

## La recuperación de la información en los archivos en línea<sup>1</sup>

[\[Versió catalana\]](#)

VICENT GIMÉNEZ-CHORNET   
Profesor Titular de Biblioteconomía y Documentación  
Universitat Politècnica de València  
[vigicho@har.upv.es](mailto:vigicho@har.upv.es)

### Opciones

 Imprimir  Recomandar  Citació  Estadístiques  <meta /> Metadades  Similars

### Resumen [\[Abstract\]](#) [\[Resum\]](#)

**Objetivo.** Analizar las propuestas de recuperación de la información de los archivos que han puesto a disposición de los usuarios en Internet sus fondos documentales, para reflexionar sobre los cambios de paradigma de los archivos físicos a los archivos digitales.

**Metodología.** Identificación y análisis de los sitios web de archivos que permiten a los usuarios diversos planteamientos para la recuperación de la información. Valoración de la implementación de las buenas prácticas en la recuperación de la información en la arquitectura del sitio web.

**Resultados.** La evolución de las Tecnologías de la Información y la Comunicación permiten implementar eficaces instrumentos de recuperación de la información en los archivos que ponen a disposición en línea sus catálogos. El análisis permite reflexionar sobre las prácticas de la recuperación de la información en los archivos físicos, y de su evolución, y de las posibilidades de aplicarlo en la Web y en la Web 3.0.

## 1 Introducción

Ha sido bastante habitual, cuando se mencionan las funciones que debe realizar un archivero, incluso en el entorno de la administración electrónica, no destacar entre sus tareas los aspectos relacionados con la recuperación de la información. Karen Dawley Paul, en los años 80, incidía en que los gestores documentales tenían las funciones de planificación, incluyendo los aspectos de creación, utilización, protección y eliminación final, tras la evaluación, de los documentos electrónicos en un sistema de información, abarcando incluso aspectos como los servicios de reprografía, los procesadores de textos, el procesamiento de datos, el correo electrónico, el reconocimiento óptico de caracteres, las telecomunicaciones, la micrografía, la gestión de documentos y archivos, y también servicios bibliotecarios, mencionando que la gestión eficaz de todo ello jugaba un papel importante en el establecimiento de la estructura de información de las organizaciones y que ello afectaría a la información que se conservaba y que sería accesible para futuras investigaciones (Paul, 1988, 37).

La teoría archivística española centró sus investigaciones, en los años 80, en aspectos relacionados con la clasificación, la descripción, la valoración, las tipologías documentales, la autenticidad, o el ciclo vital de los documentos generados por las organizaciones, pero poco sobre la "búsqueda" o recuperación de la información de ellos en los archivos. En parte esto tiene cierta explicación. Los profesionales, como por ejemplo Antonia Heredia (1988), en el libro *Archivística general, teoría y práctica*, se centraban en los aspectos que más preocupaban a los archiveros en un entorno de carencias tecnológicas con el cual trabajan la mayoría. Los estudios realizaban una visión de conjunto y sistematizaban los conocimientos archivísticos, dando a conocer, incluso, las novedades en la automatización de los archivos. Los temas que se trataban hacían referencia a las ciencias auxiliares de la archivística, la historia de los archivos, las características de los documentos de archivo, las transferencias, la valoración documental, la administración de archivos,

la clasificación y la ordenación, la descripción y los instrumentos de descripción (guías, inventarios y catálogos), considerando a las fichas índice como los instrumentos auxiliares de descripción, los tesauros, las listas alfabéticas o los libros registro, la accesibilidad documental y el servicio de documentos. Si en los manuales de archivística de los años 80 y 90 no hay un capítulo destinado a la recuperación de la información, no es porque no interese el tema, sino porque el entorno físico y la ausencia de implementación tecnológica es bastante generalizada (excepto excepciones, como el caso de la automatización del Archivo General de Indias), y la recuperación de la información se contempla desde otro enfoque.

Antes del entorno tecnológico (fundamentalmente bases de datos, Internet e intranets) la recuperación de la información, es decir, encontrar los documentos tras una petición, se basaba en una buena organización del archivo y en unos buenos instrumentos de descripción, de ahí que no existiesen capítulos específicos dedicados a esta faceta en nuestra literatura archivística, ya que la respuesta estaba implícita en la buena gestión de los documentos archivísticos. Otra cuestión es cuando aparece Internet, y trasladamos este mismo asunto a la arquitectura de la información del sitio web, y a sus funcionalidades (los OPACs).

El interés por el proceso de "búsqueda de documentos" no es nuevo. Ya en 1835, en una memoria realizada por Jorge García sobre el Archivo del Reino de Valencia, dedicó un capítulo titulado "De la busca de documentos". Según dicha memoria: *la busca pues, de los documentos depende en la actualidad o de las noticias ciertas y circunstanciadas que suministren los interesados, o de las que arrojen los índices que se mencionarán, o de la luz que el archivero y oficiales puedan dar a beneficio de sus observaciones y experiencias*. A pesar de la antigüedad de estas apreciaciones, recalamos los aspectos que nos parecen trascienden a la actualidad. Para la recuperación de la información en los archivos habría, pues, que tener en cuenta, según aquellos parámetros del siglo XIX:

1. El vocabulario y los términos de la consulta.
2. El análisis documental y la indización.
3. La tecnología en la búsqueda.

Literatura archivística más reciente no ha contemplado dichos aspectos. Es más, algunas obras que han sido muy positivas en algunos aspectos, en otros han impregnado una opinión, desde nuestro punto de vista, errónea. Es el caso de las afirmaciones de Elio Lodolini (1993, p. 207-208, 213-214.), que tanta influencia ha teniendo en algunos archiveros españoles. Según este autor la descripción individualizada de los documentos de archivo y su análisis documental, lo que se entiende por *catálogo*, no se debe realizar en los archivos:

*En efecto, no tendría sentido preguntar, en un archivo, ¿qué hay sobre tal tema? y ¿qué hay sobre tal personaje? (como, por desgracia, bastante a menudo hacen los que se dirigen a un archivo sin una suficiente preparación específica). Es necesario, por el contrario, preguntar cuál era, en los diversos momentos, la oficina competente para tratar el tema que interesa y qué procedimientos usaba, es decir, cómo producía y organizaba sus propios documentos, o bien, cómo un personaje ha tenido contactos con la autoridad pública (si la investigación se desarrolla en un archivo público), en calidad de juez, de acusado [.]*

*El inventario es, pues, un instrumento compuesto; el único medio que permite realizar la búsqueda en un archivo. Del todo inútil sería, por el contrario, un catálogo de documentos, entre otras cosas carente de sentido... sin poner de relieve el aspecto fundamental, archivístico, de los documentos mismos, es decir, su interdependencia [.]*

En este contexto algunos autores españoles han considerado que el catálogo era el menos recomendable de todos los instrumentos de descripción, desde un punto de vista práctico y de servicio, debido a las carencias de medios materiales y humanos existentes en la mayoría de los archivos. En consecuencia, siguiendo el planteamiento de Lodolini, la ausencia de descripción individualizada y análisis documental de los documentos imposibilita su búsqueda específica dado que no hay indización ni descripción individual del documento. Estas opiniones del siglo XX, curiosamente, contrastan con la consideración del archivero de 1835, que tenía muy claro que las fichas índices (donde estaba también la descripción individualizada de los documentos) era uno de los pilares básicos para la "búsqueda" de los documentos. De hecho, en muchos archivos españoles se realizaron en el siglo XIX y principios del XX muchas fichas índice, que han servido reiteradamente a los historiadores.

En este contexto, ¿la producción teórica en archivística en otros países se ha preocupado por la indización y la recuperación de la información en los archivos? Richard J. Cox (1992) ya señaló que frente a la definición tradicional de *gestión documental* —como el uso del control sistemático y

científico de toda la información registrada que una organización necesita en su actividad empresarial— la aparición en los últimos años —años 80— del concepto de gestión de recursos de información, *information resources management* (IRM), era algo más: la información es un recurso institucional, que necesita de la tecnología para ser gestionada, y necesita de personas para tratar eficazmente ambas cosas. La Gestión de Recursos de Información (IRM) incluye el esfuerzo de tener el control de la total diversidad de recursos de información, hardware y software, equipamiento de telecomunicaciones, equipamiento de automatización de la oficina y de los aparatos de reprografía, así como de la información transmitida a través de esta infraestructura y las personas. La IRM incluye, también, una variedad de técnicas que son muy similares a aquellas empleadas por los gestores documentales y creadas en torno al concepto de *ciclo de vida*. Estas afirmaciones las señaló Richard Cox cuando aún no había nacido Internet, tal y como lo conocemos ahora. Posteriormente, Michael Cook y Margaret Procter (2000) señalaron la existencia de diferentes estrategias de recuperación de la información por parte de los usuarios. Las tres principales estrategias empleadas por éstos (incluyendo al personal propio de la entidad como usuarios) para localizar los documentos archivísticos más relevantes para su investigación son: la identificación directa, el hojearo y la búsqueda o exploración. La identificación directa ocurre cuando el usuario sabe una o más de las características de identificación de los documentos (por ejemplo un nombre específico, una fecha o un código de referencia). Para ello el requisito principal es que los datos que pertenecen a cada entidad deben estar claramente identificables. En la estrategia de hojear, los usuarios leen las páginas del asistente de búsqueda para seleccionar cualquier información o tema que consideren como útiles. Esta estrategia se emplea con descripciones archivísticas de texto libre, tales como la historia administrativa, la custodia o las áreas de alcance y contenido. En la estrategia de búsqueda, el objetivo de la recuperación debe estar más o menos bien definido. La acción de la exploración o búsqueda se basa en el uso de palabras claves. Hay que tener presente que los usuarios persiguen rapidez en la presentación de los resultados —están habituados a los resultados de Google—. En esta estrategia se requiere una OPAC correctamente construida. Suele haber un cierto grado de conflicto entre la configuración del sitio web y la estructura de la bases de datos o OPAC más conveniente para cada modalidad de recuperación de la información, por ello es importante que se establezcan formas de ayudas apropiadas para cada estrategia, separando las referidas al hojearo de la búsqueda o exploración.

También Cook y Procter (2000) han señalado la importancia de la indización en los archivos como parte de la recuperación de la información. Opinan que la creación de los índices son partes esenciales de una descripción archivística y que son puntos de acceso necesarios en un sistema de recuperación de la información. La importancia de los índices en la descripción archivística ha aumentado en estos últimos años. El vocabulario de los índices debe estar hecho con antelación, y deben ser una parte de un sistema integrado de apoyo en la recuperación de la información. Los usuarios tienen que disponer de acceso directo a los índices. Los índices son ayudas secundarias de la recuperación. Ello permite llevar a los usuarios de forma certera a las descripciones que en el sistema están indizadas, permitiendo recuperar directamente los documentos que se desean. Se recomienda que los índices sean palabras claves dentro de los asistentes de búsqueda.

## **2 Instrumentos a nuestro alcance para la recuperación de la información**

En la actualidad disponemos de diversos instrumentos a nuestro alcance para resolver de la forma más eficiente la recuperación de la información en los archivos que han decidido poner sus fondos a disposición de cualquier usuario, o a usuarios de una comunidad específica. Estos instrumentos se basan en software y en hardware, así como en aplicación de estándares. No vamos a tratar del hardware ni del software específico, sino de forma genérica que nos ofrecen las TICs para optimizar la gestión, y por tanto, la recuperación de la información.

Hemos de dejar claro que todo el trabajo que se realiza en la administración de los archivos y en la gestión de los documentos converge en una sola finalidad: poder consultar los documentos tras una petición. Los documentos se clasifican, se ordenan, se describen, se garantiza su óptima conservación, para poder resolver la consulta en un momento dado.

Entre los instrumentos a nuestro alcance destacamos:

### **2.1 Base de datos**

La base de datos y el sistema de gestión de bases de datos han revolucionado la gestión de los

documentos. Es, en definitiva, la convergencia de un sistema de gestión basado en muebles con cajones que contienen fichas descriptivas (ordenadas tradicionalmente por lugares, fechas, o asuntos) a un sistema de gestión electrónica donde identificamos campos en los que se almacenan datos o información.

Mientras que las bases de datos son un conjunto o colección de datos, un sistema de gestión de bases de datos —SGBD— es un software que permite la creación, mantenimiento y explotación de la base de datos (Abadal; Codina, 2005, 18-19). Una base de datos es una colección organizada de datos para uno o más propósitos, y que por lo general, en las últimas décadas, la encontramos en forma digital. El término base de datos se aplica a los datos y estructuras de datos, y no al SGBD que requiere de un software para gestionarlos, dado que la estructura de una base de datos suele ser demasiado compleja para ser manejada sin su SGBD. La base de datos SQL dispone del estándar ISO/IEC 9075 —*Information technology, Database languages, SQL*— actualizado con frecuencia.

## Descripción archivística

La normalización en la descripción archivística —estándares como ISAD (G), ISAAR-CPM, EAD o EAC-CPF— permite estructurar la descripción de la información y la descripción de sus productores. Ello supone otro gran avance porque ha abierto la posibilidad de hacer interoperables distintas bases de datos. La aplicación de los estándares va a permitir —al igual que en las bibliotecas lo permitió la ISBD o los distintos formatos MARC— interrelacionar las bases de datos con campos definidos para la descripción y el análisis documental. Queremos hacer hincapié en tres partes que caracterizan a la descripción archivística, por las cuales se debe poder realizar la recuperación de la información.

1. *Descripción del contexto.* Son campos destinados a describir al productor de la documentación archivística, especialmente mediante las normas ISAAR-CPF o la EAC-CPF. En opinión de la comisión que elaboró la ISAAR-CPF, con ello se consigue al mismo tiempo facilitar la recuperación de información de las descripciones archivísticas, así como que la recuperación de la información se vea mejorada por el uso de puntos de acceso (o términos índice), y que los puntos de acceso funcionen mejor cuando están estandarizados por medio de un registro de autoridad (Thibodeau, 1995). El interés por describir el contexto en la documentación archivística es algo más: para saber que el expediente es auténtico debe estar identificado su productor en el mundo, se ha de especificar el sistema que lo gestiona, se tienen que nombrar sus creadores y los procesos por los que se identifica. La descripción del contexto, a partir de los metadatos generados, contribuye a garantizar su autenticidad (Cumming, K., 2007). Desde el punto de vista de la recuperación de la información es mucho más relevante describir al productor en un archivo que recoge documentación de muchos productores, que en un archivo de un único productor, puesto que una petición de búsqueda puede ser la recuperación de los documentos producidos por un productor dado. Desde el punto de vista de la autenticidad y las medidas de conservación se debe describir el productor.
2. *Descripción de la unidad archivística.* Bien hablemos de Unidad de Información, bien de Unidad Archivística, en este apartado lo entendemos como *unidad de descripción*, tal y como lo define la norma ISAD (G), es decir, como un documento o conjunto de documentos, cualquiera que sea su forma física, tratado como un todo y que como tal constituye la base de una única descripción. La norma identifica 26 elementos para describir cualquier unidad (fondo, serie, unidad documental compuesta o simple, etc.). La traslación de los elementos a campos de una base de datos hace posible su recuperación automatizada, con todas las características que puedan ofrecer estos campos: numéricos (para procesar números, máscaras de bits, fechas u horas) o alfanuméricos. Igualmente, los etiquetados de la EAD sirven para ser trasladados a bases de datos.
3. *Indización.* Utilizada en los archivos, al menos desde época medieval, de forma progresiva se va incorporando a las bases de datos archivísticas. A principios del siglo <sup>xxi</sup> eran escasas, en el entorno español, las bases de datos que habían incorporado la indización (bien sea con lenguaje documental o con lenguaje libre) como campos para la recuperación de la información. Hemos de resaltar la diferencia que existe entre índice y productor. El primero como uno de tantos puntos de acceso a la información registrada en ciertos campos de la base de datos, donde se aconseja la utilización de un lenguaje documental, que lo utilizan tanto las bibliotecas como los archivos para indicar lugares, materias, organismos o personas. Mientras que el del productor en los archivos es un concepto más amplio —como

hemos indicado anteriormente— y dispone de su norma específica para describir este contexto.

## Interoperabilidad

La *interoperabilidad* la entendemos como la propiedad de un producto o de un sistema que es capaz de conseguir la utilización de software por distintos sistemas informáticos (sistemas operativos y aplicaciones de software), interconectados por diferentes tipos de redes, para el intercambio de información o de datos. En los archivos, inicialmente —años 70 y 80—, la introducción de software y hardware fue para la automatización de distintas unidades de descripción en bases de datos, muchas veces con la realización de índices, para favorecer la recuperación de la información. Primero se introdujo en local (Bell, 1975), posteriormente apareció el interés de compartir esta información en línea (Arad; Bell, 1977-1978), como ocurría en las bibliotecas y, finalmente, la automatización iba a contemplar la totalidad de la gestión en una organización (Vázquez de Parga, 1986). Desde la perspectiva de la recuperación de la información nos interesa destacar la interoperabilidad desde dos aspectos:

1. *Intercambio de datos*. El intercambio de datos e información se efectúa principalmente mediante una estructura de datos (los campos de las bases de datos, o mediante *scripts* creados para la transferencia de los datos). A nivel europeo, el programa IDA (Intercambio de Datos entre Administraciones) ha sido una gran iniciativa que, desde 1998, ha generado experiencias, estándares y aplicaciones para hacer posible la interoperabilidad de las redes telemáticas transeuropeas destinadas al intercambio de datos entre administraciones (Unión Europea, 1998), con ello hemos alcanzado que los archivos sean interoperables (Klischewski, 2004). En España, en 2010, empieza a regularse el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica (España, 2010). Sin embargo, en los archivos históricos que han decidido poner a disposición de los usuarios sus fondos, distintos sistemas de redes y software ya existentes han permitido, desde hace más de una década, el completo intercambio de datos e información en Internet.
2. *Datos enlazados o vinculados*. En el entorno de alcanzar la web semántica, la aparición en 1999 de la especificación RDF —Resource Description Framework— de la World Wide Web Consortium (W3C) fue el inicio que permitió poner datos o metadatos en la web para su procesamiento, proporcionando interoperabilidad entre aplicaciones que intercambian información legible por máquina en la web. Tim Berners-Lee (2006) acuñó más tarde el concepto de *Linked Data*, señalando que con los datos vinculados se pueden encontrar en la web otros datos relacionados. No se trata de la Web Hipertexto, sino de la definición de una sintaxis abstracta basada en RDF, que sirve para vincular su sintaxis concreta a su semántica formal incluyendo, entre otros, el tratamiento de referencias URI (Universal Resource Identifier).

## 3 Análisis a partir de casos

La recuperación de la información de los archivos que han puesto a disposición de los usuarios la descripción de sus contenidos en línea se debe poder realizar bien interrogando el productor —ISAAR (CPF) o EAC (CPF)—, bien interrogando los campos de la descripción —principalmente los estándares ISAD (G) o EAD—, bien interrogando su indización —lenguajes documentales, especialmente tesauros, para las materias, geográfico, entidades o personas— o bien sondeando en su interoperabilidad —especialmente a partir de los metadatos RDF—.

### 3.1 Interrogación sobre el productor

En los archivos se describe al productor —bien sean instituciones, personas o familias— con la finalidad de controlar el contexto que ha producido un fondo documental. No tiene sentido realizar esfuerzos en describir productores que no se enlacen, o no se vayan a enlazar, con las descripciones de los fondos documentales. Ni tiene sentido describir en sobremanera a un productor —casi una investigación científica de historia de las instituciones— si esa información no sirve directamente para identificar el contexto y para recuperar la información de los fondos documentales archivísticos. Con la identificación y descripción del productor garantizaremos

también la autenticidad y, si aplicamos, por ejemplo, un OAIS —Open Archival Information System (ISO 14721, 2003)— conseguimos también la conservación de la documentación en nuestro sistema de gestión documental.

En España la principal base de datos que utiliza la ISAAR (CPF), y al mismo tiempo la EAC-CPF, está disponible en el [Censo Guía de Archivos de España e Iberoamérica](#). Contiene un total de 4.324 descripciones, y es posible su búsqueda por personas, familias o instituciones. Las descripciones se pueden también visualizar según la codificación EAC-CPF.



Imagen 1. Descripción de Autoridades en el Censo-Guía (España)



Imagen 2. Visualización codificada EAC

En esta base de datos se debe mejorar la unificación de criterios en algunas, consideramos escasas, descripciones de entidades, como por ejemplo en el caso del *Consejo de Aragón* y el *Consejo Supremo de Aragón*, ya que nos referimos al mismo productor y debe disponer, en consecuencia, de un identificador y no de dos identificadores (ES28079MCU193 y ES.8019.ACA/2). Pero la mejora más destacada que se ha de realizar está relacionada en la misma finalidad de la descripción del productor, recuperar los fondos que éste ha producido. En este sentido, salvo algunas excepciones, la identificación del productor no enlaza con la base de datos que describe los fondos documentales.

A nivel internacional uno de los proyectos más interesantes es [SNAC](#) (Social Networks and Archival Context Project). Se aprovecha del estándar EAC-CPF y de tecnología digital para "desbloquear" las descripciones de las personas desde las fuentes secundarias e integrarlas en nuevas utilidades para, por una parte, crear eficientes herramientas de código abierto que permitan a los archiveros separar el proceso de describir las personas del de describir los documentos o unidades de descripción, y por otra, crear un prototipo integrado de los recursos históricos y sistema de acceso

que permitirá el enlace mutuo de las descripciones de las personas con las descripciones realizadas en los archivos, bibliotecas y museos. Dicho prototipo ya está en funcionamiento (<http://socialarchive.iath.virginia.edu/xtf/search>) y permite efectuar búsquedas sobre personas, familias o instituciones, y enlaza con las descripciones de las instituciones en línea que disponen de esos fondos documentales.



Imagen 3. Prototipo SNAC

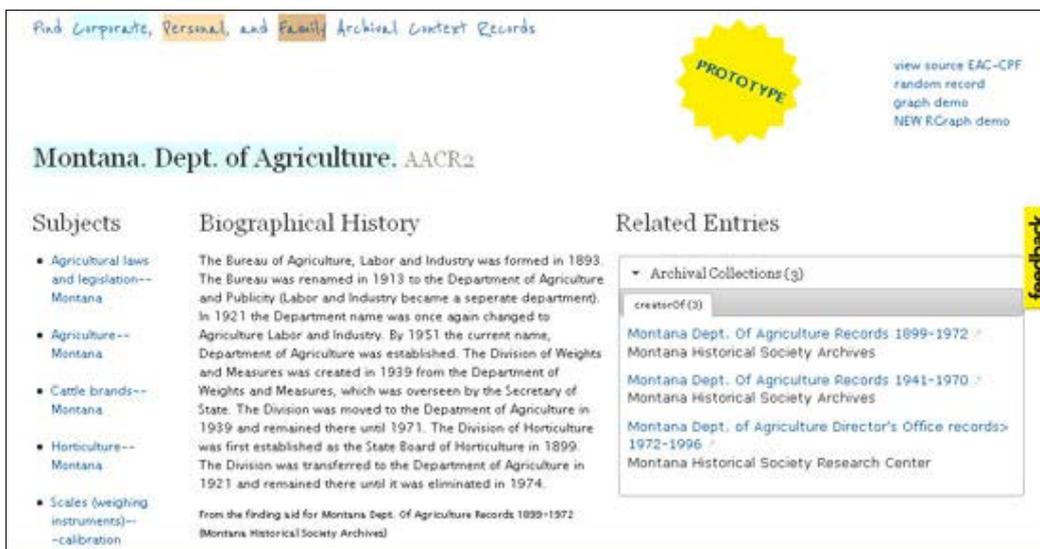


Imagen 4. Prototipo SNAC, ejemplo de descripción de institución y enlace con las unidades de descripción relacionadas

### 3.2 Interrogación sobre los campos de descripción

En los archivos, los campos de las unidades de descripción se regulan principalmente por los estándares ISAD (G) o por las EAD. Una de las principales ventajas de estos estándares es que permiten la descripción multinivel, otra cuestión distinta es que el software que lo soporta sea capaz de integrar la descripción multinivel en su base de datos. La otra ventaja de los estándares es que, al fin, se consiguió en archivos algo que ya ocurría en bibliotecas, hacer posible la interoperabilidad de datos, ya que se ha consensuado una estructura única de campos, y definir la función de esos campos —como en bibliotecas el formato MARC—. Cualquier alteración del número de los campos o de la función de los mismos rompe con uno de los objetivos de estos estándares, hacer posible la creación de redes entre archivos, o la interconexión entre diferentes bases de datos.

En España el [Portal de Archivos Españoles \(PARES\)](#) es el sitio web más importante por el volumen de unidades de descripción realizadas y, como valor añadido, por la disposición de documentos digitalizados en abierto. La utilización de los estándares ha facilitado la interconexión entre los diferentes archivos españoles que están en esta red. La búsqueda avanzada no permite

individualizar la interrogación por todos los campos de la ISAD (G), pero dispone de suficientes cajas de búsqueda para el rastreo de las palabras elegidas en el lenguaje natural utilizado en la descripción, o para el rastreo en los campos fecha, o incluso permite realizar la búsqueda en archivos específicos de la red.

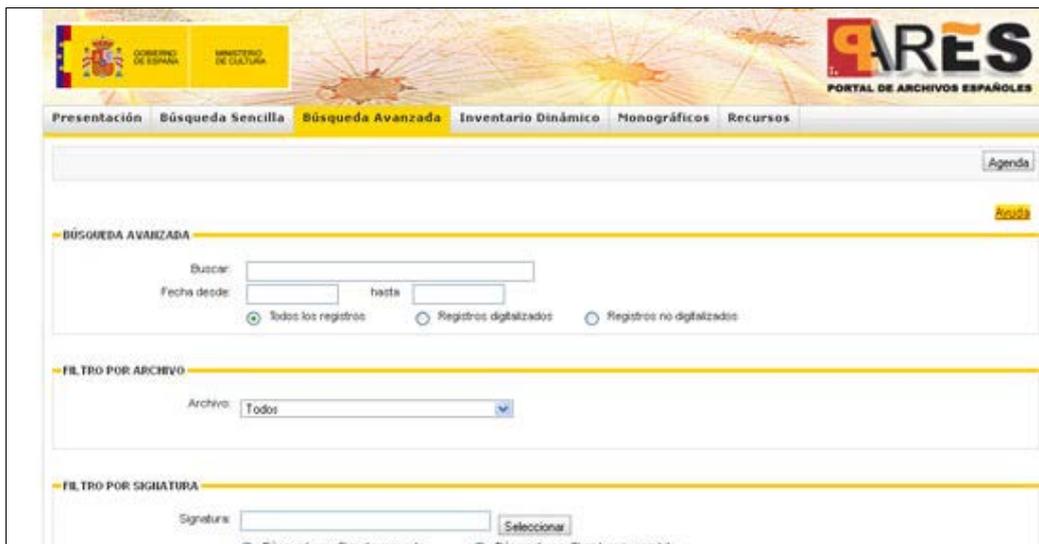


Imagen 5. PARES, Portal de Archivos Españoles

Como ejemplos extranjeros, el [Archivo Nacional de Colombia](#), en su sitio web de consulta documental, permite hacer una consulta por cualquier campo de la ISAD (G) mediante su selección en la casilla "campo", y recuperar por varios campos y según niveles de descripción y fondos seleccionados. Este es un ejemplo del mayor aprovechamiento o integración de los campos ISAD (G), que están en la base de datos, con los campos de recuperación de la información, que están en el OPAC.

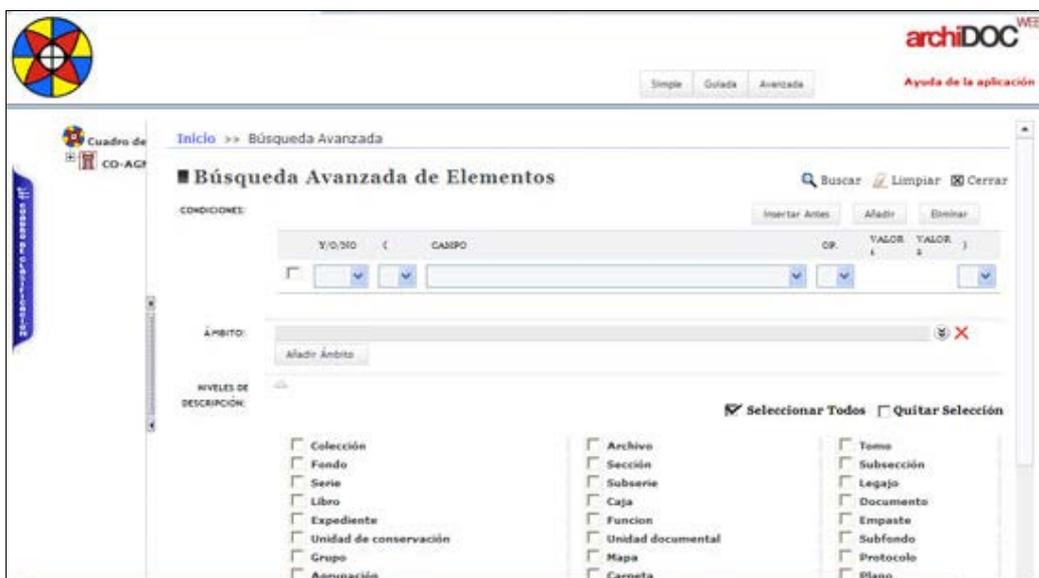


Imagen 6. Archivo Nacional de Colombia. Búsqueda avanzada

### 3.3 Interrogación por la indización

De la misma forma que en los archivos históricos existen o existían las fichas índice, ordenadas alfabéticamente por un vocablo referido a la regesta de la unidad de descripción, en un sistema de gestión documental también es posible y conveniente indizar las descripciones de esas unidades. En bibliotecas es bastante frecuente utilizar una lista como lenguaje documental que controla el vocabulario de la indización, sin embargo, en archivos la práctica que tiene más garantía de eficacia es la utilización de tesauros como lenguaje documental para la indización de materias, entidades o instituciones, geografía o personas (aunque este último, dada la escasa importancia de la jerarquía, muchas veces se utiliza el mismo software para su control, pero excluyendo la jerarquía). Para la indización de los archivos producidos por las organizaciones se están creando

tesauros específicos, como el [EUROVOC](#) para la documentación de la Unión Europea, utilizado también por algunas administraciones autonómicas, u otros tesauros para archivos históricos (Giménez; Escrig, 2011).

Un buen ejemplo de utilización del tesoro para las descripciones con la ISAD (G) es [AIM25](#). Es un sitio web que proporciona acceso en línea a las descripciones realizadas en los archivos de más de cien instituciones, tanto de educación superior, como sociedades científicas, organizaciones culturales o empresas que están situadas en el área metropolitana de Londres. Estas instituciones describen su documentación a diferentes niveles, pero AIM25 sólo recoge las descripciones a nivel colección o fondo. Es un proyecto con más de diez años que crece continuamente. Esta interconexión entre AIM25 y las respectivas entidades —más de 100— es posible porque mantienen la estructura de los campos de la ISAD (G) y su funcionalidad. Pero, además, han conseguido incorporar un tesoro único que controla el vocabulario de los registros catalográficos de todas las instituciones, estructurado por nombres de personas, materias, geográficos y entidades. El sitio web permite navegar por el vocabulario controlado de los tesauros, además de acceder por la institución que tiene depositado el archivo. Este sistema de indización y análisis documental permite recuperar la documentación pertinente.



Imagen 7. AIM25, navegación por palabras clave

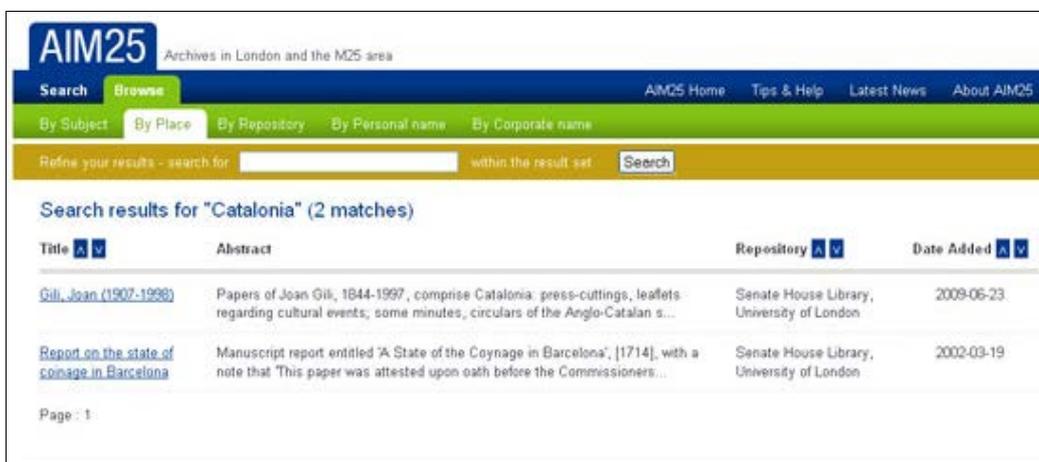


Imagen 8. AIM25. Ejemplo de resultados indizados por "Catalonia"

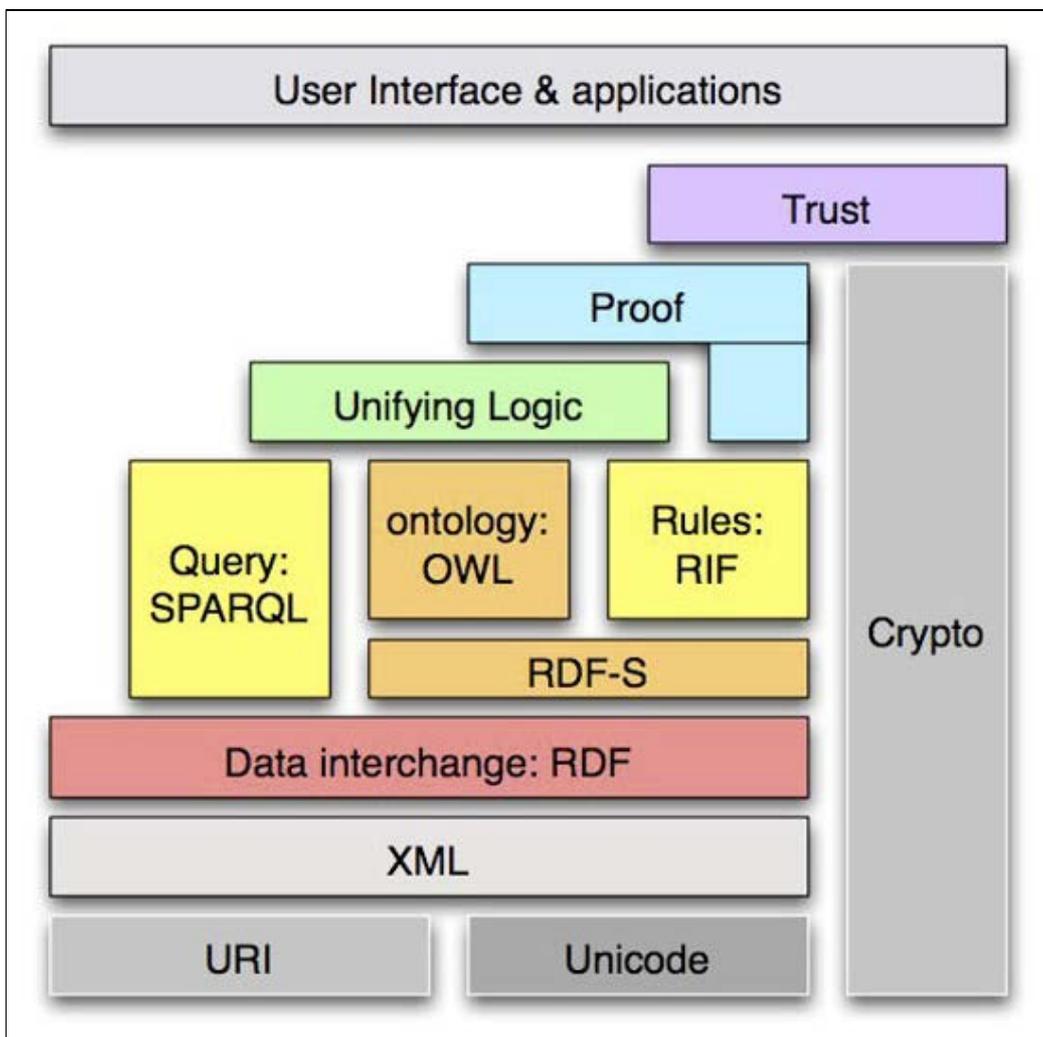
En España, el sitio web PARES también dispone de un instrumento para la búsqueda por la indización. Sin embargo su funcionamiento indica que los diferentes archivos españoles están utilizando lenguaje natural para su indización. No hay referencias en el sitio web de que dispongan de un lenguaje documental. La utilización de un lenguaje natural induce a la sinonimia, ya no que no se da la univocidad de los descriptores. En un sistema de gestión documental, por bases de datos, imposibilita la recuperación total de los documentos pertinentes. Por ejemplo, el concepto *Morisco* lo encontramos indizado como "MORISCOS", "Morisco", "Moriscos de Marbella", "Moriscos" o "Bandolerismo. Moriscos de Marbella" —en este último término han unido tres conceptos, dos materias (*bandolerismo* y *moriscos*) y uno geográfico (*Marbella*)—. Esto imposibilita la recuperación de la información pertinente, dado que el sistema no da los mismos resultados, por ejemplo, para "Moriscos" que para "Bandolerismo. Moriscos de Marbella". Para solucionar el problema debe existir un lenguaje controlado que sea utilizado en todo el sistema.



Imagen 9. PARES, búsqueda por índices

### 3.4 Interoperabilidad Web

Nos referimos a la Web Semántica. Ya tenemos a nuestra disposición instrumentos que hagan posible la interoperabilidad de datos en Internet, simplemente hacen falta las buenas prácticas y el uso de dichos instrumentos para hacer posible que recuperemos información a partir de datos vinculados —Linked Data— de las unidades de descripción. Para ello disponemos de unos mecanismos específicos destinados a las máquinas (Álvarez Espinar, 2008) con los siguientes objetivos: para evitar la ambigüedad en la identificación (URI), para describir los recursos (RDF), para modelar antologías (OWL), para realizar búsquedas en bases de datos (SPARQL), y para expresar las reglas y su intercambio (RIF) —estas especificaciones se pueden encontrar en <http://www.w3.org>—.



Desconocemos si existe algún archivo que haya implementado los mecanismos para la interoperabilidad Web. Sí que se está llevando a cabo en instituciones documentales o bibliotecarias y, evidentemente, lo utilizan también para la documentación procedente de archivos que está depositada en ella. El proyecto más interesante, donde participan instituciones españolas, es [Europeana](#). En Europeana se pueden efectuar búsquedas sobre los fondos de unas 1.500 instituciones colaboradoras. La interoperabilidad implementada permite a las personas explorar los recursos digitales existentes en instituciones como museos, bibliotecas, archivos y colecciones audiovisuales de Europa, todo ello en una red multilingüe. Se pueden encontrar más de 15 millones de artículos, entre los que se incluyen imágenes —pinturas, dibujos, mapas, fotos e imágenes de objetos de museo—, textos —libros, periódicos, cartas, diarios y documentos de archivo—, sonidos —música y palabra hablada en cilindros, cintas, discos y emisiones de radio— y vídeos —películas, noticiarios y programas de TV—. Esta diversidad de documentos y formatos se puede recuperar desde una única plataforma, gracias a las buenas prácticas en el trabajo de descripción realizado y a la implementación de los instrumentos de interoperabilidad y web semántica.

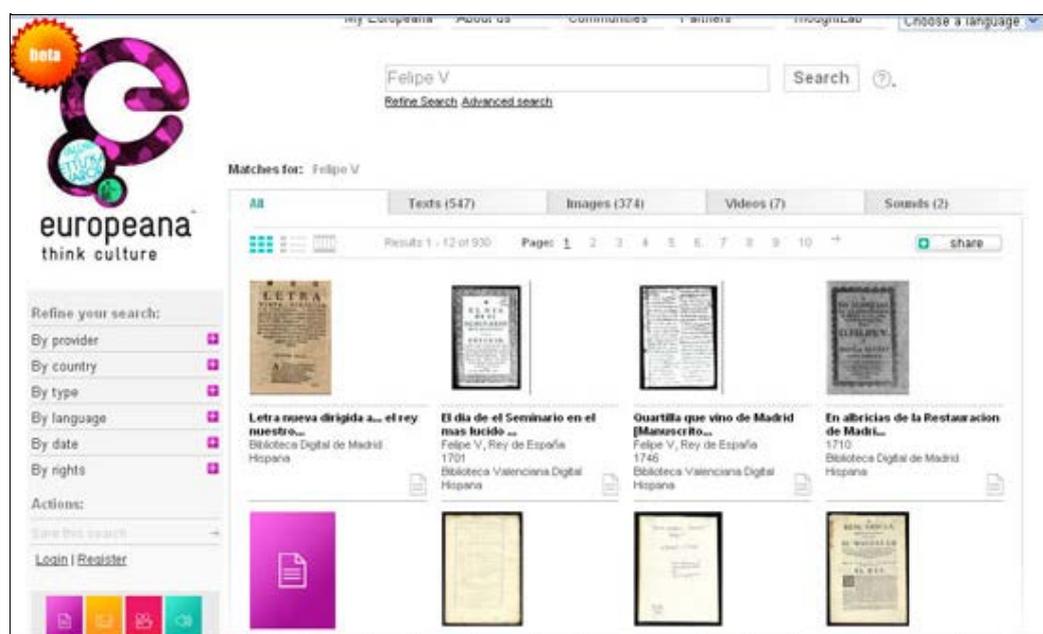


Imagen 11. Europeana. Ejemplo de recuperación de la información mediante el término *Felipe V*: 547 textos, 374 imágenes, 7 audiovisuales y 2 audios

## 4 Conclusión

El trabajo que se ha estado realizando de forma tradicional en los archivos, para poder encontrar aquellos documentos requeridos por una petición, es perfectamente trasladable al entorno Web. Ahora bien, el éxito de una eficaz recuperación de la información a partir de los OPACs de archivos va a depender fundamentalmente de dos factores: de las buenas prácticas en la gestión de documentos —destacaríamos unas correctas clasificaciones para la navegación multinivel, y unas correctas descripciones e indizaciones para la recuperación de la información pertinente— y uso de las TIC, con especial hincapié en su adaptación a la Web Semántica —es el futuro, que va a condicionar la presencia de unas organizaciones o su omisión en Internet—. Esto supone un cambio de paradigma, ahora centrado en el usuario, el cual impone a los archiveros el diseño de los sistemas de acceso a la información y a la creación de servicios orientados a diversos perfiles de usuarios en el marco de la Web (Sebastià Salat, 2009).

El uso de las TIC no es sólo importante para el entorno Internet, sino también para cualquier organización que esté gestionando documentos en intranets. De hecho, la mayor parte de las organizaciones, en su eAdministración, no solamente van a tener la necesidad de depositar documentos digitales en sus servidores o sistemas de almacenamiento, sino también de diseñar la arquitectura de la información y el sistema de recuperación de la información. No es ningún descubrimiento si decimos que, al igual que en los archivos físicos, lo que no se describe no se recupera —aunque utilicemos en uno u en otro otros mecanismos más costosos para recuperar lo

que queremos—, y unas deficientes prácticas en la descripción documental dificultan o imposibilitan la recuperación de los documentos deseados. Disponemos de suficientes instrumentos tecnológicos, incluidos los recientes estándares de la W3C, y de suficientes instrumentos científicos para aplicarlos en nuestras organizaciones, y que no continúe ocurriendo el caso siguiente: las organizaciones depositan grandes volúmenes de documentos (físicos o digitales) sin los adecuados instrumentos para la recuperación de la información.

Si guardamos los documentos, y realizamos diversos trabajos técnicos, es para poder recuperarlos ante las peticiones. Si la recuperación de la información no funciona de una forma eficaz para la organización o los usuarios —en pertinencia y rapidez— todo el trabajo hecho se vuelve inútil.

## Bibliografía

Abadal, E; Codina, LI. (2005). *Bases de datos documentales: características, funciones y método*. Madrid: Síntesis.

Álvarez Espinar, M. (2008). "Interoperabilidad semántica en la Web". *Congreso Nacional de BPMS*. Madrid: W3C. <<http://www.w3c.es/Presentaciones/2008/0220-semanticaBPMS-MA/>>. [Consulta: 18/09/2011].

Arad, A., Bell, L. (1977-1978). "Descripción Archivística. Un sistema general". *Boletín ADPA*, vol. 2, nº 2-3, pág. 35-42.

Bell, L. (1975). "Una investigación sobre el Tratamiento de Datos Archivísticos". *Boletín ADPA*, vol. 1, nº 3, pág. 15-26.

Berners-Lee, T. (2006). *Linked Data*. <<http://www.w3.org/DesignIssues/LinkedData.html>>. [Consulta: 9/09/2011].

Cook, M.; Procter, M. (2000). *Manual of archival description*. Vermont: Gower

Cox, Richard J. (1992). *Managing institutional archives: Foundational Principles and Practices*. Connecticut: Greenwood press

Cumming, Kate (2007). "Purposeful data: the roles and purposes of recordkeeping metadata". *Records Management Journal*, Vol. 17 Iss: 3, pp.186-200

EAC-CPF, *Encoded Archival Context - Corporate Bodies, Persons, and Families*. <<http://eac.staatsbibliothek-berlin.de/>>. [Consulta: 5/09/2011].

España (2010). "Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica". *Boletín Oficial del Estado*, nº 25, 29 de enero de 2010, pág. 8139-8156.

Giménez, V; Escrig, M. (2011). "Designing a Thesaurus to Give Visibility to the Historical Archives in the Archivo del Reino in Valencia". *Knowledge Organization*, 38, Nº 2, p. 153-166.

Heredia Herrera, A. (1988). *Archivística General, Teoría y Práctica*. Sevilla: Diputación Provincial.

ISO 14721:2003 *Space data and information transfer systems - Open archival information system - Reference model*. <[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683)>. [Consulta: 17/09/2011].

Klischeswski, R. (2004). "Information Integration or Process Integration? How to Achieve Interoperability in Administration". *Lecture Notes in Computer Science*, 2004, Vol. 3183, pp. 57-65.

Lodolini, Elio (1993). *Archivística. Principios y problemas*. Madrid: Anabad.

Paul, K.D. (1988). "Archivist and Records Management". *Managing archives and archival institutions*. Chicago: The University of Chicago Press.

Sebastià Salat, M. (2009). "La transformación de los archivos y de la Archivística". *Tabula*, Nº 12, pp. 17-30.

SNAC, *Social Networks and Archival Context Project*.

<<http://socialarchive.iath.virginia.edu/index.html>>. [Consulta: 17/09/2011].

Thibodeau, S. (1995). "Archival Context as Archival Authority Record: The ISAAR (CPF)". *Archivaria* 40, p. 75-85.

<<http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12097/13084>>. [Consulta: 5/09/2011].

Unión Europea (1998). Propuesta de decisión del Consejo por la que se adopta un conjunto de acciones y medidas al objeto de garantizar la interoperabilidad de las redes telemáticas transeuropeas destinadas al intercambio de datos entre administraciones (IDA), así como el acceso a las mismas, *Diario Oficial* n° C 054 de 21/02/1998 p. 0012. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:51997PC0661%2802%29:ES:HTML>>. [Consulta: 9/09/2011].

Vázquez de Parga, M. (1986). "El P.I.A.: Plan de Informatización de Archivos". *Boletín ANABAD*, vol. 36 (1-2), pág. 79-83.

W3C (1999). *Resource Description Framework (RDF) Model and Syntax Specification*.

<<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>. [Consulta: 9/09/2011].

Fecha de recepción: 30/09/2011. Fecha de aceptación: 01/11/2011

---

## Notas

<sup>1</sup> El presente trabajo ha sido realizado en el proyecto Infoscopos (La nueva ecología de la información y la documentación en la sociedad del conocimiento: desarrollo de una métrica sistémica, planificación de un observatorio para su seguimiento e identificación de tendencias básicas y retos estratégicos) es un proyecto I+D subvencionado por el Ministerio de Ciencia e Innovación (CSO2009-0761)

