

Abrir y enlazar datos: la reutilización de información *de, por y para* las bibliotecas

Gema Bueno de la Fuente, Universidad Carlos III de Madrid (e: gbueno@bib.uc3m.es / tt: @gema_bueno)

La reutilización de la información pública: hacia una economía del conocimiento

La reutilización de la información en el sector público (RISP), entendida como “*el uso por parte de personas físicas o jurídicas, de información generada por organismos del sector público, con fines comerciales o no*”¹ es algo latente y vinculado a la transparencia y a la noción de “lo público”. Sin embargo, esta reutilización se ha potenciado o propiciado por el desarrollo tecnológico, no sólo en cuanto a las posibilidades de acceso e intercambio de información que ofrecen las TIC, sino especialmente, por su evolución en las capacidades de análisis, explotación y tratamiento de datos, que permite la creación de servicios y aplicaciones basados en el uso, la agregación o la combinación de datos.

Esta reutilización es una tónica común a todos los Estados Miembros de la Unión, promovida desde las administraciones públicas a instancia del propio Parlamento Europeo y del Consejo², con su Directiva 2003/98/CE³. En España, se ha plasmado en la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público⁴, que se desarrolla en el Real Decreto 1495/2011⁵, y cuenta con un fuerte apoyo en la promoción, difusión y sensibilización del Proyecto Aporta (ahora *datos.gob.es*)⁶ del Plan Avanza, continuado por la Estrategia 2011-2015 del Plan Avanza 2⁷. Países de todo el mundo están llevando a cabo iniciativas similares, como es el caso de Estados Unidos con su *Open Government Directive*, de 2009⁸ y su portal *data.gov*.

¹ Proyecto Aporta (2009). *Guía Aporta sobre reutilización de la información del sector público*. Madrid: Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información. Disponible en:

² *Public Sector Information - Raw Data for New Services and Products*. Comisión Europea, Sociedad de la Información (sitio web). Disponible en: http://ec.europa.eu/information_society/policy/psi/index_en.htm

³ Parlamento Europeo, Consejo de la Unión Europea (2003). Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público. *Diario Oficial*, n. L 345, 31/12/2003 p. 0090 – 0096. Disponible en: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:ES:HTML>

⁴ Jefatura del Estado (España) (2007). Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. *Boletín Oficial del Estado*, n. 276, de 17 de noviembre de 2007, pp. 47160 a 47165. Disponible en: <http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>.

⁵ Ministerio de la Presidencia (España) (2011). Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal. *Boletín Oficial del Estado*, núm. 269 de 8 de noviembre de 2011, pp. 116296-116307. Disponible en: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560

⁶ *De Proyecto Aporta (2009) a datos.gob.es*. Blog del portal *datos.gob.es*, 30 de julio de 2012. Disponible en <http://datos.gob.es/datos/?q=node/1836>.

⁷ Ministerio de Industria, Turismo y Comercio, Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (España) (2010). *Estrategia 2011-2015 del Plan Avanza 2*. Disponible en: https://www.planavanza.es/InformacionGeneral/Estrategia2011/Documents/Estrategia_2011-2015_PA2.pdf

⁸ Peter R. Orszag (2009). *Open Government Directive. Memorandum for the Heads of Executive Departments and Agencies*. Disponible en: <http://www.whitehouse.gov/open/documents/open-government-directive>

La Directiva 2003/98/CE pretende catalizar un cambio de cultura a favor de la reutilización de la gran variedad y cantidad de información, fiable y de calidad, que producen y gestionan los organismos públicos. Con ello se persigue responder al compromiso de transparencia de la administración con los ciudadanos, al tiempo que se posibilita e incentiva la explotación comercial de esta información, allanando el camino a una economía del conocimiento basada en los contenidos digitales. Para ello, es preciso que los datos se desbloqueen y estén disponibles para su reutilización.

¿Qué papel juegan las bibliotecas en la reutilización de la ISP?

La Directiva 2003/98/CE resalta la relación de la RISP con otras políticas y normas de la Unión Europea, entre las cuales las dos últimas están directamente ligadas con las funciones de buena parte de las bibliotecas: por un lado, con las iniciativas de acceso abierto a la información científica, incidiendo directamente en aquellas bibliotecas responsables de la gestión de este tipo de información (principalmente universidades, centros de investigación...); y por otro, con las políticas de digitalización y patrimonio cultural, que afectan especialmente a las bibliotecas de carácter patrimonial pero también a muchas otras.

A estos dos aspectos destacados por la directiva, cabría añadir la vertiente educativa de las bibliotecas, principalmente universitarias y escolares, que en la línea de la apertura y la reutilización enlaza directamente con el movimiento a favor de los recursos educativos abiertos (REA) u *open educational resources* (OER), término acuñado hace ya más de una década por la UNESCO⁹. A pesar de que no exista aún una política a nivel europeo o nacional en este sentido, la Declaración de París de 2012¹⁰, insta a los gobiernos a la publicación mediante licencias abiertas de los recursos educativos financiados con fondos públicos. Se espera que a corto plazo los gobiernos de distintos países se adhieran a la declaración, y con ello, que los recursos educativos se contemplen en las políticas europeas y nacionales de reutilización y apertura de la información. La Comisión Europea está realizando ya una consulta¹¹ para constituir una iniciativa europea de promoción de la apertura de la educación, el uso de las tecnologías y los recursos educativos abiertos, anticipando posibles acciones como la definición de un marco legal común o el establecimiento de un mandato de publicación en abierto de aquéllos REA financiados con fondos públicos.

⁹ UNESCO (2002). *Forum on the Impact of Open Courseware for Higher Education in Developing Countries. Final report*. Paris, 1-3 julio 2002. Disponible en: http://portal.unesco.org/ci/en/files/2492/10330567404OCW_forum_report_final_draft.doc/OCW_forum_report_final_draft.doc

¹⁰ UNESCO (2012). *Declaración de París de 2012 sobre los REA. Congreso Mundial sobre los Recursos Educativos Abiertos (REA) UNESCO*, París, 20-22 de junio de 2012. Disponible en: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Spanish_Paris_OER_Declaration.pdf

¹¹ Comisión Europea, Directorado General de Educación y Cultura (2012). *Public Consultation on "Opening up Education – a proposal for a European Initiative to enhance education and skills development through new technologies"*. Disponible en: http://ec.europa.eu/dgs/education_culture/consult/open_en.htm

Más allá de estas vinculaciones con las funciones y actividades de las bibliotecas, la propuesta de revisión de la directiva, publicada en 2012¹², llena el enorme vacío que significa la exclusión de las instituciones de la memoria, y amplía su ámbito de aplicación a las bibliotecas (incluidas las universitarias), archivos y museos. Reconoce así que los materiales de archivo, colecciones de libros, y obras de arte constituyen también parte del patrimonio de información y contenidos de los organismos públicos. Con esta revisión, entran en juego un gran número de instituciones públicas, y con ellas, un enorme volumen de información difícil de calcular. El directorio de las bibliotecas de titularidad estatal¹³ registra más de diez mil bibliotecas, entre bibliotecas públicas, Nacionales y Centrales de CCAA, dependientes de instituciones de enseñanza superior, y bibliotecas especializadas, en las que tienen cabida las bibliotecas de centros de investigación, de empresas, de archivos y museos, de hospitales y otros centros sanitarios, o de los propios organismos de la Administración. La complejidad y riqueza del universo de bibliotecas con un compromiso público representa una gran diversidad y tipología de colecciones y de datos potencialmente reutilizables¹⁴.

Si bien la Directiva sólo menciona explícitamente los libros, y la Ley 37/2007 se refiere genéricamente a los documentos, entendido como *“toda información cualquiera que sea su soporte, ya sea escrito en papel, almacenado en forma electrónica o como grabación sonora, visual o audiovisual”*, las bibliotecas también generan datos y productos de información propios potencialmente reutilizables. Así, además de sus propias colecciones y objetos digitales, algunas de las áreas en las que las bibliotecas pueden aportar datos para la reutilización, incluyen datos bibliográficos, datos de registros de autoridades, vocabularios, archivos y datos heterogéneos, citación, usos sociales y usos innovadores¹⁵.

Aunque desde hace ya décadas se viene hablando de un cambio de paradigma para las bibliotecas, de la tradición en papel a la biblioteca digital, del átomo al bit, ahora más que nunca esta transición se está haciendo realidad, especialmente en lo que afecta a las bibliotecas académicas. Como afirma Stefan Gradmann en su ponencia *“From Containers, to Content, to Context”*¹⁶, las bibliotecas deben pasar de ser contenedoras a contextualizadoras de la información. Es una oportunidad que no pueden perder si no quieren correr el riesgo de volverse insignificantes. Gradmann apunta las claves de esta transición comenzando con la necesidad de cambiar la terminología (ver Figura 1).

¹² Comisión Europea (2011.) *Propuesta de Directiva del Parlamento y del Consejo por la que se modifica la Directiva 2003/98/CE relativa a la reutilización del sector público*. Cod. 2011/0430. [Bruselas]: Comisión Europea, 2011. Disponible en: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/es.pdf

¹³ *Directorio de las Bibliotecas españolas* (sitio web). Ministerio de Educación, Cultura y Deporte. Secretaría de Estado de Cultura. Disponible en: <http://directoriobibliotecas.mcu.es>

¹⁴ Cabe destacar, no obstante, que no todos los fondos de las bibliotecas pueden ser abiertos y reutilizados, ya que la legislación no aplica a aquéllos documentos y datos con derechos de propiedad intelectual, aunque estén en la biblioteca.

¹⁵ Estos son algunos de los casos de uso analizados por el Library Linked Data Incubator Group del W3C, traducido al español: Vila Suero, D. (trad.) (2011). *Grupo Incubador de Datos Vinculados de Bibliotecas: Casos de uso. Informe de Grupo Incubador del W3C de 25 de octubre de 2011*. Disponible en: <http://www.larramendi.es/LAM/Incubador/llid/XGR-llid-usecase-20111025.html>

¹⁶ Stefan Gradmann (2012). *From Containers to Content to Context: the Changing Role of Libraries in eScience and eScholarship*. *Bielefeld Conference*, April 24 2012. Disponible en: http://conference.ub.uni-bielefeld.de/programme/presentations/Gradmann_BC2012.pdf

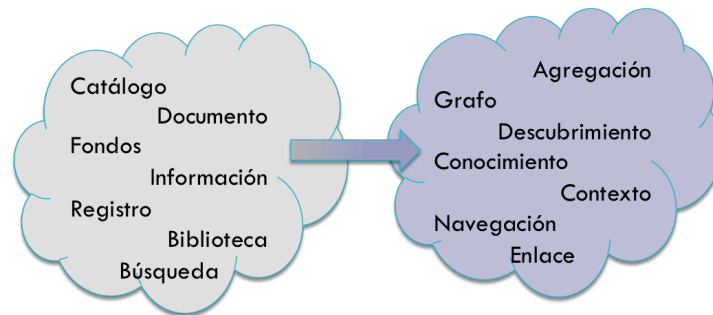


Figura 1. Cambio de terminología en las bibliotecas (trad. de Stefan Gradmann, 2012, p. 41)

En este contexto, la contribución de las bibliotecas a la reutilización de la información se hace aún más evidente y necesaria. Además de ofrecer sus contenidos y sus datos en abierto para la reutilización por terceros, las bibliotecas pueden, y deben, jugar un papel cada vez más activo y asumir por sí mismas los roles de consumidoras, reutilizadoras y de *infomediarias*. Esta denominación, que se atribuye a las empresas explotadoras de la información pública, define la esencia de la función de los bibliotecarios y profesionales de la información, pues constituyen el nexo entre la información, el dato, y los usuarios, mediando en el proceso de comunicación y transferencia de la información. Una vez más, es un reto y una oportunidad que no deberían dejar escapar.

A modo de síntesis, apuntamos los aspectos en los que las bibliotecas se pueden vincular a la reutilización de la información en el sector público:

- Publicación y apertura de sus documentos y sus datos en formato digital, preferiblemente como datos abiertos y enlazados.
- Registro y mantenimiento de catálogos de conjuntos de datos disponibles (selección, organización, descripción, difusión, monitorización). Este registro puede realizarse con un enfoque *bottom-up*, desde lo local a lo internacional, e incluso, y servir a modo de depósito legal que identifique los conjuntos de datos abiertos publicados.
- Creación de servicios de información de valor añadido a partir de la reutilización de fuentes de datos propias y ajenas, seleccionadas y dirigidas a sus respectivas comunidades de usuarios.
- Asesoría en la publicación de datos en abiertos en cuanto a licencias, formatos, estándares, etc.
- Participación en el desarrollo de estándares, recomendaciones y guías buenas prácticas para la publicación, identificación, descripción, difusión, localización, o citación de ISP.
- Difusión, concienciación y formación sobre la existencia, derechos y posibilidades de acceso y reutilización a la ISP, dirigida a los ciudadanos, entidades y otros colectivos.
- Preservación digital de conjuntos de datos propios, y de los organismos de los que dependa la biblioteca, entre otros.

¿Cómo deben publicarse los datos para que sean realmente reutilizables?

Siguiendo lo expuesto, no cabe duda de que las bibliotecas deben abrir sus contenidos y sus datos, y que pueden contribuir a apoyar y asesorar a otras iniciativas de apertura, así como ser reutilizadoras en sí mismas. Teniendo en cuenta la legislación en torno a la RISP, tres son las condiciones fundamentales para los datos puedan ser reutilizados, y es que sean publicados:

- i. en la web, en acceso libre y abierto,
- ii. bajo una licencia abierta que permita su reutilización incluso con fines comerciales, con las mínimas restricciones posibles pero que exija el reconocimiento de la fuente a los reutilizadores,
- iii. y mediante formatos legibles por máquina junto con sus metadatos, de forma que garanticen su interoperabilidad.

Estos principios son, inicialmente, compartidos con los requisitos de los datos abiertos enlazados o *Linked Open Data* (LOD), tal y como los define Tim Berners-Lee en su esquema incremental de cinco estrellas¹⁷. Añade, además, que los datos: estén estructurados, en formatos no propietarios, conforme a estándares del W3C, y enlazados con otros conjuntos de datos. Estos requisitos, se explicitan de forma más técnica como los Principios de *Linked Data*¹⁸:

1. Usar URIs como nombres para las cosas.
2. Usar URIs HTTP, de modo que la gente pueda buscar y localizar esos nombres.
3. Cuando alguien busque empleando un URI, proporcionar información útil, utilizando estándares (RDF, SPARQL).
4. Incluir enlaces a otros URIs, para que se puedan descubrir y localizar más cosas.

Si realmente se pretende llegar a una reutilización y explotación eficiente de los datos públicos el objetivo debería ser cumplir con los principios LOD más allá de la mera publicación de contenidos, en formatos propietarios no preparados ni pensados para la reutilización.

Para las bibliotecas, el camino de la RISP no debe enfocarse únicamente en la explotación comercial, en la construcción de una *economía del conocimiento*, sino más bien en el desarrollo de la *Web del conocimiento*: la Web Semántica. Una web, que como la define Eva Méndez¹⁹, es, ni más ni menos, una web más bibliotecaria.

¹⁷ Tim Berners-Lee (2006). *Linked Data*. Disponible en: <http://www.w3.org/DesignIssues/LinkedData.html>

¹⁸ Tom Heath, and Christian Bizer (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool. Disponible en: <http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>.

¹⁹ Eva M^a Méndez Rodríguez. La Web Semántica: una web "más bibliotecaria". *Boletín CLIP*, 41. Disponible en: http://www.sedic.es/p_boletinclip41_confirma.htm.

¿Qué están haciendo las bibliotecas a nivel nacional e internacional?

A pesar de la reciente incorporación de las bibliotecas en las normas, iniciativas y estudios de RISP en Europa y España, y de la incipiente adopción y reconocimiento de la importancia de crear y publicar LOD en el ámbito de las bibliotecas (el *W3C Library Linked Data Incubator Group* se constituyó en mayo de 2010)²⁰, son numerosas las iniciativas que se pueden identificar en ambas.

El Catálogo de Información Pública de la Administración General del Estado, en el portal *datos.gob.es*²¹, recoge una alta proporción de proveedores del ámbito bibliotecario, archivístico y documental, incluyendo principalmente bibliotecas, pero también hemerotecas, mediatecas, centros de documentación, videotecas, fototecas o cartotecas. Como productos de información, abundan los catálogos de toda índole: de bibliotecas de diversas instituciones (Congreso, Senado, ministerios, institutos, centros de estudios y de investigación, museos); colectivos como el del Patrimonio Bibliográfico o los de redes bibliotecarias como REBIUN, REBECA, CSIC o CEDEX; bibliográficos como el de la BNE o el CSIC; de autoridades; de bibliotecas digitales; así como otros específicos de monografías, revistas, publicaciones, últimas adquisiciones, etc., de distintos organismos públicos. Junto a los catálogos, se pueden encontrar directorios de bibliotecas, de recursos digitales, guías y censos de archivos, revistas digitales y digitalizadas, portales, y servicios de recolección como Hispana y metabúsqueda como el Buscón de la BNE. Si bien aún no existe un servicio equivalente a nivel europeo, se está trabajando en el desarrollo de este Pan-European Open Data Portal²², que aglutinará los proveedores de datos de los distintos países y de la propia Comisión Europea.

Ahora bien, de todos estos conjuntos de datos públicos, que se han abierto con el objetivo último de ser reutilizables, en contados casos se han publicado en formatos abiertos y conforme al resto de los principios LOD. A pesar de afirmar que el catálogo ofrece información en formatos estructurados (XML, CSV, XLS, RSS, JSON, RDF), con el fin de automatizar su reutilización²³, los formatos mayoritarios son XHTML, HTML, y PDF, y muy pocos conjuntos de datos se ofrecen en formatos estructurados como XLS, o CSV, y menos aún en XML o RDF. Aún con todo, un 10,6% de las empresas reutilizadoras participantes en el último estudio del sector infomediario²⁴, basan su actividad en la información procedente de museos, bibliotecas y archivos.

En cuanto a las aportaciones desde el ámbito GLAM (galerías, bibliotecas, archivos y museos) en España a la nube de datos enlazados, o *Linked Data Cloud* (ver figura 2), son todavía escasas. Destacan los esfuerzos de la Biblioteca Nacional Española, a través del portal

²⁰ W3C Library Linked Data Incubator Group: <http://www.w3.org/2005/Incubator/lld/>

²¹ Catálogo de Información Pública de la Administración General del Estado, Portal Datos.gob.es: <http://datos.gob.es/datos/>

²² Open Data Portal: http://ec.europa.eu/information_society/policy/psi/open_data_portal/index_en.htm

²³ Datos.gob.es, Preguntas frecuentes: <http://datos.gob.es/datos/?q=faq#5>

²⁴ *Estudio de caracterización del sector infomediario. Edición 2012*. Ministerio de Hacienda y Administraciones Públicas, Secretaría de Estado de Administraciones Públicas; Ministerio de Industria, Energía y Turismo, Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información; Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información; Red.es. Disponible en: <http://datos.gob.es/datos/?q=node/1840>.

datos.bne.es, con la publicación de sus catálogos bibliográficos y de autoridades; y de la propia Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación Cultura y Deporte, con la lista de encabezamiento de materias (LEM) de las bibliotecas públicas; e Hispana y su contribución al proyecto Europeaana Linked Open Data²⁵, aportando 1,5 de los 2,4 millones de objetos digitales que aglutina el proyecto.

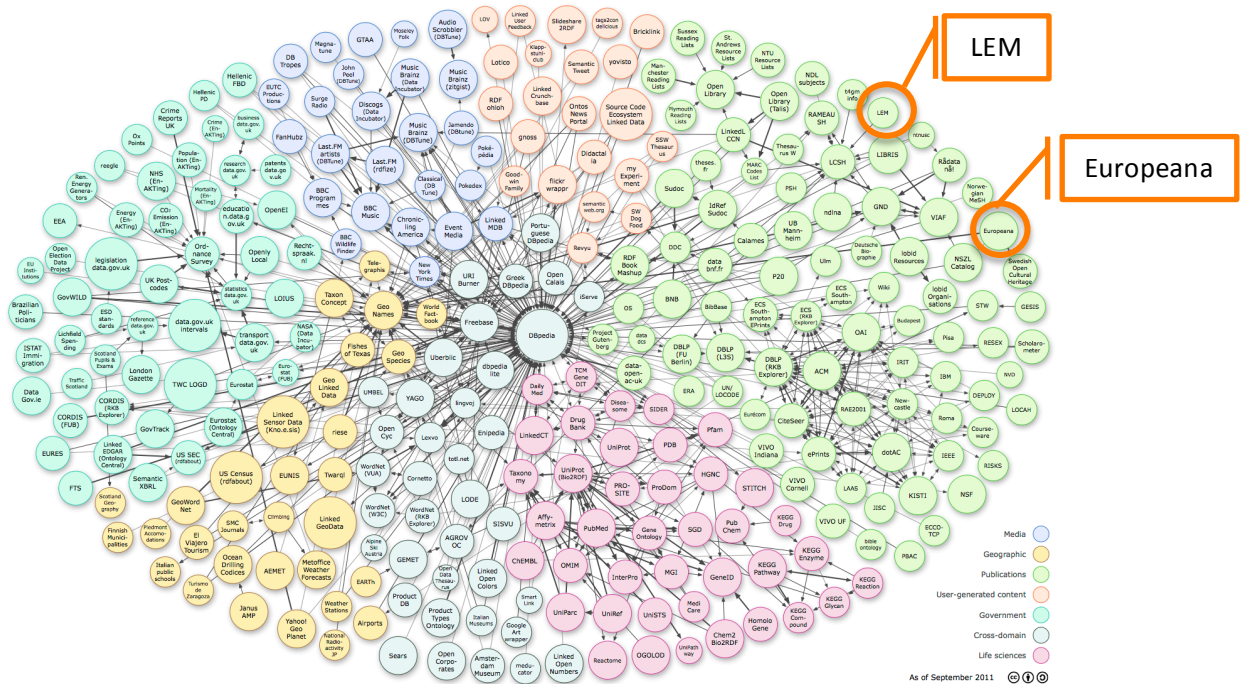


Figura 2. Visualización de la Linked Data Cloud, sept. 2011

En el plano internacional, otras bibliotecas nacionales están llevando a cabo iniciativas similares, como es el caso del Reino Unido (British Library), Estados Unidos (Library of Congress), Francia (Bibliothèque nationale de France), Alemania (Deutsche Nationalbibliothek), o Hungría (National Széchényi Library). Y desde una perspectiva más amplia, el informe de casos de uso del Grupo Incubador de Datos Vinculados de Bibliotecas del W3C analiza más de sesenta iniciativas en las múltiples áreas LOD anteriormente señaladas (datos bibliográficos, datos de registros de autoridades, alineamiento de vocabularios, colecciones, objetos digitales, archivos y datos heterogéneos, citación, usos sociales y usos innovadores).

Todo ello nos ofrece un panorama de lo que es, y de lo que puede llegar a ser, la publicación y reutilización de datos e información en el ámbito bibliotecario. Nos lleva a imaginar nuevos nodos en la nube de datos enlazados. Que se incorporen todos los proveedores GLAM de nuestro catálogo de datos abiertos. Que lo público y reutilizable vaya de la mano de lo abierto y enlazado. En definitiva, lograr que la información, que los datos, sean públicos, abiertos, enlazados y reutilizables.

²⁵ Data.europeana.eu, Europeaana Linked Open Data (LOD). Disponible en: <http://pro.europeana.eu/linked-open-data>.