

## F. 10. ISO-THES: Ampliando Skos a partir de la norma de tesauros ISO 25964

Juan-Antonio Pastor-Sánchez

29 octubre 2012

**Pastor-Sánchez, Juan-Antonio** (2013). "ISO-THES: Ampliando Skos a partir de la norma de tesauros ISO 25964". *Anuario ThinkEPI*, v. 7, pp. 189-193



**Resumen:** La publicación de la nueva norma de tesauros *ISO 25964* ha coincidido con un alto grado de implantación de la ontología *Skos* desarrollada por el *W3C*. Es preciso solucionar el problema que suponen las diferencias entre ambas propuestas. Actualmente se está trabajando en el desarrollo de una extensión de *Skos* que permitiría utilizar muchas de las funciones que ofrece el modelo de datos de la nueva norma.

**Palabras clave:** *ISO 25964*, *ISO-THES*, Tesauros, *Skos*.

**Title:** *ISO-THES: Enlarging Skos through the thesaurus standard ISO 25964*

**Abstract:** The publication of the new *ISO 25964* thesaurus standard has coincided with a high level of deployment of the *Skos* ontology developed by the *W3C* and should solve the problem posed by the differences between them. There are currently works in progress to create a *Skos* extension that would use many of the functions provided by the data model of the new standard.

**Keywords:** *ISO 25964*, *ISO-THES*, Thesauri, *Skos*.

### Introducción

Durante muchos años las normas *ISO 5964:1985* e *ISO 2788:1986* fueron las únicas referencias normativas para la creación y mantenimiento de tesauros. Si bien es cierto que más adelante se hicieron algunas revisiones parciales, los cambios introducidos fueron poco sustanciales en cuanto a la estructura y elementos de los tesauros (*ANSI/NISO*, 2005). Hubo que esperar a las cinco partes de la norma británica *BS 8723* (2005-2008) y más recientemente a las dos partes de la norma *ISO 25964:2011*, para poder hablar de cambios de envergadura en este sentido. El resultado de estos trabajos ha tenido como fruto unas propuestas realmente innovadoras, en especial en lo referido a la aplicación de los tesauros en los procesos de recuperación de información y la interoperabilidad con otros vocabularios.

Durante este lapso de tiempo se han realizado aplicaciones en la Web que requerían aplicar el concepto de tesoro, pero para las que la noción normativa existente en aquel entonces era insuficiente. *Skos* (Miles; Bechhofer, 2009) se desarrolló como un medio sencillo para representar en la web semántica sistemas de organización

del conocimiento de todo tipo. Actualmente son innumerables las representaciones y conjuntos de datos disponibles de una amplia variedad de vocabularios controlados (tesauros incluidos) en la Web mediante esta ontología. Por otro lado, la nueva norma ofrece un modelo de datos muy alejado de RDF y excesivamente orientado a su implementación mediante bases de datos relacionales. Podría decirse que la nueva norma *ISO 25964* ha aparecido una vez que *Skos* se ha convertido en el estándar de facto para la publicación de tesauros en la web semántica.

Por lo tanto, se plantea una serie de dificultades que compatibilicen la aplicación de *ISO 25964* con la publicación de tesauros desarrollados bajo dicha norma en la web semántica. En esta nota se analiza una solución que está llevándose a cabo actualmente: la definición de una extensión de *Skos* para adaptar esta ontología a la nueva norma *ISO*.

### *ISO 25964* y *Skos*

Uno de los aspectos en los que *Skos* diverge de las primeras normas de tesauros es en el planteamiento de dos niveles de representación:

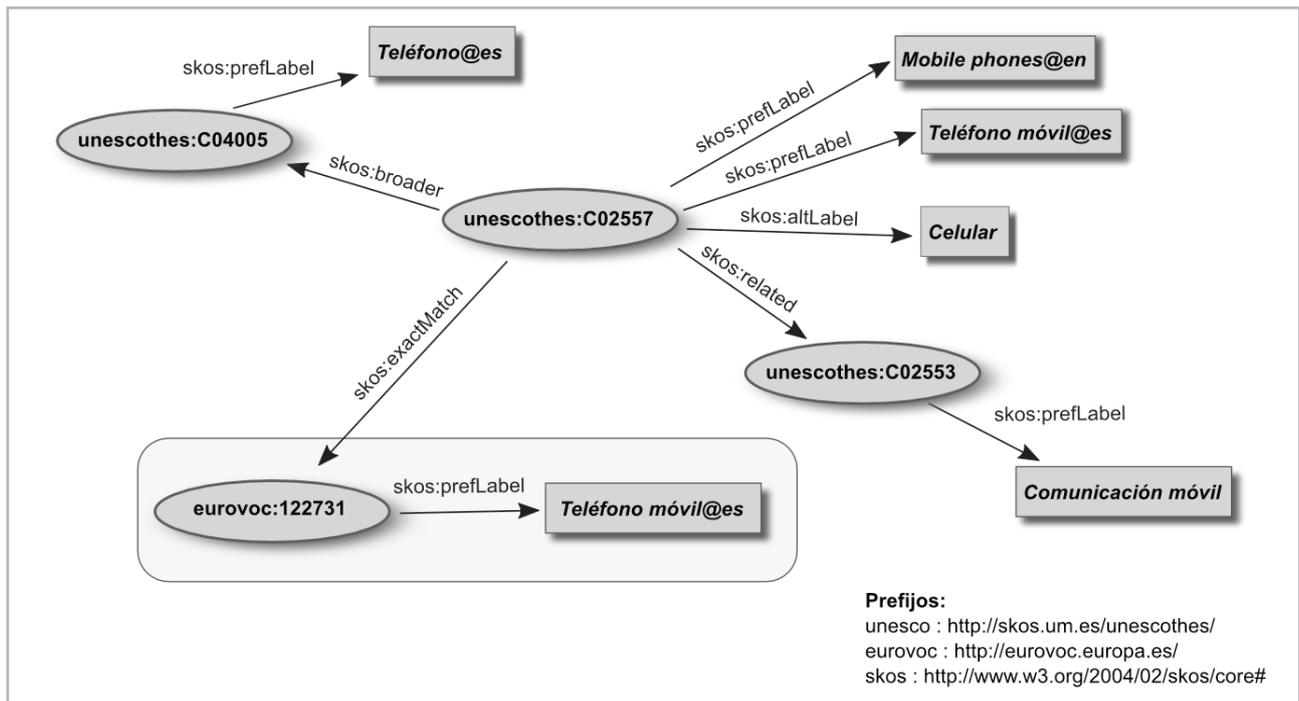


Figura 1. Correspondencia Tesauro de la Unesco y Eurovoc

uno conceptual y otro léxico. Con *Skos* los conceptos se definen como un recurso de la clase "skos:Concept" y se identifican mediante un uri<sup>1</sup>. Una vez identificado el recurso y definido su tipo es posible realizar declaraciones sobre diferentes aspectos del mismo, entre los que se encuentran su etiqueta preferente (término descriptor), etiquetas alternativas (términos no-descriptores), así como las relaciones semánticas (jerárquicas y asociativas) con otros recursos que a su vez se definen como conceptos. *Skos* dispone de dos clases que permiten crear estructuras de agrupación: esquemas de conceptos y colecciones, junto con las correspondientes propiedades para relacionar los conceptos con estas estructuras.

La normas *ISO* y *Ansi/Niso* anteriores a *ISO 25964* ofrecían únicamente un enfoque léxico: se definía una serie de términos descriptores y no-descriptores entre los que se establecían relaciones semánticas tales como "Usado por", "Término específico", "Término genérico", "Término relacionado", etc. Este enfoque resulta problemático desde el punto de vista del mantenimiento de los tesauros y de los procesos de indización: con *Skos* los documentos se asocian a conceptos del tesauro, mientras que con un tesauro elaborado bajo el paradigma léxico, estos mismos documentos se asocian a términos. Los cambios de términos descriptores y no-descriptores no afectan a la indización si se utiliza *Skos*, puesto que únicamente suponen modificaciones en la asignación de etiquetas a los conceptos.

Sin embargo, esto no sucede así con los tesauros terminológicos, puesto que los cambios en

la estructura de descriptores y no-descriptores conlleva la actualización de los índices que relacionan los términos con los documentos e incluso de las relaciones que se hayan podido definir con términos de otros tesauros<sup>2</sup>.

La figura 1 muestra un ejemplo abreviado de representación de un concepto del *Tesauro de la Unesco* con *Skos* y su conexión con un concepto de *Eurovoc*. En la misma puede observarse como incluso los términos equivalentes en diferentes idiomas se representan como etiquetas asociadas al mismo concepto. Puede verse como este enfoque resulta mucho más sencillo para representar tesauros multilingües. En resumen: con *Skos*, un cambio en el etiquetado de los conceptos no implica ningún cambio en la estructura de relaciones semánticas.

La nueva norma *ISO 25964* está mucho más cercana a *Skos* por varios motivos:

1. Define un modelo de datos que facilita el desarrollo de aplicaciones informáticas y su aplicación en procesos de recuperación de información, tal y como se muestra en la figura 2<sup>3</sup>.
2. En dicho modelo se establecen un nivel conceptual y un nivel terminológico (léxico).
3. Además de los conceptos y los términos, se definen varias clases de estructuras superiores para representar tesauros ("Thesaurus"), grupos de conceptos ("ConceptGroup") y listas estructuradas de conceptos ("ThesaurusArray").
4. Los elementos de anotación son prácticamente idénticos a los de *Skos*.

La gestión de la estructura a nivel léxico de

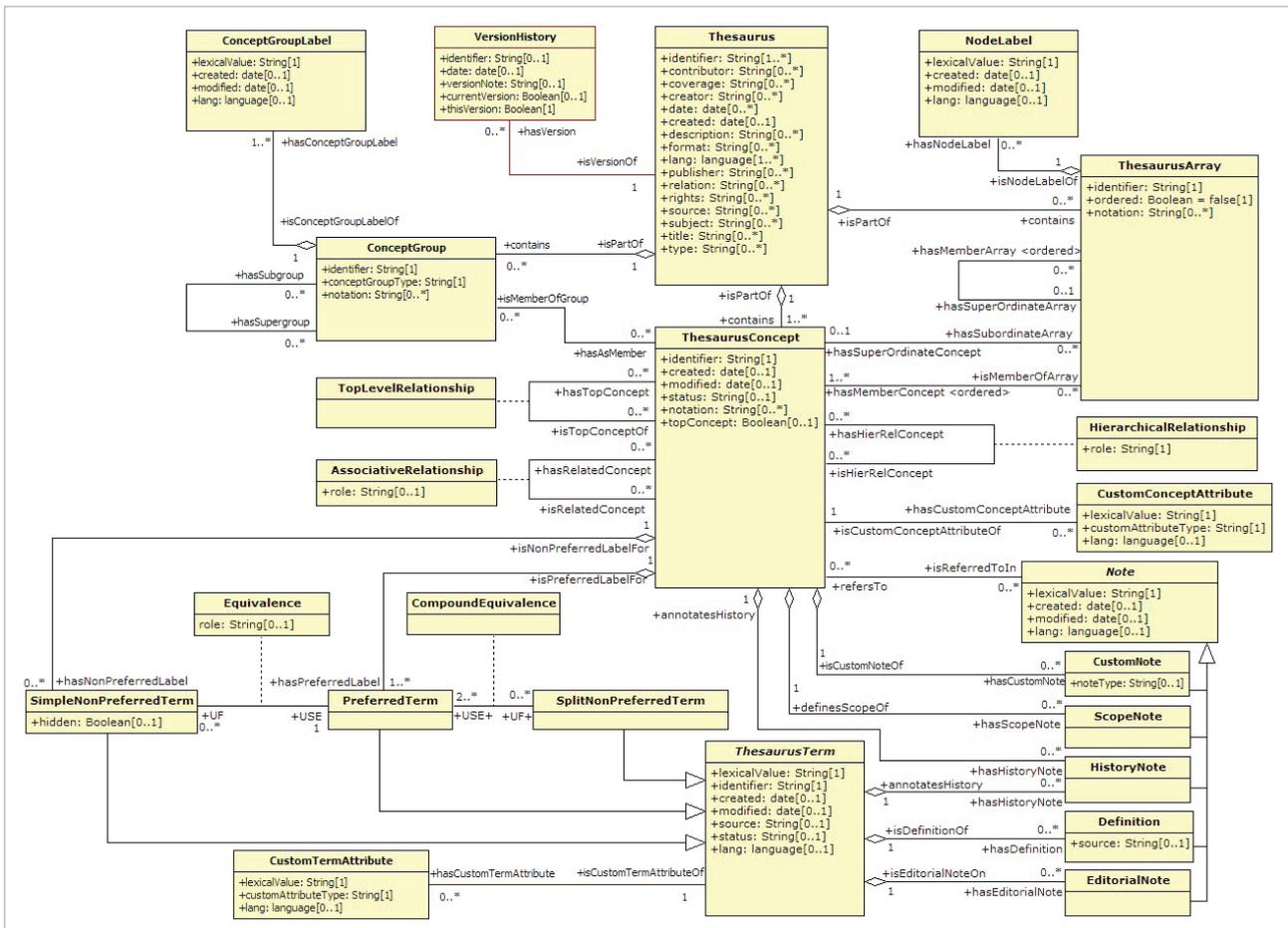


Figura 2. Modelo de datos definido por la ISO 25964

la norma ISO es especialmente compleja. Esto se debe a que se definen relaciones de equivalencia entre términos (“USE/USE+” y “UF/UF+”) que complementan las que se establecen entre términos y conceptos (“isNonPreferredLabelFor”, “isPreferredLabelFor”).

Dichas relaciones permiten la equivalencia de un término compuesto no-preferente con varios términos preferentes a partir de una relación de equivalencia compuesta. En Skos no es posible definir estas equivalencias, ni siquiera recurriendo a la extensión Skos-XL (Skos eXtension for Labels)<sup>4</sup>.

La segunda parte de la norma ISO 25964 aborda la interoperabilidad de los tesauros con otros vocabularios. Se definen diferentes relaciones y técnicas para el mapeado de equivalencia entre conceptos de un tesoro con los de otro tesoro, clasificación, taxonomía, encabezamientos de materia, ontologías, terminologías, listas de autoridades o anillos de sinónimos. Estas equivalencias pueden ser:

a) Exacta: se trata de conceptos que representan exactamente la misma idea. Por ejemplo: “enfermedad de las vacas locas”, “encefalopatía espongiiforme bovina”.

b) Inexacta: conceptos que pueden ser equi-

valentes en determinados contextos o tener diferencias de connotación. Por ejemplo: “sillas” y “asientos”.

c) Parcial: son conceptos con un significado muy cercano aunque uno de ellos es ligeramente más genérico/específico que otro. Por ejemplo: “Aeronaves” y “Aeroplanos”.

d) Jerárquicas: un concepto es claramente más genérico o específico que otro. Por ejemplo: “Países Bajos” es más genérico que “Holanda”.

e) Asociativas: los conceptos no son equivalentes pero están relacionados entre sí. Por ejemplo: “Fotografías” y “Fotógrafos”.

Las equivalencias pueden definirse entre dos conceptos de tesauros diferentes (equivalencia simple) o entre un concepto de un tesoro y varios de otro (compuesta). En este último caso la equivalencia puede establecerse a partir de la intersección o la unión del significado de los conceptos del tesoro de destino. Las relaciones de equivalencia no están reflejadas explícitamente en el modelo de datos de la norma, aunque pueden derivarse a partir de relaciones jerárquicas y asociativas.

En este aspecto la aplicación de Skos ha acumulado una larga experiencia, puesto que es ampliamente utilizado para la interconexión de

vocabularios. Skos dispone de propiedades para definir equivalencias exactas, próximas, jerárquicas y asociativas. No obstante dichas propiedades únicamente permiten definir relaciones de equivalencia simples y no compuestas, tal y como se establece en la *ISO 25964:2011-2*.

## La extensión ISO-THES

Como ya se apuntó al principio, la nueva norma *ISO* llega a un escenario en el que *Skos* se ha convertido en el estándar para la publicación de todo tipo de vocabularios controlados.

Para facilitar la interoperabilidad entre ambas soluciones, el grupo de trabajo responsable de la norma *25964* y **Antoine Isaac** (co-editor de *Skos*) han editado un documento que define las correspondencias entre ambos modelos (*ISO; Isaac*, 2012).

A grandes rasgos las propuestas de dicho documento pueden resumirse en los siguientes puntos:

- Se definen correspondencias de los elementos básicos de *Skos* (conceptos, esquemas y colecciones)<sup>5</sup> con los elementos del modelo de datos de la norma *ISO*.

- Identifica los elementos de *Dublin Core* a utilizar para describir ciertas propiedades, tanto de los tesauros como de los conceptos.

- Se establece un amplio conjunto de clases y propiedades que no están presentes de forma nativa en *Skos* para la definición de equivalencias entre términos.



Antoine Isaac

### Correspondence between ISO 25964 and SKOS/SKOS-XL Models

#### Contributors

This document was created by:

- ISO TC46/SC9/WG8 working group for the ISO 25964 standard about Thesauri
- Antoine Isaac: Co-Editor of the SKOS Recommendations and MADS

Work was completed between 2012-02-20 and 2012-06-23.

#### Objective

This document is intended as a correction and/or update to the Appendix “Correspondences between ISO-2788/5964 and SKOS constructs” of the [SKOS Simple Knowledge Organization System Primer](#). This update is needed because [ISO 25964-1:2011](#) has been published, replacing the earlier ISO standards [ISO 2788:1986](#) and [ISO 5964:1985](#).

[http://www.niso.org/apps/group\\_public/download.php/9507/Correspondence\\_ISO25964-SKOSXL-MADS-2012-09-16.pdf](http://www.niso.org/apps/group_public/download.php/9507/Correspondence_ISO25964-SKOSXL-MADS-2012-09-16.pdf)

- Igualmente propone clases y propiedades para ampliar las estructuras de agrupación de conceptos para adaptar *Skos* a la nueva norma. Más concretamente, crea las clases y las propiedades necesarias para representar con *Skos* listas estructuradas de conceptos y grupos de conceptos.

- Se proponen relaciones para definir jerarquías de grupos de conceptos que son de gran utilidad para la representación de tesauros organizados en dominios de conocimiento y microtesauros.

- Resuelve la problemática de las ediciones diferentes de un mismo tesoro, principalmente pensando en su explotación a través de un *Sparql Endpoint*. Para ello también se incluye una serie de clases y propiedades que permiten identificar versiones distintas para su publicación en forma de datos abiertos enlazados.

Como puede verse, esta extensión aborda la problemática de la representación de jerarquías de microtesauros. Varios tesauros (como *Eurovoc*, el *Tesoro de la Unesco* o *Gemet*) utilizan estas estructuras en la que los conceptos se organizan en grupos diferentes que a su vez están dentro de dominios de conocimiento o grupos más amplios. Hasta ahora esta necesidad se había suplido modelando los dominios y los microtesauros como colecciones y definiendo la pertenencia de estos últimos a los primeros.

Otra alternativa muy utilizada ha sido la definición de artificios en la que un tesoro se modelaba como diferentes esquemas de conceptos que se jerarquizaban mediante una propiedad *ad hoc*.

En cualquier caso, hasta ahora no se disponía de una técnica normalizada para representar este tipo de estructuras. La sinergia entre *ISO 25964* y *Skos* propone una solución mediante la aplicación de la clase “*iso-thes:ConceptGroup*” y sus correspondientes propiedades.

Sin duda uno de los aspectos más complejos de la norma *ISO* y de *ISO-THES* es la equivalencia entre términos, y muy especialmente la equivalencia compuesta, aunque su uso es innecesario en muchos tesauros. Sin embargo, conforma un

nuevo camino que sin duda deberá ser recorrido por tesauros que, como *Eurovoc*, desarrollaron sus propias ontologías para cubrir necesidades concretas en este aspecto.

## Conclusiones

*ISO-THES* constituye una solución idónea para la interoperabilidad semántica entre vocabularios publicados con *Skos* y los tesauros que sigan la norma *ISO 25964*. Aunque la norma ofrezca un esquema xml<sup>6</sup> para la representación de tesauros, lo cierto es que dicho esquema resulta poco útil para su publicación en la web semántica. Con *ISO-THES* se tienden puentes entre dicho esquema y RDF, de manera que la práctica totalidad de los elementos de la norma tienen su equivalente en *Skos* sin necesidad de definir una nueva ontología. De este modo, los editores de un tesoro creado conforme a la norma *ISO* disponen de un mecanismo casi directo para su publicación como un conjunto de datos RDF.

No obstante, quedan varios aspectos por resolver. Tal vez el más significativo sea la ausencia de mecanismos para definir relaciones de equivalencia compuesta entre conceptos de diferentes vocabularios. Esta ausencia podría deberse a la indefinición de *ISO 25964*, ya que el modelo de datos no incorpora de un modo explícito toda la riqueza de relaciones de mapeado que recoge la segunda parte de la norma. Por otro lado se echa de menos todo un conjunto de clases y propiedades que permitan la representación de información de gestión sobre conceptos y términos, que permitan prescindir del uso de propiedades de anotación OWL, de elementos de *Dublin Core* o de los propios mecanismos de reificación<sup>7</sup> RDF.

Todavía queda un paso importante: la elaboración de un vocabulario RDF (y su correspondiente "namespace") con los diferentes elementos de la extensión. Ciertamente, es muy posible que en este proceso se plantee la necesidad de realizar cambios y definir nuevos elementos de *ISO-THES*. Pero en definitiva, se trata de una propuesta que se encuentra en el buen camino y que sin duda reforzará el papel de *Skos* en la web semántica.

## Notas

1. También es posible utilizar identificadores que no se refieran a un uri mediante la definición de nodos en blanco. Más información en: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#structuredproperties>

2. Se podría argumentar que estos cambios en los tesauros terminológicos únicamente precisan de consultas de actualización que pueden ser automatizadas. Sin

embargo, resulta mucho más consistente adoptar el enfoque conceptual, ya que no solamente evita este tipo de operaciones, sino que dota a los tesauros de una mayor flexibilidad para la interconexión con otro tipo de vocabularios, tal y como establece la propia norma *ISO 25964*.

3. Fuente:

[http://www.niso.org/schemas/iso25964/Model\\_2011-06-02.jpg](http://www.niso.org/schemas/iso25964/Model_2011-06-02.jpg)

4. Más información sobre *Skos-XL* en:

<http://www.w3.org/TR/skos-reference/skos-xl.html>

5. Los elementos de *Skos* están precedidos del prefijo "skos:".

6. Dicho esquema está disponible en:

[http://www.niso.org/schemas/iso25964/iso25964-1\\_v1.4.xsd](http://www.niso.org/schemas/iso25964/iso25964-1_v1.4.xsd)

7. Reificación: Declaración de una declaración.

## Referencias

ANSI/NISO (2005). *Z39.19: 2005 Guidelines for the construction, format and management of monolingual controlled vocabularies*. NISO Press: Bethesda, MD.

BS (2005). *BS 8723-2:2005 Structured vocabularies for information retrieval. Guide. Thesauri*. British Standards Institution: London.

BS (2007). *BS 8723-4:2007 Structured Vocabularies for information retrieval. Guide. Interoperability between vocabularies*. British Standards Institution: London.

BS (2008). *BS 8723-5:2008 Structured vocabularies for information retrieval. Guide. Formats and protocols for data exchange*. British Standards Institution: London.

ISO (1985). *ISO 5964:1985. Guidelines for the establishment and development of multilingual thesauri*. Geneva: International Organization for Standardization.

ISO (1986). *ISO 2788:1986. Guidelines for the establishment and development of monolingual thesauri*. Geneva: International Organization for Standardization.

ISO (2011a). *ISO 25964-2:2011. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneva: International Organization for Standardization.

ISO (2011b). *ISO 25964-2:2011. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization.

ISO TC46/SC9/WG8; Isaac, Antoine (2012). "Correspondence between ISO 25964 and Skos/Skos-XL models". <http://www.niso.org/schemas/iso25964/correspondencesSKOS>

Miles, Alistair; Bechhofer, Sean (18 Aug. 2009). *Skos (simple knowledge organization system) reference. W3C recommendation*. <http://www.w3.org/TR/skos-reference>