

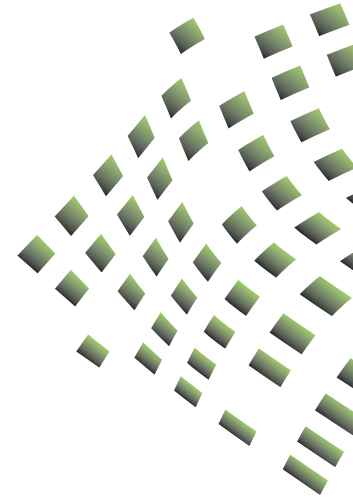
Los metadatos en Alfresco se almacenan en una Base de Datos independiente, separados de los ficheros electrónicos a los que se refieren. Esta característica puede resultar problemática a la hora de transferir ficheros a otro Archivo Electrónico ya que habría que implementar la forma de adjuntar los metadatos junto con los documentos a transferir.

6.4. Diseño de la implantación piloto

A partir de los resultados de la comparativa entre las distintas soluciones analizadas, se ha considerado que Fedora Commons ofrece una serie de características que facilitarán el diseño de una implantación piloto:

- Es una herramienta basada en el modelo OAIS, lo que favorece su alineamiento con el modelo conceptual de Archivo Electrónico.
- Cuenta con el soporte de una nutrida comunidad.
- El diseño de Fedora Commons se orientado hacia una arquitectura modular, lo que facilita el desarrollo de nuevos componentes y su integración con los ya existentes.
- Implementa interfaces vía Web Services, por lo que puede integrarse dentro de una arquitectura SOA.

Esto no quiere decir, que la implantación del sistema de Archivo Electrónico deba estar basada en Fedora Commons. Este diseño tecnológico es principalmente un ejercicio teórico, en el que se toma una aplicación tipo de repositorio digital, en este caso Fedora Commons, y se contrasta con el modelo conceptual presentado anteriormente.



A partir de este contraste, se determinarán las carencias funcionales típicas y esperables, y se identificarán componentes de software de fuentes abiertas que puedan solventarlas.

Hecha esta puntualización, **nada impide que, llegada la hora de abordar un diseño real, sea preferible adoptar otro sistema diferente, como por ejemplo iArxiu, como núcleo del futuro sistema.**

Entre las carencias típicas fruto del contraste del modelo con la aplicación de repositorio digital cabe esperar las siguientes:

- Implementación de Cuadros de Clasificación para la catalogación de los documentos.
- Existencia de interfaces adecuados de usuario.
- Mecanismos de transferencia de documentos entre distintos Archivos.
- Mecanismos para la monitorización y evaluación de formatos de fichero.
- Sistemas de control de accesos suficientemente restrictivos.
- Visualizadores adecuados para distintos tipos de contenidos digitales.

Para el análisis se utilizará la definición funcional del modelo presentada en el capítulo anterior y que segmentaba el sistema de Archivo en módulos con funcionalidad similar.

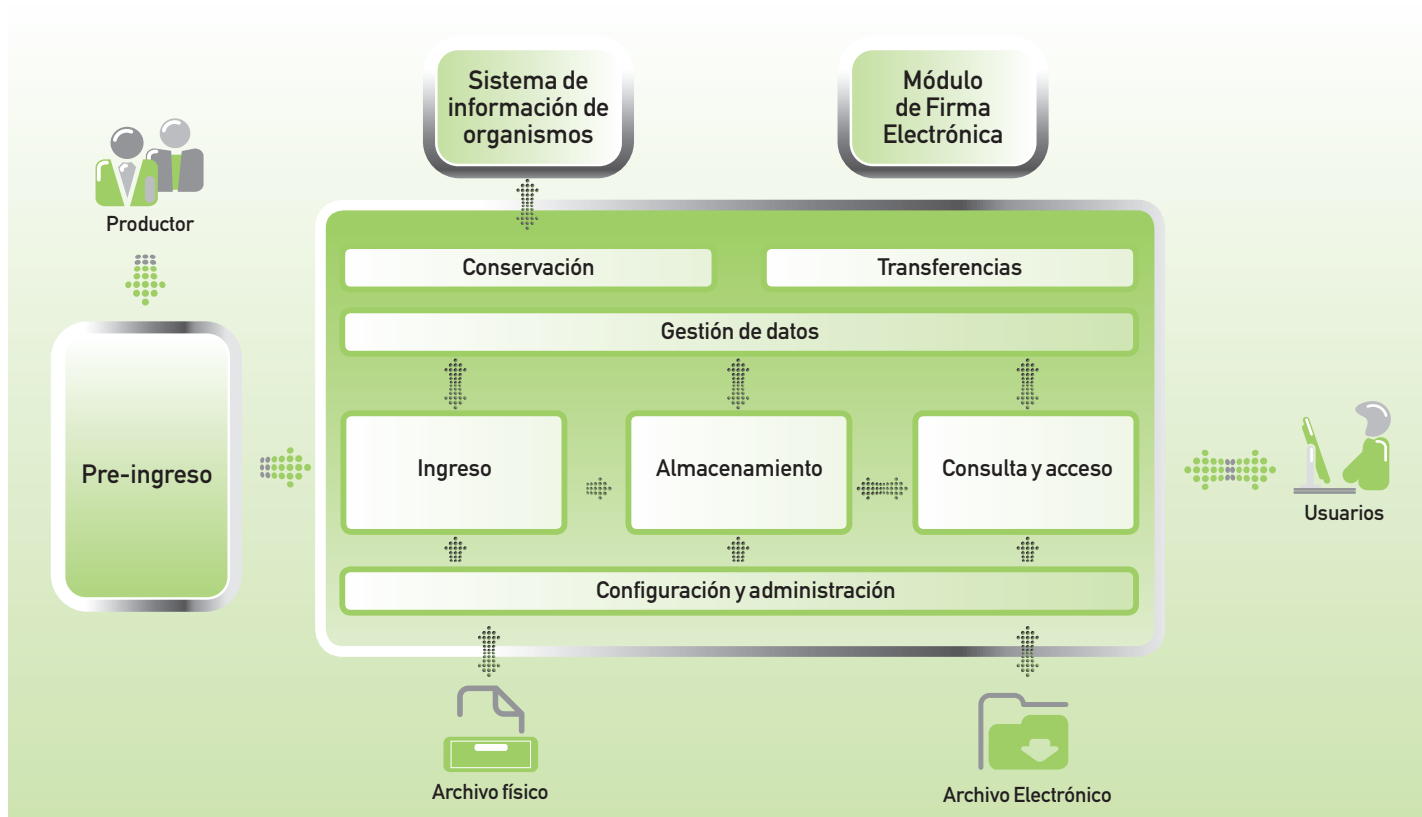
De este modo, es posible analizar en que grado la herramienta de repositorio cubre los requerimientos de cada uno de los módulos y determinar aquellas funcionalidades que se deberán implementar a través de componentes adicionales de software de fuentes abiertas.

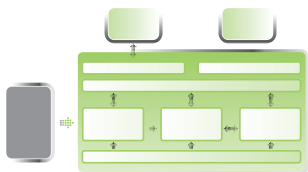
Como se ha comentado antes, la elección de las componentes concretos pretende ser meramente orientativa, prefiriéndose la adopción de estándares reconocidos que definan procedimientos para el intercambio de datos entre los componentes del sistema.

En las secciones siguientes, se aborda cada uno de los distintos módulos funcionales del modelo.



Modelo funcional de un sistema de Archivo Electrónico.





6.4.1. Módulo de pre-ingreso

Su propósito es facilitar la transferencia de datos entre los productores y el Archivo, estableciendo los criterios formales para la admisión de los contenidos en el sistema. Este módulo no forma parte del sistema de Archivo Electrónico. Sus funciones se integrarán dentro de las aplicaciones de Gestión Documental de la organización o través de una aplicación dedicada que sirva como interfaz de usuario para el ingreso manual de documentos.

La entrada del módulo serán los ficheros electrónicos con los contenidos a preservar, así como información complementaria (metadatos) que facilite la clasificación e identificación de los contenidos para su conservación y recuperación futura.

De forma genérica, la información requerida al proveedor será:

- Metadatos descriptivos: nombre del objeto, nombre de la unidad que lo produce, tipología, etc.
- Metadatos técnicos: relativos al formato de fichero empleado, a la firma digital, etc.
- Metadatos estructurales: referentes a la composición interna del documento o a las relaciones con otros documentos almacenados en el archivo.

- Metadatos preservación: correspondientes a periodos de vigencia, pistas de auditoría, etc.

Con objeto de favorecer la interoperabilidad y el intercambio de información entre repositorios, se recomienda el uso de esquemas de metadatos estandarizados y de amplia difusión.

La herramienta permitirá su configuración con distintos esquemas. Así, aunque el estándar de facto para metadatos descriptivos es Dublin Core, debería tenerse en cuenta la posibilidad de incluir otros complementarios. Del mismo modo, se considerará el uso de esquemas diferentes en función de la naturaleza del contenido a ingresar.

Tanto los contenidos como los metadatos se encapsularán juntos dentro de una estructura predeterminada, llamada paquete SIP (Submission Information Package), y que será la unidad de información que se ingresará en el Archivo.

Este paquete tendrá una estructura predefinida y adecuada para facilitar su procesamiento automatizado por el sistema de Archivo. En este sentido, la tendencia generalizada es el mercado es el encapsulamiento

to XML, principalmente basado en la especificación METS.

Con objeto de facilitar la preservación y recuperación de los contenidos en el futuro, se recomienda restringir el ingreso de documentos en el sistema a un conjunto reducido de formatos definidos previamente y consensuados entre productor y archivo.

Al final del documento se proporciona una lista de formatos recomendados, cuyas características los hacen adecuados para la preservación de los datos a lo largo del tiempo.

Asimismo, es importante destacar la problemática asociada al ingreso en el archivo de documentos firmados electrónicamente. La validez jurídica de estos documentos puede llegar a perderse si en el futuro no se dan las condiciones que permitan verificar la autenticidad de la firma.

Para evitar esto, se propone el resellado de toda la documentación recibida en el sistema de Archivo Electrónico, previa verificación de la validez de la firma digital. Para ello, el paquete SIP deberá incorporar tanto la firma del documento como los datos necesarios para su validación.

Este módulo deberá permitir clasificar los documentos a ingresar dentro de un esquema jerárquico (cuadro de clasificación).

Fedora Commons dispone de un módulo que facilita la creación de paquetes METS.

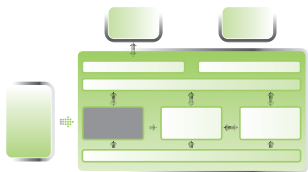
Asimismo, Fedora Commons no impone restricciones respecto a los distintos esquemas de metadatos que puede gestionar.

Estándares



Componentes open-source





6.4.2. Módulo de ingreso

Este módulo gestiona el ingreso de información en el sistema de archivo. Esta información debe recibirse correctamente estructurada en un paquete SIP.

A partir de la recepción del paquete SIP, el módulo realizará una serie de validaciones para asegurar la integridad y adecuación de la información a ingresar:

- Comprobación formal de la estructura del paquete SIP.
- Validez del objeto digital de acuerdo al formato de fichero indicado.
- Antivirus.
- Información de los metadatos requeridos.

A continuación, se validará la autenticidad de la firma de negocio (si existiese). En caso positivo, se aplicará un resellado temporal sobre el documento firmado y sobre la información empleada para la validación de la firma.

Con ello, el Archivo podrá “da fe” de la validez jurídica del documento en el momento del ingreso, de modo que quede constancia de su autenticidad incluso si en el futuro no pudiese repetirse esta verificación.

En los SIP que hayan superado el proceso, los objetos digitales será procesados para la extracción de nuevos

metadatos, principalmente de carácter técnico, que completen la información provista por el productor.

De nuevo, se recomienda el uso de esquemas de metadatos estandarizados como:

- Metadatos técnicos: MIX (para imágenes), AudioMD (para audio), TEI (para texto)
- Metadatos de preservación: PREMIS

Finalmente, el paquete SIP será transformado en un paquete AIP (Archival Information Package). El paquete AIP es la estructura en la que la información será almacenada por el repositorio. Este enfoque garantiza la preservación de los contenidos y sus metadatos asociados como de una misma unidad.

A cada paquete AIP se le asignará un identificador único y perdurable (PID) que se usará para referenciarlo en el futuro dentro del sistema.

Hay que mencionar que la correspondencia entre SIP y AIP no es necesariamente uno a uno, siendo posible que un SIP genere varios AIP y viceversa.

Fedora Commons implementa los servicios necesarios para el ingreso del SIP y la generación del AIP. Sin embargo,

deberá abordarse la integración de nuevos componentes que implementen las distintas validaciones propuestas, así como la extracción automatizada de metadatos.

Hay que destacar también entre las carencias identificadas, el hecho de que Fedora Commons no cumple actualmente con el estándar **JSR-170** (Content Repository for Java).

Esta norma especifica un interfaz estándar para repositorios de contenidos, lo que permite integrarlos fácilmente con cualquier aplicación de gestión de contenidos compatible. De este modo, se favorece la separación entre las aplicaciones de cliente y el repositorio, favoreciendo la sustitución de componentes y alargando la vigencia de la solución final.

JSR-170 define dos niveles de compatibilidad. En este caso, sería necesario implementar tanto el nivel 1 (lectura) como el nivel 2 (escritura en el repositorio).

La aproximación que se considera más factible sería el desarrollo de una capa intermedia entre la interfaz propietaria de Fedora y la API Java definida en el estándar.

Estándares



AUDIOMD

TEI

JSR-170

XACML

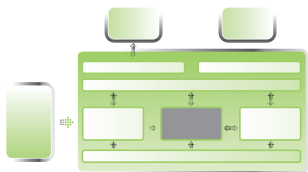
Componentes open-source



VTLS
Metadata
Extraction

NLNZ
Metadata
Extraction





6.4.3. Módulo de almacenamiento

El módulo de almacenamiento es el responsable de la conservación de los paquetes AIP, por lo que proporcionará mecanismos adecuados de back-up y de redundancia que aseguren la disponibilidad de la información.

Así mismo, establecerá los procesos adecuados para gestionar la recuperación de los paquetes a partir de su PID conocido.

Complementariamente, este módulo proporcionará los medios para realizar consultas sobre el contenido de los paquetes AIP, dando la posibilidad de realizar búsquedas por texto.

Esta característica debe estar complementada con un riguroso control de accesos a los datos almacenados, dado el carácter privado de una parte de la documentación archivada, limitándose su acceso a los funcionarios al cargo y a los propios interesados.

Idealmente, el módulo deberá ser capaz de interactuar con distintos tipos de almacenamiento físico (discos, cintas LTO, SAN), lo que permitiría optimizar el acceso utilizando medios más rápidos para los datos de consulta más frecuente.

En la actualidad, Fedora Commons no implementa una gestión avanzada de medios, por lo que sería recomendable estudiar su integración con un gestor de almacenamiento.

Componentes open-source



6.4.4. Módulo de gestión de datos

Con objeto de optimizar las consultas, el sistema extraerá parte de los metadatos incluidos en los paquetes AIP y los registrará en un Índice de Recursos. Este módulo será responsable de la gestión de esta información, así como de establecer las relaciones entre los campos almacenados y los paquetes AIP a los que se refieren.

La información manejada por el módulo será:

- Metadatos estructurales (definen la relación entre documentos almacenados en el archivo).
- Metadatos descriptivos (principalmente basados en el esquema Dublin Core).
- Referencias a los paquetes AIP almacenados en el repositorio (PID).

Por defecto, Fedora Commons delega estas funciones en una Base de Datos Mulgara ya integrada. Mulgara es una BD RDF y difiere de una BD relacional en que está optimizada para establecer relaciones en formato tripleta.

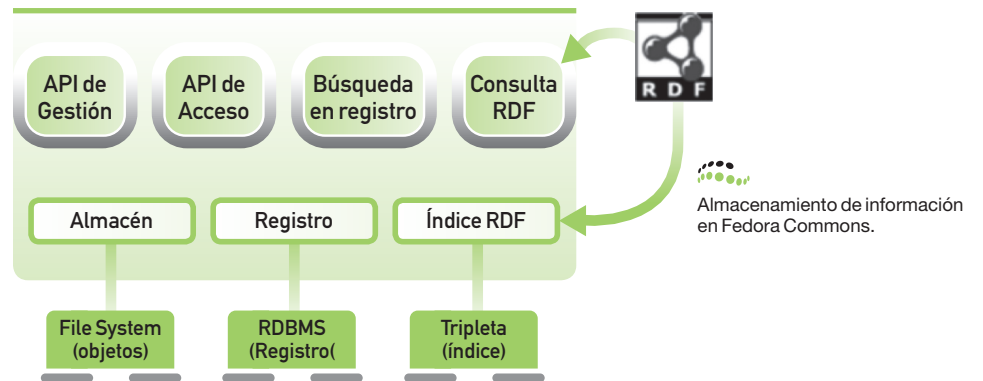
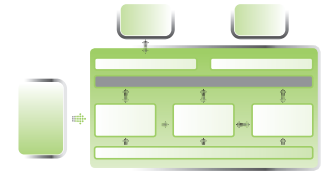
Una tripleta se compone de tres valores:

1. Recurso (identificador del objeto digital archivado)
2. Propiedad (propiedad o tipo de metadato)
3. Valor (característica asociada al metadato)



Representación XML de una tripleta.

```
<rdf:Description rdf:about="paqueteFedora.100">
  <dc:title> "Documento_n123" </dc:title>
</rdf:Description>
```

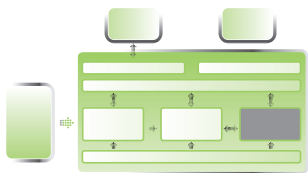


Estándares



Componentes open-source





6.4.5. Módulo de consulta y acceso

Los usuarios finales harán uso de los servicios de este módulo para realizar consultas sobre la información almacenada y acceder a una visualización de los documentos almacenados en el Archivo Electrónico.

Estas consultas podrán realizarse:

- Sobre la BD de metadatos incluida en el módulo de Gestión de Datos.
- Sobre el contenido de los paquetes AIP almacenados en el repositorio.

Como resultado de estas búsquedas se presentará un listado de los paquetes que cumplan con los criterios de búsqueda.

A partir de los resultados de las búsquedas, los usuarios podrán solicitar la consulta de los documentos archivados (siempre en modo lectura). Para ello, el sistema generará una copia del paquete con la información requerida.

A esta copia se le denomina Dissemination Information Package (DIP) y puede presentar una encapsulación distinta al AIP. El hecho de proporcionar una copia del paquete implica que, además de los contenidos, también se distribuyen los metadatos.

Esta aproximación da la posibilidad de tratar los objetos digitales solicitados previamente a su representación, abriendo las posibilidades a la implementación de visores y emuladores.

El modelo de datos de Fedora Commons está diseñado para la generación de distintas representaciones de un mismo objeto digital.

Sin embargo, Fedora no dispone una interfaz de usuario (IU) propia por lo que será necesario la integración de alguna de las IU existentes o el desarrollo de una a medida. Al igual que en el módulo de ingreso, se recomienda la adopción del estándar JSR-170.

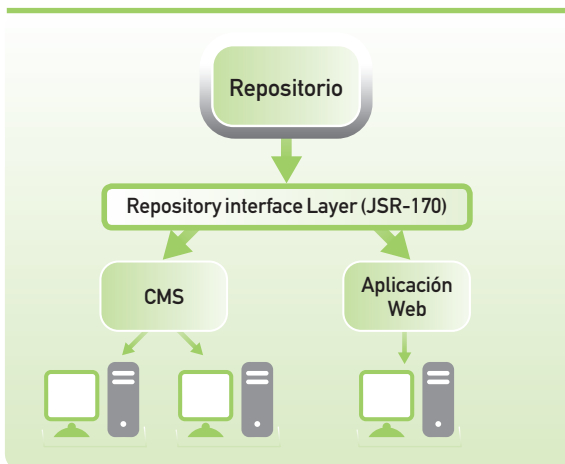
Con objeto de proteger la confidencialidad de la información registrada en el Archivo, este módulo deberá implantar mecanismos eficaces de control de acceso. Esta restricción debe imponerse tanto a la visualización de documentos como a la recuperación de resultados de búsquedas.

En este sentido, se propone adoptar el estándar XACML, que define la sintaxis y los procesos para la definición y aplicación de políticas de autorización complejas y flexibles.

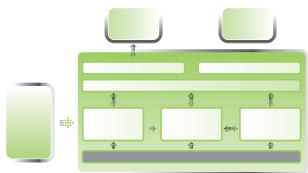
Estándares



Componentes open-source



Interoperabilidad JSR-170.



6.4.6. Módulo de configuración y administración

La finalidad de este módulo será la gestión de la aplicación de Archivo Electrónico, siendo sus usuarios principales los administradores del sistema.

Las principales tareas que comprende:

- Gestión de usuarios.
- Definición de políticas de acceso a los objetos del repositorio.
- Ejecución de procesos (copias de seguridad, migración de ficheros, actualización de firmas de virus).
- Monitorización del sistema y generación de informes.
- Auditorías del sistema.

Como se ha ido apuntando en las secciones anteriores, el control sobre el acceso a la información exclusivamente por los usuarios interesados es un aspecto vital del sistema. Por ello, se deberá hacer un esfuerzo para la definición y mantenimiento de políticas de seguridad apropiadas.

La funcionalidad de Fedora Commons no parece suficiente para proporcionar el nivel de control adecuado sobre el sistema por lo que se considera necesario complementarla con otros componentes, como un gestor de directorio LDAP o un programador de procesos.

Fedora Commons proporciona un interfaz web para la administración del sistema.

Estándares

XACML

LDAP

Componentes open-source



JavaLogin API

6.4.7. Módulo de preservación

Este módulo se orientará a tareas relacionadas con la monitorización de la evolución de formatos de archivo, así como a la planificación y definición de políticas orientadas a la preservación de los contenidos archivados.

Su finalidad es proporcionar mecanismos para asegurar que la información almacenada puede ser representada a los usuarios incluso después de largos periodos de tiempo. Los métodos que actualmente se consideran más viables para ello son la migración de formatos y la emulación.

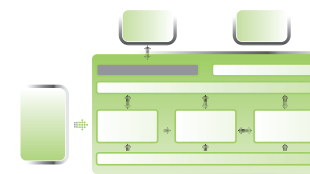
Complementariamente, deberá hacerse un seguimiento de la caducidad de los certificados digitales usados para hacer el resellado temporal de los paquetes que han ingresado en el sistema. En caso de que se acerquen al final de su vida útil, deberá gestionarse su renovación y planificar un nuevo resellado de los paquetes.

Fedora Commons permite el almacenamiento de distintas versiones de un mismo fichero dentro del paquete AIP, lo que da la posibilidad de acometer migraciones a nuevos formatos manteniendo los originales.

Sin embargo, no proporciona ninguna funcionalidad orientada a la evaluación de la obsolescencia de los ficheros empleados.

Durante el estudio, no ha sido posible identificar una aplicación en el mercado que automatice esta tarea.

Sin embargo, sí existen aplicaciones y servicios de información sobre formatos de archivo que pueden servir de referencia a un experto para la definición de políticas de conservación eficaces.



Estándares

P R E M I S

Componentes open-source

JHOVE

DROID
Digital Record Object Identification

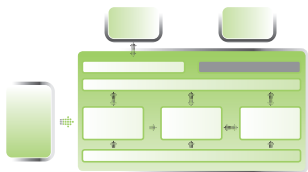
PKP Open Archive Harvester

Quartz
Enterprise Job Scheduler

The technical registry
PRONOM

GDFR⁽¹⁾

⁽¹⁾ GDFR se encuentra en fase experimental, por lo que sus creadores no recomiendan su uso en entornos productivos.



6.4.8. Módulo de gestión de transferencias

Este módulo tiene como propósito facilitar la gestión de transferencias de documentación entre los distintos tipos de Archivo Electrónico de la Administración Pública.

Es un añadido al modelo OAIS, que no lo contempla, y se incorpora al modelo para identificar una serie de funciones específicas proporcionadas por los distintos módulos y que en este caso tienen un peso suficiente como para ser agrupadas en un módulo aparte. Esta funcionalidad comprenderá:

- Generación de listados de paquetes a transferir.
- Identificación de los documentos que han excedido su periodo de vigencia en el archivo.
- Revisión de la documentación previa a su ingreso en el siguiente archivo.
- Edición de metadatos para facilitar una clasificación más adecuada de la documentación.
- Registro de las autorizaciones y rechazos sobre las transferencias realizadas.

Los paquetes identificados para su transferencia serán generados como paquetes DIP en el fichero origen y enviados al fichero destino en forma de paquete SIP.

En principio, se recomienda mantener el mismo formato para paquetes DIP y SIP con objeto de evitar la necesidad de un convertidor intermedio.

El proceso podría automatizarse a través del empleo de una herramienta de workflow, lo que permitiría automatizar las tareas de envío y aceptación de la documentación. Se recomienda que estas herramientas cumplan con la recomendación WfMC.

Estándares



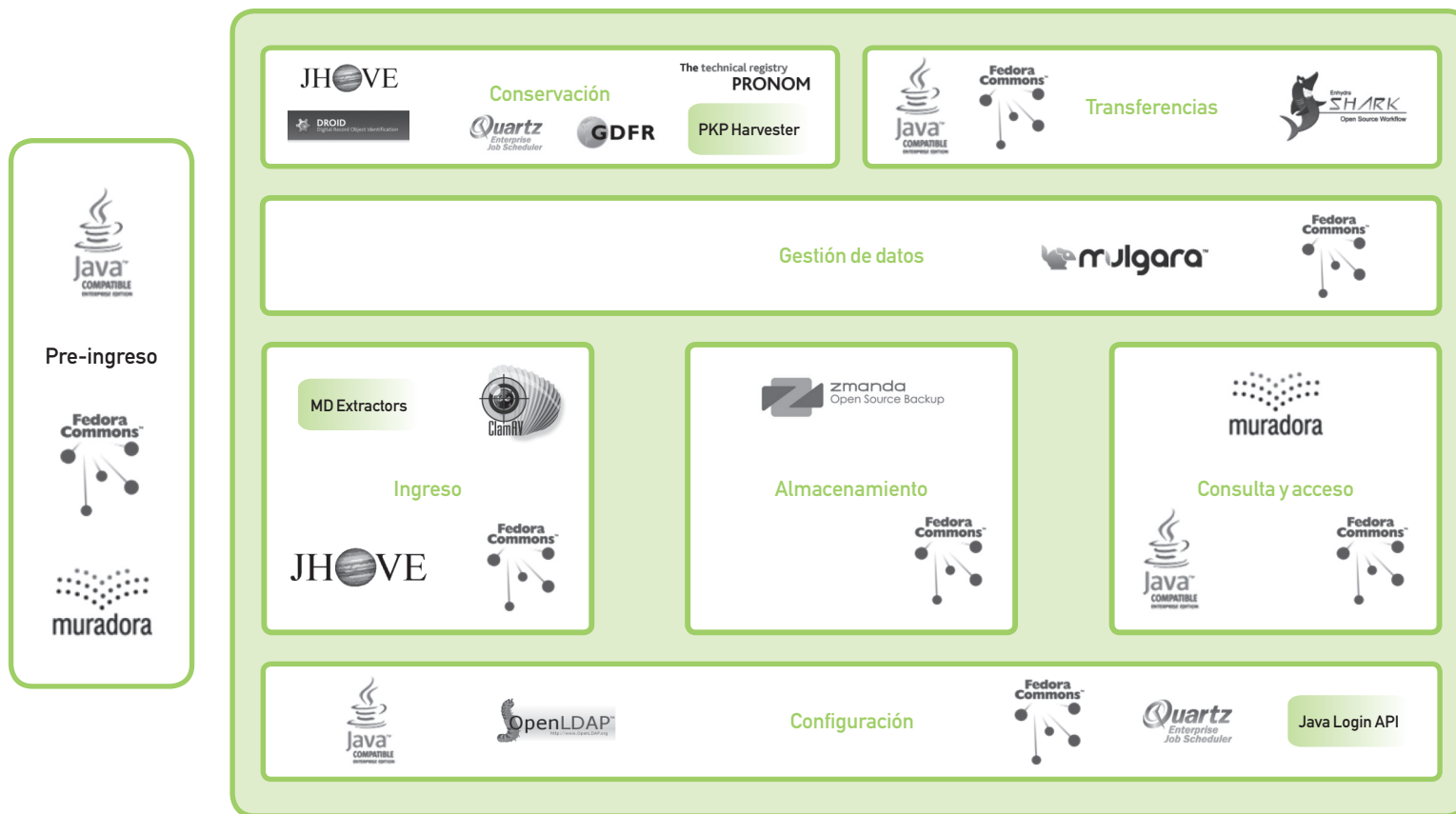
WfMC

Componentes open-source



6.4.9. Esquema de componentes del piloto

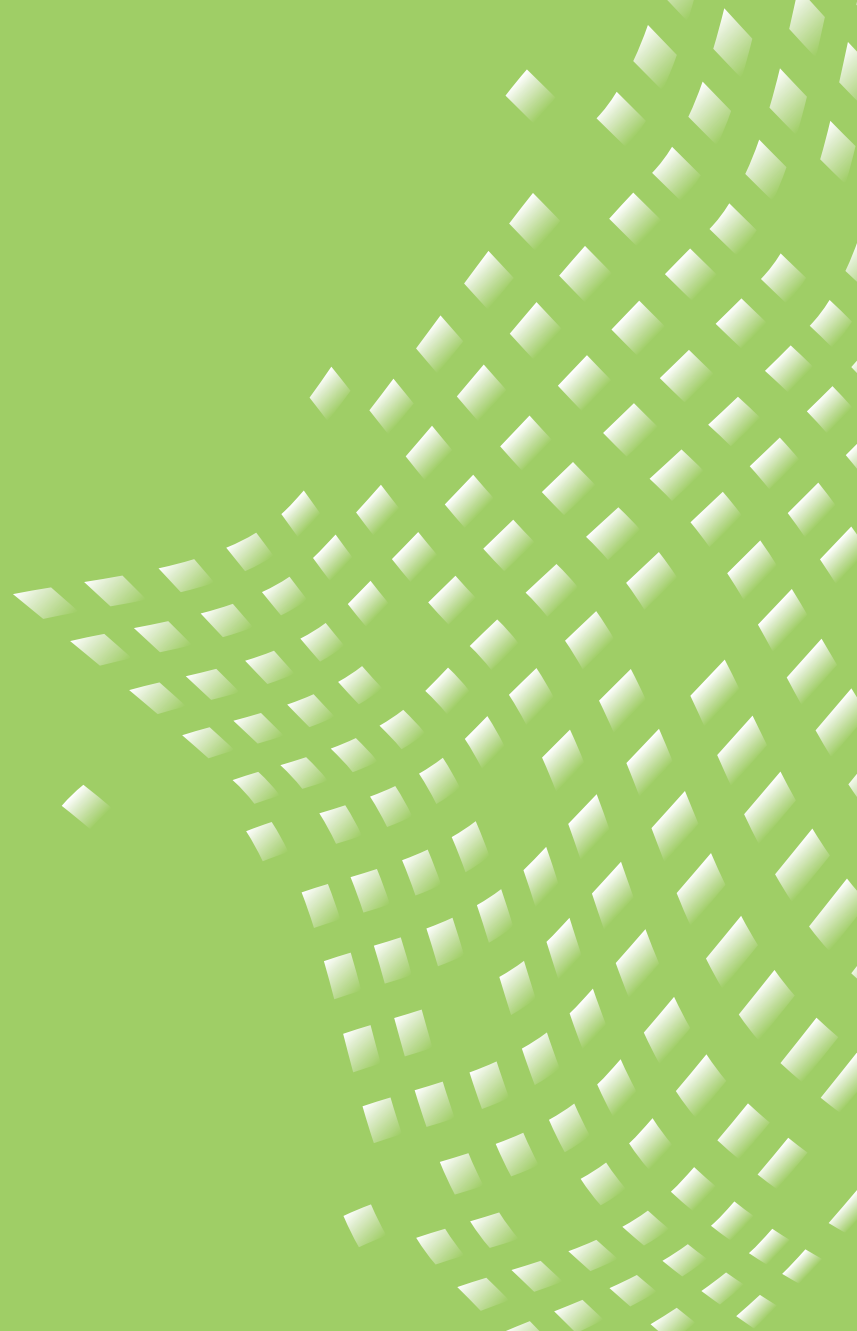
A continuación se presenta como queda el esquemas de componentes del piloto.





07.

Mapa de componentes



7. Mapa de componentes

El presente mapa de componentes es el resultado de un estudio del mercado, cuyo objetivo ha sido identificar aplicaciones y componentes de software de fuentes abiertas susceptibles de completar la funcionalidad de una aplicación de repositorio digital tipo, de modo que se pueda conformar una implementación del Modelo Conceptual propuesto.

Debe tenerse en cuenta que el presente proyecto no persigue generar un diseño cerrado del sistema de Archivo Electrónico, sino más bien comprobar su viabilidad en el ámbito del software de fuentes abiertas. Por ello, las soluciones incluidas en el mapa de componentes deben considerarse sólo como una recomendación, siendo susceptibles de ser sustituidas por otras si se considerase apropiado cuando se aborde el diseño de un prototipo real.

En este sentido, se ha intentado identificar distintos productos para cada uno de los componentes a integrar, otorgando a los desarrolladores la posibilidad de elegir la que crean más adecuada.

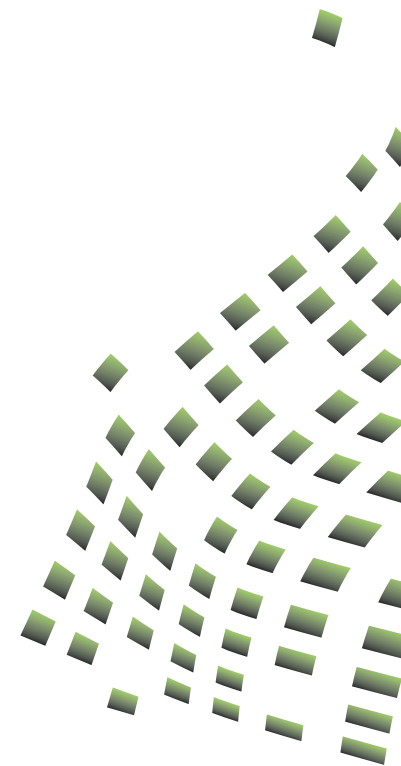
Los criterios utilizados para la selección de los productos incluidos en el mapa de componentes han sido:

- Adecuación funcional a los requerimientos del Archivo Electrónico de las AAPP.
- Facilidad de integración con otros componentes y aplicaciones.
- Grado de madurez tecnológica del producto.

7.1. Identificación de categorías de componentes

Para dar solución a los gaps funcionales, se han identificado las siguientes categorías de componentes.


- Validador de formato de fichero
- Planificador de tareas
- AOI-PMH harvester
- Gestor de Bases de Datos RDF
- Gestor de directorio LDAP
- Registro de logs
- Interfaz gráfica para Fedora Commons
- Framework para workflow
- Framework para el desarrollo de aplicaciones web
- Extractores de metadatos
- Registro de metadatos de formatos de fichero
- Antivirus
- Gestor de backup



7.1.1. Validador de formato de fichero


Este tipo de software permite identificar el formato de fichero y validarlo contra sus especificaciones públicas.

Estos módulos tienen su utilidad como verificadores de los ficheros digitales que ingresan en el sistema, así como comprobadores de la correcta ejecución de las acciones de migración.

JHove	Droid
<ul style="list-style-type: none"> • Es un software patrocinado por la Library of Congress (USA). • Presenta una API que facilita su integración en distintos sistemas. • Permite la validación de múltiples formatos de archivo, entre otros: <ul style="list-style-type: none"> • Imagen fija: GIF, PNG, JPEG, TIFF • Texto / etiquetado: ASCII, HTML, XML • Documentos: ODF, PDF, OfficeXML • Audio: AIFF, WAV • Actualmente se considera la solución de validación de formato más extendida. • Se distribuye bajo licencia GNU Lesser General Public License (LGPL) • La versión 2 se espera para principios de 2010. Este nuevo desarrollo permitirá nueva funcionalidad como la validación de ficheros de acuerdo a políticas locales. 	<ul style="list-style-type: none"> • Desarrollado por The National Archives (UK). • Usa como fuente de formatos el PRONOM Technical Registry. • Implementa un API bien documentado para su integración con otros sistemas. • Permite la interacción a través de: <ul style="list-style-type: none"> • Línea de comandos. • Un interfaz gráfico de usuario. • Se distribuye bajo licencia BSD.
	

7.1.2. Planificador de tareas

Su función es la ejecución automatizada de procesos de acuerdo a una planificación previa.

Quartz Enterprise Job Scheduler	Jcrontab
<ul style="list-style-type: none"> • Software soportado por OpenSymphony. • Se distribuye bajo licencia Apache v.2.0. • Permite dos opciones para el lanzamiento de las tareas: <ul style="list-style-type: none"> • Informando la fecha exacta y/o periodo de ejecución (ejemplo, cada lunes a las 10:30). • Como respuesta a un evento generado por una rutina programada (ejemplo, cada vez que se reciba un número determinado de ficheros). 	<ul style="list-style-type: none"> • Es una implementación en Java de la funcionalidad proporcionada por el comando cron de Unix. • Sólo permite el lanzamiento de tareas en momentos predeterminados (ejemplo, una fecha exacta o un periodo de tiempo). • La definición de eventos se hace a través de un fichero de configuración. • Se distribuye bajo licencia GNU Lesser General Public License (LGPL).

7.1.3. AOI-PMH harvester

La finalidad de este componente es facilitar la publicación de metadatos a otros repositorios, a través del protocolo AOI-PMH, implementando una caché que libera al repositorio origen de la tarea de responder las consultas solicitadas por otros sistemas.

Adicionalmente, los dos componentes considerados permiten también la realización consultas de información sobre otros repositorios.

Metalis	PKP Open Archive Harvester
<ul style="list-style-type: none"> • Distribuido bajo licencia GNU GPL. • Compatible con AOI-PMH v.2. • Tiene soporte para metadatos DC, pero no para MODS. • Permite dos tipos de búsquedas: <ul style="list-style-type: none"> • Simples: keywords en todos los campos. • Complejas: keywords y frases limitados sólo a algunos campos. 	<ul style="list-style-type: none"> • Distribuido bajo licencia GNU GPL. • Compatible con OAI-PMH v.1.1 y v.2 • Tiene soporte para metadatos DC y MODS. • Permite dos tipos de búsquedas: <ul style="list-style-type: none"> • Simples: keywords en todos los campos. • Complejas: keywords y frases limitados sólo a algunos campos.

7.1.4. Gestor de Bases de Datos RDF

Gestiona la Base de Datos RDF donde se almacenan los metadatos descriptivos y estructurales, estableciendo referencias perdurables a los paquetes AIP almacenados en el repositorio.

Esta Base de Datos mantiene índices de acuerdo a los valores de campos de metadatos concretos para acelerar su búsqueda.


Debido a su alto grado de integración con Fedora Commons, no se recomienda ninguna solución para este componente distinta de Mulgara RDF, ya que cualquier otra tecnología precisaría de un complejo trabajo de integración.



7.1.5. Gestor de directorio LDAP

Permite la implantación de un directorio LDAP para la gestión de cuentas de usuario.

Esta información define a los usuarios del sistema y es requerida para la implementación de políticas de acceso y seguridad en el sistema.

OpenLDAP	Apache Directory Server
<ul style="list-style-type: none"> • Distribuido bajo licencia propia (OpenLDAP Public Licence). • Desarrollado por The OpenLDAP Project. • Posiblemente es la implementación LDAP open-source más popular debido a su adopción por multitud de distribuciones Linux. 	<ul style="list-style-type: none"> • Distribuido bajo licencia Apache v.2. y desarrollado por la Apache Software Foundation. • Codificado en Java. • Además de a LDAP, da soporte a otros protocolos como Kerberos.
	

7.1.6. Registro de logs

Componente de librería que proporciona funciones para ampliar las capacidades de trazabilidad y registro de eventos del sistema.

La incorporación de este componente posibilita una mayor flexibilidad a la hora de generar trazas en respuesta a circunstancias específicas.

Log4J	Java.util.logging
<ul style="list-style-type: none"> • Código liberado por la Apache Software Foundation. • Distribuido bajo licencia Apache v.2.0. • Presenta una API que puede ser invocada desde distintas aplicaciones para la generación de logs con distintos niveles de especificación. 	<ul style="list-style-type: none"> • No es una aplicación independiente sino una librería dentro de la plataforma Java para la generación de logs. • Disponible a partir de la versión 1.5. del Java SE Development Kit.



7.1.7. Interfaz gráfica para Fedora Commons

Existen distintas iniciativas de software open-source para solucionar la carencia de una interfaz gráfica de usuario integrada con Fedora Commons.

A continuación se presentan dos de los GUI's más extendidos:

Fez	Muradora
<ul style="list-style-type: none"> • Código liberado por The University of Queensland Library. • Distribuido bajo licencia GNU GPL. • Implementado en PHP. • Mejora el sistema de autorización de Fedora. • Permite la autorización a través de servicios single sign-on como Shibboleth. • Permite la implementación de flujos de revisión y publicación. • Permite administrar vocabularios controlados (definición y mantenimiento de códigos para la clasificación de documentos). • Permite la integración con Jhove para la validación de formatos de fichero. • Implementa su propio índice de búsqueda para ampliar las posibilidades de consulta que proporciona Fedora. 	<ul style="list-style-type: none"> • Código Java liberado por el Macquarie University E-learning Centre of Excellence. • Distribuido bajo GNU Lesser General Public Licence. • Persigue mejorar las funciones de autenticación de Fedora (permite integración con Shibboleth, servicio federado de autenticación). • Posibilita una mayor flexibilidad para la implantación de políticas de autorización. • Permite definir colecciones de objetos digitales de una forma sencilla. • Implementa métodos básicos de gestión y mantenimiento de estos objetos. • Implementa distintos métodos de consulta del repositorio: <ul style="list-style-type: none"> • Simple: por texto. • Complejas: por campos de metadatos. • Por colección.



7.1.8. Framework para workflow

Facilitan mecanismos establecer procesos de trabajo relacionados con el envío y revisión de transferencias de paquetes entre sistemas de Archivo.

Permiten informatizar el registro de los movimientos de información entre los distintos tipos de Archivo.

Enhydra Shark	ObjectWeb Bonita
<ul style="list-style-type: none"> • Código liberado por Enhydra. • Distribuido bajo licencia GNU LGPL. • Es un framework que permite la implantación de servidores de workflow integrados. • Su implementación se basa en las especificaciones WfMC y XPDL. 	<ul style="list-style-type: none"> • Código liberado por ObjectWeb bajo licencia GNU LGPL. • Es una herramienta de workflow flexible basada en WfMC. • Proporciona una interfaz web. • Permite su integración con otras aplicaciones a través de SOAP y XML.
	

7.1.9. Framework para el desarrollo de aplicaciones web

Se ha considerado apropiado la inclusión en el mapa de componentes de un framework para el desarrollo de aplicaciones java, dado que posiblemente sea necesario el desarrollo ad-hoc de cierta funcionalidad (p.e. nuevos interfaces gráficos de usuario).

En la actualidad existen distintos frameworks basados en software de fuentes abiertas y la elección de uno u otro suele estar condicionada por las preferencias y experiencia previa del equipo programador.

Por ello, más que proponer productos concretos, se ha considerado más conveniente recomendar la aplicación de estándares de desarrollo reconocidos que garanticen la calidad de las aplicaciones resultantes.

Esto, además de favorecer la interoperabilidad, dota de una mayor libertad a los desarrolladores que podrán elegir entre aquellas que cumplan con los estándares.

El diseño deberá seguir una arquitectura en capas, de modo que se establezca una separación de los componentes de la aplicación y se favorezca la reutilización del código.

La división típica es: capa de presentación –capa de lógica de negocio– capa de integración de sistemas corporativos (EIS).

Para la capa de presentación se recomienda la adopción de Java Server Faces (JSF), cuya primera versión está especificada por el estándar JSR-127.

Esta especificación está implementada en la plataforma Java Enterprise Edition a partir de la versión 5.

Respecto a la capa de integración de sistemas, donde se implementa el mapeo entre objetos lógicos y bases de datos relacionales, la sugerencia es programar en base a la Java Persistence API (JPA), estandarizada en la norma JSR-317.

Esta especificación también está implementada en la plataforma Java EE 5, así como en otros productos de amplia difusión como Hibernate.

A partir de estas consideraciones, la elección del IDE a utilizar quedará a la elección del equipo desarrollador.



La aplicación de esta tecnología se identificará en el mapa de componentes como:



7.1.10. Extractores de metadatos

Componentes que permiten la generación automatizada de metadatos de tipo técnico a partir de la inspección del objeto digital.

Dado que no existe un único sistema que extraiga todos los metadatos que pueden ser necesarios, se propone la implementación de las soluciones en paralelo.

VTLS Metadata Extraction Service	NLNZ Metadata Extraction
<ul style="list-style-type: none"> • Servicio Web proporcionado por VTLS • Permite la captura de metadatos para imágenes fijas según el esquema de metadatos MIX. • Facilita su integración con Fedora ya que empaqueta los metadatos recopilados con un formato adecuado para su almacenamiento como datastream. • Permite lanzar validaciones sobre Jhove. 	<ul style="list-style-type: none"> • Aplicación desarrollada por la National Library of New Zealand (NLNZ) • Distribuido bajo licencia Apache v.2.0. • Permite el análisis de distintos formatos de objetos digitales (documentos Word y OpenOffice, PDF, MP3, WAV y otros) • La información de salida se entrega en un formato XML estándar. 

7.1.11. Registro de metadatos de formatos de fichero

Se trata de un servicio remoto que mantiene información actualizada sobre las características técnicas y la vigencia de distintos formatos de fichero electrónico.

Su consulta permite la identificación de formatos que han quedado obsoletos, así como la recomendación sobre políticas de migración de estos formatos a otros formatos con mayor vigencia.

No implica la instalación de nuevos paquetes de software, pudiéndose acceder a varios de ellos en paralelo con objeto de recopilar la información más completa posible.


Hay que tener en cuenta que el GDFR aún se encuentra en fase de desarrollo y sus desarrolladores no recomiendan su uso en sistemas productivos.


7.1.12. Antivirus

Como medida preventiva para garantizar la seguridad del sistema, se deberá asegurar que los objetos digitales que ingresen en el archivo estarán libres de virus.

Para ello, se requerirá un escaneo de los ficheros incluidos en el paquete SIP durante las validaciones efectuadas por el módulo de ingreso del sistema.

En el mercado existen actualmente varios antivirus open-source que pueden utilizarse para este cometido. La elección de uno u otro estará determinada, además de por la propia funcionalidad del producto, por la existencia de un soporte post-implantación adecuado (principalmente relacionado con la actualización de firmas de virus).



Clam Antivirus	UNA Antivirus Toolkit
<ul style="list-style-type: none"> • Toolkit de antivirus desarrollado para sistemas Unix. • Liberado bajo licencia GPL. • Es con mucho la opción más extendida. 	<ul style="list-style-type: none"> • Soportado sólo por sistemas Unix. • Licenciado bajo licencia GPL. • Soporte limitado ya que cuenta con una comunidad reducida.

Pronom	Global Digital Format Registry
<p>The technical registry</p> <h1>PRONOM</h1>	

7.1.13. Gestor de backup

Software que proporciona mecanismos para generar copias de seguridad del repositorio de paquetes almacenados por el Archivo.

Su implantación es vital para la restauración de la información en caso de que se corrompan los datos o el software que les da soporte.

Amanda Network Backup	Bacula
<ul style="list-style-type: none"> • Desarrollado por Zmanda. • Código Perl distribuido bajo licencia BSD. • Permite, desde un servidor, la realización de copias de backup y restauraciones en múltiples hosts. • Permite la realización de copias de backup en múltiples soportes (discos, cintas, SAN). 	<ul style="list-style-type: none"> • Software distribuido bajo licencia GNU GPL. • Conjunto de aplicaciones para realizar copias de seguridad y restaurarlas en red. • Da la posibilidad de volcar las copias de seguridad en distintos formatos (discos, cintas, SAN) • Responde a una arquitectura cliente / servidor. 

7.2. Mapeo funcionalidad – componente









En esta sección, se presenta un listado de la funcionalidad requerida al sistema de Archivo Electrónico de las AAPP agrupada por módulos.

Sobre este listado, se establece la correspondencia con los componentes open-sources identificados para su implantación. La columna marcada como core representa la aplicación de repositorio digital que servirá de núcleo al sistema (en el diseño se ha optado por Fedora Commons).

Esta correspondencia se marca a través de los siguientes iconos:

- ✔ El componente cubre la totalidad de la funcionalidad.
- ⊕ El componente cubre la funcionalidad parcialmente.
- ✕ El componente no proporciona la funcionalidad señalada y se requerirá un desarrollo ad-hoc, bien para implantar dicha funcionalidad, bien para propiciar la integración con otros componentes.

7.2.1. Módulo de pre-ingreso

Funcionalidad	Componentes	
	Interfaz de usuario	
	Muradora	Fez
Ingreso de objeto digital		
Ingreso de mandatos		
Relación entre paquetes		
Categorización por directorios		

7.2.2. Módulo de ingreso

Funcionalidad	Componentes						
	Core	Validador de formato		Extractor de MD		Antivirus	e-firma ^(*)
		Jhove	Droid	VTLS	NLNZ	ClamAV	
Validación formato		✓	✓				
Validación virus	✗					✓	
Validación firma de negocio	✗						✓
Validación catalogación							✓
Extracción automática de metadatos				✗	✗		
Sellado de tiempo	✗						✓
Creación paquete AIP	✓						
Generación PID	✓						

* Componente de firma electrónica incluido en la plataforma de Administración Electrónica del organismo.

7.2.3. Módulo de almacenamiento

Funcionalidad	Componentes		
	Core	Gestor de backup	
		Amanda	Bacula
Ingreso paquete AIP	<input checked="" type="checkbox"/>		
Indexación paquete AIP	<input checked="" type="checkbox"/>		
Conservación paquete AIP	<input checked="" type="checkbox"/>		
Recuperación de AIP a partir de PID	<input checked="" type="checkbox"/>		
Backup de paquete AIP		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

7.2.4. Módulo de gestión de datos

Funcionalidad	Componentes	
	Core	Base de datos RDF
		Mulgara
Registro de metadatos		<input checked="" type="checkbox"/>
Relación de Metadatos		<input checked="" type="checkbox"/>
Indexación de metadatos		<input checked="" type="checkbox"/>
Lanzamiento de consultas	<input checked="" type="checkbox"/>	

7.2.5. Módulo de consulta y recuperación

Funcionalidad	Componentes		
	Core	Interfaz de usuario	
		Muradora	Fez
Consulta por MD		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Consulta por texto completo		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Consulta por PID		<input type="checkbox"/>	<input type="checkbox"/>
Vista de resultados de búsqueda		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Generación de paquetes DIP	<input checked="" type="checkbox"/>		
Visualización de los contenidos		<input type="checkbox"/>	<input type="checkbox"/>
















7.2.6. Módulo de administración y configuración

Funcionalidad	Componentes						
	Core	Planificador de tareas		Gestor LDAP		Registro de Logs	
		Quartz	Jcrontab	OpenLDAP	Apache Directory	Log4J	Java API
Gestión de usuarios							
Control de accesos							
Planificación de procesos							
Transformaciones							
Auditorías y monitorización							

7.2.7. Módulo de preservación

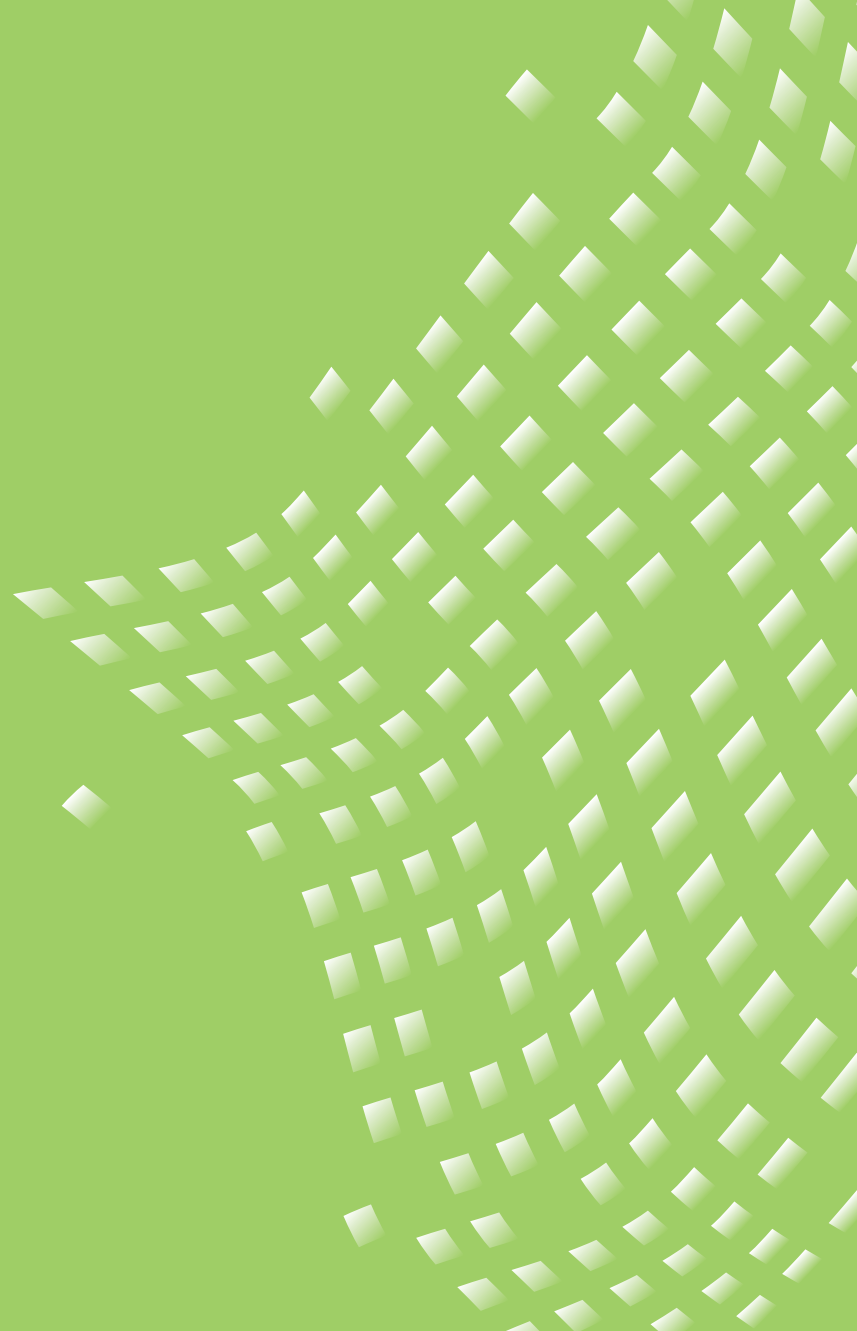
Funcionalidad	Componentes			
	Core	Registros de formatos digitales		
			Pronom	GDFR
Monitorización de la tecnología				
Políticas de transformación				
Políticas de emulación				
Identificación de paquetes objetivo				
Monitorización de certificados digitales				
Políticas de resellado				

7.2.8. Módulo de gestión de transferencias

Funcionalidad	Componentes			
	Core	Gestores de workflow		
		Enhyndra Shark	ObjectWeb Bonita	
Identificación de paquetes a enviar				
Relación de transferencia				
Revisión en archivo destino				
Aceptación de la transferencia				
Transferencias de paquetes				
Rechazo de la transferencia				
Gestión de rechazos en origen				

08.

Glosario



8. Glosario

AIP

(Archival Information Package)

- Unidad de almacenamiento del repositorio digital definida por la norma OAIS.

AudioMD

(Audio Technical Metadata Extension Schema)

- Esquema de metadatos utilizado para la descripción de ficheros digitales de audio.

DIP

(Dissemination Information Package)

- La norma OAIS lo define como un empaquetamiento de contenidos y metadatos empleado para su distribución a los usuarios.

Dublin Core

- Esquema de metadatos elaborado por la DCMI para la descripción genérica de objetos digitales.
- Define campos como título, autor, organización, etc.
- Considerado como el estándar de facto para metadatos descriptivos.

Exif (Exchangeable Image File Format)

- Especificación de formatos digitales usada principalmente en cámaras fotográficas digitales.
- Se basa en la modificación de formatos de archivo existentes como JPEG o TIFF, a los que se agregan etiquetas de metadatos.

GDFR (Global Digital Format Registry)

- Servicio soportado desde la Harvard University Library para el almacenamiento, monitorización y evaluación de información de representación sobre archivos digitales.

IPTC

(Internacional Press Telecommunications Council)

- Hace referencia al estándar técnico de metadatos promovido por esta organización para favorecer el intercambio de noticias.

JSR-170

(Content Repository for Java Technology API)

- Especificación pública que define una API Java estandarizada para el acceso a distintos repositorios de contenidos independientemente de su tecnología.
- Establece dos niveles de compatibilidad y algunas funciones complementarias aparte de ellos:
- Nivel 1: Navegación por el repositorio y lectura
- Nivel 2: Agregación de datos al repositorio
- Complementos: entre otros versionado, bloqueo de ficheros, consultas SQL.

LDAP

(Lightweight Directory Access Protocol)

- Protocolo estandarizado para el acceso a un servicio de directorio distribuido.
- Permite la consulta de propiedades de los usuarios tales como identificador, contraseña, grupos a los que pertenece, etc.

METS

(Metadata Encoding & Transmission Standard)

- Estándar para la codificación de metadatos descriptivos, administrativos y estructurales referentes a objetos almacenados en repositorios digitales.
- Se basa en la definición de una estructura XML.

MIX

(Metadata for Images in XML)

- Esquema de metadatos para la definición de imágenes fijas promovido por la Librería del Congreso americana.

MODS

(Metadata Object Description Schema)

- Esquema de metadatos para la descripción genérica de archivos digitales.
- Actualmente es la alternativa más clara a Dublin Core.

NLNZ

(Nation Library of New Zealand)

- Organismo público impulsor de iniciativas para la preservación de información electrónica.

OAI-PMH

(Open Archives Initiative - Protocol for Metadata Harvesting)

- Framework de interoperabilidad, basado en XML, para la compatición de información entre repositorios digitales.

OAIS

(Open Archival Information System)

- Modelo de referencia que define un sistema de archivo electrónico orientado a la preservación de la información digital.
- Está normalizado según la norma ISO 14721:2003.

PREMIS

(PREservation Metadata Implementation Strategies)

- Esquema de definición de metadatos orientados a la preservación de ficheros digitales durante largos periodos de tiempo.

RDF

(Resource Description Framework)

- Framework para el intercambio de metadatos.
- Basado en el concepto de tripleta, cuyos componentes son:
- Recurso: cualquier entidad que pueda tener una URI
- Propiedad: característica normada del recurso (metadato)
- Valor: valor asociado al metadato.

SIGEM

- Plataforma de Administración Electrónica impulsada por el Ministerio de Industria, Turismo y Consumo.
- Esta solución está a disposición de todas aquellas Administraciones Públicas que quieran implantarla.

SIP

(Submission Information Package)

- Paquete de ingreso en el repositorio digital de acuerdo a la especificación OAIS

SOA

(Service Oriented Architecture)

- Concepto de arquitectura software basada en el uso de servicios para dar soporte a los requisitos de negocio.
- La publicación e invocación de estos servicios se realiza a través de procedimientos estándar, lo que facilita la integración entre distintos tipos de sistemas, tanto propios como ajenos.

TEI

(Text Encoding Initiative)

- Esquema de metadatos para la representación de textos en formato digital.

WfMC

(Workflow Management Coalition)

- Modelo de referencia para favorecer la interoperabilidad entre sistemas de workflow

XACML

(eXtensible Access Control Markup Language)

- Lenguaje para la declaración de políticas de control de accesos implementado sobre XML y modelo de procesado para la interpretación de estas políticas
- Se basa en dos componentes o entidades:
- PEP (Policy Enforcement Point) – recibe las peticiones sobre los recursos y da acceso en función de la respuesta dada por el PDP.
- PDP: Policy Decision Point – evalúa las políticas de acceso definidas en el sistema y decide si el PEP debe permitir que el solicitante acceda a la información

XMP

(eXtensible Metadata Platform)

- Contenedor XML de metadatos que puede incorporarse a un archivo digital.
- Desarrollado por Adobe