

PLAN DE INVESTIGACIÓN (PROYECTO DE TESIS DOCTORAL)

PROGRAMA DE DOCTORADO EN FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO

UNIVERSIDAD DE SALAMANCA

TÍTULO: Visibilidad e impacto de la literatura gris científica en repositorios institucionales de acceso abierto. Estudio de caso bibliométrico del repositorio Gredos de la Universidad de Salamanca

AUTOR: M^a del Tránsito Ferreras Fernández

DIRECTOR: José Antonio Merlo Vega

Francisco José García Peñalvo

FECHA: 20 de marzo de 2015

INTRODUCCIÓN Y JUSTIFICACIÓN DEL TEMA OBJETO DE ESTUDIO

Las tecnologías de la información y especialmente Internet han cambiado el modo en el que producimos, publicamos y comunicamos la información hacia modelos en los que la información se produce fundamentalmente en formato digital y se consume a través de medios online. Gracias a Internet, la distribución de documentos electrónicos es fácil, simple y barata. Una de las formas más importantes y eficientes para el intercambio de conocimientos es la filosofía del acceso abierto al conocimiento, porque el conocimiento es absolutamente indispensable para la evolución social en un mundo globalizado (García-Peñalvo et al., 2010a).

Según Peter Suber (2004) el movimiento de acceso abierto depende de las tecnologías de Internet y del consentimiento del autor o titular de los derechos de autor (Arlitsch and O'Brien, 2012). Los repositorios institucionales fueron uno de los productos de este movimiento; estos capturan los resultados intelectuales del profesorado, de los investigadores y de los alumnos de las instituciones académicas y aseguran un acceso perpetuo y libre a estos resultados, salvando los periodos de embargo u otras restricciones del editor.

Así pues, los repositorios institucionales (RIs) a menudo, y prácticamente en todos los casos, difunden tesis electrónicas inéditas (Schöpfel, 2013), porque las tesis doctorales no dejan de ser el trabajo de investigación primigenio de todo investigador. Hasta ahora, esta falta de difusión de estos trabajos de investigación inéditos hizo que fueran considerados como literatura gris. En este sentido, los repositorios son una poderosa herramienta para la difusión de la investigación llevada a cabo en las universidades a través de la difusión de las tesis defendidas en el seno de la institución, y por ello consideramos que las tesis en abierto deberían dejar de calificarse como literatura gris si se difunden en acceso abierto.

Durante la última década los RIs están alcanzando el peso suficiente para ser fuentes viables para las publicaciones de investigación. También parecen contribuir de manera significativa a las tasas de citación del autor. Los repositorios institucionales pueden desempeñar un papel crucial en la medición de resultados de la investigación, y a su vez pueden afectar a los rankings universitarios (The Times Higher Education, 2010).

En relación con el tema del aumento de la visibilidad y de la citación de los trabajos en acceso abierto a través de repositorios (vía verde) hay un número importante de estudios de prestigiosos autores comprometidos con el acceso abierto, como Stevan Harnad y Tim Brody (2004), Xia y Nakanishi (2012), Lawrence (2001), Piwovar y Vision

(2013). La mayoría de estas investigaciones han demostrado que los documentos de acceso abierto tienen una ventaja en el número de citas (Gargouri et al., 2010).

En el presente estudio, nos planteamos otra pregunta: ¿Tienen las tesis en acceso abierto una ventaja en citación sobre las tesis que no están en acceso abierto? De acuerdo con Richard Jones y Theo Andrew (2005) la práctica de hacer versiones digitales de tesis disponibles en línea está creciendo a nivel internacional. Los Repositorios de Tesis y Disertaciones Electrónicas (ETDs) son habituales en las universidades de todo el mundo. Los datos de estos repositorios sugieren un aumento espectacular en el uso y la citación de tesis doctorales en la actividad de la investigación actual, y esto se relaciona con el aumento del descubrimiento de recursos y la facilidad de acceso que confieren los sustitutos digitales.

En la Universidad de Salamanca, el depósito de tesis en el repositorio fue regulado por mandato en 2008. Desde 2011 el "Real Decreto 99/2011" también incluye en su capítulo 14 la obligación de difundir las tesis en acceso abierto.

El presente estudio se desarrolla en este contexto. Consideramos que tanto el crecimiento del movimiento de acceso abierto, como el desarrollo de nuevos esquemas de metadatos para facilitar la interoperabilidad y la preservación a largo plazo (García-Peñalvo et al., 2010b), como la adopción de políticas de mandato de acceso abierto por parte de las instituciones (Harnad, 2009), incrementan la visibilidad de la investigación depositada en los repositorios institucionales, incrementando de esta forma la citación y el impacto.

HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS A ALCANZAR

El trabajo de investigación que se presenta pretende dar respuesta a la pregunta sobre la visibilidad, usabilidad e impacto de las tesis doctorales depositadas en repositorios de acceso abierto. ¿Tienen las tesis en acceso abierto un mayor número de citas que las tesis que no son de acceso abierto?

El principal objetivo de este trabajo es demostrar que la difusión de las tesis a través de repositorios institucionales tienen más visibilidad y por tanto más citación e impacto que las tesis que no son de libre acceso. En el desarrollo de esta hipótesis debemos plantearnos además otra pregunta: ¿qué factores intrínsecos y extrínsecos influyen en la visibilidad de las tesis depositadas en repositorios institucionales de acceso abierto?

Como hemos indicado existen estudios que demuestran el aumento de citación de los trabajos de investigación publicados y puestos en acceso abierto frente a aquellos trabajos publicados y que no están en acceso abierto. La novedad de nuestra investigación radica en el estudio de la literatura gris científica, campo mucho más difícil de explorar a la hora de obtener datos de uso y citación que el de los trabajos publicados por los canales tradicionales de edición.

El objetivo principal implica una serie de sub-objetivos:

1. Demostrar que la interoperabilidad de los repositorios (aspecto técnico) y las políticas de mandato afectan positivamente a la visibilidad y al impacto de la investigación depositada en los repositorios institucionales.
2. Definir el estado de la cuestión de los repositorios institucionales de acceso abierto tanto desde un punto de vista estratégico-político como tecnológico.
3. Definir un modelo de interoperabilidad para la visibilidad y el mayor impacto de la literatura gris científica.
4. Realizar un estudio cuantitativo en el repositorio Gredos que tome como elemento de estudio las tesis doctorales de la Universidad de Salamanca.
5. Conocer y aplicar métricas alternativas para la realización del estudio.
6. Establecer un modelo bibliométrico viable para la medición del uso y de la citación de las tesis depositadas en acceso abierto.

METODOLOGÍA A UTILIZAR (APORTAR CONFORMIDAD/INFORMES/PROTOCOLOS GARANTIZANDO BIOÉTICA/BIOSEGURIDAD SI EL TIPO DE EXPERIMENTACIÓN LO REQUIERE)

La presente investigación se compone de dos partes: una teórica y otra práctica.

Teórica

En la parte teórica se realizará un estudio pormenorizado del estado de la cuestión sobre el conocimiento científico en acceso abierto, sobre los repositorios institucionales, como base para el caso de estudio del repositorio Gredos de la Universidad de Salamanca, sobre la literatura gris científica y sobre el impacto de la investigación en acceso abierto.

Se identificará la bibliografía especializada en las siguientes áreas:

- El conocimiento científico en acceso abierto
- Repositorios institucionales: políticas de depósito en acceso abierto, interoperabilidad de los repositorios y diseminación de los contenidos, visibilidad y rankings web de los repositorios.
- Citación e impacto de la investigación en acceso abierto: estadísticas de uso, índices de citas e impacto y métricas alternativas.

Para ello se llevará a cabo una revisión sistemática bibliográfica de las principales fuentes de la literatura científica en la web, en los catálogos de bibliotecas y en las bases de datos especializadas. Determinando, dentro de las áreas anteriormente expuestas, palabras clave y autores relevantes en la materia. La gestión de la bibliografía se realizará con Mendeley.

Se revisarán los estudios sobre visibilidad, citación e impacto de la investigación en acceso abierto, a través de bibliografía sobre el tema.

Se analizará la relación entre buenas prácticas de gestión técnica y el aumento de visibilidad e impacto de los repositorios institucionales de acceso abierto.

Se analizarán cuantitativamente los aspectos técnicos en la implementación y mejoras en el repositorio Gredos.

Empírica

La parte práctica estará basada en el análisis cualitativo del repositorio Gredos, a través del estudio de la experiencia de usuario, el seguimiento de citas y estudios de impacto y del análisis de rankings de repositorios institucionales. Esta investigación se dirige a demostrar que las tesis de la Universidad de Salamanca que se difunden en acceso abierto a través del repositorio Gredos alcanzan valores más altos de visibilidad, citación e impacto que aquellas otras tesis que no se difunden en acceso abierto.

El desarrollo de esta práctica se basará en:

- Realización de un análisis cuantitativo de la colección de investigación en el repositorio Gredos "Tesis Doctorales" a través del propio repositorio y de las estadísticas que este ofrece.

- Realización de un análisis cuantitativo de las tesis de la Universidad de Salamanca que no están depositadas en el repositorio.

- Estudio de diseminación y visibilidad de las tesis de la Universidad de Salamanca depositadas en Gredos, a través de los recolectores u otros portales de tesis existentes (por ejemplo TDR, DART-Europe y Google Scholar)

- Seguimiento de citas de los autores de las "Tesis Doctorales" a través de las herramientas en 2.0 que se ofrecen para la investigación en acceso abierto y a través de sistemas de métricas alternativas.

- Estudio de la evolución del repositorio Gredos en los rankings mundiales de repositorios.

Una vez revisados los datos se realizará un análisis y comparación de resultados para determinar que:

- las tesis doctorales depositadas en acceso abierto en Gredos tienen mayor citación e impacto derivados de una mayor visibilidad en la Web;

- la investigación en acceso abierto de la Universidad de Salamanca da mayor visibilidad a la institución y por tanto mayor impacto;

- y que las políticas de acceso abierto (mandato) de las instituciones favorecen la visibilidad de la institución y por lo tanto su prestigio.

Asimismo podremos establecer un sistema de análisis bibliométrico viable para la medición del uso, la visibilidad, la citación y el impacto de las tesis depositadas en repositorios institucionales de acceso abierto.

Finalmente se establecerán las conclusiones de la investigación.

MEDIOS Y RECURSOS MATERIALES DISPONIBLES

Se ha elaborado una base de datos con la muestra de tesis a analizar de la Universidad de Salamanca, que recoge una serie de variables e indicadores sobre el uso, la visibilidad y la citación de las mismas.

La información para elaborar la base de datos ha sido extraída de la base de datos del Ministerio de Educación, Cultura y Deporte, TESEO, y del repositorio GREDOS.

Los datos estadísticos de visitas y descargas se han extraído del módulo de estadísticas de la Universidad de Tasmania, utilizado en el repositorio GREDOS

Se han extraído también datos a través de las herramientas Google Analytics y AddThis utilizadas por GREDOS.

Para el trabajar con los datos hemos utilizado Excel.

Se han utilizado herramientas 2.0, y métricas alternativas para la medición y análisis de los datos, como Google Scholar, Almetrics, Google Analytics, Publish or Perish y Mendeley.

Para la realización de la revisión de la literatura y elaboración de la bibliografía se han utilizado recursos suscritos por la Universidad de Salamanca y recursos en acceso abierto.

PLANIFICACIÓN TEMPORAL AJUSTADA A TRES AÑOS / CINCO AÑOS (Tiempo parcial)

1) Estudio de la bibliografía. Se está llevando a cabo un estudio detallado de la bibliografía existente a través de la revisión sistemática de la literatura científica sobre los temas planteados. La bibliografía se está gestionando con Mendeley.

Periodo: desde mayo 2012 hasta marzo 2015.

2) Elaboración de la base de datos de la muestra a partir de TESEO y el repositorio GREDOS. Ya se ha realizado.

Periodo: desde enero de 2013 hasta abril de 2013.

3) Toma de datos de los indicadores sobre uso, visibilidad y citación utilizados en la muestra. Ya se ha realizado.

Periodo: agosto de 2014.

4) Estudio de los datos de la muestra: normalización y estudio con Excel. Ya se ha realizado este estudio sobre una muestra aleatoria del total de la población a estudio. En la actualidad se está realizando sobre la muestra completa.

Periodo: desde septiembre de 2014 hasta abril de 2015.

5) Elaboración de una propuesta de análisis bibliométrico de aplicación a las tesis depositadas en repositorios institucionales de acceso abierto con el fin de medir el uso, la visibilidad, la citación y el impacto de las mismas.

Periodo: abril/mayo 2015.

6) Redacción de los capítulos de la tesis: febrero /junio 2015. En la actualidad se está realizando la fase de redacción.

7) Lectura de la tesis: septiembre 2015.

BIBLIOGRAPHICAL REFERENCES

Arlitsch, K. & O'Brien, P.S. (2012). Invisible institutional repositories: addressing the low indexing ratios of IRs in Google. *Library Hi Tech*, 30(1), 60-81.

García-Peñalvo, F.J., Figuerola, C.G. de & Merlo, J.A. (2010a). Open knowledge: challenges and facts. *Online Information Review*, 34, 520-539.

García-Peñalvo, F.J., et al. (2010b). Qualified Dublin Core Metadata Best Practices for GREDOS. *Journal of Library Metadata*, 10 (1), 13-36.

Gargouri, Y. et al., 2010. Self-selected or mandated, open access increases citation impact for higher quality research. *PloS one*, 5(10).

Harnad, S. and Brody, T. (2004). Comparing the impact of open access (OA) vs non-OA papers in the same journals. *D-Lib Magazine*, 10 (6). Available at: www.dlib.org/dlib/june04/harnad/06harnad.html.

Harnad, S. (2009). Integrating Universities' Thesis and Research Deposit Mandates. In 12th International Symposium on Electronic Theses and Dissertations, University of Pittsburgh, Pittsburgh Pennsylvania, 10 - 13 Jun 2009.

Jones, R. and Andrew, T. (2005). Open access, open source and e-theses: the development of the Edinburgh Research Archive. *Program: electronic library and information systems*, 39(3), 198-212.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411, 521.

Piwohar H.A. and Vision T.J. (2013). Data reuse and the open data citation advantage. *Peer J.*, 1(e175). Available at: <http://dx.doi.org/10.7717/peerj.175>

Schöpfel, J., 2013. Adding value to electronic theses and dissertations in institutional repositories. *D-Lib Magazine*, 19(3-4).

Suber, P. (2004). Very brief introduction to open access. Available at: www.earlham.edu/~peters/fos/brief.htm (accessed 15 October 2011).

The Times Higher Education (2010). The Times Higher Education World University Rankings 2010-2011. Available at: www.timeshighereducation.co.uk/world-university-rankings

Xia, J. and Nakanishi, K. (2012). Self-selection and the citation advantage of open access articles. *Online Information Review*, 36 (1), 40- 51.