

**Buenas prácticas y recomendaciones
estratégicas sobre formatos de
intercambio aplicables al Servicio de
Registro de Bibliotecas y Entidades
Relacionadas (SERBER) en un contexto
de datos abiertos y enlazados**

Informe

Madrid, 4 de septiembre de 2012



Informe realizado por Infor@rea por encargo de la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte

Preparación

Miquel Térmens Graells

Miquel Centelles Velilla

(Dept. Biblioteconomía y Documentación. Universidad de Barcelona)

Coordinación

Elisa García-Morales (Infor@rea)

Domingo Arroyo Fernández (Subdirección General de Coordinación

Bibliotecaria. Ministerio de Educación, Cultura y Deporte)

Madrid, 4 de septiembre de 2012

© Ministerio de Educación, Cultura y Deporte, 2012



El presente informe pertenece a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y está bajo una licencia de **Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 España** (<http://creativecommons.org/licenses/by-nc-sa/3.0/es/>) y por ello está permitido copiar, distribuir y comunicar públicamente esta obra bajo las condiciones siguientes:

Reconocimiento — Debe reconocer los créditos de la obra de la manera especificada haciendo referencia expresa a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y al identificador <http://hdl.handle.net/10421/6501> en <http://travesia.mcu.es>. Dicho reconocimiento no podrá en ningún caso sugerir que el Ministerio de Educación, Cultura y Deporte presta apoyo a dicho tercero o apoya el uso que hace de su obra.

No comercial — No puede utilizar esta obra para fines comerciales.

Compartir bajo la misma licencia — Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

Contenido

Resumen ejecutivo	5
1. Antecedentes y objetivos de la publicación de datos abiertos del SERBER según el estándar ISO 2146:2010.....	7
1.1 Antecedentes.....	7
1.2 Objetivos	7
1.3 Formatos	9
1.4 Propuesta estratégica	10
2. Marco normativo aplicable a la publicación de datos abiertos en el sector público.....	10
3. Recomendaciones internacionales y nacionales sobre formatos para la publicación de datos abiertos en el sector público.....	12
3.1 Qué formatos específicos responden a la definición de estándares abiertos.....	12
3.2 Cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER.....	14
4. Revisión de soluciones aplicadas por otros servicios de registro y directorios de bibliotecas y organizaciones relacionadas	18
5. Descripción y comparación de las alternativas posibles.....	24
6. Recomendaciones estratégicas de futuro: de la publicación de datos abiertos a la publicación de datos enlazados.....	26
7. Bibliografía	27
8. Glosario	29

Resumen ejecutivo

En el marco del diseño e implementación de un modelo estratégico para el Servicio de Registro de Bibliotecas y Entidades Relacionadas con el Ministerio de Educación, Cultura y Deporte (MECD), según la norma ISO 2146:2010, la Subdirección General de Coordinación Bibliotecaria del MECD encargó a la empresa Infor@rea un estudio sobre la situación y buenas prácticas en relación a los formatos de intercambio de datos a publicar en dicho registro (desde ahora SERBER).

En el primer apartado, se establecen las condiciones que el proyecto impone a la propuesta de formatos de publicación de datos, y las fuentes de información utilizadas para el estudio. En lo que respecta a las condiciones, es necesario abarcar, a largo plazo, los datos vinculados a los cinco tipos de objetos del modelo conceptual de ISO 2146:2010 –además de la clase principal “Objeto Registro”, las subclases Partes (personas y organizaciones), Actividades, Colecciones, y Servicios– y la limitación de los formatos de publicación de datos al marco de los datos abiertos (open data). También se especifican las tres cuestiones básicas que el estudio trata de responder: qué son los estándares abiertos; qué formatos específicos responden a la definición de estándares abiertos, y cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER.

En el apartado 2, se define y delimita el concepto de estándar abierto a partir del marco normativo europeo y estatal de interoperabilidad y de reutilización de datos en el sector público. Estas fuentes permiten, además, detallar las circunstancias de aplicación de estándares no abiertos, y los criterios de selección entre diferentes estándares.

En el apartado 3, y sobre la base de recomendaciones de especialistas nacionales e internacionales, se analiza qué formatos específicos responden a la definición de estándares abiertos, y cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER. Los tres formatos de referencia seleccionados para los propósitos del proyecto de SERBER son CSV, XML y RDF (con sus diferentes modalidades de serialización). Este último formato permite la incorporación de formatos de vinculación de datos, hecho que exige una profunda planificación previa que, entre otros aspectos, analice y seleccione los vocabularios idóneos para la descripción de los datos.

En el apartado 4 se presenta el resultado de la búsqueda exhaustiva y del análisis de servicios de registro y directorios de bibliotecas y organizaciones relacionadas a nivel internacional que publican datos mediante formatos de intercambio. El resultado ha sido la identificación de una decena de casos de diferentes tipos, que se clasifican en tres grupos: servicios de registro de bibliotecas e instituciones similares (*registries of libraries and related organizations*) que aplican la norma ISO 2146:2010, y cuyo objeto principal de descripción en el modelo conceptual son las partes (personas y organizaciones); registros de servicios de colección (*collection service registries*) en pleno funcionamiento, que, como en el grupo anterior, aplican la norma ISO 2146:2010, pero se diferencian de ellos porque el objeto principal de descripción en el modelo conceptual son las colecciones; y, finalmente, directorios diseñados a partir de modelos diferentes a la norma ISO 2146:2010 y que, por lo tanto, no aplican su modelo conceptual.

En los dos últimos apartados, se exponen las recomendaciones para la publicación de datos abiertos del proyecto de SERBER, expresadas en dos etapas sucesivas: una primera etapa (apartado 5) en la que los datos se publicarán en dos formatos estructurados, CSV y XML; y una segunda etapa (apartado 6), marcada por la transición estratégica hacia los datos enlazados en el contexto de la web semántica. En este segundo apartado, se presentan los pasos concretos a llevar a cabo, y los beneficios inmediatos que comportarían para el MECD.

1. Antecedentes y objetivos de la publicación de datos abiertos del SERBER según el estándar ISO 2146:2010

En el marco del diseño e implementación de un modelo estratégico para el Servicio de Registro de Bibliotecas y Entidades Relacionadas con el Ministerio de Educación, Cultura y Deporte, según la norma ISO 2146:2010, se plantea la necesidad de analizar la situación y buenas prácticas en relación a los formatos de intercambio de datos a publicar en dicho registro (desde ahora SERBER).

1.1 Antecedentes

La Subdirección General de Coordinación Bibliotecaria, de acuerdo con el *Real Decreto 257/2012, de 27 de enero, por el que se desarrolla la estructura orgánica básica del Ministerio de Educación, Cultura y Deporte*, art. 10 s) tiene la función de "la obtención, explotación y utilización de datos de bibliotecas". Entre los productos de esta actividad figuran, entre otros, los distintos directorios de personas y entidades que utiliza la Subdirección para su gestión operativa. La parte pública de estos datos constituyen el Directorio de bibliotecas españolas actualmente disponible en <http://directoribibliotecas.mcu.es/portada.html>.

En el año 2009 se inició el proyecto de creación de un nuevo directorio unificado partiendo de las siguientes condiciones:

1. Cumplimiento de la norma ISO/DIS 2146:2010. En el marco de este estándar, se adopta la variante de "registry of libraries and related organizations", caracterizado por la posición preeminente del objeto Grupo (organizaciones y profesionales) en el modelo conceptual del *SERBER*, alrededor del cual se sitúan los otros objetos: actividades, colecciones y servicios (*Proyecto DIRECTORIO ISO BIBLIOTECAS*, p. 5 y 22).
2. Adaptación a la norma MARC 21 para posibilitar la exportación e importación de información al esquema MARCXML. Dicha norma define el intercambio de información que permite estructurar e identificar los datos de tal forma que puedan ser reconocidos y manipulados por un ordenador (*Proyecto DIRECTORIO ISO BIBLIOTECAS*, p. 5).
3. Interoperabilidad e integración con otros sistemas de información y, especialmente, con el Punto de Consulta Único donde el formato prioritario de los registros de autoridades, bibliográficos y de *holdings* están expresados en formato MARC 21, en alguna modalidad de codificación (ISO2709 o MARCXML).

1.2 Objetivos

Las posibilidades de descarga de datos abarcan todos los elementos vinculados a los cinco tipos de objetos del modelo conceptual ISO 2146:2010. A saber: además de la clase principal "Objeto Registro", las subclases Partes (personas y

organizaciones), Actividades, Colecciones, y Servicios, aunque en el estadio actual del proyecto queda excluida la información sobre actividades y colecciones de las bibliotecas y organizaciones relacionadas, así como las estadísticas referidas a los diferentes tipos de objetos. El modelo conceptual de ISO 2146:2010 permitirá la incorporación de todos estos elementos inicialmente no contemplados en etapas posteriores.

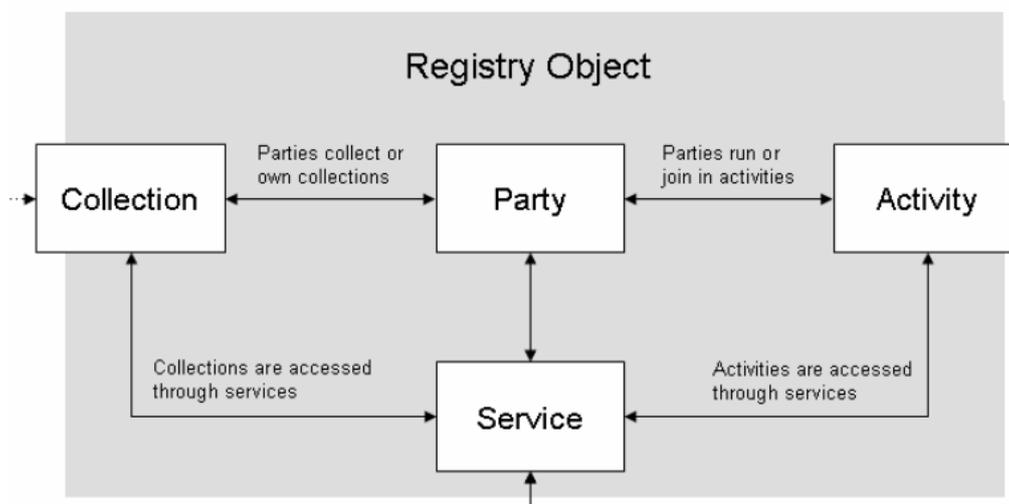


Diagrama: Modelo conceptual ISO 2146 clase Objeto Registro y subclases

Un aspecto clave en el tema que nos ocupa es la identificación de usos que se pretende dar a los datos publicados en el servicio de registro, que determina en gran medida los modelos funcionales de gestión de los datos y, en consecuencia, los formatos a implementar. Los usos son diferentes en función de los tipos de clientes establecidos. En el caso del *SERBER*, se han establecido tres tipos de clientes:

- Los usuarios internos del Ministerio de Cultura, Educación y Deporte. Los diferentes directorios de organismos y profesionales que la Subdirección General de Coordinación Bibliotecaria mantiene en las actuaciones en los ámbitos de la cooperación y coordinación bibliotecaria, utilizarán *SERBER* como fuente única y fiable de datos.
- Los proveedores de datos sobre bibliotecas y entidades relacionadas (usuarios terceros relacionados), en concreto los Servicios de Bibliotecas de las Comunidades Autónomas, y
- Los usuarios externos sin vinculación a la gestión del Directorio o a la provisión de sus datos.

En los dos primeros casos, las finalidades son previsibles y por lo tanto los requerimientos funcionales y los desarrollos y formatos exigidos por éstos. En el caso de los usuarios externos, los requerimientos funcionales son poco previsibles (y, hasta cierto punto, es importante que sea así). En el contexto del *SERBER*, debemos pensar en finalidades como las siguientes:

- Proporcionar ficheros de autoridades para bases de datos y catálogos bibliográficos.
- Facilitar la (re)utilización de registros completos y de datos en aplicaciones vinculadas a diversos dispositivos, y, especialmente, para servicios de información basados en geolocalización.
- Impulsar la difusión de colecciones y servicios mediante el uso de datos cuantitativos y cualitativos.
- Etc.

1.3 Formatos

En definitiva, el *SERBER* se plantea el reto de identificar e implementar formatos de publicación de datos, adecuados a la expresión de información de un servicio de registro de bibliotecas y organizaciones relacionadas, que puedan adaptarse a necesidades de usuarios externos, conocidas y no conocidas en la actualidad y en el futuro.

En el marco de este informe, limitamos la identificación de formatos de publicación de datos al marco de los datos abiertos (*open data*), definida como una filosofía y una práctica que requiere que ciertos datos sean de libre acceso para todos, sin limitaciones técnicas o legales. De entrada, todos los formatos son aceptables; ahora bien, cuanto más estructurados y enriquecidos estén los datos, mayor será la posibilidad de reutilización y de diseñar nuevas aplicaciones que los procesen de forma automática. Dicho de otra forma: no todos los formatos son igualmente valiosos para un proyecto de datos abiertos específico.

Nuestro informe trata de responder tres cuestiones básicas, que deben responderse en el orden en que se indican:

- Qué son estándares abiertos
- Qué formatos específicos responden a la definición de estándares abiertos
- Cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER

Afortunadamente, disponemos de algunos asideros en los contextos interno y externo del proyecto: el marco normativo europeo y estatal sobre interoperabilidad y reutilización de información del sector público; recomendaciones internacionales y nacionales sobre publicación de datos abiertos en el sector público; y (algunas) soluciones aplicadas por otros servicios de registro y directorios de bibliotecas y organizaciones relacionadas.

No son objeto de este informe los aspectos relacionados con la naturaleza de las aplicaciones informáticas que se utilizarán para la explotación de los datos, que pueden ser de fuentes abiertas (software abierto) o propietarias (software privativo). Es importante tener en cuenta que el modelo de estándares abiertos que el marco normativo estatal prioriza (como veremos en el apartado 2) se alinea claramente con el software de fuentes abiertas en el sector público. En cambio, el consumidor privado es libre de tomar sus decisiones por lo que respecta a la naturaleza de las aplicaciones para la explotación de los datos.

1.4 Propuesta estratégica

Antes de concluir este apartado de presentación, es importante destacar que una finalidad estratégica del proyecto de *SERBER* es constituir un modelo de referencia para la publicación de datos abiertos relativos a bibliotecas, profesionales, e instituciones similares de España en el portal nacional de acceso al catálogo de información pública, [-datos.gob.es-](https://datos.gob.es), y a otros portales similares del ámbito autonómico. Elementos relacionados con este objetivo:

- Los datos ofrecidos deben ser completamente fiables y actuales. La calidad se garantiza por la inmensa actividad de coordinación de la información que se realiza con las comunidades autónomas, el INE y otras organizaciones y grupos de trabajo.
- Los datos ofrecidos deberían tener carácter de exclusividad y, en este sentido, es altamente recomendable una alianza estratégica y, a ser posible, una integración con el servicio de datos actualmente presente en datos.gob.es: el Directorio de Bibliotecas y Hemerotecas Españolas.

2. Marco normativo aplicable a la publicación de datos abiertos en el sector público

La respuesta a la pregunta "Qué son estándares abiertos", en el contexto de un proyecto de publicación de datos de la administración pública, debe tener en cuenta el marco normativo estatal sobre la reutilización y la interoperabilidad de la información del sector público. En las disposiciones normativas que los regulan, ambos conceptos están profundamente interrelacionados, como observaremos a continuación.

En el ámbito de la reutilización de información, las disposiciones normativas de referencia son la "Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público" y el "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"¹. Esta segunda disposición establece, en su artículo 2.2, que los documentos reutilizables se pondrán a disposición del público

"de una manera estructurada y usable para los interesados e interesadas y preferentemente en bruto, en formatos procesables y accesibles de modo automatizado correspondientes a estándares abiertos"

El concepto de "estándar abierto" es definido en la "Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos", anexo, letra k, de la siguiente forma:

"Aquel que reúna las siguientes condiciones:

¹ Tal y como se indica en su preámbulo, esta ley "incorpora a nuestro ordenamiento jurídico la Directiva 2003/98/CE, de 17 de noviembre de 2003, del Parlamento Europeo y del Consejo, relativa a la reutilización de la información del sector público, establece el régimen jurídico general para la reutilización de dicha información."

- sea público y su utilización sea disponible de manera gratuita o a un coste que no suponga una dificultad de acceso,
- su uso y aplicación no esté condicionado al pago de un derecho de propiedad intelectual o industrial.”

Para una mayor concreción de la definición y alcance del concepto de “estándar abierto” se remite al “Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica”². En el anexo se define el “coste que no suponga una dificultad de acceso” en los siguientes términos:

“Precio del estándar que, por estar vinculado al coste de distribución y no a su valor, no impide conseguir su posesión o uso.”

El artículo 11 de este reglamento regula los estándares aplicables a la interoperabilidad técnica. Establece que los documentos y servicios de administración electrónica que los órganos o Entidades de Derecho Público emisores pongan a disposición de los ciudadanos o de otras Administraciones públicas se encontrarán disponibles, como mínimo y de forma prioritaria, mediante estándares abiertos (artículo 11.1). Como estándares complementarios a los abiertos, el Real Decreto hace referencia a los estándares de uso generalizado por los ciudadanos, es decir, usados por casi todas las personas físicas, personas jurídicas y entes sin personalidad que se relacionen o sean susceptibles de relacionarse con las Administraciones públicas españolas.

Los objetivos que persigue esta prelación de estándares son dos: la independencia en la elección de alternativas tecnológicas por los ciudadanos y las Administraciones públicas y la adaptabilidad al progreso de la tecnología.

En el artículo 11.2 se establece que el uso en exclusiva de un estándar no abierto sin ofrecer una alternativa basada en un estándar abierto se limitará a aquellas circunstancias en las que no se disponga de un estándar abierto que satisfaga la funcionalidad satisfecha por el estándar no abierto en cuestión y sólo mientras dicha disponibilidad no se produzca.

En el artículo 11.3 se establecen los criterios de selección de estándares y establecimiento del catálogo de estándares. Entre estos criterios se incluye la vinculación de los estándares a determinados tipos de normalización: estándares abiertos y estándares cuyo coste no suponga una dificultad de acceso; normas y especificaciones técnicas en los términos establecidos por la “Directiva 98/34/CE del Parlamento Europeo y del Consejo”; y especificaciones formalizadas. Además incluye otros criterios como son:

- adecuación del estándar a las necesidades y funcionalidad requeridas;
- condiciones relativas a su desarrollo, uso o implementación, documentación disponible y completa, publicación, y gobernanza del estándar;
- condiciones relativas a la madurez, apoyo y adopción del mismo por parte del mercado, a su potencial de reutilización, a la aplicabilidad

² Esta disposición normativa desarrolla la “Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos”.

multiplataforma y multicanal y a su implementación bajo diversos modelos de desarrollo de aplicaciones.

Como podemos observar, el marco normativo en el que se inscribe el desarrollo del *SERBER* establece criterios claros de selección de formatos de publicación de datos, que deben responder al carácter de estándares abiertos. En el apartado 3, este requerimiento legal se alinea con las recomendaciones de especialistas a nivel internacional.

Síntesis de aspectos a tener en cuenta en el proyecto SERBER a partir del marco normativo europeo y estatal de interoperabilidad y reutilización de datos en el sector público:

- Definición y alcance de estándares abiertos en el contexto de la normativa europea y estatal sobre interoperabilidad y reutilización de datos.
- Circunstancias de aplicación de estándares no abiertos.
- Criterios de selección entre diferentes estándares.

3. Recomendaciones internacionales y nacionales sobre formatos para la publicación de datos abiertos en el sector público

En este apartado nos proponemos responder a dos cuestiones básicas en el problema que nos atañe:

1. Qué formatos específicos responden a la definición de estándares abiertos
2. Cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER

3.1 Qué formatos específicos responden a la definición de estándares abiertos

Antes de identificar y presentar las recomendaciones internacionales y nacionales sobre publicación de datos abiertos en el sector público, debemos atender a la posición de la norma ISO 2146:2010 sobre este aspecto. Si bien en la "Introducción" (ISO 2146:2010, p. v) se autodefine como un modelo orientado a objetos que puede ser convertidos en formatos legibles por máquina como XML, declara que no es su intención prescribir un formato específico para el intercambio de datos entre sistemas. Posteriormente, insiste en este aspecto indicando que si bien su intención no es remplazar los estándares existentes para facilitar el intercambio de objetos de registro entre sistemas, puede utilizarse una versión del directorio de elementos de datos en forma de esquema

XML. En el apartado 4 volveremos a esta cuestión para presentar esquemas XML elaborados y publicados por perfiles de aplicación de la norma ISO 2146:2010, siguiendo las recomendaciones a que hemos hecho referencia.

En el ámbito de las iniciativas que surgen del sector de la administración pública, debe tenerse en cuenta, en primer lugar, la guía *Publishing Open Government Data* (Bennett y Harvey, 2009) elaborada y publicada a instancias del W3C, que enumera las siguientes recomendaciones para la selección de formatos idóneos:

- “El principal formato para la publicación de datos legibles por humanos es (X)HTML.
- Es más probable que los datos en bruto sean producidos utilizando formatos adaptados a los datos en concreto, a las herramientas empleadas, y a los estándares de la industria. W3C ha sido pionero en XML y RDF, que resultan convenientes para la manipulación excelente y para conjuntos de herramientas normalizadas. Se puede acceder a los contenidos de los archivos RDF y XML como si fueran bases de datos, utilizando SPARQL, XQuery, JavaScript y muchos otros lenguajes informáticos.
- Cuando sea posible, deben utilizarse estándares abiertos establecidos, y herramientas que permiten la producción y publicación de datos de forma sencilla y eficiente.
- También es importante tener en cuenta el poder de los datos enlazados.”

Estas recomendaciones son corroboradas y, en algunos aspectos, complementadas por la clasificación “de una a cinco estrellas” propuesta por Berners-Lee (2010) en lo que se refiere al grado de implementación de datos abiertos enlazados (*open linked data*):

- * Haz que tu material esté disponible en la web (en el formato que sea)
- ** Haz que esté disponible como datos estructurados (por ejemplo, es preferible Excel a una imagen de una tabla)
- *** Utiliza un formato no propietario (por ejemplo, es preferible CSV a Excel)
- **** Utiliza URLs para identificar cosas, de forma que la gente pueda apuntar a tu material
- ***** Enlaza tu material con los datos de otras personas para proporcionar contexto



Cyganiak (2011) sobre la base de la clasificación “de una a cinco estrellas” del grado de implementación de datos abiertos enlazados concreta algunos de estos criterios, estableciendo relaciones de formatos específicos para el sector público.

Por un lado, recomienda publicar en formatos procesables de forma automática, ya que permite a otras instancias procesar, analizar y visualizarlos de acuerdo con sus intereses específicos, y, por qué no, generar nuevos servicios y nuevas ideas. A partir de este criterio, establece una gradación de formatos:

- Buenos: MS Excel, CSV, XML, JSON, Microdata
- No muy buenos: sitios web a secas, MS Word
- Malos: PDF
- Realmente malos: gráficos o mapas sin números

Por otro lado, recomienda utilizar un formato abierto, ya que éstos son adecuados para herramientas y aplicaciones que la gente puede utilizar directamente. Entre estos formatos abiertos, destacan los siguientes: CSV, KML, RSS, XML, JSON, RDF.

En el contexto español, entre las diferentes iniciativas para facilitar la aplicación de las disposiciones normativas sobre interoperabilidad y reutilización de datos en el sector público, debe tenerse en cuenta la *Guía de aplicación del Real Decreto 1495/2011 por el que se desarrolla la Ley 37/2007 sobre Reutilización de la Información del Sector Público*. En el apartado “¿Cómo se preparan y representan los datos en el mejor formato posible?” (p. 24), se citan los siguientes formatos:

- Texto plano
- XML
- CSV
- HTML (HiperText Markup Language)
- RDF/XML (Resource Description Framework / eXtensible Markup Language)
- RDF/N3 (Resource Description Framework / Notation 3)
- TURTLE (Terse RDF Triple Language)
- JSON (JavaScript Object Notation)

3.2 Cuáles son los formatos abiertos idóneos para los propósitos del proyecto de SERBER

El marco normativo aplicable a la publicación de datos abiertos en el sector público de España, y las recomendaciones internacionales permiten establecer un elenco de formatos de publicación de datos, tal y como hemos observado en

el apartado anterior. Las características del proyecto de SERBER, donde es fundamental la representación de datos altamente estructurados, permiten establecer una lista de formatos preferentes reducida a tres: CSV, XML³ y RDF (con sus diferentes modalidades de serialización).

CSV (*Comma-Separated Values*): Formato de fichero para representar datos (textuales y numéricos) en forma de tabla, donde las columnas se separan por comas (o punto y coma) y las filas por saltos de línea. Texto del estándar: *Common Format and MIME Type for Comma-Separated Values (CSV) Files (RFC 4180)* <<http://tools.ietf.org/html/rfc4180>>.

XML (*eXtensible Markup Language*): Es un lenguaje de marcado que define un conjunto de reglas para la codificación de documentos en un formato que es a la vez legible y procesable. Deriva del lenguaje SGML y permite definir la gramática de lenguajes específicos (de la misma manera que HTML es a su vez un lenguaje definido por SGML) para estructurar documentos grandes. Está definido en *Extensible Markup Language (XML) 1.0 (Fifth Edition)* <<http://www.w3.org/TR/xml>> producida por la W3C, y diversas especificaciones relacionadas.

RDF (*Resource Description Framework*): Es una familia de especificaciones de la W3C originalmente diseñada como modelo de metadatos. Ha llegado a ser utilizado como método general para la descripción conceptual o modelado de información que se implementa en recursos web, utilizando una variedad de formatos de sintaxis. Su modelo de representación basado en tripletas sujeto-predicado-objeto constituye uno de los fundamentos de la web semántica.

La transición entre los formatos CSV y XML, por un lado, y el formato RDF, dibuja un salto cualitativo entre los tres primeros niveles de la clasificación de Berners-Lee (2010), anteriormente reproducida, y los niveles cuatro y cinco. Dicho de otra forma: significa una transición desde un conjunto de datos relacionados únicamente entre ellos ("una isla"), a datos susceptibles de relacionarse, más allá del conjunto en que han sido definidos y creados, con todos los datos disponibles en la web, con los que, originalmente, no mantenían ninguna afinidad. La publicación de un SERBER en CSV o XML permite la realización de consultas o de operaciones limitadas, fundamentalmente, a los datos incorporados en dicho registro. En cambio, la aplicación de la filosofía de datos enlazados, permite la interrelación de los datos con otros datos disponibles en la web, de forma que el directorio se convierte en un componente de una inmensa base de datos donde los resultados de las consultas y de otras operaciones no se reducen a la información disponible en el registro.

Los beneficios del paso a la web semántica son inimaginables para los depositarios y usuarios de los datos. La contrapartida es que adaptar el servicio de registro al modelo de datos enlazados requiere la aplicación de cuatro principios básicos. Los dos primeros nos sitúan en el nivel 4 de la clasificación de Berners-Lee (2010):

³ La alternativa JSON, si bien está ganando terreno, se encuentra todavía en un estadio de menor difusión que XML.

1. Utilizar URIs como nombres para las cosas.
2. Utilizar las HTTP de las URIs para que la gente pueda localizar y consultar dichos nombres."

Los dos últimos principios, culminan el nivel 5:

3. Proporcionar información útil, utilizando los estándares (RDF, SPAQL), para cuando alguien consulte una URI.
4. Incluir enlaces a otras URI relacionadas, de forma que sea posible descubrir más cosas."

La aplicación de dichos principios exige tomar decisiones y diseñar procesos, instrumentos, etc. tanto en el diseño de los datos a publicar –*"how one shapes and structures data to fit neatly in the Web"*, en palabras de Heath y Bizer (2010)–, como en las propias modalidades de publicación. Especialmente crítica es la selección y uso de vocabularios para la descripción de los datos. RDF proporciona un modelo de datos genérico, abstracto, para describir recursos utilizando tripletas *sujeto-predicado-objeto*. No obstante, no proporciona términos propios de un dominio determinado para describir clases de objetos del modelo ISO 2146:2010 (Partes, Servicios, Colecciones, Actividades) ni cómo se relacionan entre ellos. Esta función es realizada por los vocabularios de descripción de datos. Como norma general, es preferible la reutilización de uno o más vocabularios ya existentes; aquéllos que, en conjunto, permiten la representación de todos los objetos y relaciones entre objetos del modelo conceptual.

Heath y Bizer (2011, p. 84) aconsejan tener en cuenta los siguientes criterios para la selección de vocabularios a reutilizar:

1. "Uso y captación. ¿Es un vocabulario de uso generalizado? ¿El uso de este vocabulario motivará que un conjunto de datos sea más o menos accesible a las actuales aplicaciones de datos enlazados?
2. Mantenimiento y administración. ¿El vocabulario es activamente mantenido de acuerdo con un proceso de administración claro? ¿Con qué frecuencia y en qué condiciones se llevan a cabo las actualizaciones?
3. Cobertura. ¿Cubre el vocabulario una proporción suficiente del conjunto de datos para justificar su adopción en términos de "compromisos ontológicos"?
4. Expresividad. ¿El grado de expresividad en el vocabulario es apropiado para el conjunto de datos y para el escenario de aplicación? ¿Es demasiado expresivo o no suficientemente expresivo?"

A partir de la consulta de diferentes directorios de vocabularios y, especialmente, *Linked Open Vocabularies (LOV)*, es posible localizar vocabularios que cumplen con los criterios indicados y que abarcan los objetos y relaciones entre objetos del modelo de datos del *SERBER*:

AD - Address Schema
<http://labs.mondeca.com/dataset/lov/details/vocabulary_ad.html>
DIR - Directory Schema <<http://schemas.talis.com/2005/dir/schema#>>
RDAFRBR - FRBR Entities for RDA
<<http://rdvocab.info/uri/schema/FRBRentitiesRDA>>
FOAF - Friend of a Friend vocabulary <<http://xmlns.com/foaf/0.1>>
MADS - Metadata Authority Description Schema
<<http://www.loc.gov/mads/rdf/v1>>
SCHEMA - The schema.org terms in RDFS+OWL <<http://schema.org>>
SCSV - Schema.org CSV <<http://vocab.deri.ie/scsv>>
VCARD - An Ontology for vCards
<<http://www.w3.org/2006/vcard/ns-2006.html>>
WHOIS - Who's who description vocabulary
<<http://www.kanzaki.com/ns/whois#>>
Open Metadata Registry <<http://metadataregistry.org>>

Como observaremos en el apartado 4, no disponemos de muestras de servicios de registro o directorios de bibliotecas que hayan aplicado formatos de publicación de datos enlazados, y que estén en pleno funcionamiento. El proyecto más ambicioso es, sin duda, *lobid-organisations* <<http://lobid.org/organisation>>, vinculado al North Rhine-Westphalian Library Service Center (hbz) Linked Open Data Service. Aunque todavía no funciona a pleno rendimiento, constituye una interesante experiencia a tener en cuenta por el *SERBER*, ampliamente documentada a través de *Linked-Open-Data-Wiki des hbz*. Es importante destacar que el proyecto se inició en 2010 y ha exigido la inversión de importantes recursos económicos y humanos, así como consensos entre diferentes instancias implicadas. En este sentido, la propia *Guía de aplicación del Real Decreto 1495/2011 por el que se desarrolla la Ley 37/2007 sobre Reutilización de la Información del Sector Público* advierte sobre la necesidad de aumentar los esfuerzos y recursos requeridos para llevar a cabo un aumento de nivel en el grado de implementación de datos abiertos enlazados según la clasificación “de una a cinco estrellas”.

Síntesis de aspectos a tener en cuenta en el SERBER a partir de las recomendaciones de especialistas a nivel internacional:

- Deben aplicarse diversos formatos, y no uno en exclusiva.
- Deben excluirse los formatos propietarios en beneficio de formatos libres que se pueden crear y manipular por cualquier software, libres de restricciones legales.
- Deben preferirse los formatos de datos estructurados o procesables de forma automática en diversos tipos de aplicaciones, frente a los formatos de datos "fijos" (por ejemplo, PDF).
- Los formatos más cotizados en las fuentes consultadas y que resultan apropiados para el contexto del Directorio de bibliotecas son CSV, XML y RDF.
- Sin renunciar a facilitar el uso de los datos en cualquier contexto, los formatos deben asegurar aplicaciones especializadas en servicios de registro y en gestión documental.
- Sin excluir el uso de estándares que faciliten el intercambio de objetos de registro entre sistemas, puede utilizarse una versión en esquema XML del directorio de elementos de datos de la norma ISO 2146:2010.
- La incorporación de formatos de datos enlazados exige una profunda planificación previa que, entre otros aspectos, analice y seleccione los vocabularios idóneos para la descripción de los datos.

4. Revisión de soluciones aplicadas por otros servicios de registro y directorios de bibliotecas y organizaciones relacionadas

Mediante la utilización de diferentes fuentes y estrategias de consulta, se ha llevado a cabo una búsqueda exhaustiva de servicios de registro y directorios de bibliotecas y organizaciones relacionadas a nivel internacional, que publican datos mediante formatos de intercambio. El resultado ha sido la identificación de una decena de casos de diferentes tipos, que podemos clasificar en tres grupos:

- Grupo A: Servicios de registro de bibliotecas e instituciones similares (registries of libraries and related organizations) que aplican la norma ISO 2146:2010, y cuyo objeto principal de descripción en el modelo conceptual son las partes (personas y organizaciones). Los objetos relativos a organizaciones y profesionales se sitúan en una posición subordinada. La norma se refiere a ellos expresamente en el "Annex C".
- Grupo B: Registros de servicios de colección (collection service registries), que, como en el grupo anterior, aplican la norma ISO 2146:2010, pero se diferencian de ellos porque el objeto principal de descripción en el modelo conceptual son las colecciones. La norma se refiere a ellos expresamente en el "Annex D".

- Grupo C: Directorios diseñados a partir de modelos diferentes a la norma ISO 2146:2010 y que, por lo tanto, no aplican su modelo conceptual.

En los grupos A y B encontramos los perfiles de aplicación de la norma ISO 2146:2010.

A continuación se presentan las muestras de cada uno de los tres grupos.

Grupo A.

En este grupo se sitúa el WorldCat Registry <<http://www.oclc.org/registry>>, vinculado a OCLC.

Es el perfil de aplicación de la norma ISO 2146:2010 que, desde el punto de vista del modelo conceptual, se aproxima más al *SERBER*. Es definido como *"a free Web tool that provides a single location from which any library can view data that describes its institutional identity and services, and update it in OCLC Service Configuration. Free Web services available through the WorldCat Affiliate site distribute Registry data across the Web, enhancing Web discovery of libraries' rich content and services."* (*WorldCat Registry at a glance* <<http://www.oclc.org/registry/about/default.htm>>).



Entre las características del servicio de registro destaca la posibilidad de descargar registros como ficheros XML a una unidad local o en red, para la importación en aplicaciones de datos estructurados y para el mantenimiento y revisión fuera de línea. Con esta finalidad, se ha creado un esquema XML para el perfil de OCLC, denominado *WorldCat Registry Schema* <<http://www.oclc.org/developer/documentation/registry/xml-schemas>>, a partir del cual se generan registros XML para cada uno de los objetos disponibles en el directorio. Tiene una estructura modular, segmentada en un esquema básico, *institution.xsd*, que abarca los elementos de datos de identificación de los objetos, y 11 componentes adicionales, de forma que pueden mezclarse y combinarse de acuerdo a la naturaleza de la institución que va a ser representada. Los registros en formato XML son descargables mediante la opción *"Download this Profile as XML"*, que acompaña la descripción de cada objeto del directorio. Puede observarse en el caso de la Biblioteca del Banco de España <<http://www.worldcat.org/registry/Institutions/111783>>.

Grupo B

En este grupo de experiencias situamos tres servicios de registro vinculados al Global Registries Initiative <<http://www.globalregistries.org/index.html>>: Australian National Data Service <<http://ands.org.au>>; Information Environment Service Registry <<http://iesr.ac.uk>> y Ockham Digital Library Services Registry <<http://www.ockham.org>>⁴.

⁴ Durante el periodo de realización del informe, ha sido imposible la consulta de este repositorio.



Australian National Data Service <<http://ands.org.au>> es una infraestructura para facilitar a los investigadores australianos la publicación, descubrimiento, acceso y utilización de datos resultantes de la investigación. El perfil de aplicación de la norma ISO 2146:2010 prioriza los objetos de colecciones y servicios, tal como se observa en el modelo conceptual <<http://www.ands.org.au/guides/rif-cs-awareness.html>>. La descripción de los objetos y sus relaciones se expresa mediante el esquema XML RIF-CS (Registry Interchange Format – Collections and Services) <<http://ands.org.au/resource/rif-cs.html>>, que fue desarrollado como formato de datos para facilitar el intercambio de metadatos de colecciones entre las instituciones participantes, y el envío de metadatos de colecciones al registro central (ANDS Collection Registry). Más concretamente, el esquema ha sido diseñado con el propósito de exponer metadatos de colecciones al registro central vía un proveedor de datos OAI-PMH.

Tiene una estructura modular, segmentada en 6 componentes correspondientes a los cinco objetos del modelo conceptual –Registry Objects Schema, Activity Schema, Collection Schema, Party Schema y Service Schema–, y un componente adicional dedicado a Type Definitions.

Existen dos entornos diferenciados de acceso a este servicio de registro: como usuario que contribuye con registros y datos, mediante *ANDS online services* <https://services.ands.org.au/home/login.php?logout=logout&page=%2Fhome%2Forca%2Fuser%2Fcollection_add.php>; o como usuario final, mediante *Research Data Australia* <<http://services.ands.org.au/home/orca/rda/index.php>>. En el primer caso, se ofrece la posibilidad de visualizar o descargar los registros de RIF-CS para cada uno de los objetos descritos. Véase, por ejemplo, el caso de University of Sydney Library <<https://services.ands.org.au/home/orca/view.php?key=http%3A%2F%2Fnla.gov.au%2Fnla.party-531570>>. En cambio, en el entorno de búsqueda para el usuario final (Research Data Australia) los registros de partes (personas y organizaciones) no permiten la descarga de datos en formatos de intercambio. Únicamente si accedemos al registro correspondiente en el repositorio que provee los datos, son posibles las descargas. Por ejemplo, en el caso del repositorio Trove <<http://trove.nla.gov.au/people/783443?c=people>>, los registros permiten la descarga de datos en formatos bibliográficos como EndNote, XML i BibTeX.



Information Environment Service Registry <<http://iesr.ac.uk>>. De acuerdo con *IESR: Metadata* <<http://iesr.ac.uk/metadata/>>, este repositorio contiene información sobre colecciones, servicios y partes; y, más concretamente, sobre:

- Las colecciones de recursos de información, los servicios asociados que facilitan el acceso a las colecciones, y las partes que disponen de las colecciones y/o administran los servicios.

- Los servicios transaccionales.

El modelo conceptual de este perfil puede consultarse en *IESR: Application Profile* <<http://iesr.ac.uk/profile/index.html>>. Cabe destacar que los metadatos para colecciones se basan en RSLP Collection Description <<http://www.ukoln.ac.uk/metadata/rsdp/schema>>.

En el momento actual, no se ofrece la posibilidad de descargar o visualizar registros con formatos de intercambio; no obstante, y de acuerdo con la noticia "Linked Data 29 July 2011" <<http://iesr.ac.uk/news/>>, IESR está preparando el salto a los datos enlazados, para lo que, como primer paso, ha convertido los metadatos en tripletas RDF, utilizando microformatos RDFa incrustados en las páginas web.

Grupo C

Fuera del ámbito de aplicación de la norma ISO 2146:2010, encontramos directorios nacionales vinculados a agencias de asignación de códigos de identificación en el entorno bibliográfico (especialmente orientados a los servicios de préstamo e intercambio bibliotecario). Los dos ejemplos más destacables son los directorios de las agencias que asignan "MARC organization codes" (anteriormente conocidos como NUC codes) y el International Standard Identifier for Libraries and Related Organizations (ISIL).

En el caso de los directorios vinculados a los códigos MARC⁵, únicamente la agencia del Reino Unido (United Kingdom MARC Organization Codes <<http://www.bl.uk/bibliographic/marcagency.html>>) proporciona la descarga del directorio completo en formato PDF.

Un mayor compromiso con el intercambio de datos se observa en el caso de los directorios vinculados a las agencias de asignación del International Standard Identifier for Libraries and Related Organizations (ISIL). La ISIL Registration Authority <<http://biblstandard.dk/isil/>>, reúne 26 de estos directorios creados y mantenidos por otras tantas agencias nacionales, de los que cinco ofrecen alguna modalidad de descarga y visualización de registros en formatos de intercambio de datos. Se detallan a continuación:

Alemania. El directorio *ISIL- und Sigelverzeichnis online* <<http://dispatch.opac.d-nb.de/DB=1.2>> responsabilidad de Staatsbibliothek zu Berlin – Zeitschriftendatenbank. Este directorio incluye todos los tipos de códigos asignados a las bibliotecas alemanas por la Staatsbibliothek zu Berlin (códigos ISIL y MARC). Estos códigos son utilizados para la identificación de bibliotecas y organizaciones relacionadas en catálogos colectivos, sistemas de préstamo interbibliotecario, etc. Permite la descarga de registros en formato etiquetado, mediante mensajería electrónica o la visualización en pantalla de un registro individual o un conjunto de registros. Es posible seleccionar los campos del registro que se incluyen en la descarga.

⁵ El único servicio próximo a un directorio de directorios es Library of Congress code agency *Search the MARC Organization Codes Database* <<http://www.loc.gov/marc/organizations>>.

Austria. El directorio *Adressen-, ISIL- und Sigelverzeichnis* <http://aleph20-prod-acc.obvsg.at/F?CON_LNG=ger&func=find-b-0&local_base=acc09> es responsabilidad de Die Österreichische Bibliothekenverbund und Service GmbH <<http://www.obvsg.at>>. Permite la descarga de un registro individual o un conjunto de registros mediante mensajería electrónica o en una unidad de disco. El formato es etiquetado, y puede incluir todos los campos del registro o una selección. En el caso de la descarga en unidad de disco, la extensión del fichero es .sav, que corresponde a una aplicación SPSS. La codificación de caracteres es en ASCII, en las dos modalidades de descarga, en ISO 8859-1 (Roman character sets), en el caso de la descarga por mensajería electrónica, y en Unicode / UTF-8 (non-Roman character sets), en el caso de la descarga en unidad de disco.

Francia. El directorio *Répertoire des centres de ressources du SUDOC* <<http://www.sudoc.abes.fr>> es responsabilidad de l'Agence Bibliographique de l'Enseignement Supérieur <<http://www.abes.fr>>. Permite la descarga por mensajería electrónica o la visualización en pantalla de un registro individual o un conjunto de registros. El formato es etiquetado, y puede incluir todos los campos del registro o una selección.

Italia. El directorio *Anagrafe biblioteche italiane* <<http://anagrafe.iccu.sbn.it/index.html>> es responsabilidad del Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche. Actualmente, no permite a los usuarios la visualización o descarga de registros en formatos de intercambio de datos. Según *Anagrafe delle biblioteche italiane* <http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/naz/pagina_78.html?sessionid=2DB1B10DDB6E0E569403CE646204C535>, esta posibilidad se aplica únicamente para la importación y exportación de datos para la entidades que contribuyen en el mantenimiento del directorio. El formato aplicado es XML, cuyo esquema actual puede consultarse en *Anagrafe delle biblioteche italiane - Formato di scambio XML* <http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/naz/pagina_365.html>.

Nueva Zelanda. El directorio *Directory of New Zealand libraries* <<http://directory.natlib.govt.nz/library-symbols-web/Home.html>> es responsabilidad de la National Library of New Zealand Te Puna Mātauranga o Aotearoa <<http://www.natlib.govt.nz/en/services/6docsupply.html#sect1>>. Permite la descarga del directorio completo en los formatos PDF y MS-Word, y la descarga de las últimas actualizaciones del directorio en los mismos formatos.

Suiza. El directorio *HelveticArchives* es responsabilidad de la Swiss National Library <<https://www.helveticarchives.ch/suchinfo.aspx>>. Permite la visualización y descarga de registros individuales o de un conjunto de registros en un formato etiquetado en PDF.

Mención aparte merece el proyecto *lobid-organisations* <<http://lobid.org/organisation>>, al que hemos hecho referencia en el apartado 2 de este informe. Es un servicio de *lobid* <<http://lobid.org>> ("Linking Open Bibliographic Data"), que se autodefine como un directorio internacional de bibliotecas y organizaciones relacionadas que sigue los principios de los datos enlazados. El modelo conceptual de este registro no sigue los dictados de la norma ISO 2146:2010, aunque mantiene profundas similitudes estructurales. Entre abril y agosto de 2010, se diseñó un prototipo de la infraestructura técnica para la publicación, descripción y actualización de datos bibliográficos. Se detectó entonces la necesidad de describir bibliotecas y organizaciones

relacionadas, así como las actividades, servicios y colecciones de éstas, en el marco de la infraestructura de datos abiertos. Siguiendo los principios de los datos enlazados, se planificaron dos procesos:

- Identificación de objetos mediante URIs apropiados para datos enlazados (URIs HTTP). Las URIs se basan en el [International Standard Identifier for Libraries and Related Organizations \(ISIL\)](#) que también pueden actuar como [MARC Organization Codes](#).
- Provisión de información útil para las URIs consultadas, mediante el uso de estándares. Inicialmente, los datos se incorporaron desde dos fuentes: la [address database for german libraries](#) y la [MARC Organization Codes Database](#). Para el mantenimiento y actualización de los datos, se descartaron las opciones de recibirlos de las bibliotecas e instituciones y trasladarlos al registro de forma centralizada, o de proporcionar opciones de edición a terceros. La solución escogida fue la agregación de datos provenientes directamente desde las web institucionales mediante un modelo de descripción RDF incrustada en los contenidos HTML: RDFa.

Como puede observarse, las soluciones implementadas presentan diferencias fundamentales respecto al proyecto de *SERBER*:

- Diferente modelo conceptual.
- Diferente modelo de recopilación, actualización y mantenimiento de los datos.

No obstante, sí sirve de ejemplo sobre los procesos que son necesarios realizar y las decisiones que hay que tomar para realizar el salto desde la publicación de datos para el consumo externo, a la implementación de los principios de datos enlazados, y, especialmente:

- El sistema de identificación de todos los componentes del modelo conceptual (objetos y relaciones entre objetos) mediante URIs apropiadas.
- El modelado de los objetos y relaciones entre objetos en forma de ontología.
- La selección y aplicación de vocabularios de descripción de los datos.

En cuanto a los vocabularios de descripción de datos aplicados, los responsables ponen en evidencia las dificultades planteadas, ya que en la versión de prueba (*Howto – Describing libraries, their collections and services in RDF* <<https://wiki1.hbz-nrw.de/display/SEM/Howto+-+Describing+libraries%2C+their+collections+and+services+in+RDF>>, deben combinar hasta 5 vocabularios diferentes para representar los metadatos de todos los objetos y relaciones del modelo conceptual:

- *[^vcard](#) for representing address and contact information.*
- *[Good Relations](#) for representing information about services like opening hours etc.*

- [DC Terms](#)
- *For collection description, the [Dublin Core Collections Application Profile](#), including the [Collection Description Terms](#) and [DCMI Type Vocabulary](#).*
- [Organization Ontology](#) *for representing hierarchical relations between organizations and their units."*

Una muestra del tipo de registros de descarga de datos enlazados es <<http://lobid.org/organisation/DE-605/about>>, que aplica los formatos RDFa, Turtle, RDF/XML.

Síntesis de aspectos a tener en cuenta en el SERBER a partir de las experiencias de otros servicios de registro y directorios de bibliotecas y organizaciones relacionadas:

- Los servicios de registro vinculados a la norma ISO 2146:2010 manifiestan una preferencia por el desarrollo de perfiles de aplicación de la norma, basados en esquemas XML, siguiendo la recomendación de la propia norma ISO.
- En las experiencias vinculadas a sistemas de codificación de bibliotecas se aplican soluciones poco avanzadas desde el punto de vista de la filosofía de datos abiertos y, más específicamente, una preferencia por PDF y por formatos de etiquetado ad hoc.
- Las experiencias de formatos de datos enlazados se sitúan aún en una fase de desarrollo y experimentación. El caso específico de lobid-organisations es especialmente interesante como modelo de buenas prácticas en el diseño y ejecución de procesos, y en la toma de decisiones críticas para la implementación de los principios de datos enlazados en un registro de bibliotecas y organizaciones relacionadas.

5. Descripción y comparación de las alternativas posibles

Sobre la base del marco normativo aplicable a la publicación de datos abiertos en el sector público, en los ámbitos de reutilización e interoperabilidad de la información (apartado 2), y teniendo en cuenta las recomendaciones internacionales y nacionales sobre formatos para datos abiertos (apartado 3) y las soluciones aplicadas por servicios de registro y directorios de bibliotecas y organizaciones relacionadas (apartado 4), realizamos la siguiente propuesta de actuación:

De forma inmediata, se propone la implementación de los siguientes formatos:

- CSV (Comma Separated Values). Es un formato orientado específicamente a datos organizados en forma de directorio o de hojas de cálculo que es soportado por muchas aplicaciones, incluyendo gestores de correo electrónico como Thunderbird, Gmail y Hotmail.
- XML (eXtended Markup Language). Amplía las prestaciones por su especial vinculación a la norma ISO 2146:2010, y por ser un punto de partida para opciones avanzadas en el marco de la web semántica y de la automatización de bibliotecas. La especificación de este metalenguaje general a las condiciones del directorio se realizará de dos formas:
 - Mediante el diseño y publicación de un esquema XML que conformará el perfil de aplicación del SERBER, siguiendo los modelos de los esquemas para el perfil de WorldCat Registry y, en segunda opción, de RIF-CS.
 - Mediante la aplicación del formato MARC XML.

A partir de las experiencias de referencia, es posible indicar algunas recomendaciones sobre la gestión del esquema XML:

- Organización del esquema en forma modular, a partir de un componente que incorpore los elementos de datos relativos a identificación de los objetos.
- Rigor en el mantenimiento del esquema, que asegure la incorporación de nuevos objetos y relaciones conforme se amplíe o modifique el perfil de aplicación; que refleje los posibles cambios en elementos de datos y vocabularios controlados, etc.
- Clara y correcta localización del esquema XML en la sede del SERBER.
- Provisión de servicios de valor añadido del esquema XML, como es el mapeo con esquemas de otros perfiles de aplicación de referencia (WorldCat Registry Schema y RIF-CS), y con esquemas de otros modelos como DCMI Collections AP.

Finalmente, se recomienda que el acceso a los datos se realice en las siguientes condiciones:

- Debe permitirse tanto la visualización como la descarga de datos.
- Debe permitirse la visualización y descarga de un único registro o de un conjunto de registros.
- Debe permitirse la selección de datos específicos de un único registro o de un conjunto de registros.
- Debe permitirse la sindicación de la actualización de datos en los diversos formatos.

La aplicación del formato RDF, en alguna de sus formas de serialización, situaría el proyecto del *SERBER* en el marco de la web semántica y, más

específicamente, permitiría la vinculación de datos. Este paso requiere, no obstante, un análisis más profundo que se trata en el siguiente apartado.

6. Recomendaciones estratégicas de futuro: de la publicación de datos abiertos a la publicación de datos enlazados

Una vez asegurada la publicación de datos abiertos en las modalidades propuestas en el apartado 5, el tránsito hacia la publicación de datos enlazados exige la realización de cuatro procesos básicos de preparación de los datos.

- En primer lugar, debe concretarse el modelo conceptual del perfil de aplicación de la norma ISO 2146:2010 con el que se fundamenta el SERBER. En este modelo conceptual deben quedar perfectamente definidos y representados los objetos y las relaciones entre los objetos. Este aspecto queda asumido en el Modelo de información y análisis funcional definido en el marco de este proyecto [RDF]
- En segundo lugar, debe establecerse el sistema de identificación de los diferentes tipos de objetos representados en el modelo conceptual mediante URIs apropiadas para datos enlazados, es decir, URIs de HTTP. En el caso de las partes y, más específicamente las bibliotecas y organismos similares, la base de las URIs podría ser alguno de los sistemas internacionales de identificación y, principalmente, los identificadores asignados por el ISIL, o los identificadores MOC asignados por la Library of Congress.⁶ Adicionalmente, debería establecerse el sistema de identificación mediante URIs para el resto de objetos representados en el SERBER: personas físicas, actividades, colecciones, y servicios.
- En tercer lugar, deben identificarse y seleccionarse los vocabularios de descripción de datos necesarios para la representación de objetos y relaciones, y coherente con el modelo conceptual del Registro. En este paso, es importante tener en cuenta la norma general de reutilización de vocabularios ya existentes, y los criterios de selección propuestos por Heath y Bizer (2011, p. 84), que hemos indicado en el apartado 3. Es especialmente aconsejable tener en cuenta la experiencia previa del proyecto lobid-organisations <<http://lobid.org/organisation>>, que hemos presentado en el apartado 5.
- En cuarto lugar, deberá diseñarse un esquema de codificación de los datos del Registro en una sintaxis de ontología, preferentemente RDFS. Además, se debe asegurar la correcta publicación y mantenimiento del esquema.

Hecho esto puede avanzarse en una nueva dimensión que constituye el enriquecimiento de datos y que incluye el diseño de los procesos funcionales de creación de enlaces dentro del SERBER y entre el Registro y otros conjuntos de

⁶ Una alternativa local es utilizar como identificador de referencia el NIDEN que otorga el INE. El problema es que no es un valor estable y ya ha ocurrido en ocasiones que el INE ha modificado códigos previamente asignados.

datos, y/o la selección e implementación de una o varias "recetas" para la publicación de datos vinculados:

- Servir datos enlazados como archivos RDF/XML estáticos.
- Servir datos enlazados como RDF integrado en archivos HTML.
- Servir RDF y HTML con scripts del servidor del cliente.
- Servir datos enlazados desde bases de datos relacionales.
- Servir datos enlazados desde depósitos de tripletas RDF.
- Servir datos enlazados a partir de aplicaciones de empaquetamiento (wrapping) existentes o de APIs de webs.

Se considera que el avance hacia un modelo de datos enlazados y enriquecidos es objetivo estratégico para el MECD por los siguientes motivos:

- Implica posicionar la iniciativa del directorio en las últimas tendencias de la Web Semántica, y es un impulso de este tipo de proyectos en el mundo de la cultura.
- Hay un hueco en "el mercado" para el proyecto directorio; hay pocos proyectos en esta línea y puede ser núcleo de otras iniciativas de interés a nivel Iberoamericano o internacional.
- Facilitará el posicionamiento del directorio a nivel internacional y su reutilización por aplicaciones inteligentes y por los grandes buscadores.
- Dentro de la AGE es función de la Subdirección la coordinación en el campo semántico.

7. Bibliografía

Bennett, Daniel; Harvey, Adam. *Publishing Open Government Data: W3C Working Draft 8 September 2009*, 2009 <<http://www.w3.org/TR/gov-data>>. [Consulta: 29 junio 2012]

Berners-Lee, Tim: "The 5 stars of open linked data". *inkdroid: paper or plastic?*, June 4, 2010. <<http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data>>. [Consulta: 29 junio 2012]

Cyganiak, Richard. *How to publish Open Data*. Presentado en el Opening Up Government Data. Galway, 8 Nov 2011 <<http://www.slideshare.net/cygri/how-to-publish-open-how-to-publish-open-data>>. [Consulta: 29 junio 2012]

"Directiva 98/34/CE del Parlamento Europeo y del Consejo de 22 de junio de 1998, por la que se establece un procedimiento de información en materia de las normas y reglamentaciones técnicas y de las reglas relativas a los servicios de la sociedad de la información". *Diario Oficial de la Unión Europea*, nº L 204, de 21/7/1998, p. 37-48.

"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 de la Unión Europea*, nº L 345 de 31/12/2003, p. 90 -96.

Directorio de Bibliotecas Españolas
<<http://directoriobibliotecas.mcu.es/portada.html>>. [Consulta 29 de junio 2012]

Directorio de Bibliotecas y Hemerotecas Españolas
<<http://www.bne.es/es/Servicios/DirectorioBibliotecas>>. [Consulta: 29 junio 2012]

Guía de aplicación del Real Decreto 1495/2011 por el que se desarrolla la Ley 37/2007 sobre Reutilización de la Información del Sector Público. 1ª edición electrónica. Madrid: Ministerio de Industria, Energía y Turismo; Ministerio de Hacienda y Administraciones Públicas, febrero 2012. 79 p.
<http://www.datos.gob.es/datos/sites/default/files/files/PLANCISP-GRD-07_3_2.doc>. [Consulta: 29 junio 2012]

Hausenblas, Michael. "5 ★ Open Data". <<http://5stardata.info>>. [Consulta: 29 junio 2012]

Heath, Tom; Bizer, Christian. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011. (Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1), 136 p. <<http://linkeddatobook.com/editions/1.0>>. [Consulta: 29 junio 2012]

ISO 2146:2010: Information and documentation: Registry services for libraries and related organizations. Geneva: ISO, cop. 2010.

"Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos". *BOE*, nº 150, 23 de junio de 2007, p. 27150-27166.

"Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, nº 276, 17 de noviembre de 2007, p. 47160-47165.

Linked-Open-Data-Wiki des hbz <<https://wiki1.hbz-nrw.de/display/SEM/Home>>. [Consulta: 29 junio 2012]

Linked Open Vocabularies (LOV)
<<http://labs.mondeca.com/dataset/lov/index.html>>. [Consulta: 29 junio 2012]

Linking Open Bibliographic Data <<http://lobid.org>>. [Consulta: 29 junio 2012]

Peset, Fernanda, Ferrer-Sapena, Antonia, Subirats-Coll, Imma: "Open data y linked open data: Su impacto en el área de bibliotecas y documentación". *El Profesional de la Información*, 20(2), 2011, p. 165-173.

"Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica". *BOE*, nº 25, 29 de enero de 2010, p. 8139-8156.

"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". *BOE* nº 269, 8 de noviembre de 2011, p. 116296-116307.

Towards the Australian Data Commons: A proposal for an Australian National Data Service. Canberra: The ANDS Technical Working Group, 2007.

8. Glosario

Datos abiertos (*open data*): Filosofía y práctica que requiere que determinados datos sean de libre acceso para todo el mundo, sin restricciones de derechos de autor, patentes u otros mecanismos de control. Para distinguirse de los términos más formalmente definidos de código abierto y estándar abierto, los datos abiertos hacen hincapié en el acceso y la reutilización de los datos científicos y el gobierno como un medio para ampliar la colaboración, facilitar la rendición de cuentas del gobierno a los ciudadanos, y acelerar el ritmo de los descubrimientos y de la innovación.

Datos abiertos gubernamentales (*open government data*): Datos abiertos generados en las actividades de entidades de la administración pública.

Datos enlazados (*linked data*): Conjunto de buenas prácticas para la publicación y la interconexión de datos estructurados que facilitan el acceso de los humanos y las máquinas a mediante la utilización de la familia RDF de sintaxis (por ejemplo, RDF / XML) y los URIs de HTTP. Los datos enlazados pueden ser publicados por una persona u por una organización, detrás de un firewall o en la Web pública. Si los datos vinculados se publican en la Web pública, son denominados datos enlazados abiertos (linked open data). Un término sinónimo es datos vinculados.

Esquema Nacional de Interoperabilidad: Comprende los criterios y recomendaciones de seguridad, normalización y conservación de la información, de los formatos y de las aplicaciones que deberán ser tenidos en cuenta por las Administraciones públicas para asegurar un adecuado nivel de interoperabilidad organizativa, semántica y técnica de los datos, informaciones y servicios que gestionen en el ejercicio de sus competencias y para evitar la discriminación a los ciudadanos por razón de su elección tecnológica.

Estándar abierto (*open standard*): Estándar que reúne las siguientes condiciones: es público y su utilización es disponible de manera gratuita o a un coste que no supone una dificultad de acceso; su uso y aplicación no está condicionado al pago de un derecho de propiedad intelectual o industrial.

Formato de intercambio de datos (*data exchange standard*): Una especificación para el almacenamiento, acceso y transmisión de datos. Proporciona la sintaxis necesaria para la expresión de conjuntos de datos. Se distingue de la descripción y representación semántica de los datos, que es llevada a cabo por el vocabulario de datos. En el contexto de este informe son considerados sinónimos los términos formato de intercambio de datos, estándar de intercambio de datos, formato de exposición de datos y formato de publicación de datos.

Interoperabilidad (*interoperability*): Capacidad de los sistemas de información, y por ende de los procedimientos a los que éstos dan soporte, de compartir datos y posibilitar el intercambio de información y conocimiento entre ellos.

Modelo conceptual (*conceptual model*): Proceso y resultado del análisis y representación de los conceptos y relaciones entre conceptos en un dominio determinado. El resultado del modelo conceptual se representa mediante términos de un vocabulario de datos.

Objeto registro (*registry object*): Clase principal del modelo de datos de la norma ISO 2146:2010, que incluye cuatro tipos de objetos diferentes: actividad (operación que ocurre en el tiempo y produce uno o varios resultados); colección (conjunto de objetos físicos o digitales); parte (organismo o persona); y servicio (sistema que proporciona alguna funcionalidad al usuario).

RDFa: Sistema para la incorporación de datos RDF en páginas web codificadas en (X)HTML.

Registro de bibliotecas y organizaciones relacionadas (*registry for libraries and related organizations*): Conjunto de descripciones objetos de registro que han sido compilados para apoyar la actividad o negocio de una comunidad determinada

Reutilización (*reutilization*): Uso de documentos que obran en poder de las administraciones y organismos del sector público, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública. Queda excluido de este concepto el intercambio de documentos entre Administraciones y organismos del sector público en el ejercicio de las funciones públicas que tengan atribuidas.

Vocabularios de datos (*vocabulary to describe data*): Conjuntos de términos que describen conceptos y relaciones entre conceptos, generalmente, en el contexto de un dominio particular. Abarcan los esquemas de metadatos y las ontologías (que maximizan las relaciones contextuales entre conceptos, las restricciones semánticas, etc.)

Web semántica (*semantic web*): Una extensión de la Web actual, que ofrece una manera más fácil de encontrar, compartir, reutilizar y combinar información. Se basa en información legible por máquina y se construye sobre la capacidad de la tecnología XML para definir esquemas de etiquetado personalizados y el enfoque flexible de RDF para representar datos. La Web Semántica proporciona formatos comunes para el intercambio de datos (frente a la Web actual, que se limita al intercambio de documentos). También proporciona un lenguaje común para representar cómo los datos se relacionan con objetos del mundo real, lo que permite a una persona o a una máquina partir de una base de datos, y luego transitar a través de una serie interminable de bases de datos que no están conectados por medio de cables, sino por referirse a la misma cosa.

Para la confección del glosario se han tenido en cuenta las siguientes fuentes:

- *ISO 2146:2010: Information and documentation: Registry services for libraries and related organizations*. Geneva: ISO, 2010.
- "Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos". *BOE*, nº 150, 23 de junio de 2007, p. 27150-27166.
- "Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, nº 276, 17 de noviembre de 2007, p. 47160-47165.
- *Linked Data Glossary: W3C Editor's Draft 26 April 2012* <<https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html>>
- "Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica". *BOE*, nº 25, 29 de enero de 2010, p. 8139-8156.
- *W3C eGovernment (Printable) Glossary* <<http://www.w3.org/egov/wiki/Glossary>>
- *Webopedia* <http://www.webopedia.com/TERM/S/Semantic_Web.html>
- Zeng, Marcia L.; Qin, Jian. *Metadata*. New York: Neal-Schuman, 2008.