

## Sobre la interoperabilidad semántica en las descripciones archivísticas digitales

*On the semantic interoperability of digital archival descriptions*

Salvador Sánchez Alonso\*, Miguel Ángel Sicilia Urbán\* y Gonzalo de Rato Leguina\*

**Resumen:** Como en otras disciplinas científicas, en la archivística resulta de gran importancia la existencia de estándares, tanto para facilitar el intercambio de información entre entidades como para normalizar el acceso a dicha información. Tomando como punto de partida las normas vigentes para la elaboración de descripciones archivísticas, ISAD(G) e ISAAR (CPF), este artículo presenta un trabajo orientado a facilitar la interoperabilidad semántica (concepto que cobra su pleno sentido en el contexto de la *Web Semántica*) de las descripciones archivísticas. Este propósito se pretende conseguir mediante la representación de ISAD(G) en un lenguaje de ontologías específicamente orientado a la Web, y estableciendo correspondencias con ontologías de conocimiento general. El objetivo es proporcionar servicios y funcionalidades de acceso avanzadas que mejoren las que actualmente proporcionan los puntos de acceso normalizado, formalizando las descripciones archivísticas para su manejo por programas de computadora.

**Palabras clave:** descripciones archivísticas, web semántica, ISAD (G), ontología.

**Abstract:** The international archival community relies on the existence of standards as the means to both reliably interchange information and to promote its accessibility. Departing from current standards for the description of archival entities and authority records, namely ISAD(G) and ISAAR(CPF), the work described herein establishes a means to make semantic interoperability possible for archival descriptions under ISAD(G) in the context of the *Semantic Web*. For the purpose of this work, semantic interoperability is understood as the use of explicit semantics to facilitate archival descriptions integration with the main

---

\* Information Engineering Research Unit, Universidad de Alcalá.  
Correo-e: {salvador.sanchez, msicilia}@uah.es , gonzalorato@gmail.com  
Recibido: 25-11-05; 2.<sup>a</sup> versión: 16-1-06; 3.<sup>a</sup> versión: 9-11-06; 4.<sup>a</sup> versión: 3-7-07.

objective of fostering the automated or semi-automated use of the information. This paper introduces a mapping of ISAD(G) archival descriptions to an ontology written in a Web-oriented ontology language, as the main artefact to promote semantic interoperability in archival descriptions. It has been achieved through the introduction of a common ground for the definition of concepts, based on the use of shared definitions from in widely-used upper ontologies. The final aim is to provide advanced services and access facilities which can serve as improved access points in archival description.

**Keywords:** archival descriptions, semantic web, ISAD(G), ontology.

## 1. Introducción

En el campo de la archivística, la normalización resulta clave para el intercambio de información. Al igual que en otras disciplinas, existe la necesidad de un lenguaje y unas reglas de descripción, que en este caso permiten normalizar el conjunto de documentos que componen un archivo. El objetivo es superar momentos históricos donde, tal y como apuntaba Antonia Heredia, se daba prioridad a la descripción y clasificación de los documentos sobre las necesidades de recuperación: «...*Esta situación se agrava cuando por medio están las fuentes documentales, de las que introductores y usuarios se valen y no se utiliza un mismo lenguaje para entenderse...*» (Heredia, 1992). Para alcanzarlo se han realizado numerosos esfuerzos de normalización y estandarización, algunos de los más relevantes de los cuales son:

- El Estándar internacional para la descripción general de archivos, ISAD(G), que proporciona directrices para la elaboración de descripciones archivísticas normalizadas. El objetivo de la descripción es identificar y explicar el contexto y contenido de los documentos de archivo con el propósito de hacerlos accesibles. ISAD(G) contiene reglas generales para la descripción archivística que pueden aplicarse con independencia del tipo de documento o del soporte físico. Dichas reglas identifican y definen 26 elementos que pueden combinarse entre sí para constituir la descripción de una entidad archivística.
- La Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias, ISAAR(CPF), complemento de ISAD(G) cuya finalidad es proporcionar reglas para normalizar descripciones de personas y entidades productoras y gestoras de documentación de archivos. Concedida para permitir compartir los registros de autoridad, facilita entre otros el acceso a los archivos a partir de la descripción de su contexto de producción y la identificación precisa de productores de documentos.
- Las reglas para la creación de *Descripciones Archivísticas Codificadas* (EAD) que permiten transferir la información normalizada, adecuando las normas de descripción al proceso técnico necesario para su tratamiento y difusión informatizada. EAD define una estructura de datos normalizada que permite inter-

cambiar instrumentos de descripción y acceder a ellos, determinando un formato de comunicación que facilita a los archivos distribuir electrónicamente instrumentos de descripción. Al proponer principios independientes de las plataformas informáticas utilizadas, facilita la aparición de herramientas para buscar, recuperar, mostrar y navegar por instrumentos de descripción. EAD es completamente compatible con ISAD(G), utiliza el lenguaje SGML (*Standard Generalized Markup Language*) y es conforme con XML, lo cual facilita su tratamiento en Internet.

Las normas y estándares anteriores cubren diferentes aspectos, importantes todos ellos, de la normalización en el área de la archivística. Tomando estos esfuerzos como punto de partida, el objeto del presente trabajo es avanzar un paso más, hacia la denominada *interoperabilidad semántica*. Se trata de un concepto que tiene sentido en el contexto de la *Web Semántica* (Berners-Lee et al., 2001), donde las descripciones (en el caso presente, las descripciones archivísticas) están preparadas para su manejo por parte de programas de computadora que puedan hacer uso de representaciones del conocimiento con el objeto de proporcionar servicios y funcionalidades de acceso avanzadas. Para ello, es preciso contar con representaciones formales del conocimiento especialmente orientadas a su tratamiento automatizado (lo que en terminología inglesa se conoce como *machine-consumption oriented data*).

En este artículo se presenta un trabajo orientado a facilitar la interoperabilidad semántica de las descripciones archivísticas mediante la representación de ISAD(G) en un lenguaje de ontologías. Algunos trabajos anteriores ya han puesto de manifiesto la importancia de aplicar los conceptos de la Web Semántica a la integración de archivos, transformando datos meramente informativos en datos preparados para ser manejados por mecanismos de recuperación avanzados. Sin embargo, el enfoque propuesto en dichos trabajos ha sido sólo someramente comentado en lo tocante a ISAD(G) y únicamente aplicado a dominios muy específicos tales como un sistema de gestión parlamentario (Costilla et al., 2004), un museo de fotografía (Hänninen, 2005) o un museo de ciencias naturales (Chi, 2006). A diferencia de dichas investigaciones, la ontología resultado del trabajo aquí expuesto es pública y puede ser libremente accedida en <http://www.cc.uah.es/ie/ISADG.owl>, está escrita en OWL (un lenguaje específicamente orientado a la Web), y las definiciones del dominio de la archivística se han enlazado y hecho corresponder con definiciones de términos más generales incluidos en ontologías fundamentales (*upper ontologies*, término inglés, frecuentemente traducido como ontología superior u ontología fundamental; en este artículo se ha preferido la segunda de estas acepciones) con objeto de proporcionar mayor universalidad e independencia de usos o aplicaciones específicas.

El resto de este artículo se estructura de la siguiente forma. La sección 2 muestra las carencias de los modelos de descripción actuales cuando se trata de que el software extraiga información relevante de los mismos, y define el concepto de semántica computacional como paso siguiente en el perfeccionamiento de las descripciones archivísticas estandarizadas. La sección 3 muestra los aspectos más rele-

vantes de la representación de ISAD(G) mediante una ontología y sirve de base para la correspondencia con bases de conocimiento ontológico de carácter general que se presenta en la sección 4. Finalmente, la sección 5 resume las conclusiones del estudio y apunta diversas líneas de trabajo futuro.

## 2. Semántica computacional para la archivística

La normalización proporcionada por la norma ISAD(G), junto a formatos para el intercambio electrónico como EAD, permiten el intercambio de descripciones archivísticas gracias a que la información se encuentra estructurada en campos predeterminados. Es posible hablar aquí de *interoperabilidad para el intercambio o portabilidad*. No obstante, los contenidos de la mayoría de los elementos de que constan estas descripciones no son otra cosa que texto escrito en lenguaje natural, de los cuales un programa de computadora difícilmente puede deducir o extraer ninguna información adicional.

Veamos un ejemplo. Uno de los elementos de una descripción archivística conforme con la norma ISAD(G) es el campo *accrual*, donde según la norma debe reseñarse información sobre adquisiciones adicionales de documentos a incluir en el archivo. El siguiente texto, incluido como ejemplo en el estándar ISAD(G), es un contenido válido para el campo *accrual*:

*«It is understood that further Mabó papers are still in the possession of the Mabó family and may be transferred to the Library in the future.»*

Indudablemente, este texto aporta información a un lector humano que sea capaz de entender el lenguaje en que está escrito, en este caso el inglés. Así, es posible deducir de dicho texto que *el fondo descrito puede recibir aportaciones adicionales, pero no hay una periodicidad establecida ni tampoco un plazo para ello*. Sin embargo, un programa de computadora, incluso si utiliza las más avanzadas técnicas de procesamiento del lenguaje natural, difícilmente será capaz de obtener esa información. Así, no podrá clasificar el fondo como cerrado o abierto, ya que simplemente no puede deducirlo a partir de la información disponible, y no será capaz por tanto, de hacer búsquedas basadas en el criterio de si el fondo está o no cerrado.

Tampoco podrá ningún programa controlar de manera automática el acceso a una descripción archivística en función del contenido del campo *Conditions governing access*, a pesar de que las informaciones para dicho campo son a menudo detalladas, pues el lenguaje para la descripción es, una vez más, lenguaje natural y por tanto imposible de interpretar por el software. Un ejemplo incluido en el estándar es el siguiente:

*«Use of audio, video, or film materials may require production of viewing copy.»*

Otros ejemplos notables son el conocimiento geográfico y las fechas. En las descripciones de fechas de la norma son válidas anotaciones tales como «1791-1964, predominantemente 1879-1963». Un programa de computadora tendrá serias dificultades para interpretar qué fechas realmente determinan los límites temporales del fondo, y desde luego no será capaz de interpretar *predominantemente*, desaprovechando la posibilidad de que búsquedas que utilizan la época como criterio principal localicen de forma precisa los fondos. En el ámbito geográfico, la procedencia puede determinarse por la entidad creadora o por apuntes en la historia administrativa, pero una vez más esta información no podrá ser explotada por las máquinas, incapaces de interpretar las narrativas que aparecen habitualmente en las descripciones. Si, por el contrario, tuvieran la capacidad para interpretarlas, podrían hacerse consultas (expresadas en lenguaje natural) tales como «fondos creados por instituciones del ámbito católico iberoamericano en el siglo XVIII», y el software podría encontrarlas con precisión, teniendo en cuenta incluso los límites geográficos de la nación española en el período histórico al que la consulta hace referencia.

Los ejemplos anteriores muestran cómo las descripciones conformes con las actuales normas de descripción estructuran bien dónde encontrar la información en un registro dado siempre que quien deba interpretarlas sea un ser humano, si bien los programas de computadora tienen dificultades para obtener información adicional de ellas. Este problema, similar al que existe en otros estándares de metadatos, ha sido estudiado con anterioridad respecto del estándar *Dublin Core* (Sicilia, 2005).

El siguiente escalón, por tanto, es dotar a las representaciones archivísticas de semántica computacional adicional, con el objeto de que el software sea capaz de llegar más allá en la interpretación de los datos, y con el consiguiente valor añadido a la hora de hacer búsquedas o de comprobar la consistencia de las descripciones. En este artículo se emplea el término *semántica computacional* para denominar aquella propiedad de una descripción archivística que permite a los programas de computadora *entender* la información que contiene, y por tanto realizar tareas de tratamiento de dicha información o inferidas a partir de ella. Este concepto va íntimamente unido al de *interoperabilidad semántica*, que denota la capacidad de una información para ser interpretada, compartida e intercambiada por diferentes sistemas de tratamiento basados en el paradigma de la Web Semántica.

Para alcanzar la semántica computacional es necesario:

1. Que las descripciones (o al menos parte de ellas) se representen mediante lenguajes orientados a las máquinas, basados en lógicas formales. En el caso particular de la archivística, las descripciones formalizadas deberían ser compatibles con ISAD(G) o basarse en dicho estándar, siendo un complemento (y no un sustituto) de las descripciones *orientadas a humanos*.
2. Que exista una representación del conocimiento en ese mismo lenguaje. Actualmente las ontologías formales (Gruber, 1995) son la manifestación más utilizada de este tipo de representaciones del conocimiento.

Cabe preguntarse si la provisión de la semántica computacional no conllevará un aumento de la complejidad y dificultad del proceso de descripción que lo haga impracticable. Posiblemente sea así si se quisiera codificar *computacionalmente* absolutamente todo el contenido de las descripciones, pero no lo es si se dan las siguientes condiciones:

- Se presupone la existencia de ontologías del dominio adecuadas para describir el contexto, organismos y aspectos de las descripciones. Aunque aún hoy no existen para todos los dominios, los esfuerzos en ontologías de conocimiento general, las versiones formalizadas como ontologías de grandes tesauros (Wielinga et al., 2001), y la gran cantidad de ontologías de dominio que se encuentran disponibles puede resultar en un acervo adecuado en un futuro próximo.
- Se proporcionan las interfaces de usuario apropiadas, que oculten en cierta medida al archivero el lenguaje formal utilizado para las descripciones.

Con las condiciones anteriores, la labor de descripción del archivero produce un resultado que, además de estandarizado, posee un cierto grado de semántica computacional más allá de la simple estructuración en campos de texto que proporcionan los estándares actuales. En el resto de este trabajo se describe el resultado de proporcionar un esquema para conseguir dotar de semántica computacional a las descripciones archivísticas normalizadas según el estándar ISAD(G). La pretensión del mismo no es la exhaustividad, sino la ilustración, por lo que no agota todas las posibilidades de provisión de semántica, que pueden ser objeto de trabajos ulteriores.

### **3. Representación de ISAD(G) en una ontología**

Esta sección presenta inicialmente los conceptos en que se basa la representación ontológica, los cuales serán utilizados para la definición de los elementos del estándar ISAD(G). Posteriormente describe el proceso de construcción de una ontología que cubre todos los elementos de descripción de archivos que conforman el estándar ISAD(G). Para la definición formal de los elementos citados se ha elegido el lenguaje OWL (McGuinness y Harmelen, 2004), principalmente porque es el lenguaje de descripción de ontologías recomendado por el consorcio W3C (<http://www.w3.org/>). Los resultados obtenidos permitirán, en la siguiente sección, proponer un esquema de interoperabilidad semántica basada en la correspondencia entre los términos que conforman la ontología ISAD(G) y los definidos en una ontología fundamental.

#### **3.1. Breve introducción a las lógicas de descripciones y OWL**

Las *lógicas de descripciones* (Baader et al., 2002) constituyen una familia de formalismos lógicos para la representación de conocimiento. Basadas en la lógica de

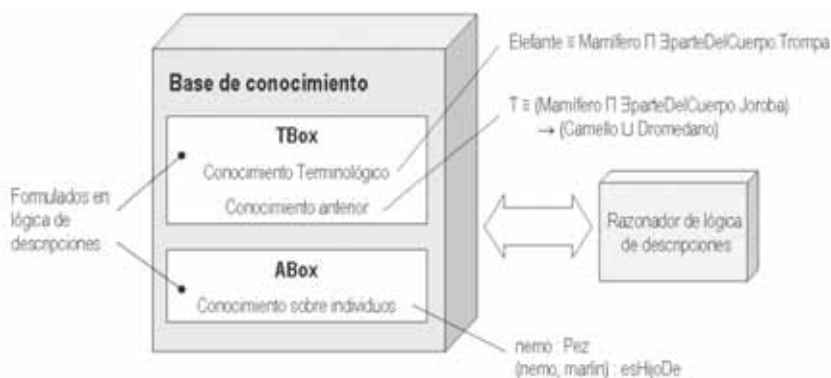
marcos y en las redes semánticas, se utilizan para representar el conocimiento de un dominio de aplicación mediante la definición de los conceptos más relevantes del mismo, las relaciones entre los mismos y las propiedades de los individuos. Dentro de las lógicas de descripciones, una de las más representativas es el cuerpo ALC, que es la lógica en la que se basa OWL.

Las representaciones construidas con lógicas de descripciones permiten llevar a cabo razonamientos automáticos a partir del conocimiento representado explícitamente en un modelo (cuando se habla de razonamiento automático nos referimos al proceso de extraer conclusiones a partir de información en forma de axiomas). De igual modo, soportan la inferencia de patrones, lo cual permite estructurar y comprender el dominio de aplicación mediante la clasificación de conceptos e individuos, característica muy utilizada en sistemas inteligentes de procesamiento de información tales como los basados en ontologías.

Los sistemas basados en lógica de descripciones están formados por dos componentes básicos denominados T-Box y A-Box, que comprenden la terminología (vocabulario del dominio de la aplicación) y las aserciones (individuos definidos según el vocabulario), respectivamente. La representación del dominio se lleva a cabo mediante la clasificación de conceptos, lo que determina las relaciones jerárquicas concepto/supraconcepto dentro de una terminología. Así por ejemplo, pertenece a la T-Box la definición que establece que un *elefante* es un tipo de *mamífero* con *trompa*, hecho que lo diferencia de otros mamíferos (ver Figura 1). Un razonador de lógica de descripciones (un programa de computadora específicamente diseñado para realizar razonamiento automático) es capaz de inferir, con este conocimiento de base y otros similares, que aquellos mamíferos que tienen una parte del cuerpo denominada joroba pertenecen, bien a la clase *Camello* o bien a la clase *Dromedario*. Y esto será así dado que en la T-Box no habrá ningún otro concepto definido que cumpla dicha condición. Por otra parte, la clasificación de individuos permite determinar si un individuo es una instancia de un concepto, tomando como base la descripción del individuo y la definición del concepto. En el ejemplo de la figura 1, *nemo* se declara mediante la aserción *nemo:Pez* como instancia (es decir, individuo) de la clase *Pez*. Las relaciones entre instancias (como por ejemplo la declaración de que el pez *nemo* es hijo del pez *marlin*) posibilitan la aplicación de reglas que inserten hechos adicionales en la base de conocimiento. Así, toda consulta posterior sobre la base de conocimiento que inquiera cuáles son los hijos de *marlin* arrojará (al menos) el resultado *nemo*.

OWL (Web Ontology Language, <http://www.w3.org/TR/owl-ref/>) es un lenguaje de marcado semántico para publicar y compartir ontologías en Internet, específicamente orientado a aquellas aplicaciones que necesiten procesar el contenido de la información en una ontología Web, y no sólo representarla de manera inteligible para usuarios humanos. Las ontologías definidas en OWL pueden utilizar cualquiera de los tres sub-lenguajes disponibles: *OWL-Lite*, *OWL-DL* y *OWL-Full*. El lenguaje elegido dependerá de la expresividad necesaria para representar el conocimiento. *OWL-Lite* es el sub-lenguaje menos expresivo y su uso está limitado a contextos donde

**Figura 1**  
**Componentes de un sistema de representación de conocimiento basado en lógica de descripciones**



exista una jerarquía de conceptos sencilla y propiedades simples. *OWL-DL* es una extensión del anterior basada en lógica de descripciones, lo cual permite realizar razonamiento automatizado a partir de la ontología. Éste es el lenguaje utilizado para codificar formalmente la ontología que resulta del presente trabajo.

### 3.2. Presentación de la ontología ISAD(G)

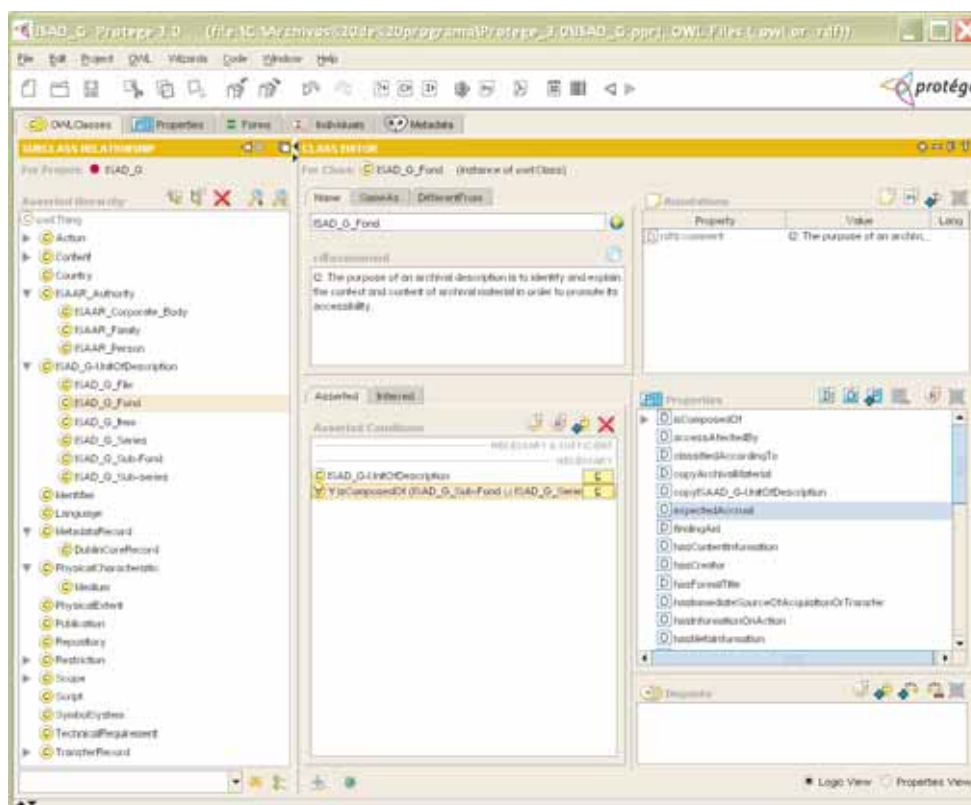
Según la definición del *Standard Upper Ontology Working Group* de IEEE (SUO WG, <http://suo.ieee.org/>), una ontología está formada por «un conjunto de conceptos, axiomas y relaciones entre términos que describen un dominio de interés». Las ontologías se utilizan frecuentemente en el ámbito de la inteligencia artificial para la representación del conocimiento (Uschold y Grueninger, 1996). La ventaja fundamental que aporta el uso de ontologías como medio de representación formal del conocimiento reside en el hecho de que el software puede utilizar el conocimiento de la ontología para llevar a cabo razonamiento inductivo, buscar y recuperar información, o resolver diversos problemas, todo ello de manera automatizada.

La existencia de la ontología que se describe en esta sección permite anotar los recursos (en este caso las descripciones archivísticas) con informaciones tales como «Correspondencia manuscrita del gobernador de la Pampa». A diferencia de las descripciones que actualmente pueden llevarse a cabo utilizando ISAD(G), las anotaciones basadas en ontologías se expresan de modo formal en términos de conocimiento ontológico, lo que elimina las ambigüedades derivadas de, por ejemplo, la existencia de varios lugares en el mundo que concuerdan con un mismo topónimo. De este modo, se evitarían resultados indeseados en las búsquedas como consecuencia de hechos como la coincidencia de que hasta 17 lugares (según el tesauruso



de nombres geográficos de AAT, [http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)) responden al topónimo *Pampa*: el río Pampa (en Brasil), el pico Pampa (en el condado de Kern, California), el lago Pampa (en el condado de Whitman, Washington)... y por supuesto la región de la Pampa en el sur de Argentina. La determinación exacta y sin ambigüedad de un elemento de información facilita que el software pueda realizar búsquedas en la Web del estilo: «busca toda la correspondencia entre gobernadores de colonias españolas con la metrópoli durante el siglo XVII», proporcionando resultados cualitativamente superiores a los proporcionados por las búsquedas basadas en palabras clave que proporcionan los motores de búsqueda convencionales en la Web tales como Google.

**Figura 2**  
Vista parcial de los términos definidos en la ontología ISAD(G)



Esta sección muestra las decisiones tomadas durante el proceso de creación de una ontología que cubre las definiciones de todos los elementos y áreas incluidas en la norma ISAD(G) en su versión original en inglés, así como los resultados del cita-

do proceso. El método de trabajo ha consistido en analizar por separado cada elemento de la norma para, utilizando el editor de ontologías Protégé (Noy et al., 2001), ir construyendo la ontología a medida que se identificaban conceptos y propiedades que dieran respaldo a los conceptos e informaciones que cada elemento ISAD(G) permite almacenar. Simultáneamente, se ha ido formalizando en OWL el resultado de cada decisión. Finalmente, se ha procedido a evaluar la consistencia lógica del modelo, así como la posibilidad de definir como instancias en la ontología un pequeño número de casos escogidos de entre los ejemplos incluidos en ISAD(G).

El término fundamental de la ontología es el que da soporte a la unidad de descripción, denominado *unit of description* y también referido en la norma como *archival material*. Para definirlo en la ontología se ha incluido el término genérico ISAD\_G-UnitOfDescription, que representa los elementos abstractos comunes a todas las unidades de descripción enumeradas en la norma: fondo, subfondo, serie, subserie, fichero e ítem. Para completar la descripción de un material de archivo se han definido propiedades de cada instancia del término ISAD\_G-UnitOfDescription, tal y como se muestra en la figura 2.

El resultado final del proceso descrito en este apartado (un documento escrito en lenguaje OWL con la definición de la ontología completa) es demasiado extenso como para ser incluido completo en este artículo, por lo que a continuación se muestran sólo las decisiones de alto nivel que han permitido crear los términos de la ontología y los conceptos que le dan soporte.

### 3.3. Representación de los elementos de ISAD(G) en una ontología

Como ya se ha señalado, la norma ISAD (G) establece un conjunto de áreas de información que conforman el registro o descripción archivística de los documentos de archivo:

1. De *identificación*: Identifica de forma única la unidad de descripción y establece el vínculo con la descripción que la representa.
2. De *contexto*: Consigna el nombre de la(s) entidad(es) o persona(s) física(s) responsables de la producción, acumulación y conservación de los documentos de la unidad de descripción. El nombre debe especificarse de manera normalizada según los principios de la norma ISAAR (CPF).
3. De *contenido y estructura*: Proporciona la información necesaria para apreciar el valor potencial de la unidad de descripción, dando una visión de conjunto (por ejemplo, por periodos de tiempo, áreas geográficas...) y realizado un resumen de contenido de la unidad de descripción.
4. De *condiciones de acceso y utilización*: Informa sobre la situación jurídica y otras normativas que restrinjan o afecten el acceso a la unidad de descripción. Si procede, indica el período de tiempo durante el cual la documentación permanecerá inaccesible y la fecha en que la documentación lo estará.

5. De *documentación asociada*: Si la unidad de descripción está formada por copias, indica la existencia, disponibilidad, y/o eliminación de los originales.
6. De *notas*: Información complementaria no reseñada en las áreas anteriores.
7. De *control de la descripción*: Indica quién ha elaborado la descripción, y cómo.

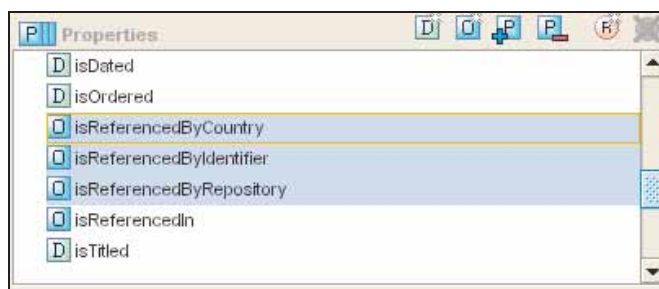
El trabajo aquí reseñado ha consistido en la construcción de una ontología que da soporte a todos los elementos de todas las áreas de la norma ISAD(G). Se muestran a continuación los elementos de la misma, a menudo acompañados por las decisiones tomadas al modelarlos. El código de dos cifras que va entre corchetes y que precede cada término refiere el área al que pertenece el elemento y el número que hace el elemento dentro del área.

[1.1] *Reference code(s)*

El código (o códigos) de referencia/signatura se compone de: código de país, código de repositorio y código de referencia local (número de control u otro identificador único). Tal y como se muestra en la figura 3, esta información se ha modelado con las propiedades `isReferencedByCountry`, `isReferencedByIdentifier` e `isReferencedByRepository` del término `ISAD_G-UnitOfDescription`, que enlazan sus instancias, respectivamente, con instancias de `Country`, `Identifier` y `Repository`.

El código de país seguirá generalmente el formato establecido en el estándar ISO 3166-1, por lo que se ha definido una propiedad de datos para la clase `Country` denominada `ISO3166-CountryCode` que referencia una cadena de dos caracteres. No obstante es posible establecer otro sistema específico de nomenclatura creando nuevas propiedades `Std-CountryCode` según las circunstancias.

**Figura 3**  
Propiedades que soportan la información del código de referencia/signatura



### [1.2] Title

El título, que puede ser formal o asignado, ha sido modelado mediante la propiedad `isTitled`, y sus subpropiedades `hasSuppliedTitle` y `hasFormalTitle` del término `ISAD_G-UnitOfDescription`, que enlazan con elementos de datos de tipo texto (`string`).

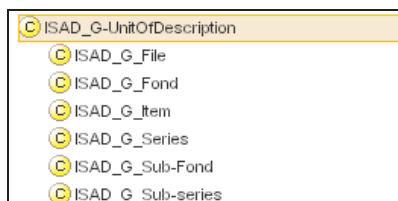
### [1.3] Date(s)

Modelado mediante subpropiedades de `isDated` (propiedad de datos de tipo fecha) del término `ISAD_G-UnitOfDescription`. El estándar hace referencia a fechas de creación y de acumulación, lo cual puede modelarse con propiedades específicas: `creationDate` y `accumulationDate`. El dominio de todas las propiedades que manejan tipos fecha es `string`, pues en ISAD(G) las fechas pueden tomar valores textuales tales como «1833-1998 (la mayor parte de 1833-1874)». Sin embargo, sería deseable utilizar un concepto específico para fechas o rangos de fechas que pudieran haber sido definidos en ontologías de conocimiento general.

### [1.4] Level of description

Es el término fundamental de la ontología. Modelado mediante un término genérico `ISAD_G-UnitOfDescription`, es la base para la creación de los términos subsumidos `ISAD_G-Fond`, `ISAD_G-Sub-Fond`, `ISAD_G-Series`, `ISAD_G-Sub-Series`, `ISAD_G-File` e `ISAD_G-Item`. La figura 4 muestra esta jerarquía, donde el nivel de cada unidad de descripción es implícito y viene definido por la pertenencia de una instancia a alguna de las clases de la jerarquía y por las relaciones entre los conceptos de la misma. Así por ejemplo, un fondo está formado (`isComposedOf`) por un conjunto de Subfondos, o Series o Subseries o Ficheros o Ítems o mezcla de los anteriores según un criterio de agregación.

**Figura 4**  
**Jerarquía del término `ISAD_G-UnitOfDescription`**



La formalización (parcial) en lenguaje OWL de esta jerarquía de conceptos es la siguiente:

```
<owl:Class rdf:about="#ISAD_G-UnitOfDescription">
</owl:Class>

<owl:Class rdf:ID="ISAD_G_Fond">
<rdfs:subClassOf rdf:resource="#ISAD_
G-UnitOfDescription"/>
<rdfs:subClassOf>
<owl:Restriction>
<owl:allValuesFrom>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<owl:Class rdf:ID="ISAD_G_Sub-Fond"/>
<owl:Class rdf:ID="ISAD_G_Series"/>
<owl:Class rdf:about="#ISAD_G_Sub-series"/>
<owl:Class rdf:about="#ISAD_G_File"/>
<owl:Class rdf:about="#ISAD_G_Item"/>
</owl:unionOf>
</owl:Class>
</owl:allValuesFrom>
<owl:onProperty>
<owl:ObjectProperty rdf:about="#isComposedOf"/>
</owl:onProperty>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

Así, la definición de una propiedad de un concepto se codifica de la siguiente manera:

```
<owl:ObjectProperty rdf:ID="languageOfmaterial">
<rdfs:range rdf:resource="#Language"/>
<rdfs:domain rdf:resource="#ISAD_
G-UnitOfDescription"/>
</owl:ObjectProperty>
```

Sin embargo, y con la intención de clarificar el artículo, no se incluirán a partir de aquí las codificaciones en OWL de los elementos de la ontología. Remitimos al lector interesado al documento OWL con la completa definición de todos los conceptos y propiedades que puedan accederse en <http://www.cc.uah.es/ie/isadg.owl>.

[1.5] *Extent and medium of the unit of description (quantity, bulk, or size)*

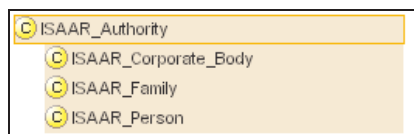
Modelado mediante la propiedad `hasPhysicalExtent` que relaciona instancias del término `ISAD_G-UnitOfDescription` con instancias del término `PhysicalExtent`.

[2.1] *Name of creator(s)*

[2.2] *Administrative / Biographical history*

Siguiendo la recomendación incluida en la norma ISAD(G), estas informaciones han sido modeladas como propiedades (`createdBy`, `accumulatedBy`, `maintainedBy`) que enlazan instancias de `ISAD_G-UnitOfDescription` con instancias del término `ISAAR_Authority` (cuya jerarquía se representa en la figura 5), donde se representa y da soporte a los conceptos de la norma ISAAR(CPF). Cada una de las propiedades hace referencia al rol de la autoridad referenciada, tal y como se especifica en el apartado 2.1. de la norma ISAD(G): creador, persona que acumula y persona que mantiene la unidad de descripción.

**Figura 5**  
**Jerarquía del término `ISAAR_Authority`**

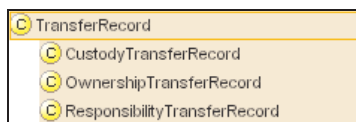


[2.3] *Archival history*

[2.4] *Immediate source of acquisition or transfer*

Información modelada mediante las propiedades `hasTransferRecord` y `hasImmediateSourceOfAcquisitionOrTransfer` que enlazan las instancias del término `ISAD_G-UnitOfDescription` con instancias del término `TransferRecord`, cuya jerarquía se representa en la figura 6.

**Figura 6**  
**Jerarquía del término `TransferRecord`**

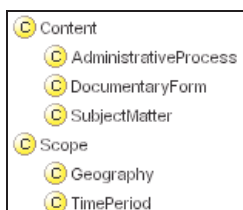


Es necesario destacar que si bien las propiedades `hasTransferRecord` y `hasImmediateSourceOfAcquisitionOrTransfer` relacionan instancias de las mismas clases, la primera admite en su rango múltiples instancias mientras que la última es funcional, esto es, relaciona cada instancia de la clase `ISAD_G-UnitOfDescription` con una única instancia del término `TransferRecord`. En particular, la instancia referida desde esta propiedad será necesariamente una de las referidas con la propiedad `hasTransferRecord`, probablemente la última de ellas, ya que se trata de la última fuente de adquisición o transferencia de la unidad de descripción.

### [3.1] *Scope and content*

Información modelada mediante las propiedades `hasContentInformation` y `hasScopeInformation`, que relacionan las instancias del término `ISAD_G-UnitOfDescription` con instancias de términos de las jerarquías `Content` y `Scope` tal y como se muestra en la figura 7.

**Figura 7**  
**Jerarquía de los términos `Content` y `Scope`**

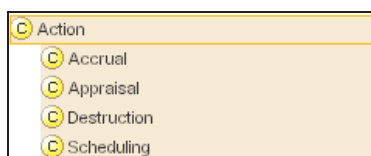


### [3.2] *Appraisal, destruction and scheduling information*

#### [3.3] *Accruals*

Información modelada mediante la propiedad `hasInformationOnAction` que relaciona las instancias del término `ISAD_G-UnitOfDescription` con instancias del término `Action`, cuya jerarquía se representa en la figura 8.

**Figura 8**  
**Jerarquía del término `Action`**



#### [3.4] *System of arrangement*

Se emplean tres textos (o cadenas de caracteres, *strings*) asociados a las instancias de `ISAD_G-UnitOfDescription` mediante las propiedades `internallyStructured`, `isOrdered` y `classifiedAccordingTo`.

Ésta no es, posiblemente, la solución óptima. Una solución alternativa podría ser, por ejemplo, incluir dicha información dentro de la estructura interna de fondo, serie, etc. en donde se podría añadir a cada elemento una propiedad *next* para especificar el siguiente elemento indicando el orden, si bien se considera aceptable para una primera versión de la ontología.

#### [4.1] *Conditions governing access*

Usamos la propiedad `accessAffectedBy` que relaciona el término `ISAD_G-UnitOfDescription` con `Restriction`. `Restriction` tiene cuatro términos subsumidos: `LegalStatus`, `Contract`, `Regulation` y `Policy`. También tiene dos propiedades de datos: `extentOfPeriod` y `terminationDate`.

#### [4.2] *Conditions governing reproduction*

Usamos la propiedad `reproductionAffectedBy` que relaciona el término `ISAD_G-UnitOfDescription` con `Restriction`, que tiene `Copyright` como término subsumido.

#### [4.3] *Language/scripts of material*

Usamos las propiedades `symbolSystemOfMaterial`, `scriptOfMaterial` y `languageOfMaterial` que relacionan el término `ISAD_G-UnitOfDescription` con los términos `SymbolSystem`, `Script` y `Language` respectivamente.

#### [4.4] *Physical characteristics and technical requirements*

Usamos la propiedad `hasPhysicalCharacteristic` que relaciona el término `ISAD_G-UnitOfDescription` con los términos derivados de `PhysicalCharacteristic`, que a su vez están relacionados con `TechnicalRequirement` mediante la propiedad `entails`. Para el software, se supone que el término `Medium` (subsumido de `PhysicalCharacteristic`) ya enlaza con el `TechnicalRequirement` adecuado.



#### [4.5] *Finding aids*

Para modelar esta información se ha utilizado la propiedad `findingAid` del término `ISAD_G-UnitOfDescription`.

#### [5.1] *Existence and location of originals*

Modelado mediante la propiedad `originalISAD_G-UnitOfDescription` del término `ISAD_G-UnitOfDescription` que enlaza con otra `ISAD_G-UnitOfDescription` (en caso de que exista y esté descrita) o con un original descrito de otra forma no conforme con ISAD(G) a través de la propiedad `originalArchivalMaterial`. También se emplea la propiedad `originalStatus` como propiedad de datos para indicar si está disponible, destruido, etc.

#### [5.2] *Existence and location of copies*

Se emplea bien la propiedad `copyISAD_G-UnitOfDescription` del término `ISAD_G-UnitOfDescription` que enlaza con otra `ISAD_G-UnitOfDescription` (en aquellos casos en que exista y esté descrita) o bien `copyArchivalMaterial` cuando la copia descrita se encuentra en formato no ISAD.

#### [5.3] *Related units of description*

Se ha creado la propiedad `relatedISAD_G-UnitOfDescription` del término `ISAD_G-UnitOfDescription` que lo relaciona con otra `ISAD_G-UnitOfDescription`.

El modelado de la información relatada como «*Use appropriate introductory wording and explain the nature of the relationship*» es comprometido, ya que requiere decisiones sobre el carácter de las relaciones. Constituye por tanto, material para un trabajo futuro o una revisión posterior.

#### [5.4] *Publication note*

La propiedad `isReferencedIn` de `ISAD_G-UnitOfDescription`, enlaza instancias de dicha clase con las instancias del término `Publication`.

#### [6.1] *Note*

Se emplea la propiedad de datos (tipo *string*) `hasnote` de `ISAD_G-UnitOfDescription`.

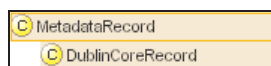
[7.1] *Archivist's Note*

[7.2] *Rules or Conventions*

[7.3] *Date(s) of descriptions*

Para el modelado de toda la información en esta categoría, información de metadatos al fin y a la postre, se ha incluido en la ontología la propiedad `hasMetadataInformation` de `ISAD_G-UnitOfDescription` que apunta a un `MetadataRecord`. Puede ser un registro `DublinCore`, por ejemplo, tal y como se muestra en la jerarquía tentativa de la figura 9.

**Figura 9**  
**Jerarquía (no definitiva) del término `MetadataRecord`**



Hay que resaltar que los registros `DublinCore` no permiten información tan específica sobre archivos como la que podría derivarse de una lectura amplia de las ISAD(G) en este apartado. Sin embargo, la orientación de los registros `Dublin Core` a la descripción de información sobre todo tipo de materiales permite encajar las informaciones de este apartado en las distintas categorías (denominadas elementos en `Dublin Core`) propuestas por dicha especificación.

#### **4. Correspondencia de ISAD(G) con ontologías de conocimiento general**

Lo descrito hasta el momento, si bien interesante desde la perspectiva de la Web Semántica, no es sino una traducción (más o menos directa y estructurada) del estándar ISAD(G) a lógica de descripciones y plasmado en un lenguaje de ontologías. Ya se mencionó en la introducción la existencia de esfuerzos similares, si bien no tan prolijamente detallados.

Las siguientes secciones muestran la parte del trabajo que realmente aporta novedad con respecto a otros similares, pues describe cómo enlazar la ontología de dominio particular (en este caso del dominio de la archivística) con ontologías de conocimiento general con el objetivo de obtener interoperabilidad semántica.

##### **4.1. Introducción**

Las ontologías fundamentales o superiores incluyen definiciones de entidades generales y no están, por tanto, orientadas a un dominio específico. Aunque en la actualidad ninguna ontología fundamental constituye un estándar *de facto* que pueda ser referenciado universalmente, los esfuerzos más importantes están siendo estu-

diados por el *SUO WG* con el objeto de desarrollar un estándar que especifique una ontología fundamental que incluya definiciones universalmente aceptadas.

Formalizar en una ontología escrita en OWL un conjunto de descripciones no es suficiente para proporcionar semántica computacional rica a dichas descripciones. El paso más obvio en la dirección correcta consiste en integrar los conceptos de la nueva ontología con conceptos de mayor nivel cuya definición ha sido consensuada universalmente. Es indudable que uno de los objetivos de la Web Semántica como sistema de representación de conocimiento distribuido es el de reutilizar representaciones. Por ello, cuando se desarrolla una nueva ontología formal como la que se ha descrito en la sección anterior, tratar de relacionar sus conceptos con los declarados en ontologías superiores de conocimiento general es primordial, de manera que todo el conocimiento representado en éstas últimas pueda ser utilizado en nuestra nueva ontología (Sicilia et al., 2006). La utilidad de elaborar la citada correspondencia técnica puede pues resumirse en los siguientes puntos:

1. Revisar y comprobar el trabajo de ingeniería de ontologías realizado. El hacer corresponder a los términos de nuestra nueva ontología formal con los de la ontología fundamental es un proceso intelectual que en muchas ocasiones lleva a clarificar conceptos, o a descubrir inconsistencias.
2. Poner la ontología fundamental a disposición de las aplicaciones que vayan a utilizar nuestra nueva ontología.
3. Como resultado adicional, mejorar la interoperabilidad de nuestra ontología, ya que la ontología fundamental sirve como marco común público para que otras personas (o programas de computadora) utilicen el conocimiento que hemos representado.

El proyecto *Cyc* (Lenat, 1995) representa uno de los intentos más importantes de construir una base de conocimiento de un volumen significativamente mayor al de otros sistemas actuales. Sus promotores creen que el éxito de la inteligencia artificial requiere la realización de enormes bases de conocimiento para que los sistemas basados en ellas puedan aprovecharse del conocimiento *común o general* (*common-sense*), y no limitarse a conocimientos específicos de ciertas situaciones o dominios. El proyecto *OpenCyc* (<http://www.opencyc.org>) no es otra cosa que una versión reducida de *Cyc* que puede descargarse gratuitamente a través de Internet. En la fecha de redacción de este artículo, *OpenCyc* contiene más de 300.000 conceptos que conforman, según sus autores, «una ontología fundamental de toda la realidad consensual humana», junto con millones de aserciones sobre esos conceptos. Si bien existen otras ontologías fundamentales disponibles a través de Internet, en lo que sigue se utiliza *OpenCyc* por su importancia.

## 4.2. Descripción de la técnica

La técnica de correspondencia con OpenCyc (válida para cualquier otra ontología fundamental) parte de una descripción estructurada como ISAD(G), o de un modelo ontológico como el descrito en la sección anterior, y lleva a cabo un proceso de revisión sistemática de cada término y sus descripciones, produciendo una correspondencia con la ontología fundamental. Los pasos pueden resumirse en lo siguiente:

- Encontrar en la ontología fundamental uno o más términos que subsuman a la categoría que se está considerando.
- Comprobar que la correspondencia es consistente, incluyendo al resto de los supra-términos y predicados de éstos en la ontología fundamental.
- Proporcionar los predicados adicionales necesarios, en su caso, para caracterizar al nuevo concepto.
- Editar en Protégé (u otro editor de ontologías) la definición formal.
- Iterar las fases anteriores, con la comprobación de diferentes expertos o grupos de ellos.

**Figura 10**  
Ejemplo de búsqueda en *OpenCyc*

The screenshot shows the OpenCyc search interface. At the top, there is a search bar with the text "country" and buttons for "Search" and "Clear". To the right of the search bar are several icons representing different search functions: History, Create, Assert, Compose, Query, and Doc. Further right, it says "You are: CycAdministrator" with links for "Logout", "Preferences", and "Tools".

The main content area displays "Search results for '\*country\*'" and indicates "74 matching terms found:". Below this, there are three columns of results: "Collections", "Relations", and "Individuals".

Collections	Relations	Individuals
<a href="#">AfricanCountry</a>	<a href="#">CityInCountryFn</a>	<a href="#">CountryLakeFoods-Corporation</a>
<a href="#">AlternativeCountryMusicGenre</a>	<a href="#">EmbassyBuildingOfCountryFn</a>	<a href="#">CountrysideLifesaving-NCR</a>
<a href="#">AlternativeCountryMusicPerformer</a>	<a href="#">EmbassyMissionsOfCountryFn</a>	<a href="#">CountrywideMortgageInvestments-Corporation</a>
<a href="#">CasorianCountry</a>	<a href="#">TermParaphraseFn-CityWithCountryAbbreviation</a>	<a href="#">CrackerBarrelOldCountry-Corporation</a>
<a href="#">CaucasusCountry</a>	<a href="#">TermParaphraseFn-CityWithCountryName</a>	<a href="#">Cyprus-TheCountry</a>
<a href="#">CentralAmericanCountry</a>	<a href="#">countryCanProduceObject</a>	<a href="#">Malta-TheCountry</a>
<a href="#">CityTypeByCountry</a>	<a href="#">countryCodeDigraph</a>	<a href="#">Nima-Country-CMLS</a>
<a href="#">Country</a>	<a href="#">countryName-LocalLongForm</a>	<a href="#">Nima-CountryName-KS</a>
<a href="#">CountryAndWestern-StyleOfDance</a>	<a href="#">countryName-LocalShortForm</a>	<a href="#">Nima-CountryName-PS</a>
<a href="#">CountryMusic</a>	<a href="#">countryName-LongForm</a>	

At the bottom of the page, there is a status bar with the text: "Update Comm: Storing Only Agenda: Sleep KB: 2 System: 1.8023".

La correspondencia resultante de este proceso tiene como ventaja el estar sujeta a comprobación o evaluación racional desde la perspectiva meramente formal (lógica) pero también conceptual (ontológica) de acuerdo a las definiciones preexistentes en la ontología fundamental. Para la búsqueda de conceptos, predicados o individuos existentes se utilizan las interfaces de búsqueda de *OpenCyc* (Figura 10).

El resultado del proceso inicial puede entonces exponerse a la consideración de otras personas, o a procesos de evaluación en grupo, pero siempre el corpus lógico-racional de la ontología fundamental deberá ser el referente y el marco de formulación de las posibles objeciones, dando por tanto a la discusión un carácter preciso, y eliminando en gran medida las subjetividades o ambigüedades de los usos del lenguaje natural.

Es importante reseñar que la ontología descrita en la sección 3 está escrita en OWL mientras que la ontología Opencyc lo está en un lenguaje propio denominado CycL, no compatible con OWL. A continuación se definirán las principales correspondencias entre la ontología ISAD(G) y Opencyc aunque para explotar el conocimiento de Opencyc en nuestra ontología ISAD(G) sería necesario bien definir esta última en CycL o bien exportar el conocimiento de Opencyc a OWL.

### 4.3. Fundamentales correspondencias

#### *Materiales de archivo*

No existe en OpenCyc un concepto específico para los materiales de archivo, pero el término Information Bearing Object (IBO) puede utilizarse como super-concepto de ISAD\_G-UnitOfDescription. A continuación se reproduce la descripción del término (se han dejado las definiciones en el inglés original dado que las traducciones pueden desvirtuar el propósito original de la descripción):

*Each instance of InformationBearingObject (or «IBO») is a physical object that can be interpreted to yield information. InformationBearingObject includes: (1) artifacts made solely for the purpose of conveying information (e.g., a newspaper, or a children's science video); (2) artifacts that convey information in addition to their intended function (e.g., Neolithic pottery); and (3) non-artifacts, such as a person's fingerprints or cellular DNA, which may be interpreted to yield information.*

De la definición anterior se desprende que es la interpretación (1) la que subsume a los documentos de archivo. Sería necesario añadir a la misma una definición formal para incluir el carácter testimonial de los documentos de archivo, y así diferenciarlos de otros tipos de documentos. No obstante, esta definición no es trivial y en principio no influye como criterio de búsqueda dentro del dominio archivístico, por lo que no se ha trabajado en ella. La relación entre los diferentes niveles de organización del fondo pueden representarse gracias al predicado `compositeParts` de OpenCyc, que ya es por defecto transitivo, binario y antisimétrico, como se requiere en la definición de ISAD(G).

El uso de ese predicado cumple directamente la regla 2.3. sobre descripción multinivel incluida en ISAD(G) acerca de la *vinculación de las descripciones* y permite automatizar varias funciones relativas a la descripción multinivel de ISAD(G):

1. Comprobar la consistencia de las descripciones, en caso de que aparezcan a más de un nivel. Por ejemplo, que el ámbito geográfico del nivel de fondo cubra los ámbitos geográficos proporcionados en los subniveles.
2. Comprobar la no repetición de la información (regla de descripción multinivel 2.4), y en su caso eliminar o factorizar las informaciones que sean redundantes.
3. Otras comprobaciones que pueden ser específicas de ciertos organismos o dominios. Por ejemplo, en archivos de administraciones actuales, pueden incorporarse como reglas las actuales de funcionamiento.

Estas funciones pueden automatizarse fácilmente con un nivel adecuado de semántica computacional.

### *Identificación*

El *título* no requiere correspondencia adicional, y el *nivel* está implícito en la propia jerarquía incluida en el modelo ontológico OWL. Respecto al código de *referencia*, al menos la parte de consignación del país debería representarse mediante enlaces a países representados como entidades geopolíticas en la ontología. Para ello, OpenCyc proporciona el término `IndependentCountry`: para modelar estos países.

*A collection of geopolitical entities and a specialization of Country. Each instance of IndependentCountry is a distinct, independent geopolitical entity generally recognized by the international community. An independent country typically has a relatively stable government, enforced borders, its own currency, laws, and culture.*

Debe observarse que el término OpenCyc más general `Country` no es apropiado en este contexto, ya que incluye a regiones no completamente independientes, como Escocia. Es importante resaltar la utilidad de esto en las consultas, ya que la información de los países en la ontología incluye conocimiento general sobre su ubicación relativa, datos demográficos o políticos, etc. que en determinadas ocasiones podrían resultar de utilidad en la búsqueda. Por ejemplo, se podrían solicitar «fondos albergados en países gobernados por regímenes democráticos», «fondos en países de habla hispana», o «fondos en países con renta per cápita superior a un cierto nivel en cierto intervalo temporal».

Respecto a las fechas, además de las distinciones ya hechas en el esquema OWL, en general puede utilizarse `TimeInterval` para su representación. La descripción se proporciona a continuación.

*Each instance of TimeInterval is an intangible temporal thing that is characterized fully by its temporal extent. In this way, time intervals differ*

*from Situations such as Events. For example, the year 1969 C.E. is a TimeInterval; although many interesting things happened during that year, the year itself is completely defined by its temporal extent. Conversely, Neil Armstrong's walking on the Moon is an event and not a time interval, since it is not fully characterized by its temporal extent or any other purely temporal attributes (e.g. its taking place on the Moon and being done by Armstrong are not temporal attributes).*

Nótese en el texto de descripción anterior que en ocasiones podría ser adecuado describir este intervalo temporal como *situación* o *evento*, conceptos que no se determinan exclusivamente por la extensión temporal. Por ejemplo, *La guerra civil española* sería un evento predominantemente reconocido como descripción temporal, pero que además determina un ámbito y un contexto histórico. Nótese de nuevo que este tipo de modelos sirve a innovadoras modalidades de consulta para la investigación, por ejemplo «fondos creados en situación de guerra en un país».

Respecto a la descripción de volumen y soporte de la unidad de descripción, la gran variabilidad posible en esta información requeriría un importante esfuerzo adicional. Sólo a modo de ideas preliminares, se puede mencionar que el predicado de *OpenCyc* `volumeOfObject` serviría para describir volúmenes, y que muchos tipos de formatos físicos, tales como `VideoCassette` o `Microfilm`, ya aparecen modelados en la ontología.

### *Contexto*

Respecto a la creación de las unidades documentales, *OpenCyc* proporciona un tipo de evento concreto denominado `IBOCreation` que permite describirlo en función de los agentes (`IntelligentAgent`) que participan en su creación. No se han encontrado elementos en *OpenCyc* explícitos para la conservación o creación, por lo que habría que extender su lenguaje de eventos en este caso.

Respecto a las descripciones normalizadas de ISAAR(CPF), se pueden hacer las siguientes correspondencias:

- Las personas pueden representarse mediante instancias de `Person`.
- Las organizaciones pueden representarse bien por `Organization`, o bien por `CommercialOrganization` si se consideran como organizaciones con ánimo de lucro.
- Las familias son consideradas un subtipo de organización denominado `Family-Human`, definido como «*Organization of individuals, usually related by birth and mating, often cohabiting*».

Las tres categorías anteriores son subclases de `IntelligentAgent`, un concepto definido en *OpenCyc* de la siguiente forma:

*An agent is an IntelligentAgent if and only if it is capable of knowing and acting, and capable of employing its knowledge in its actions. An intelligent agent typically knows about certain things, and its beliefs concerning those things influences its actions. As with agents generally, an intelligent agent might either be a single individual, such as a person, or a group consisting of two or more individual agents, such as a business or government organization.*

De este modo, los diferentes tipos de agentes pueden participar en eventos de creación, circulación, gestión, etc. de las unidades documentales. La información en la ontología sobre estos agentes contendría la historia institucional/reseña biográfica descrita de manera formal, por lo que no es necesario introducir más elementos. En caso de que sólo algunas de estas informaciones fueran relevantes para la descripción de una unidad archivística determinada, éstos podrían consignarse mediante *meta-predicados*.

La *historia archivística* hace referencia a los *traspasos sucesivos de propiedad, responsabilidad o custodia* entre otras informaciones. Los eventos de transferencia de propiedad en *OpenCyc* se detallan en el concepto *TransferringOwnership*:

*A collection of events; a subcollection of TransferringPossession. In each instance of TransferringOwnership the ownership of some object passes from one agent to another.*

Otros tipos de eventos pueden modelarse con conceptos de manera similar. Como todos los eventos, se definen mediante los *roles de agentes* apropiados para determinar los participantes, y la extensión temporal queda también determinada mediante predicados. La *forma de ingreso* sería un evento de este tipo, pero con el caso especial de que la transferencia viene del productor; si bien esto no requiere ningún modelo adicional.

### *Contenido y estructura*

El elemento *3.1. Alcance y contenido* incluye informaciones de muy diversa índole. Algunas de las informaciones que tienen relevancia en este apartado son las siguientes:

- Las regiones geográficas pueden representarse como *GeographicalRegions* (*Each instance of GeographicalRegion is a tangible spatial region that includes some piece of the surface of a planet*).
- Los períodos temporales con referencia histórica también pueden modelarse como *HistoricalPeriod* (*The collection of significant periods of history. Examples include TheRenaissance*).



Para la descripción del contenido pueden utilizarse aproximaciones de enlaces semánticos que ya se han discutido en el ámbito de los contenidos didácticos (Sicilia y García Barriocanal, 2005).

Como eventos pueden también modelarse las acciones de valoración, selección y eliminación. En el caso de los *nuevos ingresos*, la frecuencia puede modelarse mediante instancias de `Frequency`, de modo que también la periodicidad de adquisición pueda convertirse en criterio de búsqueda.

La *Organización* del fondo está en parte determinada por la propia estructura multinivel. No obstante, para la representación de los criterios de ordenación en ocasiones puede utilizarse el concepto `CataloguingWrittenMaterial`, definido como:

*A specialization of IBOCreation and SortingSomething. Each instance of this collection is an event in which someone (such as a Librarian) creates a Catalogue of written materials (by recording and/or filing information about those materials).*

No se encuentra en *OpenCyc* un modelo de tipos de ordenación.

#### *Condiciones de acceso y utilización*

Las condiciones de acceso en *OpenCyc* deben modelarse como tipos específicos de *situaciones*, a las cuales se llega mediante la aserción de predicados `rights-Granted(Evento, Situación)`. El evento (ej. una compra) inicia la posesión de derechos de un agente modelada por una situación. Por ejemplo, la situación de `ExclusiveUserRights` se define de la siguiente manera:

*Each instance of this collection is a situation that consists of an agent having an exclusive right to some object. Consider the exclusive user rights I enjoy with respect to my car. This is a particular instance of Situation. It started when I bought the car and ends when I sell it or otherwise transfer ownership.*

En este tipo de modelado se registran tanto los eventos como las situaciones de derecho, combinándose por tanto con parte del área anterior. No obstante, *OpenCyc* no proporciona un modelo completo de representación de derechos de reproducción que cubran todas las posibilidades. Esto podría ampliarse considerando especificaciones existentes como la *Open Digital Rights Language*, ODRL (<http://odrl.net/>).

Respecto a las lenguas utilizadas, estas están representadas en *OpenCyc* como `HumanLanguage` y `NaturalLanguage` (es importante reseñar que ésta última excluye los lenguajes con creación artificial).

No hay en OpenCyc modelos sobre requisitos de preservación en general, aunque los requisitos técnicos de software sí pueden modelarse mediante instancias de `SoftwareObject-Individual`.

#### *Documentación asociada*

OpenCyc proporciona predicados espaciales para las localizaciones, pero en los sistemas de archivos se suelen utilizar códigos para esta localización, por lo que no parece tener una gran utilidad.

La conexión con unidades de información relacionadas puede establecerse como una forma especial de *enlaces con tipo* (Sicilia et al., 2005), donde éstos especifican la naturaleza de la relación, y opcionalmente pueden tener un grado de fuerza u otra información relacionada.

#### *Notas*

Dado que en la norma el objetivo se define por exclusión como *información que no esté en otras áreas*, se puede asimilar a cualquier otro predicado descriptivo en el que participe algún individuo de tipo `ISAD_G-UnitOfDescription`.

#### *Control de la descripción*

En esta parte se consigna información sobre el creador de la información que hemos descrito anteriormente. Los archiveros pueden representarse, como ya hemos visto, como instancias de `Person`. Pero para relacionar a estas personas con la descripción hay que relacionarlas con predicados existentes, lo que implica *meta-descripción* dentro de la ontología. Para consignar la descripción puede utilizarse un evento de un tipo nuevo `MetadataRecordCreation`, que debe indicar:

- La persona que lo ha creado.
- Las fechas en las que se ha creado, que pueden ser un `TimeInterval` definitorio de la extensión temporal del evento.
- Las normas utilizadas. En este caso hace falta más elaboración para determinar qué tipo de normas, por ejemplo podría utilizarse `Specification` para modelar las mismas.

## **5. Conclusiones y trabajo futuro**

El proceso general de normalización en el área de la archivística ha de ser comprendido dentro de los cambios producidos por el crecimiento y especialización de

la información, lo que ha obligado a buscar accesos más rápidos, eficaces y profundos a la información, evitando ruidos y/o silencios que dificultan el acceso a la misma. En este contexto se entiende la aparición de bases de datos y catálogos colectivos, y los importantes avances en el tratamiento de los archivos, que comienzan a plantearse no sólo la mejora en los medios de transferencia de su información, sino la obligación de abordar una normalización del contenido de los documentos para facilitar el acceso a la información.

Se ha abordado la provisión de semántica computacional para una representación formal de la norma ISAD(G), utilizando como base la lógica de descripciones, el lenguaje de ontologías Web (OWL) y como soporte para definiciones de términos de propósito general, la base de conocimientos de Opencyc, que se puede clasificar dentro de las denominadas *ontologías fundamentales*. La correspondencia con Opencyc permite que otros investigadores corrijan y extiendan el trabajo realizado a partir del conocimiento codificado en Opencyc.

El grado de estructuración y detalle de la norma permite hacer la *traducción* entre ISAD(G) y OWL, consiguiendo un alto grado de formalización que puede utilizarse para proporcionar servicios avanzados de comprobación y consulta. No obstante, algunas partes de la descripción requieren de un mayor trabajo futuro en la especificación semántica. La ontología creada pretende ser una propuesta inicial que provoque la discusión de la comunidad, con el objetivo inmediato de alcanzar consenso sobre las definiciones más comprometidas o menos evidentes a la hora de ser clasificadas. En un futuro, se desarrollarán aplicaciones que desarrollen los mecanismos de inferencia y las capacidades adicionales que la existencia de la ontología y su puesta a libre disposición sugieren.

### Agradecimientos

Este trabajo se enmarca dentro de los temas de interés y actividades del proyecto Sere-OC (ref. CCG06-UAH-TIC-0606) de la Universidad de Alcalá, del que recibe financiación parcial.

### Referencias

- BAADER, F., CALVANESE, D., McGUINNESS, D., NARDI, D. y PATEL-SCHNEIDER, P. (2002). *The Description Logic Handbook*, Cambridge University Press.
- BERNERS-LEE, T., HENDLER, J., LASSILA, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43.
- CHI, Y. (2006). Constructing Ontologies for Sharing Knowledge in Digital Archives. En D.S. Yeung et al. (Eds.): *Advances in Machine Learning and Cybernetics, 4th International Conference, ICMLC 2005*, LNAI 3930, pp. 295–304.
- COSTILLA, C., CREMADES, J., CALLEJA, A., FERNÁNDEZ, R. y PALACIOS, J.P. (2004) Integración de archivos digitales en la Web a partir del sistema de gestión parla-

- mentario SIAP. *Actas del IX Congreso Nacional de Internet, Telecomunicaciones y Movilidad, Mundo Internet 2004*, 41–57.
- EAD: Encoded Archival Description, version 2002. En línea (último acceso: 30 de junio de 2007): <http://www.loc.gov/ead/eadschema.html>
- GRUBER, T. (1995). Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43 (5/6), 907–928.
- HÄNNINEN, H. (2005). Creating and implementing a large ontology (RDF schema) for managing photography collections for the Finnish Museum of Photography (FMP). En actas de XML Finland 2005 Conference.
- HEREDIA, A. (1992). Thesaurus indiano en Europa e Iberoamérica. Cinco siglos de intercambios. IX Congreso Internacional de Historia de América. Sevilla, Asoc. de Historiadores Latinoamericanos, 255-260.
- ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, Second Edition. En línea (ultimo acceso 30 de junio de 2007): <http://www.ica.org/sites/default/files/ISAAR2EN.pdf>
- ISAD(G): General International Standard Archival Description, Second Edition. Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999. En línea (ultimo acceso 30 de junio de 2007): [http://www.ica.org/sites/default/files/isad\\_g\\_2e.pdf](http://www.ica.org/sites/default/files/isad_g_2e.pdf)
- LENAT, B. (1995). CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM*, 38(11), 32-38.
- McGUINNESS, D.L. y HARMELEN, F. (2004). OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004. En línea: <http://www.w3.org/TR/owl-features/>
- NOY, N.F., SINTEK, M., DECKER, S., CRUBEZY, M., FERGERSON, R.W., MUSEN, M.A. (2001). Creating semantic web contents with Protege-2000. *IEEE Intelligent Systems*, 16, 60-71.
- SICILIA, M.A. (2005). On the Use of Existing Upper Ontologies as a Metadata Integration Mechanism. En *Proceedings of the 2005 Dublin Core Conference*, Madrid.
- SICILIA, M.A. y GARCÍA-BARRIOCANAL, E. (2005) On the Convergence of Formal Ontologies and Standardized e-Learning. *Journal of Distance Education Technologies*, 3(2), 12-28.
- SICILIA, M.A., LYTRAS, M., RODRIGUEZ, E. y GARCIA-BARRIOCANAL, E. (2006) Integrating descriptions of knowledge management learning activities into large ontological structures: A case study. *Data & Knowledge Engineering*, 57(2), 111-121.
- SICILIA, M. A., GARCÍA, E., AEDO, I., DÍAZ, P. (2005) Using Links to Describe Imprecise Relationships in Educational Contents. *International Journal for Continuing Engineering Education and Lifelong Learning* 14(3), pp. 260-275.
- USCHOLD, M., GRUENINGER, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review* 11, 93-155.
- WIELINGA, B.J., SCHREIBER, A., WIELEMAKER, J. y SANDBERG, J.A.C. (2001). From thesaurus to ontology. *Actas de First International Conference on Knowledge Capture (K-CAP 2001)*, October 21-23, 2001, Victoria, BC, Canada, pp. 194-201.