



TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL
XIII CONGRESSO ISKO ESPANHA

Universidade de Coimbra, 23 e 24 de novembro de 2017

Com a coordenação de

Maria da Graça Simões, Maria Manuel Borges

TÍTULO

Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES

Maria da Graça Simões
Maria Manuel Borges

EDIÇÃO

Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN

978-972-8627-75-1

ACESSO

<https://purl.org/sci/atas/isko2017>

COPYRIGHT

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

OBRA PUBLICADA COM O APOIO DE



FLUC FACULDADE DE LETRAS
UNIVERSIDADE DE COIMBRA

2



CEIS 20
CENTRO DE ESTUDOS
INTERDISCIPLINARES
DO SÉCULO XX
UNIVERSIDADE DE COIMBRA

FCT

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJETO UID/HIS/00460/2013



RECORDS IN CONTEXTS Y LA PUBLICACIÓN DE CONJUNTOS DE DATOS ARCHIVÍSTICOS INTEROPERABLES

Juan-Antonio Pastor-Sánchez¹, Dunia Llanes Padrón²

¹Universidad de Murcia, 0000-0002-1677-1059, pastor@um.es

²Universidad de La Habana, 0000-0001-8639-4706, duniallp@yahoo.es

RESUMEN Introducción: Este trabajo analiza el modelo conceptual Records in Contexts (RiC-CM) propuesto por el Consejo Internacional de Archivos (ICA) para la descripción archivística, así como su implementación en una ontología OWL y su aplicación para la publicación de datos en el ámbito Linked Data. **Metodología:** Se ha estudiado la estructura y elementos de RiC-CM en relación con normas de descripción ya existentes para la elaboración de la ontología. También se han considerado, las condiciones necesarias para la elaboración y mantenimiento de conjuntos de datos interoperables que contemplen la aplicación de buenas prácticas para la publicación de datos estructurados en la web. **Resultados:** Se muestra un análisis del modelo conceptual RiC-CM junto con una propuesta de ontología OWL para su representación. También se detalla una lista de buenas prácticas para la publicación de conjuntos de datos archivísticos. **Conclusiones:** El trabajo revela el carácter integrador del modelo conceptual de Records in Contexts, así como las posibilidades que ofrece una mayor interoperabilidad de los datos de descripciones archivísticas.

PALABRAS CLAVE Descripción Archivística, Records in Contexts, web semántica, conjuntos de datos, Linked Data

ABSTRACT Introduction: This work analyzes the conceptual model Records in Contexts (RiC-CM) proposed by the International Council of Archives (ICA) for the archival description, as well as its implementation in an OWL ontology and its application for the publication of data in the field Linked Data. **Methodology:** The structure and elements of RiC-CM have been studied in relation to existing norms of description for the elaboration of the ontology. The necessary conditions have also been considered for the development and maintenance of interoperable data sets that include the application of good practices for the publication of structured data on the web. **Results:** An analysis of the RiC-CM conceptual model is presented together with an OWL ontology proposal for its representation. A list of good practices for the publication of archival data sets is also detailed. **Conclusions:** The work reveals the integrative nature of the conceptual model of Records in Contexts, as well as the possibilities offered by greater interoperability of archival descriptions data.

KEYWORDS Archival Description, Records in Contexts, semantic web, datasets, Linked Data

COPYRIGHT Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

INTRODUCCIÓN

El proceso de representación archivística ha cambiado en lo que respecta a sus conceptualizaciones y prácticas debido, principalmente, a las tecnologías basadas en internet y a las nuevas necesidades de acceso a la información por parte de los usuarios. Las nuevas formas de representación de la información abarcan a bibliotecas, archivos, museos y galerías de arte y es la interoperabilidad semántica quien permite relacionar datos estos diferentes ámbitos, ayudando a eliminar las líneas de separación que existen entre dichas instituciones. La elaboración de modelos conceptuales para la descripción archivística es un paso importante para lograr la interoperabilidad semántica de todos los recursos de información.

Después de la intensa actividad normativa desarrollada por el *International Council Archives* (ICA) y algunas comunidades nacionales en la década de los 90 del pasado siglo, comenzaron a surgir los primeros modelos conceptuales para la Archivística. Los nuevos preceptos no sólo incluyen las relaciones entre las sub-entidades del documento (descripción multinivel) sino que analizan los vínculos de estos con su contexto, analizando su producción, gestión y conservación. Estos modelos son conocidos como pluridimensionales pues incluyen diversas entidades, sus relaciones y definen propiedades.

Los primeros modelos nacionales para la descripción de objetos archivísticos fueron publicados en Australia, Nueva Zelanda, España y Finlandia¹ y tuvieron gran impacto tanto a nivel nacional como en el contexto internacional. El ICA, motivado por el auge de la normalización y por la aplicación de los estándares, en el 2012, formó un grupo de expertos para desarrollar un modelo conceptual internacional en materia de archivos. En 2016 se publica el primer borrador de “*Records in Contexts: A Conceptual Model for Archival Description*” (RiC-CM). Este modelo completa, amplía e interrelaciona las normas ISAD(G), ISAAR(CPF), ISDF e ISDIAH y se adapta a las nuevas posibilidades que brindan las tecnologías de la información y las comunicaciones para integrar datos y servicios en la web.

Los modelos conceptuales archivísticos, en general, representan e interrelacionan las entidades involucradas en el proceso descriptivo, facilitando la interoperabilidad entre registros de información de archivos, bibliotecas, museos o cualquier otra institución cultural. En este sentido, la interoperabilidad semántica permite que los sistemas de información intercambien datos directamente (Sheth, 1999, p. 4; Hernández-Pérez et al., 2009, p. 68) facilitando la reutilización de la información y aumentando la capacidad para generar y transformar datos, contenidos, conocimiento, productos y servicios, armonizando datos y procesos (Tolk y Muguira, 2003, p. 2).

¹ Australian Government Recordkeeping Metadata Standard. (Australia); Technical Specifications for the Electronic Recordkeeping Metadata Standard. (Nueva Zelanda); Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones. (España); Finnish Conceptual Model for Archival Description (Finlandia).

La web ofrece diversas tecnologías para acceder, publicar y compartir información siguiendo los principios de la interoperabilidad semántica. RDF es un modelo de datos que describe las propiedades y relaciones de los recursos de información y junto con RDF Schema y OWL permiten definir ontologías con un alto grado de expresividad semántica. El uso de RDF para publicar datos archivísticos permitiría una interoperabilidad mayor que la que ofrecen formatos como EAD o EAC. Los modelos conceptuales, como Records in Contexts, pueden implementarse con RDFS y OWL para lograr la interoperabilidad semántica en la web de los conjuntos de datos archivísticos. Además, las tecnologías de la web semántica y los servicios basados en Linked Data ofrecen una serie de ventajas evidentes para los Archivos (Baker et al., 2011; Sulé et al., 2016, p. 12; Hidalgo-Delgado et al., 2016, p. 2; McKenna, 2013, p. 352).

Este trabajo analiza el Modelo Conceptual Records in Contexts (RiC-CM) de Descripción Archivística elaborado por el International Council of Archives (ICA) al tiempo que se propone su adaptación en forma de ontología OWL y la publicación de conjuntos de datos archivísticos adecuados a las buenas prácticas de publicación de datos en la web. Para ello, se realiza un estudio de RiC-CM de los propósitos, entidades y relaciones de este. También se presentan los criterios de modelado de la ontología OWL propuesta y la aplicación de las buenas prácticas para la publicación de datos en la web recomendada por el W3C.

METODOLOGÍA

En un primer momento se procedió a la contextualización del trabajo en relación a la actividad previa realizada con respecto a la normativa de descripción. Igualmente se revisaron las tecnologías más relevantes de la web semántica y su vinculación con la interoperabilidad de conjuntos de datos. Se ha realizado un análisis del borrador del modelo conceptual propuesto por el ICA, abarcando las entidades, las propiedades y las relaciones de RiC-CM.

A partir de dicho análisis se ha procedido al modelado de la ontología OWL. En una primera versión se realizó una implementación directa del modelo conceptual. Una segunda versión de la ontología se centró en la optimización de propiedades y relaciones. Para ello se unificaron algunas propiedades de datos y de objeto de la primera versión y se definieron doblemente algunas propiedades de datos para dar soporte a la representación de datos a partir de la vinculación de recursos de términos de listas y vocabularios controlados.

Finalmente se procedió a un análisis de las buenas prácticas de datos en la web y se identificaron aquellas cuya aplicación resultan más pertinentes para la publicación de conjuntos de datos archivísticos.

RECORDS IN CONTEXTS: EL NUEVO MODELO CONCEPTUAL PARA LOS ARCHIVOS

En RiC-CM se establecen los propósitos esenciales de la descripción archivística en los contextos actuales de posmodernidad. El ICA propone tres roles fundamentales: Gestión Documental (*Management of Records*); Preservación de Documentos (*Preservation of Records*) y Reutilización de datos/información (*Ongoing Use and Reuse of Records*), (ICA, 2016, p. 14-15). La descripción se adapta a los nuevos tiempos, donde la representación de la información archivística a partir de modelos

de datos es determinante para la gestión, la preservación digital y la reutilización e interoperabilidad de datos en la web semántica. Representar las funciones, actividades y transacciones que dan origen a la documentación, también, facilita el uso continuo de los registros y conserva la memoria colectiva de las naciones, aspectos determinantes para la gestión. El ICA, al pronunciarse por una descripción encauzada a la gestión, sitúa a esta actividad en un lugar indispensable dentro de la gestión integrada de procesos en las organizaciones.

Según el RiC-CM, al describir entidades archivísticas, a través de metadatos, se posibilita el mantenimiento de la autenticidad e integridad de los documentos. En el modelo se expone que “documentar el contexto describiéndolo es esencial para la conservación de los registros” (ICA, 2016). Por último, el texto hace referencia a la importancia de la descripción para la reutilización de los datos archivísticos en entornos web. Además, la reutilización de información está asociada al acceso, la calidad y la usabilidad de datos públicos abiertos (*Open Data Governmet*) permiten la creación de nuevos datos, contenidos, conocimientos, productos o servicios.

El ICA, partiendo de sus cuatro estándares y de otros modelos de la Ciencias de la Información ha propuesto un conjunto de entidades que representan la realidad de la descripción archivística: documento, componente documental, conjunto documental, agente, ocupación, posición, función, función abstracta, actividad, norma, forma documental, fechas, lugar, concepto/asunto. El RiC-CM contempla las entidades definidas en ISAD(G) (Documento), en ISAAR(CPF) e ISDIAH (Agente) y en ISDF (funciones - actividades), unificando todos los componentes del conjunto normativo. Asimismo, se puede apreciar que RiC-CM tiene entidades coincidentes con los *Functional Requirements for Bibliographic Records* y en el *Conceptual Reference Model (CRM) del International Council of Museums (ICOM)* y en los modelos de Australia, Nueva Zelanda, España y Finlandia; las entidades documento, agente, función, norma, lugar, período y concepto son denominadores comunes en los modelos conceptuales desarrollados en las CI y constituyen entidades principales.

Este modelo presta gran importancia al contexto de producción de los documentos y así lo deja explícito tanto en sus propósitos como con las entidades función, función abstracta, ocupación, actividad, posición o mandato. Con ello se define el contexto de creación de los records, garantizando la veracidad, integridad y autenticidad de los documentos archivísticos. Además, define entidades como fecha, lugar, concepto o asunto esenciales para la creación de los puntos de acceso normalizados determinantes en el proceso de búsqueda y recuperación de información.

Las 69 propiedades definidas en el modelo son otro aspecto determinante en *Records in Contexts*, para la descripción de las 14 entidades archivísticas, reduciendo de esta forma la cantidad de atributos, con respecto al conjunto normativo. Por lo tanto se pretende presentar descripciones completas pero a su vez sucintas. El esquema de elementos de RiC-CM está basado en los trabajos normativos internacionales, nacionales y regionales desarrollados durante más de 20 años, además de las experiencias prácticas a través del uso de determinados software.

La ejemplificación de las relaciones entre las entidades es uno de los temas más trabajados y conseguidos en este nuevo estándar conceptual. La determinación de una amplia variedad de relaciones garantiza la interrelación de los contenidos de los documentos con los diversos ambientes en los que fueron producidos. En el RiC-CM se definen cerca de 800 relaciones (pueden aumentar en el texto definitivo) para definir las asociaciones entre entidades para incrementar la capacidad de búsqueda, recuperación y uso de la información. La figura siguiente muestra un ejemplo de descripción archivística conforme las relaciones y las entidades definidas en el RIC-CM.

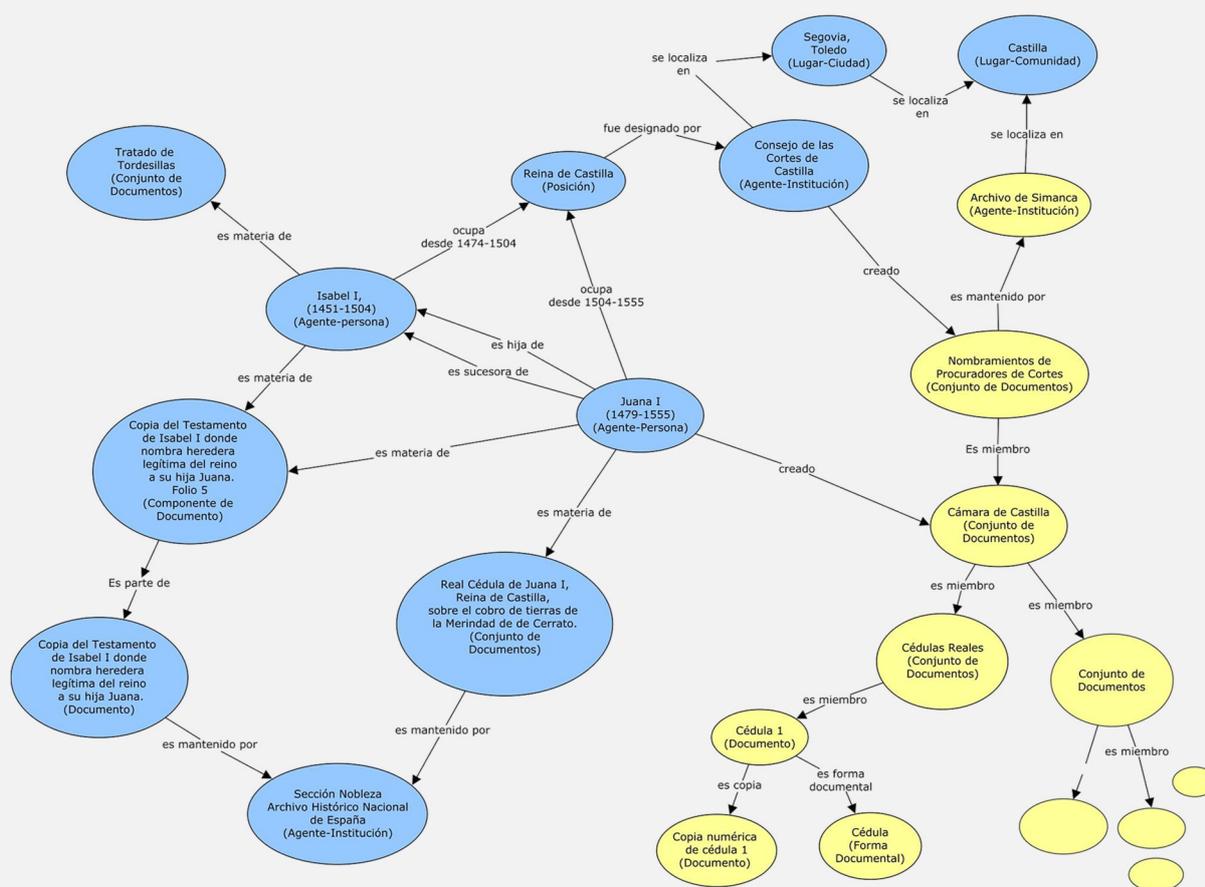


Figura 1. Ejemplo de descripción archivística a partir del RIC-CM.

Fuente: Llanes-Padrón & Moro-Cabero, 2017, p. 532

La descripción archivística, a partir del RIC-CM, facilita la consulta de recursos diversos entrelazados a partir de un enfoque único y como resultado permite descubrir y compartir más información en un modelado de red. El desarrollo de una ontología de este modelo, basada en RDF y OWL es el punto concluyente para posicionar a los recursos archivísticos en la dinámica de la publicación directa en la web, los datos abiertos entrelazados, la reutilización de datos y la interoperabilidad con sistemas de información de instituciones diversas.

INTEROPERABILIDAD, REUTILIZACIÓN Y TECNOLOGÍAS DE LA WEB SEMÁNTICA

En entornos web, las descripciones de recursos de información se realizan mediante modelos descriptivos y sus respectivos vocabularios cuya implementación suele realizarse a través de esquemas de metadatos y ontologías. Los metadatos son descripciones estructuradas y codificadas para describir características y propiedades de objetos y recursos (Gilliland-Swetland, 2003). Las ontologías proporcionan un modelo lógico-conceptual común en un área de conocimiento, definiendo, a diferentes niveles de formalización, el significado de las clases de objetos, las relaciones entre ellos y las propiedades que los caracterizan (Gruber, 1995).

Los metadatos y las ontologías requieren de tecnologías concretas para su aplicación práctica. Bajo la denominación de web semántica se agrupan un conjunto de especificaciones, tecnologías, desarrollos y aplicaciones para acceder, publicar y compartir información siguiendo los principios de la interoperabilidad semántica que permiten representar el significado de los datos. Se sigue haciendo uso del protocolo HTTP (*HyperText Transfer Protocol*) para establecer las conexiones entre clientes y servidores y las referencias IRI (*Internationalized Resource Identifier*) permiten identificar de forma inequívoca cualquier recurso. Cuando una IRI puede utilizarse para localizar y acceder a un recurso se dice que está desreferenciada como URL (*Uniform Resource Locator*).

RDF (*Resource Description Framework*) es un modelo de datos que aporta un marco semántico para describir recursos mediante sentencias que forman tripletas (*triplets*) compuestas por un sujeto, un predicado y un objeto. Tanto los recursos como las propiedades se identifica de manera unívoca mediante referencias IRI definidas dentro de un espacio de nombres XML propio (que pueden abreviarse mediante prefijos). Las descripciones se realizan mediante la asignación de valores a determinados atributos mediante propiedades de datos, o con el establecimiento de relaciones entre dos recursos utilizando propiedades de objeto (Cyganiak et al., 2014). Las sentencias RDF se almacenan utilizando algún formato legible por máquina como RDF/XML, Turtle, JSON-LD o RDFa entre otros (Gandon y Schreiber, 2014; Prud'hommeaux et al., 2014; Sporny et al., 2014; Herman et al., 2015). RDF se amplía con RDF Schema y con OWL (*Ontology Web Language*) para definir jerarquías de clases y propiedades, restricciones de cardinalidad, equivalencia entre clases, relaciones inversas, etc (Brickley, Guha y McBride, 2014; Hitzler et al., 2012).

PROPUESTA DE ONTOLOGÍA OWL BASADA EN EL MODELO CONCEPTUAL DE RECORDS IN CONTEXTS

La definición de ontologías OWL a partir de modelos conceptuales han sido propuestas ampliamente para FRBR (Ciccarese y Peroni, 2011; Shotton y Peroni, 2016), RDA (Hillmann et al., 2010) y CIDOC-CRM (Reinhardt, 2009). El EGAD está desarrollando una ontología OWL denominada RiC-O para implementar el modelo conceptual. La propuesta recogida en este trabajo puede contribuir a un debate que haga de RiC-O una ontología más eficiente y se encuentra disponible para su descarga y referencia en el espacio de nombres <http://purl.org/ica/ric>².

La ontología OWL modela las entidades del modelo conceptual mediante 14 clases principales sobre las que se definen las correspondientes propiedades. El borrador del RiC-CM especializa la entidad “RiC-E4 Agent” al definir las relaciones (ICA, 2016, p. 30-90). De esta forma se han definido tres subclases específicas para la entidad “RiC-E4 Agent” (agentes delegados, grupos y personas). Teniendo en cuenta estas consideraciones la Tabla 1 muestra una propuesta de taxonomía de clases para la ontología.

² La documentación de la ontología se encuentra disponible en <http://skos.um.es/TR/ric>

Tabla 1. Modelado de clases OWL a partir de las entidades de RiC-CM. Fuente: elaboración propia.

Entidad RiC	Denominación	Clase OWL
RiC-E1	Record	ric:RiC-E1-Record
RiC-E2	Record Component	ric:RiC-E2-Record Component
RiC-E3	Record Set	ric:RiC-E3-Record Set
RiC-E4	Agent	ric:RiC-E4-Agent
	Delegate Agent (subtipo de Agent)	ric:RiC-E4-Agent-DelegateAgent
	Person (subtipo de Agent)	ric:RiC-E4-Agent-Person
	Group (subtipo de Agent)	ric:RiC-E4-Agent-Group
RiC-E5	Occupation	ric:RiC-E5-Occupation
RiC-E6	Position	ric:RiC-E6-Position
RiC-E7	Function	ric:RiC-E7-Function
RiC-E8	Function (Abstract)	ric:RiC-E8-FunctionAbstract
RiC-E9	Activity	ric:RiC-E9-Activity
RiC-E10	Mandate	ric:RiC-E10-Mandate
RiC-E11	Documentary Form	ric:RiC-E11-DocumentaryForm
RiC-E12	Date	ric:RiC-E12-Date
RiC-E13	Place	ric:RiC-E13-Place
RiC-E14	Concept/Thing	ric:RiC-E14-ConceptThing

Fuente: Elaboración propia.

La definición de tipos más específicos para las subclases que representan personas, grupos y agentes-delegados se realiza mediante la propiedad “RiC-P32 Type” (ric:RiC-P32-type en la ontología propuesta).

La mayor parte de las propiedades de RiC-CM son específicas para ciertas entidades y algunas de ellas tienen una semántica similar e incluso idéntica. En estos casos, se han modelado como una única propiedad cuyo dominio es la unión de las clases de las entidades correspondientes. Este es el caso de las propiedades “Type”³ cuya semántica no es exactamente la misma puesto que su dominio es distinto. Sin embargo, estas propiedades tienen un objetivo común: la categorización de un recurso dentro de una clasificación o lista de tipos específica de cada entidad. En consecuencia, la ontología propuesta considera que la semántica de dichas propiedades es la misma. De esta forma se define una única propiedad ric:RiC-P-type cuyo dominio es la unión de las oportunas clases (ver Tabla 2).

3 Los números de estas propiedades en el modelo conceptual son: RiC-P23, RiC-P32, RiC-P42, RiC-P45, RiC-P48, RiC-P52, RiC-P55, RiC-P58, RiC-P61, RiC-P63 y RiC-P66.

Tabla 2. Definición del dominio de la propiedad ric:RiC-P-type

1	ric:RiC-P-type rdf:type owl:ObjectProperty ;
2	rdfs:domain [rdf:type owl:Class ;
3	owl:unionOf (ric:RiC-E10-Mandate
4	ric:RiC-E11-DocumentaryForm
5	ric:RiC-E12-Date
6	ric:RiC-E13-Place
7	ric:RiC-E14-ConceptThing
8	ric:RiC-E3-RecordSet
9	ric:RiC-E4-Agent
10	ric:RiC-E5-Occupation
11	ric:RiC-E6-Position
12	ric:RiC-E7-Function
13	ric:RiC-E9-Activity)
14] .

Fuente: *Elaboración propia.*

El rango de las propiedades de datos se representan mediante literales RDF (rdfs:Literal). Sin embargo, RiC-CM establece el rango de algunas propiedades a partir de los términos de un vocabulario controlado. Por ello, y considerando que OWL no considera mutuamente excluyentes a owl:DatatypeProperty y owl:ObjectProperty, se han definido este tipo de propiedades como miembros de ambas clases representando las 69 propiedades de RiC-CM mediante 38 propiedades de datos y 15 propiedades de objeto.

También se reduce el número de relaciones de RiC-CM puesto que muchas de ellas tienen una semántica prácticamente idéntica, pudiendo definirse en OWL como una única propiedad de objeto. Esto puede verse en “is member of” que se utiliza con las entidades que representan documentos, conjuntos de documentos y agentes para la que el modelo conceptual define tres relaciones (Tabla 3).

Tabla 3. Relaciones “is member of” de RiC-CM

Número de relación	Nombre	Dominio	Rango
RiC-R18	is member of	RiC-E1 Record	RiC-E3 Record Set
RiC-R143	is member of	RiC-E3 Record Set	RiC-E3 Record Set
RiC-R266	is member of	RiC-E4 Agent	RiC-E4 Agent (of type group)

Fuente: *elaboración propia a partir de ICA, 2016.*

Esta relación, únicamente permite vincular documentos con conjuntos de documentos, conjuntos de documentos con otros conjuntos de documentos y agentes con otros agentes del tipo grupo. En este caso, la ontología propuesta define una única propiedad de objeto ric:RiC-R-isMemberOf cuyo dominio sería la unión de las clases ric:RiC-E1-Record, ric:RiC-E3-RecordSet y ric:RiC-E4-Agent.

Tabla 4. Modelado de las relaciones “is member of”.

1	ric:RiC-R-isMemberOf rdf:type owl:ObjectProperty ;
2	rdfs:domain [
3	a owl:Class ;
4	owl:unionOf (
5	ric:RiC-E1-Record
6	ric:RiC-E3-RecordSet
7	ric:RiC-E4-Agent
8)
9] ;
10	rdfs:range [
11	a owl:Class ;
12	owl:unionOf (
13	ric:RiC-E3-RecordSet
14	ric:RiC-E4-Agent-Group
15)
16] ;
17	rdfs:comment "Relations RiC-R18, RiC-R143, RiC-R266"@en ;
18	rdfs:label "is member of"@en .

Fuente: Elaboración propia.

No obstante, puesto que el rango de las relaciones puede variar en función de las clases sobre las que se definen la unificación de este tipo de relaciones con OWL requiere el establecimiento de restricciones para limitar el uso de una propiedad a los objetos de las clases correspondientes (Tabla 5). En consecuencia, las 792 relaciones del borrador del modelo conceptual se definen con 166 propiedades de objeto en la ontología propuesta.

Tabla 5. Definición de restricción de la propiedad ric:RiC-R-is Member Of sobre la clase RiC-E1-Record.

1	ric:RiC-E1-Record rdfs:subClassOf [rdf:type owl:Restriction ;
2	owl:onProperty ric:RiC-R-isMemberOf ;
3	owl:allValuesFrom ric:RiC-E3-RecordSet
4]

Fuente: Elaboración propia.

Por su parte, las propiedades “RiC-P68 Date” y “RiC-P69 Place” permiten describir las relaciones entre entidades. Por ello se han modelado como propiedades de anotación para su uso mediante reificaciones RDF.

PUBLICACIÓN DE CONJUNTOS DE DATOS ARCHIVÍSTICOS

Además de una ontología basada en RiC-CM también se precisa definir una serie de características que deben reunir los conjuntos de datos archivísticos. Un conjunto de datos debe ser una unidad autodescriptiva con datos estructurados disponibles en varios formatos para su descarga completa o acceso selectivo. Los metadatos que describen diversos aspectos de un conjunto de datos deben proporcionar información sobre su contenido y estructura, así como sobre sus diferentes versiones y distribuciones, para facilitar la confianza en los datos y su reutilización.

Los principios Linked Open Data (LOD) tienen como objetivo la interconexión de conjuntos de datos, utilizando IRIs desreferenciables (convertibles a URLs) durante la identificación y acceso a la

descripción de recursos mediante el protocolo HTTP (Berners-Lee, 2006; Haslhofer y Schandl, 2010). Sin embargo los escenarios de acceso y consumo de datos RDF son muy variables: en algunos casos se precise descargarlo en su totalidad y en otros utilizar SPARQL para un acceso selectivo (Harris, Seaborne y Prud'hommeaux, 2013) o utilizar una API especializada.

Esta diversidad de casos de uso exige un análisis detenido del proceso de publicación de un conjunto de datos. El W3C ha creado un grupo para trabajar en la identificación de diversos escenarios y retos a partir de los cuales se han identificado y descrito 35 buenas prácticas para la publicación de datos en la web (Farias, Burle y Calegari, 2016) que pueden resumirse en los siguientes puntos:

- Uso de IRIs persistentes para identificar tanto a las distintas versiones de un dataset como a los datos que contiene.
- Reutilización de vocabularios de datos para incrementar la interoperabilidad semántica y escogiendo niveles adecuados de modelado y formalización.
- Uso preferente de vocabularios y formatos estándar para que los datos sean legibles tanto por personas como por máquina, haciendo más flexible su reutilización.
- Enriquecimiento de los datos para aportar valor añadido a través del uso de otros conjuntos, bases de datos u otros recursos de información.
- Proporcionar acceso en tiempo real a datos actualizados, no solamente descargando todo el conjunto de datos, sino también mediante servicios web y APIs bien documentadas.
- Garantizar la privacidad y la seguridad de los datos sensibles, identificando qué partes de los conjuntos de datos no están disponibles públicamente e indicando el motivo.
- Informar a los usuarios y consumidores de los datos acerca del modo para contactar con los editores y publicar los resultados de dicha retroalimentación como datos abiertos.
- Incluir una licencia de uso del conjuntos de datos para que los usuarios estén informados sobre las limitaciones y posibilidades de reutilización de los datos.
- Incluir información sobre la procedencia del contenido del conjuntos de datos para que pueda evaluarse con la finalidad de medir su fiabilidad y confianza.
- Gestionar adecuadamente las distintas versiones de un conjunto de datos, evitando modificar APIs o servicios de acceso a los datos que obliguen a introducir cambios en los clientes que los utilicen.

La aplicación de estas buenas prácticas aporta una serie de beneficios en la publicación de datos en la web y una mejor comprensión de su estructura, naturaleza, significado, y mecanismos de acceso (Lee, Farias y Archer, 2015). Todo esto, deberá ser tenido en cuenta a nivel técnico, para la implementación o adopción de plataformas de publicación y consulta de conjuntos de datos archivísticos.

CONCLUSIONES

RiC-CM expone la importancia de la descripción para la reutilización de los datos archivísticos en la web y sienta las bases para que los archivos se beneficien de las ventajas de las tecnologías de la web semántica en lo concerniente a interoperabilidad, reutilización de datos y servicios basados en Linked Open Data. Además, la mayor parte de las entidades se encuentran en otros modelos conceptuales en el ámbito de las Ciencias de la Información lo que acentúa la idea de crear un modelo semántico integral

que facilite la interoperabilidad de los contenidos de todas las instituciones de información en el contexto web apoyados en desarrollos ontológicos.

El desarrollo de una ontología de este modelo, basada en el Esquema de Datos para la Web Semántica (RDF) es el punto concluyente para posicionar a los recursos archivísticos en la dinámica de la publicación directa en la web, los datos abiertos entrelazados, la reutilización de datos y la interoperabilidad con sistemas de información de instituciones diversas. Los conjuntos de datos archivísticos interoperables y de confianza son una aportación muy valiosa en el universo Linked Open Data y a su vez pueden enriquecerse enormemente mediante enlaces con vocabularios controlados SKOS, entidades de DBpedia, registros bibliográficos, etc. También será preciso alinear cualquier ontología basada en RiC-CM con otros vocabularios de metadatos y ontologías ya existentes como FOAF, SKOS o Dublin Core.

Cuando el EGAD termine sus trabajos y el ICA publique las versiones definitivas del modelo conceptual y de la ontología será el momento en el que las herramientas de gestión y descripción archivística deberán adaptarse a esta propuesta. El futuro dirá la repercusión de Records in Contexts en la publicación de datos archivísticos en la web. La publicación final de RiC trae consigo un nuevo desafío para la Ciencia de la Información que transita rumbo a la creación de un único modelo conceptual de descripción de recursos de información. Cada vez se avanza más en la ruptura de las líneas teóricas y prácticas que separan los procedimientos de trabajo en archivos, bibliotecas, museos y otras instituciones culturales. Sin lugar a dudas, los modelos conceptuales, las ontologías y los formatos son el camino para lograr la compatibilidad e interoperabilidad total de la información patrimonial y cultural en el contexto de la web semántica.

REFERENCIAS BIBLIOGRÁFICAS

Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., ... Zeng, M. (2011). Library Linked Data Incubator Group Final Report. Recuperado de <https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>

Berners-Lee, T. (2006). Linked Data. Recuperado de <http://www.w3.org/DesignIssues/LinkedData.html>

Brickley, D., & Guha, R. V. (2014). RDF Schema 1.1. Recuperado de <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>

Brickley, D., Guha, R.V., & McBride, B. (2014). RDF Schema 1.1. Recuperado de <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>

Ciccarese, P., & Peroni, S. (2011). Essential FRBR in OWL2 DL. Recuperado de <http://purl.org/spar/frbr>

CNEDA. (2017). NEDA-MC. Modelo conceptual de descripción archivística: Entidades, relaciones y atributos. Recuperado de <https://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/NEDA-MC-2017/NEDA-MC-2017.pdf>

Cyganiak, R., Wood, D., & Lanthaler, M. (2014). RDF 1.1 Concepts and Abstract Syntax.

Farias Lóscio, B., Burle, C., & Calegari, N. (2016). Data on the Web Best Practices. Recuperado de <https://www.w3.org/TR/2017/REC-dwbp-20170131/>

Gandon, F., & Schreiber, G. (2014). RDF 1.1 XML Syntax. Recuperado de <http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>

Gilliland-Swetland, A. J. (2003). Metadata - Where Are We Going?. En G.E. Gorman (ed.) *International yearbook of library and information management 2003-2004: Metadata applications and management* (pp. 17-33). London: Facet Publishing.

Gruber, T. R. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies* 43(5-6), 907-928. Recuperado de <http://doi.org/10.1006/ijhc.1995.1081>

Gueguen, G., Fonseca, V. M. M. da, Pitti, D. V., & Sibille-de Grimouard, C. (2013). Hacia un Modelo Conceptual Internacional de Descripción Archivística: Un informe preliminar del Grupo de Expertos en Descripción Archivística del Consejo Internacional de Archivos. *Métodos de información*, 4(7), 155-172. <https://doi.org/10.5557/IIMEI4-N7-155172>

Haslhofer, B., & Schandl, B. (2010). Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies archive*, 5(1), 17-31. Recuperado de <http://dx.doi.org/10.1504/IJMSO.2010.032648>

Harris, S., Seaborne, A., & Prud'hommeaux, E. (2013). SPARQL 1.1 Query Language. Recuperado de <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/>

Herman, I., Adida, B., Sporny, M., & Birbeck, M. (2015). RDFa 1.1 Primer: Rich Structured Data Markup for Web Documents (Third Edition). Recuperado de <http://www.w3.org/TR/2015/NOTE-rdfa-primer-20150317/>

Hernández-Pérez, T., Rodríguez-Mateos, D., Martín-Galán, B., & García-Moreno, M. A. (2009). El uso de metadatos en la administración electrónica española: los retos de la interoperabilidad. *Revista española de Documentación Científica*, 32(4), 67-91.

Hidalgo-Delgado, Y., Senso, J. A., Leiva-Mederos, A., & Hípola, P. (2016). Gestión de fondos de archivos con datos enlazados y consultas federadas. *Revista española de Documentación Científica*, 39(3), 145.

Hillmann, D., Coyle, K., Phipps, J., & Dunsire, G. (2010). RDA Vocabularies: Process, Outcome, Use. *D-Lib Magazine*, 16(1/2). <https://doi.org/10.1045/january2010-hillmann>

Hitzler, P., Krötzsh, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012). OWL 2 Web Ontology Language Primer (Second Edition). Recuperado de <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

ICA. (2000). ISAD(G): General International Standard Archival Description. Ottawa: International Council on Archives. Recuperado de http://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf

- ICA. (2016). Records in Contexts: A conceptual model for archival description (Consultation Draft v0.1). Recuperado de <http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>
- Lee, D., Farias Lóscio, B., & Archer, P. (2015). Data on the Web Best Practices Use Cases & Requirements. Recuperado de <https://www.w3.org/TR/2015/NOTE-dwbp-ucr-20150224/>
- Llanes-Padrón, D., Moro-Cabero, M. (2017). Records in Contexts: un nuevo modelo para la representación de la información archivística en el entorno de la web semántica. *El profesional de la información*, 26(3), 525-533.
- McKenna, G. (2013). Linked Heritage Experience in Linking Heritage Information. *JLIS.it*, 4(1), 325. <https://doi.org/10.4403/jlis.it-6304>
- Prud'hommeaux, E., Carothers, G., Beckett, D., & Berners-Lee, T. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. Recuperado de <http://www.w3.org/TR/2014/REC-turtle-20140225/>
- Reinhardt, S. (2009). CIDOC CRM in OWL 2. Recuperado de <http://bloody-byte.net/rdf/cidoc-crm/>
- Schneider, M. (2012). OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition). Recuperado de <https://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/>
- Sheth, A. P. (1999). Changing focus on interoperability in information systems: from system, syntax, structure to semantics. En M. Goodchild, M. J. Egenhofer, R. Fegeas, & C. Kottman (Eds.), *Interoperating geographic information systems* (pp. 5-29). Norwell, MA, USA: Kluwer Academic Publishers. Recuperado de <http://knoesis.org/sites/default/files/S98-changing.pdf>
- Shotton, D., & Peroni, S. (2016), FaBiO, the FRBR-aligned Bibliographic Ontology. Recuperado de <http://purl.org/spar/fabio/>
- Sporny, M., Kellogg, G., Lanthaler, M., & Lindström, N. (2014). JSON-LD 1.0: A JSON-based Serialization for Linked Data. Recuperado de <http://www.w3.org/TR/2014/REC-json-ld-20140116/>
- Sulé, A., Centelles, M., Franganillo, J., & Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. *Revista española de Documentación Científica*, 39(1), 121. <https://doi.org/10.3989/redc.2016.1.1268>
- Tolk, A., & Muguira, J. A. (2003). The Levels of Conceptual Interoperability Model. En *Fall Simulation Interoperability Workshop*, Orlando, Florida, September 2003. Recuperado de <https://pdfs.semanticscholar.org/f655/af160f630b9be8dbab986f6a96953aa3e986.pdf>