

Distribución de publicaciones a través de internet

el modelo RePEc

11 de febrero de 2000

José Manuel Barrueco Cruz
Biblioteca Gregori Maians
Universitat de València
46022 València
Spain

Jose.Barrueco@uv.es

<http://www.uv.es/~barrueco>

Thomas Krichel
Department of Economics
University of Surrey
Guildford GU2 5XH
United Kingdom

T.Krichel@surrey.ac.uk

<http://gretel.econ.surrey.ac.uk>

Resumen

En este trabajo hacemos una descripción de RePEc (Research Papers in Economics). Analizamos la historia, fundamentos y proyectos futuros de la que es la segunda mayor biblioteca digital de acceso gratuito a nivel mundial dedicada a proporcionar resultados de investigación.

El trabajo descrito en esta comunicación fue financiado parcialmente por el Joint Information Systems Committee en el Reino Unido a través de su Electronic Libraries Programme. Este documento está accesible electrónicamente en la dirección: <http://www.uv.es/~barrueco/tnt.pdf>.

©2000 José Manuel Barrueco Cruz y Thomas Krichel

1 ¿De dónde venimos?

Los orígenes de RePEc hemos de buscarlos en NetEc. NetEc es fue creado por Thomas Krichel en febrero de 1993 como un grupo de "proyectos" destinados a facilitar la comunicación de resultados de investigación en Economía por medios electrónicos y paralelos a los canales comerciales tradicionales.

Inicialmente estuvo basado en el servidor gopher del Manchester Computing Center. En 1994 migró a un sistema web. 1995 fue el año de la expansión mundial al inaugurarse dos *mirrors*, uno en Estados Unidos, en la Universidad de Washington en San Luis, cortesía de Bob Parks y otro en Japón, en la Hitotsubashi University, cortesía de Satoshi Yasuda.

Todos los proyectos que forman parte de NetEc comparten dos características: el acceso es gratuito y están mantenidos por voluntarios. Los dos proyectos que marcaron el inicio de NetEc fueron BibEc y WoPEc. En ambos podemos establecer dos etapas: antes y después de 1997.

- **BibEc**, mantenido por Thomas Krichel y Fethy Mili, se centró en la recogida de datos bibliográficos sobre documentos de trabajo impresos. Los datos procedían, en la primera fase, de las listas de novedades elaboradas por Fethy en el centro de documentación del departamento de Economía de la Universidad de Montreal. Estos datos eran convertidos a páginas *html* estáticas en Manchester e indizados usando una base de datos *wais*. La cobertura de los fondos llegaba hasta comienzos de los años ochenta. Hasta 1997 BibEc recopiló unos 40.000 registros.
- **WoPEc** mantenido por Jose Manuel Barrueco y Thomas Krichel nació como una colección centralizada de referencias bibliográficas de working papers cuyo texto completo se hallaba disponible en la red en formato electrónico. Así funcionó desde 1993 a 1997 y llegó a alcanzar los 5.000 documentos al final de este periodo. 1996 fue un año clave para WoPEc ya que recibió fondos del Joint Information Systems Committe a través del Electronic Libraries Programme.

Otros proyectos incluidos en NetEc durante estos años fueron:

- **CodEc** es una colección de software de libre distribución y de interés para economistas. Está mantenido por Dirk Eddelbüttel.

- **WebEc** es una colección de enlaces a recursos sobre Economía disponibles en internet. Está mantenido por Lauri Saarinen.
- **JokEc** es el más lúdico de los proyectos ya que recopila chistes sobre economistas. Está mantenido por Pasi Kuoppomäki.

Paralelamente al trabajo de NetEc existían, dentro de la disciplina de Economía, otros proyectos e iniciativas tendentes a lograr el mismo objetivo: facilitar la comunicación de resultados de investigación. Así podemos citar la de Bob Parks, quien en 1993 fundó el Economics Working Papers Archive (EconWPA) siguiendo el ejemplo del exitoso servidor de Física creado por Paul Ginsparg. Este archivo pretendía ser un sitio central donde investigadores de todo el mundo pudieran remitir sus documentos para ser distribuidos. Por otro lado, cada vez más departamentos estaban colocando en sus servidores versiones electrónicas de los documentos que antes distribuían en formato impreso.

2 ¡La idea!

En definitiva, antes de 1997, nos encontramos con un grupo cada vez más amplio de iniciativas centralizadas que intentaban alcanzar un objetivo común pero sin ninguna coordinación entre ellas. BibEc y WoPEc fueron conscientes muy pronto de la imposibilidad de mantener una colección de las dimensiones que estaba alcanzando (45.000 registros) de forma centralizada y voluntaria.

En este contexto podemos situar la "idea" de Bill Goffe que constituye el embrión de RePEc. Bill Goffe es un economista de reconocido prestigio y autor de una famosa recopilación de recursos electrónicos de interés para economistas. En 1995 escribió a la lista de discusión NetEc-Admin:

... What I would suggest is this: a distributed system with any number of sites, each mirroring each other. [...] archives could "join" the system (say it was written in perl so could run on NT as well as Unix). Then you'd have the best of both worlds [...] Such a system could easily grow with the profession's use of the net. Such a system would GREATLY benefit the profession...

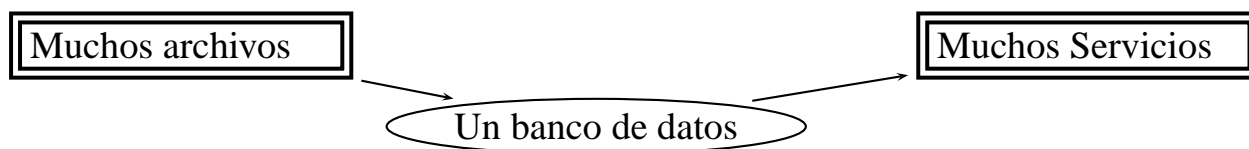
Lo que estaba proponiendo aquí era un sistema descentralizado similar al de las *news*. Con esta idea en mente, es decir, encontrar una forma de descentralizar

la recolección y distribución de datos bibliográficos sobre documentos, WoPEc convocó una reunión en Guildford en mayo de 1997 a la que asistieron representantes de dos proyectos colaboradores: DEGREE de los Países Bajos, es una biblioteca electrónica que almacena todos los working papers distribuidos por universidades de estos países y S-WoPEc, su homónimo en Suecia.

A esta reunión se la puede considerar el nacimiento de RePEc. En ella se discutieron dos propuestas que constituyen el corazón de RePEc: el protocolo de Guildford y ReDIF.

3 ¿Quiénes somos?

El modelo descentralizado para la distribución de documentos que preconiza RePEc se fundamenta en el siguiente principio:



Podemos asimilar los archivos de una forma simple a instituciones, departamentos, etc. que ofrecen información bibliográfica y, en algunos casos, el texto completo de los documentos que producen. Todos esos datos procedentes de decenas de archivos forman un único banco de datos, distribuido en decenas de máquinas alrededor del mundo. Cualquier persona o institución puede acceder a estos datos sin restricciones, dado que están ubicados en servidores *HTTP* o *FTP anónimo*, y siguen una estructura marcada por el protocolo de Guildford. Cualquiera puede tomarlos y crear unos servicios utilizables por los usuarios finales, tales como presentaciones a través de páginas *html*, listas de novedades o selección de documentos de determinadas áreas temáticas. De esta forma los usuarios pueden acceder a los datos a través de diferentes interfaces. Los archivos ofrecen los datos a todos los servicios al mismo tiempo lo que ocasiona una óptima distribución.

En Febrero de 2000 hay más de cien archivos que ofrecen datos sobre sus publicaciones. Algunos pertenecen a instituciones prestigiosas como la OCDE, el National Bureau of Economic Research (NBER), el Center for Economic Policy Research (CEPR) o los bancos federales de Estados Unidos (Fed in Print). No

obstante, la mayoría proceden de pequeñas instituciones (University of Surrey o Universitat de Valencia). También hay algunos datos, pocos, procedentes de editoriales comerciales como Springer Verlag.

Todas estas instituciones colaboran para ofrecer un banco de datos que alcanza más de 80.000 referencias bibliográficas. La tipología documental abarca documentos de trabajo, artículos publicados en revistas y software de dominio público. De esas 80.000 referencias, 22.000 hacen referencia a documentos cuyo texto completo está accesible de forma gratuita en internet.

3.1 ¿Cómo se organizan los archivos?

RePEc se asienta en dos pilares: un formato para el intercambio de información bibliográfica a través de internet denominado ReDIF y una serie de normas, recogidas en el protocolo de Guildford, que describen la forma de almacenar y compartir esa información a través de servidores *FTP* o *HTTP*.

ReDIF ofrece distintos tipos de registros dependiendo del documento que se esté describiendo como podemos comprobar en la figura 1

<i>Nombre</i>	<i>Version</i>	<i>Usado para</i>
ReDIF-Paper	1.0	working papers
ReDIF-Article	1.0	artículos en una revista
ReDIF-Chapter	1.0	capítulos de libros
ReDIF-Series	1.0	colecciones de documentos
ReDIF-Archive	1.0	archivos
ReDIF-Mirror	1.0	mirrors de archivos

Figura 1: Tipos de registros en ReDIF

Los registros bibliográficos se almacenan en ficheros ASCII. Se componen de un conjunto de campos que siguen la estructura tradicional de una base de datos de "*campo: valor*". Cada registro debe ocupar un párrafo de texto y debe comenzar con un campo obligatorio: **Template-Type**: cuyo contenido será alguno de los tipos enunciados más arriba.

El resto de los campos no importa el orden en que aparezcan. Ahora bien, ReDIF utiliza el concepto de **clusters**, por lo que los datos referidos a un mismo objeto (autor, lugar de trabajo) deben aparecer juntos.

Mención especial merece el **handle**, ya que es un campo obligatorio cuyo contenido es un código que identifica unívocamente a cada objeto descrito en RePEc.

Esta es la característica más importante del formato ReDIF, ya que una simple cadena de caracteres nos proporciona información sobre el archivo, la serie y el número de un objeto determinado. Por otro lado, a través del handle podemos crear una estructura relacional que enlace los distintos elementos que forman la descripción de un documento. Ejemplos de handles son:

Handle: RePEc:bon:bonnsf:a452
 Handle: RePEc:die:calsdi:9338

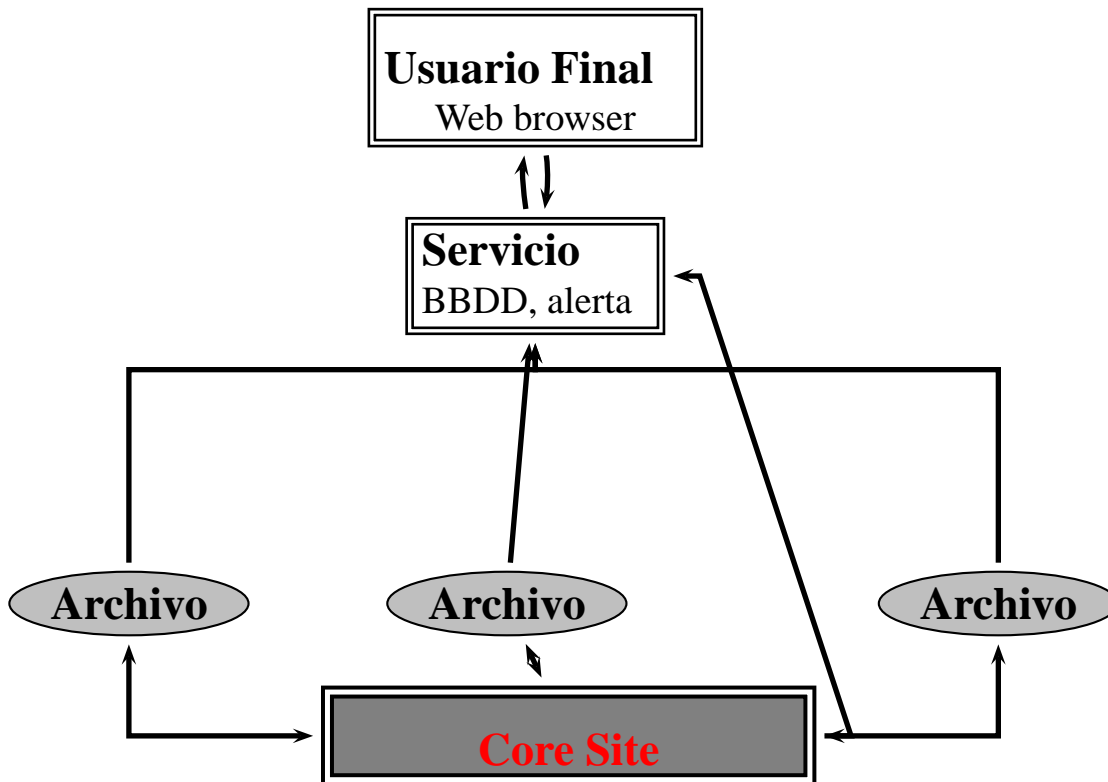
Un registro ReDIF completo quedaría como sigue:

Template-Type: ReDIF-Paper 1.0
 Title: A Discussion of the Reliability of Results Obtained with Long-Run Identifying Restrictions
 Author-Name: St-Amant, P.
 Author-Email: St-Amant@uv.es
 Keywords: CENTRAL BANKS ; MONETARY POLICY ; MACROECONOMICS
 Length: 14 pages
 Classification-JEL: C1 ; E27
 Creation-Date: 1998
 File-URL: <http://www.nber.org/papers/w6490.pdf>
 File-Format: application/pdf
 File-Restriction: Access to the full text is restricted. Look up <http://www.nber.org/wwpfaq.html> for details, or write to feenberg@nber.org. If you have no access to the full text you will be shown an abstract page instead. Anyone browsing the NBER working paper database from a site with a TLD in a non-OECD, non-OPEC country will be offered full text downloads for any paper for which the full text is available online, but only if their DNS does reverse name lookup.
 Handle: RePEc:nbr:nberwo:6490
 Price: \$5.00 per paper (plus \$10.00 postage for orders outside U.S.)

Este registro ilustra cómo la colección no está necesariamente limitada a documentos gratuitos. Dado que las condiciones de acceso a estos documentos de pago varían el texto del campo **File-Restriction** es libre.

El Protocolo de Guildford proporciona un conjunto de reglas para la publicación e intercambio de documentos y metadatos en internet. Establece dos niveles para la participación de las instituciones, el de **archivo**, de carácter pasivo, simplemente proporciona información, y el de **servicio**, activo, ya que toma la información de los anteriores y construye un "servicio" de utilidad para los usuarios finales.

Gráficamente este modelo lo vemos representado:



Como vemos en el gráfico, el esquema RePEc está compuesto básicamente por el **core site**, quien recoge y suministra a los archivos y servicios información básica sobre la localización del resto de participantes (flechas intermitentes), los archivos, encargados de almacenar la información bibliográfica, los servicios, que reciben los datos de los archivos y crean una presentación de los mismos para los usuarios finales, y los propios usuarios, que acceden al sistema a través de los servicios.

3.2 Los servicios

Servicios que actualmente están implementando este modelo son:

- **BibEc** y **WoPEc**, han continuado con su papel de recogida de documentos en formato impreso y electrónico respectivamente. Aunque mantienen su independencia pueden consultarse a través de un interface común que usa una base de datos SQL.
- **IDEAS**, gestionado por Christian Zimmermann en la Universidad de Quebec en Montreal, Ofrece información sobre todos los datos existentes en RePEc, tanto electrónicos como impresos. Permite hojear las documentos por series o buscar en una base de datos basada en Excite.

- **DECOMATE**. Basado en los Países Bajos. Ofrecen una pasarela Z39.50 a los datos contenidos en RePEc.
- **NEP** (New Economic Papers) Gestionado por Jhon S. Irons. Está compuesto por una serie de listas temáticas de novedades donde semanalmente se anuncian los nuevos documentos incluidos en RePEc. Cada lista está editada por un investigador de prestigio en la materia tratada.
- **RuPEc**. Servicio basado en Rusia. Ofrece información sobre todos los datos en RePEc en ruso.
- **HoPEc**. Es un nuevo servicio encargado que trata de unir la información bibliográfica existente en RePEc con información relativa a los autores.

4 ¿Qué es lo que no somos?

No todos los posibles tipos documentales están cubiertos por ReDIF. Este formato sólo permite, por el momento, describir documentos de trabajo, artículos, software, personas e instituciones, quedando fuera, por ejemplo, las series estadísticas o los libros.

No existe una extrategia global de conservacion de los documentos. Estos son almacenados por las instituciones que los generan y es su responsabilidad conservarlos. Si en algún momento deciden retirarlos no existe la posibilidad, por parte de los servicios, de mantenerlos en forma local. Una vez que el documento o su descripción han sido retirados del archivo que los generó lo serán también de todos los servicios.

No hay estadísticas globales del uso del banco de datos. Algo que sería muy difícil de conseguir, dado que RePEc se centra en los archivos más que en los servicios.

No hay una política de revisión de los contenidos almacenados en los archivos. Básicamente cada uno puede distribuir lo que crea conveniente.

5 ¿Hacia dónde vamos?

En conclusión RePEc es un esfuerzo de colaboración para poner a disposición de la comunidad internacional, de forma gratuita, un corpus de conocimiento

en formato electrónico. Esta última característica tiene una importancia considerable para los países en vías de desarrollo, donde los elevados costes de suscripción dificultan el acceso a las revistas tradicionales, tanto impresas como electrónicas.

Una vez que se ha conseguido una masa crítica de documentos que permite poner de manifiesto las ventajas del sistema, son cada vez más las instituciones que colaboran con RePEc. Sería un efecto de bola de nieve.

Futuras implementaciones en el mundo RePEc serán:

- Introducir un servicio de índice de citas entre los documentos distribuidos.
- Extender el modelo RePEc a otras disciplinas. Concretamente se está trabajando en tres áreas: ReSoS sería la implementación del protocolo de Guildford y ReDIF en el área de ciencias sociales, ReLIS para la Documentación y ReCMAP para la Física y las Matemáticas. De estos dos últimos casos ya existen dos servicios en una fase de experimentación: DoIS (Documents in Information Science) y BigBang.
- Introducir una versión avanzada de ReDIF que sea independiente de la sintaxis y orientada a objetos.

6 Direcciones a tener en cuenta

RePEc	http://netec.mcc.ac.uk/RePEc
NetEc	http://netec.mcc.ac.uk
BibEc	http://netec.mcc.ac.uk/BibEc
WoPEc	http://netec.mcc.ac.uk/WoPEc
CodEc	http://netec.mcc.ac.uk/CodEc
WebEc	http://netec.mcc.ac.uk/WebEc
JokEc	http://netec.mcc.ac.uk/JokEc
EconWPA	http://wuecon.wustl.edu
ReDIF	http://netec.mcc.ac.uk/RePEc/docu/ReDIF.html
Protocolo de Guildford	http://netec.mcc.ac.uk/RePEc/GuilP.html
IDEAS	http://ideas.uqam.ca
DECOMATE	http://www.kub.nl/dbi/demomate/repref.htm
NEP	http://netec.mcc.ac.uk/NEP
RuPEc	http://www.ieie.nsc.ru/RuPEc
DoIS	http://dois.mcc.ac.uk/DoIS
BigBang	http://bigbang.wustl.edu