

## **Detección y eliminación de duplicados en los registros de autoridad para los archivos de historias clínicas de Castilla y León, España**

### **Detection and elimination of duplicate authority records for files histories of Castilla y León, Spain**

**María del Carmen Rodríguez López**

Doctora en Historia. Profesora Titular. Departamento de Patrimonio Artístico y Documental. Facultad de Filosofía y Letras. Universidad de León. España.

---

#### **RESUMEN**

Con el objetivo de definir un fichero de pacientes que actuara como un catálogo de historias clínicas, es decir, como una relación ordenada del conjunto de materiales que existen en una unidad documental, con indicación mediante un símbolo (signatura o número de historia clínica del lugar que ocupan los documentos), se desarrolló un mecanismo que sirviera, tanto para su identificación, como para su ordenamiento y acceso lógico a partir de sus descripciones. El mecanismo se sirve de la asignación de los puntos de acceso, que bajo la forma de encabezamientos ofrecen una mayor flexibilidad al fichero. El presente proyecto buscó determinar el tipo de información que puede incluirse en un registro de autoridad de archivos y servir de orientación sobre la forma de integrar estos registros en un sistema descriptivo controlado de historias clínicas. El contenido de la información proporcionada en el registro de autoridad se detalla con ayuda de las *reglas nacionales de catalogación*, que se basan en normas, directrices, especificaciones y modelos de referencia internacionales.

**Palabras clave:** Archivo de historias clínicas, recuperación de información, encabezamientos, autoridades, detección de duplicados, resolución de duplicados, complejo asistencial de León, España.

---

#### **ABSTRACT**

In order to define a file of patients to act as a catalogue of clinical records, an ordered relationship of all materials in a documentary unit, indicating by symbol: signature or number clinical history the location of documents developed a mechanism serve both for identification as to their logical access from descriptions. The mechanism uses the allocation of access points, which in the form of headings

offer greater flexibility to the file. This project sought to determine the type of information that can be included in a register of authority files and provide guidance on how to integrate these records controlled in a descriptive system of medical records. The content of the information provided in the authority record is detailed with the aid of the *National Cataloguing Rules*, based on standards, guidelines, specifications and international reference models.

**Key words:** Clinical records archive, information retrieval, headings, authorities, detection of duplications, resolving duplications, health care complex of Leon, Spain.

---

El informe que ahora se presenta se corresponde con un proyecto de investigación, recientemente finalizado, financiado por la Junta de Castilla y León, titulado *Registros de autoridad para los archivos de historias clínicas de Castilla y León: el camino a la informatización de la historia clínica. Paso previo a la eliminación de duplicados*, y que fue uno de los proyectos que la Dirección General de Planificación y Ordenación de la Junta de Castilla y León subvencionó en la *Convocatoria de ayudas de la Junta de Castilla y León para proyectos de investigación en materia de gestión sanitaria en el año 2005* (ORDEN SAN/163/2005, de 9 de febrero, por la que se convocan subvenciones para la realización de proyectos de investigación en materia de gestión sanitaria en el año 2005).

Además, es continuación de un proyecto anterior, desarrollado por este equipo investigador y dirigido por la Doctora *Rodríguez López*, quien trató sobre la *Evaluación del Sistema Archivístico Sanitario de León. Ponderación del valor económico de su gestión*, cuyos objetivos fueron:\*

- Conocer la tendencia en la indización de las historias clínicas en los hospitales castellano-leoneses.
- Establecer su influencia en la adecuada recuperación de la información.
- Censar el volumen de duplicados generados anualmente.
- Relacionar la incorrecta asignación de registros de autoridad con la generación de duplicados.
- Establecer unas bases generales para asignar registros de autoridad en el archivo de historias clínicas del Complejo Hospitalario de León, relacionadas con los patronímicos de los principales usuarios del sistema sanitario.

## MÉTODOS

Se diseñó el trabajo en dos secuencias diferentes: la primera, destinada al acopio de información; la segunda, al procesamiento de los datos. El acopio de información, se subdividió en tres fases:

1. Revisión y validación de la legislación y bibliografía según los repertorios de jurisprudencia, bases de datos, páginas Webs y centros de información bibliográfica. Revisión de la normativa internacional emanada de los organismos de normalización como IFLA y CIA (Consejo Internacional de Archivos).
2. Manejo de la lista de duplicados en el Archivo del Hospital de León, obtenido entre dos fechas: 24 de julio de 2004 y 25 de octubre de 2005.

3. Consulta de archivos de centros habituados al trabajo con usuarios procedentes de otros países: archivos municipales, registro civil, etcétera.

A partir de los datos obtenidos, se trató de establecer algunas correlaciones hasta determinar, entre otros, el volumen de duplicados de historias clínicas generados y resueltos al año, y cuáles son los datos especialmente críticos para alcanzar los objetivos y la misión corporativa. La revisión de la documentación permitió determinar una serie de necesidades como:

- Normas generales para la elección del encabezamiento.
- Elección del elemento inicial.
- Resolución de la casuística hallada.
- Alfabetización de los apellidos precedidos de partículas.
- Resolución de los casos de pacientes europeos.
- Alfabetización de los apellidos de las poblaciones de inmigración mayoritaria: la casuística de los nombres árabes.
- Otros: nombres: hebreos, chinos, birmanos, hindúes, indonesios, tailandeses, turcos.
- Transliteración de caracteres procedentes de alfabetos diferentes: cirílico, etcétera.

## PROCESAMIENTO DE LA INFORMACIÓN

El trabajo se centró en la listado de duplicados del archivo de historias clínicas del Complejo Asistencial de León, que aglutina tres hospitales: San Antonio Abad, Hospital Princesa Sofía y Residencia Virgen Blanca. Además, proporciona servicio al Hospital Monte San Isidro y Santa Isabel. La muestra se consideró suficientemente representativa, por tratarse de un centro con un volumen documental importante (486 526 historias clínicas) y una antigüedad que se prolonga en el tiempo más que suficiente (el 9 de octubre de 1919 se erigió la primera piedra del Hospital de San Antonio Abad). Se trata de un complejo hospitalario de 850 camas, al que desde 1991 corresponde la asistencia del área sanitaria de León que, según los datos del Insalud de 1999, abarca una población de 366 385 habitantes, de los que 172 898 corresponden a zonas urbanas (León capital, La Bañeza y Astorga) y 193 487 a municipios rurales.

La finalidad principal de su archivo de historias clínicas es la custodia, organización, recuperación y préstamo de las historias clínicas, porque estas aseguran la prestación de la asistencia sanitaria a los pacientes y su continuidad. Otra función primordial del archivo hospitalario es el mantenimiento y ordenamiento de la colección documental abierta, de manera que integre todos los documentos relacionados con el paciente. Para cumplir estas funciones, el principal recurso es el propio fondo del hospital, compuesto por una única serie integrada por cada historia clínica personal que se le abre a un paciente, con motivo de su asistencia al hospital. Cada historia está integrada por un conjunto de documentos generados en los distintos episodios por los que transcurre el paciente en su paso por el hospital, distribuidos en ocho bloques. Cuando se habla de duplicados en el presente informe, se hace referencia a la existencia de más de una historia clínica por paciente; así, esta segunda o tercera (e incluso puede darse en mayor número) historia constituye un duplicado de la original.

Una vez determinada cómo quedaría integrada la muestra, el procesamiento primario continuó con un segundo paso, consistente en el examen de los mecanismos de recuperación de la información empleados por el propio archivo. El archivo tiene informatizado el control de la información, pero no la información en

sí. Así, por ejemplo, están informatizados los ficheros de pacientes, de préstamos, etcétera. Es decir, que se informatiza el total de los registros que permiten y facilitan el control de la información. El fichero más utilizado es el de pacientes, en el que se graban todos los datos de ellos y el correspondiente número de historia clínica. El principal problema detectado en el funcionamiento de este fichero es que no detecta los duplicados, y existe un nutrido número de estos por causa de la diversidad de hábitos de alfabetización existentes en la comunidad médica.

Los datos incluidos en el fichero de pacientes son los necesarios para la asistencia, distribuidos en dos conceptos: datos básicos y datos de contacto.

- *Datos básicos:* se asignan al paciente, salvo excepciones, para toda la vida. Le acompañan en todos los episodios asistenciales. Son, a saber: fecha y lugar de nacimiento, DNI/NIF o número de pasaporte.
- *Datos de contacto:* por el contrario, son variables en función de las vicisitudes vitales del paciente. En este grupo se registran: régimen y número de la seguridad social, domicilio y teléfono.

En tercer lugar, se procedió al acceso a los campos homogeneizados de la lista de duplicados, formado por las siguientes entradas:

Apellido o 1	Apellido o 2	Nombre	Sexo	No. H.C.	Fecha de nacimiento	Provincia de nacimiento	Provincia de residencia	Domicilio	DNI
-----------------	-----------------	--------	------	-------------	------------------------	-------------------------------	-------------------------------	-----------	-----

El volumen de duplicados que se resuelven al año en el Complejo Asistencial de León asciende a 1 000, pero el volumen generado permaneció sin cuantificar hasta el momento de la investigación. El muestreo se hizo sobre una lista de duplicados, obtenido en dos fechas diferentes:

1. Vaciado de los duplicados, listados el 26-07-2004. Letras: C, D, E, F, G, H, I, J, K, L, M, N, Ñ, O, P, Q, R, sólo hasta Rodríguez García.
2. Vaciado de los duplicados, listados el 25-10-2005, Letra Z.

Las dos listas cuentan con un total de 183 páginas. Cada página consta de 17 casillas con 2 entradas. Así se obtiene el número total de entradas revisadas, que es de 6 222. Es pertinente aclarar que no todas las entradas contenían duplicados, porque se pudo apreciar que en muchos casos no se producía coincidencia, aunque formaban parte de la lista.

El recuento de datos fue verificado únicamente sobre los campos que afectan el control de autoridades:

Apellido 1	Apellido 2	Nombre
---------------	---------------	--------

## RESULTADOS

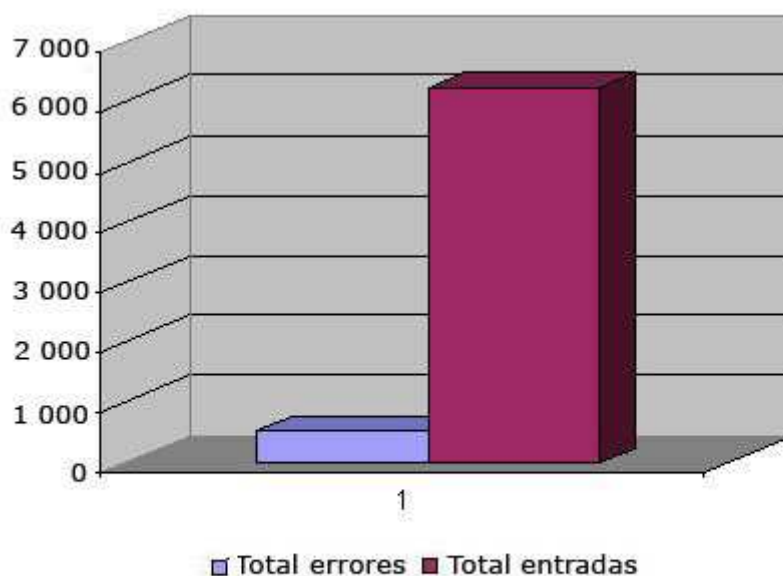
### DETECCIÓN DE ERRORES

Una vez detectados los problemas que causan duplicados, se agruparon en siete *ítems*. Estos *ítems* se pueden subdividir en los que corresponden al nombre o nombres, es decir, nombres compuestos, nombres problemáticos, errores tipográficos, nombres con iniciales solamente y registros en los que no consta el nombre. En lo que respecta a los apellidos, los principales problemas son por errores tipográficos y apellidos compuestos (tabla).

**Tabla.** *Ítems* que generan duplicados

Errores más comunes	Número de errores	%
Nombres compuestos	214	40,00
Nombres problemáticos (similares a otros)	161	30,09
Errores tipográficos en el nombre	67	12,52
Iniciales del nombre sin desarrollar	17	3,18
Registros sin nombre	25	4,67
Errores tipográficos en el apellido	45	8,41
Apellidos compuestos	6	1,12
Total	535	–

El total de historias clínicas duplicadas por las causas mencionadas es de 535. Es perceptible su baja proporción en relación con el total de errores detectados (figura 1). Es necesario hacer una llamada de atención a la escasa casuística producida, contra todo pronóstico, por los apellidos. Son, entonces, los nombres los causantes del mayor número de incidencias. La mayor problemática está en los nombres, frente al control de los apellidos que parece correcto.



**Fig. 1.** Proporción de errores sobre el total de las entradas.

#### DUPLICADOS ATRIBUIBLES A NOMBRES COMPUESTOS

Se desprende como primera observación, que el problema de generación de duplicados radica en los nombres, sobre todo en los nombres compuestos. Estos suponen el 40 % de los casos (214). Refleja la práctica habitual de proporcionar en unas ocasiones el primer nombre y en otras los dos (figura 2).



**Fig. 2.** Porcentaje de nombres y apellidos.

#### DUPLICADOS ATRIBUIBLES A NOMBRES PROBLEMÁTICOS

Esta es la segunda causa, con 30,09 % de los casos (161). Existe una amplia casuística, como la que se muestra a continuación:

- Los generados por aquellos nombres que pueden asimilarse sin dificultad, por corresponder a nombres que se emplean de forma indistinta en el lenguaje habitual: *Isidro-Isidoro/ Ascensión-Asunción/ Consolación-Consuelo*.
- Otro grupo corresponde a nombres que han recibido una nueva versión más moderna: *Victoriano-Victorino/ Victorina-Victoria/ Antonino-Antonio*.
- c) Aquellos nombres que se han modificado en su pronunciación en el uso diario: *Benedicta-Benita*.
- d) Uso de diminutivos o reducción del nombre, que puede dar lugar a cambios significativos hasta el punto de convertirse en nombres diferentes: *Carol-Carolina/ Rosa-Rosario/ Josefina-Josefa/ Eva-Evangelina*.

#### DUPLICADOS ATRIBUIBLES A ERRORES TIPOGRÁFICOS EN EL NOMBRE

Nos referimos a aquellos errores que se cometen derivados de la lectura rápida de los datos del DNI o de otros impresos. Este tipo de errores ocupa el tercer lugar, con 67 casos, que suponen el 12,62 % del total de duplicados. Como ejemplos se encuentran los nombres de: *Amalia-Amelia/Argelina-Arselina/ Licinio-Lucinio/ Eufemia-Eugenia/ Enedías-Enedina*.

#### DUPLICADOS POR APARECER SOLO LAS INICIALES DEL NOMBRE

Como consecuencia de este tipo de errores, sólo nos encontramos 17 casos, que representan 3,18 % del total de duplicados. No se requiere de aclaración, salvo en el caso de *María*, que puede aparecer desarrollado solamente en inicial, e incluso no se graba. *El Manual de normas del archivo y uso de la historia clínica del Hospital de León*, aconseja "no grabar María". Una de las normas básicas adoptadas internacionalmente es la de consignar el nombre completo, y sobre todo aquel que sea el preferido por el usuario. María podría contribuir a deshacer casos de homonimia.

#### DUPLICADOS POR LA AUSENCIA DE NOMBRE

Se detectaron 25 casos (4,67 % de los errores). Estos son fácilmente explicados por la rapidez con que a veces se toman los datos. Para estos no existiría una recomendación. Responden a un error humano, que se ha de asumir por ser fácilmente subsanables si el propio sistema no dejara continuar, por definirse este campo (nombre) como obligatorio. Sin embargo, tan drástica decisión sería problemática igualmente porque, al explorar un poco más, resultó fácil concluir que en muchos casos se trataba de neonatos, sin nombre de pila establecido.

#### DUPLICADOS ATRIBUIBLES A ERRORES TIPOGRÁFICOS EN EL APELLIDO

Los duplicados por errores en los apellidos no son muy numerosos; concretamente se detectaron 45 casos que sólo suponen el 8,41 % del total. Entre los casos detectados se encuentran:

- La ausencia de acentos, los errores tipográficos y la permuta de letras: *Diez Morcillo-Díez Morillo/ Ferrer Cela-Ferrer Celar/ Franco Escudero-Franco Escuredo*.
- Escasa homogeneidad ortográfica: *Domínguez Borge-Domínguez Borja/*

## DUPLICADOS EN APELLIDOS COMPUESTOS

El grupo menos numeroso parece el más problemático, si no se consideraran los guiones de separación entre el primer y segundo apellidos. Solo se detectaron 6 casos, que constituyen un escaso 1,12 % de ellos. Como ejemplo: Díaz Rada Martín-Díaz/ Rada Martín Navarrete.

En este grupo conviene hacer una llamada sobre los apellidos con preposición. Si bien aparentemente no se encuentra ningún caso, sí es posible detectar su existencia: por ejemplo, García Río. Lo más probable es que se trate de García del Río, pero en el caso del Hospital de León, *El manual de normas del archivo*, citado antes indica "no hay que grabar los de" tratando (creemos entender) de responder a la regla que establece que artículos y preposiciones no alfabetizan, pero esa regla no resulta equivalente a la aplicada en el *Manual de normas del archivo*, es decir, que se han de consignar las preposiciones detrás del nombre, pero no deben obviarse.

## DUPLICADOS GENERADOS POR NOMBRES EXTRANJEROS

Como hipótesis de partida, se había establecido que la casuística en los nombres o apellidos serían los que más problemas causarían por la tipografía y por la distinta alfabetización que se producen en los distintos países. Sin embargo, sólo se detectaron 9 casos. Estos se corresponden también a las comunidades más establecidas en León. Al final se comportaron como un ítem más; la problemática que presentan es idéntica a la de los nombres nacionales: variantes de los nombres de pila, sobre todo, como tales están computados con todos los demás. Sin embargo, recogemos un catálogo de esta casuística:

- Portuguesa: con la partícula «Dos» como apellido: *Dos Anjos, José Carlos/Dos Anjos Correia, Azucena*.
- China, por tomar a veces el nombre fonéticamente o errores tipográficos: *Kwon Lee, Hyunsluk/ Kwon Lee, Hyun Chuk/ Zhou Ye, Shen Ce/ Zhou Ye, Shence*.
- Árabes, también, por las mismas causas: *Mahmad, Mrahri/ Zaaoti, Sami/ Zragdgui, Rachida/ Zradgui, Rachira*.
- De países del este europeo: *Zabek, Wieslaw Jan/ Zagoruyko, Oleksamdr /Zagoruyko, Oleksandr*.

Excepción hecha del caso portugués, en que el problema radica en el apellido iniciado por la partícula "dos", los demás casos reproducen la problemática de los nombres oriundos: el duplicado tiene su origen en el nombre, no en el apellido. Para el caso portugués, la solución se presenta fácilmente, con solo acudir a las *Reglas de catalogación, Partículas. 15.1.2Ba*): "Portugués: se comienza con la parte que sigue a la partícula. Ej. Figueredo, Fidelino de; Almeida, Julio Lopes de; Santos, Joao dos".<sup>1</sup>

## DUPLICADOS ATRIBUIBLES A OTROS FACTORES AJENOS AL CONTROL DE AUTORIDADES

Reiteramos que el problema parte de los campos alfabéticos, no de los numéricos. El campo "nombre" produce el 90 % de los duplicados atribuibles al control de autoridades, pero entre estos y los producidos por los campos "apellido 1" y "apellido 2" resulta un escueto 9 % de la lista (figuras 3 y 4). Es el restante campo



alfabético el auténtico creador de duplicados: el campo destinado a "Domicilio", porque la "Provincia de nacimiento" y la "Provincia de residencia" se indican con un código numérico.



Fig. 3. Representación proporcional de los errores.

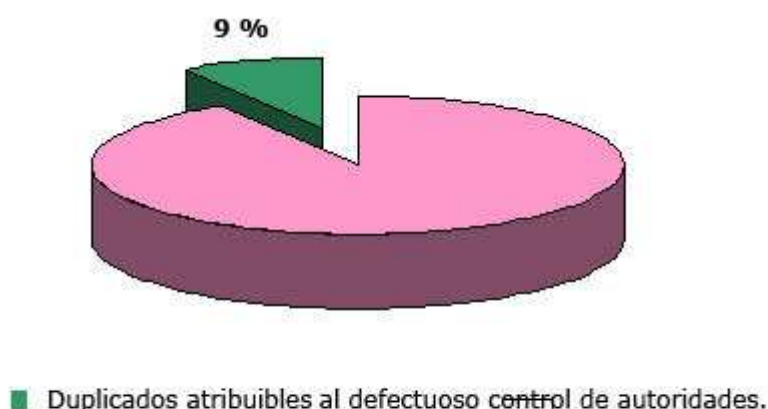


Fig. 4. Porcentaje de duplicados generados por un defectuoso control de autoridades.

Del examen riguroso de los datos introducidos en el domicilio se desprende que el sistema automatizado interpreta cada bloque alfabético como una palabra. Si al introducir los datos se cambia alguno de estos, el sistema lo interpreta como un paciente diferente. Así, el sistema interpreta de forma diferente las indicaciones: *Calle, C/* o *C*. Del mismo modo que resulta diferente que los datos vayan precedidos o seguidos de una coma: *Calle X, número 1* o *Calle X número 1*. Por supuesto, tampoco identifica como idénticas las indicaciones: *nº, número* o *n*.

Si se tienen en cuenta los elementos señalados al introducir los datos del paciente, estamos en condiciones de determinar que, en buena medida, una de las causas por las que un mismo usuario puede ser identificado como dos personas diferentes

es el uso habitual de dos residencias. Se ha constatado, en algunos casos, que si se produce la necesidad de asistencia durante los meses de verano, el domicilio que se proporciona es el del pueblo; si, por el contrario, se produce en los meses invernales, se comunica como domicilio el de la vivienda de la ciudad. Un tercer domicilio se proporciona cuando, por la avanzada edad, ha resultado necesario el trasladado para vivir con los hijos o parientes directos; y en algunos casos se ha podido rastrear un cuarto domicilio que corresponde a una residencia geriátrica.

## CONCLUSIONES

Se concluye que no existen ficheros de autoridades acordes con la normativa existente. La entrada de datos sobre un mismo paciente puede producirse bajo diferentes encabezamientos. También existe una relación directa entre los dos puntos anteriores y la generación de duplicados. Es muy escaso el porcentaje de duplicados generados por nombres no españoles. El mayor porcentaje de duplicados se produce en los nombres frente a los apellidos. En estos el porcentaje es reducido. El problema principal radica en los campos alfabéticos destinados a domicilio. La homonimia se desharía con los datos numéricos: DNI o número de afiliación a la seguridad social. Existen dificultades para recuperar la información, que redundan en el costo de la estancia del paciente en el centro.

## Agradecimientos

A María Luisa Alvite Diez, Ángela Diez Diez, Josefa Gallego Lorenzo, María Antonia Morán Suárez, Blanca Rodríguez Bravo y Lourdes Santos de Paz, pertenecientes a los departamentos de Patrimonio Artístico y Documental y de Ingeniería Eléctrica y Electrónica de la Universidad de León.

## REFERENCIAS BIBLIOGRÁFICAS

1. Dirección General del Libro Archivos y Bibliotecas. Reglas de catalogación. Madrid: Dirección General del Libro Archivos y Bibliotecas; 1999.

Recibido: 7 de enero de 2010.  
Aprobado: 26 de enero de 2010.

Dra. *María del Carmen Rodríguez López*. Departamento de Patrimonio Artístico y Documental. Facultad de Filosofía y Letras. Universidad de León. Campus de Vegazana s/n 24071 León. España. Correo electrónico: [mcrodl@unileon.es](mailto:mcrodl@unileon.es)

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS<sup>1</sup>  
REGISTROS MÉDICOS; SISTEMAS DE INFORMACIÓN; PROCESAMIENTO

AUTOMATIZADO DE DATOS.  
MEDICAL RECORDS, INFORMATION SYSTEMS; AUTOMATIC DATA PROCESSING.

Según DeCI<sup>2</sup>  
REGISTROS ELECTRÓNICOS.  
ELECTRONIC RECORDS.

<sup>1</sup>BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2009. Disponible en: <http://decs.bvs.br/E/homepagee.htm>

<sup>2</sup>Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Rodríguez López MC. Detección y eliminación de duplicados en los registros de autoridad para los archivos de historias clínicas de Castilla y León, España. Acimed. 2010;21(2). Disponible en dirección electrónica de la contribución [consultado: día/mes/año].

---

\*Convocatoria de ayudas de la Junta de Castilla y León para proyectos de investigación en materia de gestión sanitaria en el año 2004 (Orden SAN/526/2004, de 7 de abril, B.O.C. y L. de 16 de abril de 2004). Este proyecto se ha difundido en forma variada: dos pósteres: uno de ellos titulado *Evaluation of the Archivist System of the Hospital de León (España): establishment and operation*. Presentado en el *9th European Conference of Medical and Health Libraries*, celebrado en Santander los días 20 a 25 de septiembre de 2004. El segundo bajo el título: *Ponderación del valor económico de la gestión del archivo del Hospital de León puesta en relación con el Servicio de Urología*, presentado en el *IX Congreso de Metodología de las Ciencias Sociales y de la Salud*. El planteamiento inicial se presentó en los *IX Encuentros Internacionales sobre Sistemas de Información y Documentación (IBERSID)*, celebrados en Zaragoza del 2 al 4 de noviembre de 2004, en una comunicación sobre *Evaluación del Sistema archivístico sanitario de León*. Finalmente, un artículo que vio la luz en el mes de marzo de 2006 en la revista *El Profesional de la Información* con título homónimo al anterior pero con los resultados finales de la investigación. El equipo que ejecutó estos proyectos está formado por la profesora que firma el artículo y las que figuran en los agradecimientos, además de Raquel Vega Ampudia, Licenciada en Filología, Archivera-Bibliotecaria del Hospital de León y Tania Cid González, Licenciada en Arte y Diplomada en Biblioteconomía y Documentación, Ayudante de Archivos de la Junta de Castilla y León.