

Un análisis de los principales sistemas de identificación y perfil para el personal investigador

Nieves Lorenzo-Escolar y Fátima Pastor-Ruiz

Universidad del País Vasco. UPV/EHU

Se analizan las diferentes iniciativas que han surgido para identificar de forma exacta al investigador y su currículum, puestas en marcha por entidades gubernamentales, bibliotecarios y documentalistas, productores y distribuidores de bases de datos y consorcios con participación de todas las entidades involucradas en la producción científica. Del examen de los más representativos, se proponen consideraciones de utilidad para el desarrollo y mejora de tales sistemas, así como orientaciones clarificadoras para el personal investigador, ya que su implantación es un problema complejo que implica un gran número de interesados con puntos de vista opuestos sobre algunas de las cuestiones que deben abordarse, tales como el acceso a la información, la privacidad, los modelos de negocio y de colaboración, o los derechos de propiedad intelectual. Se concluye la necesidad de adoptar sistemas que integren identificador y perfil de autor, que permitan la interoperatividad y recojan soluciones ya refutadas en el mundo bibliotecario, así como promover su aceptación y uso generalizado por parte de la comunidad científica.

Palabras clave: Identificador de autor, perfil de autor, sistemas de información científica, registro de investigadores, Control Bibliográfico Universal.

Analysis of the main systems of identification and profiles of research staff. This paper analyses different initiatives that have come to identify the researcher and his/her curriculum; these systems have been implemented by government agencies, librarians and archivists, producers and distributors of databases and consortia with the participation of all entities involved in the scientific production. From the review of the most representative initiatives, several considerations are proposed for the development and improvement of such systems; in addition, some clarifying guidelines for researchers are presented, taking into account that their implementation is a complex problem involving a large number of stakeholders with opposing views on some of the issues to be addressed such as access to information, privacy issues, business models and collaboration, or intellectual property rights. It can be concluded that there is a need for systems that integrate author identifier and profile, allowing interoperability solutions that have been tested in the library sector, and promoting their acceptance and widespread use by the scientific community.

Keywords: Author Identifier, author profile, current research information systems, research record, Universal Bibliographic Control.

La identificación universal del nombre de los autores no es un tema nuevo. Las bi-

bliotecas tienen una larga historia en el uso de los registros de autoridad para identificar sistemáticamente personas y organizaciones, y vincular las diferentes variantes de los nombres dentro de sus catálogos. De hecho, desde hace ya tiempo varios organismos e instituciones internacionales vienen desarrollando iniciativas alrededor del control

Fecha de recepción: 30-04-2012 • Fecha de aceptación: 23-05-2012

Correspondencia: Fátima Pastor-Ruiz.

Universidad del País Vasco. Unidad de Cienciometría de los Servicios Generales de Investigación (SGIKER). Vicerrectorado de Investigación. Edificio Rectorado, Barrio Sarriena, s/n Leioa (Vizcaya). C.P. 48940. Correo-e: fatimaana.pastor@ehu.es

de autoridades, especialmente en el marco del hoy remozado Programa de Control Bibliográfico Universal. La mayoría de los sistemas bibliotecarios han creado sus ficheros de autoridades propios o en red, aunque más para normalizar los puntos de acceso que para identificar claramente las distintas autoridades.

Con el crecimiento exponencial de la producción científica y su cada vez mayor presencia en la red, ya sea en forma de bases de datos o sin estructurar, el problema de identificar correctamente y sin ambigüedades al autor de un documento se ha agravado aún más. Si ya en 1974 Eugene Garfield, creador del *Institute for Scientific Information* (ISI) hoy más conocido por su plataforma *Web of Knowledge*, se refería a la dificultad que supone la identificación de autores (Garfield, 1974), en la actualidad son diversas las voces que reclaman desde todos los sectores la adopción de una solución global similar a la elegida para identificar el documento en soporte digital con el DOI (*Digital Object Identifier*), al objeto de superar los inconvenientes derivados de la existencia de autores diferentes con el mismo nombre, firmas y afiliaciones cambiantes y procesos de vaciado e indización automáticos, incapaces de solventar correctamente la ingente casuística relacionada con la autoría (Qiu, 2008), tales como número, orden y frecuencia de nombres y apellidos en otras culturas, grafías no latinas, y demás (FECYT, 2007; Universidad del País Vasco/EHU, 2011).

Las primeras iniciativas en este sentido, se remontan a la década de 1940, cuando la Sociedad Americana de Matemáticas hizo intentos para identificar a los autores de los documentos enumerados en la base de datos *Mathematical Reviews*. Desde entonces, se han multiplicado las iniciativas públicas y privadas, nacionales e internacionales, así como por parte de los productores y distribuidores de bases de datos generales y especializadas, que intentan paliar la situación desarrollando tanto sistemas de identificación de autor como de elaboración de perfiles donde se registre toda la producción científica del investigador.

Conscientes de la importancia de recoger información fidedigna y exacta de cada científico, necesaria para concurrir en convocatorias de investigación, someterse a procesos de evaluación, asignar y justificar recursos, generar indicadores bibliométricos, además de las lógicas ventajas derivadas de registrar la información sobre los protagonistas de la investigación (búsqueda y localización de expertos, mapeo de áreas fuertes de cada institución y país, etc), distintos agentes presentes en el proceso de creación y distribución de la información científica han implantado métodos propios de identificación del personal investigador.

Actualmente la mayoría de los proyectos trabajan bajo la filosofía de datos abiertos entrelazados. La aparición de proveedores de datos y de servicios usando el protocolo OAI-PMH (*Open Archives Initiative / Protocol for Metadata Harvesting*) desarrollando sus propios sistemas, convierte a la normalización e identificación de nombres en un aspecto clave para la recuperación de la información.

En este sentido, el Control Bibliográfico Universal ha cobrado una nueva perspectiva a través de varias iniciativas internacionales que pretenden reducir el coste del control de autoridades mediante la cooperación. Estas iniciativas contribuyen también a facilitar la búsqueda en entornos heterogéneos y a la desambiguación de términos dentro de la búsqueda libre (autores que aparecen referenciados en los buscadores, en los depósitos de acceso abierto y en las bases de datos internacionales de diferentes formas dependiendo de cómo lo haya interpretado el productor de la base de datos o recurso).

Se trata por tanto de un momento histórico en el que se está replanteando la identificación principal de los autores científicos. Hay diversas posibles soluciones y el problema es especialmente crucial para los investigadores, que basan su reconocimiento personal en su producción científica.

Actualmente se constata una mayor conciencia sobre la necesidad de utilización de los identificadores únicos de autor (Nature, 2009). Los investigadores tienen que inter-

actuar con una variedad de sistemas de autenticación para que los resultados de su investigación estén disponibles, perfectamente identificados y sin ambigüedades en el repositorio local, en los sistemas comerciales que la revista y los editores de libros y actas de congresos utilizan, en la maquinaria de los sistemas de información de investigación, en los índices de citas, en los datos compartidos, en las propias bases de datos y en todas las demás manifestaciones que se desprenden de estos sistemas.

La visión general de los sistemas existentes no sólo es útil para describir el status quo, sino también para entender los diferentes enfoques empleados en la identificación de la autoría y su repercusión en la actividad investigadora.

Objeto de estudio

Con el fin de clarificar dos conceptos que a menudo aparecen mezclados en gran parte de la literatura existente al respecto, se definen en primer lugar los términos de *identificador de autor* y de *perfil de autor*. Por *identificador de autor* se entiende un código con forma numérica o alfanumérica que se asigna a un autor para identificar de forma inequívoca su producción científica, con independencia de cómo firma o en qué institución trabaja. El término *perfil de autor* se aplica al conjunto de datos que recoge de forma estandarizada bien únicamente su producción científica o toda su actividad investigadora (puestos desempeñados, proyectos de investigación, contratos, etc.).

En este trabajo se exponen las principales características y modos de funcionamiento de los sistemas más representativos desarrollados hasta la fecha y que permiten generar un identificador de autor/a o un perfil profesional por cada investigador o ambos a la vez.

Se clarifica el panorama de tales sistemas, con el objetivo de aportar ciertas consideraciones a tener en cuenta tanto por parte del investigador, como de los editores, distribuidores, gestores de bases de datos y repositorios, agencias de evalua-

ción y acreditación, instituciones financieras e entidades involucradas en la política científica.

Metodología

La cada día mayor necesidad de identificar correctamente la producción científica de un autor está originando un auténtico debate en diferentes ámbitos: bibliotecario (en relación con los repositorios), documental, editorial e institucional (agencias financiadoras, evaluadoras, investigadoras), lo que sin lugar a dudas resulta altamente positivo para intercambiar experiencias, enriquecer las propuestas y lograr avances, pero también de forma coyuntural, está propiciando una multiplicación de diferentes sistemas que puede generar cierta confusión para el personal investigador.

De ahí que, como primer enfoque metodológico para abordar este tema, se proponga la siguiente clasificación:

- Hay unos Sistemas de Identificación Puros (SIP), puesto que únicamente propugnan el desarrollo de un identificador, tipo ISNI o IraLIS.
- Otros entrarían en la categoría de Sistemas de Perfil Puros (SPP) que, de momento, se centran en el desarrollo de un formato de currículum vitae normalizado, sin asignarle ningún identificador alfanumérico, tipo LATTES en Brasil o CVN de la Fecyt en España a nivel nacional. Este segundo grupo se equiparía con los denominados Sistemas de Información Científica (SIC), más conocidos (GrandIR, 2011) por su denominación inglesa de Current Research Information Systems (CRIS).
- Un tercer grupo aúna los dos anteriores al adjudicar un identificador y un perfil a cada autor, serían los Sistemas Mixtos (SM) como Research ID o Scopus Author Identifier.
- El cuarto grupo está conformado por los Sistemas Globales (SG), de última generación que integran identifi-

cadores y perfiles elaborados por cualquier operador, tipo ORCID y VIVO.

A partir de esta clasificación, se muestran los resultados del análisis realizado sobre los principales sistemas escogidos como más representativos de los cuatro grupos. Existen más ejemplos de los aquí expuestos, muchos de ellos nombrados en la bibliografía aportada (Jones, Ramsey, MacGillivray, Adams y O'Steen, 2011), e incluso es previsible que aparezcan más. No obstante, se considera que la muestra analizada refleja de forma esencial la filosofía que los impulsa así como las ventajas y contrapartidas de su implementación para la comunidad científica.

Resultados

Sistemas de Identificación Puros

Entre las iniciativas que proponen la asignación de un código identificador para cada autor hay que señalar en primer lugar el *Internacional Standard Name Identifier* o ISNI, con vocación normalizadora, amplia cobertura temática y geográfica y, presumiblemente, de pago; e IRALIS, de ámbito más local y carácter gratuito.

ISNI es un desarrollo de la norma ISO 27729, y diseñada para identificar de forma única a todos los creadores de obras creativas, ya sean individuos o grupos, por lo que tiene un amplio alcance. Aunque no se refiere de forma exclusiva al personal de investigación, sí puede aplicarse a este ya que los artículos pueden ser considerados como una obra original.

De igual manera que la producción bibliográfica está identificada con el ISBN (*International Standard Book Number*), y la hemerográfica con el ISSN (*International Standard Serial Number*), el ISNI se compone de 16 dígitos que son utilizados para la identificación exacta del autor, y pueden servir también como nexo de unión entre otros sistemas donde se almacene más información al respecto.

Esta iniciativa es auspiciada por la Organización ISNI, a su vez compuesta por la

Autoridad de Registro gestionado por la Agencia Internacional ISNI y las Agencias de Registro que proporcionan los servicios ISNI a los usuarios. Los principales partidarios son fundamentalmente las organizaciones de derechos de autor, con un fuerte componente de participación de las bibliotecas a través de la Conferencia de Bibliotecas Nacionales Europeas.

En diciembre de 2011 el sistema ISNI empezó su andadura con una base de datos inicial de más de un millón de nombres creados por los datos consolidados de VIAF (*Virtual International Authority File*), <http://viaf.org>, sustentado por 19 de las mayores bibliotecas del mundo, así como datos de otros 15 grupos, incluyendo 300 organizaciones de derechos de autor, sociedades profesionales, organizaciones gubernamentales de ayudas y subvenciones, y proveedores comerciales.

El ISNI ha sido diseñado como un identificador "puente", que permite a varios socios de una industria intercambiar información relacionada con uno de ellos sin necesidad de revelar información confidencial. Por ello el ISNI solo mantiene un mínimo de metadatos necesarios para diferenciar a dos Identidades Públicas y evitar la ambigüedad. Otra información relevante se mantiene en las bases de datos propietarias.

Después de registrar los metadatos de una identidad pública, la base de datos ISNI mantiene el URI (*Uniform Resource Identifier*) de cada una de las bases de datos propietarias, almacenando la información adicional en la identidad pública. Esas bases de datos propietarias pueden intercambiar información en un Grupo dado remitiendo a sus ISNI's.

Como los ISNI's se asignan a las entidades que tienen VIAF ID, parece lógico que se establezcan estrechos vínculos entre los dos archivos. Por otro lado, la participación de la Biblioteca Nacional de Francia y la Biblioteca Británica en ISNI ha influido en que OCLC (*Online Computer Library Center* -servicio cooperativo de más de 57.000 bibliotecas en 112 países que usan los servicios de OCLC para localizar, adquirir, pres-

tar y preservar material bibliográfico-) haya implementado VIAF.

Además, cuando el procesamiento ISNI descubre errores en el archivo VIAF (por ejemplo, dos nombres que deberían fusionarse), ISNI informará a VIAF.

Otra iniciativa de Sistema de Identificador de autor Puro, lo constituye IraLis (*International Registry of Authors-Links to Identify Scientists*). Iralis es una plataforma apoyada por la FECYT cuyo objetivo es crear un registro de autoridades que ayude a localizar las diferentes variantes de autor existentes y a mejorar la recuperación de información y la visibilidad de los autores (Baiget, Rodríguez-Gairín, Peset, Subirats y Ferrer, 2007).

IraLIS ha creado una base de datos, cuyos registros son generados por los propios autores rellenando un sencillo formulario, a partir del cual se les adjudica un código numérico que permite su identificación. En estos momentos (abril 2012) hay 3.297 autores registrados.

Sistemas de perfiles de autor puros

Los dos sistemas de perfil puros analizados responden a la iniciativa de los respectivos gobiernos encargados de la política científica de Brasil y España, y son Lattes y CVN (*Curriculum Vitae Normalizado*) respectivamente, ambos operativos a nivel nacional. No obstante, hay que señalar que en España existen también sistemas similares en ámbitos más locales como en Andalucía con el Sistema de Información Científica (SICA 2), <http://sica2.cica.es>, y el Descriptor de la Recerca i la Activitat Acadèmica de la Universitat Politècnica de Catalunya (DRAC), <http://drac.upc.edu>.

En el caso de Lattes, es el Consejo Nacional para el Desarrollo Científico y Tecnológico de Brasil (el CNPq) quien en el año 1999 puso en marcha un sistema de información curricular, con el objetivo de recabar información acerca de todos los actores institucionales involucrados en el desarrollo científico. Tal sistema, que recibe el nombre del famoso físico brasileño, Cesare Mansueti Giulio Lattes, se utiliza para la evaluación

de las competencias de los candidatos a las becas y ayudas a la investigación; la selección de consultores, miembros de comités y grupos de asesoramiento; y la subvención de la evaluación de los brasileños de posgrado y la investigación.

El sistema brasileño proporciona datos de alta calidad de más de 1,6 millones investigadores, de los que un 8% pertenecen a doctores, y alrededor de 4.000 instituciones. Este importante volumen de datos recopilado obedece a la obligatoriedad establecida en el año 2002 por las autoridades del país sudamericano para que todos los investigadores registren sus curriculum en Lattes, ya que la no inclusión del mismo impide los pagos y renovaciones de becas y ayudas.

Una de los aspectos más relevantes de Lattes es su vinculación con las bases de datos de otras instituciones así como con los datos bibliométricos proporcionados por la Web of Science de Thomson Reuters. Esta plataforma además es un servicio de base de datos de libre acceso, tanto para visualizar como para recuperar información a través de la red.

En España, el Curriculum Vitae Normalizado (CVN), es un estándar de currículum vitae permitido en convocatorias del Plan Nacional de I+D+I, y se está trabajando para que sea también aceptado en las convocatorias de carácter autonómico. CVN, utiliza un protocolo empleado para trasladar el currículum de una base de datos a otra sin tener que volver a redactar nuevamente el currículum.

El proyecto CVN es una apuesta de la FECYT y el Ministerio de Economía y Competitividad, para la implantación racional de las tecnologías de la información en la gestión de la actividad investigadora, evitando que el investigador tenga que completar de nuevo su currículum en los formularios/bases de datos de las convocatorias.

La implantación de este proyecto se inició en el año 2006 en las instituciones origen (universidades, organismos de investigación, etc.), adaptando la información ya cargada de los investigadores al estándar CVN.

Gracias al formato unificado, las bases de datos de las diferentes instituciones relacionadas con la investigación (ANEP, CNEAI, ANECA, universidades, etc.) se podrán conectar entre sí. De esta manera, cuando el investigador actualice su currículum, todas las instituciones tendrán constancia de la actualización. La administración pasa así a ser proactiva: en lugar de pedir al investigador su currículum cada vez que quiera optar a una convocatoria de ayudas a proyectos, le preguntará si los datos de que dispone son correctos y están actualizados.

La información se protege de acuerdo con las normas de la Ley Orgánica de Protección de Datos. Sólo el interesado podrá actualizar los datos. Ningún organismo ni institución dispone de ninguna versión curricular sin el consentimiento expreso del investigador propietario

Actualmente el CVN no tiene un vínculo con las bases de datos capaces de proveer el número de citas recibidas, y algún indicador de la revista de publicación, aunque esta utilidad está contemplada, al objeto de enlazar las publicaciones registradas por los autores con las bases de datos bibliográficas, de forma que se puedan elaborar indicadores bibliométricos que complementen su evaluación.

Sistemas mixtos

Los dos sistemas mixtos que se analizan responden a una filosofía netamente comercial y se caracterizan por ofrecer un código alfanumérico asociado a un perfil, aunque como se expondrá aquí cada uno posee un enfoque diferente.

Puesto en marcha en el año 2008, ResearchID, es un producto abierto a todos los investigadores impulsado por Thomson Reuters que plantea que cada investigador pase a tener un único número identificador que permita realizar las búsquedas bibliográficas sin confusiones de identidad (Cals, y Kotz, 2008).

Cada perfil profesional está asignado a un número identificador único y una dirección URL permanente a un perfil personal para facilitar la visibilidad y el ac-

ceso a sus trabajos y métricos de citas. Este identificador se compone de una letra, cuatro dígitos y el año que se registra. Si se tiene suscripción con la Web of Science se puede buscar por este código para localizar las publicaciones agrupadas por el autor.

El investigador puede añadir sus nuevas publicaciones y semanalmente se actualiza el número de citas, promedio e índice h. Actualmente en España hay 8.665 investigadores registrados. Es habitual que los propios investigadores incluyan su Researcher-id en su página web o incluso en el membrete de sus correos electrónicos.

Además de la lista de publicaciones de dicho autor, proporciona más informaciones acerca de éste, tales como afiliaciones anteriores y actuales, colaboradores, e intereses.

En principio, para registrarse y obtener un código y una página con el perfil, no se requiere una suscripción a los servicios de Thomson Reuters. Así, los usuarios no suscritos a WOK pueden buscar el registro para ver perfiles y encontrar colaboradores potenciales y revisores, aunque hay ciertas ventajas de ser un suscriptor de WOK, como el hecho de acceder al documento original. En cualquier caso, las citas importadas de WOK tienen un enlace de vuelta al registro de base de datos y generan los siguientes indicadores bibliométricos de las publicaciones introducidas desde WOS: Índice H, distribución de citas por año, total de citas y promedio de citación. Además, la herramienta *Research Labs*, ofrece recursos dinámicos para explorar redes de colaboración y de citas según autor, categoría, país o institución, mostrando los datos en forma de gráficos.

Researcher ID depende del auto-registro y auto-identificación de los artículos del investigador. No hay verificación independiente de la autoría de los artículos en este sistema y no hay métodos para evitar que un autor se registre varias veces. El autor tiene un control total sobre la información que aparece en su perfil público. Requiere por tanto una mezcla de responsabilidad individual, un sistema de honor, y para el trabajo

crítico, la verificación por parte del buscador.

Aunque se puede argumentar que el carácter comercial (Wolinsky, 2008) de esta base de datos limita su uso como estándar, el sistema tiene una ventaja clara para la investigación científica y de evaluación: el perfil resultante se pone a disposición del dominio público en cuanto lo decide el autor.

Por su sencillez de uso es un estándar destinado a triunfar, aunque su nivel de implementación actual es muy bajo y plantea la necesidad de contar con la suscripción institucional a la WOK para una explotación plena.

Por su parte, la otra gran base de datos de producción científica, SCOPUS, cuenta desde el año 2006 con las herramientas Scopus Author Identifier y Author Profile. A diferencia del Researcher ID, el sistema implementado por la editorial Elsevier, nominalmente distingue los dos conceptos; el identificador del autor –Scopus Author Identifier y el perfil del autor- author profile, aunque luego se presenten en una misma página. Ambos desarrollados para autores con artículos publicados en revistas indexadas en su base de datos.

Scopus Author Identifier es también un sistema comercial que tiene como objetivo proporcionar identificadores a los autores. El recurso asigna un número identificador exclusivo a cada investigador firmante de artículos indexados por Scopus, sin la aquiescencia del autor.

Para asignar el identificador, que es un código numérico, Scopus utiliza un algoritmo que hace coincidir los nombres de autor sobre la base de su afiliación, dirección, área temática, título de la fuente, las fechas de las citas de publicaciones y co-autores. Así, aunque las entradas no estén normalizadas, puede establecerse una estrecha relación entre las diferentes formas del nombre de una persona utilizando los datos de afiliación, aunque también puede darse el caso que un mismo autor posea más de un identificador y perfil (Aerts, 2008).

Un enlace de retroalimentación en la página Author Details (Detalles del Autor) per-

mite a los autores verificar la información sobre ellos y comunicar a la base de datos si dicha información necesita ser modificada.

Sistemas globales

Se exponen aquí dos iniciativas que han aparecido recientemente y que responden a la intención de crear sistemas globales e interoperativos, que favorezcan la integración del resto: son ORCID y VIVO.

A nivel internacional, destaca la iniciativa ORCID (*Open Researcher and Contributor ID*), lanzada oficialmente como una organización sin fines de lucro en agosto de 2010, y que pretende establecer un registro abierto e independiente asignando identificadores únicos enlazables a la producción científica del autor, independientemente del portal científico en el que aparezca. ORCID trasciende las disciplinas, las fronteras geográficas, nacionales e institucionales

El sistema quiere ayudar al proceso científico y mejorar la eficiencia de las ayudas a la investigación y la colaboración entre la comunidad científica mediante la creación de un registro central de los identificadores únicos de autor, incluyendo un mecanismo abierto y transparente de enlace entre ORCID y otros sistemas actuales de identificación.

Estos identificadores y las relaciones entre ellos, pueden estar vinculadas a la producción de los investigadores para mejorar el proceso de descubrimiento científico, la eficiencia de la financiación de la investigación y la colaboración dentro de la comunidad de investigación.

ORCID es el único servicio que trata de asociar su propio identificador con otros identificadores de autor existentes. Esta integración es necesaria para que los identificadores de autor específicos, como LATTES, Researcher-id o ID ArXiv Autor se puedan utilizar en paralelo con los identificadores universales.

ORCID cuenta con el apoyo y participación de los principales pesos pesados de la industria editorial (Elsevier, Springer, Wiley, NPG, etc), además de algunas bibliotecas universitarias (Viena, UCL, Manches-

ter), prestigiosas instituciones y asociaciones científicas como SLAC, CERN o EMBL, Thomson Reuters y CrossRef como gestor del sistema DOI cuya filosofía de identificación es común con este proyecto (García-Gómez, 2012).

La plataforma de software de ORCID es Open Source. El servicio de registro ORCID se pondrá en funcionamiento en la primavera de 2012. Hasta la fecha (abril de 2012) hay 9 instituciones españolas registradas en ORCID: la Fundación Ciudad de la Energía, el Colegio Oficial de Psicólogos de Madrid, Dialnet, iMedPub, IraLIS, la Universidad de Cádiz, la Universidad del País Vasco, la Universidad Politécnica de Cataluña y El Profesional de la Información.

Por su parte, VIVO, es una aplicación de origen norteamericano de código abierto que permite conocer la investigación en todas las disciplinas en una institución en particular y más allá de ella, mediante la utilización de Linked Open Data -una filosofía y práctica que persigue que determinados datos estén disponibles de forma libre a todo el mundo, sin restricciones de copyright, patentes u otros mecanismos de control.

La aplicación que soporta VIVO fue desarrollada en la Universidad de Cornell en 2003 y desde entonces se ha extendido en otros lugares a través de su software libre.

La ontología VIVO de interoperabilidad semántica, permite a las instituciones que participan mediante la instalación de su software VIVO, ofrecer sus datos en abierto (perfiles detallados de los profesores e investigadores, elementos, tales como publicaciones, docencia, servicio, y sus afiliaciones profesionales) y que estos sean reutilizados, participando en el desarrollo de aplicaciones colaborativas que proporcionan una búsqueda mejorada, nuevas capacidades de colaboración, la agrupación, la búsqueda y el mapeo de los científicos y su trabajo.

Con la excepción notable de VIVO, todos los sistemas de identificación de investigadores examinados son gestionados por una organización central. Corresponde a cada universidad mantener sus identificadores

que, posteriormente, ceden a VIVO que es quien asume la desambiguación de URIs (*Uniform Resource Identifier*) para la visualización de la información de un investigador en un perfil con una funcionalidad de búsqueda de gran alcance

Dado que un objetivo clave tanto de VIVO como de ORCID, es la creación de una infraestructura que haga que la información de los investigadores y sus trabajos académicos estén abiertamente disponibles, ambos están estudiando los beneficios de una colaboración tecnológica más estrecha.

Conclusión

Tras exponer las características principales de los sistemas representativos de la tipología elaborada *ad hoc*, se resaltan los siguientes aspectos:

Los Sistemas de Identificación Puros resultan de gran eficacia para evitar ambigüedad y ruido, son fáciles de utilizar, pero carecen de vinculación alguna con datos relativos al perfil profesional. En cualquier caso, un identificador único de autor en sí mismo tiene un valor limitado. Tenemos que añadir significado a él por la asociación de la identificación con la información biográfica y bibliográfica.

Por su parte, los Sistemas de Perfil Puro solventan la laguna de los anteriores, pero carecen de una clave de identificación exacta que favorezca su interrelación con otros sistemas.

Los Sistemas Mixtos examinados aúnan identificador y perfil pero tienen el gran handicap de responder a iniciativas netamente comerciales y estrechamente ligadas a sus propias bases de datos, además de carecer de la necesaria interoperatividad entre los mismos.

Los sistemas globales responden a un sistema más abierto y completo, aunque queda en suspenso la cantidad de información que puedan ofrecer al incorporar datos de diversas fuentes.

Resulta evidente que cada día es más necesaria la asociación de los registros de autor a través de los diferentes sistemas que fa-

cilitan la creación de servicios y herramientas que operan a través de internet. Por ese motivo, el enfoque para abordar el problema del nombre del autor pasa por involucrar a la comunidad de creación. Esta es la idea central de cualquier proyecto que pretenda identificar de manera inequívoca a un autor y su producción científica. De ahí que ORCID, tal y como está planteado actualmente, parece responder a esta filosofía, al integrar a todos los agentes del proceso comunicativo, y englobar los diferentes sistemas desarrollados hasta la fecha, lo que constituye su principal fortaleza, aunque a la vez plantea dificultades para navegar entre los intereses contrapuestos de sus participantes.

Algunas de las cuestiones que deben ser abordadas por un sistema identificador de autor son las siguientes: a) El identificador único del investigador debe apoyar la creación de un registro académico de manera clara e inequívoca, capaz de permitir una recolección proactiva de los datos del investigador actual, pero también retrospectiva de aquellos pretéritos; b) El identificador debe trascender las instituciones, disciplinas y fronteras nacionales; c) El identificador debe interoperar con sistemas de identificadores de investigadores que ya existen, pero son más limitados en su alcance; d) Un requisito clave es que cualquier identificador y los datos asociados a él debe ser seguro y utilizable por los sistemas de la web semántica para garantizar la interoperabilidad en el futuro; e) Sería igualmente deseable que este sistema recopilara por mecanismos automáticos las nuevas publicaciones de un autor y, previa validación por él, actualizara su registro personal.

En la actualidad, las tecnologías de la información hacen posible este tipo de herramientas, por lo que ésta es la parte más fácil de solventar. El verdadero reto es conseguir su aceptación y su uso generalizado y asiduo por la comunidad científica.

En definitiva, la identificación y construcción de un sistema de identificador de autor es un problema complejo, e implica un gran número de interesados que, a veces, tienen puntos de vista opuestos sobre algu-

nas de las cuestiones que deben abordarse: exige decisiones sobre apertura, privacidad, modelos de negocio de colaboración, y otras cuestiones fundamentales.

Por otra parte, están los aspectos relacionados con el control que tienen los autores sobre sus datos, el efecto que tendrá un registro de autor sobre los derechos de propiedad intelectual, cómo prevenir los ataques de *phishing*, cómo abordar las publicaciones de autores muertos, cómo tratar los nombres ficticios y los seudónimos...

Pero quizás es importante recalcar que el mayor desafío será cultural. Si ORCID o algún otro sistema de identificación de autor se convierte en la norma aceptada, las nuevas métricas posibles tendrán que ser tomadas en serio por todos los involucrados en el sistema académico de recompensa - los organismos de financiación, las administraciones universitarias, y los comités de promoción-. Cada aportación a la ciencia deberá ser reconocida y recompensada, no sólo la de aquellos que producen publicaciones de alto perfil.

Habrà muchos retos en el camino, pero parece evidente que en la construcción de sistemas de identificación y perfiles debe tenerse muy en cuenta el uso de herramientas como los ficheros de autoridad por las mismas razones que éstos se utilizan en los catálogos bibliográficos.

El mundo bibliotecario actual ha reformulado las bases de la normalización e identificación bibliográfica, sobre todo por medio de herramientas de software libre que gestionan registros de autoridad y empiezan a dar cuenta de ello en sus interfaces de consulta para el usuario.

A menudo se olvida que una gestión de autoridades dentro del entorno web no es sino una extensión de las funcionalidades de un catálogo bibliográfico (Tillett, 2000). No utilizar herramientas propias de la gestión de bibliotecas tradicionales, lleva en ocasiones a reinventar la rueda, normalmente cuando se descubren las ventajas para el tratamiento de la información de las viejas técnicas bibliográficas (Rodríguez-Yunta, 2012).

Así lo ha entendido OCLC al promover y desarrollar el uso cooperativo de los ficheros de autoridad en el entorno de la Web, a nivel nacional e internacional, según la filosofía de datos abiertos entrelazados *Linked Open Data* (Peset, Ferrer Sapena y Subirats-Coll, 2011) que permite ofrecer de forma abierta datos y relaciones entre datos, en lugar de documentos HTML, de manera que otras aplicaciones pueden reutilizarlos para nuevas formas de presentación de información.

Liberar datos suficientemente refutados del fichero de autores, abrirlos a su uso y explotación públicos, fomentar la reutilización de información abriendo los catálogos, directorios, vocabularios controlados u otro tipo de productos resultado del procesamiento de datos a otros agentes y lograr la difu-

sión de los mismos de forma desagregada, compartiéndolos de forma masiva y estable ayudará en la detección de duplicados, la desambiguación terminológica, el enriquecimiento de los datos o en suministrar su información en formatos más transparentes a los usuarios y permitirá una combinación más efectiva y eficiente con los identificadores y perfiles de autor.

Agradecimientos

Una de las autoras del artículo agradece al Ministerio de Economía y Competitividad por financiar su labor como técnico de apoyo a la Investigación dentro de los Servicios generales de Investigación (SGIker) de la UPV/EHU.

Referencias

- Aerts, R. (2008). Digital identifiers work for articles, so why not for authors? *Nature*, 453(19), 979.
- American Mathematical Society (1940). *Mathematical Reviews*. Recuperado el 18 de abril de 2012, de <http://www.ams.org/mr-database>
- Baiget, T., Rodríguez-Gairín, J., Peset, F., Subirats, I., y Ferrer, A. (2007). Normalización de la información: La aportación de IraLIS. *El Profesional De La Información*, 16(6), 636-643.
- Cals, J. y Kotz, D. (2008). Research identification: The right needle in the haystack. *The Lancet*, 371(9631), 2152-2153.
- FECYT (Fundación Española para la Ciencia y Tecnología) (2007). *Recomendaciones para la normalización de nombres de autores e instituciones en las publicaciones científicas*. Recuperado el 18 de abril de 2012, de http://www.accesowok.fecyt.es/wp-content/uploads/2009/06/normalizacion_nombre_autor.pdf
- Fenner, M. (2011). Author Identifier Overview. *Libreas, Library Ideas*, 18. Recuperado el 18 de abril de 2012, de <http://edoc.hu-berlin.de/libreas/18/fenner-martin-18/PDF/fenner.pdf>
- García-Gómez, C. (2012). Orcid: Un sistema global para la identificación de investigadores. *El Profesional De La Información*, 21(2), 210-212.
- Garfield, E. (1974). Errors-theirs, ours and yours. *Current Contents*, 2(25), 5-6.
- GrandIR (2011). *Jornada técnica sobre CRIS y repositorios*. Recuperado el 16 de abril, de <http://www.grandir.com/es/sesiones-tecnicas/jornada-tecnica-sobre-cris-y-repositorios-en-barcelona>
- Jones, R., Ramsey, M., MacGillivray, M., Adams, S., O'Steen, B. (2011). *Research Identifiers*. Recuperado el 10 de abril de 2012 de http://ie-repository.jisc.ac.uk/572/1/ResearchIdentifiers_DatasourcesReport.pdf
- Nature (2009). Credit where credit is due. *Nature*, 462. doi:10.1038/462825a.
- Peset, F., Ferrer Sapena, A., y Subirats-Coll, I. (2011). Open data y linked open data: Su impacto en el área de bibliotecas y documentación. *El Profesional De La Información*, 20(2), 165-173.
- Qiu, J. (2008). Scientific publishing: Identity crisis. *Nature*, 451. doi: 10.1038/451766^a
- Rodríguez-Yunta, L. (2012). Control de autoridades, una herramienta desaprovechada en los sistemas de recuperación. Recuperado el 18 de abril de 2012, de <http://www.thinkepi.net/control-de-autoridades-una-herramienta-desaprovechada-en-los-sistemas-de-recuperacion>
- Tillett, B. (2000) *Bicentennial Conference on Bibliographic Control for the New Millennium*. Recuperado el 18 de abril de 2012, de http://www.loc.gov/catdir/bibcontrol/tillett_paper.html
- Universidad del País Vasco/EHU (2011). *Producción científica de la UPV/EHU: Denomina-*

ción Bibliográfica Única. Recuperado el 18 de abril de 2012, de http://www.ehu.es/p200-content/eu/contenidos/noticia/18012012_produccion_cientifica/es_noticia/adjuntos/Uso_denominacion_UPV_EHU.pdf

Wolinsky, H. (2008). What's in a name? *EMBO Reports*, 9(12), 1171-1174.

Apéndice
Cuadro descriptivo de los principales sistemas de identificación y perfil de autor

NOMBRE	ORGANIZACIÓN	TIPO^a	CARACTERÍSTICAS	DISCIPLINAS	PAÍSES	AÑO INICIO	ENLACE
ISNI	ISNI International Organization	SIP	Norma ISO 27729 Ejemplo:1422 4586 3573 0476	Todas	Todos	2012	http://www.isni.org
IraLIS	IraLIS. International Registry of Authors-Links to Identify Scientists.	SIP	Es un sistema de estandarización de las firmas de los autores científicos. Su objetivo es crear un registro de autoridades. Ejemplo: ID3303	Todas	Todos	2007	http://www.iralis.org/
LATTES	National Council for Scientific and Technological Development (CNPq)	SPP	Parte de varias bases de datos que cubren muchas actividades académicas. Proporciona datos de organizaciones y personas. Ejemplo: K4775259P6	Todas	Brasil	1999	http://lattes.cnpq.br
CVN	FECYT	SPP	Establece un formato normalizado de currículos en soporte digital que permite mantener e intercambiar información científica curricular	Todas	España	2006	https://cvn.fecyt.es
Researcher ID	Thomson Reuters	SM	Integrado con la base de producción científica Web Of Science. Ejemplo A-3456-2012	Todas	Todos	2008	http://www.researcherid.com
Scopus Author ID	Elsevier	SM	Integrado con la base de datos de producción científica Scopus. Ejemplo: 7004464361	Todas	Todos	2006	http://www.scopus.com
ORCID	ORCID	SG	Integra con varias bases de datos y otros sistemas de identificador de autores.	Todas	Todos	2009	http://www.orcid.org
VIVO	Universidad de Cornell	SG	Aplicación de código abierto que utiliza el Linked Open Data	Todas	EEUU	2003	http://vivoweb.org/

^aSIP por Sistema de identificación Puro; SPP por Sistema de Perfil Puro; SM por Sistema Mixto, y SG por Sistema Global, siguiendo la clasificación establecida en la metodología.