
ESTUDIOS / RESEARCH STUDIES

Análisis de la variabilidad de nombres de autores españoles en depósitos digitales universitarios de acceso abierto: un estudio por áreas de conocimiento

Enrique Orduña-Malea*, Fernanda Peset*, Antonia Ferrer-Sapena*

Resumen: Los autores que firman con su nombre oficial español (dos apellidos y a veces dos nombres) o que varían su firma a lo largo del tiempo, tendrán más dificultades para reunir su producción científica. A partir de la experiencia con *IraLIS (International Registry for Authors: Library and Information Science = Links to Identify Scientists; <http://www.iralis.org>)*, este trabajo estudia cómo aparecen registradas las autoridades personales en dos proyectos españoles, de entorno no propietario, que cumplen con el protocolo *OAI-PMH: Recercat y E-Prints Complutense*.

Se descarga la base de datos de autoridades de los dos repositorios estudiados, se clasifica el estilo de firma —recomendando la que contiene dos únicas ristas de elementos— y se distribuyen los resultados según el sector de conocimiento, usando para ello las clasificaciones temáticas del proyecto *Intute* y de la *Agencia Nacional de Evaluación y Prospectiva (ANEP)*.

El estilo de firma más usado es mediante dos apellidos (76 % en *Recercat* y 94 % en *E-Prints Complutense*). En cuanto a áreas temáticas, las ciencias sociales son las que presentan un mayor porcentaje de registros de autor con estilos de firma recomendados (un 20% del total de registros pertenecientes a esta área).

El nivel de uso del estilo de firma recomendado en ambos repositorios es bajo, especialmente en *E-Prints Complutense*. El índice de firma recomendado por sector científico nos muestra el paisaje real de los mayores repositorios de acceso abierto españoles. Con esta información se puede diseñar una campaña de concienciación personalizada por sectores. Hipotéticamente pensamos que las ciencias más internacionalizadas o los autores vinculados a la documentación tendrán en cuenta estos aspectos de la firma, pero el estudio estadístico no lo ha corroborado, ya que son las ciencias sociales las que presentan el mayor nivel de uso de estilos de firma recomendados. Según los datos, las áreas tecnológicas, de salud y ciencias de la vida y de arte y humanidades precisan de campañas de concienciación evidentes. Sería deseable un estudio de las posibles diferencias en el uso de estilos de firma en función de quién alimenta las bases de datos.

Palabras clave: normalización, control de autoría, repositorios de acceso abierto, estilo de firma recomendada, recercat, e-prints, complutense.

* Facultad de Informática. Universidad Politécnica de Valencia. Correo-e: enorma@upv.es, mpesetm@upv.es, anfersa@upv.es

Recibido: 14-11-09; 2.^a versión: 18-3-09; 3.^a versión: 8-4-09.

Variation in Spanish authors' names in university Open Access repositories with respect to different fields of knowledge

Abstract: Spanish authors who sign with their official Spanish name (which includes two surnames and sometimes two first names) or who change their signing name over time will find it more difficult to collect together their scientific output. Making use of experience with *IraLIS* (International Registry for Authors: Library and Information Science = Links to Identify Scientists; <http://www.iralis.org>), this study examines how authorities are registered in two Spanish OAI-PMH-compliant projects: *Recercat* and *E-Prints Complutense*.

The authorities database of the two repositories analysed were downloaded. The style of signing - the recommended one should contain only two elements - was recorded by field of knowledge, using the topic classification systems of the *Intute* project and the *ANEP* agency.

The most commonly used signing style involved two surnames (76% in *Recercat* and 94% in *E-Prints Complutense*). The social sciences showed the greatest adoption of the recommended signing style (20% of all records in this area).

The adoption rate of the recommended signing style in both repositories was low, especially in *E-Prints Complutense*. The index of the use of the recommended signing style by knowledge field reveals a true picture of these Open Access repositories. This information could allow awareness campaigns to be designed for different fields of knowledge. It might be thought that the most internationalised sciences and the authors of its literature would take signing style into account, but this does not seem to be the case. Indeed, the social sciences were those which most frequently followed the recommended system. The areas of technology, health and life sciences, the arts and humanities should be the subjects of awareness campaigns. Further work should study the possible differences in the use of signing style depending on who inputs information into databases.

Keywords: standardization, authorship control, open access repositories, recommended signature style, *recercat*, *e-prints complutense*.

1. Introducción

1.1. La variabilidad en la forma de los nombres

La normalización del nombre de los autores es un tema que se ha estudiado de diferentes formas y en diferentes ámbitos, existiendo una amplia y diversa cantidad de iniciativas que lo analizan (Danskin, 2007).

Estas iniciativas suelen centrarse fundamentalmente en tres áreas: las relacionadas con el diseño de esquemas para la descripción de registros de autoridades (desde las *ISAAR CPF* hasta los más recientes esquemas de metadatos como *MADS* o microformatos como *VCard*), las relacionadas con la creación de identificadores únicos, como el *ISAN* (Snyman, 2000) o *DAI* (Spanje, 2007) y finalmente las relacionadas con procesos de desambiguación de nombres en bases de datos (Torvik, 2005; Wooding, 2006; Gálvez, 2007). Asimismo, destaca el reciente trabajo aportado por el centro de investigación *Ingenio* (Pinar, 2007).

En cambio, existe un área de estudio, generalmente olvidada, que corresponde al estilo de firma de los autores, donde destacan los trabajos relacionados con la redacción de guías de estilos recomendados de firma (FECYT, 2007). Entre éstas destaca en el ámbito español el proyecto *Iralis* (Baiget, 2007), que permite la asignación de estilos de firma recomendados, la creación de registros de variantes de firma así como la realización de oportunas campañas de concienciación entre los autores. Así mismo, en el ámbito anglosajón destaca la iniciativa *Names* (*Joint Information System Committee*, 2008).

En este punto resulta conveniente diferenciar entre la forma normalizada de un nombre de autor y el estilo de firma del mismo.

La forma normalizada (o autorizada) de un nombre de autor es aquella que resulta de aplicar los criterios de normalización establecidos por los códigos nacionales e internacionales, aspecto en el que no se centra este trabajo.

El estilo de firma, por otra parte, es la forma mediante la cual un autor firma sus trabajos. Si un autor ha adoptado un determinado estilo (desde el inicio de su carrera o a partir de un momento determinado), se habla de «nombre de pluma», siendo este concepto en el que se basa este estudio.

El problema radica en cómo un nombre español, expresado a través de un determinado estilo de firma, es interpretado como normalizado cuando es indizado, recogido o tratado por distintos entornos. Así, la forma del nombre con la que un autor aparece en las bases de datos puede variar debido, entre otras razones, a:

- La posibilidad de que un autor pueda firmar sus trabajos con estilos diferentes a lo largo de su vida, debido al desconocimiento o despreocupación de los problemas que ello puede ocasionar.
- La interpretación que los productores de bases de datos hacen del nombre, debido fundamentalmente a diferencias culturales o limitaciones técnicas.
- La posibilidad de que un usuario (autor o no) envíe un artículo a un repositorio y pueda registrar, en el metadato del campo «autor», una forma de nombre distinta a la que figura realmente en el artículo y, además, ser ésta la forma que se use para la recuperación dentro del sistema.

La estructura de los nombres en español, formada por tres elementos (dos apellidos y un nombre, cada uno de los cuales puede ser simple o compuesto), favorece tanto las variaciones de nombre como las distintas interpretaciones del mismo.

La importancia de esta área de estudio radica, por tanto, en la conveniencia y necesidad, por parte de un autor, de usar un nombre de pluma adecuado durante toda su vida científica para minimizar (que no eliminar) los problemas en la recuperación de su producción documental en los distintos proveedores de datos y de servicios donde aparezcan referenciados sus trabajos.

Con esto se pretende indicar que un autor debería elegir un estilo de firma que cumpliera unos requisitos recomendados y además mantenerlo durante toda su vida profesional.

A efectos de este trabajo, se propone como estilo de firma recomendado aquel que cumple con las especificaciones del formato *IraLIS*, esto es, un estilo de firma compuesto por dos ristas de caracteres

Este estilo de firma hace referencia al uso de dos cadenas de caracteres (generalmente «nombre apellido», por ejemplo «Ramón Martínez») y está diseñado con el objetivo de que los nombres de autores españoles puedan ser interpretados adecuadamente por las fuentes de información extranjeras, fundamentalmente anglosajonas, facilitando de ese modo los problemas en la recuperación.

En el caso de que el autor tuviera un nombre y apellido muy común, una práctica recomendada es la de unir mediante guión el primer y segundo apellido (simulando un apellido compuesto), de la forma «nombre apellido1-apellido2». Así se evita suprimir el segundo apellido para aquellos autores que no desean renunciar al mismo y además se favorece la identificación unívoca. Un ejemplo sería «Ramón Martínez-Alonso».

Algunas casuísticas que podrían darse son las siguientes:

Nombre Apellido;
Nombre Apellido1-Apellido2;
Nombre Apellido1a-Apellido1b (en el caso de apellidos compuestos);
Nombre1-Nombre2 Apellido (en el caso de nombres compuestos);
Nombre1-Nombre2 Apellido1-Apellido2;
Nombre1-Nombre2 Apellido1a-Apellido1b;

Esto es, la creación de dos secuencias de caracteres a través de la concatenación mediante un guión de todas las partes de los apellidos por un lado y todas las partes de los nombres por otra:

Nombre1-Nombre_n Apellido1-Apellido_n

De esta forma, todos los registros analizados compuestos por dos cadenas de caracteres (en cualquiera de las casuísticas anteriores) se consideran estilos de firma recomendados, dejando aparte si la elección de ese estilo es premeditada o casual por parte del autor.

A pesar de que el formato *IraLIS* da la opción de incluir una inicial entre las dos cadenas de caracteres (ej. «Ramón F. Martínez»), en este trabajo sólo se consideran como recomendados los registros que estrictamente están formados por dos ristas, considerando que el estilo de firma recomendado, en este caso, debería ser «Ramón-F Martínez» o «Ramón-F Martínez-Alonso».

Pese a las ventajas que la utilización de un estilo de firma mediante dos ristas introduce en la recuperación, hay que tener presente que su uso también puede introducir algunas limitaciones de cara a la indización en sistemas donde

la forma normalizada se construya con el apellido y sólo la inicial del nombre, como ocurre en *Web of Science (WoS)*.

Por ejemplo, «Ramón Martínez» y «Rafael Martínez», ambos nombres con estilos de firma recomendados, serían indizados como «Martínez, R». En estos casos se aconseja el uso del guión para concatenar los dos apellidos en la segunda ristra y/o los dos nombres (en caso de nombre compuesto) en la primera ristra.

En el caso de la concatenación del nombre compuesto, por ejemplo «Ramón-Fernando Martínez» (o «Ramón-F Martínez»), su indización correcta en ese sistema debería ser siempre «Martínez, R-F». Obviamente, «R-F» puede significar, además de «Ramón-Fernando», otras combinaciones como «Rafael-Francisco», pero al menos las probabilidades de que esto ocurra son menores que si se usa sólo la primera parte del nombre. Y si se concatena además la ristra de los apellidos («Ramón-F Martínez-Alonso»), las probabilidades de duplicación todavía son menores («Martínez-Alonso, R-F»).

No obstante, si la forma normalizada del nombre en el sistema usa sólo las iniciales del nombre, el margen de error nunca se podrá eliminar del todo a menos que se use algún sistema de identificación y desambiguación de nombres, aspecto en el que no se centra este trabajo.

Pese a la existencia de estudios que abordan el problema del estilo de firma en sistemas propietarios (Ruiz-Pérez, 2002; Ruiz-Pérez, 2003), se observa una carencia de análisis en entornos abiertos (no controlados ni normalizados), como son los repositorios.

En este sentido, conocer el grado de variabilidad que presentan los estilos de los distintos nombres en su forma de ser consignados en los repositorios respecto al estilo de firma recomendado puede permitir, en cierta manera, conocer el grado de concienciación previo de los autores y establecer así futuras recomendaciones que permitan minimizar este problema.

Además, al estar estos sistemas abiertos basados en el protocolo de recolección de metadatos *OAI-PMH*, un error o variante de un nombre en un repositorio se arrastrará al proveedor de servicios que recolecte información de él y, a su vez, a *Google* o cualquier otro motor de búsqueda que recolecte datos tanto de los proveedores de servicios como del repositorio directamente. Esto supone que, en ocasiones, será imposible recuperar toda la producción documental de un investigador.

Pese a la gravedad del asunto, este aspecto suele ser olvidado en las cada vez más frecuentes encuestas de satisfacción de usuarios en repositorios, donde ni los autores ni los gestores de los sistemas parecen preocupados por esta cuestión.

Con el objetivo de analizar este problema se inicia una línea de investigación centrada en el estudio del grado de utilización del estilo de firma recomendado por *IraLIS* por parte de los autores en los proyectos y repositorios que cumplen con el protocolo *OAI-PMH*, con el fin de poder conocer el grado de concienciación previo de éstos.

En 2007 se realiza un primer análisis general (Pezet, 2007), donde se estudia el porcentaje de uso del estilo de firma recomendado en los registros del repositorio *Recercat*. Los resultados muestran cómo únicamente un 25% del total de autores firma mediante un estilo recomendado.

La expansión de *IraLIS* desde las ciencias de la información y documentación a todos los campos científicos (Baiget, 2007) provoca además la conveniencia y necesidad de estudiar el estilo de firma de los autores por áreas temáticas para poder detectar las posibles diferencias o semejanzas, con el fin de poder focalizar adecuadamente posibles campañas de concienciación entre los autores.

1.2. Objetivos

El propósito de este estudio es ampliar el trabajo realizado en *Recercat* analizando el estilo de firma por área temática (Baiget, 2008) y comparar los resultados con otro importante depósito digital universitario español de similar tamaño (*E-Prints Complutense*).

De esta forma, los objetivos principales de este trabajo son los siguientes:

- Analizar el nivel de uso del estilo de firma recomendado por *IraLIS* que presentan los nombres de autor en los registros de los entornos no propietarios *Recercat* y *E-Prints Complutense*.
- Conocer la distribución de este estilo de firma por áreas del conocimiento, con el fin de realizar en el futuro campañas adecuadas de concienciación dirigidas a dichos colectivos.

2. Material y métodos

Se escogen dos repositorios para realizar el estudio: *Recercat* y *E-Prints Complutense*. Depósitos que, junto al de la *Universidad Nacional de Educación a Distancia (UNED)* y la *Universidad de Alicante (RUA)*, constituyen en la actualidad los repositorios universitarios españoles con mayor número de registros (Melero, 2008). Los datos manejados en 2006 (Pezet, 2008) daban como repositorios más numerosos otros, dado que se incluían algunos depósitos no institucionales universitarios. Por ejemplo, *Dialnet*, *Racò* o *UABPubD* se erigen en diciembre de 2007 en los más grandes en el *International Report about the situation of OAI-compliant projects*.

En el caso de *Recercat*, se trata de un repositorio colectivo de documentos digitales que incluye la literatura de investigación de las universidades y los centros de investigación de la Comunidad autónoma de Cataluña (España). Incluye artículos aún no publicados (*preprints*), comunicaciones a congresos, informes de investigación, *working papers*, proyectos de final de carrera, memorias técnicas, etc.

En el caso de *E-Prints Complutense*, se trata del repositorio institucional de la *Universidad Complutense de Madrid (UCM)*. Las tipologías documentales que incorpora son artículos, secciones de libros, ponencias, libros, informes técnicos y, fundamentalmente, tesis doctorales.

Para los autores que circulan en estos repositorios, en primer lugar, se adoptan dos tipos de clasificaciones (una general y otra exhaustiva) para las áreas de conocimiento y se establecen los mapas de correspondencia entre ambas:

- Clasificación general: idéntica a la usada en el proyecto *Intute (Ciencia y tecnología; Ciencias sociales; Arte y humanidades; y Salud y Ciencias de la vida)*.
- Clasificación específica: la considerada por la *Agencia Nacional de Evaluación y Prospectiva (ANEP)*, mostrada en la Tabla I junto a sus códigos de identificación.

TABLA I
Áreas ANEP

Área ANEP	Código ANEP
Agricultura	AGR
Biología molecular, celular y genética	BMC
Biomedicina	BMED
Biología vegetal y animal, Ecología	BVAE
Ciencia y tecnología de los alimentos	TA
Ciencias de la tierra	CT
Ciencias sociales	CS
Derecho	DER
Economía	ECO
Ciencias de la educación	EDUC
Filología Filosofía	FFI
Física y Ciencias del espacio	FI
Fisiología y Farmacología	FFA
Ganadería y pesca	GAN
Transferencia de tecnología	IND
Historia y Arte	HA
Ingeniería civil y arquitectura	ICI
Ingeniería eléctrica, electrónica y automática	IEL
Ingeniería mecánica, naval y aeronáutica	IME
Matemáticas	MTM

TABLA I (continuación)

Área ANEP	Código ANEP
Medicina clínica y Epidemiología	MCLI
Psicología	PS
Química	QMC
Tecnología electrónica y de las comunicaciones	COM
Tecnología materiales	TM
Tecnología química	TQ

Las correspondencias entre las dos clasificaciones se muestran en la Tabla II.

TABLA II*Correspondencia entre áreas generales y ANEP*

Área general	Área específica
Ciencia y tecnología	AGR, TA, INF, CT, FI, GAN, IND, ICI, IEL, IME, MTM, QMC, COM, TM, TQ
Ciencias sociales	CS, DER, ECO, EDUC, PS
Arte y humanidades	FFI, HA
Salud y Ciencias de la vida	BMC, BMED, BVBE, FFA, MCLI

La clasificación específica permite estructurar las áreas de conocimiento de la misma forma que la ANEP, agencia encargada de realizar evaluaciones científico-técnicas así como de llevar a cabo estudios y análisis prospectivos de investigación científica. Esto permite, por tanto, comparar los datos obtenidos en este trabajo con informes realizados por la agencia.

Por su parte, la clasificación general proporciona una forma simple de categorizar las distintas áreas con el fin agruparlas, de forma que se puedan analizar en conjunto y establecer patrones de normalización generales. Se opta por usar la clasificación del proyecto *Intute*, por ser éste un proyecto consolidado donde la clasificación de recursos web usada ha demostrado su eficacia.

En segundo lugar, se procede a la descarga de la base de datos de autores de estos repositorios a un fichero *Excel*:

- *Recercat*: el proceso se realiza entre el 20 y 29 de diciembre de 2007.
- *E-Prints Complutense*: el proceso se realiza durante la primera semana de marzo de 2008.

En este punto se debe reseñar que los registros de los autores proceden del repositorio y no de los trabajos originales de los autores. Este hecho es impor-

tante, pues es posible que la información que los autores (o personal de las bibliotecas) incluyan en la base de datos del repositorio cuando realizan la carga de los datos (la usada por el protocolo *OAI-PMH*) sea diferente de la que incluyen en los trabajos, debiéndose considerar este aspecto en los resultados conseguidos.

De los dos listados de autores obtenidos anteriormente se eliminan los autores extranjeros (o formas de nombre no castellanas) y los nombres institucionales.

En tercer lugar, y partiendo de la hipótesis de que la filiación de un autor a un centro (departamento o facultad) coincide con el campo de conocimiento sobre el que trabaja, se realizan las asignaciones de áreas de conocimiento a los distintos autores.

En el caso de *Recercat*, se establece un mapa de correspondencia entre sus comunidades (distintas colecciones creadas a partir del uso de *DSpace*, el *software* instalado) y las áreas *ANEP*.

Para la *UCM*, al no estar la colección agrupada en comunidades (el *software* usado es *E-Prints*, no basado en el uso de comunidades), la correspondencia se realiza a partir de los departamentos universitarios a los que pertenecen los autores, disponibles desde su sede *web*.

De esta forma, cada autor queda identificado con dos áreas de conocimiento: una general (*Intute*) y otra específica (*ANEP*).

Una vez disponible la información anterior, se asigna a cada autor un estilo de firma:

- a) Un apellido y dos nombres (Ej. Ramón Fernando Martínez).
- b) Dos apellidos (Ej. Ramón Martínez Alonso; Ramón Fernando Martínez Alonso).
- c) Dos ristas (recomendado).

Tras la asignación a cada registro de un estilo de firma, los autores se agrupan por áreas asignadas para hallar la distribución de firmas en cada una de ellas. En este caso resulta importante destacar que no se ha realizado ningún proceso para identificar y/o cuantificar las distintas variantes de nombre para un autor, sino que se han contabilizado cada una de las firmas localizadas sin importar la identificación real del autor.

Esto se debe a que el propósito es encontrar la distribución del uso de los distintos estilos de firma usados. De esta forma, un autor podrá haber firmado mediante varios estilos que cumplan el formato recomendado y otros tantos que no, contabilizando la frecuencia de aparición de cada uno de ellos (y la tipología de estilo usado) de forma independiente.

En el caso de *Recercat*, existen comunidades en las que no se puede conocer un área de conocimiento (*Universitat Politècnica de Catalunya* y *Universitat de Vic*, entre otras), bien porque no se detalla o porque el nombre de la colección es genérico (por ejemplo, «trabajos de investigación»).

En estos casos se ha consultado directamente el repositorio de la institución (*Recercat* es una colección de repositorios) para conocer la procedencia del autor (departamento o instituto de investigación).

Es importante hacer notar que la forma del nombre a considerar es siempre la que figura en la base de datos de autores de *Recercat*, pues en algunos casos difiere de la forma encontrada en otros repositorios (nombre propio en catalán y castellano, por ejemplo).

En aquellos casos en los que el procedimiento anterior no es posible (fundamentalmente para autores externos que han participado en algún trabajo o investigación pero que no pertenecen contractualmente a la universidad), se asigna un área al autor en función de los descriptores temáticos que aparecen en los trabajos donde aparece como autor.

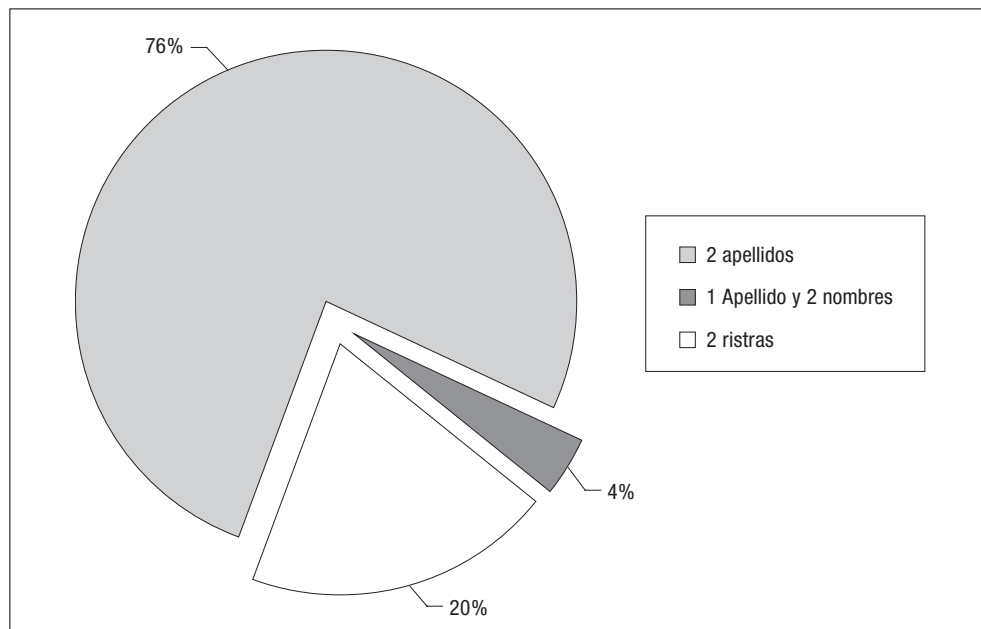
3. Resultados

3.1. *Recercat*

Se han analizado un total de 3.653 registros, en los que se han detectado 3.534 autores distintos. A continuación se muestra la distribución total de sus estilos de firma en *Recercat* (Figura 1).

FIGURA 1

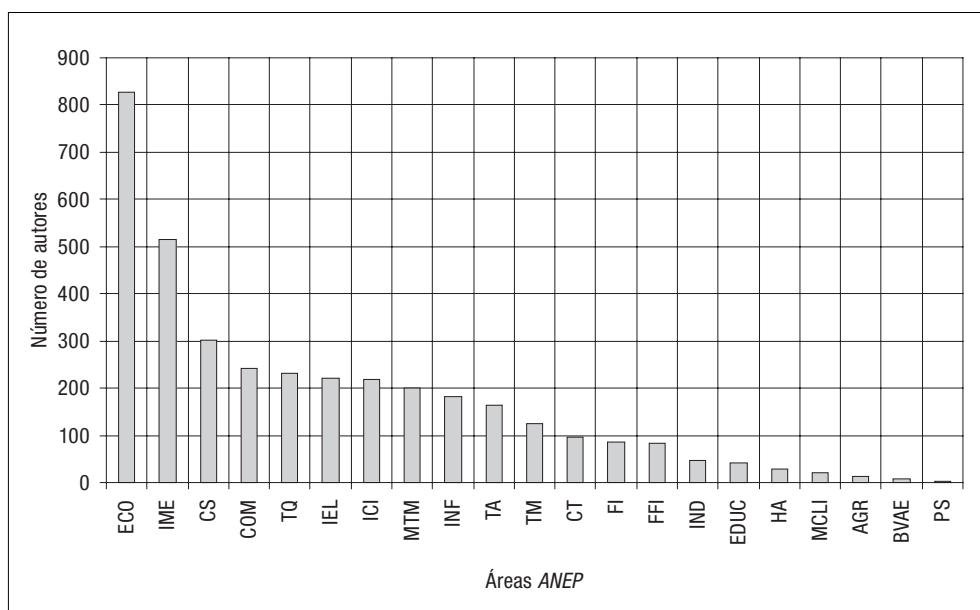
Distribución total del estilo de firma en Recercat



Se observa cómo el estilo recomendado (dos ristas) ha disminuido al 20% respecto al obtenido en el primer estudio realizado (25%), mientras que el estilo formado por nombre y dos apellidos constituye el 76% del total de firmas, constituyéndose en el estilo predominante.

En la Figura 2 se muestra la distribución de áreas ANEP en *Recercat*, con el fin de poder categorizar temáticamente el repositorio en función de los autores españoles:

FIGURA 2
Distribución de áreas ANEP en Recercat



Se observa una distribución desigual, donde el área de Economía aparece en el primer puesto de forma muy destacada. La Ingeniería mecánica, naval y aeronáutica y las Ciencias sociales le siguen en número de registros.

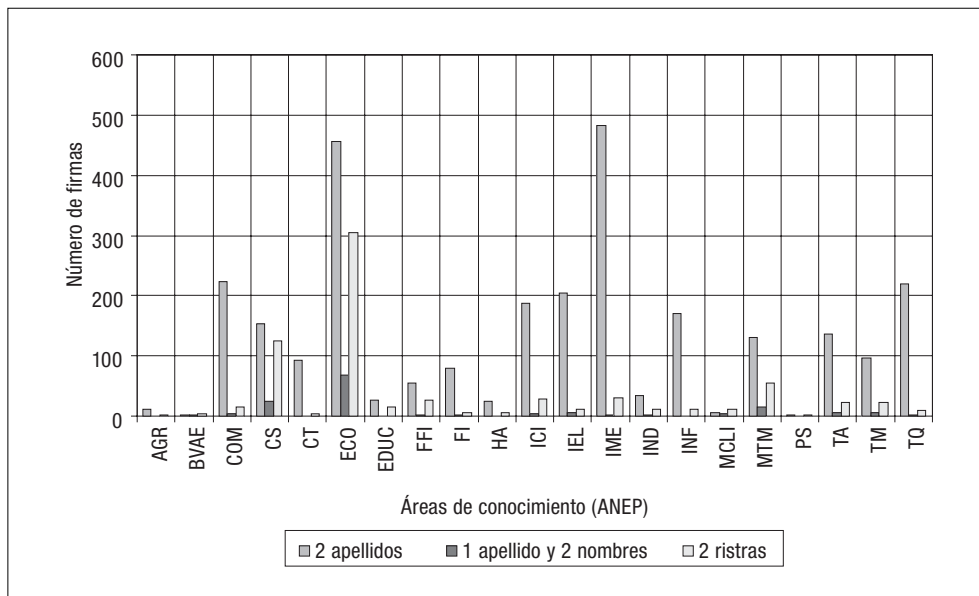
También resulta de interés observar un pequeño agrupamiento de áreas en la zona alta de la curva de distribución, concretamente del 4º puesto (243 registros) al 8º (200 registros).

Por otra parte, en la Figura 3 se analiza el estilo de firma para cada una de las áreas ANEP.

Las áreas con una mayor cantidad de firmas con el estilo recomendado corresponden a Economía y Ciencias Sociales (1ª y 3ª en número total de registros), mientras que áreas más tecnológicas presentan grados muy bajos en el uso de este estilo de firma.

FIGURA 3

Distribución del estilo de firma por área ANEP en Recercat



No obstante, estos datos están influenciados por la gran cantidad de registros en estas áreas pues, pese a todo, el estilo de firma dominante en éstas es mediante dos apellidos.

También se debe destacar la distribución obtenida para el área *IME* (Ingeniería mecánica, naval y aeronáutica), segunda en número de registros (514), pero con una adopción prácticamente inexistente del estilo recomendado.

En la Figura 4 se muestra la distribución porcentual del estilo de firma recomendado, tomando en consideración sólo a las áreas temáticas con un mínimo de 100 autores:

Se obtiene igualmente una distribución desigual de áreas. Las Ciencias Sociales y la economía son las áreas, junto a las matemáticas, con unos porcentajes de adopción del estilo recomendado más elevados. Se destaca cómo en el área de la Economía, algo más del 40% de las firmas de los autores de esta categoría firman mediante una forma recomendada (generalmente mediante nombre y apellido).

Es especialmente significativo el hecho de que únicamente en tres áreas (Ciencias Sociales, Economía y Matemáticas) se supere el 25% de firmas con un estilo normalizado, existiendo un salto cuantitativo negativo importante a partir del 4º puesto (Tecnología de los materiales).

En la Tabla III se muestra la distribución según la clasificación temática general.

FIGURA 4

Distribución del estilo de firma recomendado (ANEP)

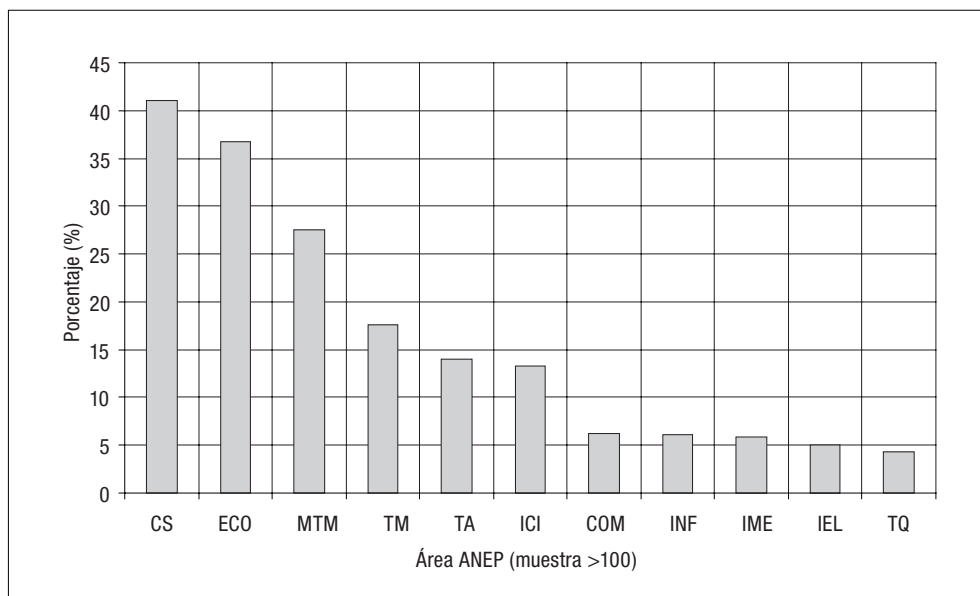


TABLA III

Distribución de áreas generales en Recercat

Área de conocimiento	2 apellidos		1 apellido y 2 nombres		2 ristas		Total
	Nº	%	Nº	%	Nº	%	Nº
Artes y humanidades	78	69,64	2	1,79	32	28,57	112
Ciencias sociales	638	54,30	93	7,91	444	37,79	1.175
Ciencia y tecnología	2.067	88,37	44	1,88	228	9,75	2.339
Salud y ciencias de la vida	7	25,93	5	18,52	15	55,56	27
Total	2.790	76,38	144	3,94	719	19,68	3.653

El área relacionada con Salud, Biología y Medicina es prácticamente inexistente en *Recercat*, pues sólo se computan 27 registros, por ello sus resultados no son significativos.

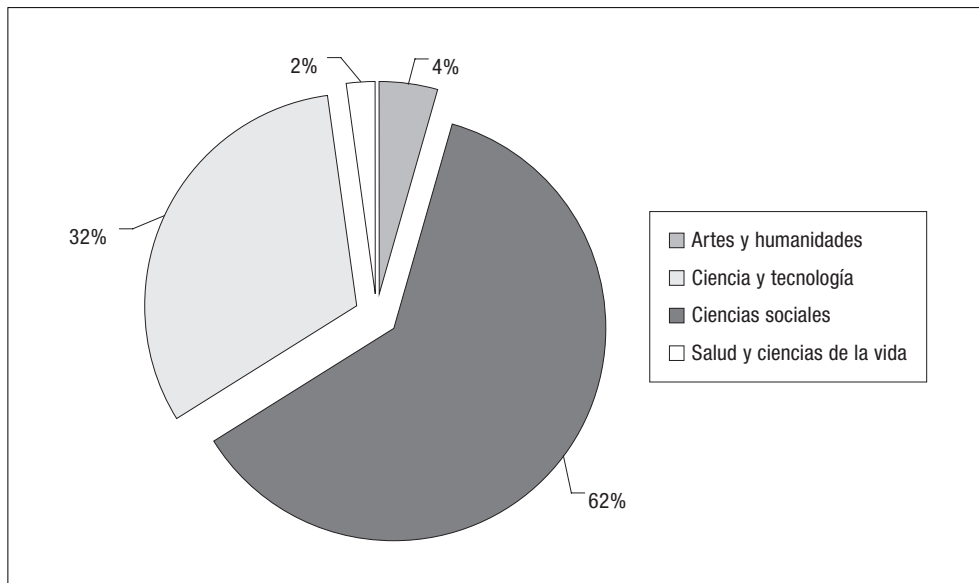
Las áreas Tecnológicas (2.339 registros) son las más presentes, junto a las Ciencias Sociales (1.175 registros). En cuanto al estilo de firma recomendado, las Ciencias Sociales presenta el porcentaje más elevado (37,79%) junto a las Artes

y Humanidades (28,57%), que supera ampliamente al área de Ciencia y Tecnología, con apenas el 9,75%.

En la Figura 5 se muestra la distribución para el estilo recomendado, de cada una de las áreas.

FIGURA 5

Distribución en Recercat del estilo de firma recomendado (áreas generales)



Como se observa en la Figura 5, el 62% de las firmas con estilo recomendado de la muestra corresponden a las Ciencias Sociales, mientras que las áreas con carga tecnológica aparecen con un 32% de las firmas con estilo recomendado. No obstante, al tener en consideración el número de registros en cada área (tabla III), el área tecnológica es la que menos adopta el estilo recomendado, como se ha indicado anteriormente.

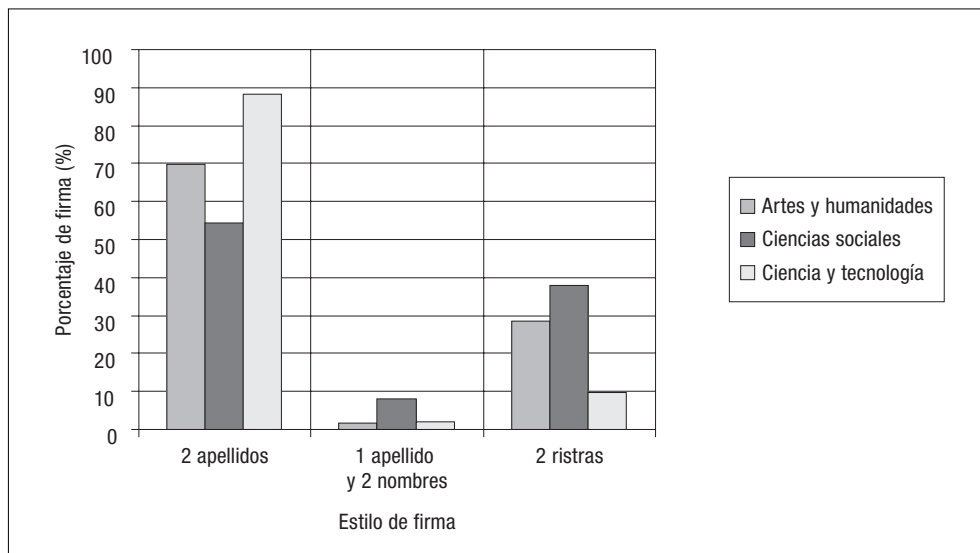
En la Figura 6 se presenta la distribución para todos los estilos de firma.

Se observa claramente cómo la forma más usual de firma es mediante dos apellidos, usada por el 88,37% de los autores del área de Ciencia y tecnología y el 69,64% de Artes y humanidades.

No obstante, es importante destacar que el hecho de que los autores hayan escogido un estilo de firma determinado no significa necesariamente un mayor o menor grado de concienciación en la importancia de la firma usada, pues la elección de un estilo recomendado puede deberse a la casualidad u otros factores. Por esta razón es tan importante usar un estilo recomendado como mantener éste durante toda la vida profesional. Es decir, firmar mediante muchos estilos

FIGURA 6

Distribución total del porcentaje de los estilos de firma en Recercat (áreas generales)



de firma recomendados es igual o peor para la recuperación que firmar únicamente con un estilo no recomendado.

Por ese motivo, el número de autores que hayan usado más de un estilo de firma puede ser un importante indicador de la mayor o menor concienciación de éstos. En ese sentido, de los 3.653 registros analizados en *Recercat* se han detectado 98 autores con al menos dos variantes en sus estilos de firma, lo que representa un 2,68% del total, cifra nada desdeñable.

Aparte del estilo de firma usado, se ha detectado una cantidad importante de variantes debidas al uso indebido del nombre propio, por ejemplo:

- Jarauta Bragulat, Eusebi;
- Jarauta Bragulat, Eusebio;

En el caso anterior, ninguna de las dos formas sería recomendada, según los parámetros seguidos en este trabajo, pero además se añade el problema del uso del nombre propio en castellano y catalán. No obstante, este trabajo se centra en la estructura de las firmas y no en su contenido, tema que merece un estudio aparte.

3.2. E-Prints Complutense

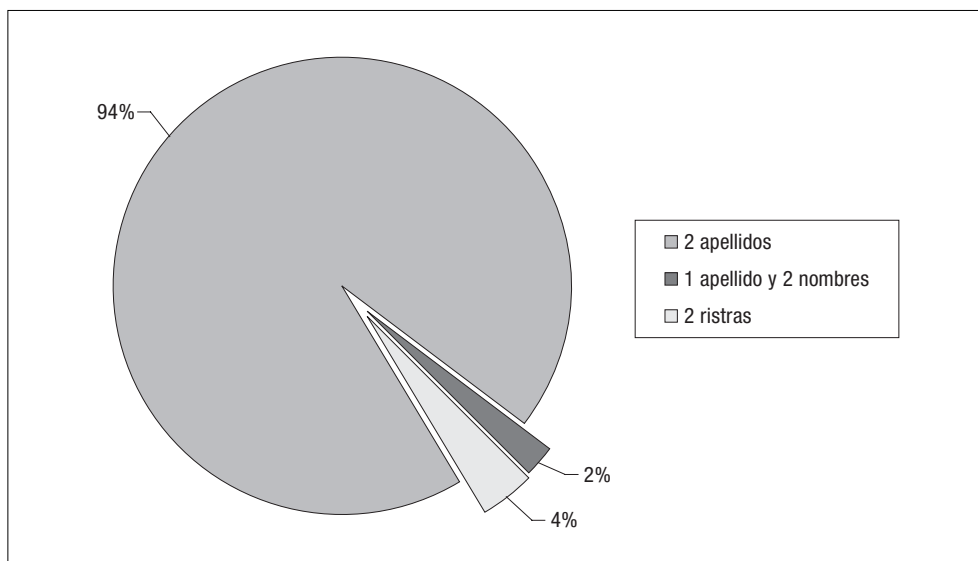
Se dispone de una muestra de 4.743 registros de nombre distintos, donde se han identificado 4.663 autores únicos. En este caso se han detectado 60 autores

con distintos estilos de firma, lo que constituye el 1,25% del total de registros, valor inferior al detectado en *Recercat*. El hecho de no existir problemas con el uso indistinto de dos idiomas (como sucede en *Recercat* con el castellano y catalán) puede ser una de las causas que expliquen este fenómeno.

La correspondiente distribución total del estilo de firma se presenta en la Figura 7.

FIGURA 7

Distribución total del estilo de firma en E-Prints Complutense



Se muestra de forma clara cómo el estilo de firma predominante es mediante dos apellidos. Los otros estilos de firma, incluido el recomendado, son tan sólo esporádicos.

La Figura 8 muestra cómo las áreas temáticas difieren de las obtenidas en *Recercat*. La distribución no es tan desigual y además existe una presencia importante de áreas relacionadas con la Medicina y Biología.

La distribución del estilo de firma recomendado por áreas *ANEP* se detalla en la Figura 9.

La Figura 9 muestra cómo los niveles de adopción del estilo de firma recomendado son escasos en todas las áreas. Destaca el área *CT* (Ciencias de la tierra) en el primer puesto (sin apenas llegar al 20% de registros de autor con el estilo que se recomienda).

La Economía y el Derecho aparecen con índices reseñables, mientras que *CS* (Ciencias sociales), presenta unos niveles casi mínimos.

En cuanto a las áreas generales, la distribución se detalla en la Tabla IV.

FIGURA 8

Distribución áreas ANEP en E-Prints Complutense

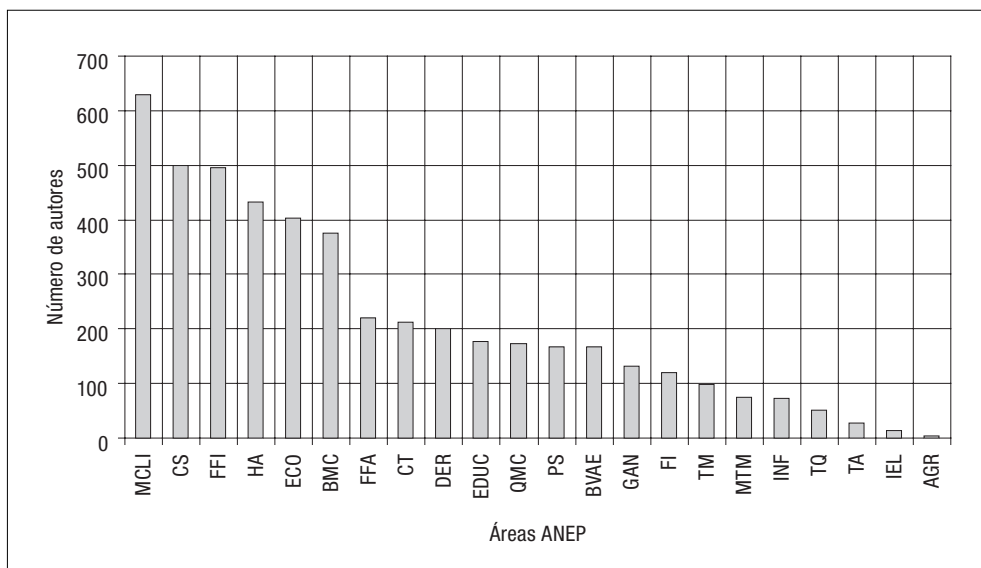


FIGURA 9

Distribución del estilo de firma recomendado en E-Prints Complutense

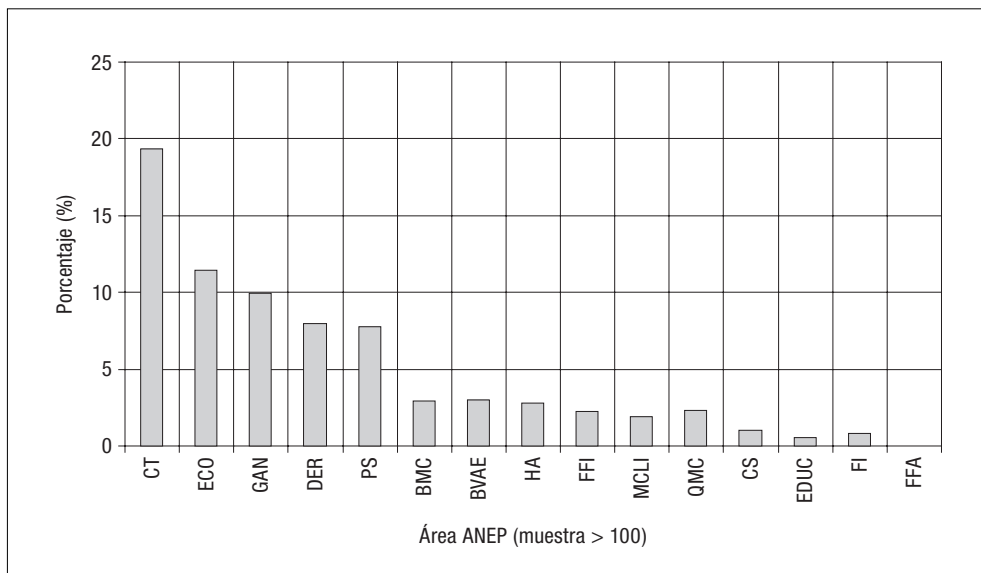


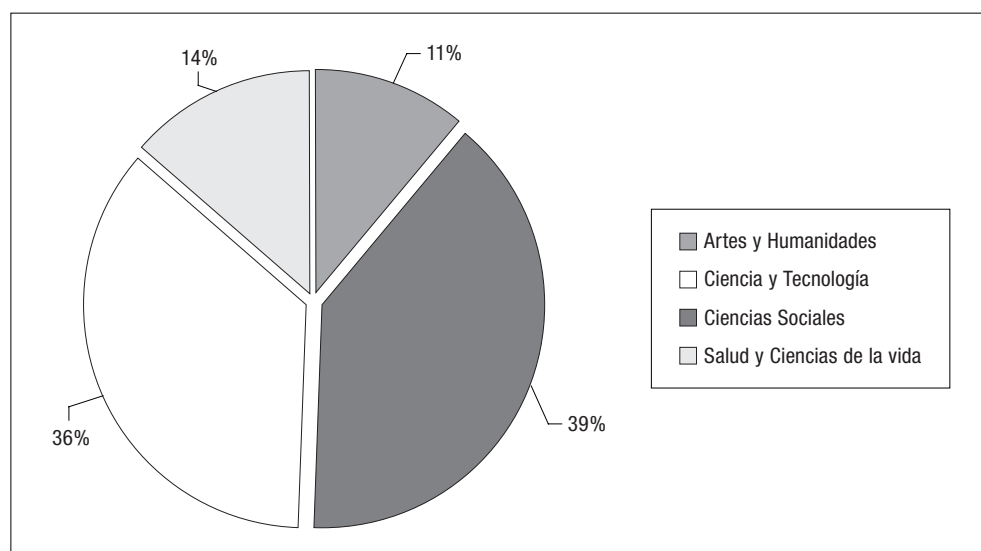
TABLA IV
Distribución de áreas generales en E-Prints Complutense

Área de conocimiento	2 apellidos		1 apellido y 2 nombres		2 ristas		Total
	Nº	%	Nº	%	Nº	%	Nº
Artes y humanidades	898	96,77	7	0,75	23	2,48	928
Ciencias sociales	1.331	91,92	36	2,49	81	5,59	1.448
Ciencia y tecnología	884	90,57	18	1,84	74	7,58	976
Salud y ciencias de la vida	1.351	97,12	12	0,86	28	2,01	1.391
TOTAL	4.464	94,12	73	1,54	206	4,34	4.743

Las áreas más presentes en el repositorio son las relacionadas con las Ciencias Sociales (1.448 registros) y las relacionadas con la Salud, Medicina y Biología (1.391 registros), área apenas presente en *Recercat*.

Se observa cómo al englobar la Economía y el Derecho dentro de las áreas sociales, los niveles de adopción del estilo recomendado suben ligeramente. Pese a ello, el porcentaje de registros con estilo recomendado para esta área es escaso (5,59%). En este caso, y al contrario de lo que sucedía en *Recercat*, el área

FIGURA 10
Distribución en E-Prints Complutense del estilo de firma normalizado (áreas generales)



con un mayor porcentaje de registros con estilo recomendado es la Ciencia y Tecnología, con un discreto 7,59%.

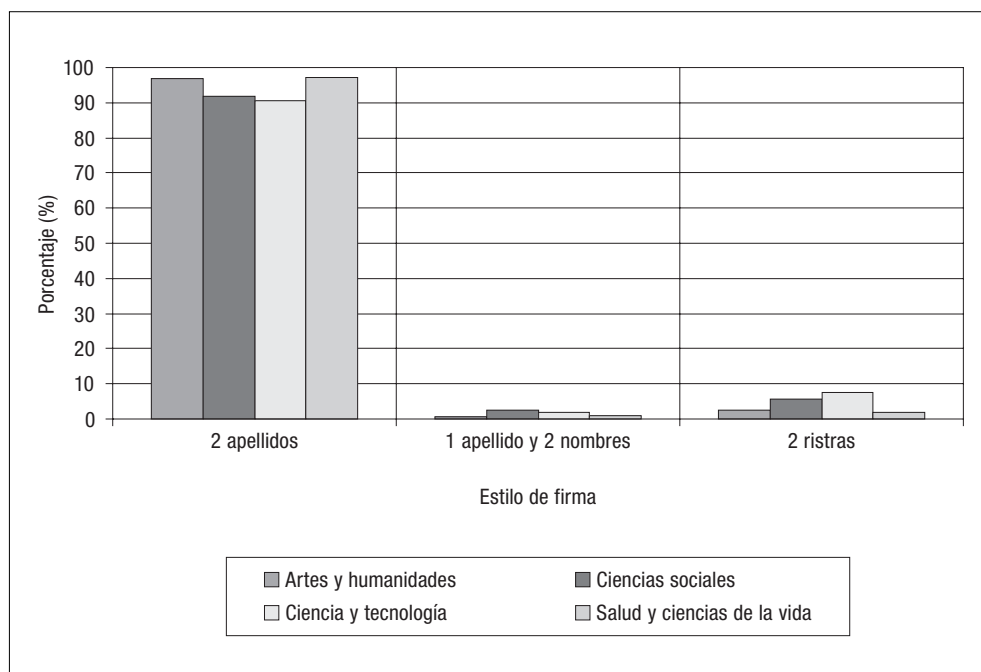
La distribución total del estilo de firma recomendado en base a estas áreas generales se muestra en la Figura 10.

La Figura 10 muestra que la distribución es más repartida que la obtenida en *Recercat*. El área tecnológica aparece aquí como la segunda con más registros con estilos recomendados (36%), mientras que las Ciencias Sociales aparecen en primera posición con un 39%, justo al contrario que ocurría en *Recercat*. No obstante, al considerar los porcentajes en función del tamaño de cada área, tal como indica la tabla IV, las posiciones se invierten y es el área tecnológica el que aparece con un mayor número de registros con estilo recomendado.

Finalmente, la distribución total de estilos de firma por área general aparece en la Figura 11.

FIGURA 11

Distribución total del porcentaje de los estilos de firma en E-Prints Complutense (áreas generales)



Se comprueba, una vez más, cómo el estilo de firma mediante dos apellidos es abrumadoramente superior al resto en todas las áreas estudiadas.

3.3. Análisis combinado

La muestra conjunta de registros de autores de los repositorios *Recercat* y *E-Prints Complutense* se muestra en la tabla V.

TABLA V
Distribución total de autores por áreas generales

Área de conocimiento	2 apellidos		1 apellido y 2 nombres		2 ristas		Total
	Nº	%	Nº	%	Nº	%	Nº
Artes y humanidades	976	93,85	9	0,87	55	5,29	1.040
Ciencias sociales	1.969	75,07	129	4,92	525	20,02	2.623
Ciencia y tecnología	2.951	89,02	62	1,87	302	9,11	3.315
Salud y ciencias de la vida	1.358	95,77	17	1,20	43	3,03	1.418
TOTAL	7.254	86,40	217	2,58	925	11,02	8.396

El propósito de la unión de los resultados es el de obtener una muestra de registros de nombres diferentes de autor suficiente (8.396) como para poder tomar en cuenta los resultados por área temática.

En lo que se refiere a las áreas temáticas, los resultados de la tabla V muestran a las Ciencias Sociales como el campo con mayor porcentaje de registros con un estilo recomendado (20,02%), seguido de los campos Tecnológicos (9,11%). Las Ciencias de la Salud ocupan el último puesto (3,03%), aunque el número de registros de los que se dispone es mucho menor (1.418 registros, frente a 3.315 de las Ciencias Sociales).

La distribución cuantitativa final de estilos de firma recomendados aparece reflejada en la Figura 12.

El modo de firma mediante dos apellidos, como se observa, es el más usual para todas las áreas. De hecho, en las Ciencias Sociales, área donde esta modalidad de firma obtiene su menor porcentaje, su valor se sitúa en torno al 80%.

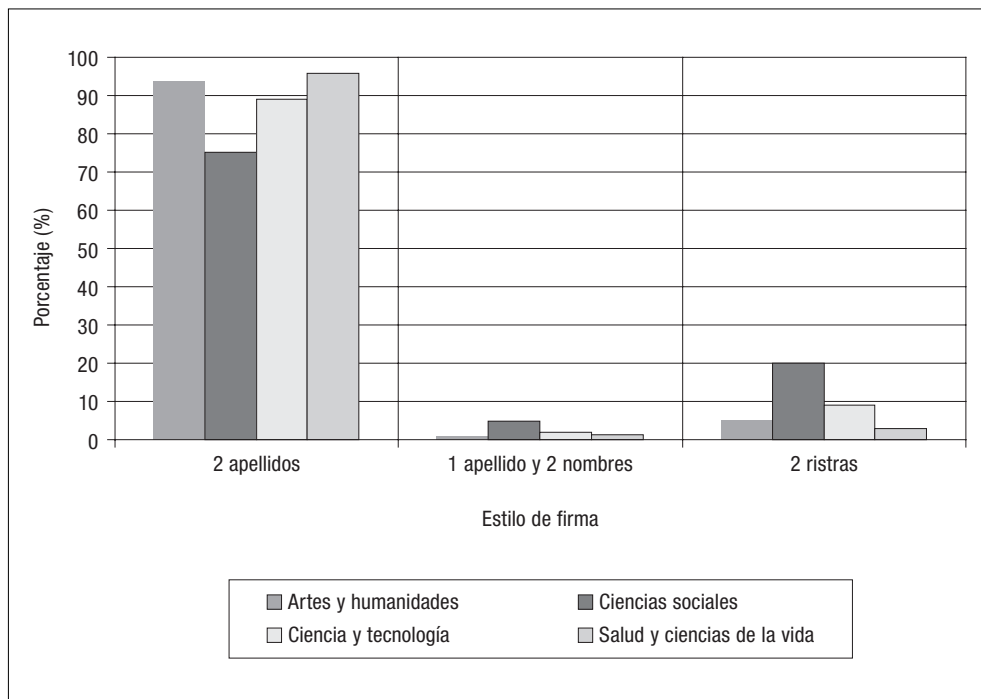
Pese a ello, es reseñable que las Ciencias Sociales sean el área con menor porcentaje de firma mediante dos apellidos y, a su vez, posean el mayor porcentaje en el uso de dos ristas. Esto muestra que en el ámbito estudiado es el campo donde la adopción del estilo de firma recomendado es mayor, mientras que el área de Salud y Ciencias de la Vida es el que presenta un menor uso del estilo formado por dos ristas.

4. Conclusiones

Los niveles de utilización del estilo de firma recomendado, tanto en *Recercat* como en *E-Prints Complutense*, son muy bajos, especialmente en este último,

FIGURA 12

Distribución total del porcentaje de los estilos de firma en Recercat y E-Prints Complutense (áreas generales)



donde únicamente es usado en el 4% de los registros. En *Recercat* se alcanza el 20%. Dato significativo si se tiene en cuenta, además, que en *Recercat* se dispone de cerca de mil registros menos.

No obstante, el formato de dos ristas es relativamente novedoso, por lo que esto puede contribuir a explicar los resultados obtenidos. De este modo, para los investigadores que no conocen este formato, puede que su tendencia natural sea la de incluir los dos apellidos. De hecho, la mayoría de los casos de dos ristas se corresponden realmente con casos de investigadores que firman con un sólo nombre y un sólo apellido (los casos de apellidos o nombres concatenados son menores). Por ello, una primera recomendación sería la de difundir a los autores el formato basado en dos ristas y sus ventajas en la recuperación.

Por otra parte, el área general de las Ciencias Sociales presenta un nivel de utilización del estilo de firma recomendado mayor del esperado. En *Recercat* se observa claramente con los resultados en las áreas específicas *ECO* y *CS*. En el caso de *E-Prints Complutense*, pese a que *CS* se encuentra muy por debajo, destaca la presencia del área *DER* (Derecho), que se encuadra en el área general de Ciencias Sociales.

La distribución del uso del estilo de firma recomendado en función de las áreas ANEP presenta una distribución desigual en *Recercat*, mientras que en *E-Prints Complutense* no ocurre así, donde se observa una distribución más equitativa.

En *Recercat* el 62% de las firmas con un estilo recomendado pertenecen al área general de las Ciencias Sociales. Este valor se reduce al 39% en el caso de *E-Prints Complutense*. No obstante se debe tener en cuenta las diferencias temáticas de ambos repositorios. Así, mientras en *Recercat* se obtienen 1.175 registros en ciencias sociales, en *E-Prints Complutense* se obtienen 1.448 (es decir, en *Recercat* se dispone de menos registros, pero con un estilo recomendado). El caso opuesto ocurre con el área general de Ciencia y Tecnología (32% de los registros con estilo recomendado en *Recercat* y 36% en *E-Prints Complutense*). Finalmente, el análisis combinado muestra a las Ciencias Sociales como el área con un mayor porcentaje de registros con un estilo basado en dos ristas (20,03% del total de los registros en esta área).

Un análisis combinado confirma que el estilo de firma más usado es mediante nombre y dos apellidos, estilo de firma no recomendado y que dificulta su visibilidad y recuperación en bases de datos internacionales.

Una segunda recomendación a los autores españoles, tras la adopción de un estilo basado en dos ristas, sería la de unir sus dos apellidos mediante un guión, de forma que sin renunciar a sus dos apellidos, puedan formar un estilo de firma recomendado mediante dos cadenas de caracteres (hecho que favorecería su identificación y recuperación) y que este estilo lo mantuviesen siempre durante toda su vida profesional.

Las diferencias de temáticas y cantidad de registros no permite la extrapolación de los resultados de un repositorio a otros, por ello deben ser considerados en el contexto de la comunidad a la que sirven. En ese aspecto, en *Recercat*, las áreas Tecnológicas presentan niveles muy bajos de implantación de estilos de firma recomendados, mientras que en *E-Prints Complutense*, las áreas de Salud y Ciencias de la Vida y Arte y Humanidades, precisan de campañas de concienciación evidentes.

Se debe reseñar que en ambos repositorios la carga de registros se realiza tanto por parte de la biblioteca como por parte de los autores. Además, pueden existir diferencias entre el estilo de firma usado en el documento original y el usado en los metadatos del repositorio (los recogidos en este trabajo), sea quien sea el responsable de su carga en el sistema.

Esto es de especial importancia, pues tal vez los autores sí están concienciados (al menos cuando firman sus trabajos originales) pero no tanto cuando archivan sus documentos en los repositorios.

Por tanto, una tercera recomendación para los autores sería la de elegir un estilo de firma recomendado y usarlo siempre, tanto para firmar como para cumplimentar los correspondientes metadatos.

Por ello, un futuro estudio de las posibles diferencias en el uso de estilos de firma en función de quién alimenta las bases de datos (autor o personal biblio-

tecario) así como de las diferencias entre los estilos usados en los documentos y en los metadatos de los repositorios es necesario para analizar más profundamente el grado de concienciación tanto de los autores como del personal encargado de introducir los registros.

5. Agradecimientos

Agradecemos enormemente tanto a Sandra Reoyo-Tudó (*Consorti de Biblioteques Universitàries de Catalunya, CBU*) como a Manuela Palafox (*Biblioteca de la Universidad Complutense de Madrid*) su cooperación en la obtención de material en las bases de datos de autores de los repositorios estudiados.

Asimismo, agradecemos a los revisores todas sus sugerencias y aportaciones, que han contribuido a mejorar el trabajo de forma muy significativa.

6. Bibliografía

- Agencia nacional de evaluación y prospectiva (ANEP): Disponible en: <http://www.micinn.es/ciencia/jsp/plantilla.jsp?area=anep&id=32> [Consulta 10/09/2008].
- Baiget, T.; Ferrer, A.; Ontalba-Ruipérez, J-A.; Orduña-Malea, E.; Peset, F.; Rodríguez-Gairín, J. M., y Subirats, I. (2008): Normalización de la firma de autores por disciplina en repositorios catalanes: una aproximación estadística. En: *11 Jornades catalanes d'Informació i Documentació, Barcelona 22 y 23 de mayo*, 637-640. Disponible en: http://www.cobdc.org/jornades/11JCD/actes11jcid/posters/pag_636.pdf [Consulta 10/09/2008].
- Baiget, T.; Rodríguez-Gairín, J. M.; Peset, F.; Subirats, I., y Ferrer, A. (2007): La normalización de la información: la aportación de IRALIS -International Registry for Authors in Library and Information Science. *El profesional de la información*, 16(6), 636-643.
- Beaubien Bennet, D. (2006): Name Authority Challenges for indexing and Abstracting Databases. *Evidence based library and information practice*, 1(1), 37-57. Disponible en: <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7/86> [Consulta 10/09/2008].
- Costas, R., y Bordons, M. (2007): Algoritmos para solventar la falta de normalización de nombres de autor en los estudios bibliométricos. *Investigación bibliotecológica*, 21(42), 13-32.
- Danskin, A. (1996): International standards in authority data control: costs and benefits. En: *62nd IFLA General Conference - Conference Proceedings - August 25-31*. Paris: INIST. Disponible en: <http://www.ifla.org/IV/ifla62/62-dana.htm> [Consulta 10/09/2008].
- Danskin, A.; Dixon, A.; Docherty, M.; Hill, A., y Moore, R. (2007): A review of the current landscape in relation to a proposed Name Authority Service for UK repositories of research outputs. Disponible en: http://names.mimas.ac.uk/documents/Names_landscape_report_1Oct2007.pdf [Consulta 10/09/2008].
- DIALNET. Disponible en: <http://dialnet.unirioja.es/> [Consulta 10/09/2008].

- Dipòsit Digital de Documents UAB (UABPubD): Disponible en: <http://ddd.uab.es/collection/uabpubd?ln=es> [Consulta 10/09/2008].
- Estrategia nacional de ciencia y tecnología: comisión interministerial de ciencia y tecnología. (2007): Madrid: FECYT.
- FECYT. (2007): Recomendaciones para la correcta identificación de las publicaciones científicas. FECYT: Madrid. Disponible en: http://www.accesowok.fecyt.es/recomendaciones_publicaciones.html [Consulta 10/09/2008].
- Ferrer, A.; Peset, F., y Baiget, T. (2008): Visibilidad de la investigación española: el proyecto IraLIS de normalización de nombres de autores. En: *V Foro sobre la evaluación de la calidad de la educación superior y la investigación. San Sebastian, 2-5 septiembre, 2008*.
- Gálvez, C., y Moya-Anegón, F. (2007): Approximate personal name-matching through finite-state graphs. *Journal of the American Society for Information Science and Technology*, 58 (13), 1960-1976.
- International Council on Archives (CIA) (2004): ISAAR (CPF). International Standard Archival Authority Record for Corporate Bodies, Persons and Families. (2nd edition). CIA: Paris. Disponible en: <http://www.ica.org/sites/default/files/ISAAR2EN.pdf> [Consulta 10/09/2008].
- INTUTE: best of the web. Disponible en: <http://www.intute.ac.uk/> [Consulta 10/09/2008].
- IraLIS: International Registry for Authors: Links to Identify Scientists. Disponible en: <http://www.iralis.org> [Consulta 10/09/2008].
- Joint Information System Committee (JISC) (2008): Names: pilot national name and factual authority service. Disponible en: http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/shared_services/project_names.aspx [Consulta 10/09/2008].
- MADS: Metadata Authority Description Schema. Disponible en: <http://www.loc.gov/standards/mads> [Consulta 10/09/2008].
- Melero, R. (2008): El paisaje de los repositorios open access en España. *Bid*, Num. 20. Disponible en: http://www2.ub.edu/bid/consulta_articulos.php?fichero=20meler4.htm [Consulta 10/09/2008].
- Peset, F.; Ferrer, A.; Subirats, I., y Muñoz, A. (2007a): Use of OAI protocol and its impact in digital libraries: a case study in Spain, Portugal and Latin America. En: *Prasad, ARD and Madalli, Devika P., Eds. Proceedings International Conference on Semantic Web and Digital Libraries*. Disponible en: <http://eprints.rclis.org/archive/00010931/> [Consulta 10/09/2008].
- Peset, F.; Ferrer, A.; Baiget, T., y Rodríguez-Gairín, J. M. (2007b): DSpace and the standardization of the information: names of Spanish authors. En: *DSpace User Group Meeting, FAO, Rome, Italy, October 17th-19th*.
- Peset, F. y Ferrer, A. (2008): Implantación de la Open Archives Initiative en España. *Information Research*, 13(4). Disponible en: <http://informationr.net/ir/13-4/paper385.html> [Consulta 10/09/2008].
- Pinar Gámez, J. L. (2007): Identificación de autores en bases de datos bibliográficas. Unpublish final project dissertation. UPV: CSIC, Valencia. Disponible en: <http://www.dsic.upv.es/~proso/resources/PinarPFC.pdf> [Consulta 10/09/2008].
- RACÒ: Revistes Catalanes en Accés Obert. Disponible en: <http://www.raco.cat/> [Consulta 10/09/2008].

- RECERCAT: Disponible en: <http://www.recercat.net/> [Consulta 10/09/2008].
- RUA: Repositorio Institucional de la Universidad de Alicante. Disponible en: <http://rua.ua.es/> [Consulta 10/09/2008].
- Ruiz-Pérez, R.; Delgado López-Cózar, E. y Jiménez-Contreras, E. (2002): Spanish personal names variations in national and international biomedical databases: implications for information retrieval and bibliometric studies. *Journal of the Medical Library Association*, num. 90, 411-430. Disponible en: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=128958> [Consulta 10/09/2008].
- Ruiz-Pérez, R.; Delgado López-Cózar, E. y Jiménez-Contreras, E. (2003): Spanish name indexing errors in international databases. *Lancet*, num. 361, 1656-1657.
- Snyman, M. M. M. (2000): The Standardization of Names to Improve Quality and Cooperation in the Development of Bibliographic Databases. *Libri*, 50, 269-279.
- Snyman, M. M. M., y Rensburg, M. J. VAN. (1999): Reengineering name authority control. *The electronic library*, 17(5), 313-322.
- Snyman, M. M. M., y Rensburg, M. J. VAN. (2000): NACO versus ISAN: prospects for name authority control. *The electronic library*, 18(1), 63-68.
- Spanje, Daniël van. (2007): Digital Author Identification (DAI). *UKSG 17-18* april. Disponible en: http://dai-uitrol.ub.rug.nl/logboek/FILES/61/DAI_UKSG_20070417_final.doc [Consulta 10/09/2008].
- Torvik, V. I.; Weeber, M.; Swanson, D. R., y Smalheiser, N. R. (2005): A probabilistic similarity metric for Medline records: a model for author name disambiguation. *Journal of the American Society for information science and technology*, 56(2), pp. 140-158.
- UCM Complutense. Disponible en: <http://eprints.ucm.es/> [Consulta 14/03/2009].
- Wooding, S.; Wilcox-Jay, K.; Lewison, G., y Grant, J. (2006): Co-author inclusion: A novel recursive algorithmic method for dealing with homonyms in bibliometric analysis. *Scientometrics*, 66(1), 11-21.
- XEP-0054: VCard-temp. Disponible en: <http://www.xmpp.org/extensions/xep-0054.html> [Consulta 10/09/2008].