



INFORME DE LA COMISIÓN DE SEGUIMIENTO SOBRE EL GRADO DE CUMPLIMIENTO DEL ARTÍCULO 37 DE LA LEY DE CIENCIA

Fecha de publicación 29 Junio 2016



RESUMEN EJECUTIVO

En los últimos años la comunidad científica, así como las entidades públicas financiadoras de la investigación, han adquirido el convencimiento de que la difusión abierta de la ciencia no sólo es justa, sino que comporta mejor ciencia. Justa porque se hace público lo que se financia con fondos públicos, mejor ciencia porque ésta llega a más gente y de forma más fácil.

El Open Access o Acceso Abierto a los resultados de la investigación están hoy en el centro de la mayoría de las instituciones con algún papel en la disseminación de la ciencia. España se ha sumado de forma activa al movimiento con el artículo 37 de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación, que marcó la obligación legal para la comunidad de científicos españoles de depositar en acceso abierto una copia de los artículos que publicaran en el marco de sus proyectos de I+D+i nacionales.

A pesar de que el OA esté en el horizonte inmediato, los diferentes agentes están buscando las mejores maneras de facilitar la transición hacia el acceso abierto y de medir sus avances con unas pautas de partida.

El objetivo de este informe es hacer una primera medición del cumplimiento de esta obligación para que sirva como referencia y punto de arranque para desarrollar una política de seguimiento y proponer un plan de actuación con las medidas necesarias para avanzar en este cumplimiento.

Pero, al igual que les sucede a otros países y organismos, esta tarea no es fácil. Para poder obtener un indicador o indicadores, se han encontrado limitaciones importantes. La principal es que aún no se cuenta con una base de datos oficial de los artículos resultantes de convocatoria de proyectos de I+D+i nacionales. La segunda es que, a pesar de que existe un agregador, RECOLECTA¹, que integra los metadatos de las publicaciones depositadas en todos los repositorios institucionales existentes de universidades y otros centros de investigación, en muchos de estos repositorios no se incluye el metadato que identifique de forma unívoca las publicaciones recogidas con los proyectos de I+D+i que han financiado esta investigación.

Frente a la disyuntiva de esperar a tener las condiciones ideales para recolectar el metadato o de dar por primera vez un indicador, aunque sea aproximado, que mida el grado de cumplimiento de la Ley, se ha optado por esta segunda. Por ello, se ha elegido una metodología de aproximación que permite obtener un primer valor de referencia cuya difusión es imprescindible para concienciar a la comunidad investigadora y poder avanzar en el cumplimiento de la publicación en abierto.

El valor obtenido es de un 9%, lo que, sin duda, resulta muy bajo. Más adelante se expondrán las medidas que se han considerado para aumentar este porcentaje, así como para lograr una mayor exactitud de la medición.

¹ www.recolecta.fecyt.es

ANTECEDENTES

La publicación de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación marcó la obligación legal para la comunidad de científicos españoles de depositar en acceso abierto una copia de los artículos que publicaran en el marco de sus proyectos nacionales de I+D+i.

Cómo hacer frente a esta obligación, recogida en el artículo 37 “Difusión en acceso abierto” de la Ley, suscitó numerosas dudas en los distintos ámbitos de aplicación que hicieron necesaria la elaboración, publicación y difusión de unas [recomendaciones](#) para su correcta aplicación. Este documento², publicado en octubre de 2014, definió los aspectos principales de la política nacional de acceso abierto y despejó el camino a todos los actores afectados. Para ello especificó cuáles eran los nuevos roles que debían adoptarse y elaboró una serie de pautas destinadas a todos los colectivos implicados en la producción y gestión del mercado de la información científica.

Una de las recomendaciones que se hizo a los gestores de ayudas públicas a la I+D+i puso de manifiesto la necesidad de constituir una comisión de seguimiento que midiese el grado de cumplimiento del mandato de depósito en abierto por parte de la comunidad investigadora. Esta comisión fue creada por FECYT en el primer trimestre del año 2015 y constituida por miembros de Universidades y CSIC con mandato en acceso abierto y con experiencia en ese ámbito de aplicación³

Desde ese momento, la comisión de seguimiento analizó la existencia de otros posibles indicadores para concluir que el indicador dado en las recomendaciones citadas anteriormente era el adecuado para medir el grado de cumplimiento del mandato. Este mandato obliga a los investigadores españoles al depósito en acceso abierto de todos los artículos que se publiquen con fondos públicos nacionales. La mejor forma de medir el mandato del depósito en acceso abierto sería aquel indicador que reflejara el porcentaje de los artículos en acceso abierto resultantes de la financiación de convocatorias de proyectos I+D+i nacionales depositados en los repositorios institucionales.

El indicador definido fue por tanto:

Artículos resultantes de convocatorias de proyectos I + D + i nacionales depositados en acceso abierto en repositorios institucionales
Artículos resultantes de convocatorias de proyectos I + D + i nacionales

² Fundación Española de Ciencia y Tecnología (FECYT) *Recomendaciones para la implementación del artículo 37 Difusión en Acceso Abierto de la Ley de la Ciencia, la Tecnología y la Innovación*. 2014.

³ Lluís Anglada (CSUC), Teresa Malo de Molina (Universidad Carlos III), Remedios Melero (CSIC) e Inmaculada Ribes (Universitat Politècnica de València)

METODOLOGÍA

El primer paso para el cálculo del indicador en el periodo 2011-2014 fue obtener los trabajos depositados en repositorios institucionales españoles que respondan a proyectos nacionales financiados en este periodo. Se contactó con los gestores de los repositorios presentes en RECOLECTA para requerir los datos y se propuso un plazo de recogida entre el 31 de julio y el 15 de septiembre de 2015. La recolección de datos debía ser homogénea en todos los repositorios: inclusión de los trabajos en acceso abierto de proyectos del plan nacional entre 2011 y 2014. Para ello se tendrían en cuenta los metadatos correspondientes a derechos de autor, puesto que habría que filtrar la búsqueda a los registros en abierto o embargados, y el metadato sobre el órgano de financiación del artículo.

Para obtener el número total de artículos publicados en el universo potencial de artículos financiados con fondos públicos de convocatorias de I+D+i, se optó por utilizar como fuente de datos lo recogido en las principales bases de datos de referencias bibliográficas.

Debido a que la base de datos SCOPUS no contiene la información necesaria respecto al campo de entidad financiadora, la comisión optó por utilizar la Web of Science (WOS), que cuenta con metadatos mejor informados y de mayor calidad asociados a las entidades financiadoras, como fuente de obtención del dato, consulta realizada 03 de junio de 2016.

Por lo tanto, el indicador quedó definido de este modo:

$$\frac{\text{Artículos resultantes de convocatorias de proyectos I+D+i nacionales depositados en repositorios institucionales, en abierto o embargados, de las instituciones que pudieron aportar datos}}{\text{Artículos resultantes de convocatorias de proyectos I+D+i nacionales financiados-indexados en WoS}}$$

Cálculo para el periodo 2011-2014

Numerador: Artículos resultantes de convocatorias de proyectos I+D+i nacionales depositados en acceso abierto o embargados en repositorios institucionales.

Para la obtención del numerador, se realizó una petición a los 68 repositorios institucionales registrados en RECOLECTA de todos los registros de tipo "artículo", financiados con cargo a proyectos de I+D+i nacionales, con fecha de publicación entre el 01/01/2011 y el 31/12/2014, tanto abiertos como embargados.

El listado debía contener: título, autor, año de publicación del artículo, ID del repositorios (handle u otros), DOI, código de proyecto, título de la fuente, dc.right (especificando si es "open" o "embargado") y patrocinio (*sponsorship*) en los casos en los que exista este campo.

Denominador: Número de artículos resultantes de convocatorias de proyectos I+D+i nacionales financiados.

Búsqueda en la WOS: Para la obtención del denominador y determinación del número de artículos resultantes de convocatorias de proyectos financiados de I+D+i nacionales, se siguió el siguiente proceso en la base de datos de Web of Science:

- Se buscó el número de artículos españoles resultantes de convocatorias de proyectos I+D+i nacionales financiados entre 2011 y 2014.
- Sentencia de búsqueda: (OG⁴=(*Institución1*" OR " *Institución2*"....) AND (FO=(Spain OR Spanish OR MINECO OR MEC OR MINCINN OR Ministerio OR Espana OR CSIC OR "Cons* Sup* Investi* Cient*" OR "Agen* Esta* Cons* Sup* Investi* Cient*" OR "Spanish National Research Council" OR ISCIII OR "Inst* Sal* Carl*" OR "Institute of Health Carlos") OR FT=(Spain OR Spanish OR MINECO OR MEC OR MINCINN OR Ministerio OR Espana OR CSIC OR "Cons* Sup* Investi* Cient*" OR "Agen* Esta* Cons* Sup* Investi* Cient*" OR "Spanish National Research Council" OR ISCIII OR "Inst* Sal* Carl*" OR "Institute of Health Carlos") OR FG=(Spain OR Spanish OR MINECO OR MEC OR MINCINN OR Ministerio OR Espana OR CSIC OR "Cons* Sup* Investi* Cient*" OR "Agen* Esta* Cons* Sup* Investi* Cient*" OR "Spanish National Research Council" OR ISCIII OR "Inst* Sal* Carl*" OR "Institute of Health Carlos"))) AND Tipos de documento: (Article)
- La búsqueda se realizó en los índices=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Período de tiempo=2011-2014.

4 Nombre de normalizado de la institución en WOS por cada una de las instituciones que han aportado datos

LIMITACIONES DEL ESTUDIO

Tras la exposición metodológica del estudio que se ha realizado, es importante tener en cuenta las principales limitaciones que se han encontrado en la medición del numerador y el denominador que constituyen el indicador de cumplimiento del artículo 37 de la ley de la Ciencia. Se detallan seguidamente:

- Respecto al numerador: En el momento de la realización del trabajo⁵ no todos los repositorios institucionales incluían algún metadato que identificara el proyecto que hubiese financiado la investigación en la que se basa la publicación, lo que redujo el número de repositorios que potencialmente podían responder a la demanda del estudio. Para futuros análisis sería deseable que todos los repositorios pudiesen ofrecer esta información.

Además, los escasos resultados que enviaron los repositorios llevan a pensar que no siempre se incluye en el metadato específico el código de proyecto, lo que significa que la cifra aportada por los repositorios estaría por debajo de la real.

Como fuente del numerador se han considerado los repositorios institucionales dejando fuera del estudio los repositorios temáticos, cuando la Ley de la Ciencia permite el depósito también en éstos.

Por otra parte, en el caso de publicaciones en coautoría, la Ley de la Ciencia considera satisfecho el acceso abierto con el depósito de la publicación por parte de tan sólo uno de los autores del mismo. De hecho en este estudio se han eliminado aportaciones duplicadas de un artículo por parte de varias instituciones. Esta actuación, si bien mide adecuadamente el grado de cumplimiento del artículo 37 de la Ley, es cierto que rebaja el porcentaje de cumplimiento de los repositorios por instituciones en el caso de artículos descartados por ser considerados duplicados.

Debido a estas limitaciones el indicador ofrecido muestra un grado de cumplimiento de la Ley de la Ciencia posiblemente inferior al real, e imposible de determinar con más precisión en el marco de este estudio.

- Respecto al denominador: debería proceder en lo sucesivo de una base de datos oficial, que recoja la totalidad de proyectos financiados y las publicaciones derivadas de los mismos, y no de una fuente proveniente de una empresa privada. Al no poder manejar datos oficiales se ha acudido a la base de datos más completa disponible, la WoS, que presenta algunas limitaciones importantes:
 - No recoge la totalidad de la producción científica de las instituciones
 - No es representativa de las áreas temáticas de Humanidades y Ciencias Sociales.

⁵ La petición de los datos se realizó en junio de 2015

- No tiene normalizado el nombre de las instituciones, por lo que hay que hacer las búsquedas combinando variantes de los nombres de las universidades y eso hace que se escapen variantes y que las búsquedas no siempre sean 100% completas.
- Se han detectado fallos en la filiación de los autores, lo que significa que no se recogen adecuadamente todos los resultados.

RESULTADOS

- De los 68 repositorios institucionales que forman parte de RECOLECTA, respondieron un total de 58 lo que representó un 85% de respuestas⁶, de los cuales enviaron datos 25 instituciones. El número total de registros brutos obtenidos es de **9.739**.
- Del total se identificaron y eliminaron los registros duplicados.
- Se normalizaron los títulos con el fin de realizar el mapeo de los registros con la base de datos WoS.
- Como resultado de este primer trabajo de normalización se obtuvieron **9.044** registros⁷ (sin eliminar los duplicados el número de registros era de **9.310**).

Mapeo de estos registros con WoS:

Los repositorios institucionales recogen potencialmente toda la producción científica de sus investigadores pero WoS solamente indexa una porción concreta: la que se publica en las revistas que forman parte de su base de datos de referencias bibliográficas. Por ello, con el objetivo de asegurar que la comparación se efectuaba entre dos poblaciones análogas, se hizo un *cruce* entre los registros aportados por los repositorios y los indexados en WoS, de modo que en el numerador solamente se incluyesen aquellos registros que estuviese incluidos en la base de datos de WoS. Para ello se siguió el siguiente proceso:

- Búsquedas de los títulos de las publicaciones (no de los registros individuales) entre 2011 y 2014 en WoS. Para ello se normalizaron los títulos de las fuentes y se realizó la búsqueda a través de los ISSN.
- Se identificaron un total de **1.874** títulos independientes de revistas en los registros aportados, de los cuales **1.570** tienen cobertura en WoS, que corresponden a los registros en acceso abierto y embargados. Por lo tanto, el grado de coincidencias a nivel de título de revista es del **83%**.

Filtrado de registros abiertos o embargados:

- Se filtraron los registros en los que el metadato dc:right indica que el documento está en acceso abierto o embargado.

⁶ En el anexo 1 se puede consultar el listado de las instituciones que respondieron a la petición. El número total de instituciones fue de 55 ya que en el momento de la petición había alguna de ellas que tenían varios repositorios

⁷ En este número se engloban también: registros enviados que no están en acceso abierto, otros que no se corresponden al periodo de 2011-2014, otros que no son artículos y por último otros que no están indexados en WoS

- Una vez aplicado el filtro se obtuvieron **5.952** registros abiertos o embargados en los repositorios participantes en el estudio, de los cuales **5.496** (un **92%**) se encontraban en WoS.

Número de artículos resultantes de convocatorias de proyectos I+D+i nacionales financiados.

El número de artículos resultantes de convocatorias de proyectos I+D+i nacionales financiados entre 2011 y 2014, con las condiciones impuestas en la ecuación de búsqueda e indexados en WoS fue de **60.293 artículos** correspondientes a las 22 instituciones indicadas en el anexo 2.

Por lo tanto con estos datos, se desprende que solo el 9% de los artículos de proyectos de I+D+i nacionales contienen alguna información sobre la financiación y se encuentran depositados en acceso abierto en repositorios institucionales. Este porcentaje podría estar por debajo del dato real teniendo en cuenta las limitaciones del método de cálculo, sobre todo por la falta de estandarización en los metadatos del código de proyecto.

CONCLUSIONES Y RETOS DE FUTURO

Teniendo en cuenta lo anteriormente expuesto cabe concluir que con los datos manejados provenientes de los repositorios, no se puede realizar una medición con mayor precisión por falta de gran parte de las contribuciones de los repositorios institucionales de España.

Habría que considerar este estudio como una primera medición que ha de servir como punto de partida para mejorar los sistemas de identificación y recuento de datos. Es importante señalar que no es más que una aproximación que debe entenderse en el seno de unas limitaciones y características que han tenido lugar en la recogida de datos, que son indispensables a tener en cuenta para su interpretación.

Una vez hecha esta primera medición y contando con una metodología concreta y adecuada, sería deseable, con vistas a recoger adecuadamente el universo del estudio, poder repetir este trabajo siguiendo los mismos pasos metodológicos pero teniendo en cuenta las siguientes consideraciones:

- Todos los repositorios deberán contar con los metadatos para poder recolectar los derechos de acceso/reutilización de los artículos, el código de proyecto, y disponer en cada uno de los ítems de un identificador único (tipo DOI). En este sentido RECOLECTA deberá actuar como dinamizadora de la mejora.
- Además se animará a los repositorios a recoger en sus registros el código de financiación del proyecto. Esta medida permitirá automatizar la recolección y el mapeo de datos para evitar el trabajo manual y los posibles errores que esto pueda producir.
- Convendrá disponer de una base de datos con todos los proyectos I+D+i financiados.
- Se solicitará al Ministerio el esfuerzo de adaptar las sedes electrónicas de gestión de proyectos de forma que incluyan en los informes anuales los metadatos de las publicaciones derivadas del proyecto, entre los que además de título, autores etc., estará la URL del trabajo depositado en el repositorio institucional o temático correspondientes.

Una vez aplicada esta metodología y comprobadas las limitaciones que manifiesta, la comisión encargada de este estudio considera que, en paralelo a la puesta en marcha de las consideraciones recién expuestas, sería adecuado acometer una FASE 2 del estudio en la que se aplicase una metodología complementaria.

En este sentido, para evitar las limitaciones expuestas en la obtención del numerador, se podría trabajar con la WOS y el set completo de registros de repositorios españoles. Para ello se requeriría seguir un método de análisis bibliométrico capaz de mapear más de 100.000 registros.

ANEXO 1.

Listado de instituciones de los repositorios que respondieron a la petición

- Agencia Estatal CSIC
- Consorci de Serveis Universitaris de Catalunya
- Fundación Juan March
- Instituto de Investigación y Formación Agraria y Pesquera
- Instituto Español de Oceanografía
- TECNALIA
- Universidad Autónoma de Madrid
- Universidad Carlos III de Madrid
- Universidad Católica San Antonio de Murcia
- Universidad CEU San Pablo
- Universidad Complutense de Madrid
- Universidad de Alcalá
- Universidad de Alicante
- Universidad de Almería
- Universidad de Burgos
- Universidad de Cantabria
- Universidad de Castilla-La Mancha
- Universidad de Córdoba
- Universidad de Extremadura
- Universidad de Granada
- Universidad de Huelva
- Universidad de Las Palmas de Gran Canaria
- Universidad de León
- Universidad de Málaga
- Universidad de Murcia
- Universidad de Navarra
- Universidad de Oviedo
- Universidad de Salamanca
- Universidad de Sevilla
- Universidad de Valladolid
- Universidad de Zaragoza
- Universidad del País Vasco/Euskal Herriko Unibertsitatea
- Universidad Europea de Madrid
- Universidad Francisco de Vitoria
- Universidad Internacional de La Rioja
- Universidad Miguel Hernández
- Universidad Nacional de Educación a Distancia
- Universidad Pablo de Olavide
- Universidad Politécnica de Cartagena
- Universidad Politécnica de Madrid
- Universidad Pompeu Fabra
- Universidad Pontificia de Salamanca
- Universidad Pública de Navarra
- Universidad Rey Juan Carlos
- Universidade da Coruña
- Universitat de Barcelona

- Universitat de Girona
- Universitat de les Illes Balears
- Universitat de Lleida
- Universitat de Valencia
- Universitat de Vic
- Universitat Jaume I
- Universitat Oberta de Catalunya (UOC)
- Universitat Politècnica de Catalunya
- Universitat Politècnica de València

ANEXO 2.

Instituciones de las que se han obtenido los datos para este estudio

- Agencia Estatal CSIC
- Instituto Español de Oceanografía
- Universidad Autónoma de Madrid
- Universidad Carlos III de Madrid
- Universidad CEU San Pablo
- Universidad Complutense de Madrid
- Universidad de Alcalá
- Universidad de Burgos
- Universidad de Cantabria
- Universidad de Sevilla
- Universidad del País Vasco/Euskal Herriko Unibertsitatea
- Universidad Politécnica de Cartagena
- Universidad Politécnica de Madrid
- Universidad Pública de Navarra
- Universidad Rey Juan Carlos
- Universidade da Coruña
- Universitat Autònoma de Barcelona
- Universitat de Girona
- Universidad de Granada
- Universitat Jaume I
- Universitat Politècnica de Catalunya
- Universitat Politècnica de València