

Importancia de la información de financiación

Relevance of funding information

Pablo De-Castro

De-Castro, Pablo (2018). "Importancia de la información de financiación". *Anuario ThinkEPI*, v. 12, pp. 258-264.

<https://doi.org/10.3145/thinkepi.2018.39>

Publicado en *IweTel* el 4 de diciembre de 2017



Resumen: La implantación del acceso abierto (OA) parece haber alcanzado actualmente una suerte de 'plateau' en muchos países. Como consecuencia de ello, en el movimiento hacia la ciencia abierta se hace a menudo más énfasis en la gestión de los datos de investigación (RDM). Existen no obstante aspectos en el área de la comunicación de los resultados de investigación donde resta aún mucho camino por recorrer. Uno de los más evidentes en este sentido es la codificación de la información de financiación en los registros bibliográficos de las publicaciones archivadas en los repositorios de acceso abierto. En tanto que el énfasis ha residido tradicionalmente en acumular el mayor número posible de registros de publicaciones en repositorios institucionales y temáticos, la inclusión de la información de financiación en los metadatos de dichos registros continúa siendo relativamente inusual. Este texto argumenta que ésta es un área de capital importancia para la evaluación del impacto

de la investigación y que los repositorios institucionales necesitan mejoras significativas en este sentido. Se argumenta también que para que dichas mejoras surtan efecto se requiere asimismo la aplicación de estándares abiertos sólidos por parte de las agencias de financiación de la investigación. Finalmente, se presentan diversas estrategias para la óptima captura y exposición de la información de financiación que se están aplicando en diversos entornos científicos y académicos.

Palabras clave: Gestión de la información científica; Información de financiación; Resultados de investigación; Metadatos; Agencias de financiación de la investigación; Impacto de la investigación; Acceso abierto.

Abstract: Open access implementation seems to have reached a plateau in many countries, and as a result the push for open science frequently places more emphasis on research data management. There are, however, whole areas around the open communication of research outputs which can still benefit from significant improvement. One of the most obvious is a better codification of the funding information in the bibliographic records for publications in repositories. While the emphasis has traditionally been to collect as many full-text records as possible about institutional and subject-based repositories, the typical way the funding information for publications is being coded into metadata records remains very poor, or is absent altogether. This paper argues that this is a key area for research impact assessment and improvement. It's also argued that such an improvement needs solid housekeeping on the research funders' side, possibly based on new and evolving standards. A few different approaches to capture and display the funding information in metadata records in an optimal fashion are discussed.

Keywords: Research information management; Funding information; Research outputs; Metadata; Research funders; Research impact; Open access.

1. Muchas lagunas por cubrir

Cuando se discuten los presentes estándares para la codificación de la información de financiación en los registros de publicaciones archivados en repositorios, es útil recordar el propósito original de los repositorios de acceso abierto: se concibieron para “almacenar, preservar y difundir” los resultados producidos en el seno de una institución dedicada a la investigación. Tradicionalmente, los repositorios han estado ligados al contenido y a su difusión en línea. Los objetivos han ido no obstante deviniendo más y más ambiciosos a medida que progresábamos, y estos días ya no estamos únicamente interesados en el contenido y su difusión en línea, sino también en muchas otras áreas tales como identificadores persistentes, métricas alternativas o revisión por pares abierta por mencionar sólo algunas de ellas. Persisten sin embargo lagunas más que evidentes en el objetivo original relativo al contenido y su exposición que impiden que los resultados sean totalmente satisfactorios. La recogida y exposición de la información de financiación de las publicaciones es una de las más claras.

“En el movimiento hacia la ciencia abierta se hace a menudo más énfasis en la gestión de los datos de investigación que en el acceso abierto”

Es cierto que los repositorios no fueron originalmente diseñados para este propósito, y que determinados estándares de metadatos pueden carecer de la necesaria profundidad para abordar esta complicada área. Pero no es menos cierto que para que la próxima generación de

repositorios sea eficaz no basta con implantar nuevas y elegantes funcionalidades sobre las plataformas actuales: además de esto, la calidad del contenido y en especial de su descripción a nivel de metadatos ha de mejorar de manera significativa. Buenas prácticas en esta materia tales como el trabajo realizado por *OpenAIRE* en el ámbito de la recogida de información de financiación para los resultados de investigación procedentes de proyectos europeos –que se están viendo ampliadas en la actualidad a determinadas agencias de financiación de ámbito nacional (*OpenAIRE*, sf)– muestran la relevancia y el potencial de esta capa adicional de metadatos en términos de adecuarse a los objetivos de las agencias de financiación de la investigación.

“Ya no estamos únicamente interesados en el contenido y su difusión en línea, sino en muchas otras áreas como identificadores persistentes, métricas alternativas o revisión por pares abierta”

Estos días las agencias de financiación se interesan de manera primordial por el impacto de la investigación. Este resbaladizo concepto continúa discutiéndose por el momento, con una refrescante aproximación empírica basada en el diseño y el análisis de estudios de caso a nivel institucional (*REF2014*, sf), conviviendo con los intentos más teóricos de definir de manera precisa qué es y en especial cómo (diablos) puede medirse el impacto de la investigación. Está muy claro en todo caso a estas alturas que uno de los elementos clave en la evaluación del impacto

de la investigación reside en el establecimiento de un vínculo entre los resultados de investigación y los proyectos financiados que han hecho posible su aparición. En una sociedad movida por los resultados económicos, esta posición tiene que ver con la rendición de cuentas y el retorno de la inversión. Críticamente, tiene asimismo que ver con la transparencia y el gobierno abierto: dado que la investigación tiende a ser una actividad financiada con fondos públicos, resulta evidente que debería

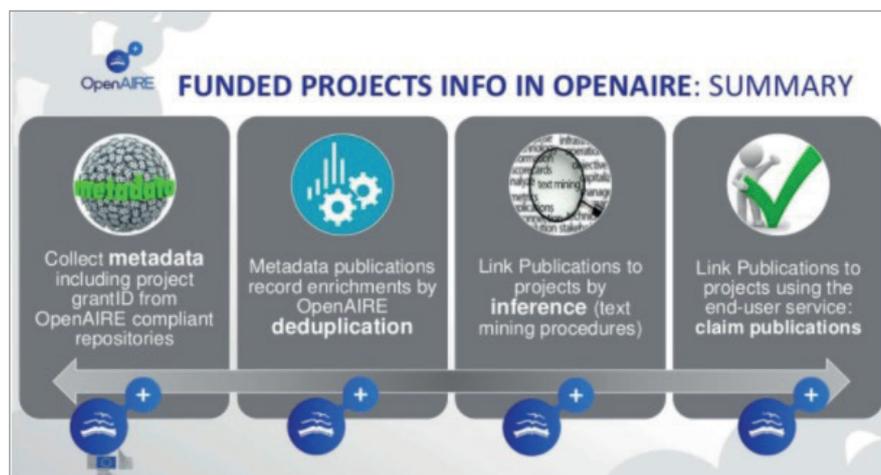


Figura 1. Líneas de actividad del proyecto *OpenAIRE* relativas a proyectos financiados

ofrecerse al contribuyente la posibilidad de verificar qué es lo que se está obteniendo como resultado de la inversión del dinero público.

Estos vínculos entre los resultados de investigación y sus fuentes de financiación no se han percibido tradicionalmente como una prioridad a nivel de la codificación de información en los metadatos de las publicaciones. Se trata de un área muy compleja que parecería pertenecer al ámbito de la administración interna tanto de instituciones como de agencias de financiación más que a algún lugar próximo a las bibliotecas de investigación. Sólo de manera relativamente reciente han comenzado las agencias de financiación públicas a emitir una serie de directrices para indicar a los investigadores el modo correcto de agradecer sus fuentes de financiación en sus publicaciones (*RIN*, 2018). Por lo que respecta a los autores, se encuentran a menudo demasiado absorbidos por su actividad investigadora como para prestar excesiva atención a la correcta descripción de la información de financiación en los agradecimientos de un manuscrito que con frecuencia es el resultado de la colaboración entre muchos coautores con diferentes afiliaciones institucionales y múltiples fuentes de financiación. Los editores (publishers) se encuentran así en una posición ideal para abordar este nivel de complejidad, e iniciativas como *FundRef* (*CrossRef*, 2017) y las mejoras introducidas por los proveedores de sistemas de procesamiento de manuscritos proporcionan significativas muestras de progreso en este campo.

“Los vínculos entre los resultados de investigación y sus fuentes de financiación no se han percibido tradicionalmente como una prioridad a nivel de la codificación de información en los metadatos de las publicaciones”

La complejidad asociada a la codificación de la información de financiación en las plataformas institucionales sigue siendo muy elevada. Aunque es cierto que los editores tienen la ocasión de tomar la iniciativa en esta área requiriendo a sus autores que especifiquen sus fuentes de financiación en el momento del envío de sus manuscritos (y convirtiendo de paso este proceso en “una pesadilla” de acuerdo con los comentarios de

Funding

This study was supported by grants from the Spanish Ministry of Science and Competitiveness (BFU2014-56692-R to AG and JMD-G) and the Junta de Andalucía (Spain; BIO-122, CVI-02487, and P07-CVI-02686 to AG and JMD-G), from the European Community’s Seventh Framework Program (FP7/2007-2013 to AG and DB) under grant agreement number 201714 (DEVANX), from the Austrian Science Fund (SFB-F43 to MAK) and the DFG (FOR2333 to MAK; SPP-1738 to MAK and DB). RS is a Boehringer Ingelheim Fonds fellow.

Figura 2. Ejemplo de agradecimientos en un manuscrito

un número significativo de autores), los sistemas en los que las agencias de financiación confían de manera tradicional para el establecimiento de los vínculos entre publicaciones y proyectos financiados son las plataformas institucionales. Es por ello que ejemplos como la sección de agradecimientos de la figura 1 –que incluye referencias a proyectos financiados a nivel nacional por parte de tres países diferentes, más financiación regional procedente de uno de estos tres países, más dos proyectos europeos y una agencia de financiación privada– muestran la magnitud del reto de codificar correctamente dicha información. ¡Y éste no es siquiera un artículo particularmente complejo!

“Las agencias de financiación de la investigación no están necesariamente versadas en el ámbito de la transparencia, de los estándares abiertos o del rol de las instituciones como garantes de la pública disponibilidad de la información científica”

2. ¿Una tarea para las bibliotecas? No way!

La adición de información de financiación a registros bibliográficos es un ámbito en el que los *Current Research Information Systems* o sistemas CRIS tienden a ser particularmente eficientes. No sólo porque éste es el tipo de proceso de gestión interna de la información científica para el que estos sistemas se diseñaron originalmente, sino sobre todo porque contienen tablas relacionales con la información de todos los proyectos financiados para la institución a la que sirven. La existencia de estas tablas convierte el proceso de inclusión física de las referencias en una tarea mucho más accesible, incluso para los autores. Pero para estos sistemas CRIS sigue habiendo problemas: cada vez que aparezca una referencia de financiación externa

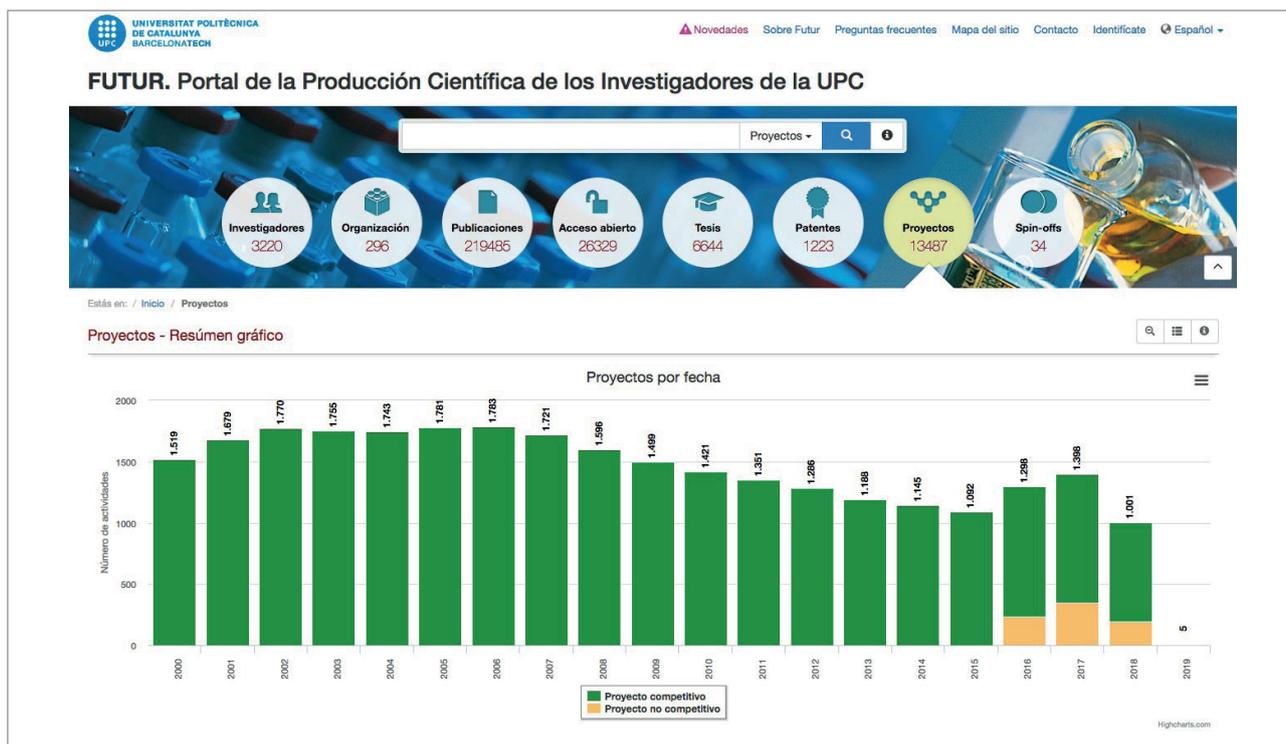


Figura 3. Información de proyectos en el portal *Futur* de la *Universitat Politècnica de Catalunya*
<https://futur.upc.edu/proyectos>

a la organización, el sistema CRIS carecerá de una entrada pre-existente para dicho proyecto, y será preciso crearla de manera específica, con toda probabilidad de manera manual. Esto se convierte por tanto en una cuestión de metadatos, un área que pertenece al campo de conocimiento de las bibliotecas. En un momento histórico en el que las bibliotecas de investigación se interrogan sobre su futuro y su valor en los procesos de difusión de la información científica, ésta es claramente un área que una dirección de biblioteca decidida podría considerar interesante. Existen ya numerosas buenas prácticas al respecto en el sector, y los resultados de la encuesta *RIM 2017* sobre la gestión de información científica (*OCLC, 2017*) recientemente lanzada por *OCLC Research* y *euroCRIS* mostrarán la (considerable) medida en que las bibliotecas se encuentran ya involucradas en el apoyo a las tareas de gestión de la información científica.

“En el marco de la implantación de la ciencia abierta, es tarea de cada país el identificar los foros y los mecanismos más eficaces para el establecimiento de un diálogo eficaz con los diversos actores involucrados en el proceso”

Ningún esfuerzo en este sentido por parte de las instituciones será completamente fructífero sin una tarea previa de normalización por parte de las agencias de financiación públicas en los diversos niveles administrativos (**De-Castro, 2017**). Las instituciones llevan mucho tiempo haciéndose cargo de procesar la información sobre publicaciones de una manera estandarizada, pero la información sobre proyectos financiados a los que es preciso enlazar dichas publicaciones es responsabilidad directa de las agencias de financiación, no de las instituciones. Estas agencias de financiación son con frecuencia (con muy honrosas excepciones) recién llegadas en el mundo del intercambio de información científica. Están acostumbradas a tratar con los investigadores a los que financian y con los investigadores que evalúan las propuestas de proyectos, pero no necesariamente versadas en el ámbito de la transparencia, de los estándares abiertos o del rol de las instituciones como garantes de la pública disponibilidad de la información. Además de esto, no existen todavía estándares técnicos reconocidos por la comunidad en el ámbito de la definición de los proyectos de investigación a nivel de metadatos.

Existen sin duda excelentes prácticas hoy día en cuanto a la puesta a disposición pública de la información de proyectos financiados por una agencia concreta: la plataforma *Cordis*

de la *Comisión Europea* para proyectos *FP7* y *H2020* y la base de datos *Gateway to Research* de los *Research Councils UK (RCUK)* son dos de los más estimulantes ejemplos de la manera correcta de abordar esta área, máxime cuando ambas agencias –a través respectivamente de *OpenAIRE* y de *ResearchFish* como infraestructuras de apoyo a nivel técnico y humano– están ya compartiendo en directo la información de las publicaciones resultantes de los proyectos que financian. Pero tanto la *Comisión* como *RCUK* son ejemplos de agencias de financiación muy solventes económicamente y constituyen más bien la excepción que la norma. Es tarea urgente para la comunidad de gestión de información científica en su conjunto la de persuadir del valor de estos avances a las agencias de financiación públicas más pequeñas o menos informadas sobre los caminos por los que transita estos días el intercambio de información científica entre las diversas organizaciones involucradas. *OpenAIRE* está realizando una labor ingente en este ámbito a nivel europeo, pero tiene limitaciones a la hora de alcanzar los niveles más profundos de las administraciones públicas. En el marco de la implantación de la ciencia abierta, es tarea de cada país identificar los foros y los mecanismos más eficaces para el establecimiento de un diálogo eficaz en este sentido.

A nivel técnico, estos avances apuntan a la eventual creación de una tercera capa de identificadores persistentes de proyecto de investigación a añadir a los identificadores de autor y de organización ya existentes. Basta consultar las bases de datos de proyectos mencionadas más arriba para comprobar que existe ya un cierto estándar de facto para la definición de los elementos de metadatos que deben emplearse en la descripción de un proyecto. El hecho de que muchas agencias de financiación permanezcan al margen de estos desarrollos no significa que no estemos ya en condiciones de alcanzar progresos significativos en la materia. Estos y otros aspectos relacionados con los identificadores persistentes se discutieron de manera concienzuda en el evento *PIDapalooza* en Girona el pasado mes de enero. <https://pidapalooza.org>

3. Estrategias para la recogida de información de financiación en los sistemas institucionales

Está en primer lugar la adición manual de metadatos con la información de financiación. Ésta es la aproximación estándar para proyectos europeos de acuerdo con las directrices de *OpenAIRE* para repositorios de publicaciones (la estrategia aplica igualmente para los datos de investigación, pero el énfasis suele estar en las publicaciones, que son las que cuentan para los ejercicios de evaluación de la actividad científica). Esta estrategia se ha aplicado de manera muy eficiente en una amplia red de repositorios de acceso abierto en Europa, lo que permite a *OpenAIRE* proporcionar a la *Comisión Europea* una impresionante colección de indicadores relativos a los resultados de investigación asociados a proyectos europeos y a los niveles de cumplimiento de las políticas de acceso abierto emitidas por la *Comisión*.

<https://www.openaire.eu/h2020-stats>

Con un buen número de iniciativas en activo orientadas a monitorizar los progresos –generalmente a nivel nacional– hacia el objetivo acordado de una implantación completa del acceso abierto para el año 2020, una estrategia basada en el enlace entre publicaciones y proyectos financiados resulta ciertamente de interés.

Más aún, las directrices de *OpenAIRE* para repositorios de acceso abierto se han adoptado a estas alturas mucho más allá de las fronteras

Showing results 1 to 20 of 2940		next >
3M		1
A. G. Leventis Foundation		35
Aarhus University Research Foundation		4
Abbott Fund		1
AbbVie Pharmaceuticals		3
Abdus Salam International Centre for Theoretical Physics		4
Abengoa		2
ABS Graduate School (Finland)		2

Figura 4. Clasificación de publicaciones por proyecto financiado en el repositorio *Digital.CSIC*

europas. Si bien por el momento la recogida de información de financiación tiene una prioridad mucho menor que la pura acumulación de contenido, esta área forma también parte de las directrices y deberá abordarse conjuntamente con la actividad de gestión de contenido que representa el núcleo de la implantación del acceso abierto. *OpenAIRE* entretanto ha comenzado a colaborar con un buen número de agencias de financiación de ámbito nacional en diferentes países (sobre todo europeos, pero no sólo) para explorar la posibilidad de reutilizar esta estrategia –que tan exitosa ha resultado a nivel comunitario– basada en la recogida de información comprensiva acerca de proyectos financiados que pueda agregarse a la creciente capa de metadatos contextuales para los registros de publicaciones.

“La estrategia automatizada basada en la minería de datos sobre los agradecimientos de los textos completos de las publicaciones almacenadas en los repositorios tiene grandes visos de devenir la aproximación por defecto en el futuro”

Existe por tanto una estrategia manual. Algunas instituciones contrataron en su momento personal específico de apoyo al repositorio con el fin de completar la introducción manual en los registros de metadatos de la información de financiación incluida en los agradecimientos de todas las publicaciones almacenadas en sus repositorios hasta ese momento. La aplicación de esta estrategia de enriquecimiento retrospectivo de registros de publicaciones puede ser factible en casos en los que el número de registros disponibles no es demasiado elevado, pero resulta escasamente sostenible a medio plazo. De hecho, es mucho más frecuente hoy día encontrar repositorios teóricamente compatibles con *OpenAIRE* cuyos registros carecen de los metadatos básicos para los proyectos europeos que repositorios cuya información de financiación cubra más allá de los proyectos *FP7* y *H2020*. No obstante, hay excepciones.

Una de ellas es *Digital.CSIC*, el repositorio institucional del Consejo

Superior de Investigaciones Científicas. La principal razón por la que es una excepción es que la plataforma *Dspace* sobre la que dicho repositorio operaba inicialmente se actualizó en un momento dado a un sistema *Dspace-CRIS* cuyo modelo de datos compatible con *Cerif* admitía la inclusión de información de financiación. Una vez que esto se verifica, la interoperabilidad entre repositorios y sistemas *CRIS* puede fácilmente hacer el resto. Sistemas *CRIS* como *ConCiencia* pueden alimentar los repositorios con la información de proyectos financiados ligados a las publicaciones, y los repositorios eventualmente cerrarán el círculo a través de *OpenAIRE*.

<https://digital.csic.es/dcl/pasarela.jsp>

Por otra parte, se han realizado ya ensayos preliminares sobre la estrategia automatizada que tiene grandes visos de devenir la aproximación por defecto en el futuro. Esta se basa en la aplicación de la minería de datos sobre los agradecimientos de los textos completos de las publicaciones almacenadas en los repositorios (comenzando como norma con los repositorios temáticos por el enorme volumen de contenidos disponible en ellos) y en la subsiguiente identificación automática de la referencia de financiación en el artículo contra una combinación de nombre de la agencia de financiación y código de proyecto almacenada en las comprensivas tablas de proyectos financiados facilitadas por las agencias de financiación a iniciativas como *OpenAIRE*. Este proceso de identificación automática requiere un



Figura 5. Investigación no financiada

considerable nivel de aprendizaje por parte de la máquina, pero la estrategia ya está aplicándose a nivel experimental con prometedores resultados (**Giannakopoulos; Foufoulas; Dimitropoulos**, 2017).

“Es mucho más frecuente hoy día encontrar repositorios teóricamente compatibles con OpenAIRE cuyos registros carecen de los metadatos básicos para los proyectos europeos que repositorios cuya información de financiación cubra más allá de los proyectos FP7 y H2020”

Tan innovadora y prometedora como esta estrategia automatizada puede parecer, no funcionará sin embargo si los autores cometen errores en los agradecimientos tales como la omisión del nombre de la agencia de financiación o la (muy frecuente) introducción de erratas en los códigos de proyecto. Éste es de nuevo un campo en el que una biblioteca involucrada en el apoyo a la investigación podría encontrar un nicho, por ejemplo colaborando con investigadores y grupos de investigación para la revisión sistemática de las referencias incluidas en los agradecimientos tan pronto como sea posible en el ciclo de publicación de modo que posibles errores se detecten y corrijan a tiempo y no lleguen a publicarse. Esta vertiente del apoyo a la investigación, que se está ensayando de manera razonablemente satisfactoria en la *University of Strathclyde*, permite a la biblioteca establecer estrechos vínculos con los investigadores a la vez que recoge una exhaustiva –y fascinante– imagen de la investigación financiada a nivel institucional.

Por otra parte, esta estrategia de colaboración con los autores se beneficia notablemente de la disponibilidad de financiación para cubrir las tasas por procesamiento de artículo (APC) desde la biblioteca: cuando estos fondos están disponibles, los investigadores en la institución se ponen regularmente en contacto con la biblioteca para consultar las opciones de financiación de las APCs para sus manuscritos aceptados. Dado que los fondos para APCs provienen en la mayoría de los casos de las propias agencias de financiación y están ligadas a los proyectos financiados, las opciones de cobertura de las APCs se confirmarán una vez

que se revise el texto completo del manuscrito aceptado para publicación y en especial cualquier referencia a proyectos financiados en su sección de agradecimientos. Si bien esta actividad no es algo que suela tenerse a la hora de establecer juicios de valor sobre el modelo de negocio basado en las tasas por procesamiento de artículos, podría en alguna medida justificar el apego que por dicho modelo existe en determinadas latitudes.

No está de más recordar finalmente junto a todas estas consideraciones acerca de la relevancia de la información de financiación el hecho de que una fracción muy significativa de la investigación realizada en las instituciones no procede de ningún proyecto financiado. Esta circunstancia, que establece grandes divisiones a nivel de oportunidades entre las ciencias “duras” y las ciencias sociales y humanidades, hace recomendable una estrategia de apoyo a la investigación que no se limite al soporte de proyectos financiados. Pero éste es tema para otro *ThinkEPI*.

4. Referencias

- CrossRef* (2017). “Funder registry”. *CrossRef*.
<https://www.crossref.org/services/funder-registry>
- De-Castro, Pablo** (2017). “The quest for funding information”. *euroCRIS blog. Current research information systems*, 18 agosto.
<http://eurocris.org/blog/quest-funding-information>
- Giannakopoulos, Anastasios; Foufoulas, Yannis; Dimitropoulos, Harry** (2017). “Funding information extraction”. En: *Open Science Fair*, 6-8 September, Athens (poster).
<http://www.opensciencefair.eu/posters/funding-information-extraction>
- OCLC (2017). “An increasing role for libraries in research information management”. *OCLC Research*.
<http://oc.lc/rim>
- OpenAIRE* (sf). “For funding agencies”. *OpenAIRE Newsletter*.
<https://www.openaire.eu/support/faq/funders-faq>
- REF2014 (sf). *REF2014 Impact case studies*.
<http://impact.ref.ac.uk/CaseStudies>
- RIN (2018). “Acknowledgement of funders in journal articles”. *Research Information Network (RIN)*.
<https://goo.gl/KLkuHV>

Pablo De-Castro
University of Strathclyde
pcastromartin@gmail.com