

Algunas reflexiones sobre acceso abierto, repositorios y cooperación

JOSÉ MANUEL BARRUECO

Universitat de València. Servei de Biblioteques i Documentació

ORCID: 0000-0001-7916-7847

En este trabajo se presenta la experiencia del autor en la puesta en marcha y desarrollo de la biblioteca digital RePEc (Research Papers in Economics). RePEc es una de las principales iniciativas en el desarrollo del acceso abierto a nivel internacional, así como una de las fuentes de información en abierto más relevante para los investigadores en el área de Economía. Además de una descripción del funcionamiento de esta iniciativa, se enumeran los factores que han marcado su éxito y que la diferencian de otros proyectos con similares objetivos. Por otro lado, tomando como referencia los puntos fuertes de RePEc, se analizan aspectos de cooperación bibliotecaria y repositorios institucionales en las bibliotecas universitarias de nuestro país, indicando algunos puntos que podrían mejorarse.

acceso abierto, repositorios, bibliotecas digitales, cooperación, RePEc

“No, pero puedo aprender” fue la respuesta allá por 1993, a un mensaje de Thomas Krichel, cuando me preguntó si conocía *Linux*, *perl*, *emacs* ... imprescindibles para empezar a colaborar en el proyecto que estaba poniendo en marcha. Entonces, recién había acabado la Diplomatura de Biblioteconomía en la Universidad de Salamanca y acababa de incorporarme a trabajar en la biblioteca de la Facultat d’Economia de la Universitat de València.

Eran años de agitación profesional. De expectación hacia las nuevas oportunidades que empezaban a apuntar con eso que se denominó nuevas tecnologías de la información y la comunicación, aka internet. De posibilidades ingentes de innovación, en un momento que la red aún estaba ligada al entorno universitario, antes de que diera el salto hacia la explotación comercial y las puntocom.

A partir de ese momento comenzó mi implicación con el movimiento de acceso abierto y los repositorios. En esos ámbitos se ha centrado toda mi vida profesional, que ha estado marcada por la puesta en marcha de RePEc (Research Papers in Economics) junto con Thomas Krichel. Thomas forma parte de ese grupo de innovadores que sentó las bases de las bibliotecas digitales y lanzó el movimiento *open access*, entre los que están también, Carl Lagoze, Herbert van de Sompel, Stevan Harnad, Paul Ginsparg, Clifford Lynch o Michael L. Nelson¹.

¹ <https://www.openarchives.org/meetings/SantaFe1999/ups-participants.htm>

El objetivo inicial que nos movió fue aprovechar los servicios de internet para facilitar el acceso a los resultados de investigación en Economía. Unos resultados que tradicionalmente se venían distribuyendo en forma de documentos de trabajo. Desde entonces, RePEc ha evolucionado hasta convertirse en una de las bibliotecas digitales más importantes a nivel internacional y, sin duda, el recurso de información en abierto más consultado por los investigadores en Economía.

Solamente unas cifras para ponerlo en contexto: proporciona metadatos sobre más de tres millones de recursos de investigación (artículos, working papers, capítulos de libros...), datos de autoridad sobre más de 15.000 investigadores registrados y cuenta con la colaboración de más de 2.000 proveedores de datos a nivel mundial. Entre ellos se encuentran organizaciones que van desde pequeños departamentos o centros de investigación universitarios hasta organismos como el International Monetary Fund o la OCDE, pasando por las grandes editoriales Elsevier, SpringerNature ...

Porque RePEc no es solamente acceso abierto... Se trata de una iniciativa que trata de englobar a todas las entidades que forman parte del ecosistema de comunicación científica de la disciplina. Y en ese ecosistema, la industria editorial juega un papel esencial. En este sentido, es particularmente interesante la simbiosis que RePEc ha tenido siempre con editoriales comerciales y que demuestra que el acceso abierto no está reñido con la distribución comercial de contenidos. Es decir, que puede existir un modelo híbrido donde, desde un mismo servicio, se puede proporcionar acceso tanto a la versión de una obra publicada bajo suscripción, como a diferentes versiones preprint, en abierto, de la misma, sin que eso suponga ningún detrimento en los derechos de las editoriales. En Economía es habitual que los autores distribuyan varias versiones preprint de sus trabajos, una en cada institución de la que son miembros. Posteriormente, ese trabajo se publica en forma de artículo en una revista, comercial o en abierto. RePEc agrega los metadatos de todas las versiones en un solo registro y es el lector el que elige qué versión descargar.

RePEc ha sido una iniciativa innovadora en muchos aspectos, como, por ejemplo:

- Su arquitectura distribuida, que fue diseñada en 1997, se basa en la recolección de metadatos desde unos proveedores de datos (o repositorios) hacia unos proveedores de servicios (o recolectores). Es la misma arquitectura que luego fue la base del protocolo OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting), desarrollado en 1999. OAI-PMH es el protocolo en el cual se fundamenta toda la infraestructura de repositorios institucionales actuales. En este sentido, RePEc representa un ecosistema propio e independiente. No es un repositorio temático como habitualmente se le categoriza, sino un ecosistema completo donde interactúan más de 2.000 entidades equivalentes a repositorios institucionales, que se agregan en un número reducido de recolectores denominados servicios (IDEAS², EconPapers³, CitEc⁴, Collec⁵, RAS⁶, ...). A través de estos servicios, el investigador interactúa con el sistema para: realizar búsquedas, recibir alertas de nuevos trabajos, consultar indicadores de calidad, o cosas más curiosas como, consultar una genealogía de la disciplina donde se

² <http://ideas.repec.org>

³ <http://econpapers.repec.org>

⁴ <http://citec.repec.org>

⁵ <http://collec.repec.org>

⁶ <http://authors.repec.org>

ofrece información de quién fue alumno de quién, dónde y cuándo... Desde ese microcosmos, existen pasarelas que permiten la recolección de contenidos desde terceras partes como OpenAIRE, Sciverse o WorldCat. Esas pasarelas constituyen el punto de unión de RePEc con otros servicios multidisciplinares.

- En el desarrollo de esquemas de metadatos. Cuando aún no existía el Dublin Core, fue necesario diseñar un esquema propio, muy sencillo, basado en texto, que pudiera ser editado por cualquier persona sin necesidad de ninguna infraestructura técnica ni conocimientos específicos. En muchos casos, los metadatos son creados por personal administrativo de los departamentos, sin conocimientos bibliotecarios.
- En el desarrollo de identificadores y procesos para la desambiguación de autores. Algo similar a lo que luego se puso en práctica con ORCID. Los identificadores persistentes son una parte esencial y están disponibles para cualquier entidad descrita en RePEc. Esos identificadores propios, basados en el sistema *handle*, permiten la identificación, localización y resolución de cualquier ítem (autor, documento, institución, revista).

El valor añadido de los repositorios a debate

Sin duda, las principales innovaciones que han diferenciado RePEc de otras iniciativas similares, han sido en el campo de los servicios de valor añadido. Cuenta, por ejemplo, con un servicio de alertas basado en el trabajo editorial de revisores temáticos, una especie de peer-review post-publicación. Estos revisores seleccionan los contenidos más relevantes en su área y los publicitan cada semana a través de boletines de novedades. Cuenta también con un servicio de índice de citas autónomo que nació bastante antes que Google Scholar Citations: CitEc (Citations in Economics). El desarrollo, puesta en marcha y mantenimiento de CitEc ha sido mi contribución personal a RePEc. Inicialmente fue gestado como mi proyecto de fin de carrera para la licenciatura de Documentación en la UPV⁷. Ha evolucionado hasta convertirse en una base de datos de más de 55 millones de referencias, que proporciona indicadores de calidad sobre documentos, autores, revistas e instituciones.

Si extrapolamos la experiencia de RePEc al resto de infraestructuras actuales de repositorios institucionales, creo que es precisamente en el campo de los servicios de valor añadido donde más carencias encontramos. Es necesario poner más trabajo por parte de las bibliotecas universitarias. Es fundamental que los repositorios proporcionen alguna utilidad a los investigadores para incentivar su implicación y compromiso. Mientras que RePEc ha puesto el foco en los autores como destinatarios de los servicios, y les ha demostrado su utilidad en el día a día de su trabajo, la utilidad que aportan los repositorios institucionales es, cuanto menos, bastante mejorable. Es cierto que hemos conseguido posicionar los resultados de investigación de nuestras universidades en Google Scholar, pero deberíamos ir mucho más allá.

Es necesario, por tanto, posicionar a los repositorios en los flujos de trabajo cotidianos de los investigadores. Que cuando necesiten información, utilicen nuestros buscadores; que cada día reciban alertas sobre las novedades en sus áreas de interés; que puedan acudir a un perfil donde seamos capaces de agregar su producción académica completa, tanto en abierto como en cerrado, de su institución actual y de toda su carrera; que puedan depositar fácilmente sus trabajos; que tengan datos integrados de popularidad (accesos/descargas) de todos sus trabajos

⁷ <https://polibuscador.upv.es/permalink/f/6lh1qt/alma2132529600003706>
<https://www.uv.es/~barrueco/PFC.pdf>

independientemente del repositorio donde se guarden; que cuando haya procesos de selección acudan a nuestros indicadores de calidad, sí, basados en citas, para evaluar sus publicaciones; que se les informe cómo se posicionan sus publicaciones respecto de las de otros compañeros. Desgraciadamente, parte de eso ya existe y no hemos sido los repositorios quienes lo hemos llevado a cabo, sin embargo, aún nos quedan posibilidades de mejora...

Ese escenario debe lograrse a un nivel superior de agregación de contenidos, por encima del repositorio, al nivel de recolectores. Si nos fijamos en los servicios proporcionados por recolectores como OpenAIRE, RECOLECTA o LaReferencia, las opciones que ofrecen a los investigadores son reducidas. Claramente, necesitamos más y mejores agregadores de contenidos y en la medida de lo posible, de carácter temático, dado que las características y necesidades de cada disciplina son diferentes. El hacer agregadores geográficos (Europa, España, Latinoamérica respectivamente) es lo más sencillo, pero no lo más útil. La pregunta para la reflexión aquí sería, ¿somos capaces de recolectar de nuestros repositorios, de forma efectiva, sencilla y sistemática, los contenidos disponibles sobre un área temática determinada? Como diría Isidro, vuestro turno...

Los problemas que tenemos planteados a mi juicio, después de dos décadas de desarrollo de repositorios institucionales, son dos. En primer lugar, los agregadores se han orientado exclusivamente hacia proporcionar motores de búsquedas, cuando esta es una carrera que tenemos perdida desde la salida a manos de Google Scholar. Mucho más cuando la calidad de nuestros metadatos no permite implementar funciones más específicas o avanzadas de lo que ofrece Google.

En segundo lugar, creo que hemos equivocado el sujeto al que dirigimos nuestras actividades. Las actuaciones de los agregadores están dirigidas hacia los propios repositorios, cuando debería ser hacia la comunidad de investigadores. En estos momentos, proporcionan servicios como la validación de metadatos, que, aunque muy útil, carece de sentido si no va acompañada de una mejora en su explotación. La normalización es crucial, pero crear buenos metadatos es costoso en recursos. Para justificar la inversión en metadatos completos y de calidad, es necesario algún incentivo adicional. En el caso de RePEc, cuanto mejores sean los metadatos, mayores puntos de acceso y, por consiguiente, mayor visibilidad para los recursos, lo cual incentiva la inversión en normalización.

Otra actividad promovida por los agregadores nacionales y europeos, que está llevando mucho tiempo a nuestros repositorios, es la identificación en los metadatos de las entidades financiadoras de la investigación. El objetivo es poder medir si se cumplen o no los mandatos institucionales de acceso abierto. De esta forma se convierte a los repositorios en fiscalizadores de la actividad de nuestros investigadores y a al personal de las bibliotecas en los guardianes del cumplimiento de la ley. Siempre me he preguntado, ¿Para quién hacemos eso? O ¿Con qué objeto? O ¿Qué retorno tenemos? No dudo que pueda ser interesante conocer esa información, pero ... ¿De verdad eso está entre las prioridades de nuestros investigadores? Quizás, la nueva funcionalidad de Google Scholar venga a liberarnos de ese trabajo⁸

La cooperación como modelo de negocio

Sin duda el aspecto más innovador de RePEc es su modelo de negocio. Básicamente, no hay negocio. RePEc funciona gracias a la comunidad de investigadores. Toma el modelo *open source*

⁸ <https://scholar.googleblog.com/2021/03/track-and-manage-your-public-access.html>

como principio fundamental, basado en esa filosofía de colaboración que guio el desarrollo de internet hasta los años 90. Se ha demostrado que es posible exportar estos principios al campo de las bibliotecas y que se puede aprovechar la potencialidad de la colaboración para crear un servicio que perdure en el tiempo sin necesidad de financiación estable.

Los principios de cooperación que guían el desarrollo de RePEc son:

- Descentralización. Tanto en el trabajo como en los datos. Se trata de dividir y facilitar la cooperación en la medida de los recursos de cada participante. Cada miembro de la comunidad aporta según sus posibilidades, que pueden ir desde las más modestas como corregir una referencia o una cita, a las más complejas como proporcionar metadatos sobre miles de artículos publicados o implementar un nuevo servicio de valor añadido. Esta descentralización, donde no existe ningún nodo ni participante esencial o gestor del sistema, garantiza, igual que en el caso de la propia internet, que el sistema esté a salvo de caídas o, lo más importante, que no pueda ser controlado o comprado por una multinacional. Nunca podrá suceder lo que ocurrió con SSRN (Social Science Research Network) cuando fue adquirida por Elsevier.
- Cooperación abierta, fundamentada en participantes con una orientación hacia los objetivos y una coordinación difusa, que interactúan para crear un servicio que se hace disponible para la comunidad. En esta cooperación, el objetivo general está claro, facilitar la difusión en abierto de resultados de investigación. No obstante, los específicos se dejan a la iniciativa e innovación de los participantes. Por su parte, la necesidad de coordinación es prácticamente inexistente. Cada participante sabe el rol que juega en el sistema y lo lleva a cabo de forma independiente. La necesidad de comunicación y colaboración con otros miembros es reducida. Finalmente, no existe ningún gestor ni administrador.
- Datos abiertos. Todos los datos generados en el seno de RePEc son distribuidos como Open Data, cumpliendo los principios FAIR (Findable, Accessible, Interoperable, Reusable)⁹. Cualquiera puede implementar un nuevo agregador a partir de los datos existentes. El disponer de datos abiertos puede ocasionar una cierta competencia entre agregadores para proporcionar más y mejores servicios.
- Hacer las cosas fáciles. Conseguir involucrar a muchos participantes, como se ha logrado en RePEc, pasa por cumplir dos requisitos: hacerles la vida fácil y proporcionar algún beneficio concreto e inmediato. Si se ha conseguido incorporar a tantos proveedores de datos es porque no se pide ninguna infraestructura técnica ni conocimientos específicos. Por comparación con la implementación del protocolo OAI-PMH que requiere de aplicaciones costosas, nosotros nos basamos en un protocolo de comunicación que no requiere ninguna instalación especial más allá de tener acceso a un servidor FTP o HTTP y en un esquema de metadatos donde prácticamente todos los elementos son opcionales. Si bien, cuanto más información aporte el proveedor, mayor visibilidad obtendrá debido a la multiplicación de puntos de acceso que supone cada nuevo elemento añadido. La visibilidad es el beneficio clave que obtienen los proveedores de datos.
- Comunidad virtual definida como una agregación de personas que interactúan alrededor de un interés común utilizando medios tecnológicos guiados por algunos protocolos y normas. En nuestro caso, el interés que nos une es la difusión en abierto de los resultados

⁹ <https://www.go-fair.org/fair-principles/>

de investigación en Economía. La comunicación se hace exclusivamente por medios electrónicos utilizando los servicios básicos de internet. El canal de comunicación habitual son los foros a través de listas de correo electrónico. Los miembros, distribuidos por todo el mundo, únicamente nos hemos reunido de forma presencial dos veces a lo largo de estos casi 25 años. La última vez para celebrar los 20 años de RePEc en una conferencia que tuvo lugar en el Federal Reserve Bank of St Louis¹⁰ (USA).

Una pregunta recurrente cuando hacemos alguna presentación sobre RePEc es, ¿Cuál es su garantía de sostenibilidad? Y la respuesta es siempre la misma, se trata de una comunidad virtual de cooperación que será sostenible mientras la comunidad la necesite. A lo largo de estos 24 años, hemos asistido al nacimiento y muerte de innumerables proyectos en el ámbito del acceso abierto, con generosa financiación pública, y grandes planes de continuidad que desaparecieron sin dejar rastro cuando se acabó la financiación. Proyectos que se perpetúan en el tiempo, en un círculo vicioso en el que crean necesidades a las que poder dar respuesta, con un retorno ínfimo en la sociedad. La diferencia con RePEc radica en que no consiguieron hacerse necesarios para la comunidad.

Por supuesto, no quiero decir que la financiación no sea importante, obviamente lo es. Lo que quiero decir es que la sostenibilidad, que es el gran problema de los proyectos con financiación pública cuando se acaba la subvención, solamente estará garantizada cuando un proyecto de soluciones de verdad a problemas concretos, en definitiva, sea útil. Obvio también.

Creo que, desde las bibliotecas universitarias, en unos momentos de indefinición de nuestra función, podemos hacer muchas cosas útiles. Debemos poner en valor la fuerza de trabajo conjunta que tenemos para llevar a cabo proyectos colaborativos exitosos e innovadores. Los repositorios y agregadores, las humanidades digitales y la digitalización de nuestro patrimonio son oportunidades que están esperando nuestras ideas.

En las bibliotecas universitarias españolas ya se ha demostrado que es posible. Tenemos ejemplos que comparten con RePEc muchos puntos en común (descentralización, buenos valores añadidos, etc.). Me refiero a Dialnet, a mi juicio, el servicio más innovador y exitoso de cooperación bibliotecaria en nuestro país, con una gran repercusión en la comunidad investigadora y que coincide con RePEc en la implementación de servicios como perfiles de investigadores, deduplicación de autores o métricas de calidad.

Finalmente, cuando hablamos de proyectos digitales, quizás la debilidad mayor que tenemos como colectivo profesional es la escasa formación tecnológica. Una base sólida en TIC es necesaria para poner en práctica nuevos servicios sin necesidad de depender de personal informático. Volviendo a la frase con la que empezaba, ¿Cuántos de nuestros recién titulados podrían contestar afirmativamente a la misma pregunta hoy en día? Mucho me temo que la respuesta sería la misma, necesitaremos aprender. El Consejo de Cooperación Bibliotecaria¹¹ recoge, en sus perfiles profesionales, la importancia de las competencias profesionales en diseño informático de sistemas de información y en desarrollo de aplicaciones. Estas competencias necesitan ser potenciadas en nuestras titulaciones, necesitan incluirse en los temarios de oposiciones para el acceso a las bibliotecas universitarias, deben estar presentes en la formación

¹⁰ <https://blog.repec.org/2017/10/18/st-louis-fed-hosted-repec-workshop/>

¹¹ https://www.libreria.culturaydeporte.gob.es/libro/perfiles-profesionales-del-sistema-bibliotecario-espanol-fichas-de-caracterizacion_1325/

continua y lo que es más importante, los profesionales tenemos que verlas como algo propio y fundamental si queremos seguir siendo útiles y necesarios en la universidad del siglo XXI.



Figura1: 20 años de RePEc. Octubre 2017. De pie, Christian Mongeau, Jose Manuel Barrueco, Bob Parks, Kit Baum, Jan Höffler. Sentados, Douglas Hanley, Thomas Krichel, Areerat Kichkha, Soledad Signago, Mahdi Moqri, Christian Zimmermann.

Sobre el autor



JOSÉ MANUEL BARRUECO

Bibliotecario en Universitat de València. Servei de Biblioteques i Documentació

Licenciado en Documentación por la Universidad Politécnica de Valencia, trabajo en el servicio de bibliotecas de la Universitat de València desde 1993. Mi puesto actual es coordinador de proyectos de acceso abierto, en concreto, gestión del repositorio institucional y la plataforma de edición de revistas OJS. Desde 1997 colaboro en el desarrollo de la biblioteca digital internacional RePEc (Research Papers in Economics) donde me encargo del portal de métricas CitEc (Citations in Economics).

- **Correo electrónico:** Jose.Barrueco@uv.es
- **Twitter:** @JMBarrueco