

El lenguaje sobre el patrimonio

Ministerio
de Educación, Cultura
y Deporte

Estándares documentales para la descripción
y gestión de colecciones

eLP

El lenguaje sobre el patrimonio

Estándares documentales para la descripción y gestión de colecciones

Catálogo de publicaciones del Ministerio: www.mecd.gob.es
Catálogo general de publicaciones oficiales: publicacionesoficiales.boe.es

Edición 2016



MINISTERIO DE EDUCACIÓN, CULTURA
Y DEPORTE

Edita:
© SECRETARÍA GENERAL TÉCNICA
Subdirección General
de Documentación y Publicaciones

© De los textos y las fotografías: sus autores

NIPO (electrónico): 030-16-459-0
DOI: 10.4438/030-16-459-0

ÍNDICE

	Pág.
<i>Prólogo</i>	6
Un vocabulario común para un patrimonio diverso: el proyecto de tesauros del patrimonio cultural de España del Ministerio de Educación, Cultura y Deporte	7
<i>María Carrillo Tundidor</i>	
Presentation RKD terminologies, RKDartists, Iconclass and (the Dutch) AAT	15
<i>Reem Weda</i>	
Normas y estándares para la elaboración de tesauros de patrimonio cultural	29
<i>Javier García Marco</i>	
Los tesauros en la web semántica: SKOS y la norma ISO 25964	47
<i>Juan Antonio Pastor Sánchez</i>	
El control terminológico en el Museo Arqueológico Nacional: evolución histórica, situación actual y perspectivas de futuro	63
<i>Javier Rodrigo del Blanco</i>	
La memoria recuperada: la búsqueda del nombre. La construcción del <i>Tesaurus de Denominaciones de Bienes Culturales</i>	76
<i>Isabel Trinidad Lafuente</i>	
Tesauros, vocabularios y listas de control terminológico. Fuentes para la construcción de tesauros culturales	93
<i>Miguel Ángel Marzal García-Quismondo</i>	
Una ontología para la gestión del conocimiento: el Tesaurus de la UNESCO y su aplicación en el Portal sobre Aprendizaje «Learning Portal» del Instituto Internacional de Planeamiento de la Educación de la UNESCO	105
<i>Asunción Valderrama Pareja</i>	
Construcción y uso de terminologías, categorías de descripción y estructuras semánticas vinculadas al patrimonio en la sociedad global de datos	115
<i>Nuria Rodríguez Ortega</i>	

La normalización terminológica aplicada a la conservación y restauración del patrimonio cultural	131
<i>Ana María Calvo Manuel</i>	
Dimensión política y cultural en la armonización terminológica multilingüe: la experiencia del tesoro para la Red de Información del Patrimonio Europeo (HEREIN)	140
<i>Teresa Abejón Peña</i>	
Creación, ejecución y desarrollo de un tesoro temático para los museos aragoneses: las tipologías arqueológicas	147
<i>José Fabre Murillo</i>	
El listado de autoridades y la normalización terminológica en el Museo del Prado	157
<i>José Luis Cueto Martínez-Pontrémuli</i>	
Normalización terminológica en bibliotecas: autoridades de materia y términos de género/forma en la Biblioteca Nacional	169
<i>Pascual Jiménez Huerta</i>	
PARES hacia la web semántica: autoridades y tesauros en los archivos estatales	181
<i>Aránzazu Lafuente Urién</i>	

Prólogo

Miguel Ángel Recio Crespo

Director General de Bellas Artes y Bienes Culturales y de Archivos y Bibliotecas

Los tesauros de patrimonio cultural son recursos que permiten normalizar el vocabulario y los conceptos utilizados para la identificación, descripción y clasificación de los bienes culturales, facilitando el uso de un lenguaje común que permite el intercambio de información entre diferentes tipos de instituciones gestoras de patrimonio. Se trata de instrumentos fundamentales para todos los profesionales que se ocupan de la documentación, investigación y gestión del patrimonio, y también constituyen un elemento que permite y facilita el acceso del público a la información.

La Subdirección General de Museos Estatales trabaja desde hace años en la elaboración y construcción de vocabularios normalizados. Por su parte, el Instituto del Patrimonio Cultural de España lleva a cabo en la Escuela de Patrimonio Histórico de Nájera (La Rioja) un programa anual de actividades formativas sobre diversos aspectos relacionados con la conservación y estudio del patrimonio. Dentro de dicho programa el IPCE organizó en noviembre de 2015, el curso: “El lenguaje del patrimonio. Estándares documentales para la descripción y gestión de colecciones”.

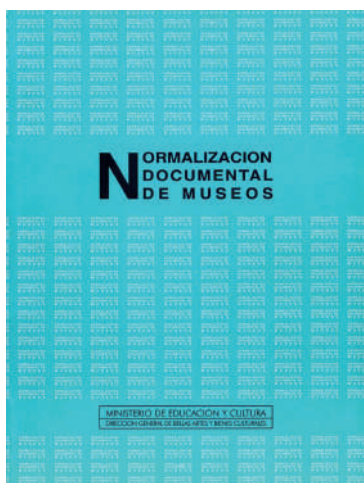
El curso ha permitido reunir aportaciones de especialistas de la universidad, la administración pública y de instituciones nacionales e internacionales que trabajan en la actualidad en el tratamiento técnico de los vocabularios, así como en su normalización, difusión y aplicación práctica a los sistemas de información de los diferentes repositorios patrimoniales actualmente disponibles. La publicación de los resultados supone, por tanto, una excelente oportunidad para acercarse a las diferentes iniciativas que se están desarrollando en este ámbito.

Un vocabulario común para un patrimonio diverso: el proyecto de tesauros del patrimonio cultural de España del Ministerio de Educación, Cultura y Deporte

María Carrillo Tundidor

Jefa del Servicio de Documentación. Ministerio de Educación, Cultura y Deporte

En el año 2016 se cumplen veinte años de la publicación del informe «Normalización documental de museos: elementos para una aplicación informática de gestión museográfica» (Carretero *et alii*, 2006), con la que se presentó en sociedad el proyecto de normalización documental del Ministerio de Educación, Cultura y Deporte. Esta publicación, que sentó las bases para el desarrollo posterior del proyecto de normalización documental y el de tesauros que tratamos aquí, recogía el análisis que había desarrollado la Comisión de Normalización Documental de Museos creada en 1994 sobre la situación en la que se encontraba la documentación en los museos estatales, e incluía las funcionalidades y requisitos que debía cumplir la aplicación informática que se desarrollaría a partir de la publicación del informe y que con el tiempo se denominaría DOMUS.



Veinte años es un tiempo suficiente para valorar las cualidades evidentes que el proyecto, en general, y la publicación que mencionamos, en particular, han tenido para el eficaz desarrollo de un sistema documental compartido en los museos. Durante estos años el proyecto ha ido creciendo en número de instituciones y, gracias a las aportaciones de los técnicos de los museos de diferentes titularidades, se ha ido enriqueciendo con más funcionalidades y evoluciones técnicas. Muchas de las cuestiones básicas sobre las que se asentó el proyecto continúan vigentes a día de hoy, y una de ellas es, sin duda, la necesidad de establecer un control terminológico sobre las tareas de catalogación, inventario y gestión de las colecciones de los museos.

Textualmente, la publicación mencionaba que debía procederse «a la unificación de normas descriptivas y terminológicas de las diversas áreas científicas, técnicas y administrativas». Y señalaba que este control terminológico debía aplicarse «a la denominación de los ítems, sus materias primas, técnicas de fabricación, denominación de componentes, etc.», y avanzaba ya la necesidad de constituir grupos de trabajo para las áreas que podrían resultar más problemáticas, refiriéndose concretamente a las clasificaciones genéricas, contextos culturales, clasificaciones geográficas e iconografía. Es curioso comprobar cómo esta profecía que adivinaba dificultades en el horizonte de algunos campos de información concretos se ha visto cumplida, si tenemos en cuenta que a día de hoy, a pesar del gran número de vocabularios aparecidos, siguen sin completarse los tesauros de clasificaciones genéricas, iconografía y contextos culturales (aunque esperamos que este último esté pronto en la calle).

Dado que la publicación era tremendamente explícita en las herramientas de control terminológico que utilizaría la futura aplicación informática (tesauros, listas y tablas de control de términos), tras su publicación se pusieron en marcha los grupos de trabajo que debían desarrollar estos tesauros, para lo cual se contó con reconocidos expertos en cada área de estudio, que comenzaron a elaborar tesauros específicos por especialidades. Fruto de este trabajo se publicaron el *Diccionario de materiales cerámicos* en el año 2002 (Cabrera *et alii*, 2002), el *Diccionario de mobiliario* en el año 2006 (Rodríguez, 2006) y el *Diccionario de numismática* en el año 2009 (Alfaro *et alii*, 2009).

Aunque redactadas por especialistas de museos diferentes, estas tres publicaciones comparten dos objetivos fundamentales: por un lado, servir como herramienta de trabajo a la hora de catalogar las colecciones patrimoniales de cada uno de los ámbitos de estudio y, por otro, aportar una sistematización terminológica que limite el empleo del lenguaje espontáneo en la catalogación. Y es precisamente la conjunción de estos dos objetivos, complementarios pero diferentes, lo que ha convertido estas publicaciones en herramientas indispensables para la catalogación, el estudio y el conocimiento de las colecciones.

Para ello, estas publicaciones siguen un mismo esquema de contenidos, que había sido previamente fijado en el *Diccionario del dibujo y la estampa* (Blas *et alii*, 1996), una publicación que, al igual que las tres que mencionamos, se había iniciado en el seno de la comisión de normalización documental constituida en 1994 y que fue finalmente publicado por la Real Academia de Bellas Artes de San Fernando en el año 1996. En esta primera publicación se habían fijado los elementos esenciales que iban a seguir el resto de diccionarios, pasando todos ellos a incluir a partir de entonces una introducción, el diccionario de términos alfabéticamente ordenados con las definiciones asociadas a cada uno de ellos y el tesoro propiamente dicho, compuesto, por una parte, por la estructura general y, por otra, por el cuerpo del tesoro. En el cuerpo del tesoro de cada uno de los diccionarios publicados se incluyen los descriptores y no descriptores con las relaciones que establecen entre ellos, identificándose términos genéricos (TG), términos específicos (TE), términos relacionados (TR), términos «usado por» (UP) y términos «use» (USE). Las publicaciones aportan además información adicional de interés para el catalogador: normas generales de catalogación, ejemplos de fichas de catalogación, figuras, láminas y bibliografía. Se empezó a trabajar en ese momento con otros ámbitos, por ejemplo, los textiles, el vidrio o el material científico, pero son tesauros que aún no han visto la luz y que se encuentran en vías de elaboración.

Aunque estos tesauros y diccionarios han supuesto una aportación indiscutible para el conocimiento y estudio de las colecciones de los museos, el Ministerio de Educación, Cultura y Deporte inició, al tiempo que la aplicación DOMUS se instalaba en los museos, otros trabajos en el área del control terminológico que completaban los ya mencionados. Efectivamente, como ya hemos mencionado, muchos de los campos de información de DOMUS estaban controlados por tesauros y listas de control terminológico, lo que quiere decir que, en el momento de cumplimentar cualquier registro de catalogación, es necesario que el término que queremos utilizar haya sido introducido previamente en alguno de estos tesauros y listas. La publicación «Normalización Documental» incluía algunas listas que podrían emplearse a la hora de empezar a trabajar, pero se referían sobre todo a campos de información administrativa, como, por ejemplo, las formas de ingreso de fondos, tipos y unidades de medida, tipos de movimiento de fondos o las series documentales del archivo administrativo.

La utilidad de estas listas ha sido fundamental y su publicación temprana fue un buen empuje para el objetivo de normalización que se quería conseguir: por poner un ejemplo, los archivos de los museos siguen en la actualidad un mismo esquema de organización documental, independientemente de su organización interna, contenido o incluso administración gestora. No obstante, fuera del ámbito de la publicación, habían quedado un gran número de campos «científicos» empleados a la hora de catalogar un fondo: clasificación genérica, denominación del objeto, título, técnica, materia, iconografía... Todos los museos empezaron a trabajar en los tesauros adaptándolos a las necesidades de sus colecciones y a la información que ya se contenía en las herramientas documentales que cada museo había utilizado con anterioridad y que se migró o se volcó en la nueva aplicación. El resultado fue que los museos fueron elaborando tesauros de forma interna a espaldas del resto de instituciones, con el único objetivo de racionalizar la información contenida en sus propias bases de datos.

Para paliar esta situación que se había empezado a producir con la instalación de la nueva aplicación DOMUS, a partir del año 2001 el Ministerio de Educación, Cultura y Deporte inició una segunda vía de actuación en la normalización terminológica, dirigida a elaborar tesauros generales que sirvieran para cualquier tipo de colección, mediante la contratación externa de personal especializado. Para ello, se escogieron áreas de trabajo que tuvieran su correspondencia en determinados campos de información de DOMUS: clasificación genérica, materias, técnicas, autores, iconografía, lugares geográficos y toponimia histórica, denominaciones de bienes culturales (para el campo «objeto»), iconografía y contextos culturales.



Fruto de este trabajo, han visto la luz hasta ahora: el *Diccionario de materias* en el año 2008 (Kroustallis, 2008); el *Diccionario de objetos asociados a ritos, cultos y creencias* en

2011 (Trinidad, 2011); el *Diccionario de objetos asociados a la expresión artística* en el año 2012 (Trinidad, 2012), y el *Diccionario de técnicas* en el año 2016 (Kroustallis, 2016). Está ya ultimando su aparición el *Diccionario de objetos asociados a la arquitectura y estructuras*, que se publicará también en el año 2016, y que, junto con los otros dos diccionarios de objetos ya publicados, forman parte del ambicioso *Diccionario de denominaciones de objetos culturales*, cuya inabarcable dimensión ha obligado a publicarlo por ramas. Se prevé que en el año 2017 se publiquen el *Diccionario de toponimia histórica* y el *Diccionario de contextos culturales*.

El conjunto de todos los tesauros que se iban elaborando, tanto los específicos de cada área como los generales, requerían de herramientas informáticas que permitieran su construcción, elaboración, publicación y posterior distribución en las bases de datos instaladas en los museos usuarios de DOMUS. Para ello, el Ministerio de Educación, Cultura y Deporte puso en marcha en el año 2004 el diseño de dos herramientas informáticas denominadas Jerartes y Convertes (Alquézar; Carrasco, 2005). Al igual que había sucedido con DOMUS, el objetivo era contar con un software que se ajustara a las necesidades específicas del proyecto, por lo que fue el Ministerio el que desarrolló las dos aplicaciones, de las que es propietario.

Jerartes es la aplicación que se emplea para la construcción de los tesauros, controla los términos que se incluyen en los mismos y las relaciones que se establecen entre ellos. Permite elaborar tesauros en paralelo, estableciendo para cada uno de ellos grupos diferenciados con bases de datos independientes. Para la elaboración de la jerarquía, pueden darse de alta tanto descriptores como indicadores clasificatorios. Estos últimos solo se introducen para facilitar la agrupación de los descriptores, por lo que no forman parte del tesoro en el que se está trabajando. Los descriptores establecen entre ellos relaciones, por lo que cada uno de ellos puede designarse, con respecto a otro descriptor incluido en el tesoro, como término genérico, término específico y término relacionado. Cada descriptor incluido lleva asociados diferentes campos de información: la definición del término, nota de alcance (condiciones de utilización del término o notas para su reconocimiento o diferenciación), referencia bibliográfica (publicaciones de las que procede la definición ofrecida, o que tengan especial interés para el conocimiento o estudio del término o aspectos directamente relacionados con él), observaciones (anotaciones sobre contexto histórico, particularidades de uso, etc.), fuente (persona o institución que propone la inclusión del término o a partir de cuya información se incluye) y fecha de incorporación o modificación del término en la base de datos. Por otra parte, la aplicación permite incluir la traducción de cada descriptor a otros idiomas, lo que facilita la elaboración de diccionarios multilingües, y la inclusión de imágenes asociadas a cada término.

Una vez finalizado un tesoro, se procede a su distribución en las bases de datos de los museos usuarios de DOMUS mediante la aplicación Convertes. Dado que, como hemos explicado anteriormente, los museos han ido trabajando en sus propios tesauros, Convertes compara las dos versiones de un mismo tesoro (la que tiene el museo y la que el Ministerio ha dado como finalizada) para determinar las coincidencias e incidencias. Para ello, los términos procedentes de los museos se importan a Convertes en formato XML. La aplicación iguala los términos procedentes de los museos con los términos que están en el tesoro finalizado, y el resto de términos que no coincidan quedan englobados en una lista de términos candidatos que hay que procesar manualmente. Los términos candidatos son estudiados y son objeto de dos tipos de acciones: rechazo o igualación con otro término. En este proceso se producen incidencias fáciles de resolver (por ejemplo, igualar singulares y plurales, cambios de género, errores tipográficos...), y otros que son más lentos de determinar: términos que no son aceptados, términos que pueden ser igualados a otros y que pasan a ser «no descriptores» y términos que pueden ser incluidos en la versión definitiva del tesoro. La aplicación genera un fichero de cambios único para cada tesoro y museo, compuesto por aquellos términos que hayan sido igualados y por los que hayan sufrido baja.

Un vocabulario común para un patrimonio diverso: el proyecto de tesauros del patrimonio cultural de España...

Pero el objetivo del proyecto de normalización terminológica no se limita a la evidente utilidad que este trabajo tiene para la catalogación en DOMUS, sino que se dirige a un público mucho más extenso. Para ello, el Ministerio de Educación, Cultura y Deporte puso en marcha en el año 2011 el Portal de Tesauros del Patrimonio Cultural de España (Alonso, 2012), que engloba tanto las publicaciones de cada uno de los tesauros como una aplicación que permite la búsqueda jerárquica y alfabética de los términos de cada uno de ellos. En la actualidad, el Portal de Tesauros del Patrimonio Cultural de España ofrece en línea más de 60 000 descriptores y más de 16 000 no descriptores pertenecientes a siete tesauros diferentes: materias, técnicas, mobiliario, numismática, cerámica, lugares geográficos y denominaciones de objetos culturales (con tres ramas publicadas hasta ahora: objetos asociados a ritos, cultos y creencias, objetos asociados a la expresión artística y objetos asociados a la arquitectura y estructuras).



En los últimos meses esta aplicación ha sido objeto de una completa renovación, convirtiéndose en el primer proyecto que la Subdirección General de Museos Estatales acomete en el ámbito de la web semántica y la publicación en datos abiertos, un camino que empiezan a recorrer muchas instituciones culturales y en el que los museos y colecciones españolas tienen mucho que aportar dada la cantidad y calidad de la información que se contiene en sus bases de datos.

El objetivo es publicar como Linked Open Data el amplio caudal de información que aportan los tesauros. Obviamente, la información ya estaba a disposición del público desde el año 2011 gracias al Portal, pero en el año 2013 empezó a verse la necesidad de que esta información estuviera a disposición no solo de las personas usuarias del portal, sino también de los rastreadores automatizados, a fin de que la información contenida en el Portal fuera accesible para otros recursos de información. Para la publicación de los tesauros en Linked Open Data era necesario, primero, convertir la información a un formato reconocible y aceptado, cumpliendo los requisitos establecidos en el Word Wide Web Consortium. Se eligió el SKOS por ser el más adecuado para la publicación en LOD de vocabularios y ontologías, siendo el sistema que había elegido previamente la Lista de Encabezamiento de Materias desarrollada por la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte, así como otras instituciones que habían empezado a trabajar en datos abiertos, como los tesauros de Getty Research Institute o el British Museum.

Los datos se han modelado a SKOS siguiendo las directrices definidas por el W3C para las «Cool URIs», mediante la asignación a cada uno de los términos de un identificador o URI persistente. Cada uno de los tesauros se ha generado con el siguiente esquema: Esquema del tesoro (skos:Scheme); Concepto (skos:Concept); Facetas (skos:Collection). Por cada descriptor se ha creado un concepto SKOS, mapeando sus campos a tripletas RDF: *skos:prefLabel* (término); *skos:inScheme* (tesauro al que pertenece); *skos:definition* (definición); *dct:bibliographicCitation* (referencias bibliográficas); *skos:scopeNote* (nota de alcance) y *dct:created* (fecha de creación del término). A cada término relacionado se le ha asignado la etiqueta *skos:related*; a los términos genéricos, *skos:broader* y *skos:topConceptOf*; y a los términos jerárquicamente dependientes, *skos:narrower* y *skos:hasTopConcept*. Por cada no descriptor se han formado tripletas en el concepto SKOS, creado para su descriptor asociado con la etiqueta *skos:altLabel*.

En una primera fase, se encargó a la empresa Digibís el modelado en SKOS del *Diccionario de materias* y una primera aplicación para su gestión. Posteriormente, la empresa Serdoc ha realizado el modelado en SKOS del resto de diccionarios y ha definido la aplicación definitiva de gestión de los tesauros en LOD. Para ello, la empresa ha utilizado una aplicación denominada BRÚJULA©, una plataforma de desarrollo de la propia empresa en la que se combina *software open source* con herramientas y programas de desarrollo propio.

La nueva página de Tesauros del Patrimonio Cultural tiene, por tanto, esta duplicidad en el tipo de cliente que debe atender. Por una parte, la nueva aplicación ofrece al usuario humano una experiencia de uso similar al de la aplicación antigua: cuando la persona usuaria realiza una petición (por ejemplo, la consulta en el apartado de «estadísticas»), el sistema recopila la información que se puede mostrar y devuelve la página al usuario. Del mismo modo, cuando el usuario realiza la petición de consultar un término de alguno de los vocabularios, el sistema responderá a la petición redirigiendo al usuario a la URI del término elegido en un formato HTML. Al igual que en el Portal anterior, la búsqueda de términos se puede realizar alfabéticamente, eligiendo la letra inicial del término que se quiere recuperar, o bien jerárquicamente, escogiendo el término en un árbol desplegable. Del mismo modo, se mantiene la redirección a la Red Digital de Colecciones de Museos de España (CER.ES), a la que se puede acceder a través de la imagen asociada al término (y que devolverá el registro completo que tiene ese objeto en CER.ES), o lanzando una búsqueda que recuperará todos los resultados de CER.ES del término que se elija.

Además de la funcionalidad de la consulta de los diccionarios y sus términos a los usuarios finales, el sistema ofrece otro tipo de información: enlaces de descargas de datos, estadísticas, punto de consulta SPARQL, y aporta un mecanismo de negociación de contenido / redirección 303 para el acceso a recursos LOD (Cool URIs). Para los rastreadores automatizados, los datos se devolverán siempre en formato RDF (rdf/xml, Notation3, rdf/Json y Json-ld).

La nueva presentación de los tesauros como LOD permite, por tanto, que recursos de información ajenos puedan vincular sus propios registros de forma automatizada con los datos que ofrece el Portal de Tesauros, si bien este quiere a su vez convertirse en un recurso de información vinculado con otras fuentes. Para ello, los tesauros incluidos en el Portal han sido enlazados con vocabularios externos que ya se han publicado en Linked Open Data. Hasta el momento, se han realizado vinculaciones con la Lista de Encabezamientos de Materia para Bibliotecas Públicas del Ministerio de Educación, Cultura y Deporte, la Llista de Encapçalaments de Matèria de la Biblioteca de Catalunya, Bibliothèque nationale de France, Getty Art and Architecture Thesaurus, Library of Congress, Deutsche Nationalbibliothek, así como con Dbpedia, Geonames y Getty Thesaurus of Geographic Names (estos últimos en el caso del *Diccionario geográfico*). En total, se han establecido 54 710 enlaces SKOS closeMatch. El proceso de conversión a SKOS y de vinculación a recursos externos no ha sido automático en su totalidad, ha requerido de la intervención humana para analizar previamente la definición, estructura y relaciones internas de cada tesoro que se skosifica.

The screenshot shows the website interface for 'tesauros del Patrimonio Cultural de España'. The main content area displays the entry for 'Cartón' with the following details:

- Término:** Cartón
- URL:** <http://museos.sedoc.es/numismatica/1180000>
- Definición:** Conjunto de varias hojas superpuestas de pasta de papel* que, en estado húmedo, se adhieren unas a otras por compresión y se secan después por evaporación. También puede estar realizado directamente con pasta de papel o con hojas encoledas de papel viejo. En Numismática*, el cartón se emplea con fines diversos como, por ejemplo, para fabricar bandejas** de almacenamiento o, en momentos de penuria, para emitir moneda de necesidad*.
- Referencias bibliográficas:** [Esto es dinero (2001: 87; 167, nº 118-119); Kroustalliz (2008: Voz cartón)]
- Término genérico:** Materiales
- Términos similares en otras colecciones:**
 - Cartón [Bibliothèque Nationale de France]
 - Paperboards [Library of Congress Subject Headings]
 - Cartón [Lista de encabezamientos de materia]
 - Cartó [Lista de encabezamientos de materia de la Biblioteca de Catalunya]
 - <http://hmg.sg.egob.mcu.es/Autoridades/LENA/3201300754/concept>
- Imágenes:**
 - Museo: MUSEO ARQUEOLÓGICO NACIONAL
 - Inventarios: 2001/141/1
 - Objeto: Cartón

The left sidebar contains a hierarchical menu for 'Diccionario de Numismática' with categories like 'Piedra de toque', 'Prensa', 'Prueba de acuñación', etc., and 'Cartón' is highlighted under 'Materiales'.

El proyecto de normalización terminológica intenta, por tanto, adaptarse a los nuevos retos tecnológicos y a las nuevas oportunidades de mejorar la calidad de la información que se ofrece a los usuarios. La incorporación de nuevos vocabularios al Portal y su distribución entre los museos usuarios de DOMUS permitirán mantener vigente el continuado esfuerzo que se ha desarrollado en este campo durante los últimos años.

Bibliografía

- ALFARO ASINS, C.; MARCOS ALONSO, C.; OTERO MORÁN, P., y GRAÑEDA MIÑÓN, P. (2009): *Diccionario de numismática*. Madrid: Ministerio de Cultura.
- ALONSO SÁEZ, R. (2012): «Tesauros de Patrimonio Cultural: Conocimiento en línea al servicio del ciudadano», *Siam. Series Iberoamericanas de Museología*. Edición de Asensio; Cabrera; Asenjo y Castro, vol. 1.
- ALQUÉZAR YÁÑEZ, E. M., y CARRASCO GARRIDO, R. (2005): «JERARTES, CONVERTES y DOMUS. Herramientas para la construcción, distribución y utilización de tesauros en los museos», *Museo*, n.º 10, pp. 1-10.
- BLAS BENITO, J.; CIRUELOS GONZALO, A., y BARRENA FERNÁNDEZ, C. (1996): *Diccionario del dibujo y la estampa*. Madrid: Real Academia de Bellas Artes de San Fernando. Calcografía Nacional.
- CABRERA BONET, P.; MAICAS RAMOS, R., y PADILLA MONTOYA, C. (2002): *Diccionario de materiales cerámicos*. Madrid: Ministerio de Educación, Cultura y Deporte.
- CARRETERO PÉREZ, A.; CHINCHILLA GÓMEZ, M.; BARRACA DE RAMOS, P.; ADELLAC MORENO, M. D.; PESQUERA VAQUERO, I., y ALQUÉZAR YÁÑEZ, E. M. (1996): *Normalización documental de museos: elementos para una aplicación informática de gestión museográfica*. Madrid: Ministerio de Educación y Cultura.

- KROUSTALLIS, S. K. (2008): *Diccionario de materias y técnicas (I. Materias)*. Madrid: Ministerio de Cultura.
- (2016): *Diccionario de materias y técnicas (I. Técnicas)*. Madrid: Ministerio de Educación, Cultura y Deporte.
- RODRÍGUEZ BERNIS, S. (2006): *Diccionario de mobiliario*. Madrid: Ministerio de Cultura.
- TRINIDAD LAFUENTE, I. (2011): *Diccionario de objetos asociados a ritos, cultos y creencias*. Madrid: Ministerio de Cultura.
- (2012): *Diccionario de objetos asociados a la expresión artística*. Madrid: Ministerio de Educación, Cultura y Deporte.

Presentation RKD terminologies, RKDartists, Iconclass and (the Dutch) AAT

Reem Weda

Information specialist terminologies

Introduction

I was invited here to tell you something about the controlled vocabularies that are maintained by my institute the RKD. These being: RKDartists, Iconclass and the Dutch translation of the AAT. I am honored and happy to be here.

Something about myself: I studied History in Groningen. I was assistant for the provincial museum consultant on the matters of collection registration and documentation. In this job I came across the issues concerning terminologies and thesauri when I talked to museum registrars and while giving the a course for standard registration and documentation in museums. The content of the course being sanctioned by the Museum Association of the Netherlands. In this course the use AAT was promoted to museums.

Since I started working on the Dutch AAT six years ago, I became more aware of the systematics behind thesauri, besides the usage as a terminology in which the terms reside that museum registrars need to describe collections. A thesaurus can be more than just a list of terms. For instance a knowledgebase and an great access point for browsing through collections. They are important for shared and connected data especially.

I know now that trying to connect all knowledge about heritage and art in a single thesaurus is probably not a task that any heritage organization can do solely by herself. It might not be possible at all, you might think, and maybe you're right. On the other hand: the AAT does already seem to be to connect an awful lot of generic concepts together, and the overlap with most other heritage terminologies is quite substantial. Linking the best and most important terminologies with the AAT should be able to provide a big improvement to semantically linking heritage data.

Question remains: Who will do this work? It still cannot be automated completely.

Content

Introduction

- The RKD
- RKD vocabularies
 - RKDartists
 - Iconclass
 - Dutch Art & Architecture Thesaurus
- The Getty Vocabularies in LOD
- Discussion

RKD Netherlands Institute for Art History

But first something about where I work. The RKD is located in The Hague, location of the Dutch parliament, and home to the international criminal court. The building you see here is our home, we share the building with the Royal Library among other, so it's not all ours. The RKD is a knowledge institute and the central resource for the study of Dutch and Flemish art in an international context. The RKD opened as a state funded institute to the public in 1932. Since 1995 the RKD has been a private foundation. However, the collections managed by the foundation remain state property. We have about 60 full time employees working at the RKD.

RKD collections

The RKD manages a unique collection of archive, documentation and library material relating to Western art from the late Middle Ages to the present. The nucleus of the collections is made up of documentation on Netherlandish art.

The other parts of the collection are the extensive library with more than 4500 000 items; The RKD's Archival Collection contains 1100 meters of documents; the datasets of the Rkd are being published online through the web portal RKD explore RKD. It provides access to the collections: visual and press documentation, the library and archives as is the biographical data in *RKDartists&*, and the attributions of works of art in *RKDimages*. New information is added to the databases every day.

Besides these collections we host several vocabularies, RKDartists as a name authority, Iconclass and we maintain the Dutch version of the AAT.

Controlled vocabularies RKD

Here we see our vocabularies.

RKDartists: biographical info 250 000 persons related to the arts.

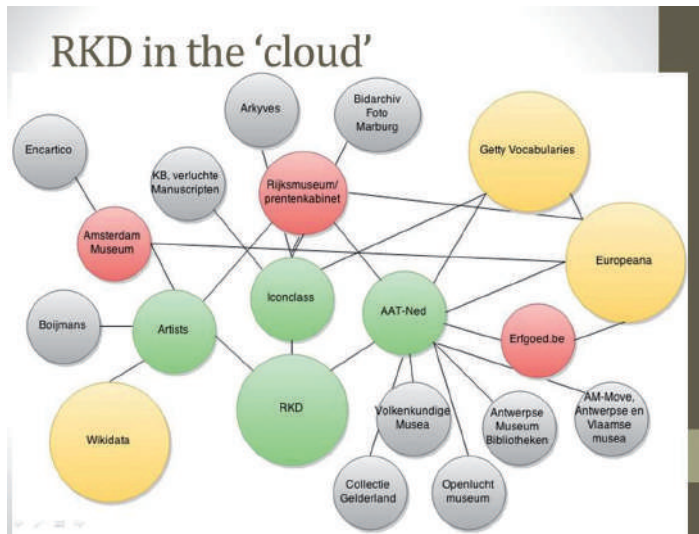
Iconclass: 28 000 classifications for art subjects.

Dutch AAT: 36 000 generic concepts for heritage collections.

I would first like to show you how the vocabularies are situated in the Dutch heritage world.

RKD in the Cloud

Here we see a visualization that I made of the RKD terminologies in the 'cloud' of Heritage data. They are connected to users in the field art like museums, but also other standards and repositories like Europeana and also Wikipedia/wikidata. Most users are only using one vocabulary, while some might use all three. This is not a complete picture but merely a suggestion of the situation.



RKDexplore

For our collections we have a search and discovery portal called RKD explore. To improve the search results and the user experience is an ongoing challenge.



We use local thesauri in all databases, we try to link as much as possible to the AAT, but RKD has built up a distinct local thesaurus over the years. Besides AAT for we use Iconclass to index iconographical subjects on the older artworks in our Images database. In this example you see a record of a painting by Frans Hals, of a portrait of a man dressed as Mars. This images shows the subject keywords and iconclass codes used.

RKDartists is available through Explore and is used by RKD as an index and access point to the other datasets and also the analog art documentation (which is still situated in card board boxes for the large part).

Why still the need?

Here are some examples for identifying synonyms and homonyms.

Why still the need for thesauri vocabs?

Identify Synonyms

- 'salinity' and 'saltiness'
- 'VHS' and 'Very High Frequency'

Identify Homonyms

- Mercury (planet)
- Mercury (metal)
- Mercury (automobile)
- Mercury (mythical being)




For describing and cataloging cultural objects

Vocabularies necessary for:

- More effective exploration and use of collections
- People find more of what they want and what they need
- Standardised cataloging is the key to better acces to collections; it helps people to understand and use other collections that are working the same way.

Example Picasso, five female nudes

Here are some examples of what we do with terminologies

Work Record

Class [controlled]: • paintings • Twentieth Century art

Title: Les Demoiselles d'Avignon Title Type: repository

Title: Five Female Nudes Title Type: descriptive

Bust by Rombout Verhulst, Jacob van reigersberg

Work Type *[link]*: • statue • bust

Title: Bust of Jacob van Reygersberg

Creator Display: Rombout Verhulst (Flemish, 1624-1698)

CCO

Within metadata standards (like CCO) vocabularies and thesauri have a important place. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO)* is a

published manual for describing, documenting, and cataloging cultural works and their visual surrogates. It's a Cross-community standard.

CC principles on vocabularies

I mention the CCO because here the vocabularies are integrated as a essential part in the practice of cataloging. Two of the ten principles of CCO are: use controlled vocabularies AND Create local authorities that are populated with terminology from standard published controlled vocabularies as well as with local terms and names. (Structure local authorities as thesauri whenever possible. Record and document decisions about local authorities).

RKDartists

The content of RKDartists&

This is our database with biographical data for Dutch and non-Dutch artists from the Middle ages till present. Other persons that are of importance to art, like scholars, collectors, and traders are also added to the dataset. First reason for the RKD to develop this controlled list is that it functions as an index to the collection of art historical documentation (pictures, library, archives).

Another important reason was that here we could make persons to one another and to images and documents through a relational database. Since then a lot of useful information about persons related to the arts in the Netherlands where and is being recorded in RKDartists.

At present, the database contains records of approximately 250,000 biographical records. Most of the artists for whom the RKD has documentation have been incorporated, but this is true for relatively few of the art collectors, art dealers and art historians. New records are added daily. We are looking into the possibilities of having trusted partners from outside the organization also work on the data, and even add new records themselves.

Content of a RKDartists record: Rembrandt

Some records have been worked on in greater detail than others. Here we see the example of the record for our most famous Dutch master Rembrandt. There are five name variants registered for him. But the preferred name is simply Rembrandt (with DT). The Artists record is not meant as an comprehensive article about a certain person, but in this case the most noteworthy events in Rembrandts life have been added to the record. Most of the work has been put in connecting the record to other parts of the collection, to books, documentation and documents but also to other artists records, for instance to those that followed in Rembrandts footsteps.

Webservice Open Search

Recently we have made available a web service for RKDartists data.

It is meant for application developers e.g. collection management systems, websites and apps. And is based on: Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF). Not all the data has been made available for search in this service. The Search methods are on a limited set of elements from RKDartists. Unfortunately we haven't put the documentation for the service online yet, and it's in Dutch, but if you're interested we can provide it. Perhaps translate it.

Search on Paulus Potter

It works through a set of standard search methods, for instance on a part of a name, or full name, you will retrieve information from the Artists set. Now if we want to find the data for Paulus Potter, a master in animal paintings, who died quite young. We will find the following data.

EAC-CPF data on Paulus Potter (1625-1654)

Here we see the EAC-CPF data on Paulus Potter. The information is based on a limited set of data, but the biographical core data and the name variants are there. You can search on periods and locations as well as on qualifications like 'printmaker, painter, draftsman... Also the Unique ID and the link to the full record on Explore are provided.

Authority file – Wikipedia – wikidata authority control

These days the world of terminologies is changing rapidly with multiple datasets becoming available online and often being connected. Combining the information from these sources would be beneficial for all. You see something like this happening on Wikipedia and the Wikidata elements that are becoming more important. Here the data is not yet connected as such but the RKD data is used as an authority file. Together with several other databases, like Worldcat, VIAF, LCSH and such. When these sets are made Open Data (if they aren't already) then the combining of information could really take off.

Iconclass

Iconclass is a classification system designed for art and iconography. It is a widely accepted scientific tool for the description and retrieval of subjects represented in images (works of art, book illustrations, reproductions, photographs, etc.) and is used by museums and art institutions around the world.

Iconclass is a **subject-specific classification system**. It is a **hierarchically** ordered collection of definitions of objects, people, events and abstract ideas that serve as the subject of an image.

And as a research tool, Iconclass is also used to identify the significance of entire scenes or individual elements represented within an image. This is done with Alfa-numerical codes with text correlates.

Alphabetical Index: 14,000 keywords used for locating the notation and its textual correlate needed to describe and/or index an image.

Bibliography: 40,000 references to books and articles of iconographical interest (not yet online).

Multilingual (English, German, French, Italian, Finnish and Portuguese, a small part even in Chinese).

Spanish translation is being worked on by 'La Dirección de Bibliotecas, Archivos y Museos (Dibam) El patrimonio de Chile, by Lina Nagel and Marisol Richter. Dutch is being worked on by RKD at the moment, but not structurally. If you're interested in joining the work please let me, or them know.

Needles to to say with some many languages it's Internationally used. Among others: the Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg (the German Counterpart of RKD), Centre d'Études Supérieures de la Renaissance (CESR Tours), Glasgow University and also the Rijksmuseum and Royal Dutch Library.

Main divisions of iconclass

The unique elements of the Iconclass system are its alphanumeric classification codes, called notations. Notations always begin with one of the digits 0 - 9, corresponding with the ten main divisions of Iconclass.

The main divisions of the Iconclass system are represented by digits 0 to 9. Of these ten 'main divisions', the numbers 1 to 5 are 'general' topics, designed to comprise all the principal aspects of what can be represented. Divisions 6 through 9 accommodate 'special' topics, coherent subject matter of a narrative nature, with an emphasis on the Bible (7) and Classical Mythology (9). A tenth division, represented by the number 0, was added in 1996 at the request of Iconclass users, to accommodate abstract art.

The text explaining the meaning of an Iconclass notation is the notation's 'textual correlate'. Several notations can be used to describe complex subjects.

Iconclass browser

The multilingual Iconclass Browser has an educational and a practical goal, which is to show how Iconclass works. At the same time, it serves as a search tool that allows an indexer to find the concepts to tag an image. Needless to say, it can also be used to establish the correct meaning of a notation.

You can search on keywords and text correlates to find the notation you are looking for. You can also follow the hierarchal path to come to the topic you need. But this might not show you every option there is. The hierarchy is functional as a thesaurus, with Every notation part of a hierarchical structure and, as such, incorporates all of its iconographical broader terms.

For instance 71D12 Joseph sold into slavery and taken to Egypt. The notations 7 (bible), 71 (old testament), 71D (genesis, the story of Joseph) 71D1 (story of Joseph part 1) are all integrated in this notation. You don't have to use the other notations if you use the Iconclass system.

In the past there where images shown from several collections with a notation, this was disabled unfortunately but we want to bring it back in a new version of the browser we hope to make next year.

Example images from the Bible: 71D12 Joseph sold into slavery and taken to Egypt

Some images that where indexed with this notation.



Examples subjects

The notations not only represent stories but also to themes, motifs, stories, situations, activities, gestures, events, ideas, messages, miracles. For instance: Also contains codes for gestures, positions, state of mind and feeling.

The AAT

The AAT is a structured vocabulary containing terms and other information about concepts. The Art & Architecture Thesaurus is a reference of over 250,000 generic terms attached to 36 000 concepts.

Terms in AAT may be used to describe art, architecture, decorative arts, material culture, and archival materials. The target audience includes museums, libraries, visual resource collections, archives, conservation projects, cataloging projects, and bibliographic projects.



Here we see the concept of 'chapels', with its attached terms in several languages.

Terms for any concept may include the plural form of the term, singular form, natural order, inverted order, spelling variants, scientific and common forms, various forms of speech, and synonyms that have various etymological roots. Among these terms, one is flagged as the preferred term, or *descriptor*. There may be multiple descriptors reflecting usage in multiple languages.

The other Getty Vocabularies

The AAT is part of the Getty Vocabulary Program which consists of the following products.... TGN, ULAN, CONA.

The four Getty vocabularies are intended to provide terminology and other information about the objects, artists, concepts, and places important to various disciplines that specialize in art, architecture, and material culture.

TGN is a structured vocabulary currently containing around 2 million names and other information about places. Names for a place may include names in the vernacular language, English, other languages, historical names, names and in natural order and inverted order.

The ULAN is a structured vocabulary currently containing around some 640 000 names and other information about artists. Names in ULAN may include given names, pseudonyms, variant spellings, names in multiple languages, and names that have changed over time (e.g., married names).

CONA contains titles and other information about architecture and movable works, which include paintings, sculpture, and a wide range of other objects. In CONA the other Getty vocabularies are systematically linked in the records for the unique items..

Catalogers and indexers who use the vocabularies typically access them by using them as implemented in a collection management system (either purchased off-the-shelf through a vendor or custom-built for their local requirements) or by copying-and-pasting from the online databases on the Getty Web site.

The data is also available via Web services, where it is updated every two weeks. The licensed files include no user interface. As of February 2014, the AAT is available as Linked Open Data (LOD).

What is the AAT?

Work on the AAT began in the late 1970s from the perspective of art libraries and art journal indexing services. Soon catalogers of museum objects and visual resource collections also expressed a need for similar controlled vocabulary to encourage consistency in cataloging and more efficient retrieval of information.

From its inception, the AAT was intended to satisfy a variety of user groups. The initial core of the AAT was gathered from terminology already being used in authority lists and the literature of art and architectural history; this terminology was approved and supplemented by a scholarly advisory team comprising people from the whole heritage spectrum. Reaching consensus among these various groups required long discussions and difficult negotiations, with the goal of creating a resource applicable to their diverse interests and requirements that would allow cross-collection retrieval. The AAT grows and changes via contributions from the user community and editorial work of the Getty Vocabulary Program. They do this using well-established editorial rules.

First publication on paper in 1990.

The AAT contains *generic terms*; it contains no iconographic subjects and no proper names. That is, each concept is a case of many (a generic thing), not a case of one (a specific thing). For example, the generic term *cathedral* is in the AAT, but the specific proper name *Chartres Cathedral* is out of scope for the AAT

There are about 100 languages in the AAT, but There are only three (almost) full translations of the AAT, From English, into Dutch and Spanish. Work is being done on a Chinese and German translation. I understand that Portuguese and Polish are starting up. A French translation is a long excising wish, and might be taken on in Belgium in the near future.

AAT is developed in line with ISO 25964 and NiSO norms. The AAT is truly a standard thesaurus what that matters.

CONA record, Eiffel tower

A critical feature that makes the vocabularies useful as authorities is that each vocabulary record is identified by an unique, persistent identifier. In this CONA record we see the concepts being linked to the other vocabularies.

Structure of AAT

Within its given scope of art, architecture, and material culture, the AAT is organized for general use; it is not organized for one particular use or according to any specific discipline.

For example, a user asked, why aren't communion cups and chalices narrower terms to church plate? This happens because **church plate** is a collective term for many different types and forms of objects used for ecclesiastical purposes, thus it is placed in the Object Genres hierarchy. Individual examples of church plate, such as communion cups and chalices or candlesticks, for instance, are found in various other places in the AAT, because the AAT's organization stresses function and form over the context in which an object is used. For this reason, communion cups and chalices are placed in Containers, and candlesticks in Furnishings, under lighting devices.

However, the AAT can link them to 'liturgical vessels' as an alternate parent using poly-hierarchical relationships

Definition of Thesaurus

Lets get back to the basic for a minute: what is a thesaurus?

Modern: a semantic network for retrieval and connecting (collection) data. Thesauri may be monolingual or multilingual

ISO 25964: "controlled and structured vocabulary in which (unique) concepts are represented by terms, that are organized in hierarchic, equivalent and associative relationships that are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms."

Relations: The defining characteristic of a thesaurus, that which distinguishes it from a flat list of terms such as a glossary, is the network of relationships among its terms and concepts. These relationships are semantic relationships, based on logical connections among the concepts, activities, and objects represented by the terms. Thesaurus construction standards identify three kinds of relationships, all of which are included in the AAT:

The Equivalence Relationship; flying buttresses

Multiple terms may refer to the same concept. The relationship between terms that represent the same concept is the equivalence relationship. Among terms that refer to a single concept, one is chosen as the *preferred term* or *descriptor*. This is the term that is most often used to refer to the concept in scholarly or professional literature.

All equivalence relationships in Getty Vocabularies are true synonyms = exact equivalents. Partial synonyms are related terms. In the AAT the terms include Descriptors, Alternate Descriptors, and Use For Terms.

Hierarchy for 'hacienda'

The hierarchical relationship links concepts to broader and narrower contexts. In the AAT, hierarchical relationships link generic classes of objects, actions, or concepts to their members or species.

Here we see the hierarchy from the facet all the way down to the concept of 'Haciendas'. It's a Spanish concept by origin, but because it is known in other languages also it needs to be disambiguated for each language

Hierarchy leading to 'hacienda'

Note: Large Spanish landed estates common to 17th century Seville used as farms, or ranches. Also applies to Spanish-American estates and the main building of such a complex.

Hierarchical Position:

Sources and Contributors:

estancias..... [VP]
 Avery Index (1963-)
 CD/MARC Subjects: LCSH (1988-)
 Webster's Third New International Dictionary (1961)

fazendas..... [VP]
 Avery Index (1963-)
 CD/MARC Subjects: LCSH (1988-)
 Webster's Third New International Dictionary (1961)

haciendas..... [VP]
 Avery Index (1963-)
 CD/MARC Subjects: LCSH (1988-)
 Webster's Third New International Dictionary (1961)

hacienda..... [CDBP-DIBAM, GCI Preferred, VP Preferred]
 AATP (database) (2002-): 124470 checked 26 January 2012
 Avery Index (1963-)
 CD/MARC Subjects: LCSH (1988-)
 Corbis, Plural del termino en singular
 Harris, Dictionary of Architecture and Construction (1975)
 Sturgis, Dictionary of Architecture and Building (1901)
 Webster's Third New International Dictionary (1961)
 hacienda..... [RID, AAT-Head]
 hacienda..... [RID, AAT-Head Preferred]
 hacienda..... [RID, AAT-Head Preferred]
 AAT-Head (1994-)

Subject: [CDBP-DIBAM, GCI, RID, AAT-Head, VP]
 AATP database (2002-): 124470 checked 26 January 2012
 AATP database (2002-): 124470 checked 26 January 2012

The sources and contributors are always add to the concept details. In this case we see DIBAM as contributor for the Spanish equivalent and the sources that were used.

Associative Relationship

Associative relationships or related terms are used primarily for retrieval, only when the concepts are *directly* and *uniquely* related, but where hierarchical relationships are inappropriate. If a thesaurus is bound together by too many associative relationships between entities that are only loosely or indirectly related, the value of the relationships in retrieval is lost.

For example, linking *brushes* to the hundreds of activities that could involve a brush should be avoided. However, linking *watercolor brushes* to *watercolor painting (technique)* would be a direct, unique, and appropriate Associative Relationship.

Values for relations.

Values are chosen from a controlled list comprising a code and phrase. Each code-plus-phrase is linked to another code, which is the reciprocal relationship.

The 7 facets

Facets constitute the major subdivisions of the AAT hierarchical structure. A facet contains a homogeneous class of concepts, the members or *children* of which share characteristics that distinguish them from members of other classes.

For example, *marble* refers to a substance used in the creation of art and architecture, and it is found in the *Materials facet*.

Impressionist denotes a visually distinctive style of art, and it is found in the *Styles and Periods facet*.

Dutch translation: website

Here we see the dutch browser for the AAT. All accepted concepts and the candidates are published here. The users can make comments or suggest another term if they want to. These comments are also shown to make clear that the vocabulary can be edited and improved with the help of the public.

The AAT-Ned or Dutch AAT is a translation of the Getty AAT and in an way extended version of it. Extended in the sense that AAT-Ned also hosts several thousand Dutch candidate concepts that are used in the Dutch heritage field. We are working together with an editorial board consisting of Flemish and Dutch professionals to get these terms validated and translated to have them accepted by the Getty.

Dutch translation organization

RKD handles translation of Getty AAT

We have translated 36 000 of the now 43 000 concepts in the Getty AAT.

There is a difference in content between the Dutch and the American AAT, in number of concepts but also scope notes and hierarchy can differ.

Reason: work continued on both sides and we don't have a regular way of exchanging data yet (perhaps soon).

We strive to make the difference as small as possible. Soon we'll work in an updated version.

New concepts via the editorial group and then send to Getty Vocab.

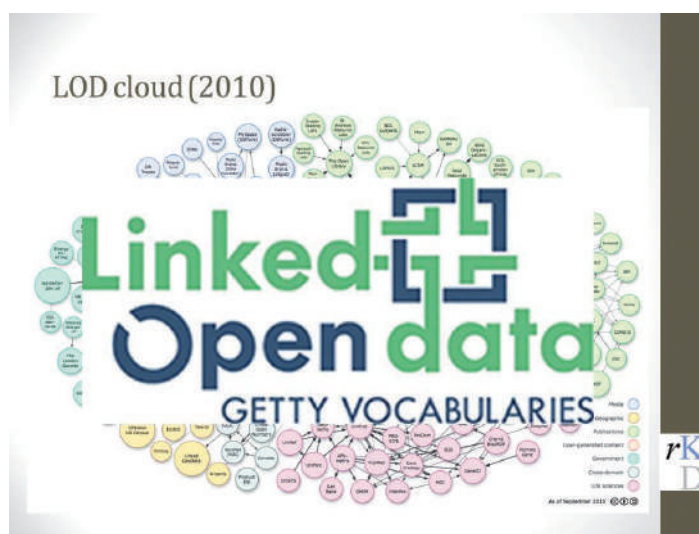
Slow process... But recently the 600th concept was added!

Spanish translation

I would like to mention the Spanish translation is being maintained by the same institute and people that work on the Iconclass translation, namely the 'La Dirección de Bibliotecas, Archivos y Museos (Dibam) El patrimonio de Chile, by Lina Nagel and Marisol Richter. They are also almost completely synchronized with the English AAT. I'm sure they are open for any input from the Spanish institutes to improve their translations.

LOD cloud

Here we have to all too familiar image of the LOD-cloud (2011). I guess nowadays it not even possible anymore to show a readable presentation of the cloud like this.



Roughly speaking, yesterday your digital resources were simple and blindly connected (on the web), today their connections with the network can have explicit meaning. The hyperlink is becoming semantic. Because the Semantic web is a world of connected pieces of knowledge rather than a network of pieces of information.

Getty web service

Data structures like JSON and OWL are available

The base URI is <<http://vocab.getty.edu/>>

SPARQL-endpoint (SPARQL Protocol and RDF Query Language)

Standard presentation is SKOS or SKOS-XL Simple Knowledge Organization System. But Several other xml

Mappings and ontology based on RDF/XML and Turtle

Published versions of lookup lists e.g., languages, roles, nationalities, place types, and bibliographic sources).

Europeana and AAT (and Iconclass)

Europeana is using the AAT and Iconclass in its datamodel. So when you send in data with links to the concepts, Europeana can switch between languages and might even be able to provide an end-user thesaurus in the future. Here we see the example of records from Erfgoeplus.be, a Belgian initiative and local aggregator for heritage. The example records is about 'stereographs' or 'estereogramas' (hope that is correct).

Discussion topic

When I started working on the Dutch AAT, I wondered why so many terminologies existed in the heritage field and why they kept being made? Isn't it more practical to use one standard terminology and work on expanding the one already made?

Of course this is no new idea, and it has been tried to promote in the Netherlands. But before the internet, digitized collections and linked data this was not very successful.

One reason why it didn't work: the 'not invented here syndrome', people want their own terminology to reflect their thinking and knowledge about a topic. Even when they don't really know how to construct a well organized thesaurus. That's why local terminologies that I have seen usually do not have a good thesaurus structure.

Reason two: Apparently not one terminology is perfect, or 'complete' enough to be used by another professional or institute. There will always be reasons to dismiss the ready-made terminology. Like 'the term we use for this type of object is not there'. Or: 'there are mistakes in this terminology'.

The resulting new terminology usually 'fixes' some issues for the users, but it will have its own limitations and errors just like the discarded one.

So, we keep on working on new terminologies that suit our needs. And we keep on inventing the wheel and quite often eventually getting stuck. To me, It's a shame we don't have a good way to bring the knowledge of terminologies together. Improve existing structures, expand them structurally with concepts from different knowledge domains, so that everyone can benefit. I know this centralized idea has its drawbacks, and it might not be possible to achieve.

Now, considering connecting knowledge, there might be hope on the horizon with the promise of LOD. It provides a method for connecting concepts between knowledge systems

through SKOS labels (CloseMatch and ExactMatch). The only problem is that this won't happen automatically. We, the domain experts and data experts, have to make the connections ourselves.

It's an impossible amount of work to connect everything to everything manually, I would say. But apparently there is already a lot to be gained if parts of the concepts between datasets are mapped. We have to start somewhere. Another solution could be, that we use a 'master thesaurus' to use as the backbone for our terminologies, one that is maintained and developed by a trusted and structurally funded party and keeps its content in line with thesaurus construction guidelines. This idea is slowly growing on people. But there are still a lot of hurdles to be taken.

In the Netherlands there is a movement that wants to bring terminologies together in a joint SKOS based environment where users are free to make these connections between terminologies. It's certainly an interesting thought to be able to search throughout all the connected terminologies at once and be able to make use of all the content and making semantic connections. If this approach will be a successful one still remains to be seen.

Information

RKDexplore (and RKDartists): <www.rkd.nl>

Website AAT-Ned: <www.aat-ned-nl>

Iconclass: <www.iconclass.nl>

Getty Vocabularies: <<http://www.getty.edu/research/tools/vocabularies/index.html>>

Spanish AAT: <www.aatespanol.cl>

Normas y estándares para la elaboración de tesauros de patrimonio cultural

Javier García Marco

Universidad de Zaragoza

1. Introducción

El objetivo del presente capítulo es analizar los recientes proyectos internacionales de reforma y ampliación de las normas sobre tesauros para su adaptación a los nuevos contextos de integración e interoperabilidad en el entorno digital; así como sus implicaciones para la elaboración de tesauros en el ámbito del control, la conservación, la protección y la difusión del patrimonio cultural.

El resultado de estos esfuerzos ha sido la nueva norma ISO 25964 sobre *Los tesauros y su interoperabilidad con otros vocabularios*¹, que consta a su vez de dos partes: la primera sobre *tesauros para la recuperación de información*, publicada en 2011; y la segunda sobre *La interoperabilidad con otros vocabularios*, publicada en 2013 (ISO, 2011, 2013).

En este capítulo se comenzará contextualizando la nueva norma en la historia de los tesauros, y, de forma más general, en la de la organización del conocimiento, analizando su papel en la reconstrucción de la conexión entre indización alfabética y sistemática. Gracias a la limpieza, solidez y simplicidad con la que se consiguió este objetivo, los tesauros se han asegurado durante décadas un papel fundamental en los esfuerzos por organizar el conocimiento registrado y por hacerlo disponible para su aprovechamiento en sistemas de recuperación de información. Además, han influido de forma decisiva en otros lenguajes documentales, como los sistemas de

¹ Los estándares ISO pueden adquirirse a través de su web: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53657> y <http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=01&ics2=140&ics3=20&csnumber=53658>. Asimismo se pueden conseguir a través de la agencia nacional en España, AENOR, que distribuye también la primera parte en español, en <<http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0053960#.Vkoq-4RnlZ8>>. Igualmente, se pueden consultar en varias bibliotecas universitarias españolas y algunas públicas, para lo que se recomienda la consulta del catálogo de REBIUN y el catálogo colectivo del Ministerio de Educación, Cultura y Deporte. Finalmente, el modelo de datos y el esquema XML están disponibles de forma abierta en el sitio web de NISO, en la dirección <<http://www.niso.org/schemas/iso25964/schema-intro/>>.

encabezamientos de materia, las listas de control de autoridades y las propias clasificaciones bibliográficas y archivísticas.

En segundo lugar, se examinan los determinantes que exigían una revisión de las normas existentes. Básicamente, las nuevas normas buscan abordar la revolución provocada por la generalización de Internet y lo que esto significa: la interconexión o interoperabilidad de diferentes fuentes de información en formato digital, de modo que puedan tener acceso de forma integrada por los usuarios, y que estos consigan una experiencia de uso similar a la de los portales de Internet a través de búsquedas centralizadas y simplificadas; pero, en el caso de las fuentes de información documentales, sin perjuicio de sus ventajas, ligadas a un control preciso y exhaustivo de la información que contienen. En particular, esta necesidad de adaptación de los tesauros para su interoperabilidad en Internet ha encontrado un entorno amigable y fértil en el proyecto de la Web Semántica (Berners-Lee; Hendler y Lassila, 2001).

En tercer lugar, se desglosa detalladamente el avance hasta la ISO 25964 a través de los proyectos de norma estadounidense y británico, así como las nuevas aportaciones realizadas en la norma internacional de tesauros. A continuación, se describen las dos partes de la norma, la primera dedicada a los tesauros propiamente dichos y la segunda, a su interoperabilidad. Finalmente, se explica la situación en España, donde se ha constituido un grupo de trabajo dentro del Comité 50 de la Asociación Española de Normalización y Certificación (AENOR), el cual ha realizado ya la traducción y adaptación al español de la primera parte de la norma, y se encuentra en el proceso de culminar la segunda parte.

La historia de los tesauros y su evolución hasta su forma actual ha sido desarrollada por un amplio conjunto de autores (entre otros, Foskett, 1982; Lancaster, 1986; Gilchrist, 1991; Dahlberg, 1991, 1992, 1995; Currás, 1991, 1992, 1998; Williamson, 1996; Gilchrist, 2003; Aitchison and Dextre Clarke, 2004; Dextre Clarke, 2008), en particular el camino hasta el nuevo estándar ISO 25964, resumido recientemente por Dextre Clarke and Zeng (2012).

2. Los paradigmas sistemático y alfabético en la organización y recuperación del conocimiento

La clasificación es una función cognitiva que está en la raíz del pensamiento humano (García y Esteban, 1993). La información que le llega al ser humano es desbordante para las limitaciones de su sistema cognitivo. Por ello, el ser humano selecciona de la realidad los aspectos que son accesibles para su sistema cognitivo y relevantes para su acción. Precisamente la clasificación es la estrategia que utiliza la mente para optimizar su capacidad de procesamiento, que en el caso de la memoria de trabajo ronda los siete conceptos (Miller, 1956). Parte del proceso de clasificación está regulado por mecanismos innatos, que producen una categorización «natural». Por encima de esta capa de conceptos naturales se desarrollan conceptos más generales (supraordenados) y más específicos (subordinados) (Lloyd y Rosch, 1978), que dependen del aprendizaje cultural.

Otro aspecto clave del ser humano es la capacidad de recordar y de compartir esa memoria con otros miembros de los grupos a los que pertenece o con los que se relaciona. Por ello, una parte notable de la evolución cultural ha sido el desarrollo de técnicas para mejorar el funcionamiento de la memoria, necesarias para incrementar la cantidad de conocimientos que pueden ser recordados por un grupo social. Un hito fundamental en esta evolución es la invención de memorias externas, los documentos (Pinto, García y Agustín, 2002, cap. 4).

El desarrollo de la representación del conocimiento sobre superficies externas se apoya también en procesos de clasificación. Para ganar complejidad sin perjuicio de su accesibilidad, los documentos se estructuran en partes y se relacionan unos con otros. Conforme el número de documentos prolifera, se hace necesario también clasificar y organizar estos de una manera sistemática, y surgen los sistemas de organización del conocimiento.

Los sistemas de organización más antiguos son los temáticos y los cronológicos (registros). Sin embargo, conforme los conocimientos proliferan, los árboles de conceptos que surgen de los

procesos de clasificación avanzados se vuelven cada vez más inmanejables; y, al extenderse los registros cronológicos, estos también se vuelven difíciles de manejar para búsquedas que no se basen en claves temporales. Por suerte, en el marco de un cuerpo cultural consolidado y estable, existen clasificaciones compartidas que permiten un acceso estable al conocimiento. Sin embargo, en las épocas de gran expansión del conocimiento –como ocurrió en época helenística, tras la invención de la imprenta, durante la revolución científica del siglo XIX, durante las guerras mundiales del siglo XX o en la expansión de Internet–, resulta difícil actualizar esas estructuras conceptuales, e incluso puede ocurrir que sus mismas bases sean puestas en cuestión.

Precisamente en época helenística se inventó una tecnología capaz de ofrecer acceso a la información sorteando los requisitos de la organización conceptual sistemática. Se trata de la aplicación del orden alfabético, inventado con fines didácticos para el aprendizaje de la escritura fenicia, a la ordenación de conceptos por sus términos. De esta manera, si se conoce el término que expresa un concepto, se puede rescatar la información de interés porque es posible predecir su lugar en cualquier lista ordenada. La forma más fácil de aplicar este principio es a los autores o los títulos. La utilización de la ordenación alfabética se documenta por primera vez en la Biblioteca de Alejandría (Daly, 1967). Las ventajas de la ordenación alfabética en el campo de la eficiencia son enormes: mientras que la clasificación u ordenación sistemática requiere un conocimiento profundo del campo del saber en cuestión, la ordenación y recuperación alfabética se puede delegar en cualquiera que conozca el sistema, en la práctica por cualquier persona alfabetizada. Eventualmente, como sucede en la actualidad, la ordenación alfabética se puede delegar en máquinas, automatizándola totalmente y ampliándola con otras herramientas de recuperación.

No obstante, la predictibilidad que ofrece el sistema alfabético en la recuperación se limita al propio orden alfabético. No permite recuperar por otras dimensiones conceptuales, en particular a través de las relaciones con otros conceptos, por ejemplo, buscando términos más específicos, genéricos o relacionados. Es decir, solo permite recuperar términos que comparten la misma raíz, limitando extraordinariamente las posibilidades de navegación por la información. Sin embargo, hemos visto que la mente humana, cuando encuentra un problema, utiliza la especificación y agrupación de conceptos como estrategia fundamental en sus procesos de inferencia, y, por lo tanto, también en los propios procesos de búsqueda documental.

Por esa razón, ambos sistemas son complementarios y necesarios, y en diferentes momentos de la historia de los sistemas de organización del conocimiento se ha buscado la interconexión entre ambos. La síntesis más conocida anterior a los tesauros es la que consiguió el bibliotecario norteamericano Charles Ammi Cutter en el último tercio del siglo XIX al interconectar el diccionario alfabético con las clasificaciones bibliográficas, en su caso la clasificación expansiva que había desarrollado.

3. El reto de la explosión de la información científico-técnica y el origen de los tesauros

Sin embargo, esta primera síntesis entre los dos monumentos de la biblioteconomía del siglo XIX –las listas de encabezamientos y las clasificaciones bibliográficas– colapsó ante el enorme crecimiento de las ciencias que se evidencia ya a finales del siglo XIX, pero que se potencia durante la primera mitad del siglo XX, en particular durante los esfuerzos de investigación aplicada que precedieron, acompañaron y siguieron a las dos guerras mundiales. En este contexto de explosión del saber, la era del libro dejó paso de forma decidida a la del artículo científico, caracterizado por su mayor especificidad e interdisciplinariedad.

Como consecuencia, los científicos naturales y los tecnólogos manifestaron su frustración con la rigidez y falta de precisión de las clasificaciones existentes, y aplicaron las nuevas tecnologías de ordenación automática y de cálculo basado en la lógica de Boole, como las máquinas de IBM, para solventar sus problemas de organización y recuperación del saber. La idea era marcar

los documentos con cuantos descriptores fuera necesario, y dejar su coordinación para el momento de la búsqueda, utilizando las capacidades de cálculo booleano de las nuevas máquinas ordenadoras. Por ello, se hablaba de postcoordinación (de los términos en la búsqueda), frente a la precoordinación de los términos que ocurría en el momento de la indización en el caso de las clasificaciones o encabezamientos de materia.

Entre las diversas propuestas que se hicieron por entonces, destacan como antecedente directo de los tesauros los descriptores de Moers (1950, 1951). Este autor propuso combinar una matriz de análisis de contenido propia para cada disciplina o campo especializado de documentación que guiara la selección de conceptos de interés, la codificación no precoordinada de los descriptores en una tarjeta perforada para cada artículo, y el ajuste de las búsquedas a las necesidades mediante un ordenador basado en el álgebra de Boole. Los descriptores podían constar de tantas palabras como fuera necesario para codificar conceptos lo suficientemente específicos, y se podían asignar a cada artículo tantos descriptores como se considerara adecuado para representar todas las necesidades de recuperación previstas.

Entre los nuevos sistemas basados en la ordenación alfabética que se propusieron, el sistema que se apartaba de forma más radical de los anteriores fue, posiblemente, el de los unitérminos de Taube (1953). El sistema de unitérminos se basaba en extraer términos relevantes y muy específicos de los textos, y marcar las tarjetas perforadas de cada artículo. Los términos debían ser simples, por lo que no había ninguna precoordinación, de ahí «uni-términos». Aunque el sistema era adecuado para ámbitos muy específicos con un vocabulario muy técnico, se mostró ineficaz para ámbitos más amplios, interdisciplinarios o con una base léxica muy compartida con el lenguaje cotidiano u otras especialidades, por lo que fue muy criticado fuera de los ámbitos muy especializados en los que se aplicó con éxito (Moers, 2003: 818; Aitchinson & Clarke, 2004: 7). Sin embargo, el enfoque de los unitérminos abrió el camino a los actuales sistemas de recuperación de la información en texto libre, que actúan bajo los mismos principios pero sin selección *a priori* de los términos y con un conjunto de órdenes de búsqueda más amplio.

Por otra parte, los esfuerzos desde la tradición bibliotecaria no se habían interrumpido. Mientras los documentalistas optaban por la postcoordinación, algunos expertos bibliotecarios críticos con las clasificaciones bibliográficas habían estado trabajando en sistemas alternativos para hacerlas más hospitalarias al crecimiento de las disciplinas científicas y la propia producción científica. Los más conocidos de esos pioneros son dos bibliotecarios universitarios, muy expuestos al crecimiento de la producción bibliográfica: el británico Henry E. Bliss (1929, 1932, 1935) y el indio Shiyali Ramamrita Ranganathan (1933, 1937), cuyas formaciones les proporcionaban una perspectiva fuera de lo común y que realizaron aproximaciones muy originales a la clasificación bibliográfica. Ambos llegaron a una aproximación semejante a la de los postcoordinacionistas, pues consideraban que cada documento podía recibir clasificaciones alternativas desde distintos puntos de vista. Estas perspectivas diferentes fueron denominadas por Ranganathan facetas, y se aproximan a los puntos de vista de Moers, pero con un sentido universal en vez de disciplinario. Sin embargo, frente a los postcoordinacionistas, tanto Bliss como Ranganathan eran clasificacionistas, esto es, creían que los procesos de indización debían generar clases con un orden de cita estable y predecible, de manera que se pudieran generar listados e índices sistemáticos.

En este contexto, los tesauros surgieron a partir de las listas de descriptores por la necesidad de proporcionar el contexto de los términos, tanto para facilitar la expansión o restricción de las búsquedas como para proporcionar contexto a los indizadores y recuperadores cuando no se conoce con precisión el potencial término de búsqueda. El modelo utilizado fue el de los diccionarios conceptuales o tesauros utilizados para mejorar los procesos de escritura, lo que no debería resultar extraño, pues la búsqueda de información es también un uso productivo del lenguaje que se apoya en inferencias categoriales, entre otros recursos. En particular, se tomó como modelo el tesoro en lengua inglesa más conocido, el Roget's Thesaurus.

El primer autor que propone utilizar el modelo del tesoro para organizar los vocabularios estructurados para la recuperación de información fue H. P. Luhn (1957), en el marco de un pro-

yecto de investigación de IBM. En 1959 se construyó el primer tesoro para controlar términos en un sistema de recuperación para Dupont, y los primeros tesauros publicados aparecieron en el cambio de década: el *Thesaurus of ASTIA Descriptors* del Defense Documentation Center de los Estados Unidos (1960), y el *Chemical Engineering Thesaurus* del Institute of Chemical Engineers (1961). Estos primeros tesauros estaban ordenados alfabéticamente, con las relaciones especificadas detrás de cada término.

El modelo estándar de tesoro que conocemos hoy en día surgió en 1967 con el *Thesaurus of engineering and scientific terms; a list of engineering and related scientific terms and their relationships for use as a vocabulary reference in indexing and retrieving technical information* del Engineers Joint Council y el Department of Defense (1967). El año anterior, el Panel de Vocabulario de la Ingeniería del Engineers Joint Council (1967) había aprobado una recomendación específica de reglas y convenciones para el proyecto, que se publicó como apéndice I del tesoro. Estas reglas contienen todas las relaciones de equivalencia, jerárquicas y asociativas de los tesauros modernos y sus abreviaturas en inglés (USE, UF, BT, NT, RT), recomendaciones para la selección de los términos preferentes entre un conjunto de sinónimos y cuasisinónimos, y sobre la forma de los términos. Sin embargo, la perspectiva sistemática se limitaba a una presentación de las veintidós categorías principales con un solo nivel de desarrollo. En esta década de desarrollo de los tesauros hasta la consolidación de su modelo fue decisivo el trabajo de Eugene Wall (Krooks y Lancaster, 1993).

A partir de este trabajo, Derek Austin, el creador de PRECIS, y P. Dale escribieron las recomendaciones de la UNESCO en 1970 (UNESCO, 1973), en el marco de un trabajo conjunto con el comité de documentación de la ISO (TC 46). Austin promovió la incorporación del concepto de tesoro sistemático y el análisis de facetas en las recomendaciones sobre tesauros, integrando así tanto la tradición de los clasificacionistas clásicos como de los revisionistas, y proponiendo reglas más claras para el tratamiento de términos compuestos, que se derivaban en buena parte del propio análisis de facetas. Por entonces, el tesoro se estaba transformando ya en un lenguaje koiné, que incorporaba tanto los principios de los nuevos sistemas postcoordinados basados en la lógica booleana y la utilización de ordenadores, como la tradición de las grandes clasificaciones bibliográficas y los avances producidos en dicho campo durante la primera mitad del siglo XX. En 1974 apareció también la primera versión de la norma ISO 2788 sobre tesauros monolingües (ISO, 1974).

Conjuntamente con las recomendaciones de la UNESCO e ISO, comenzaron a surgir las primeras normas nacionales. En 1974 apareció también la ANSI Z39.19-1974 titulada *Thesaurus Structure, Construction and Use*, publicada por el American National Standards Institute (ANSI, 1974). Fue revisada por primera vez en 1980, y en 1993 cambió su título, tras una segunda revisión, a *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*, que fue a su vez revisada y confirmada en 1998, aunque con la recomendación de que se revisara lo antes posible para acomodarse a los cambios que Internet había producido en el ámbito de la información y la documentación.

Así pues, en el crisol de las organizaciones internacionales se producen avances decisivos para los tesauros. Se consolida la síntesis entre el enfoque postcoordinado y sistemático de los tesauros, y se realizan algunas aproximaciones al diseño de tesauros de alcance universal, intentando también el acercamiento entre la perspectiva orientada a dominios específicos, propia de la documentación, y la enfocada a la organización bibliográfica universal, más característica de bibliógrafos y bibliotecarios. En 1972 se publicó la primera edición del *Macrothesaurus* de la OCDE realizado por Jean Viet (Organisation..., 1972), adoptado luego por las Naciones Unidas; y en 1977 del *UNESCO Thesaurus*, de Jean Aitchison (1977), seguidos de varios otros durante finales de los años setenta y la década de los ochenta. En la mayoría de estos tesauros internacionales, la presentación sistemática se despliega ya de una forma madura, en consonancia con su enfoque más clasificacionista. Un papel destacado de liderazgo en esta síntesis la tuvo el British Classification Group, que conectó las ideas de Bliss y Ranganathan y el movimiento de la indización postcoordinada con notable éxito.

Fruto de este esfuerzo de aplicación de los tesauros a ámbitos de cooperación internacional fueron las recomendaciones para tesauros multilingües de la UNESCO en 1980, seguidas del estándar ISO 5964 en 1985. Franceses, ingleses, alemanes y españoles, entre otros países, también fueron adoptando las nuevas normas internacionales.

4. El impacto de la revolución digital: la competencia de los modelos alternativos y la crisis de los tesauros

Sin embargo, los tesauros –que durante los años sesenta y setenta se habían destacado como la tecnología líder en recuperación de información, y no habían dejado de adaptarse y crecer– comenzaron su declive precisamente en la época en que la consolidación de las normas internacionales parecía marcar un futuro prometedor, a mediados de los años ochenta.

La razón estriba en el triunfo de dos tecnologías competitivas que se habían estado desarrollando precisamente en esa época, y que descansaban el esfuerzo de búsqueda en sistemas altamente automatizados y en los propios usuarios. Por un lado, los sistemas de recuperación basados en índices inversos y cálculo booleano se comoditizaron, y recibieron un gran refuerzo del cálculo vectorial y de la estadística, contribuyendo a mejorar la relevancia de sus búsquedas y presentaciones mediante la cuantificación de la semántica latente en grandes conjuntos de información (modelos vectorial y probabilístico). Por otro lado, la navegación a través de referencias intra, inter y extratextuales se automatizó mediante el hipertexto y su gran aplicación en red, la World Wide Web. Eventualmente, ambas tendencias confluyeron a finales de los noventa en el nuevo entorno digital con el surgimiento de los grandes buscadores, un ámbito en el que ahora Google es hegemónico.

Por debajo de estos procesos, se situaba una corriente aún más general, la ‘gran’ globalización, que se retroalimentó con la propia revolución de Internet. No es casual que en noviembre de 1989 cayera el muro de Berlín, y ese mismo año Tim Berners-Lee propusiera la World Wide Web como medio transparente de publicación y recuperación de información. En los años siguientes, EE. UU. abrió Internet al resto del mundo, la World Wide Web despegó (1993) y se transformó en el medio de comunicación por excelencia de la nueva globalización, derribando fronteras y barreras entre países y culturas. Tras varios intentos previos de éxito más limitado por parte de otras empresas, Google creó a partir de 1996 un catálogo mundial para acceder a Internet, liderando la monetarización de la recuperación de información con valor comercial y político, que es otro de los signos del nuevo paradigma en el que vivimos.

En este ambiente los tesauros parecían para muchos reliquias de un pasado casi lejano (con ser tan reciente), y su valor solo se apreciaba plenamente en entornos muy especializados, como las grandes bases de información química, farmacológica, biomédica, económica, legal o relacionadas con la preservación del patrimonio cultural o natural. Se trata de ámbitos todos ellos en los que se requiere gran precisión y exhaustividad en las búsquedas, cuyos profesionales tienen un alto coste por hora, y en los que las consecuencias de pasar por alto información relevante pueden ser muy graves.

Sin embargo, toda crisis conlleva una oportunidad; y una gran crisis como la que estamos tratando, grandes oportunidades. En primer lugar, se planteaba la necesidad de conectar y unir todos esos recursos especializados con el resto de la Internet, esto es, de hacerlos interoperables en la red global. Efectivamente, la mayoría de los recursos más valiosos disponibles en Internet están organizados con herramientas muy diferentes, lo que dificulta su recuperación conjunta: la Clasificación Decimal Universal, la Clasificación Decimal de Dewey, la clasificación de Yahoo, los Encabezamientos de Materia de la Library of Congress y sus derivados en otras lenguas, las Medical Subject Headings, y diferentes tipos de tesauros, listas de encabezamientos, cuadros de clasificación de archivos, etc.

En segundo lugar, pronto quedó claro que, a pesar del énfasis del hipertexto en las relaciones asociativas, los sitios web –y por ende la propia Internet– necesitaban estructura, no solo

enlaces, y que esas estructuras eran muchas veces incompatibles, impidiendo su explotación conjunta.

En tercer lugar, los datos de diferentes fuentes se acumulaban en Internet gracias al gran esfuerzo de publicación y digitalización (movimiento de los documentos y datos abiertos), pero en gran parte seguía siendo necesario un costoso procesamiento manual para cruzarlos y compatibilizarlos.

Por fin, los propios creadores de la World Wide Web se dieron cuenta de que, conforme la red crecía, era más difícil separar el grano de la paja, y que era necesario dotar a los ordenadores de la red de una mayor comprensión del significado de la información para los humanos, de manera que pudieran auxiliarles en el desbordante universo de la información digital hasta alcanzar el nivel del dato, que era muchas veces el objeto de la necesidad de información. Por ello, en 2001 Tim Berners-Lee y sus colaboradores propusieron una nueva arquitectura de la web en la que los documentos y los datos pudieran ser procesables de forma autónoma por las máquinas para optimizar la utilización de la información disponible. Precisamente este proyecto, al que denominaron ‘web semántica’, proporcionó una infraestructura desde la que aprovechar este conjunto de oportunidades que se abrían a los tesauros. En concreto, los tesauros y otros sistemas de organización del conocimiento enseguida se mostraron útiles como materia prima para modelar la capa ontológica de la futura web semántica, que ocupa el estrato intermedio del modelo –la definición de las entidades y sus relaciones–, especialmente a través de formalizaciones claras como SKOS.

En definitiva, aunque Internet constituye efectivamente un increíble y potentísimo entorno de información global, sin embargo, está fraccionada en multitud de culturas étnicas y corporativas, lenguas y lenguajes técnicos, agentes y repositorios inconexos, lo que impide que se realice plenamente todo su potencial. Para convertir Internet en un entorno de recuperación y computación realmente integrado, se requiere un gran esfuerzo para asegurar la interoperabilidad de los diferentes recursos y agentes; una interoperabilidad que tendrá que ser horizontal o distribuida por la propia naturaleza de la red.

5. La adaptación de los tesauros al mundo digital

La respuesta en el campo de los tesauros ha sido realizar un esfuerzo internacional concertado para desarrollar una nueva propuesta normativa en el campo de los vocabularios estructurados para la recuperación de la información, capaz de fortalecer la interoperabilidad de los recursos de información en Internet. Dicho esfuerzo se enmarca en el objetivo más amplio de adaptar las herramientas documentales desarrolladas en los dos últimos siglos al entorno digital, y en particular los tesauros, por su capacidad de integrar las diferentes perspectivas que existen en la organización del conocimiento: conceptual frente a alfabética, universal frente a local, categorización frente a taxonomización, asignación frente a extracción, y perspectiva de experto frente a perspectiva de usuario.

Los nuevos estándares sobre tesauros se conciben, desde esta perspectiva, como una pieza más dentro del conjunto de estándares que han ido apareciendo para hacer la Internet posible, eficaz y eficiente (HTTP, REST), en particular dentro del marco creado por la Web Semántica (XML, RDF, OWL, SPARQL) y los esfuerzos más especializados que se enmarcan en ella, en concreto el campo específico de la información y la documentación (Z39.50, SRU, ISO 2709, MARC 21) y más concretamente en el propio campo de los tesauros y vocabularios estructurados (Z39.19, BS 8723, SKOS, ZThes...).

Aunque Internet ha hecho más acuciente el problema y lo ha convertido en una gran oportunidad, la interoperabilidad de vocabularios de recuperación no es un problema nuevo, sino que se enmarca dentro de la tradición de discusiones sobre la heterogeneidad y compatibilidad de las herramientas de organización del vocabulario. En particular, uno de los problemas más agudos dentro de este campo de trabajo fundamental es el de la compatibilidad entre

diferentes lenguas, esto es, el multilingüismo. El reto del multilingüismo se ha incrementado con redoblada fuerza en Internet, donde el problema del acceso a la información en diferentes lenguas y sistemas de recuperación se presenta de forma contigua e inmediata.

Por otra parte, la capacidad de Internet de integrar los dos principales paradigmas de la organización del conocimiento –el precoordinacionista y el postcoordinacionista– se adapta muy bien a la propia naturaleza de la búsqueda de la información en la World Wide Web, en la que la recuperación propiamente dicha –esto es la localización de información mediante ecuaciones de búsqueda, en general sencillas y asistidas a partir de cajas de texto– coexiste con la navegación en listas, esquemas, mapas y otros recursos semejantes (browsing, mapping).

6. Un repaso a los principales estándares actuales

La primera norma de tesauros en ser adaptada al nuevo entorno digital fue la americana Z39.19. Como se ha mencionado, ya en la revisión de 1998 se había constatado la necesidad de ponerla al día ante los cambios que se estaban produciendo en el marco de la información electrónica y de la World Wide Web. Atendiendo a esta necesidad, la National Information Standards Organization (NISO) organizó un taller de trabajo nacional el año siguiente denominado *Workshop on Electronic Thesauri*, que se celebró el 4 y 5 de noviembre de 1999, con el objetivo explícito de investigar la conveniencia y posibilidad de desarrollar un estándar para los tesauros electrónicos. Dicha reunión fue apoyada por la American Psychological Association (APA), la American Society of Indexers (ASI) y la Association for Library Collections and Technical Services (ALCTS) de la American Library Association (ALA).

A partir del consenso y las recomendaciones del taller sobre tesauros electrónicos, se desarrolló la cuarta edición de la norma Z39.19, cuyo título se modificó a *Guidelines for the construction, format and management of monolingual controlled vocabularies* (ANSI, 2006). Se trata de un extenso documento de 172 páginas que se inicia con un análisis preliminar de la necesidad del control de vocabulario; presenta los objetivos, conceptos, principios y estructura de los tesauros; y define las normas de tratamiento terminológico –selección, alcance y forma de los términos simples y complejos–, las relaciones, las técnicas de presentación, los aspectos relacionados con la interoperabilidad, y las cuestiones referidas a su construcción, prueba, mantenimiento y gestión.

La norma americana innovó en tres frentes principales: en primer lugar, pretendía abarcar todo tipo de recursos de información, tanto tradicionales como electrónicos; en segundo lugar, ampliaba su alcance a todos los «vocabularios controlados» además de los tesauros propiamente dichos, incluyendo concretamente «las listas de términos controlados, anillos y sinónimos, taxonomías y tesauros»; y, finalmente, abordaba por primera vez en una norma el problema de la interoperabilidad de vocabularios en el marco de Internet. Sin embargo, también quedó con algunas limitaciones importantes: trataba únicamente el caso de los vocabularios monolingües y dejaba de lado herramientas de vocabulario controlado muy importantes, cuya interoperabilidad con los tesauros podría resultar también clave para los fines pretendidos.

La reforma de las normas británicas sobre tesauros monolingües y multilingües –BS 5723 y BS 6732, idénticas a la ISO 2788 y ISO 5964, respectivamente– se puso en marcha algo después, con el objetivo de integrar las dos normas anteriores y abordar los aspectos de interoperabilidad, apoyándose en los conceptos desarrollados para el tratamiento del multilingüismo (Aitchison y Dextre Clarke, 2004; Gilchrist, 2007). El resultado, la norma BS 8723, ha sido un documento muy amplio que consta de cinco partes, que se fueron publicando entre 2005 y 2008. La primera parte (BSI, 2005) establece las definiciones y conceptos comunes a todos los tipos de vocabularios controlados para la recuperación de la información en sistemas de información. La segunda parte (BSI, 2005) trata de los tesauros propiamente dichos –siendo equivalente a la norma precedente BS 5723 (ISO 2788)–, pero añade además pautas sobre su uso y gestión digital. La parte tercera (BSI, 2007) analiza otros vocabularios estructurados, notablemente los esquemas de clasificación,

los tesauros, las listas de encabezamientos, las taxonomías y las ontologías. La cuarta parte (BSI, 2007) trata en general la interoperabilidad entre vocabularios, así como el mapeo entre ellos; y aborda la cuestión del multilingüismo como un caso especial de interoperabilidad, por lo que esta parte abarca los aspectos considerados en la anterior norma sobre tesauros multilingües BS 6723 (ISO 5964). Por fin, la quinta parte (BSI, 2008) trata los protocolos y formatos para el intercambio de datos sobre vocabularios controlados, mediante el uso del XML y otras tecnologías semánticas.

En su conjunto, la norma británica consigue varios avances notables sobre la estadounidense. Entre ellos se debe destacar la consideración de la interoperabilidad lingüística, la distinción clara entre términos y conceptos, el desarrollo de un modelo de datos, el análisis de facetas –que resulta central en la moderna teoría de la clasificación y los tesauros–, el tratamiento de los sistemas de clasificación como herramientas complementarias de los tesauros y puntos clave de una estrategia de interoperabilidad entre los lenguajes documentales en conjunto, y el abordaje de otros lenguajes como las listas de encabezamientos de materia, cada vez más tesaurizadas, y de algunos sistemas de organización y codificación del conocimiento especialmente utilizados en Internet, como las ontologías, además de las taxonomías y anillos de sinónimos ya tratados en la última edición de la ANSI Z39.19.

Dentro de los esfuerzos de revisión y actualización de las normas de tesauros, se debe resaltar también la labor del grupo de trabajo de la International Federation of Library Associations (IFLA), que se puso específicamente en marcha con el objeto de poner al día las recomendaciones para tesauros multilingües (IFLA, 2006), y que trabajó en estrecha coordinación con los otros dos grupos estudiados en sus respectivos proyectos.

Finalmente, por iniciativa del British Standard Institution y bajo el liderazgo de la coordinadora de su grupo de trabajo de vocabularios, Stella Dextre Clarke, la norma británica se ofreció a la International Organization for Standardization como base para desarrollar un nuevo estándar internacional, la futura ISO 25964, dedicada a los tesauros y su interoperabilidad con otros vocabularios que actualizara las normas ISO 2788:1974, 1986 e ISO 5964:1985.

El grupo de trabajo se puso en marcha en 2008, en el que participaron representantes de quince países con ayuda de una secretaría sostenida por la NISO. Una de las labores más importantes del grupo de trabajo ha sido simplificar y desarrollar la norma británica en dos únicas partes: la primera dedicada a los tesauros, *Parte 1: Tesauros para la recuperación de la información* (publicada en agosto de 2011), que trata tanto los tesauros monolingües como los multilingües; y la segunda dedicada a la interoperabilidad con otros sistemas de organización del conocimiento: *Parte 2: Interoperabilidad con otros vocabularios* (publicada en 2013), que avanza notablemente en la definición de los modelos de mapeo.

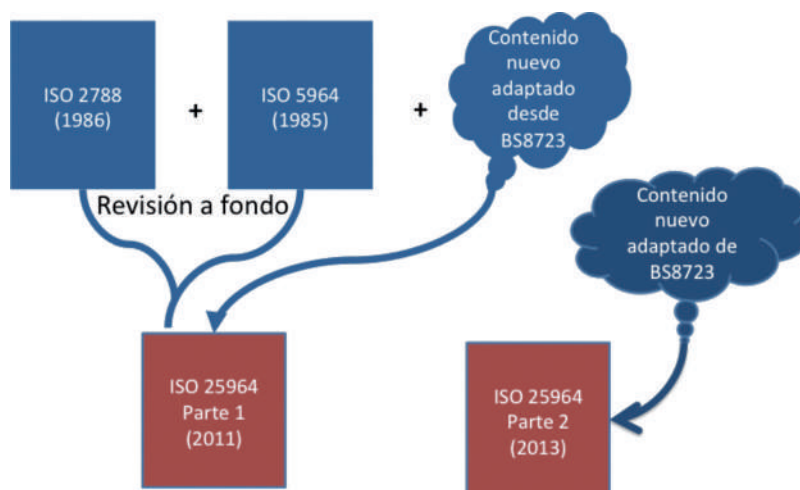


Figura 1. Genealogía de la norma ISO 25964.

7. La primera parte de ISO 25964

Los avances más importantes que se producen en la primera parte de la norma ISO 25964 se refieren, sobre todo, a tres aspectos: el multilingüismo y la multiculturalidad; la construcción, mantenimiento y difusión de los tesauros; y el aseguramiento de la interoperabilidad de los tesauros.

Un aspecto clave de la primera parte de la norma es que reunifica el tratamiento de los tesauros monolingües y multilingües en un solo código. En este sentido, supone la culminación del proceso iniciado en las dos normas anteriores, y la integración definitiva de la experiencia conseguida en los organismos internacionales por los expertos en organización del conocimiento a lo largo de varias décadas. Por otra parte, el contexto no podía ser más propicio: tanto la creciente globalización como la naturaleza internacional de la ciencia, la técnica y las humanidades, y el propio carácter multicultural de Internet exigían el pleno desarrollo del multilingüismo de forma natural.

Para facilitar la reconceptualización del tratamiento del multilingüismo, la nueva norma entroniza al concepto como unidad básica de los tesauros. Frente al énfasis en el término que se realizó durante los años setenta y ochenta, cuando diversas tradiciones en la organización temática de la información situaban el énfasis teórico en el control terminológico o el control del vocabulario, el cambio de siglo ha traído la recuperación de los conceptos como núcleo del trabajo de organización del conocimiento. Ciertamente, aunque la terminología es una herramienta fundamental, el centro de la organización del conocimiento son aquellos conceptos que son objeto del interés de una comunidad para recuperar información. Y los conceptos van cambiando en su expresión a lo largo del tiempo y, sobre todo, se expresan de formas diversas entre los distintos idiomas, para lo que la terminología resulta una herramienta limitada.

Además, la nueva norma ha mejorado su tratamiento del multilingüismo de una manera sustancial, lo que se aprecia en la ampliación del inventario de etiquetas en diferentes idiomas –que ahora se incluyen en español, inglés, francés, alemán, danés, finés, noruego, sueco, chino y ruso–, en el incremento del número de ejemplos en diversas lenguas –todavía fundamentalmente en las principales europeas–, así como en las continuas referencias a los factores culturales y lingüísticos en numerosos apartados de la norma. El capítulo nueve dedicado a la relación de equivalencia interlingüística integra y simplifica la anterior norma de construcción de tesauros multilingües. En él, partiendo de los cuatro grados de equivalencia –exacta, inexacta (cuasiequivalencia), parcial (genérico-específica) y la no equivalencia–, se abordan los diferentes tipos de problemas –cuasisinónimos, homógrafos y la ausencia de términos equivalentes en una o más lenguas–, actualizando sobre todo los ejemplos, adoptando términos más actuales.

Otro gran avance de la nueva norma ha sido clarificar y desarrollar también la doctrina sobre la construcción, mantenimiento y uso de los tesauros. En primer lugar, se amplía el tratamiento de las facetas y se establecen claramente las fronteras entre la ordenación jerárquica y por facetas. Además, los ejemplos de presentaciones se desarrollan de forma bilingüe, en inglés y español. En el capítulo trece se abordan cuidadosamente las políticas previas, la formación y mantenimiento de los equipos de trabajo y los recursos necesarios, enfatizando el requisito de un compromiso a largo plazo por parte de directivos y expertos. También destaca la importancia que se concede al respeto de los derechos intelectuales en el caso muy frecuente de la reutilización de tesauros, y a la necesidad de depositar los tesauros en repositorios especializados –como la Subject Analysis Systems Collection de la Universidad de Toronto o el registro de terminologías de la Dublin Core Metadata Initiative– para facilitar su aprovechamiento, reconocimiento y preservación.

Como la difusión y explotación de los tesauros se realiza hoy en día tanto por agentes humanos como informáticos, la norma desarrolla especialmente estas cuestiones, que hacen que los tesauros requieran hoy en día un cuidado administrativo mucho mayor y un control absoluto de todos los procesos de actualización y distribución, tanto del tesoro completo como de partes del mismo, sean árboles, conceptos, términos o relaciones.

En el capítulo catorce se especifican cuidadosamente los requisitos mínimos que debe cumplir un software de gestión de tesauros, sin el cual no cabe hoy en día plantearse la creación y sobre todo el mantenimiento de un tesoro, especialmente si va a ser sostenido por una red amplia de personas y centros cooperantes para su explotación y mantenimiento en Internet, lo cual es cada vez más la realidad cotidiana.

Por último, la última parte de ISO 25964-1 se dedica a la interoperabilidad de los tesauros, uno de sus objetivos clave, como hemos visto. Precisamente, uno de los puntos fuertes que poseen los tesauros como instrumento de recuperación de la información y de organización del conocimiento es su capacidad de integrar códigos, lenguas, terminologías y sistemas diferentes. La norma se vuelca en que los tesauros puedan seguir disfrutando de esta ventaja competitiva en el nuevo entorno caracterizado por Internet y la creciente globalización.

En el capítulo quince, se ofrece el modelo de datos, que constituye, sin duda, una de sus aportaciones más importantes. Por una parte, se modelizan formalmente en UML (Unified Modelling Language) las clases, atributos y asociaciones de los tesauros, de manera que se puedan instanciar sin ambigüedad sistemas y aplicaciones a partir de él –aunque no se especifican los comportamientos y métodos de clase–. Además, la descripción formal del modelo facilita el intercambio de datos entre sistemas de forma automatizada, asegurando la interoperabilidad.

En el capítulo dieciséis se aborda la integración de los tesauros dentro de y con las aplicaciones que desarrollan funciones de recuperación de información. Se explicitan los requisitos para navegar por el tesoro, cuando este está disponible para el usuario, y para que sea utilizable durante el proceso de recuperación de los documentos.

El capítulo diecisiete presenta los formatos de intercambio de tesauros: los formatos MARC –en particular, los de intercambio de autoridades–, SKOS, Zthes –desarrollado inicialmente para Z39.50 y ampliado a SRU (Search/Retrieve via URL)– y DD 8723-5 –desarrollado para la norma británica–. Como ninguno de ellos se consideró adecuado para expresar todas las características de un tesoro de acuerdo con el modelo expresado en el capítulo quince, se desarrolló un formato XML *ad hoc*, que se describe en el Anexo B y está disponible en Internet en la dirección <http://www.niso.org/schemas/iso25964>.

Finalmente, el capítulo dieciocho aborda los protocolos de comunicación entre sistemas, otro de los aspectos fundamentales para asegurar la interoperabilidad. En particular, se revisan tres protocolos desarrollados específicamente para tesauros –el SWAD-E SKOS API, varios protocolos basados en SOAP y REST, y el desarrollado por la Alexandria Digital Library (ADL)– así como otros de propósito general –en particular, OASIS, SPARQL y Z39.50.

8. La segunda parte de ISO 25964: interoperabilidad con otros vocabularios

La segunda parte de la norma ISO 25964 aborda la interoperabilidad de los tesauros no solo en lo que se refiere a sentar sus bases durante el proceso de diseño, mantenimiento y difusión, sino con respecto a su integración en un entorno de sistemas abiertos con otros sistemas de organización del conocimiento para facilitar búsquedas y servicios «federados». Constituye un trabajo de gran envergadura que se apoya en el éxito conseguido en la normalización de los tipos de mapeos (Dextre, 2011, 2012). Además, la interoperabilidad se aborda en el contexto de la web semántica y los datos abiertos, gracias al estrecho trabajo conjunto que realizaron los equipos de desarrollo de la ISO 25964 y del estándar W3C SKOS, Simple Knowledge Organization Systems on the Web.

Sin entrar en la descripción de cada uno de los mapeos particulares, sí que es importante trazar la problemática general del mapeo entre vocabularios. La norma acepta tres modelos estructurales de mapeo: el de unidad estructural (en el que los conceptos y las estructuras relacionales son equivalentes), el de enlace directo (en el que todos los vocabularios se enlazan entre sí), y el de 'hub' (en el que los vocabularios se mapean en ambas direcciones con respecto a uno central). Se contempla, además, un mapeo selectivo, entre partes de los vocabularios. Como se puede ver

en la Figura 2, el mapeo transitivo está desaconsejado, pues requiere mucha carga computacional para aprovecharlo y produce errores en cadena, multiplicando las fuentes de ruido y silencio.

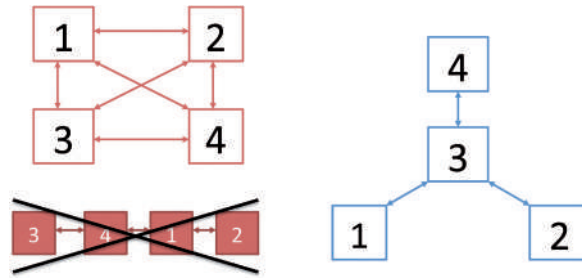


Figura 2. Tipos de mapeo previstos en ISO 25964-2.

Con respecto a las relaciones, se prevén tres tipos de mapeo básicos: de equivalencia, jerárquico y asociativo (Tabla 1). Opcionalmente, se pueden realizar también otros mapeos más específicos. En el caso del mapeo de equivalencia, se prevén dos grandes tipos: la equivalencia simple, entre pares de conceptos; y la equivalencia compuesta, en la que un concepto puede relacionarse con más de un concepto del otro vocabulario o viceversa. La equivalencia simple puede ser exacta, esto es, entre conceptos exactamente iguales en alcance semántico; o inexacta, es decir, aproximada.

Tabla 1. Tipos de mapeo previstos en ISO 25964-2.			
Tipos	Subtipos	Etiqueta en inglés	Ejemplo
Equivalencia	Simple		Ordenadores EQ Computadores
	• exacta	=EQ	Aubergines =EQ Egg-plants
	• inexacta	~EQ	Horticultura ~EQ Jardinería
	• compuesta		
	Intersectiva	+	Ejecutivas EQ Mujeres + Ejecutivos
	Acumulativa		Vías de agua interiores EQ Ríos Canales
Jerárquico	Hiperonímica	BM	Calles BM Vías
	Hiponímica	NM	Vías NM Calles
Asociativo		RM	e-Learning RM Educación a distancia

A su vez, la equivalencia compuesta puede resolverse mediante una intersección de conceptos (intersectiva), o mediante la unión de los campos semánticos (acumulativa). Se presenta un ejemplo en la Figura 3.

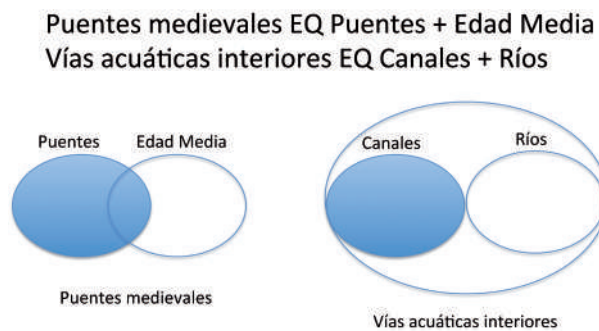


Figura 3. La equivalencia intersectiva frente a la acumulativa.

El mapeo compuesto es fundamental para asegurar la interoperabilidad de los tesauros, que están diseñados para la búsqueda postcoordinada, con los sistemas precoordinados, orientados a la búsqueda en listas jerárquicas o alfabéticas, como las que resultan respectivamente de la aplicación de clasificaciones bibliográficas o archivísticas, taxonomías o de listas de autoridades. Por ejemplo:

599 742 71(084.12) fotografías de leones (UDC)
Automobiles—Air conditioning—Maintenance and repair (LCSH)

Tras presentar las reglas generales del mapeo, ISO 25964 aborda la interoperabilidad con varios de los sistemas de organización del conocimiento más importantes, en particular, con las clasificaciones bibliográficas y archivísticas, taxonomías, sistemas de encabezamientos de materia, ontologías, terminologías, listas de control de autoridades y anillos de sinónimos. Para cada tipo de vocabulario se proporcionan sus características fundamentales, sus componentes y relaciones semánticas, y un análisis de cómo realizar el mapeo en ambas direcciones. Es importante señalar que ISO 25964 es un estándar sobre tesauros. Las descripciones de los demás sistemas de organización del conocimiento son informativas, y no se pretenden estandarizar otros tipos de sistemas de organización del conocimiento. Solo ofrece recomendaciones con carácter normativo para la interoperabilidad entre estos y los tesauros. En la Tabla 2 se presentan los resultados esperados del mapeo con algunos de los sistemas más importantes (Dextre Clarke, 2011), que, como se puede ver, ofrecen posibilidades muy diferentes.

Tabla 2. Resultados esperados del mapeo de los tesauros con otros sistemas de organización del conocimiento

Sistema objeto	Resultados esperados
Tesauros	El mapeo es relativamente sencillo
Clasificaciones bibliográficas	El mapeo conceptual resulta útil para la recuperación de información
Clasificaciones archivísticas	La precoordinación es muy frecuente
Taxonomías	
Listas de encabezamientos de materia	
Listas de autoridades de nombres	El mapeo es normalmente fácil, pero los conceptos comunes son escasos Relaciones instanciativas
Anillos de sinónimos	El mapeo conceptual no suele ser útil
Terminologías	Suele ser más útil prever usos complementarios
Ontologías	

9. La situación en España

Los tesauros llegaron a España de forma muy temprana. Fueron presentados en un libro de Lasso de la Vega de 1966, pionero de la amplia labor de difusión realizada posteriormente por Emilia Currás (ref. García, 2006). Sin embargo, la adopción formal de las normas fue relativamente tardía respecto de lo ocurrido en otros países (García Marco, 2002). En concreto, la aprobación de las correspondientes adaptaciones por la Agencia Española de Normalización y Certificación (AENOR) de las recomendaciones aprobadas por la Sociedad Internacional para la Normalización (ISO) se produjo con diez años retraso en la década de los noventa. El borrador de las Directrices para el establecimiento y desarrollo de tesauros monolingües –UNE 50-106-90, equivalente a la ISO 2788:1986, una corrección de la ISO 2788 de 1975, a su vez inspirada en las recomendaciones de la UNESCO de 1970– se publicó en la *Revista Española de Documentación Científica* en los años 1989 y 1990 (AENOR, 1989, 1990a). Finalmente, la norma fue efectivamente aprobada ese mismo año (AENOR, 1990b). Los borradores directrices para la creación y desarrollo de tesauros multilingües –UNE 50-125, un trasunto de la ISO 5964, publicada en 1985– fueron publicados en la *Revista Española de Documentación Científica* en 1996 y 1997

(AENOR, 1996, 1997a), y aprobadas y publicadas en julio de 1997 (AENOR, 1997b). Ambas pueden consultarse fácilmente en el recopilatorio de normas UNE sobre documentación publicado por la AENOR (1999).

Sin embargo, la Asociación Española de Normalización y Certificación (AENOR) se ha sumado con rapidez a los esfuerzos internacionales para revisar y ampliar el alcance de las normas sobre tesauros, y para facilitar su adaptación al entorno español. En junio de 2006 se constituyó oficialmente el grupo de Vocabularios Controlados para la Recuperación de Información del Comité 50, siguiendo la estela de la iniciativa norteamericana ANSI/NISO Z39.19-2005, y, especialmente, de la más completa propuesta por el British Standard Institution. El grupo ha participado en la elaboración de las normas ISO 25964, ha culminado la traducción y adaptación al español de la primera parte de la norma, que se publicó en noviembre de 2014, y se encuentra trabajando en la versión española de la segunda parte de la norma.

10. Conclusiones

A lo largo del presente capítulo se ha analizado cómo el proceso de digitalización y puesta en red de una enorme cantidad de recursos de información que ha supuesto Internet constituye a la vez un gran reto y una oportunidad extraordinaria para los lenguajes documentales en general y para los tesauros en particular.

Los usuarios reclaman y premian el acceso integrado a los recursos de información, especialmente si se resuelven los molestos y frecuentes resultados de recuperación indiscriminados, aportando mayor precisión y un grado de exhaustividad ajustado a las diferentes necesidades; y si se les ayuda a superar las barreras idiomáticas, culturales y disciplinares que tantas veces limitan su acceso a la información.

En este sentido resulta necesario trabajar tanto desde la perspectiva de los grandes buscadores de Internet, que funcionan como los portales efectivos para la mayoría de los usuarios, como desde el esfuerzo de integrar y hacer interoperables la multitud de bases de datos y catálogos de alta calidad que quedan ocultos, fuera del alcance de una gran parte de su público objetivo. La interoperabilidad de los sistemas de organización del conocimiento en Internet podría ser útiles a los grandes portales, como Google, para acceder con efectividad a fuentes de información indizadas con vocabulario controlado, para asegurar la interoperabilidad con bases de datos especializadas y para filtrar sus propios resultados. Recíprocamente, las bases de datos especializadas podrían tener acceso a través de portales interoperables, favoreciendo el tránsito hacia los sistemas más integrados y amigables que los usuarios reclaman.

No obstante, es necesario mantener una visión clara sobre la dificultad de la tarea. En muchos casos, como se ha visto, la interoperabilidad solo se puede conseguir de forma limitada. Además, ISO 25964 no es propiamente un estándar en el sentido más estricto de la palabra, sino que más bien ofrece una herramienta más. Como afirma Gilchrist (2007): «It should be stressed here that it is only by historical accident that these publications appear as standards; and, as such, they differ markedly from normal standards. And while they are guidelines rather than standards, nor are they textbooks. In essence they are attempts to discover best practice through a process of consensus building, made far harder by the existence of different communities of information professionals working in this area». Queda por tanto, un esfuerzo notable por delante para promover su aceptación, conjuntar voluntades, establecer alianzas y generalizar su uso.

Sin embargo, esta crítica no debe oscurecer el valor de las nuevas normas sobre tesauros, que constituyen una gran oportunidad para hacer visible la gran experiencia acumulada por la comunidad LIS en la resolución de problemas relacionados con el acceso y organización del conocimiento. Las funciones del tesoro como sistema de organización del conocimiento van más allá de la recuperación, si bien es esta su función fundamental. Cuando se construyen con la suficiente exhaustividad y sistematicidad, los tesauros permiten organizar o mapear el conocimiento utilizado en un campo de práctica o investigación, de manera que la comunidad que lo

cultiva tenga una referencia clara y sistemática de su dominio de trabajo y, a partir de ella, pueda organizar mejor su trabajo, capturar y gestionar de forma organizada su conocimiento compartido, gestionar la información independientemente de su forma, formato o canal, y organizar sus relaciones entre sí y con otros colectivos y personas. Por ello, su uso en ámbitos aplicados, más allá de su éxito en el marco de la red global, seguirá siendo un proceso clave de las estrategias de gestión del conocimiento en las organizaciones.

La protección y difusión del patrimonio cultural es uno de esos ámbitos privilegiados, y el éxito de los tesauros como herramienta de control del vocabulario para la recuperación de la información y de gestión compartida y organización del conocimiento no deja de manifestarse en la actualidad en la multitud de proyectos que se están acometiendo en muchos países, de los que dan testimonio otros capítulos de este libro. Siguen la estela de algunos de los tesauros más señeros, que precisamente han surgido del mundo del patrimonio y que han acompañado el desarrollo de esta herramienta documental desde casi sus albores.

Bibliografía

- AITCHISON, J., y DEXTRE CLARKE, S. (2004): «The thesaurus: a historical viewpoint, with a look to the future», en Roe, S. K., y Thomas, A. R., *The Thesaurus: Review, Renaissance, and Revision*. Nueva York: Haworth Press, pp. 5-21.
- AITCHISON, J., y UNESCO (1977): *UNESCO thesaurus: a structured list of descriptors for indexing and retrieving literature in the fields of education, science, social science, culture and communication*. París: UNESCO.
- AMERICAN INSTITUTE OF CHEMICAL ENGINEERS (1961): *Chemical engineering thesaurus: a wordbook for use with the concept coordination system of information storage and retrieval*. Nueva York: The Institute.
- AMERICAN NATIONAL STANDARDS INSTITUTE (1974): *American National Standard guidelines for thesaurus structure, construction and use*. Nueva York: American National Standards Institute, 1974.
- (1980): *American national standard guidelines for thesaurus structure, construction and use: approved June 30, 1980*. Nueva York: ANSI.
- ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN Y CERTIFICACIÓN (1985): *Documentación. Directrices para la creación y desarrollo de tesauros multilingüe*. Madrid: AENOR, D. L. 1997, 77 p. Normas UNE 50125. Norma equivalente a ISO 5964:1985.
- (1989): «Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües: parte 1», *Revista Española de Documentación Científica*, vol. 12, n.º 4, pp. 463-483.
- (1990): *Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, D. L. 1990, 47 p. UNE 50-106-90. Equivalente a ISO 2788-1986.
- (1990): «Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües: parte 2», *Revista Española de Documentación Científica*, vol. 13, n.º 1, pp. 601-629.
- (1995): *Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, D. L. 1995. UNE 50106:1995 ERRATUM.
- (1996): «Documentación. Directrices para la creación y desarrollo de tesauros multilingües: ISO 5964-1985, UNE 50-125», *Revista Española de Documentación Científica*, vol. 19, n.º 4, pp. 439-467.
- (1997): «Documentación. Directrices para la creación y desarrollo de tesauros multilingües: ISO 5964-1985, UNE 50-125: continuación», *Revista Española de Documentación Científica*, 1997, vol. 20, n.º 1, pp. 63-82.
- (1999): «Documentación: recopilación de normas UNE», 3ª ed. Madrid: AENOR, D. L. 1999, 580 p.
- (2014): *Información y documentación – Tesauros y su interoperabilidad con otros vocabularios – Parte 1: Tesauros para la recuperación de información*. Madrid: AENOR, D. L. 2014, 47 p. UNE 25964-1. Equivalente a ISO 25964-1.

- BERNERS-LEE, T.; HENDLER, J., y LASSILA, O. (2001): «The semantic web», *Scientific American*, 284(5), pp. 76-88.
- BLISS, H. E. (1933): *The organization of knowledge in libraries and the subject-approach to books*. Nueva York: H. W. Wilson.
— (1935): *A system of bibliographic classification*. Nueva York: Wilson.
- BLISS, H. E., y WHEELER, W. M. (1929): *The organization of knowledge and the system of the sciences*. Nueva York: Henry Holt and Company.
- BRITISH STANDARDS INSTITUTION (1978): *Draft standard guidelines for the establishment and development of monolingual thesauri*. Londres: British Standards Institution.
— (1985): *British standard guide to establishment and development of multilingual thesauri*. Londres: British Standards Institution.
— (1987): *British standard guide to establishment and development of monolingual thesauri*. Londres: British Standards Institution.
— (2005-2008): *BS 8723, Structured vocabularies for information retrieval*. Londres: British Standards Institute.
- CURRÁS, E. (1998): *Tesauros: Manual de construcción y uso*. Madrid: Kaher II.
— (2005): *Ontologías, taxonomía y tesauros: manual de construcción y uso*, 3ª ed. Gijón: Trea.
- CURRÁS, E.; AITCHISON, J., y GILCHRIST, A. (1991): *Thesaurus: lenguajes terminológicos*. Madrid: Paraninfo.
- DAHLBERG, I. (1991): «Knowledge organization, thesauri and terminology», *International Classification*, 18(3), p. 133.
— (1992): «Knowledge organization and terminology - philosophical and linguistic bases», *International Classification*, 19(2), pp. 65-71.
— (1995): «Current trends in knowledge organization», en *Organización del conocimiento en sistemas de información y documentación: actas del I Encuentro de ISKO-España, Madrid, 4 y 5 de noviembre de 1993*. Zaragoza: [ISKO-España: Universidad de Zaragoza], 7-26. http://www.iskoiberico.org/wp-content/uploads/2014/07/007-026_Dahlberg.pdf.
- DALY, L. W. (1967): *Contributions to a history of alphabetization in Antiquity and the Middle Ages*. Bruselas: Latomus.
- DEFENSE DOCUMENTATION CENTER (U.S.) (1960): *Thesaurus of ASTIA descriptors*, 1ª ed. Arlington, VA: Armed Services Technical Information Agency.
- DEXTRE CLARKE, S. G. (2008): «The last 50 years of knowledge organization: a journey through my personal archives», *Journal of Information Science*, 34(4), pp. 427-437. DOI: 10.1177/0165551508089225.
— (2011): «In Pursuit of Interoperability: Can We Standardize Mapping Types?», en *Concepts in Context – Cologne Conference on Interoperability and Semantics in Knowledge Organization, 19-20 July 2010*. Edición de Boteram, F., Goedert, W., y Hubrich, J. Colonia, Alemania: Ergon Verlag.
- DEXTRE CLARKE, S. G., y ZENG, M. L. (2012): «From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling (2012)», *Information standards quarterly*, 24(1), pp. 20-26.
- ENGINEERS JOINT COUNCIL; UNITED STATES. DEPARTMENT OF DEFENSE (1967): *Thesaurus of engineering and scientific terms; a list of engineering and related scientific terms and their relationships for use as a vocabulary reference in indexing and retrieving technical information*, 1.ª ed. Nueva York: Engineers Joint Council; Department of Defense of the United States.
- ENGINEERS JOINT COUNCIL; ENGINEERING VOCABULARY PANEL; TASK GROUP, y DODDS, R. (1967): *Thesaurus rules and conventions: approved November 9, 1966*. Nueva York: [s.n.].
- FOSKETT, A. C. (1982): *The subject approach to information*, 4.ª ed. Londres: Bingley.

- GARCÍA MARCO, F. J., y ESTEBAN NAVARRO, M. A. (1993): «On some Contributions of the Cognitive Sciences and Epistemology to a Theory of Classification», *Knowledge Organization*, 20(3), pp. 126-132.
- GILCHRIST, A. (1991): «The Standards Jungle», in *Standards for the International Exchange of Bibliographic Information: Papers Presented at a Course Held at the School of Library, Archive and Information Studies, University College London, 3-18 August 1990*. Edición de McIlwaine, L. C. Londres: Library Association.
- (2003): «Thesauri, taxonomies and ontologies - an etymological note», *Journal of Documentation*, 59(1), pp. 7-18. DOI: 10.1108/00220410310457984.
- (2007): «Revisión de las Normas Británicas BS5723 y BS6723 para el diseño y uso de tesauros: un breve informe de progreso», *Scire*, vol. 13, n.º 1.
- INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, CLASSIFICATION AND INDEXING SECTION, WORKING GROUP ON GUIDELINES FOR MULTILINGUAL THESAURI (2005): *Guidelines for Multilingual Thesauri*. IFLA, abril de 2005. URL: <http://www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf>.
- ISO (1974): ISO 2788:1974. *Documentation, guidelines for the establishment and development of monolingual thesauri = documentation, principes directeurs pour l'établissement et le développement de thesaurus monolingues*, 2.ª ed. Ginebra: ISO, 1974.
- (1985): ISO 5964:1985. *Documentation, Guidelines for the establishment and development of multilingual thesauri*. Ginebra: ISO, 1985.
- (1986): ISO 2788:1986. *Documentation, guidelines for the establishment and development of monolingual thesauri = documentation, principes directeurs pour l'établissement et le développement de thesaurus monolingues*. 2.ª ed. Ginebra: ISO, 1986.
- (2011): ISO 25964-1:2011. *Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Ginebra: International Society for Standardization.
- (2013): ISO 25964-2:2013. *Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Ginebra: International Society for Standardization.
- KROOKS, D. A., y LANCASTER, F. W. (1993): «The Evolution of Guidelines for Thesaurus Construction», *Libri*, 43(4), pp. 326-342. DOI: 10.1515/libr.1993.43.4.326.
- LANCASTER, F. W. (1986): *Vocabulary control for information retrieval*, 2.ª ed. Arlington, Virginia: Information Resources Press.
- LASSO DE LA VEGA, J. (1966): «Organización de la documentación por palabras y descriptores específicos de materias: los thesaurus». *Racionalización* 19(6), pp. 433-45.
- LLOYD, B. B., y ROSCH, E. (1978): *Cognition and categorization*. Hillsdale, N. J. Nueva York; Londres: Lawrence Erlbaum; Halsted Press division of Wiley distributor.
- LUHN, H. P. (1957): «A Statistical Approach to Mechanized Encoding and Searching of Literary Information», *IBM Journal of Research and Development*, 1(4), pp. 309-316.
- MILLER, G. A. (1956): «The magical number seven, plus or minus two: Some limits on our capacity for processing information», *Psychological Review*, 63, pp. 81-97.
- MOOERS, C. N. (1950): «Zatocoding for punched cards», *Zator Technical Bulletin*, 30.
- (1951): «Zatocoding applied to the mechanical organization of knowledge», *American Documentation*, 2(1), pp. 20-32.
- (2003): «Descriptors», en *Encyclopedia of Library and Information Science*. Marcel Dekker, pp. 813-821. DOI: 10.1081/E-ELIS 120008981.
- NATIONAL INFORMATION STANDARDS ORGANIZATION (ESTADOS UNIDOS); AMERICAN NATIONAL STANDARDS INSTITUTE (1994): *Guidelines for the construction, format, and management of monolingual thesauri: an American national standard*. Bethesda, Md.: NISO Press.

- (1995): *Developing the Next Generation of Standards for Controlled Vocabularies and Thesauri*. Bethesda, Md.: National Information Standards Organization. URL: <<http://www.niso.org/committees/MT-info.html>>.
- (2005): *Guidelines for the construction, format, and management of monolingual controlled vocabularies: an American national standard*. Bethesda, Md.: National Information Standards Organization, 2006. ANSI/NISO Z39.19-2005. Approved July 25, 2005 by the American National Standards Institute.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (1972): *Macrothesaurus; a basic list of economic and social development terms*, 1.^a ed. en inglés. París.

PINTO MOLINA, M., GARCÍA MARCO, F. J., y AGUSTÍN LACRUZ, M. C. (2002): *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos*. Gijón: Trea.

RANGANATHAN, S. R. (1933): *Colon classification*. Madrás: Madras Library Association; Londres: E. Goldston.

- (1937): *Prolegomena to library classification*. Madras: Madras Library Association; London: E. Goldston.

ROGET'S INTERNATIONAL THESAURUS (1946): Nueva York, Thomas Y. Crowell Co.

TAUBE, MORTIMER & ASSOCIATES (1953): *Studies in coordinate indexing*. Washington: Documentation Inc.

UNESCO, UNISIST (1973): *Principes directeurs pour l'établissement et le développement de thésaurus monolingues*. París: UNESCO, 34 p. (SC/WS/555).

- (1980): *Principes directeurs pour l'établissement et le développement de thésaurus multilingues*. París: UNESCO, 88 p. (PGI/80/WS/12).

Los tesauros en la web semántica: SKOS y la norma ISO 25964

Juan Antonio Pastor Sánchez

Universidad de Murcia

1. Introducción

En 1992 Tim Berners-Lee presentó la Web, que con el paso del tiempo se convertiría en la tecnología que cambió la percepción y el desarrollo de los servicios de información digital. Tanto es así que hoy día son infinidad de aplicaciones y sistemas los que utilizan la plataforma Web como un entorno de desarrollo e intercambio de datos (Guinard; Trifa y Wilde, 2010). Redes sociales, repositorios de objetos digitales, servicios financieros, blogs o comercio electrónico son, entre otros, algunos buenos ejemplos de la heterogeneidad de las aplicaciones de la web como cliente-servidor universal.

Sin embargo, actualmente estamos viviendo una revolución de la web que, al contrario de la Web 2.0, puede que no resulte visible para el gran público. El paradigma del desarrollo de la web ya no es el documento (en sus diferentes formatos), sino la publicación de datos estructurados (Heath y Bizer, 2010). Para ello se han propuesto y desarrollado múltiples estándares y tecnologías para publicar datos estructurados en la web bajo la denominación de 'web semántica' (Berners-Lee; Hendler y Lassila, 2001). Denominaciones como Linked Open Data, Hyperdata o Data on the Web se refieren a distintos «sabores» de una misma realidad: la publicación de grandes volúmenes de datos altamente estructurados en la web. En este sentido, la web semántica ha dejado de ser un proyecto de futuro para ser un hecho del presente. El principio subyacente sigue permaneciendo: el intercambio de información. Pero con una sutil diferencia: en la web original las personas generan y consumen los contenidos, mientras que en la web semántica son las máquinas¹ en primera instancia a quienes van dirigidos los datos.

¹ En la jerga de los especialistas de la web semántica el término «máquina» se utiliza para referirse a un *software* o aplicación informática. Esta distinción es importante, puesto que máquinas y personas comparten un mismo sistema (la web) y protocolo de Internet (http) para el intercambio de información.

Si se desea que las máquinas sean capaces de procesar los datos, es preciso que estos se encuentren en un formato y una estructura adecuados para ello. El concepto subyacente es el de interoperabilidad: la capacidad de un sistema de información para compartir datos de modo que sea posible el intercambio de información con otros sistemas directamente, ya sea existentes o futuros, y sin restricción de acceso o implementación (Sheth, 1999). La interoperabilidad supera los procesos de importación y exportación que conllevan la transformación de los datos previo mapeado de su estructura. La interoperabilidad también es una salvaguarda para que los sistemas de información que se desarrollen en un futuro no encuentren problemas de compatibilidad con los datos gestionados por sistemas que con el paso del tiempo quedan obsoletos. También es el primer paso para la reutilización de los datos donde se tienen en cuenta, entre otros, aspectos tales como el acceso, la calidad, la utilidad o las licencias de uso. Interoperabilidad y reutilización conforman un binomio en el que los datos se erigen en un recurso para crear nuevos productos, servicios y modelos de negocio, en el que los conjuntos de datos estructurados son la materia prima (Ramos Simón *et alii*, 2012).

Los datos estructurados precisan de modelos conceptuales que permitan representar abstracciones de ideas u objetos del mundo. Los modelos descriptivos cumplen esta función y permiten caracterizar los objetos de información mediante atributos y relaciones lógicas (Solodovnik, 2011). En el ámbito de la web semántica lo anterior se sustancia en la aplicación de vocabularios de esquemas de metadatos y ontologías. Mientras que los metadatos describen recursos de información mediante pares de atributo-valor, las ontologías clasifican dichos recursos en clases y especifican de un modo muy detallado cómo se relacionan entre sí.

En la web semántica se hace un uso intensivo de las ontologías y algunas de ellas, como SKOS, ha resultado ser esenciales en su evolución. Ello es debido a que en la web semántica es imprescindible disponer de vocabularios controlados que permitan describir el contenido de los conjuntos de datos. De este modo es posible la desambiguación terminológica y la organización semántica de conceptos, recursos y datos (Montalvo-Montalvo, 2011). SKOS permite representar, en forma de conjuntos de datos RDF, lenguajes documentales del tipo tesauros, clasificaciones, encabezamientos de materia, glosarios, etc. (Pastor-Sánchez; Martínez-Méndez y Rodríguez-Muñoz, 2009 y 2012). Así pues, con SKOS es posible representar sistemas de organización del conocimiento de forma interoperable para su posterior reutilización.

2. Arquitectura y tecnologías de la web semántica

Todo entorno tecnológico precisa de una arquitectura que describa y presente la estructura organizativa que define las interrelaciones y roles de los diferentes elementos que lo conforman. La web semántica no es ajena a lo anterior y la Figura 1 ofrece una panorámica de dicha arquitectura.

La web semántica no es una tecnología nueva. Sigue basándose en la plataforma web convencional y utiliza, por tanto, el protocolo HTTP para transmitir peticiones y datos entre clientes y servidores. UNICODE es el estándar utilizado para la definir los símbolos de cualquier alfabeto o de otro tipo (matemáticos, fonéticos, etc.) y codificarlos mediante alguna de las variantes del formato UTF². Mediante referencias URI³ se dispone de un mecanismo para referenciar de forma inequívoca recursos de información y elementos de vocabularios descriptivos. La variante IRI (Internationalized Resource Identifier) permite el uso de caracteres UNICODE para representar dicho identificador. Se dice que una URI es derreferenciable cuando puede utilizarse para acce-

² UTF: UNICODE Transformation Format. Formato de codificación de caracteres UNICODE con tres posibles variantes de 8, 16 y 32 bits (según los bits utilizados para representar un carácter) denominados UTF-8, UTF-16 y UTF-32, respectivamente.

³ URI (Uniform Resource Identifier): son cadenas de caracteres estandarizadas [RFC3986] que conforman espacios de identificadores, que permiten referenciar cualquier tipo de recurso abstracto o localizable en Internet.

der a un recurso mediante una conexión HTTP como si se trata de una URL⁴ normal o corriente. En otros casos el recurso no está accesible en línea, por lo que se identifica mediante una URN⁵.

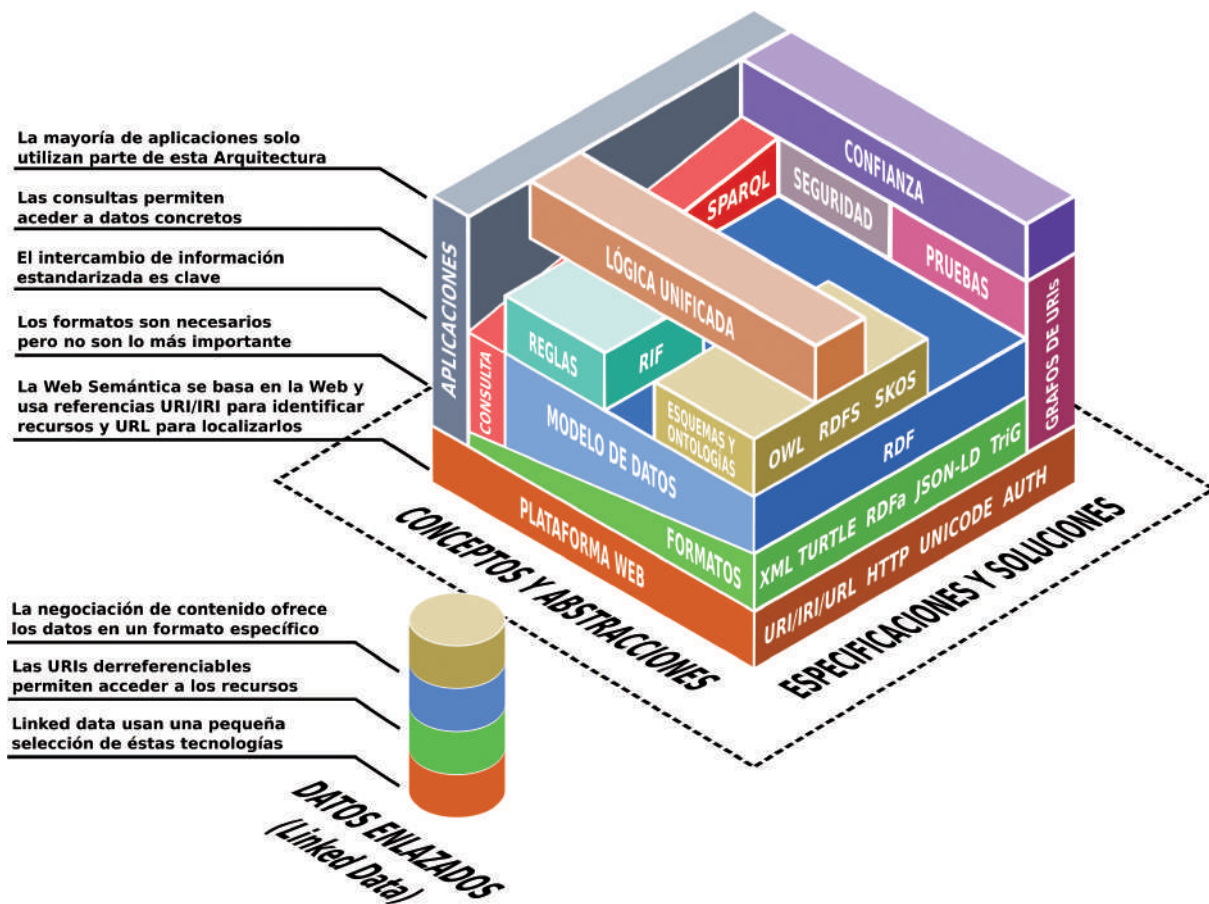


Figura 1. La arquitectura de la web semántica. Traducido y adaptado de: <http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake>.

La familia de tecnologías XML (*eXtensible Markup Language*) aporta una sintaxis que permite el intercambio y procesamiento de datos. Dentro de este conjunto de tecnologías los espacios de nombres XML (XML namespaces)⁶ permiten utilizar varios vocabularios en un mismo documento XML resolviendo posibles conflictos entre nombres. De este modo, se puede reutilizar de forma conjunta varios esquemas de metadatos, configurando diferentes soluciones ajustadas a dominios de datos o aplicaciones específicas.

También se precisa un modelo de datos que identifique recursos y describa sus propiedades y relaciones con otros recursos. Mediante RDF dicha descripción se realiza en forma de tripletas del tipo recurso-propiedad-valor o recurso-relación-recurso. Cada recurso se identifica de manera unívoca mediante una URI y se describe asignando valores a determinados atributos o estableciendo relaciones entre dos recursos (W3C, 2014). En el primer caso los recursos se vinculan con los valores mediante propiedades de datos. En el segundo caso los recursos se relacionan entre sí a través de propiedades de objeto. Cada propiedad está asociada con una URI, de manera que, al

⁴ URL (Uniform Resource Locator, Localizador Uniforme de Recurso) [RFC2718]. URL es un subconjunto de URI que permiten localizar y acceder a un recurso cuya aplicación está muy extendida debido a su uso en la web convencional.

⁵ URN (Uniform Resource Name) subconjunto de URI [RFC1737] que permite identificar recursos pero sin que impliquen su disponibilidad en red, como, por ejemplo, urn:isbn:0-123-45678-9.

⁶ Más información en: <http://www.w3.org/TR/xml-names/>.

igual que los recursos, se identifican de manera unívoca. Por lo tanto, las tripletas de descripciones conforman grafos, en donde los recursos se representan como nodos, los valores como literales y las propiedades mediante arcos. Las referencias URI pueden acortarse con prefijos (ver Figura 2).

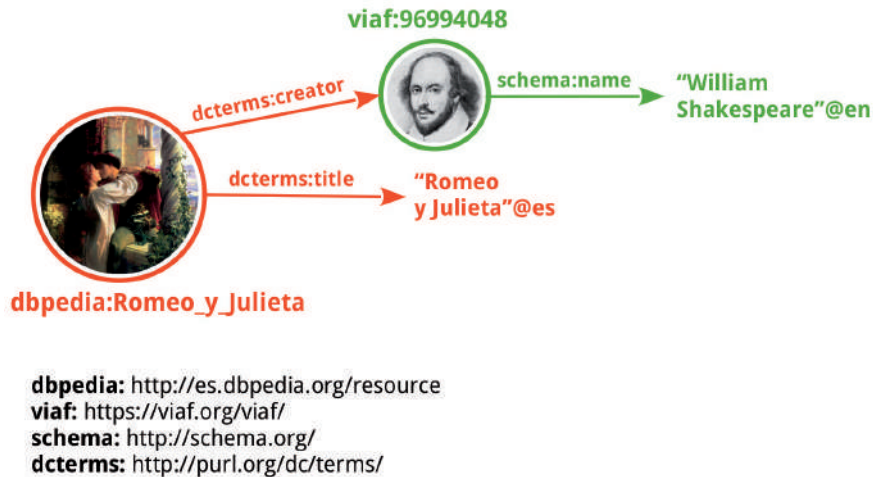


Figura 2. Ejemplo de grafo RDF. Fuente: elaboración propia.

Los grafos de sentencias RDF tienen un alto grado de expresividad para las personas, pero lo cierto es que, para que se puedan intercambiar y procesar por máquinas⁷, es preciso utilizar una representación textual. Este proceso de codificación (serialización) precisa aplicar un formato para expresar las sentencias RDF según una sintaxis normalizada. Existen diversas alternativas para la serialización de RDF, tales como RDF/XML⁸, Turtle⁹, JSON-LD¹⁰ o RDFa¹¹, entre otros.

RDF/XML	<pre><rdf:Description rdf:about="http://www.wikidata.org/entity/Q83186"> <dcterms:creator rdf:resource="https://viaf.org/viaf/96994048"/> <dcterms:title>Romeo and Juliet</dcterms:title> </rdf:Description></pre>	Permite aplicar las tecnologías XML (XSLT, XPath, XQuery, etc.)
Turtle	<pre><http://www.wikidata.org/entity/Q83186> dcterms:creator <https://viaf.org/viaf/96994048> ; dcterms:title "Romeo and Juliet"@en .</pre>	Compacto y ligero. Ideal para su uso por máquinas.
JSON-LD	<pre>@id:"http://www.wikidata.org/entity/Q83186", "dcterms:creator": "https://viaf.org/viaf/96994048", "dcterms:title": {"@language": "en", "@value": "Romeo and Juliet"}</pre>	Diseñado para el consumo de RDF por JavaScript.
RDFa	<pre><div resource="http://www.wikidata.org/entity/Q83186"> <h1 property="dcterms:title">Romeo and Juliet</h1> <p>Author: William Shakespeare</p> </div></pre>	Marcado semántico de documentos HTML insertando sentencias RDF.

Figura 3. Formatos para la serialización de RDF. Fuente: elaboración propia.

⁷ En este contexto por «máquina» se entienden programas informáticos capaces de tomar ciertas decisiones de forma autónoma durante el procesamiento de datos en función del contenido de los mismos.

⁸ <http://www.w3.org/TR/rdf-syntax-grammar/>.

⁹ <http://www.w3.org/TR/turtle/>.

¹⁰ <http://www.w3.org/TR/json-ld/>.

¹¹ <https://www.w3.org/TR/rdfa-primer/>.

Es posible establecer jerarquías de recursos mediante la definición de clases y crear nuevas propiedades a partir de otras utilizando RDF Schema (RDFS). Los desarrollos que utilizan RDF y RDFS permiten el desarrollo de lo que se conoce como «ontologías ligeras». Para elaborar ontologías con mayor contenido semántico y nivel de formalización, se precisa además el uso de OWL (Web Ontology Language). Una de las principales aplicaciones de las ontologías es la ejecución de procesos de inferencia para descubrir información mediante la definición de reglas. RIF (W3C, 2013) es un lenguaje que permite expresar este tipo de reglas para que puedan intercambiarse.

Con estas premisas, la selección y obtención de información acerca de descripciones individuales de recursos requiere un lenguaje de búsqueda específico. Algo parecido al papel que desempeña SQL cuando se representa la información siguiendo el modelo relacional de bases de datos. En la web semántica esta tarea se basa en la búsqueda y recuperación de patrones de tripletas RDF a través de SPARQL (W3C, 2013b). SPARQL define tanto el lenguaje de recuperación como el formato en el que se obtienen los datos recuperados. Una persona o una máquina puede utilizar unos servicios web específicos conocidos como SPARQL Endpoint para enviar una consulta y recuperar la información solicitada.

Una característica de las aplicaciones realizadas bajo el paraguas de la web semántica es la transparencia que para el usuario supone la combinación y uso de todas estas tecnologías. Tal vez el éxito de la web semántica es que el usuario es ajeno a su existencia, percibiendo únicamente la existencia de nuevos y mejores servicios de información. Además, estas tecnologías son escalables en función de las necesidades de un determinado proyecto, siendo posible utilizar RDF y RDFS sin que sea obligatorio desarrollar o aplicar una ontología OWL o implementar un SPARQL Endpoint.

3. SKOS

SKOS es una ontología OWL-Full (W3C, 2009) para representar sistemas de organización del conocimiento mediante RDF. En SKOS los recursos RDF pueden ser de tres clases:

- **Conceptos (skos:Concept)**. Son los elementos fundamentales de SKOS. Se trata de entidades abstractas, existentes en nuestra mente con independencia de los términos (unidades léxicas) usados para etiquetarlos. Pueden contemplarse como una idea, una noción, una referencia a un objeto u hecho real, etc.
- **Esquemas de conceptos (skos:ConceptScheme)**. Es una agregación de conceptos. Aunque pueden definirse como entidades independientes, suelen reunirse en vocabularios como tesauros, clasificaciones, etc. Los esquemas de conceptos permiten identificar el vocabulario en su totalidad o en parte.
- **Colecciones (skos:Collection / skos:OrderedCollection)**. SKOS permite definir grupos o «colecciones» de conceptos cuando un conjunto de estos comparten uno o varios aspectos en común. En estos casos es conveniente agruparlos bajo una etiqueta propia. En ciertas colecciones los conceptos pueden ordenarse según un criterio determinado.

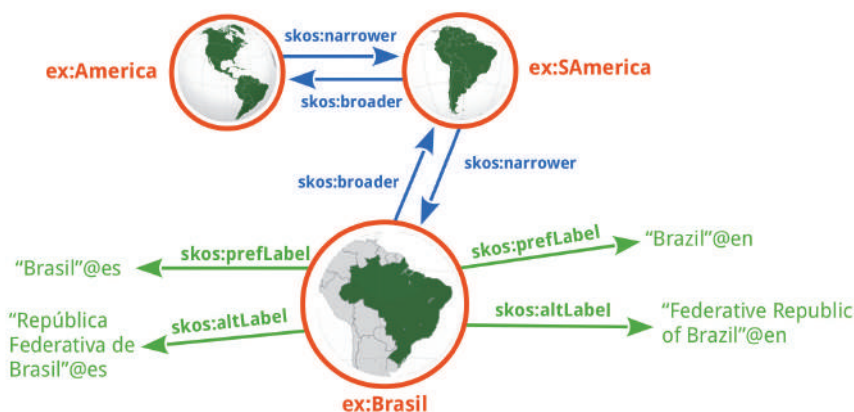
Con respecto a los conceptos que tienen asignadas etiquetas en uno o varios idiomas, existen diferentes tipos de propiedades para el etiquetado:

- **Etiquetas preferentes (skos:prefLabel)**. Un recurso únicamente puede tener una etiqueta preferente en cada idioma. Dos conceptos tampoco deberían compartir la misma etiqueta en un mismo idioma. En la terminología de tesauros podrían identificarse con los «descriptores».
- **Etiquetas alternativas (skos:altLabel)**. Permiten expandir la asignación de una sola etiqueta preferente en un idioma para un concepto. Esto resulta de gran utilidad para representar sinónimos, cuasisinónimos, abreviaturas, acrónimos, etc. En la terminología de tesauros podrían identificarse con los «no descriptores».

- **Etiquetas ocultas (skos:hiddenLabel)**. Es una cadena de caracteres accesible para las aplicaciones que procesan un vocabulario SKOS, pero no es visible directamente para los usuarios. Las etiquetas ocultas permiten incluir variantes u errores ortográficos de otras etiquetas léxicas. De esta forma, si el usuario comete un error al teclear una consulta, el sistema podría encontrar conceptos relevantes si hay alguna correspondencia entre el texto introducido y una etiqueta oculta.

También existen una serie de propiedades que permiten vincular conceptos a través de la definición de relaciones semánticas:

- **Relaciones jerárquicas (skos:broader / skos:narrower)**. Permiten establecer cuándo el significado de un concepto es más genérico o específico que otro. Son propiedades inversas entre sí. Es decir, cuando un concepto X es más genérico que el concepto Y, entonces el concepto Y es más específico que el concepto X.
- **Relaciones jerárquicas transitivas (skos:broaderTransitive / skos:narrowerTransitive)**. Las propiedades skos:broader y skos:narrower no están definidas explícitamente como transitivas. Esto significa que su semántica no soporta inferencias de manera formal. Para aplicaciones que precisen este tipo de semántica, SKOS dispone de las propiedades skos:broaderTransitive y skos:narrowerTransitive.
- **Relaciones asociativas (skos:related)**. Representa no jerárquicas como las que se definen entre un tipo de evento y las entidades que participan en él. Esta propiedad es simétrica, de forma que si un concepto X está relacionado con un concepto Y, entonces el concepto Y está relacionado con X. SKOS no define esta propiedad como transitiva.



```
@Prefix ex: <http://www.example.org/>.
@Prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@Prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
```

```
ex:America rdfs:type skos:Concept;
skos:narrower ex:NAmerica;
skos:narrower ex:SAmerica.
```

```
ex:SAmerica rdfs:type skos:Concept;
skos:broader ex:America;
skos:narrower ex:Brasil.
```

```
ex:Brasil rdfs:type skos:Concept;
skos:prefLabel "Brazil"@en;
skos:prefLabel "Brasil"@es;
skos:altLabel "Federative Republic of Brazil"@en;
skos:altLabel "República Federativa do Brasil"@es;
skos:broader ex:SAmerica.
```

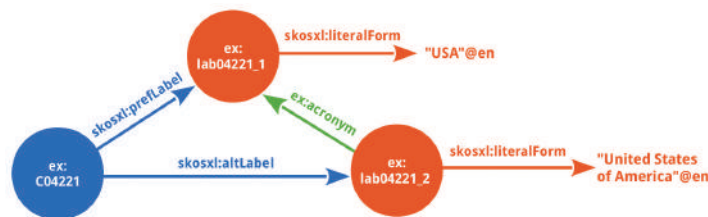
Figura 4. Ejemplo de grafo SKOS con su correspondiente serialización. Fuente: elaboración propia.

Los conceptos pueden vincularse a uno o más esquemas de conceptos utilizando **skos:inScheme**. Para indicar que un concepto es cabecera dentro de un esquema de conceptos, se utiliza la propiedad **skos:topConceptOf**, mientras que para definir el sentido inverso de la relación, es decir, aquellos conceptos cabecera del esquema, se usa **skos:hasTopConcept**. Para establecer los conceptos que pertenecen a una colección o a una colección ordenada, se utilizan las propiedades **skos:member** y **skos:memberList**, respectivamente.

SKOS dispone de diversas propiedades de documentación:

- **Notas de alcance (skos:scopeNote)**. Proporcionan información sobre la aplicación y límites del concepto generalmente durante el proceso de indexación.
- **Definiciones (skos:definition)**. Permite ofrecer una explicación completa del significado de un concepto.
- **Ejemplos (skos:example)**. Mediante esta propiedad es posible representar ejemplos concretos a los que se refiere el concepto.
- **Notas editoriales (skos:editorialNote)**. Ofrece información sobre procesos administrativos y trabajo de edición y validación del vocabulario.
- **Notas históricas (skos:historyNote)**. Proporciona a los usuarios del vocabulario cambios sustanciales en el significado o la forma del concepto.
- **Notas sobre cambios (skos:changeNote)**. Permite describir cambios sobre un concepto con fines de administración y mantenimiento.

En algunos casos es necesario definir relaciones entre etiquetas. Sin embargo, las propiedades de etiquetado léxico de SKOS (**skos:prefLabel**, **skos:altLabel**, **skos:hiddenLabel**) no pueden ser sujeto de ninguna sentencia RDF, puesto que son cadenas literales y no recursos RDF. Por tanto, no es posible definir relaciones entre ellos. SKOS eXtension for Labels (SKOS-XL) permite identificar, describir y definir vínculos entre entidades léxicas. Para ello hay que utilizar la clase **skosxl:Label**, que permite definir etiquetas como si fueran recursos RDF. A los recursos de dicha clase se asigna una etiqueta literal mediante la propiedad **skosxl:literalForm**. Para vincular conceptos SKOS con recursos de la clase **skosxl:Label**, se utilizan las propiedades **skosxl:prefLabel**, **skosxl:altLabel** y **skosxl:hiddenLabel**. Con esta propiedad pueden definirse relaciones entre recursos **skosxl:Label**. De esta forma es posible definir relaciones personalizadas, como subpropiedades de **skosxl:labelRelation**, entre recursos de la clase **skosxl:Label**.



```
@Prefix rdf: <http://id.sgcb.mcu.es/Autoridades/>.
@Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@Prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@Prefix skosxl: <http://www.w3.org/2008/05/skos-xl#>.
@Prefix ex: <http://www.example.org/>.

ex:C04221 rdf:type skos:Concept;
skosxl:prefLabel ex:lab04221_1;
skosxl:altLabel ex:lab04221_2.

ex:lab04221_1 rdf:type skosxl:Label;
skosxl:literalForm "USA"@en.
ex:lab04221_2 rdf:type skosxl:Label;
skosxl:literalForm "United States of America"@en.

ex:acronym rdfs:subPropertyOf skosxl:labelRelation.
ex:lab04221_1 ex:acronym ex:lab04221_2.
```

Figura 5. Ejemplo de SKOS-XL con su correspondiente serialización. Fuente: elaboración propia.

SKOS también dispone de un conjunto de propiedades que permiten configurar redes entre diferentes KOS mediante relaciones de mapeado que se establecen entre conceptos de diferentes esquemas:

- **Equivalencia jerárquica (skos:broadMatch / skos:narrowMatch):** se utiliza cuando se desea indicar que un concepto de un esquema tiene un significado más genérico o específico que un concepto de otro esquema.
- **Equivalencia asociativa (skos:relatedMatch):** cuando dos conceptos de diferentes esquemas están relacionados semánticamente.
- **Equivalencia exacta (skos:exactMatch):** usado cuando dos conceptos tienen un significado prácticamente idéntico siendo intercambiables en una gran variedad de aplicaciones de recuperación de información. Es una propiedad transitiva.
- **Equivalencia cercana o próxima (skos:closeMatch):** usado cuando dos conceptos tienen significado similar siendo intercambiables en ciertas aplicaciones de recuperación de información.

Adicionalmente, SKOS define una serie de condiciones de consistencia y reglas –especialmente referidas a la transitividad de las relaciones semánticas– que delimitan el ámbito de aplicación de los vocabularios en procesos lógicos de inferencia.

4. La norma ISO 25964

La nueva norma ISO 25964 sobre tesauros actualiza y reemplaza a las anteriores normas ISO 5964:1985 e ISO 2788:1986. Se basa en los trabajos de la norma británica, contemplando y armonizando las propuestas que se realizaron en la misma (García-Marco, 2014). La norma ISO 25964 se estructura en dos partes:

- Parte 1: Tesauros para la recuperación de la información (ISO 25964-1:2011).
- Parte 2: Interoperabilidad con otros vocabularios (ISO 25964-2:2013).

Como puede verse en la Figura 6, la primera parte de ISO 25964 engloba las normas originales ISO 2788 y 5964. De igual forma, se corresponde con las dos primeras partes de la norma británica BS 8723. La segunda parte se centra en la interoperabilidad entre vocabularios, y en tal sentido adopta el enfoque de las partes tercera, cuarta y quinta de la BS 8723.

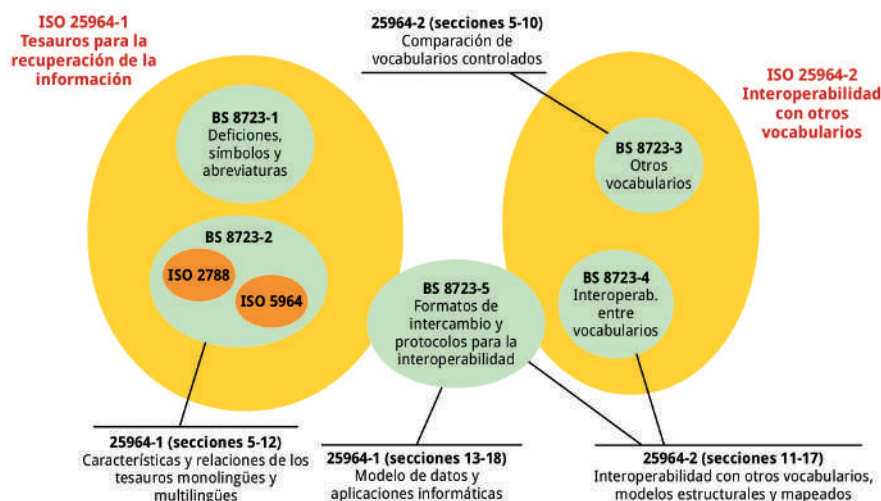


Figura 6. Correspondencia de ISO 25964 en el contexto normativo previo. Traducido y adaptado de: <https://www.reseau-canope.fr/savoirscdi/uploads/pics/image4_01.png>

El alcance de la norma ISO 25964 es realmente amplio, puesto que define detalladamente la estructura de los tesauros, sus elementos, los diferentes tipos de equivalencias intralingüísticas e interlingüísticas o las relaciones semánticas. También se centra en los aspectos de gestión, como la creación y mantenimiento de los tesauros, su aplicación, visualización, el uso de formatos y el mapeado e interoperabilidad con otros vocabularios. Uno de los aspectos más relevantes es que también se define un modelo de datos basado en UML y un esquema XML para representar tesauros.

Según el modelo de datos de ISO 25964 (Figura 7), un tesoro (*Thesaurus*) está formado por conceptos (*ThesaurusConcept*) que pueden agruparse en grupos de conceptos (*ConceptGroup*) o listas de conceptos hermanos (*ThesaurusArray*). Tanto los grupos de conceptos como los arrays de términos pueden etiquetarse y anidarse. Asimismo, es posible definir diferentes versiones de un tesoro (*VersionHistory*). Los conceptos pueden relacionarse entre sí mediante relaciones semánticas jerárquicas (*HierarchicalRelationship*), asociativas (*AssociativeRelationship*) o con sus correspondientes conceptos cabecera de nivel superior (*TopLevelRelationship*). Las relaciones jerárquicas y asociativas pueden especializarse a través del atributo «role», que permite definir distintos subtipos de relaciones semánticas de estos tipos. Pueden definirse atributos personalizados para los conceptos (*CustomConceptAttribute*) y establecer de esta forma características adicionales para los mismos.

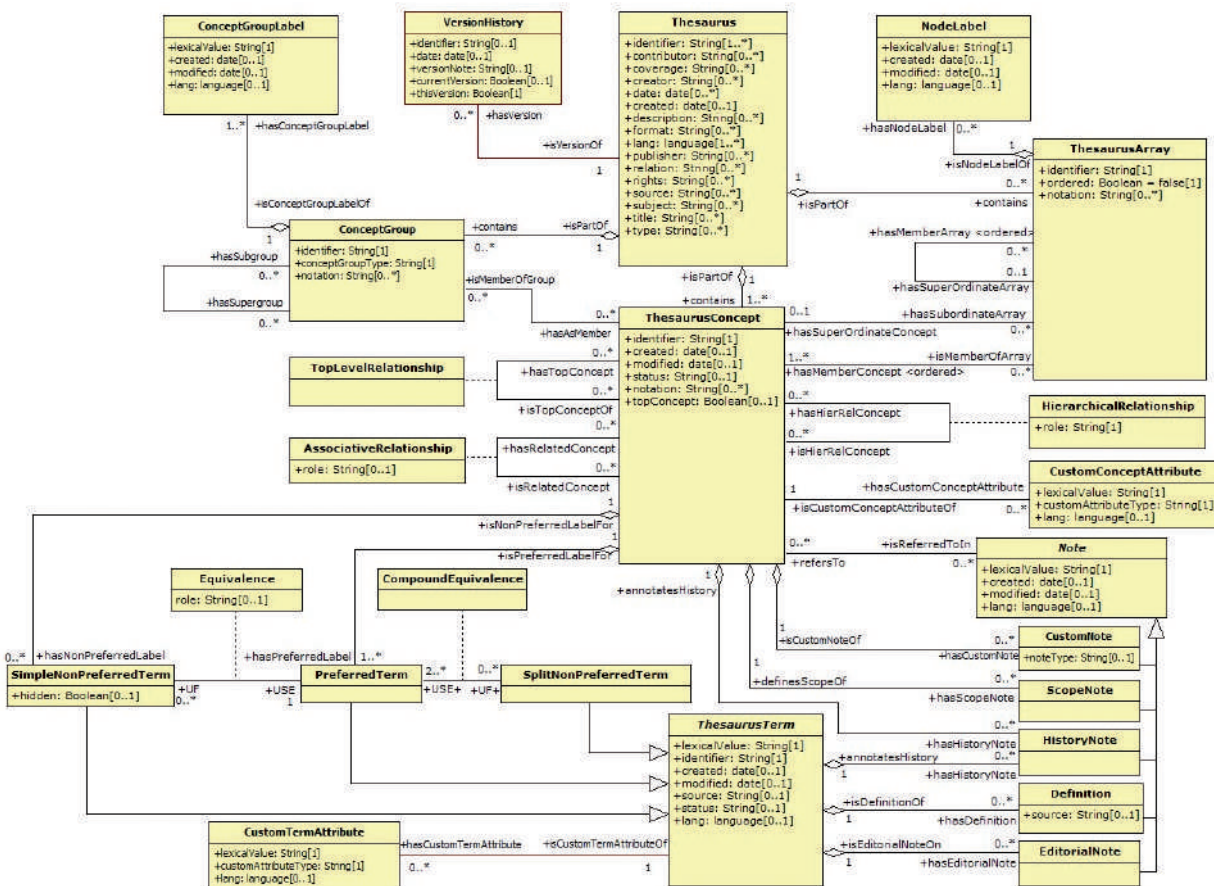


Figura 7. Modelo de datos de ISO 25964 para la representación de tesauros.

Fuente: <http://www.niso.org/schemas/iso25964/Model_2011-06-02.jpg>

Los términos preferentes y no-preferentes pueden relacionarse directamente con un concepto. También es posible definir equivalencias entre los propios términos. Se pueden definir

equivalencias compuestas para los casos en los que un término no-preferente tenga más de un término preferente. Las tareas de definición y mantenimiento del nivel léxico de un tesoro con la nueva norma ISO 25964 pueden resultar relativamente complejas, debido a las 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 (*SplitNonPreferredTerm*) con varios términos preferentes (*PreferredTerm*) mediante una relación de equivalencia compuesta (*CompoundEquivalence*).

La segunda parte de la norma ISO 25964 aborda la interoperabilidad de los tesauros con otros vocabularios, definiendo una amplia tipología de equivalencias entre conceptos de diferentes vocabularios:

- Exactas: conceptos que representan exactamente la misma idea. Por ejemplo: «enfermedad de las vacas locas», «encefalopatía espongiiforme bovina».
- Inexactas: conceptos que pueden ser equivalentes en determinados contextos o tener diferencias de connotación. Por ejemplo: «sillas» y «asientos».
- Parciales: conceptos con un significado muy cercano, aunque uno de ellos es ligeramente más genérico o específico que otro. Por ejemplo: «Aeronave» y «Aeroplanos».
- Jerárquicas: un concepto es claramente más genérico o específico que otro. Por ejemplo: «Países Bajos» es más genérico que «Holanda».
- 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 se reflejan de forma explícita en el modelo de datos de la norma, aunque pueden derivarse a partir de relaciones jerárquicas y asociativas.

5. Representación de tesauros mediante ISO-THES

Tal y como se ha expuesto anteriormente, existen muchas similitudes entre la nueva norma ISO 25964 y SKOS:

- 1.º Ofrecen modelos de representación aplicables para el desarrollo de aplicaciones informáticas de recuperación de información, como puede observarse en la Figura 2.
- 2.º Contemplan un nivel conceptual y un nivel terminológico (léxico).
- 3.º Permiten definir estructuras superiores de agrupación de conceptos: *Thesaurus*, *Concept-Group* o *ThesaurusArray*, para representar tesauros, grupo de conceptos y arrays de conceptos en el caso de ISO 25964. Con respecto a SKOS, estas estructuras son las colecciones y los esquemas de conceptos.
- 4.º Los elementos de anotación son casi idénticos: notas generales (*Note*), notas de alcance (*ScopeNote*), notas de historial (*HistoryNote*), definiciones (*Definition*), notas de edición (*EditorialNote*) y notas personalizadas (*CustomNote*).

Sin embargo, SKOS carece de elementos nativos para la composición de equivalencias, grupos y arrays de conceptos y personalización de relaciones semánticas. En cualquier caso hay que considerar que la nueva norma ISO se publica en un momento de amplio despliegue de SKOS para publicar vocabularios controlados.

Para facilitar la interoperabilidad entre ambas propuestas, se ha desarrollado la ontología ISO-THES que reutiliza ciertos elementos de SKOS para representar tesauros en el ámbito de

la web semántica conforme a ISO 25964 (Isaac y De Smedt, 2015)¹². ISO-THES reutiliza SKOS y SKOS-XL y define nuevas propiedades y clases para representar tesauros conformes al modelo de datos establecido por la norma ISO 25964. De este modo, define una serie de elementos propios que permiten expresar la riqueza semántica del modelo de la norma ISO. Las características generales de esta ontología podrían resumirse en los siguientes puntos:

- Se reutilizan las clases de SKOS para representar conceptos, esquemas de conceptos y colecciones. Se definen las clases **iso-thes:ConceptGroup** e **iso-thes:ThesaurusArray** para representar los grupos de conceptos y los arrays.
- Se reutilizan las propiedades de objeto que permiten representar las relaciones semánticas entre conceptos. No obstante, se especifica que para alcanzar el potencial del modelo de datos de la norma ISO es necesario definir subpropiedades a partir de **skos:broader**, **skos:narrower** y **skos:related** para representar de forma flexible diferentes tipos de relaciones jerárquicas y asociativas.
- Se definen relaciones jerárquicas genéricas, partitivas e instanciales. Son relaciones que expresan una semántica más precisa que **skos:broader/skos:narrower**, contemplada directamente por la norma ISO.
- Las propiedades de etiquetado de SKOS se reutilizan para el etiquetado simple de los conceptos, mientras que SKOS-XL se usa para definir las equivalencias simples o compuestas entre etiquetas y la asignación de las mismas a los conceptos.

Para plantear un caso práctico para la aplicación de SKOS y la ontología ISO-THES, se puede recurrir a un escenario clásico en el que un tesoro se modela mediante una estructura de áreas o micro-tesauros como la del siguiente modelo:

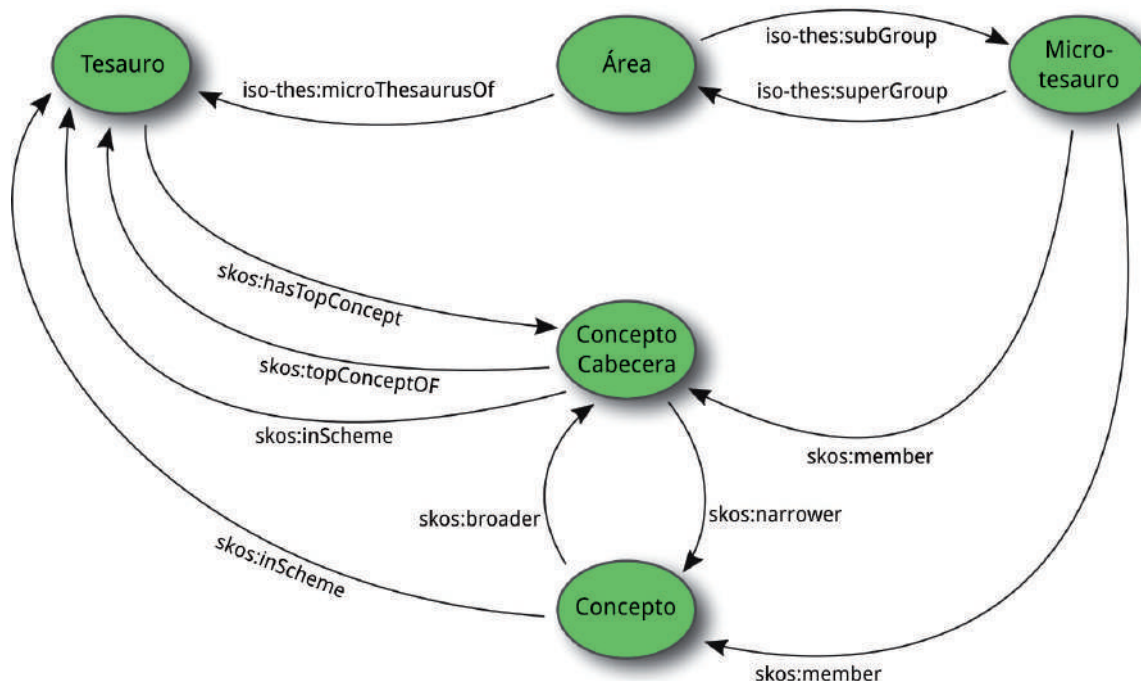


Figura 8. Modelo de representación de un tesoro estructurado en áreas y micro-tesauros mediante SKOS e ISO-THES. Fuente: elaboración propia.

¹² Existe un trabajo previo en el que se definen las correspondencias entre SKOS y los elementos del modelo de ISO 25964: <<http://www.niso.org/schemas/iso25964/correspondencesSKOS/>>.

Los tesauros se representarían como esquemas de conceptos (**skos:ConceptScheme**), mientras que las áreas y los micro-tesauros como grupos de conceptos (**skos:ConceptGroup**). Los conceptos, independientemente que fueran conceptos cabecera o no, se representarían mediante **skos:Concept**. Las áreas se vincularían con el tesoro a través de la propiedad **iso-thes:microThesaurusOf**. A su vez los micro-tesauros se vincularían con las áreas utilizando **iso-thes:superGroup**, al tiempo que se usaría **iso-thes:subGroup** para dicho vínculo pero en sentido inverso (desde el área al micro-tesauro). Para expresar los conceptos que pertenecen a un micro-tesauro, se usaría la propiedad **skos:member**. La pertenencia de los conceptos al tesoro se representaría con **skos:inScheme**. La definición de los conceptos cabecera se realizaría mediante las correspondientes propiedades **skos:topConceptOf** y **skos:hasTopConcept**.

Tomando como ejemplo el Tesoro de la Unesco:



Se podría realizar la representación del siguiente modo, expresado en sintaxis Turtle:

```

@Prefix ex: <http://www.example.org/>.
@Prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@Prefix iso-thes: <http://purl.org/iso25964/skos-thes#>.
@Prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

ex:tesauroUnesco rdfs:type skos:ConceptScheme.

ex:areaCiencia rdfs:type skosthes:ConceptGroup;
iso-thes:microThesaurusOf ex:tesauroUnesco
iso-thes:subGroup ex:microtresBiología.

ex:microtresBiología rdfs:type iso-thes:ConceptGroup;
iso-thes:superGroup ex:areaCiencia.
  
```

ISO 25964 permite la composición de conceptos complejos mediante la combinación de términos. La ontología ISO-THES reutiliza SKOS-XL para definir las clases y propiedades que permiten desarrollar esta funcionalidad. El siguiente ejemplo:



Se expresaría en Turtle como se indica a continuación:

```

@Prefix ex: <http://www.example.org/>.
@Prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@Prefix iso-thes: <http://purl.org/iso25964/skos-thes#>.
@Prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@Prefix skosxl: <http://www.w3.org/2008/05/skos-xl#>.

ex:carbon rdfs:type iso-thes:PreferredTerm;
skosxl:literalForm "Carbón"@es;
iso-thes:plusUFTerm ex:mineriaCarbon.

ex:mineria rdfs:type iso-thes:PreferredTerm;
skosxl:literalForm "Minería"@es;
iso-thes:plusUFTerm ex:mineriaCarbon.

ex:mineriaCarbon rdfs:type iso-thes:SplitPreferredTerm;
skosxl:literalForm "Minería del Carbón"@es;
iso-thes:plusUseTerm ex:carbon;
iso-thes:plusUseTerm ex:mineria.

```

Sin duda, ISO-THES complementa en gran medida a SKOS, pero ambos tienen ciertas limitaciones de representación, en especial cuando se aborda el modelado para el consumo de RDF por parte de aplicaciones basadas en el descubrimiento de datos a partir de las relaciones entre recursos.

Tabla 1. Comparativa de limitaciones de SKOS e ISO-THES.

Aspecto analizado	SKOS	ISO-THES
Representar micro-tesauros o subdivisiones del esquema de conceptos	NO	SÍ
Composicionalidad de etiquetas léxicas de un concepto	NO	SÍ
Enlace desde un concepto hacia la colección a la que pertenece	NO	NO
Definición explícita de jerarquías de colecciones	En un solo sentido	SÍ
Enlace desde el esquema de conceptos hacia las colecciones que contiene	NO	NO
Enlace desde una colección hacia el esquema de conceptos al que pertenece	De forma no estándar	SÍ
Definición explícita de conceptos como puntos de acceso en una colección	NO	NO

Fuente: elaboración propia.

En consecuencia, con ISO-THES continúa siendo imposible averiguar:

- Cuáles son los conceptos cabecera de un recurso de la clase **iso-thes:ConceptGroup**, puesto que la propiedad **skos:hasTopConcept** y **skos:topConceptOf** únicamente permite representar los términos cabecera de un esquema de conceptos, y en el caso de **iso-thes:ConceptGroup** estamos ante una subclase de **skos:Collection**.
- Cuál es el micro-tesauro al que está asociado un concepto. Es importante distinguir este hecho de la posibilidad de indicar los conceptos que contiene un micro-tesauro, para lo cual se utiliza **skos:member**.

Una posible solución a las anteriores limitaciones podría consistir en el uso del vocabulario UNESKOS (PASTOR-SÁNCHEZ, 2016). Dicho vocabulario incluye un conjunto de propiedades desarrolladas en un contexto para la representación del Tesauro de la UNESCO, pero que podrían aplicarse a cualquier proyecto que precise de ciertas relaciones inversas de las que carecen tanto SKOS como ISO-THES¹³. Dichas propiedades son las siguientes:

¹³ Namespace y vocabulario RDF: <http://purl.org/umu/uneskos>. Especificación: <http://skos.um.es/TR/uneskos>.

- **uneskos:contains**: vincula un esquema de conceptos con cualquier elemento de SKOS.
- **uneskos:hasMainConcept**: identifica desde una colección aquellos conceptos que son puntos de acceso a una colección.
- **uneskos:hasMicroThesaurus**: identifica desde un esquema de conceptos aquellos grupos de conceptos que contiene.
- **uneskos:mainConceptOf**: identifica desde un concepto aquellas colecciones de las que es un punto de acceso.
- **uneskos:memberOf**: desde un concepto identifica aquellas colecciones a las que pertenece.

Con el uso conjunto de SKOS, ISO-THES y UNESKOS, el modelo de representación de un tesoro de la Figura 9 quedaría del siguiente modo:

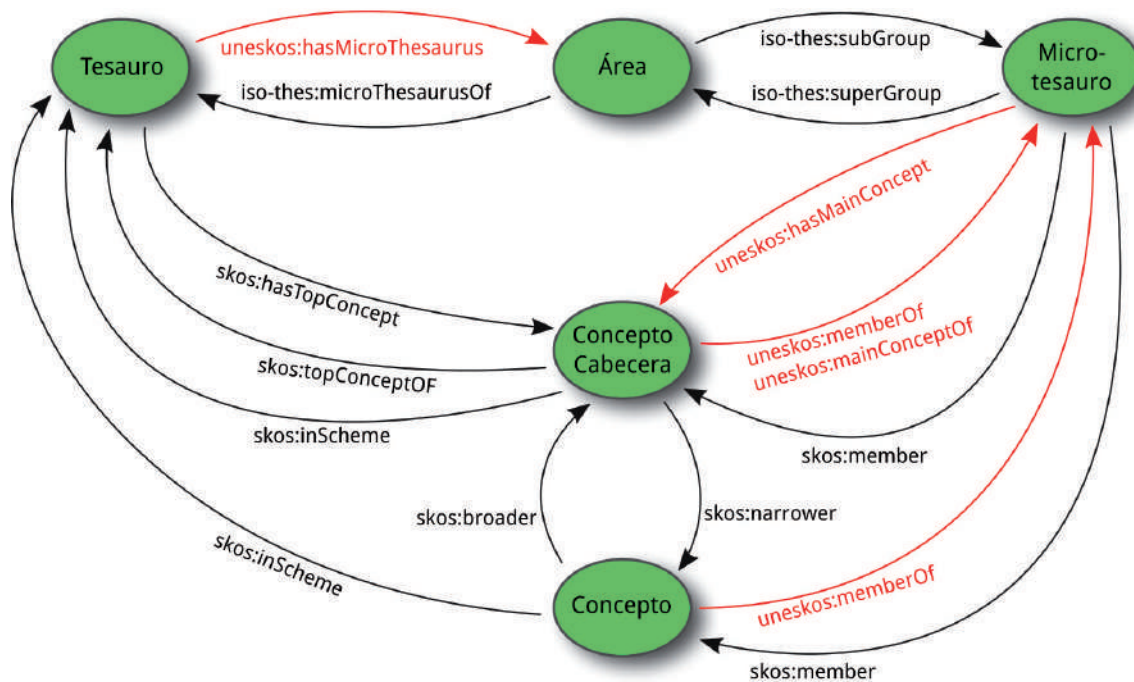


Figura 9. Modelo de representación de un tesoro estructurado en áreas y micro-tesauros mediante SKOS e ISO-THES, y complementado con el vocabulario UNESKOS. Fuente: elaboración propia.

6. Conclusiones

La nueva norma ISO 25964, en sus dos partes, plantea nuevos retos para la representación de estos instrumentos mediante tecnologías de la web semántica. Sin ninguna duda, la ontología ISO-THES, pese a encontrarse en una fase de desarrollo, posee suficientes elementos para modelar tesauros conforme a dicha norma. En este sentido, ISO-THES ofrece soluciones totalmente válidas para representar una tipología de tesauros muy amplia, en especial aquellos que se organizan en torno a micro-tesauros, facetas o áreas de conocimiento.

Sin embargo, SKOS presenta serias limitaciones para representar el mapeado compuesto entre conceptos de diferentes vocabularios. Estas limitaciones se unen a otras relacionadas con el consumo de datos mediante clientes RDF. La imposibilidad de definir las relaciones desde los conceptos hacia las colecciones (o grupos de conceptos) a los que pertenecen o los conceptos superiores en la estructura jerárquica de grupos de conceptos obligan a utilizar tecnologías,

como SPARQL, que en ocasiones puede que no estén a disposición de los editores de los conjuntos de datos de los tesauros.

Es posible definir vocabularios *ad hoc*, como UNESKOS, que permitan solventar estas limitaciones y que puedan utilizarse en diferentes contextos. No obstante, es posible que SKOS pueda ser revisado, como tantas otras tecnologías de la web semántica, para incluir algunas nuevas clases y propiedades que permitan modelar los sistemas de organización del conocimiento para facilitar su reutilización a partir de la estructura de conexiones entre elementos.

Finalmente, SKOS ofrece varios caminos para representar un mismo tesoro. Resultaría de gran utilidad trabajar sobre el establecimiento de un conjunto de buenas prácticas a partir del análisis de diferentes alternativas para modelar elementos concretos del tesoro.

Bibliografía

- BERNERS-LEE, T.; HENDLER, J., y LASSILA, O. (2001): «The semantic web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities», *Scientific American*, v. 284, n.º 5, pp. 28-37.
- GARCÍA-MARCO, F. J. (2014): «La nueva norma de tesauros UNE ISO 25964-1», *Anuario ThinkEPI*, v. 8, pp. 159-163.
- GUINARD, D.; TRIFA, V., y WILDE, E. (2010): «A resource oriented architecture for the Web of Things», *Proceedings of Internet of Things (Tokyo, 29 noviembre-1 diciembre, 2010)*. DOI: 10.1109/IOT.2010.5678452. <<https://vs.inf.ethz.ch/publ/papers/dguinard-things-2010.pdf>> [Consulta: 10 de diciembre de 2015].
- HEATH, T., y BIZER, C. (2010): *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool publishers, 2011. DOI: 10.2200/S00334ED1V01Y201102WBE001. <<http://linkeddata-book.com/book>>. [Consulta: 10 de diciembre de 2015].
- ISAAC, A., y DE SMEDT, J. (2015): ISO 25964 SKOS extension (iso-thes). <<http://pub.tenforce.com/schemas/iso25964/skos-thes/>>. [Consulta: 10 de diciembre de 2015].
- ISO (2011): ISO 25964-1:2011. *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Ginebra: International Organization for Standardization.
- (2013): ISO/DIS 25964-1:2013 *Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Ginebra: International Organization for Standardization.
- MONTALVO-MONTALVO, M. (2011): «LCSH, FAST y DELICIOUS: vocabularios normalizados y nuevas formas de catalogación temática», *Anales de Documentación*, v. 14, n.º 1. <<http://revistas.um.es/analesdoc/article/view/120141>>. [Consulta: 10 de diciembre de 2015].
- PASTOR-SÁNCHEZ, J. A. (2016): «Proposal to represent the UNESCO Thesaurus for the Semantic Web applying ISO-25964», *Brazilian Journal of Information Studies: Research Trends*, v. 10, n.º 1, pp. 1-8. <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/5684>>. [Consulta: 15 de enero de 2016].
- PASTOR-SÁNCHEZ, J. A.; MARTÍNEZ-MÉNDEZ, F. J., y RODRÍGUEZ-MUÑOZ, J. V. (2009): «Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. Information Research», v. 14, n.º 4, paper 422. <<http://InformationR.net/ir/14-4/paper422.html>>. [Consulta: 10 de diciembre de 2015].
- (2012): «Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data», *El profesional de la información*, v. 21, n.º 3, pp. 245-253.
- RAMOS SIMÓN, L. F., *et alii* (2012): «De la reutilización de la información del sector público a los portales de datos abiertos en Europa», *BiD: textos universitarios de biblioteconomía i documentació*, n.º 29. <<http://bid.ub.edu/29/pdf/ramos2.pdf>>. [Consulta: 10 de diciembre de 2015].

- SHETH, A. P. (1999): «Changing focus on interoperability in information systems: from system, syntax, structure to semantics», en *Interoperating geographic information systems*. Edición de Goodchild, M. F. *et alli*. Kluwer Academic Publishers, pp. 5-29. <<http://knoesis.wright.edu/library/download/S98-changing.pdf>>. [Consulta: 10 de diciembre de 2015].
- SOLODOVNIK, I. (2011): «Metadata issues in Digital Libraries: key concepts and perspectives», *JLIS.it Italian Journal of Archives and Information Science*, v. 2, n.º 2. DOI: 10.4403/jlis.it-4663. <<http://dialnet.unirioja.es/descarga/articulo/3945416.pdf>>. [Consulta: 10 de diciembre de 2015].
- W3C (2009): *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation 18 August 2009. <<http://www.w3.org/TR/skos-reference>>. [Consulta: 10 de diciembre de 2015].
- (2013): *RIF Primer (Second Edition)*. W3C Working Group Note 5 February 2013. <<http://www.w3.org/TR/rif-primer>> [Consulta: 10 de diciembre de 2015].
- (2013b): *SPARQL 1.1 Query Language*. W3C Recommendation 21 March 2013. <<http://www.w3.org/TR/sparql11-query/>>. [Consulta: 10 de diciembre de 2015].
- (2014): *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation 25 February 2014. <<http://www.w3.org/TR/rdf11-concepts/>>. [Consulta: 10 de diciembre de 2015].

El control terminológico en el Museo Arqueológico Nacional: evolución histórica, situación actual y perspectivas de futuro

Javier Rodrigo del Blanco

Museo Arqueológico Nacional

Existen muchas definiciones de documento, pero todas coinciden en que se trata de información contenida en un soporte físico o electrónico. Los sistemas documentales son tan antiguos como la propia escritura, pues esta surge por la necesidad de crear documentos que eviten la pérdida de información y hagan posible su posterior recuperación. Así, los primeros documentos escritos son asientos contables relacionados con el pago de impuestos, lo que supone una estructuración de la información por la necesidad de saber, al menos, quién, cuánto y por qué concepto se debe tributar ante una autoridad determinada. Y esta idea de organizar la información no es ajena a un concepto que creemos mucho más actual: la base de datos.

No comenzaremos esta evolución histórica desde un momento tan temprano de nuestra historia, sino que lo haremos desde el siglo XIX, cuando se formaliza la profesionalización del personal que presta sus servicios en archivos, bibliotecas y museos a partir de tres normas:

- 1.º Real Decreto de 7 de octubre de 1856 (*Gaceta de Madrid* de 9 de octubre), por el que se crea la Escuela Diplomática, con la función de enseñar los conocimientos necesarios para jefes y oficiales de archivos y bibliotecas.
- 2.º Real Decreto de 17 de julio de 1858 (*Gaceta de Madrid* de 18 de julio), por el que se crea el Cuerpo Facultativo de Archiveros-Bibliotecarios.
- 3.º Real Decreto de 12 de junio de 1867 (*Gaceta de Madrid* de 15 de junio), por el que se añade una Sección de Anticuarios al Cuerpo Facultativo de Archiveros-Bibliotecarios, que pasa a denominarse Cuerpo Facultativo de Bibliotecarios, Archiveros y Anticuarios, a cuyo cargo estarán las bibliotecas públicas, los archivos generales y los museos de antigüedades o arqueológicos.

Entre estos museos de antigüedades se encontraba el Museo Arqueológico Nacional (MAN), que se había creado por Real Decreto de 20 de marzo de 1867 (*Gaceta de Madrid* de 21 de marzo) y cuyas colecciones iniciales fueron los objetos arqueológicos o numismáticos exis-

tentes en la Biblioteca Nacional, el Museo de Ciencias Naturales y la Escuela Diplomática. Este real decreto también preveía la creación de museos arqueológicos en aquellas provincias que contaran con numerosos e importantes objetos de esta naturaleza.

Es relevante que en todas estas normas se haga referencia exclusiva a museos con objetos arqueológicos, definidos como «los pertenecientes a la Antigüedad, a los tiempos medios y al Renacimiento, que sirvan para esclarecer el estudio de la historia, del arte o de la industria en las indicadas épocas» (Real Decreto de 20 de marzo de 1867, artículo 2). Y lo es porque todas las normas relacionadas con la Sección de Anticuarios no afectarán a museos de bellas artes, de antropología o de ninguna otra especialidad distinta de la arqueológica. Es más, la profesionalización del personal al servicio de museos no arqueológicos va a ser un hecho relativamente reciente, salvo en lo que hoy denominamos departamentos de investigación, lo que explica la distinta importancia que se otorga a cuestiones documentales dentro de los distintos museos.

A lo largo de la segunda mitad del siglo XIX, y comenzando desde el propio Real Decreto de 12 de junio de 1867, se van a ir dictando instrucciones para regular la actividad del Cuerpo Facultativo de Bibliotecarios, Archiveros y Anticuarios¹. Sin embargo, estos reglamentos no van a aportar nada en el campo que nos ocupa, pues en todos ellos se habla de una necesaria normalización para homogeneizar los instrumentos documentales existentes en cada centro (catálogos, índices e inventarios) y hacer que todos los de una misma clase (archivos, bibliotecas y museos arqueológicos) tuvieran los mismos². Reiterativa es también la llamada a la labor que en este sentido debía hacer la Junta Consultiva de Archivos, Bibliotecas y Museos, tarea que nunca vio la luz³. La consecuencia evidente es que cada centro adoptó los instrumentos documentales que estimó convenientes, organizando en ellos la información de acuerdo con las necesidades de cada establecimiento.

Por Real Orden de 12 de septiembre de 1901 (*Gaceta de Madrid* de 13 de septiembre) se dispone que la Junta Facultativa de Archivos, Bibliotecas y Museos redacte reglamentos independientes para cada Sección, así como que dicte instrucciones para la redacción de los catálogos que deba haber en cada clase de establecimiento, tanto generales como especiales. Esta mayor especialización se plasma en:

- 1.º Real Decreto de 22 de noviembre de 1901 (*Gaceta de Madrid* de 26 de noviembre), por el que se aprueba el reglamento para el régimen y gobierno de los Archivos del Estado.
- 2.º Real Decreto de 29 de noviembre de 1901 (*Gaceta de Madrid* de 3 de diciembre), por el que se aprueba el reglamento para el régimen de los Museos Arqueológicos del Estado.
- 3.º Real Orden de 31 de julio de 1902 (*Gaceta de Madrid* del 5 al 19 de agosto), por la que se aprueban las instrucciones para la redacción de los catálogos en las Bibliotecas Públicas del Estado.

De estas tres disposiciones, las dos más completas fueron las de archivos y bibliotecas, pues no solo especificaban mejor la denominación y finalidad de los distintos instrumentos documentales, sino que además proporcionaban modelos de los mismos. En el ámbito bibliotecario, estas instrucciones fueron el primer intento serio para normalizar la catalogación de fondos bibliográficos en España y tuvieron una larga vida, pues se mantuvieron en vigor hasta el año 1985, cuando el Ministerio de Cultura publicó la primera parte de las «Reglas de catalogación», adaptación nacional de las normas ISBD (International Standard Bibliographic Description). Y la

¹ Se trata de los reglamentos orgánicos, que incluyen cuestiones de naturaleza técnica y administrativa y que se van a ir sucediendo en el tiempo: 1871, 1881, 1884, 1885 y 1887.

² En todos estos reglamentos se habla de inventarios y de índices circunstanciados de los fondos de cada institución, así como de libros de entrada y salida de material científico, libros que estaban a cargo de los secretarios de cada establecimiento.

³ Esta Junta quedará disuelta por Real Decreto de 16 de abril de 1875 (*Gaceta de Madrid* de 17 de abril), siendo sustituida por la Junta Facultativa de Archivos, Bibliotecas y Museos. Con este cambio, se quería dotar a este órgano de un carácter más ejecutivo, si bien no tuvo efectos prácticos en el terreno documental.

normalización de contenidos también había recibido mayor atención en este ámbito que en el de los archivos y museos⁴.

En cambio, el reglamento para museos arqueológicos se limitaba a enumerar los instrumentos documentales (libro de entrada de objetos por compra o donación, libro de entrada de objetos en depósito, inventario general, catálogo sistemático y catálogo monográfico) y, en el mejor de los casos, dar unas indicaciones genéricas acerca de su contenido y finalidad, pero sin ofrecer ningún modelo que pudieran seguir los distintos museos. Por ello, los cambios que esta norma supuso en la documentación de estos centros fueron inapreciables.

Habrá que esperar a 1942 (75 años después de la creación del MAN y de la Sección de Anticuarios) para tener una definición exacta de cómo debían ser los instrumentos documentales enumerados en el reglamento de 1901. Se aprobaron por Orden del Ministerio de Educación Nacional, de 16 de mayo de 1942 (BOE de 6 de junio) y serán conocidas como Instrucciones de Navascués⁵, en honor a su redactor. Sobre su importancia, simplemente decir que más del 90% de los campos contenidos en los sistemas documentales actuales ya están recogidos en estas Instrucciones.

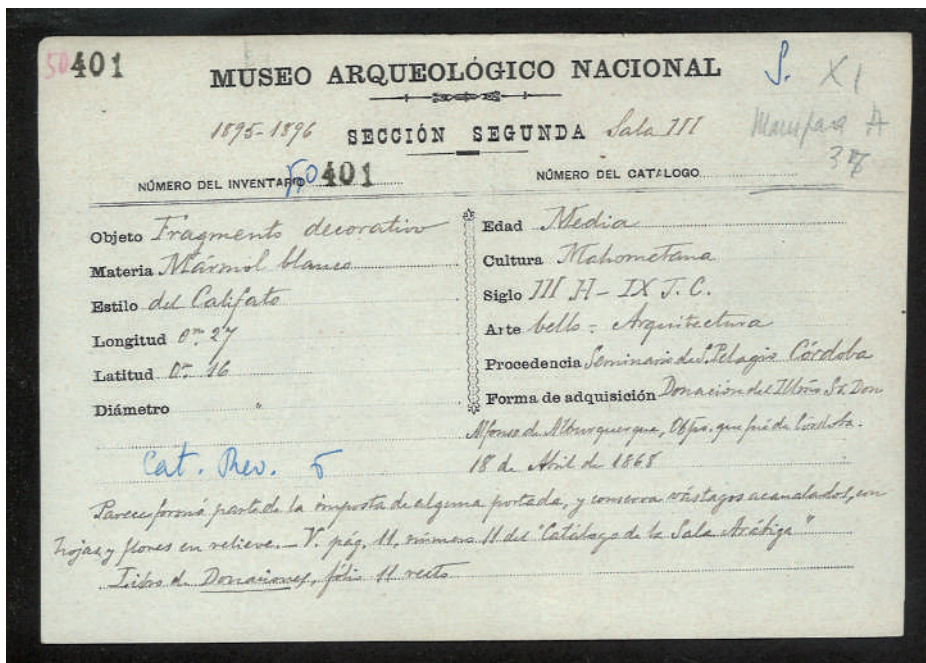
Llegados a este punto, es necesario preguntarse cómo documentaban los museos sus colecciones antes de 1942, que es cuando reciben instrucciones claras y concretas acerca del modo de proceder en este ámbito por primera vez. En el MAN, esa documentación va a estar estrechamente relacionada con la organización de los fondos, ya que se establecía una primera división de los mismos en función de su naturaleza y cronología, que explican las cuatro Secciones en que se dividía el MAN:

- Sección 1.^a: objetos arqueológicos desde la Prehistoria a la Edad Antigua.
- Sección 2.^a: objetos arqueológicos y artísticos de la Edad Media y de la Edad Moderna.
- Sección 3.^a: objetos numismáticos, con independencia de su cronología.
- Sección 4.^a: objetos etnográficos, también con independencia de su cronología.

A partir de esta primera clasificación, todos los objetos arqueológicos y artísticos van a organizarse por su edad (Edad de Piedra, Edad del Bronce, Edad Media...); por su industria, cultura o estilo artístico (magdalenense, hispano-romana, visigoda, gótico, barroco...); por su yacimiento de procedencia (solo para materiales arqueológicos, no para los artísticos); por cuestiones materiales (piedra, metal...), técnicas (textiles), funcionales (iluminación, recipientes, caza, minería...) o tipos de producción (cerámica, cestería...); y por su tipología. Los objetos numismáticos se ordenarán principalmente por emisores, incluyendo la clasificación por periodo histórico y sustituyendo el yacimiento de procedencia por la ceca productora; en tanto que los etnográficos se agruparán por fuente de ingreso y por materiales o lugares de procedencia

⁴ Clasificar significa describir contenidos en el ámbito bibliotecario. Las clasificaciones universales o enciclopédicas tratan de abarcar todo el conocimiento humano, sistematizándolo en entradas concretas. Entre estas clasificaciones hay que destacar la Clasificación Decimal de Dewey (1876), la expansiva de Cutter (1891), la de la Library of Congress (1897, aunque su plan general no se publicó hasta 1942), la Clasificación Decimal Universal (1905) y la bibliográfica de Bliss (1935).

⁵ Su denominación oficial era «Instrucciones para la formación y redacción del inventario general, de los catálogos y registros en los museos servidos por el Cuerpo Facultativo de Archiveros, Bibliotecarios y Arqueólogos dependientes del Ministerio».



Ficha de inventario anterior a 1942. Museo Arqueológico Nacional (N.I. FA50401).


Esta ordenación obedece a cuestiones prácticas de investigación y de conservación, pues se diferencian los materiales orgánicos de los inorgánicos, cada uno con sus propias necesidades. Y el hecho de tener reunidos unos objetos con unas características comunes en cuanto a tipología, técnica, materia y cronología facilita su comparación y estudio. Son estos criterios los que van a seguir las Instrucciones de Navascués a la hora de clasificar los fondos en el catálogo sistemático, cuyo fin era «conocer los objetos del Museo en relación con el cuadro artístico, arqueológico o histórico del territorio o comarca a que alcanza el área de acción del Museo» (instrucción 39). Dado que ese orden científico no se podía establecer para todos los museos, Navascués ofrece ejemplos de clasificación para fondos arqueológicos y numismáticos, estableciendo los niveles recogidos en la siguiente tabla, del más general al más específico⁶:

Niveles	Fondos arqueológicos	Fondos numismáticos
1	Edad	Edad
2	Cultura	Serie
3	Sección	Grupo
4	Serie	Soberano, Estado o ciudad

Los términos recogidos bajo estos conceptos son los que componen dos de los campos más complejos a la hora de abordar su normalización: clasificación genérica y contexto cultural. Sin embargo, el control terminológico iba más allá, pues se establece el empleo del vocablo en castellano para denominar los objetos seguidos, en su caso, del nombre erudito entre paréntesis (instrucción 12)⁷.

⁶ Para la definición de cada término, ver la instrucción 40 de Navascués.

⁷ Antecedentes directos de los actuales campos «Nombre común» y «Nombre específico» del módulo «Fondos museográficos» de la aplicación DOMUS.

MUSEO ARQUEOLOGICO NACIONAL		CATALOGO SISTEMATICO	
R.º de E.º en propiedad N.º	Edad <u>MEDIA</u>	Fecha <u>s IX-X</u>	
deposito	Cultura <u>ANDALUS</u>		
INVENTARIO GRAL. N.º <u>50.401</u>	Sección <u>medallas</u>		
CAT.º MONOGRAF. N.º	Serie <u>RELIEVES</u>	N.º	
Mod. n.º 4			
	Objeto <u>Relieve (Frag.)</u>		
	Materia <u>Caliza</u>		
	Dimensiones <u>Alt. = 16 ; Anch. = 26'5</u>		
	<u>Gm. = 7'5 cm.</u>		
	Peso		
	Conservación <u>R</u>		
Negativos <u>6200/6 ; Dep. 12. 178</u>			
Escuela, Taller, Autor, Marcas o firmas			
Asunto <u>Motivo vegetal</u>			
Procedencia <u>Seminario de San Pelagio, Córdoba</u>			
Descrito en	<u>REVILLA VIELVA, R., (1932), El Patio árabe del MAH...</u>		
Variante de			
Tomo	Pág. <u>15</u>	Lám.	N.º <u>6</u>
Observaciones <u>Donación de Alfonso de Albuquerque</u>			
<u>exp. 1868/103A-V. Libro de donaciones, fol. 11</u>			
<u>Catálogo de la Sala árabe nº 11 de mayo 18/04/1868</u>			
<u>Topografía antigua Pl. XI Hoja A-37</u>			
Referencia Topográfica <u>Comp. 42 D-2</u>			

Anverso de la ficha del catálogo sistemático. Museo Arqueológico Nacional.

Aplicadas en un principio a museos arqueológicos, las Instrucciones de Navascués van a ser la norma documental de museos de otras especialidades a medida que estos vayan pasando a depender de la Dirección General de Bellas Artes, en cuyo seno se crea el organismo autónomo Patronato Nacional de Museos por Decreto 2764/1967, de 27 de noviembre (BOE de 28 de noviembre). Eran lógicas las críticas de estos centros hacia estas Instrucciones por el carácter excesivamente arqueológico de las mismas, pero también es cierto que se redactaron pensando en los museos arqueológicos, que eran los únicos servidos por miembros del Cuerpo Facultativo de Archiveros, Bibliotecarios y Arqueólogos en el momento de su redacción. Por ello, estas Instrucciones tuvieron un desigual cumplimiento en museos artísticos o etnográficos, que muchas veces siguieron utilizando sus propios instrumentos documentales.

La década de 1980 fue importante porque asistimos a los comienzos de la informatización en los museos, que afectará enormemente al ámbito documental. No obstante, antes de introducir la informática ya se estaban planteando cuestiones acerca de la normalización terminológica, como las listas de términos más o menos jerarquizados para el contenido de algunos campos y la elaboración de unas fichas con términos aceptados para nombres de objetos (Porta *et alii*, 1982). Y las Instrucciones de Navascués también habían adelantado algo al respecto, como al tratar la referencia a las fotografías (instrucción 11) o la elaboración de un índice de grabadores como complemento al catálogo sistemático de medallas (instrucción 58).

Otro proyecto que planteó cuestiones sobre normalización terminológica fue el denominado «Documentación Etnográfica Española», iniciado en 1983 y concebido como un amplio plan de actuación en el campo etnográfico, que, por diversas causas, quedó reducido «al estableci-

miento de unas normas generales para la recogida y sistematización de documentación de interés para la realización de futuros estudios etnográficos en España» (Carretero, Quinto y Timón, 1993: 9). Entre sus aportaciones, de gran interés por tratarse de una visión de la documentación desde un punto de vista distinto del arqueológico, estuvo la de tratar de establecer unos tesauros e índices clasificatorios comunes para catalogar todos los fondos: objetos, publicaciones, fotografías, filmaciones, etc.

El Reglamento de Museos de titularidad estatal, aprobado por Real Decreto 620/1987, de 10 de abril (BOE de 13 de mayo), introdujo cambios mínimos con respecto a las Instrucciones de Navascués, si bien volvía a una práctica que ya vimos en el siglo XIX: la remisión a futuras normas técnicas para la elaboración de inventarios y catálogos (art. 13). Las modificaciones se centraban más en cuestiones administrativas, por el nuevo reparto competencial en materia de cultura, que en aspectos técnicos⁸. Lo que sí introdujeron estos libros de registro fue una cierta confusión, pues se indicó que el campo «Objeto» se cumplimentara de acuerdo a la naturaleza del mismo (pintura, escultura, mosaico, arca, etc.), mezclando clasificaciones genéricas con nombres comunes, conceptos que antes aparecían bien diferenciados⁹.

El gran cambio en los sistemas documentales llegará con las bases de datos informáticas, que unificarán los distintos repertorios documentales en una única aplicación, lo que incrementará su volumen. Esta informatización agilizará la introducción de datos, su edición y la recuperación de registros, tanto en formatos de salida impresos como en pantalla, pero obligará a una mayor exactitud en las cadenas de búsqueda para ofrecer los resultados esperados. Esto supone un cambio radical en el modo de realizar las consultas, pues antes se iba a un repertorio documental concreto y se buscaban registros físicos. Con las bases de datos informáticas, se pueden hacer búsquedas por uno o más campos de manera simultánea, pero el término introducido debe coincidir con el contenido en los distintos registros de esa base de datos, pues, de lo contrario, los registros que no lo contengan de la misma manera no aparecerán entre los resultados de nuestra consulta. La normalización terminológica pasaba a ser una necesidad.

Los primeros sistemas documentales informatizados aparecieron a finales de la década de 1980 en el Museo Arqueológico Nacional (Cacho *et alii*, 1995) y en el Museo del Pueblo Español (Carretero, 1990). En ambos casos, la normalización terminológica estaba presente en unos campos muy concretos y sus descriptores eran fiel reflejo de sus sistemas anteriores: el sistema del Museo del Pueblo Español aprovechaba el trabajo realizado con motivo del proyecto «Documentación Etnográfica Española», en tanto que las diversas aplicaciones instaladas en el MAN eran herencia directa de las Instrucciones de Navascués, con el añadido administrativo impuesto tras el Reglamento de Museos de titularidad estatal de 1987. En estos sistemas, especialmente en los del MAN, se diferencian completamente los términos que responden a la clasificación genérica de los fondos y a su contexto cultural, que aparecían entrelazados en el catálogo sistemático de Navascués y que se configuran ahora como dos campos distintos.

Consciente de los problemas derivados de la adopción de una aplicación informática distinta para cada museo, la Dirección General de Bellas Artes y Conservación y Restauración de Bienes Culturales reunió a una serie de técnicos con la triple misión de 1) definir y unificar los procesos documentales museográficos y administrativos, 2) la normalización de terminologías, y 3) el desarrollo de su aplicación informática. Este grupo, denominado Comisión de Normalización Documental de Museos, se constituyó oficialmente el 25 de octubre de 1994, aunque venía reuniéndose desde 1993. Los resultados de su trabajo quedaron plasmados en una publicación

⁸ El libro de registro de entradas en propiedad quedaba como estaba, pero el libro de registro de depósitos debía desglosarse en dos: uno para bienes depositados por otros órganos e instituciones de la Administración General del Estado y otro para bienes depositados por otras personas físicas y jurídicas, tanto públicas (comunidades autónomas y corporaciones locales) como privadas.

⁹ Ver las instrucciones dictadas al respecto por la Dirección General de Bellas Artes y Archivos del Ministerio de Cultura y comunicadas a los museos en marzo de 1988 (Caballero, 1988).

(Carretero *et alii*, 1996) y en la aplicación DOMUS para la gestión documental integral en los museos dependientes de dicha Dirección General.

De los trabajos de esta Comisión nos interesan ahora los dos últimos puntos: la normalización terminológica y su plasmación en un sistema documental informático. Esta Comisión tomó como punto de partida las experiencias que ya se habían hecho realidad en el MAN y en el Museo Nacional del Pueblo Español. A partir de esas experiencias, se configuraron tres herramientas principales para el control terminológico:

- 1.º Tesoros. Listados de descriptores y de indicadores clasificatorios, con distintas relaciones entre ellos: jerárquica, asociación y sustitución. Empleados en campos con contenidos complejos, como los de clasificación genérica, contexto cultural o nombres de objeto.
- 2.º Tablas. Listados de descriptores con una relación jerárquica rígida entre ellos. Se emplean en estructuras que difícilmente cambian, como las que indican ubicaciones.
- 3.º Listas. Relación de descriptores ordenados alfabéticamente. Estas listas pueden ser de dos tipos: abiertas o cerradas, en función del grado de libertad que se quiera dar al usuario a la hora de introducir términos. Serán listas abiertas aquellas que sean difíciles de sistematizar, como los títulos de las obras o sus lugares específicos de procedencia, pero serán cerradas aquellas para las que puedan definirse unas entradas determinadas, como las formas de ingreso o los departamentos en que se estructura cada museo.

Hay un sistema adicional de control terminológico y es el de aquellos campos que toman sus contenidos de campos existentes en otros módulos del sistema. Así, por ejemplo, se establecen relaciones entre los campos «Inventario» o «Expediente», que están contenidos en distintos módulos: «Archivo administrativo», «Documentación gráfica», «Fondos museográficos»...

En cuanto a la normalización terminológica, se planteaba la creación de un Área de Vocabularios Técnicos en el seno de la Comisión para «la unificación de normas descriptivas y terminológicas de las diversas áreas científicas, técnicas y administrativas» (Carretero *et alii*, 1996: 5-6). En esta publicación aparecen recogidas ya muchas de las tablas y listas cerradas de la aplicación DOMUS, pero se remite a los resultados de grupos de trabajo específicos que ya estaban tratando este asunto sobre distintas materias. El trabajo de estos grupos ha ido viendo la luz bajo la denominación de «Diccionarios DOMUS»¹⁰, en los que distintos especialistas abordan un campo concreto.

No obstante, y salvo el de Kroustallis, estos «Diccionarios DOMUS» no pueden tomarse sin más como herramientas para control terminológico, pues cada uno de ellos aborda su ámbito de estudio sin tener en cuenta la distribución de los términos que contienen en distintos campos del sistema. Por ello, es imprescindible un proceso de depuración que identifique un significado con un único significante¹¹ y que ajuste los descriptores a los campos que realmente deban contener esa información¹².

El MAN inició la revisión de diversas herramientas de control terminológico de DOMUS en 2010, pues los descriptores contenidos en ellas no son solamente fruto de la herencia recibida del pasado, sino también de migraciones informáticas desde otros sistemas, que han introducido no pocos errores de volcado. La primera actuación en este sentido consistió en la actualización de todo lo relacionado con la gestión de reproducciones, que implicaba a distintos instrumentos:

- 1.º Tesoro «Series documentales», incorporando entradas adaptadas a la normativa vigente en materia de procedimiento administrativo, ya que los expedientes se agruparían por la

¹⁰ Se han publicado hasta el momento los siguientes: dibujo y estampa (Blas, 1996), cerámica (Padilla, Maicas y Cabrera, 2002), mobiliario (Rodríguez, 2006), materias (Kroustallis, 2008), numismática (Alfaro *et alii*, 2009), objetos asociados a ritos, cultos y creencias (Trinidad, 2011) y objetos asociados a la expresión artística (Trinidad, 2012).

¹¹ «Recipientes» en el diccionario de cerámica, «Mobiliario contenedor» en el de mobiliario y «Elementos para almacenamiento numismático» en el de numismática son tres descriptores que hacen referencia a un mismo significado: el de un objeto cuya finalidad es guardar o almacenar otros objetos o sustancias.

¹² Trinidad (2011) incluye nombres comunes, nombres específicos y tipologías en un vocabulario destinado a nombres de objetos, cuando estos elementos aparecen en tres campos distintos y tienen sus tesauros diferenciados.

forma de finalización de dichos procedimientos: caducidad, desistimiento o renuncia y resolución favorable o denegada. Además, se añadía otra serie para recoger los pedidos que el MAN hacía a otras instituciones.

- 2.º Listas cerradas¹³ «Estado de reproducción», «Estado de tramitación», «Finalidad de reproducción», «Formatos», «Tarifas de reproducción» y «Tipos de reproducción», adaptándose a la normativa vigente sobre propiedad intelectual y precios públicos.

La primera gran depuración se hizo con la lista cerrada «Formas de ingreso», que contenía entradas distintas con el mismo significado. Tomando como base los modos de adquisición recogidos por la Comisión de Normalización Documental (Carretero *et alii*, 1998: 17-23), se asignó una entrada única a cada forma de ingreso, eliminando duplicados¹⁴; se mantuvieron algunas entradas, que habrá que tratar con mayor detenimiento para evitar la pérdida de información¹⁵; y se crearon otras nuevas para responder a realidades existentes hoy día, como los ingresos de fotografías como contraprestación por haber permitido su toma en el Museo a fotógrafos profesionales ajenos al mismo.

La segunda gran depuración afectó a instrumentos documentales vinculados con el módulo «Fondos bibliográficos», en especial las listas abiertas «Bibliografías autores» y «Bibliografías títulos». Con respecto a los autores, había una pluralidad de entradas (hasta cinco) que hacían referencia a una misma persona, aunque variaban elementos como utilizar acentos o no, poner uno o sus dos apellidos, el nombre completo o su abreviatura, separar apellidos y nombre mediante coma o no, o finalizar la entrada con un punto o sin él. Se dieron de alta 84 nuevos autores y se editaron 73 entradas, que quedaron reducidas a 31¹⁶.

Con los títulos ocurría algo similar a lo descrito para los autores: distintas entradas para un mismo título por diferencias en el modo de transcribirlo. Sin embargo, en este caso era más evidente el peso de una migración realizada de manera incorrecta, pues una gran cantidad de títulos llevan incorporadas unas comillas [“x”], herencia de sus bases de datos de origen, que no fueron eliminadas antes de dicha migración. El problema aquí es que DOMUS está configurado para incluir el contenido de este campo entre comillas, por lo que dicho contenido aparece con dobles comillas [““x””] en los formatos de salida de datos: unas corresponden a las del sistema y otras a las del contenido del campo. Se han ido eliminando estas comillas a lo largo del tiempo, pero quedan aún cerca de 2000 títulos con ellas, de los 4489 existentes actualmente.

Desde entonces, se ha venido trabajando sobre tesauros concretos, a fin de contar con la terminología necesaria para el registro, inventario y catalogación de los fondos, tanto documentales como museográficos. En este sentido, Fernando Fontes ha trabajado sobre el módulo «Directorio» y el tesoro «Autoridades» para, al igual que se hizo con los autores de bibliografía, normalizar nombres de personas y de instituciones y evitar duplicados y registros incompletos, unificando toda la información disponible sobre una persona en un registro. También a Fernando Fontes se debe la agrupación de descriptores del tesoro «Lugares geográficos» bajo unos indicadores clasificatorios concretos, enorme labor que facilitará la posterior depuración de este instrumento.

¹³ Los usuarios no podemos eliminar algunas entradas de las listas cerradas, con independencia del nivel de permisos. No parece lógico que solo los informáticos que mantienen DOMUS tengan esta facultad.

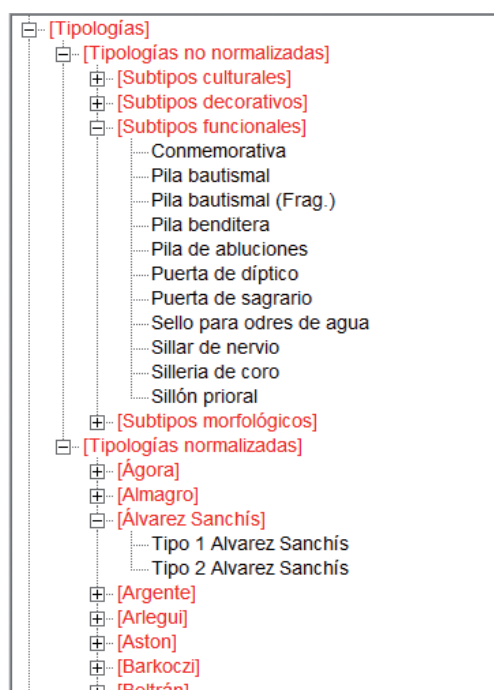
¹⁴ Por ejemplo, las entradas «Adquisición por compra», «Asignación por compra» y «Compra» quedaron unificadas en un único descriptor, eliminándose los dos sobrantes.

¹⁵ «Donación por excavación anterior a 1985» nos indica que un objeto ha sido donado, pero es importante no perder el dato de procedencia de ese objeto: una excavación científica. El motivo es que hasta la Ley de 13 de mayo de 1933, relativa al Patrimonio Artístico Nacional (Gaceta de Madrid de 25 de mayo), se permitía que los objetos procedentes de hallazgos fortuitos o de excavaciones arqueológicas legales quedaran en poder de sus descubridores (art. 40). En cambio, la Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español (BOE de 29 de junio), impone que todo objeto arqueológico debe ser entregado a la Administración, con independencia de si fue descubierto en el transcurso de excavaciones o prospecciones arqueológicas autorizadas (art. 42) o de un hallazgo fortuito (art. 44).

¹⁶ Este trabajo se hizo en el año 2011, en el marco de una campaña de revisión de datos de fondos museográficos a partir de catálogos de 96 exposiciones. La campaña fue coordinada por quien suscribe y en ella participaron cinco licenciados, contratados por el Servicio Público de Empleo Estatal.

Se ha hecho una primera revisión del tesaurus «Tipología», que alimenta el campo «Tipología / Estado»¹⁷. Este campo debería desglosarse en dos, tipología y estado, pues hace referencia a dos cuestiones totalmente distintas. Hasta ese momento, se agrupan los términos propuestos bajo dos indicadores clasificatorios bien diferenciados dentro de este tesaurus. El que hace referencia al estado, a la condición de la obra, no ofrece mayores problemas; pero el de tipología obliga a plantear una división entre:

- Tipologías normalizadas. Clasificaciones sistemáticas de un objeto o de un grupo de objetos, normalmente expresadas con notaciones codificadas, realizadas por investigadores y publicadas. Se aplica también a denominaciones codificadas de productos comerciales. Este indicador respondería a lo que la Comisión de Normalización Documental citaba como «clasificación científica» y «denominación comercial».
- Tipologías no normalizadas. Denominaciones de tipos, generalmente formales, decorativos o funcionales, que se han impuesto por su uso a lo largo del tiempo pero que no han sido objeto de normalización, clasificación y publicación por parte de ningún especialista, pese a que su uso está extendido y aceptado por la mayoría de los investigadores. Este indicador respondería a lo que la Comisión de Normalización Documental de Museos definía como «subcategoría funcional o formal» y «formas adjetivadas u otros elementos de precisión que permiten precisar el nombre o las características del objeto dentro de su grupo».



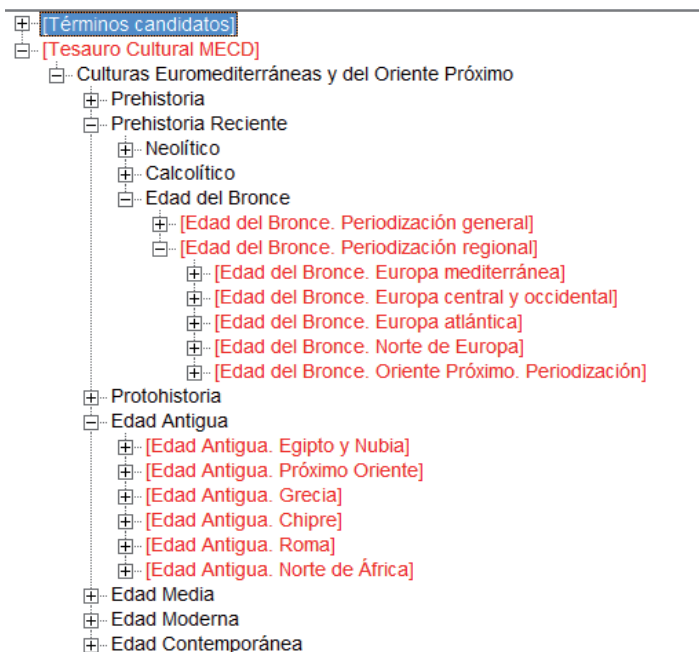
Ejemplo para tipologías normalizadas y no normalizadas. Museo Arqueológico Nacional.

Y llegamos a los dos grandes tesaurus sobre los que se ha trabajado en el MAN y para los que ya la propia Comisión avanzaba que sería necesaria la creación de grupos de trabajo específicos: «Clasificación genérica» y «Contexto cultural» (Carretero *et alii*, 1996: 6). Este último fue facilitado por la Subdirección General de Museos Estatales para su revisión y adaptación por parte del MAN,

¹⁷ Este campo recoge la «clasificación científica, subcategoría funcional o formal, condición de la obra (copia, impronta, etc.), denominación complementaria o comercial, formas adjetivadas u otros elementos de precisión que permiten precisar el nombre o las características del objeto dentro de su grupo» (Carretero *et alii*, 1996: 73).

tarea que coordinó Fernando Fontes con los departamentos de investigación de este Museo¹⁸. Este tesoro se encuentra en una fase muy incipiente, pues, pese a la enorme diversidad de nuestras colecciones, será imprescindible la aportación de técnicos de museos de otras especialidades para completarlo. En su nivel más general, este tesoro debería contar con las siguientes entradas:

- Áreas culturales. Definidas como ámbitos geográficos en los que se desarrollan culturas con unas características similares, son fundamentales a la hora de contextualizar objetos etnográficos, aunque su uso puede extenderse a otro tipo de objetos, como los arqueológicos de época prerromana en el caso de nuestro país.
- Creencias. La importancia que tienen los distintos sistemas de creencias en los grupos humanos hace que muchos objetos hayan sido fabricados para su empleo dentro de un sistema concreto, pudiendo ser específicos de determinados ritos. El carácter ritual de un objeto se define en su clasificación genérica, pero aquí se recoge a qué sistema pertenece y con qué ritos se asocia dicho objeto.
- Cronología geológica. Complementaria de la histórica en algunas fases, como en época prehistórica, es básica a la hora de tratar restos minerales, paleobotánicos y paleontológicos.
- Cronología histórica. Convencionalismos en los que dividimos la historia de la humanidad, desde sus inicios hasta nuestros días. Estas divisiones serán válidas para ámbitos geográficos concretos (América prehispánica, Europa, Extremo Oriente...), pues se deben respetar las particularidades que presenta cada uno de ellos.
- Estilos artísticos. Objetos vinculados a determinados periodos, estilos y movimientos artísticos se definen por sus características culturales en esta entrada.
- Grupos étnicos. Al igual que las áreas culturales, la determinación del grupo étnico es vital a la hora de contextualizar un objeto etnográfico, especialmente de aquellos de procedencia extraeuropea, aunque también puede aplicarse a pueblos prerromanos y germánicos.



Esquema general del tesoro de contexto cultural. Museo Arqueológico Nacional.

¹⁸ El campo «Contexto cultural» debe contener «periodo geológico o histórico, grupo cultural, estilos, u otros hechos o indicadores históricos o culturales con los que se vincula el objeto» (Carretero *et alii*, 1996: 87).

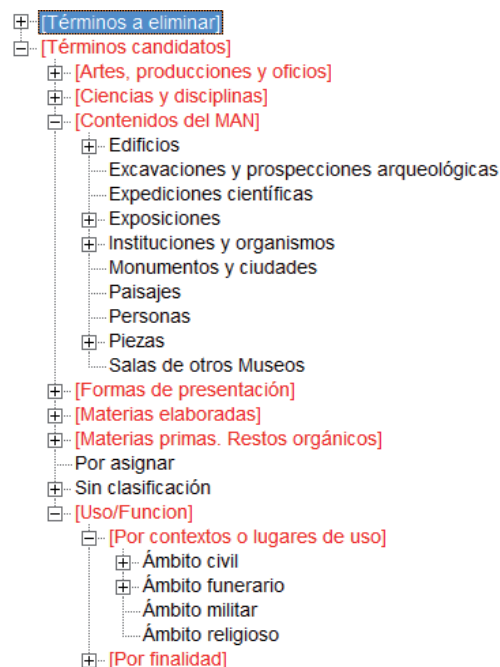
El trabajo realizado sobre este tesoro en el MAN se ha centrado fundamentalmente en la cronología histórica del ámbito denominado «Culturas Euromediterráneas y del Oriente Próximo». Esta denominación ya plantea algunos interrogantes, pues las culturas de Europa septentrional, atlántica, central u oriental no pueden depender jerárquicamente del término «euromediterráneo», ámbito con el que poco tienen que ver. Además, si queremos englobar las culturas mediterráneas, nos sobra el prefijo «euro» o nos falta el referido a África, pues Asia quedaría incluida en «Oriente Próximo». No obstante, este mismo término es también conflictivo, pues expresa conceptos distintos en nuestro país y en el mundo anglosajón.

El segundo problema fue la denominación de fases o culturas, ya que existen distintas propuestas y no hay unanimidad entre los especialistas a la hora de decidirse por una en concreto en detrimento de las demás. Por ejemplo, Bronce A, Bronce I o Bronce Inicial.

El tercer problema consistía en diferenciar unos periodos comunes con características regionales distintas, como, por ejemplo, Edad del Bronce o Edad del Hierro. La solución vino de la mano de indicadores clasificatorios que permitieran distinguir entre una periodización general, para recuperar todos los fondos de un periodo concreto, y una periodización regional, de la que ya pudieran depender jerárquicamente los descriptores referidos a cada fase o cultura concreta (Tercera dinastía de Ur, Hallstatt, Ibérica...). Quizá el indicador clasificatorio «Áreas culturales» podría contribuir a mejorar la organización de los descriptores en este sentido.

Y la cuarta cuestión era decidir el nivel al que se quería descender (jefaturas de Estado) y cómo incluir otros acontecimientos de enorme relevancia histórica, como la Revolución francesa o las guerras carlistas, que harían más completo el tesoro, pero también mucho más complejo.

Con respecto al tesoro «Clasificación genérica», conviene decir que presentaba la misma falta de relaciones entre sus términos que los demás, fruto de las migraciones masivas y del trabajo de personal eventual con escasa supervisión en esta parte de sus tareas, pues introducían términos para poder utilizarlos, no para crear un tesoro estructurado¹⁹. Analizados los términos del tesoro, estos se agruparon bajo las siguientes entradas:



Esquema general del tesoro de clasificación genérica. Museo Arqueológico Nacional.

¹⁹ El campo «Clasificación genérica» está destinado a contener información sobre categorías superiores al nombre del objeto o documento (Carretero *et alii*, 1996: 72).

- Artes, producciones y oficios. Clasificación empleada para artes (arquitectura, escultura, pintura...), producciones (eboraria, glíptica, joyería...) y oficios (alfarería, cantería...), incluyendo producciones muy específicas de una época, como «Arte mueble» o «Industria lítica», que serían casi imposibles de agrupar sin estos descriptores.
- Ciencias y disciplinas. Agrupación de bienes culturales en función del tipo de disciplina que se encarga de su estudio, ya que tradicionalmente han servido como elemento de clasificación para los mismos, como ocurre con los materiales numismáticos o los epigráficos.
- Contenidos del MAN. Agrupación de bienes culturales por la temática que abordan o los contenidos que tratan o representan. Se aplica fundamentalmente a documentos escritos, gráficos, sonoros o audiovisuales.
- Formas de presentación. Agrupación de bienes culturales por la forma en que se nos presentan. También se aplica generalmente para documentos escritos, gráficos, sonoros o audiovisuales.
- Materias elaboradas. Se incluyen en esta categoría los objetos que se encuentran en una fase intermedia entre la materia prima y la configuración final del mismo una vez finalizada su producción. Es una entrada transitoria, que debe desaparecer con el tiempo.
- Materias primas. Restos orgánicos. Agrupación de objetos naturales que adquieren un valor cultural por ser indicadores de la evolución de las especies o por ser usados en su estado natural por un grupo humano. Este apartado contiene lo que Kroustallis (2008) establecía como materias primas de origen natural, pues los productos realizados con las materias elaboradas ya se incluyen en otras entradas.
- Uso o función. La clasificación funcional es quizá uno de los criterios más claros a la hora de agrupar bienes culturales²⁰. Los descriptores incluidos en esta categoría se estructuran bajo dos indicadores clasificatorios, dependiendo de la pregunta a que respondan:
 - Por contextos de uso, esto es, ¿dónde se utilizan?, ¿en qué ámbito dentro de una sociedad?
 - Por finalidad, es decir, ¿por y para qué se fabrican o se usan?

Conclusión

La normalización documental es una de las grandes asignaturas pendientes que tenemos los museos. Han sido muchos años los que los técnicos de museos hemos tenido que hacernos nuestros propios instrumentos documentales, o adaptarlos a nuestras necesidades, y solo desde la implantación de DOMUS se han compartido normas y estándares entre un número significativo de museos. Además, el control terminológico solo ha sido efectivo desde que los administradores del sistema somos las únicas personas con permisos para creación, edición y eliminación de entradas en tesauros y listas cerradas.

Una vez se haya hecho la primera gran depuración de entradas en las herramientas de control terminológico y se hayan establecido las relaciones correctas entre sus contenidos, tendremos entonces que poner ese trabajo en común, para ver su aplicación en los distintos museos, e ir construyendo instrumentos comunes pero válidos para todos ellos. Y, superada esta fase, llegará el momento de revisar estos instrumentos con especialistas procedentes de otras instituciones y especialidades (universidades, centros de investigación...), a fin de conseguir unas herramientas útiles, no solo para el personal de cada museo, sino también para personas ajenas al mismo, ya que nuestro trabajo no tiene ningún sentido si no se pone a disposición de la sociedad, tanto de especialistas e investigadores como del público en general.

²⁰ Existe el campo «Uso / Función», que debería quedar para un contenido más descriptivo del modo en que se manejan los bienes culturales.

Bibliografía

- ALFARO ASINS, C. *et alii* (2009): *Diccionario de numismática*. Madrid: Ministerio de Cultura.
- BLAS BENITO, J. (coord.) (1996): *Diccionario del dibujo y la estampa: vocabulario y tesoro sobre las artes del dibujo, grabado, litografía y serigrafía*. Madrid: Real Academia de Bellas Artes de San Fernando.
- CABALLERO ZOREDA, L. (1988): «La documentación museológica», *Boletín de ANABAD*, vol. XXXVIII, n.º 4, ANABAD, pp. 455-496.
- CACHO QUESADA, C.; MARTÍ, J.; MAICAS RAMOS, R., y ARMADA, M. (1995): *Informatización y documentación en el Museo Arqueológico Nacional*. Madrid: ANABAD.
- CARRETERO PÉREZ, A. (1990): «Informatización y catálogos documentales del Museo Nacional del Pueblo Español», *Anales del Museo del Pueblo Español*, tomo III. Madrid: Ministerio de Cultura, pp. 239-293.
- CARRETERO PÉREZ, A.; QUINTO ROMERO, M.^a L. de; TIMÓN TIEMBLO, M.^a P. (1993): «Sistema de clasificación de documentación etnográfica española», *Anales del Museo del Pueblo Español*, tomo V. Madrid: Ministerio de Cultura, pp. 7-376.
- CARRETERO PÉREZ, A. *et alii* (1996): *Normalización documental de museos: elementos para una aplicación informática de gestión museográfica*. Madrid: Ministerio de Educación y Cultura.
- KROUSTALLIS, S. K. (2008): *Diccionario de materias y técnicas (I): materias*. Madrid: Ministerio de Cultura.
- PADILLA MONTOYA, C.; MAICAS RAMOS, R., y CABRERA BONET, P. (2002): *Diccionario de materiales cerámicos*. Madrid: Ministerio de Educación, Cultura y Deporte.
- PORTA, E.; MONSERRAT, R. M.; MORRAL, E. (1982): *Sistema de documentación para museos*. Barcelona: Departament de Cultura de la Generalitat de Catalunya.
- RODRÍGUEZ BERNIS, S. (2006): *Diccionario de mobiliario*. Madrid: Ministerio de Cultura.
- TRINIDAD LAFUENTE, I. (2011): *Tesoro y diccionario de objetos asociados a ritos, cultos y creencias*. Madrid: Ministerio de Educación, Cultura y Deporte.
- (2012): *Tesoro y diccionario de objetos asociados a la expresión artística*. Madrid: Ministerio de Educación, Cultura y Deporte.

La memoria recuperada: la búsqueda del nombre. La construcción del *Tesouro de Denominaciones de Bienes Culturales*

Isabel Trinidad Lafuente

A pesar del poder de las imágenes, o tal vez por ello, rara vez los pueblos, culturas o civilizaciones pasadas o presentes han creado sus objetos simplemente para ser vistos; cada uno de ellos suele tener un significado, un porqué, un para qué, un mensaje que puede estar más o menos velado pero que siempre existe. Porque la cultura es creación, no una mera copia de la realidad, sea cual sea su motivación: cultural, ritual, artística o estética. Pero para llegar a comprenderla y estimar el valor de conservar el producto de esta creación, es necesario recuperar de nuestra memoria y de la de aquellos que nos han precedido la historia de cada uno de estos objetos, conocer los diferentes avatares que han sufrido y que han quedado grabados en su estructura como parte de su memoria impresa. Y no podría ser de otra forma, porque, como decía Adolfo Colombes (1991), sin ella la cultura acabaría por convertirse «en una parodia de sí misma, en algo que se exhibe, no que se vive», y afortunadamente hoy ya los museos han dejado atrás para siempre la sacralidad del museo mausoleo dirigido a unos pocos para convertirse en un espacio vivo, dinámico, cuyo fin es la investigación, el disfrute, participación y enriquecimiento de todos, un espacio que permite el desarrollo de una educación no formal, porque el correcto conocimiento del pasado nos permitirá una mejor comprensión del presente.

Para ello es necesario no solo ver los objetos sino aprender a mirarlos y sobre todo, a escucharlos. Porque ellos hablan y cuentan historias, tantas como épocas, culturas, hábitos y lugares han atravesado desde que fueron creados por sus artífices, hombres y mujeres. Pero para que esa herencia que hemos recibido de nuestros antepasados podamos seguir legándola a generaciones futuras en las mejores condiciones, debemos transmitirla con su propio nombre, y cada objeto tiene el suyo propio, aquel con el que en su día fueron bautizados, porque solo así nosotros y los que nos sucedan podrán y podremos identificar cada uno de ellos para comprenderlo, conservarlo, documentarlo, estudiarlo, clasificarlo y comunicarlo como hoy está demandando a los museos la sociedad del conocimiento.



Figura 1. Alegoría de la atención. Weiss, Rosario. Museo Nacional del Romanticismo, n.º inv.: CE7126.

Ya en 1737, el taxonomista sueco Carl von Linné manifestaba que «si ignoras el nombre de las cosas, desaparece también lo que sabes de ellas».

Por eso, el *Tesaurus de Denominaciones de Bienes Culturales* elaborado por la Subdirección General de Museos Estatales de la Dirección General de Bellas Artes y Bienes Culturales y de Archivos y Bibliotecas, compuesto por todos aquellos objetos museables o susceptibles de serlo de cualquier área o disciplina, muebles e inmuebles, se propuso desde un principio recuperar no solo aquellos nombres genéricos (TG), sino también los de su larga familia, visualizándolos con su nombre específico (TE). Así, por ejemplo, el término más amplio de máscara englobará otros más concretos, y estos, a su vez, a terceros con el nombre dado por cada una de las culturas a la que pertenezcan:

TG Máscara

TE1 Máscara-casco

TE2 Máscara apasa

TE2 Máscara banda

TE2 Máscara janus

TE2 Máscara egungun

Y nos propusimos recuperar su nombre específico porque no nos podíamos quedar en un simple contenedor de ofrendas, porque estaríamos amputando parte de la rica cultura material que a lo largo de los siglos y de las culturas ha ido modulando el ser humano para acercarse a lo divino; y teníamos que darles nombre para que el *kernos* griego del siglo IV a. C. o el *panch patra* hinduista del XIX tuvieran también su sitio en nuestra memoria.



Figura 2. (Izquierda) Kernos. Museo Arqueológico Nacional, n.º inv. 11621; (derecha) Panch Patra. Museo Nacional de Antropología, n.º inv. CE1999/3/55.

Tesouro es una palabra de origen griego (*thesauros*) que significa ‘tesoro’, especialmente el tesoro de un templo. Referido tanto al continente como al contenido, es decir, tanto al templo construido en Delfos por los atenienses en el siglo V a. C., como a las armas conservadas en su interior y que estos toman de los persas durante la batalla de Maratón para ofrecerlas a Apolo.



Figura 3. Saqueo del Templo de Delfos. Coelemans, Jacques. Museo Lázaro Galdiano, n.º inv. 11196.

Y si *thesauros* significa tesoro, «thesaurizar» significa atesorar, amasar riquezas y, en sentido figurado, acumular bienes culturales, amasar información y, por lo tanto, conocimiento. Y eso es lo que pretendemos mientras construimos y damos forma al *Tesouro de Denominaciones de Bienes Culturales*. Desde un principio, el objetivo fue el de convertirlo en un acceso temático, en la memoria del museo, una memoria con un lenguaje jerarquizado y normalizado, para que, en una segunda fase, termine siendo la herramienta más eficaz en el ámbito museístico compartida por todos y de la misma forma a través de la web del conocimiento.

El *Tesouro de Denominaciones de Bienes Culturales* está compuesto, hasta el momento, por 12900 descriptores y por 4260 no descriptores. Objetos relacionados con los cultos, pasando por juegos o juguetes, armas, circo, cine, fotografía, instrumentos musicales, dibujo, pintura o escultura, la arquitectura y sus elementos o estructuras, herramientas y útiles o todo aquello

que hombres y mujeres hayan podido crear para el disfrute propio o el ajeno, para la ofrenda, promesas o protección natural, humana o divina.



Figura 4. De izquierda a derecha y de arriba abajo: Urna funeraria, Museo de América, n.º inv. 02589; «Navidad en el circo», Paredes Jardiel, José, Museo del Teatro, n.º inv. Es-1; Livia de Paestum, Museo Arqueológico Nacional, n.º inv. 2737; Colotipo, Museo Nacional del Romanticismo, n.º inv. CE36190; Lampadoscopio, Museo del Traje, CIPE, n.º inv. CE038021A; «El accidente», Ponce de León, Alfonso, Museo Nacional Centro de Arte Reina Sofía, n.º inv. AS00745; Zanfoña, Museo del Traje, CIPE, n.º inv. CE085722; Capitel románico, Museo Arqueológico Nacional, n.º inv. 50004.

Y todo ello sin limitaciones temporales, culturales ni geográficas, desde el ídolo oculado del III milenio a. C. hasta las instalaciones o *performances* del siglo XXI. Pero siempre a través de un camino común capaz de aunar disciplinas o manifestaciones temporales e históricas tan diferentes, un camino común, un hilo conductor que no podía ser sino el propio objeto, único elemento en el que coinciden los yoruba de Nigeria, los griegos del siglo IV a. C. o los cristianos de hoy.

Es decir, nuestro Tesoro debía abarcar el universo, pero intentar englobarlo de una sola vez es cuanto menos complejo y además sus resultados se verían a muy largo plazo. Por esta razón, se decidió publicar el trabajo por ramas conforme estas fueran completadas. En la actualidad, el Ministerio de Educación, Cultura y Deporte está estudiando los medios técnicos para volcarlas a los museos, con el fin de que estos puedan tener así acceso a los tesoros a medida que las diferentes disciplinas vayan siendo terminadas, como ya se ha hecho con otros tesoros, como el de *Materias* o el *Geográfico*.

Hasta ahora la Subdirección General de Museos Estatales ha publicado:

- *Tesoro y Diccionario de objetos asociados a ritos, cultos y creencias*, sistematizando, describiendo y contextualizando los objetos de cultos animistas, precolombinos, los clásicos (Grecia, Roma, Etruria o los pueblos íberos), Egipto, los relativos a las conocidas como las religiones del libro (ortodoxos, católicos o protestantes), musulmanes y judíos, así como las religiones asiáticas: budistas o hinduistas. Compuesto por un total de 1671 descriptores y por 739 no descriptores.

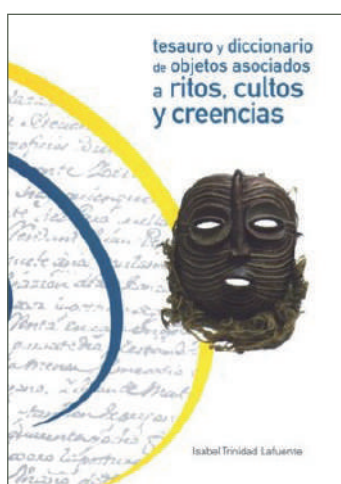


Figura 5. Portada del *Tesouro y Diccionario de objetos asociados a ritos, cultos y creencias*.

- *Tesouro y Diccionario de objetos asociados a la expresión artística*, compuesto por un total de 2356 descriptores y por 770 no descriptores y que reúne objetos vinculados a las:
- Artes de la imagen: cine, fotografía y arte gráfico.
 - Artes escénicas: música, artes de la representación (como teatro o danza) y el circo.
 - Artes plásticas: dibujo, escultura y pintura.



Figura 6. Portada del *Tesouro y Diccionario de objetos asociados a la expresión artística*.

Ambos consultables en línea <<http://tesauros.mecd.es/tesauros/tesauros>>.

- Por último, en fase de publicación, el *Tesouro y Diccionario de arquitectura y estructuras*, compuesto por 2296 descriptores y por 1179 no descriptores. En el que se sistematizan los términos desde su elaboración y preparación, como el dibujo técnico o las maquetas arquitectónicas, pasando por los relacionados con su realización y creación, como las construcciones jerarquizadas según su función: civil (públicas o privadas), conmemorativas, funerarias, militares o religiosas, o los elementos arquitectónicos, como aperturas

y vanos, cierres o separaciones, circulación, cubrición o techado, de fábrica y obra, de revestimiento, sustentación u ornamentos, hasta las estructuras no edificatorias, como los jardines.



Figura 7. Maqueta de la portada del *Tesaurus y Diccionario de arquitectura y estructuras*.

Por otra parte, y con respecto a los criterios de elección para ir acometiendo las diferentes ramas o especialidades, estos han sido en primer lugar su carácter práctico, es decir, que pudieran ser de utilidad para la mayor parte de los museos. Caso evidente en los dos primeros tesauros, ya que la mayor parte de las colecciones estables de fondos museísticos, en cualquiera de los diferentes tipos de museos, sean estos antropológicos, arqueológicos, casas museo, de artes decorativas o de bellas artes o incluso de ciencia y de tecnología, cuentan entre sus fondos con objetos que pueden ser incluidos en cualquiera de los dos primeros casos.

Al primero de ellos, el *Tesaurus y Diccionario de objetos asociados a ritos, cultos y creencias*, se le unía además el hecho de que, hasta ese momento, no existía en el ámbito científico un tesaurus del conjunto de creencias con esta amplitud y, sobre todo, que tuviera como protagonista al objeto y no a la historia de la religión o de la iconografía religiosa.

Tras los dos primeros, especialmente importante a la hora de seleccionar de entre el resto de las ramas abordables de bienes culturales fue el carácter pluridisciplinar del escogido, el *Tesaurus y Diccionario de arquitectura y estructuras*, además de por su aportación a labores tan importantes como las desarrolladas por los Planes Nacionales para la defensa del Patrimonio Cultural inmueble. Ya que, partiendo del estudio y de la unificación de estructuras descriptivas y clasificatorias de las colecciones y de la terminología técnica utilizada, se facilita un óptimo intercambio y recuperación de información en la red, permitiendo así racionalizar y optimizar los recursos destinados a su catalogación, conservación y difusión. Objetivos estos perseguidos por los diferentes Planes Nacionales desde su puesta en marcha en la segunda mitad de la década de 1980: Plan Nacional de Catedrales, Patrimonio Industrial, Arquitectura Defensiva, Paisaje Cultural, Abadías, Monasterios y Conventos, Arqueología Subacuática o Arquitectura tradicional.

Así, buscando el nombre de nuestros vestigios más antiguos o recientes, salimos esta vez extramuros de los museos.



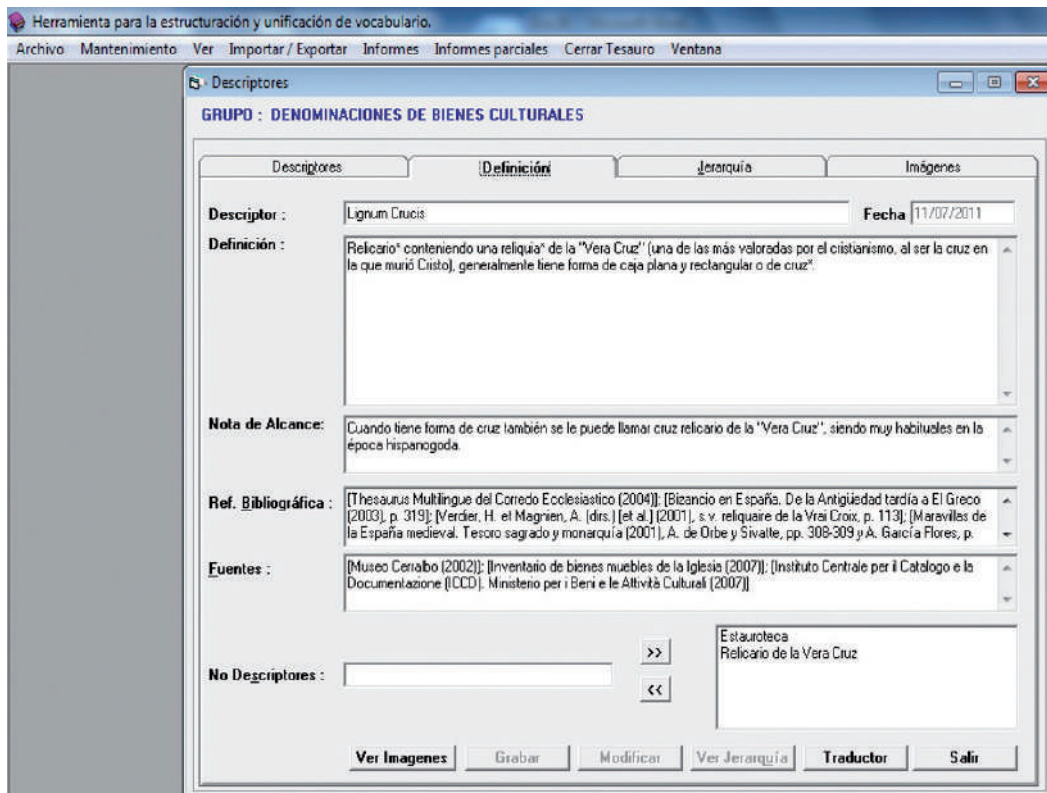
Figura 8. Scaena. Museo Nacional de Arte Romano, n.º inv. 55-XVI-8.

Con el desarrollo de esta tercera rama se ha dado un paso más en el apoyo de la filosofía con la que se crearon los grupos de trabajo para la elaboración de los tesauros, perseguida desde que, en 1993, se iniciara el proyecto a través de la Comisión de Normalización Documental de Museos, con vistas ya no solo a su utilización como herramientas de control terminológico en DOMUS, sino también enmarcada en el proyecto «Sistemas de Información del Patrimonio Histórico en la Red», que condujo a la actual CERES. Ello posibilita una tercera vía de utilización de los tesauros enmarcados en la Web Semántica (web del conocimiento), uniéndose así a la comunidad científica mundial, aportando el trabajo desarrollado en este campo y enriqueciéndose, a su vez, con la experiencia de otros.

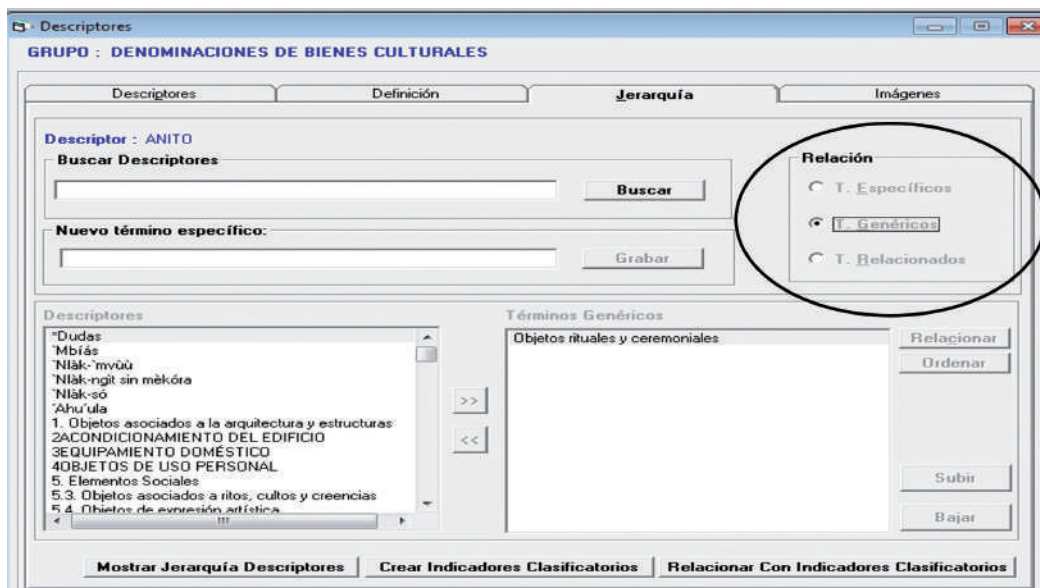
The image shows the homepage of the 'tesauros' website. At the top left, there are logos for the Spanish Government and the Ministry of Education, Culture and Sports. The main header features the 'tesauros' logo and the text 'del Patrimonio Cultural de España'. Below this is a navigation menu with links: 'Acceso y/o consulta', 'Buscador', 'Linked Open Data', 'Condiciones de uso', 'Solicitud reproducción', and 'Ayuda'. The main content area is divided into two columns. The left column is titled 'Tesauros-Diccionarios del patrimonio cultural de España' and contains a paragraph about the digital publication and a red button labeled 'ACCEDER A LOS TESAUROS'. The right column is titled 'Linked Open Data' and contains text about the availability of dictionaries as Linked Open Data collections. At the bottom, there are logos for various standards like SPARQL, RDF, SKOS, and Open Data, along with 'Créditos' and 'Contacto' links.

Pero han sido varias las fases y múltiples las dificultades con las que nos hemos ido encontrando en su desarrollo. Así, antes de la elaboración propiamente dicha, se tuvieron en cuenta las siguientes consideraciones.

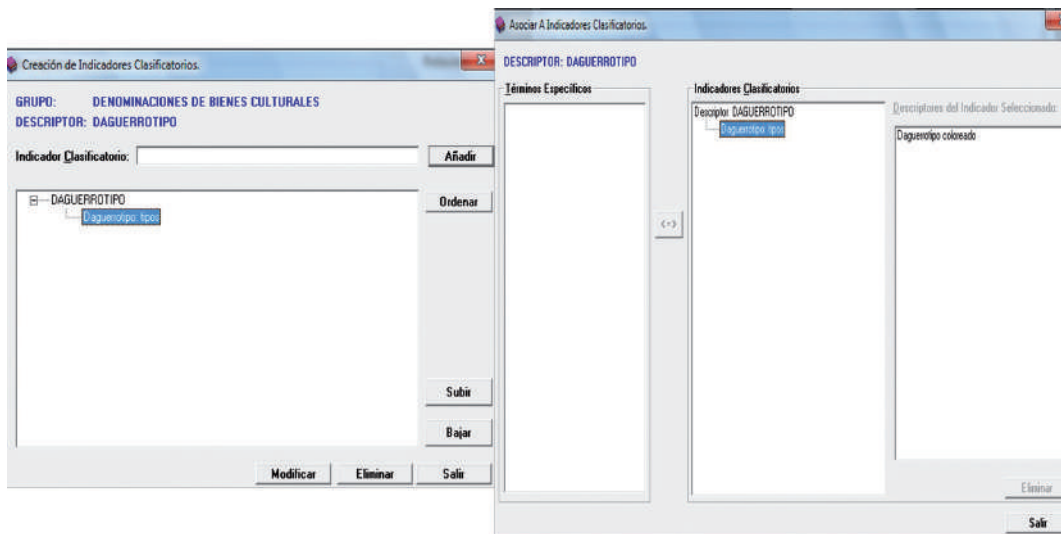
- Dar de alta no descriptores o términos no preferentes, sin información pero que nos remiten a los anteriores.



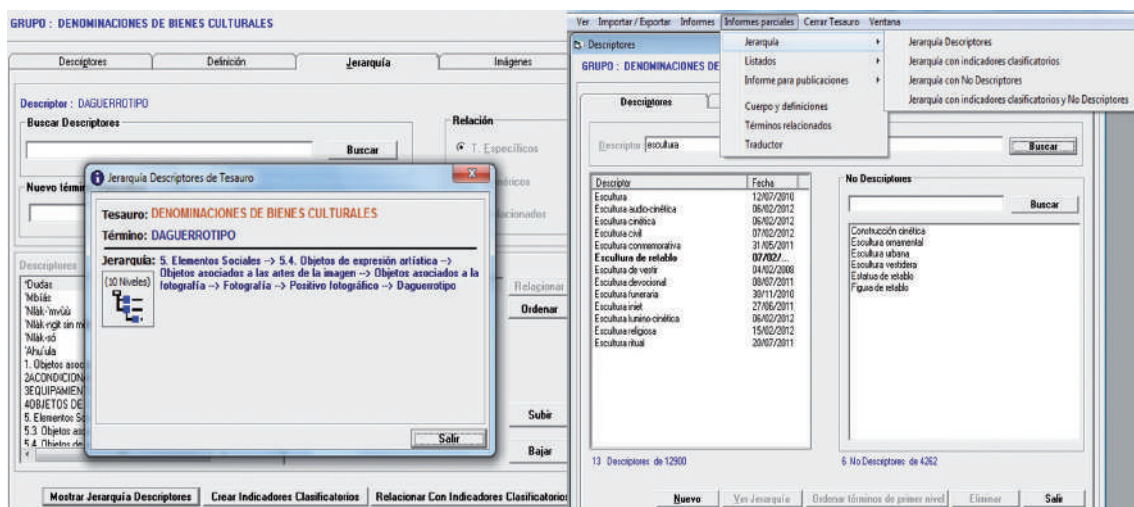
- Crear relaciones entre los términos de equivalencia, jerárquicas y asociativas.



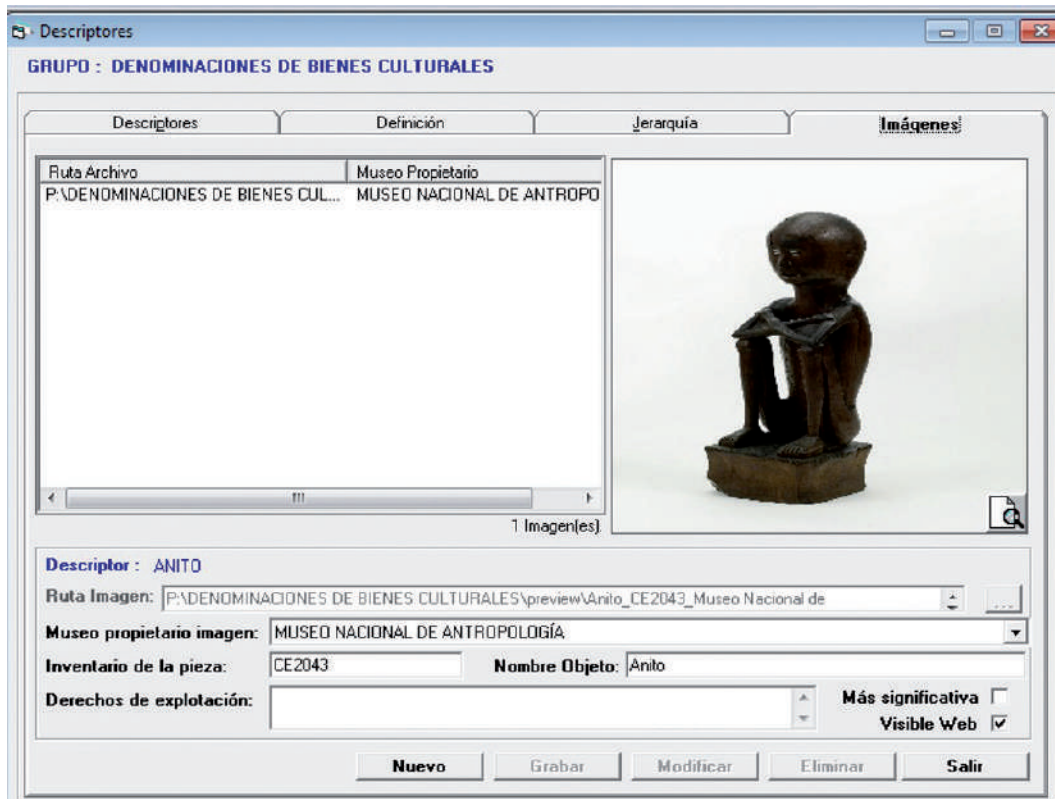
- Crear indicadores clasificatorios y relacionar los descriptores con estos.



- Jerarquizarlos y ver dicha jerarquía en la ficha de cada descriptor o generar informes de estas jerarquías.



- Vincular imágenes con cada uno de los descriptores.



- Generar todos los informes en formato texto necesarios para su publicación: cuerpo del tesoro, donde se reflejan todas las relaciones creadas entre nuestros términos y el Diccionario.

Cuerpo del tesoro
 Acetre
 UP Caldereta
 Cubo de agua bendita
 TG Objetos rituales y ceremoniales
 TE Sítula
 TR Hisopo

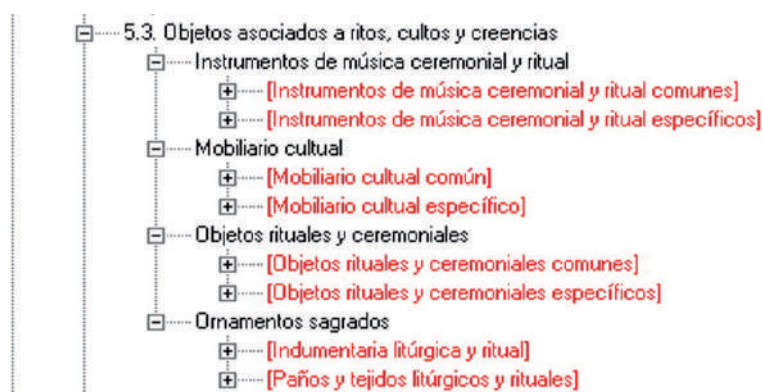
- En cuanto a la construcción propiamente dicha del tesoro, se establecieron las siguientes fases:
- El primer paso fue el de la delimitación del campo terminológico: Tesoro de Denominaciones de Bienes Culturales, Muebles e Inmuebles.
 - El segundo, el de la determinación de campos temáticos, es decir, las distintas ramas del tesoro, así como la distribución de descriptores dentro de estas. El método de trabajo ha sido el mismo para cada uno de ellos, evidentemente y como ya se ha mencionado, cada uno tratado con las especificidades propias de cada disciplina que nos conduce a la creación de relaciones y campos temáticos propios.

En aras de una mayor claridad, nos centraremos en un ejemplo concreto, el de *Ritos, cultos y creencias*. En un principio, contamos con un sinfín de piezas procedentes de los distintos museos e instituciones, a los que le buscamos y acordamos el término aceptado (descriptor)

y los no descriptores, desarrollando así las relaciones de equivalencia: USE (Use) / UP (Usado por), y ello mediante la consulta y consenso con distintas fuentes: especialistas en cada una de las culturas tratadas, instituciones museísticas, centros de investigación, universidades y fuentes bibliográficas².

Ya teníamos los descriptores (piezas) y las fuentes (los museos o instituciones que nos proporcionaban estos objetos). El siguiente paso fue el desarrollo de la estructura jerárquica básica del tesoro, es decir, determinar las grandes áreas temáticas (macrocategorías) o familias semánticas en que quedaría dividido. De esta manera se delimitó el campo de trabajo y se distribuyó este por áreas más específicas, lo que permitió agrupar los descriptores de forma precisa en el contexto semántico correspondiente, así como desarrollar las relaciones jerárquicas de manera precisa. Comenzamos entonces a elaborar las relaciones de jerarquía: Término Genérico (TG) / Término Específico (TE). Para ello se crearon cuatro Términos Genéricos, que agruparían los TE en cuatro grandes familias: Instrumentos de música ceremonial y ritual, Mobiliario cultural, Objetos rituales y ceremoniales y Ornamentos sagrados.

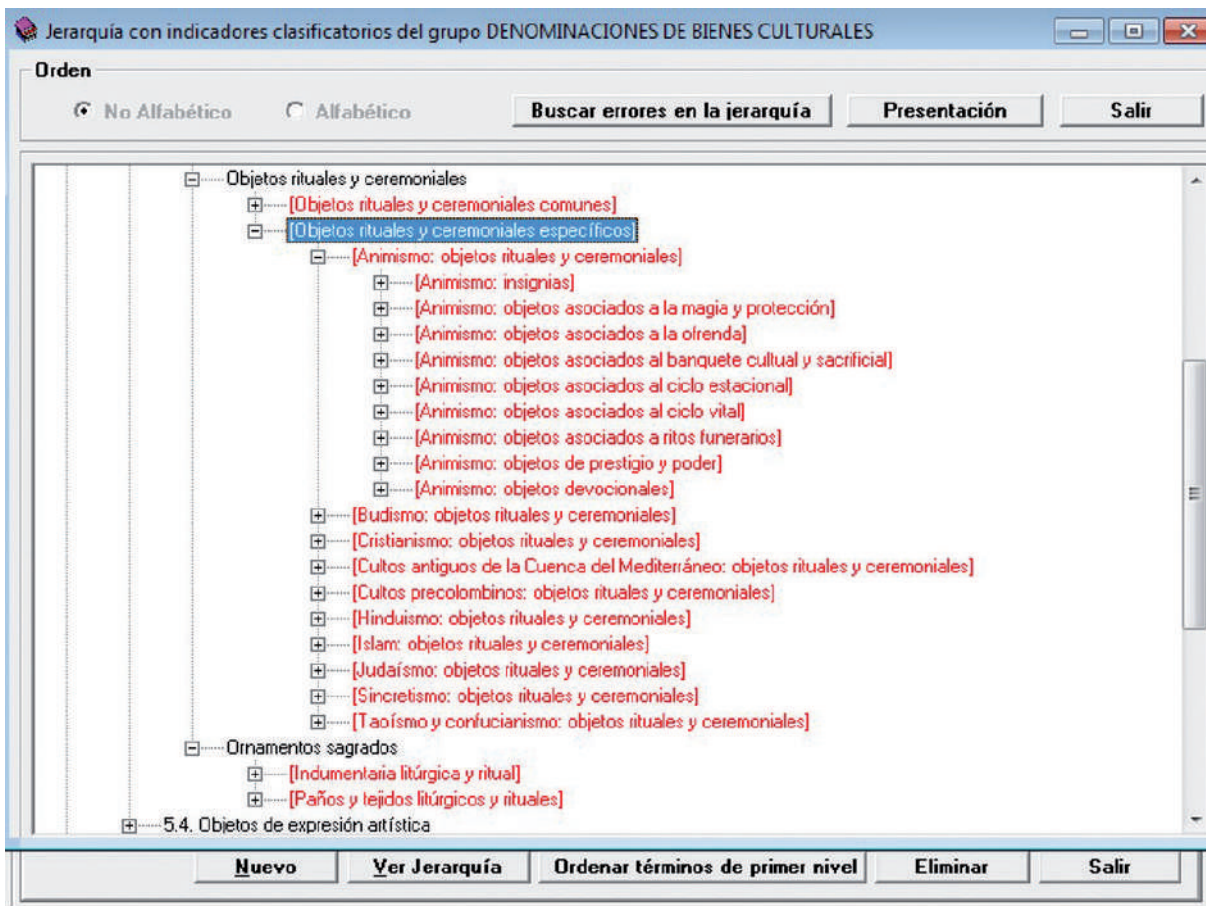
Llegado a este punto, se planteó la siguiente dificultad: aunque algunas piezas pertenecían de manera exclusiva a un culto determinado, muchas otras podían ser utilizadas, prácticamente con el mismo carácter o función, por diferentes cultos; recordemos los contenedores de ofrendas. Así, se decidió crear dos grandes indicadores clasificatorios: comunes y específicos³.



El siguiente paso consistió en crear los indicadores dentro de los cuales pudiéramos organizar los descriptores según la función que tuvieran en sus respectivos cultos; por ejemplo, en el caso de los objetos relacionados con el animismo.

² A lo largo de todos estos años han sido innumerables los museos, instituciones y, sobre todo, personas las que nos han brindado su apoyo y asesoramiento y que han colaborado con nosotros. Ricardo Olmos (CSIC), Alfonso Rodríguez Gutiérrez de Ceballos (UAM), Juan Miguel Sánchez Vigil (UCM), Pedro Rocamora (Consejo Estatal de las Artes Escénicas y de la Música), Juan Antonio Pérez Millán (Filmoteca de Castilla y León), musicólogos como Eduardo Paniagua; arquitectos como Juan Gimeno o circólogos como José A. Gallego; y, por supuesto, técnicos de museos: Francisco de Santos, Pilar Romero de Tejada, Araceli Sánchez, Ana Verde, Marigel Castellano, Paloma Cabrera, Margarita Moreno Conde, Ángela Franco, Asunción Cardona, Carolina Miguel, Lorena Delgado, Concha García-Hoz, Víctor Cageao y tantos otros sin los cuales jamás hubiéramos podido culminar este trabajo.

³ Los indicadores clasificatorios son en realidad como los cajones de nuestro escritorio, que nos permiten ordenar de manera coherente y precisa nuestras pertenencias, o los archivadores en los que podemos normalizar los distintos expedientes por fecha correlativa de ingreso o por agrupaciones, lo que facilita y acorta el tiempo de búsqueda.



Por otra parte, los tesauros nos permiten utilizar distintas opciones que evitan el vacío de información. Para ello creamos las relaciones asociativas: Términos Relacionados (TR). De esta forma, un objeto clasificado en el culto animista cuya función sea la magia o protección estará ligado a otro hinduista con el mismo uso, dando así al investigador la posibilidad de ampliar o enriquecer el sujeto de su trabajo. A veces puede llegar a sorprender que los inuit de Alaska tengan las mismas motivaciones para modelar un objeto que los antiguos senadores romanos cuando protegían a sus hijos con *bullae* de oro.

Evidentemente, una máscara ceremonial participa de distintas creencias, por lo tanto se agrupará en el Indicador Clasificador Comunes, pero a su vez se le da la opción al investigador de encontrar, a través de ella y de sus Términos Relacionados, las diferentes máscaras que, con este mismo carácter, reciben un nombre específico según el rito, pueblo o cultura a la que pertenezcan. La *mahâkâla* es una máscara ceremonial protectora del budismo tibetano, como la *ekeke* lo es para la cultura fang africana.



Así, lo que en un principio nos llegó como un maremágnum de palabras y de objetos, se fue ordenando de manera jerarquizada y normalizada para que todos y todas podamos recuperar nuestros objetos al llamarlos con el mismo nombre, hasta conformar la estructura de la presentación del tesoro, último de los pasos de su elaboración.

<p>[Grecia y Roma: objetos rituales y ceremoniales]</p> <p>[Grecia y Roma: objetos asociados a concursos rituales]</p> <p>*** Anfora panatenaica</p> <p>[Grecia y Roma: objetos asociados a las abluciones]</p> <p>*** Lebeta nupcial</p> <p>[Lebeta nupcial: tipos]</p> <p>**** Lebeta nupcial atica "tipo 1"</p> <p>**** Lebeta nupcial atica "tipo 2"</p> <p>**** Lebeta nupcial suritalica</p> <p>[Grecia y Roma: objetos asociados a ritos funerarios]</p> <p>*** Ascós funerario</p> <p>*** Copa de pie calado</p> <p>*** Lécito</p> <p>[Lécito: tipos]</p> <p>**** Lécito de fondo blanco</p> <p>[Grecia y Roma: objetos asociados a ritos de iniciación]</p> <p>*** Coe</p> <p>*** Hidria</p> <p>*** Kalathos</p>	<p>[Grecia y Roma: objetos asociados al banquete cultural y sacrificial]</p> <p>*** Aryster</p> <p>*** Cantaros</p> <p>*** Cratera</p> <p>[Cratera: tipos]</p> <p>**** Cratera de caliz</p> <p>**** Cratera de campana</p> <p>**** Cratera de columnas</p> <p>**** Cratera de volutas</p> <p>*** Enócoe</p> <p>*** Escifo</p> <p>[Escifo: tipos]</p> <p>**** Glauca</p> <p>*** Estamno</p> <p>*** Hethmos</p> <p>*** Karchesion</p> <p>*** Keras</p> <p>*** Kylix</p> <p>*** Oscillum</p> <p>*** Ritón</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

De esta forma le pudimos poner de manera correcta sus rostros a nuestros términos, a nuestros nombres.



Figura 9. (Izquierda) Kylix. Museo Arqueológico Nacional, n.º inv. 10942; (derecha) Ritón. Museo Arqueológico Nacional, n.º inv. 1999/99/130.

Pero lo novedoso y enriquecedor de este proyecto, que en su día emprendiera la Subdirección General de Museos Estatales, fue la integración en el tesoro de un diccionario que permite, tanto al especialista como al neófito, obtener un conocimiento descriptivo y contextual de cada uno de los objetos integrantes de cada una de las áreas tratadas. De este modo, la aplicación nos permite: describir el objeto, contextualizarlo en su nota de alcance, además de señalar las citas de las fuentes bibliográficas utilizadas. Y todo ello acompañado con la riqueza visual de un anexo gráfico.

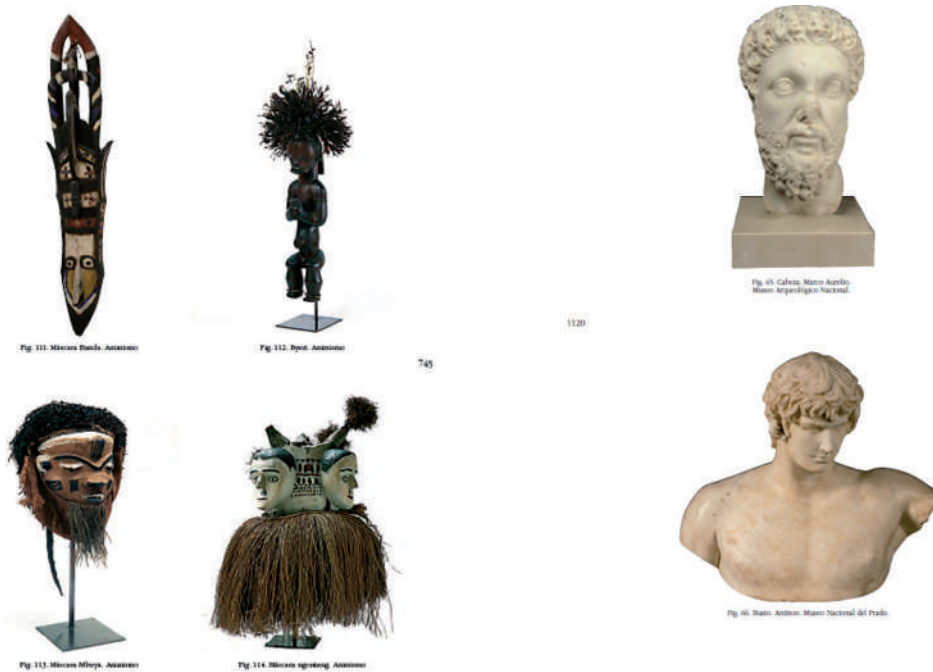


Figura 10. De izquierda a derecha y de arriba abajo: Máscara Banda, Museo Nacional de Antropología, n.º inv. CE19005; Byeri, Museo Nacional de Antropología, n.º inv. CE947; Cabeza de Marco Aurelio, Museo Arqueológico Nacional, n.º inv. 2009/36/1; Máscara Mbuya, Museo Nacional de Antropología, n.º inv. CE11778; Máscara Ngontang, Museo Nacional de Antropología, n.º inv. CE11058; Busto de Antino, Museo Nacional del Prado, n.º inv. E00060.

En cualquier caso, siempre se ha de tener en cuenta, como ya hemos señalado a lo largo del texto, que un tesoro no es una lista cerrada, es algo vivo, dinámico, que necesita de una constante actualización, en el que se podrían incluir los términos propuestos por los museos o instituciones a los que este va dirigido, modificarlos o incluso eliminarlos. Se trata de establecer una herramienta integradora que favorezca la interconexión de sistemas, que sea lo suficientemente amplia en su cobertura y, al mismo tiempo, específica en la materia como para permitir conceptualizar una temática concreta, posibilitando así mejorar la organización, la comunicación y la difusión de las colecciones del museo. En definitiva, convertirlo, de este modo, en una herramienta útil para todos aquellos que tengan la inmensa suerte de trabajar con ellos, con los objetos, a los que ya conocemos y podemos llamar con su propio nombre y apellidos. Porque esa es la importancia de cualquier tesoro cuyo fin sea el de conservar, enriquecer y facilitar la contextualización de los bienes culturales: recuperar esos nombres del pasado y los de su filiación, quedando así ya para siempre impresos como parte de nuestra memoria histórica.



Figura 11. Alegoría de la atención. Weiss, Rosario. Museo Nacional del Romanticismo, n.º inv. CE7126.

Bibliografía

- HARPRING, P. (2010): *Introduction to Controlled Vocabularies: Terminology for Art, Architecture and Other Cultural Works*. Los Ángeles: The Paul Getty Trust.
- NAGEL, L. (2008): *Manual de Registro y Documentación de Bienes Culturales*. Santiago de Chile: Centro de Documentación de Bienes Patrimoniales (Dibam).
- NAGEL, L., y MILLER, CH. (2013): «The Tesoro de Arte & Arquitectura and Tesoro Regional Patrimonial: Tools for Describing and Enhancing Access to Latin American Cultural Resources Online», *Getty Research Journal*, n.º 5. Los Ángeles: The Paul Getty Trust, pp. 149-156.

PERRIN, J. (1999): *Thesaurus des objets religieux du culte catholique*. Edición trilingüe: inglés, francés e italiano. París: Editions du Patrimoine.

VV. AA. (2002): *Thesaurus multilingue del corredo ecclesiastico*. Istituto Centrale per il Catalogo e la Documentazione.

Webgrafía

Arte & Architecture Thesaurus Nederlandstalig: <<http://browser.aat-ned.nl/>> [Consulta: 4 de febrero de 2016].

British Museum: <http://www.collectionstrust.org.uk/assets/thesaurus_bmon/Objintro.htm>.

Centro de Documentación de Bienes Patrimoniales, Dirección de Bibliotecas, Archivos y Museos (Dibam): <<http://www.aatespanol.cl/taa/publico/buscar.htm>> [Consulta: 4 de febrero de 2016].

Historic England, Forum on Information Standards in Heritage (FISH) Thesauri: <<http://thesaurus.historicengland.org.uk/newuser.htm>> [Consulta: 4 de febrero de 2016].

Iccd, Istituto Centrale per il Catalogo e la Documentazione: <<http://www.iccd.beniculturali.it/index.php?it/1/home>> [Consulta: 4 de febrero de 2016].

Ministère de la Culture et de la Communication: <[http://www.culture.fr/Ressources/Bases-du-patrimoine/\(theme\)/54944](http://www.culture.fr/Ressources/Bases-du-patrimoine/(theme)/54944)> [Consulta: 4 de febrero de 2016].

Ministerio de Educación, Cultura y Deporte: <<http://tesauros.mecd.es/tesauros/tesauros>> [Consulta: 4 de febrero de 2016].

Tesouro Regional Patrimonial de Chile: <<http://www.tesouroregional.cl/trp/publico/buscar.htm>> [Consulta: 4 de febrero de 2016].

The Getty Research Institute: <<http://getty.edu/research/tools/vocabularies/>> [Consulta: 4 de febrero de 2016]; <<http://getty.edu/research/tools/vocabularies/aat/index.html>>; [Consulta: 4 de febrero de 2016].

UNESCO: <<http://unesdoc.unesco.org>> [Consulta: 4 de febrero de 2016].

Tesauros, vocabularios y listas de control terminológico. Fuentes para la construcción de tesauros culturales

Miguel Ángel Marzal García-Quismondo

Profesor Titular de Biblioteconomía y Documentación. Universidad Carlos III de Madrid

1. Introducción

No cabe duda de que las TIC han transformado, y van a revolucionar, todo el entorno humano en casi todas sus facetas. Entre los proyectos de investigación que se están desarrollando en el Instituto del Futuro de la Humanidad, de la Universidad de Oxford, uno de ellos se denomina *Global Priorities Project* (<<http://globalprioritiesproject.org/>>), cuyo objetivo es reflexionar y analizar los mayores desafíos que en el futuro inmediato amenazan la supervivencia de la humanidad, y al menos dos tienen relación directa con las TIC: la Inteligencia Artificial con la Robótica y la Nanotecnología. El modo de aprender, conocer y saber del ser humano, naturalmente, también se verá radicalmente alterado.

Como es bien sabido, en todo proceso cognitivo los datos y la información son elementos primordiales. Para la efectividad de este proceso, ambos elementos necesitan una «corporeización» en un producto u objeto, que son los documentos en soporte material, digital o virtual, de los que se nutren las unidades de información (bibliotecas, archivos, centros de documentación), clásicas, híbridas, digitales o virtuales. A veces estos documentos u objetos no solo soportan y transmiten datos e información para conocer y saber, sino que portan valores estéticos, erigiéndose en objetos artísticos, cuya unidad, en este caso, es el museo. Se constituye así una amplia y variopinta «colección documental» que se proyecta en un *patrimonio histórico, artístico y documental*. Precisamente las TIC, tal como venimos argumentando, han alterado la noción de documento, con el desarrollo del hipertexto, los hipermedia, la realidad virtual y la aumentada, pero también de forma muy notable al objeto artístico, como demuestra la eclosión del *Media Art*, derivado de la convergencia de los medios con los soportes digitales y que utilizan materiales y las TIC para crear, presentar y difundir el arte.

La evolución expuesta ha llevado a plantearse el Museo Virtual, entendido como aquel que utiliza los medios digitales para mostrar, preservar, diseminar, guardar y reconstruir la cultura material e inmaterial de la humanidad. En el proceso, ha ido emergiendo el concepto, muy

interesante, de *Museografía Web*. Es muy importante señalar, además, que este proceso también está afectando a la evolución de los archivos y su colección documental, que también está experimentando una evolución hacia la «virtualización» de sus fondos, que precisan de vocabularios y lenguajes web propios para su comunicación, difusión y virtualización. Este será el objetivo de este trabajo.

2. Contexto para las fuentes en la construcción de los lenguajes documentales: tesauros y cultura

Consecuencia de lo que llevamos argumentado es el creciente interés por la elaboración y aplicación inteligente de los lenguajes documentales, especialmente el tesauro, sobre la colección de archivos y museos. Caso prototípico ha sido la publicación el 19 de octubre de 2015 por parte de CER.ES del Ministerio de Educación, Cultura y Deporte de dos vocabularios para el patrimonio cultural, el *Diccionario de lugares geográficos* y el *Diccionario y Tesauro de materias* (<<http://tesauros.mecd.es/tesauros>>), mediante una versión en *datos abiertos enlazados*. Vocabulario y tesauro nacen con el objetivo de representar por descripción y catalogación los bienes culturales, para ser compartido por la comunidad profesional y científica por los datos abiertos, además permite enlazar con las colecciones de otras poderosas instituciones, con unos tesauros y vocabularios sumamente poderosos. Otros museos y archivos están en un proceso de similares características y proporciones, como ha quedado de manifiesto en el curso *El lenguaje del patrimonio: Estándares documentales para la descripción y gestión de colecciones*, celebrado en Nájera entre los días 18 y 20 de noviembre de 2015.

Es evidente que el campo de estos proyectos es la Web semántica, el objeto son los documentos y objetos museísticos, el objetivo es la virtualización, y el método, la generación de tesauros para su representación y elaboración, un método que reclama unas fuentes terminológicas de construcción apropiadas, un proceso de selección que debe atender a un contexto con sus condicionantes, algo fundamental en cualquier «sistema de comunicación».

2.1 El *valor* en la sociedad del conocimiento

La información como materia prima y fuente energética, un recurso insoslayable dentro de la economía digital, ha subvertido una de las leyes de bronce del mercado capitalista de oferta y demanda: si, en este caso, el valor de las cosas tiene como variable básica la superioridad de su demanda, convirtiéndola en «escasa», siendo un objeto más precioso cuanto más raro y escaso; en la economía digital el valor de un objeto va en relación directa con su mayor (universal) disponibilidad y acceso. Los objetos y documentos de archivo, en este contexto, cuanto más representados, visibles y capturables en web más valiosos se hacen y, con ellos, su museo y archivo.

Se abre paso con fuerza, pues, la *cultura de la evaluación*, impulsando, entre otros aspectos, la relevancia del *impacto*. Todo objeto digital se ve evaluado y medido su «valor» por la cantidad de accesos, visitas, referencias, citas que recibe: la unidad de medida es el indicador y el análisis de datos se efectúa en el contexto de *rankings*. Científicos e investigadores deben permanecer muy atentos a su impacto, de modo que las bibliotecas que los apoyan suelen ofrecer un servicio muy elaborado para asegurarse el impacto y gestionarlo para su currículo, visible en bibliotecas universitarias, caso de la biblioteca de la Universidad Carlos III de Madrid <http://www.uc3m.es/ss/Satellite/Biblioteca/es/TextoMixta/1371212970020/Indices_de_impacto>.

La comunicación en medios y la socialización adquieren así una dimensión nueva y diferente mediante la constitución de *comunidades digitales* en torno a un «tópico» en web, que tienen como principios básicos el trabajo y aprendizaje colaborativo, la innovación y constante generación de contenidos mediante las herramientas de la web social (wikis, blogs, folksono-

mías, etc.). J. Sowa (Sowa, 2000) llegaba así a la noción de Red Semántica a partir de un sistema de notaciones gráficas que permiten definir un conjunto de relaciones asociativas de términos en los textos. La Red Semántica determina la significación de conceptos, palabras, frases, a través de nodos y palabras conceptuales conectadas la una a la otra, un auténtico «clúster» semántico, un modo de adaptarse a las estrategias que el cerebro utiliza para aprender, entre ellas la asociación y la mediación, esta última una competencia que trata de transformar algo complejo en algo más sencillo de recordar a través de enlaces, asociaciones y estructuras organizadas, como, por ejemplo, taxonomías jerárquicas.

Derivación lógica de las redes semánticas es la Red de Conocimiento, entendida como un mecanismo y confluencia social de organizaciones o individuos para promover el intercambio de información, compartir técnicas, metodología y buenas prácticas de trabajo, cooperar en acciones propias para la capacitación, investigación e innovación, acumular conocimiento tácito y explícito basado en la complementariedad y reciprocidad (Malerba *et alii*, 1999). Para su más sano desarrollo, ha ido emergiendo el Conectivismo y sus principios (Siemens, 2010).

El escenario apropiado para la adquisición e intercambio de conocimiento es la red social. De acuerdo con la teoría de redes sociales, tanto la imbricación (*embeddedness*) estructural como la relacional determinan la estructura de la red y la forma de la transferencia de conocimiento. Las redes sociales sin un adiestramiento, una competencia en lectura digital, pueden tener efectos indeseados (Carr, 2011), como dispersión, hiperactividad o una pobre asimilación. En las redes sociales prima la «lectura digital de comunicación», que busca noticias (nunca discursos), busca tópicos (nunca argumentaciones), busca novedades (nunca una reflexión). No cabe duda de que la lectura digital en redes reclama sus propias competencias como primar los conectores del discurso (vínculos, enlaces, navegación) y generar su propia secuencia lectora (multisequencialidad de los hiperdocumentos) etiquetando los tags, etiquetas semánticas del discurso electrónico, como puede permitirnos el software NodeXL.

Esta constatable evolución hacia un conocimiento en redes ha tenido tres inmediatos efectos para su mejor desarrollo: a) para las organizaciones la *reputación en Web*, una técnica de *marketing* que busca el modo de proyectar más y mejor la influencia en la red, convertirse en autoridad y referencia, así como prevenir tendencias, para lo que existen software utilizables como Social Mention (<<http://www.socialmention.com/>>), Whos Talkin (<<http://whostalkin.com/>>) o SamePoint (<<http://www.genbeta.com/web/samepoint-buscador-de-conversaciones-en-internet>>); b) la *identidad digital*, que no solo pretende mecanismos para «distinguir» al individuo en la red, sino que desarrolla competencias contra la permanencia indeseada de información propia en red, así como evitar los excesos de la copia (Cerezo, 2011); c) los *marcadores sociales* que, mediante el uso de tags, permiten categorizar la información de la web de forma organizada para luego recuperarla, reutilizarla o compartirla. Sin duda, son efectos que determinan el progreso, caracteres y propiedades de los lenguajes y vocabularios en red.

2.2 La concepción cambiante de *cultura*

Otro factor contextual para seleccionar fuentes en la elaboración de «tesauros culturales» es, sin duda, la concepción de *cultura*, no desde una perspectiva antropológica y filosófica (debate en el que no parece adecuado entrar ahora), sino como «dominio» para un tesauro, tratándose además de un concepto cambiante en esta faceta.

En un contexto terminológico para la elaboración de lenguajes documentales web, el debate se desarrolla desde un planteamiento dicotómico, entre la definición por binomios bien diferenciados: a) la distinción entre «civilización», asentada sobre mentalidades, tradiciones, categorías vitales, y «cultura», en tanto que comportamientos, actitudes, manifestaciones y expresiones del espíritu; b) la diferencia entre «educación», entendida como la transmisión de principios, valores y códigos de conocimientos, deslocalizado y ubicuo, y «enseñanza», sistema curricular y reglado de aprendizaje de conocimientos y saberes, bajo el diseño y control de instituciones de

enseñanza; la división de «conocimiento», como conjunto ordenado y asimilado de conceptos y nociones cognitivas en la mente del individuo, y «saber», que es una manifestación del conocimiento, por su aplicación para la toma eficaz de decisiones en la resolución de un problema o hipótesis determinado en un momento y ambiente concretos.

Desde un análisis comunicativo, la «comunidad cultural», también es importante abordar la *alteridad*, esto es, la representación del otro en el contexto comunicativo no solo como tolerancia a sus manifestaciones, sino como aceptación igualitaria, simbiótica y sinalagmática, en un contexto de globalización (Bernabé, 2012). El concepto de cultura y su terminología, por tanto, se ha enriquecido con las nociones y precisiones de *diversidad cultural*, como multiplicidad de expresiones culturales, *pluralismo cultural*, como instrumento de cohesión social por la diversidad cultural, *multiculturalidad* por coexistencia diferenciada de diversas culturas en un mismo espacio, *interculturalidad*, que da a la cultura un sentido dinámico (no un valor eterno inmutable), pues se entiende como la presencia e interacción equitativa de diversas culturas, con la posibilidad de generar expresiones culturales compartidas, por medio del diálogo y el respeto mutuo (Cantle, 2013).

2.3 Dimensión poliédrica del contenido en un corpus documental

Tradicionalmente, un texto quedaba manifestado y corporeizado en un «documento» bajo distintos soportes y tipologías. Las TIC, sin embargo, fueron añadiendo a esta noción inicial los conceptos de «documento electrónico», «documento digital», «documento virtual», hasta llegar al hipertexto (hipermedia si incorpora códigos audiovisuales de toda naturaleza) e hiperdocumento, expresión de unos contenidos que nos interesan especialmente a efectos de este trabajo.

Hipertexto puede entenderse como una estructura digital consistente en una red de nodos interconectados, que contiene conexiones predefinidas con otros documentos, pero su navegación y generación depende completamente de las necesidades y estrategias de su usuario. Esta noción del hipertexto como infraestructura tecnológica se ve implementada por el concepto de *hiperdocumento*, conjunto de documentos digitales soportados por una red hipertextual. El hiperdocumento sería un conjunto de «documentos» interconectados con un diseño de volumen poliédrico en el ciberespacio, un *objeto digital*, utilizado en diferentes contextos y códigos semióticos según los objetivos, las estrategias y las necesidades del usuario. Esta perspectiva prioriza el contenido (el almacenamiento de información en «documentos digitales», que contienen «nodos»), percibiendo el hiperdocumento como una función.

El concepto de hiperdocumento apuntado implica su naturaleza «dinámica» por integrar textos de diferentes códigos semióticos, como la escritura, imagen estática y móvil, y sonido (reclama un conjunto de signos que procesen su «imagen»), así como integrar enlaces e hipervínculos a los que remite y por los que es enlazado, como también estar compuesto por un conjunto de «átomos» o mosaico de información digital, que le convierten en reusable. Estas propiedades hacen que el hiperdocumento precise sus propios signos de representación, que no son otros que los *metadatos*, otorgados en el código fuente o adscritos externamente, así como sus propios vocabularios derivados de los lenguajes de marcado, a lo que se unen, naturalmente, las *nubes de etiquetas*, *tags* y demás formas de etiquetación de contenidos en redes sociales (Marzal y Gonzales, 2010).

Las potencialidades de la edición en web han ido impulsando con fuerza la inoculación de la imagen como elemento semántico fundamental, incorporando un lenguaje icónico por el desarrollo de la *3D*, *la realidad virtual y la realidad aumentada*. Desde el principio, la *virtualidad* en web fue una base de datos gráficos interactivos, explorable y visualizable en tiempo real mediante imágenes tridimensionales de síntesis, capaces de provocar la sensación de inmersión en la imagen mediante la visión estereoscópica y la *propioceptiva*. La Realidad Virtual supone incorporar la simulación multisensorial para la comprensión de procesos difícilmente visualizables. Este sistema se ha visto implementado por la Realidad Aumentada, que permite mejorar la

visión del mundo real recibiendo información adicional computerizada sobre el objeto y tarea en la que está trabajando, lo que implica la fusión de dos imágenes, una proporcionada por el vídeo en el entorno real y otra proporcionada por una videocámara enfocada al mundo virtual que se solapa con la imagen captada del mundo real. Sus interesantes aplicaciones se hacen patentes en software como LearnAr (<<http://www.learnar.org/>>), Wikitude World Browser (<<http://www.wikitude.com/>>), Layar (<<https://www.layar.com/>>), ARToolkit (<<http://artoolkit.org/download-artoolkit-sdk>>), Junaio, entre otros. Los *big data* están ampliando con fuerza los horizontes.

Estas potencialidades de edición en web han impulsado la conversión de los documentos de archivo y de los objetos museísticos en objetos digitales y virtuales en web, entre otros, lo que plantea el atractivo desafío de su difusión y diseminación en el ciberespacio, un desafío que precisa responder tanto a la «representación» en web (cómo se edita y publica el objeto museístico y documento de archivo), como a la «recuperación» (cómo es buscado y cómo capturarlo por los motores y buscadores en web). El problema, pues, se plantea dentro del ámbito de las Ciencias de la Documentación, la formulación, elaboración y uso de unos *vocabularios y lenguajes documentales* que permitan la representación y la recuperación de manera apropiada y eficaz. Estos vocabularios y lenguajes documentales deben responder a los «lenguajes de especialidad» de los dominios (archivístico y museístico) donde se aplicarán, por lo que un uso especializado de las fuentes para su elaboración es esencial.

3. Campo y objeto para las fuentes terminológicas en la construcción de tesauros

3.1 El corpus documental

La edición de un tesoro necesita un estudio de oportunidad, la determinación de un dominio bien definido (área de conocimiento, especialidad, etc.), adscripción a las normas universales, por tanto interoperables, de elaboración (para establecer la forma de presentación, partes, extensión, idioma, los tipos de relaciones, visualización, uso, gestión y mantenimiento, tipologías). El tesoro, en tanto que lenguaje documental convencional, arbitrario, coherente, postcoordinado, subidiomático, comienza con un inflexible diseño apriorístico. Esta naturaleza, que le convierte en un poderosísimo lenguaje documental, se asienta sobre un elemento esencial: el *descriptor*.

El descriptor es la parte sustantiva en la definición del tesoro, porque es producto de una metódica depuración semántica para su óptima efectividad en la recuperación y representación de contenidos. La depuración semántica reconoce unas fases para la relevancia y pertinencia semánticas: los textos son la suma de palabras, que sometidas a un análisis de especificidad semántica en un contexto determinado, permite establecer cuáles son las «palabras semánticamente ricas», que son categorizadas en asociaciones o clústeres semánticos. Cada clúster queda representado por una palabra y así se convierte en *palabra clave*. Cuando esta palabra clave, sometida a un proceso de homologación y normalización, es interpretada bajo el prisma de lenguajes de especialidad de un dominio determinado (glosarios, diccionarios, enciclopedias, etc.), se convierte en término de indización. Por fin, cuando estos términos de indización se ven sometidos a las reglas anteriormente apuntadas para la elaboración de tesauros, pero sobre todo se les dota de unas relaciones fijas y bien definidas, entonces llegamos al descriptor.

Este proceso, sin embargo, es un esfuerzo metódico que acometen los documentalistas desde sus técnicas, pero, por la misma razón, su funcionalidad no se hace visible a los usuarios y se origina una brecha entre la funcionalidad que tienen las listas terminológicas para unos y otros, lo que ha provocado una tendencia a determinar con eficacia la funcionalidad semántica en KO (Organización del Conocimiento) de listas de términos, taxonomías, tesauros y ontologías (Pieterse y Kourie, 2014).

De otro lado, el proceso de depuración semántica para lograr la coherencia del lenguaje mediante el control de vocabulario ha debido adaptarse, necesariamente, a la estructura horizontal y asociativa de la web. Si el método antes reposaba más en un sistema jerárquico, ahora se ensayan modos más «conectivistas» a través de un listado de palabras clave y términos, que se convierten en el fundamento de un tesoro, y solo después los descriptores pueden ser enlazados por metadatos con los códigos de la CDU, lenguaje clasificatorio por excelencia (Von Doorn y Polman, 2010).

El corpus documental como materia prima para la elaboración de lenguajes documentales también ha tenido que al «valor» de estos lenguajes en la web social y web semántica, especialmente referido a tres funcionalidades: a) la atención fundamental a las estrategias de búsqueda del usuario (pieza clave en la web), lo que ha llevado a una presencia cada vez mayor del lenguaje natural (por la facilidad de uso) por su mayor eficacia a causa de las TIC, pero también ha llevado a considerar las búsquedas como una expresión de «comportamiento», por lo que se hacen muy relevantes los análisis y estudios de las «consultas» como pilar para las búsquedas semánticas (Sancho; Fernández y Boulat, 2012); b) la evolución de los lenguajes documentales, entre ellos, el tesoro, desde su carácter de instrumento para recuperar información a recuperar conocimiento mediante un adecuado control de la gestión de los contenidos digitales (Junior; Carvalho y Azevedo, 2013); c) la expansión de la reusabilidad de los objetos digitales en web, para lo que necesitan instrumentos adecuados en lenguajes documentales, razón por la que su elaboración va derivando desde la palabra convertida en término-descriptor a la *extractación* de los contenidos digitales basándose en la frase como elemento básico del lenguaje documental, y, con este sesgo, surgen herramientas como GenEx, MLP, KP-Miner o KEA (Martín-Moncunill; García-Barriocanal; Sicilia y Sánchez-Alonso, 2015).

3.2 Los lenguajes del corpus

Algo que determina a los lenguajes documentales en general y a los tesauros en particular es su referencia a un *dominio* para su uso y aplicación; por ello, para la selección terminológica a partir del corpus, son por tanto muy útiles las *lenguas de especialidad*, esto es, el lenguaje que es propio de un área de conocimiento o profesión, cuyos términos pueden derivar de la lengua natural, pero que en esa área tienen una semántica específica. Estas «lenguas» aluden al lenguaje específico que utilizan los profesionales y expertos para transmitir información y, cuando es preciso, para refinar los términos, los conceptos y los saberes de una determinada área de conocimiento, confirmando los ya existentes, matizando el ámbito de su aplicación o modificándolos total o parcialmente; son, por tanto, variedades de una determinada lengua destinadas a cumplir un fin comunicativo en un marco profesional o académico concreto (Varó, 2006). Tienen como principales caracteres: a) un vocabulario muy singular; b) unas tendencias sintácticas y estilísticas muy idiosincrásicas; c) unos géneros profesionales propios e inconfundibles.

Las lenguas de especialidad, dentro del método de transformación de las palabras clave a término y su preparación y depuración semántica hacia descriptor, cumplen dos funcionalidades muy relevantes: su poder de desambiguación, ya que permiten establecer una relación semántica biunívoca muy poderosa entre la noción-concepto y una voz-palabra; y, dependiendo de la lengua de especialidad, la determinación de las áreas o zonas de mayor densidad semántica dentro de la organización de un texto.

La argumentación textual, modo de organizar el mapa mental del autor reflejado en un texto, junto con el valor indudable del análisis de los *conectores de discurso*, precisa detenerse en un análisis apropiado del *discurso electrónico*, con su propia sintaxis y gramática, derivadas también de los requisitos tecnológicos en las conexiones del discurso, en el que sus actores se comunican en entornos digitales, dentro de los que además el discurso incorpora imágenes audiovisuales, en su más lata tipología (Celis, 2012).

Esta percepción del discurso electrónico inclusiva de imágenes audiovisuales ha hecho considerar desde la óptica de la «lingüística de Corpus» al iconismo, el lenguaje icónico y su lectura propia. La lectura icónica es aquella que se practica sobre el hipermedio, entendido como aquel hiperdocumento cuya estructura hipertextual asocia imágenes en todas sus manifestaciones así como sonidos. La lectura icónica se orienta a comprender, asimilar, conocer y aprender en red su «discurso visual», en el que coexisten distintos códigos semióticos y no solo el textual. El corpus y sus lenguajes deben ampliarse, pues, a la eclosión y desarrollo de imágenes 3D, YouTube, Realidad Virtual y Aumentada, GIS y otros entornos emergentes donde la imagen y sonido son los vehículos básicos (incluso privativos) de los contenidos y conocimiento. Sobre el impacto de esta nueva lectura y entornos existen estudios interesantes, como los llevados a cabo por C. Ramón en 2014 en su tesis doctoral, L. Martínez en 2012 y E. G. Insuasty (2013), una interesante reflexión, recogida en su tesis doctoral, sobre la inoculación de esta lectura en los «objetos de aprendizaje» y, por ende, en la obtención de conocimiento.

Los contenidos digitales se convierten también en contenidos digitales icónicos, donde la imagen y sonido no apoyan (o no solo) al texto escrito, sino que la imagen audiovisual es el texto. Se hacen necesarios, entonces, nuevos modos para etiquetar semánticamente estos contenidos para representarlos, identificarlos, recuperarlos y extraerlos, configurándose en un lenguaje documental específico. Las técnicas documentales han debido evolucionar rápidamente, desde los métodos de análisis de la imagen en el texto propios del análisis de contenido, que reconocía un nivel identificativo (soporte, formato, extensión, presentación, luz, color, plano, ángulo, posición, derechos), un nivel descriptivo (contenido, descripción onomástica, toponímica y de materia), y otro nivel interpretativo con los mismos ítems del descriptivo pero desde una perspectiva de análisis contextual.

Un paso adelante, para el tratamiento educativo de la imagen la daría, en 1999, J. A. Ortega, en «La alfabetización visual y su incidencia en el desarrollo de capacidades perceptivo-cognitivas», donde establecía un análisis cognitivo de la imagen a partir del reconocimiento de cinco niveles de tratamiento: morfológico (percepción), morfosintáctico (reconocimiento de sus elementos), semántico (interpretativo), estético y ético. Un salto cualitativo muy importante lo dieron Gómez Díaz y Agustín Lacruz (2010), cuando propusieron el tratamiento de la imagen desde un punto de vista «documentalista», esto es, reconocer las capacidades y habilidades necesarias para tratar y comprender la imagen, en tanto que elementos alfabetizadores, como fundamento de conocimiento en la nueva sociedad y educación competencial. Profesionales y sus asociaciones, sea el caso de la Association of College & Research Libraries (ACRL), acusarían recibo del nacimiento de un nuevo compendio de técnicas y capacidades para el análisis y tratamiento de la imagen, la *Visual Literacy*, para las que editaría en 2011 las *Visual Literacy Competency Standards for Higher Education*, con definición de las normas y de indicadores *ad hoc*.

3.3 «Fuentes vehiculares» para la construcción de tesauros

En Análisis Documental siempre ha existido el debate sobre ventajas e inconvenientes del uso de herramientas de control terminológico para generar un lenguaje documental, o la aplicación del lenguaje natural, tratado con las nuevas posibilidades que ofrecen las TIC. Este debate se ha desplazado también a la construcción de tesauros en la web. En este debate, sin ser en modo alguno exclusivos tal como evolucionan las herramientas, se proponen dos vehículos muy útiles en esta construcción. La Terminología y el Procesamiento de Lenguaje Natural (PLN).

Los tesauros precisan de un sistema de significación cuyos elementos son un léxico (con términos depurados), una red paradigmática (con relaciones esenciales y estables entre los descriptores) y una red sintagmática (con unas relaciones contingentes destinadas a coordinar descriptores). La Terminología se muestra como un óptimo instrumento para la elección de palabras clave y su evolución a descriptor (por método de depuración semántica), la organización de los elementos y redes indicados, de estructuras funcionales de concepto, según aporta D. Fraga en 2011.

En palabras de M. Cabré (2000), la Terminología cumple una doble función: representar el conocimiento mediante el trabajo descriptivo (ilustración del discurso) y transmitirlo mediante el trabajo prescriptivo (garantiza unidades unívocas de expresión y comunicación). Una mirada más reciente y con un sesgo interesante hacia la «garantía» y la «cultura» como vectores de la Terminología en la Organización del Conocimiento lo ofrece M. Barité (2011).

Un uso eficaz del lenguaje natural lo proporciona el Procesamiento de Lenguaje Natural, entendido como un conjunto de modelos e investigaciones cuyo objetivo es generar instrumentos de lenguajes documentales capaces de reconocer el lenguaje natural. Su punto de partida es la definición de la *Lingüística de Corpus*, que selecciona una muestra de la lengua a la que se dota de objetivos definidos, mediante el etiquetado (evidencia morfosintáctica), parentizado (evidencia de sus constituyentes sintácticos) y análisis (evidencia de sus relaciones sintácticas). Para este cometido el PLN acude a técnicas de ingeniería lingüística orientadas a la desambiguación y al análisis sintáctico, atendiendo a: la *morfología*, por medio de un analizador morfológico con definición de *tagsets* (sistema de etiquetas) y de un desambiguador basado en reglas o estadísticas; a la *sintaxis*, bien por analizadores sintácticos parciales o superficiales (*chunkers*), bien por analizadores sintácticos completos en *treebank*; a las *redes semánticas* para interpretación de sentidos mediante ontologías o fuentes léxicas, una dimensión para la que ha tenido una función primordial WordNet.

El origen de WordNet se sitúa en las teorías psicoanalíticas y los principios de la Psicolingüística, que buscaban fundamentar un modelo y una base de datos léxico-conceptual, constituidos tanto por unidades léxicas como por las relaciones entre ellas y estructurados en forma de red semántica relacional, cuyo objetivo era generar un instrumento idóneo para la desambiguación de significados, el etiquetado semántico y la recuperación de información. WordNet se comporta como un diccionario digital, que integra términos, bien sean simples o compuestos, organizados en agrupaciones de sinónimos denominados *synsets* de glosas, insertando cada clase de términos en una glosa definicional. WordNet ha desarrollado una conectividad en tres sentidos: *nominalizaciones*, mediante el enlace a pares, morfológica y semánticamente, entre sustantivos y verbos; *desambiguación* de términos mediante *synsets*; *organización clusterizada de tópicos*, relacionando sentidos de palabras. Cada término tiene una breve descripción y en ocasiones una frase relacionada que muestra su empleo. Esta frase procede del Brown Corpus. En función del número de significados diferentes que tiene un término, hace un recuento denominado recuento de la polisemia.

La funcionalidad y útiles aplicaciones de WordNet proceden de su estructura en red léxica, con un alto número de relaciones entre términos y entre *synsets*: a) Relación Temática a partir de grandes agrupamientos por temas; b) Relación de equivalencia mediante relación entre términos sinónimos (*synsets*) y relaciones de similitud (permite obtener los *synsets* más parecidos y los *synsets* más frecuentemente relacionados con determinado término); c) Relaciones de jerarquía entre *synsets* por hiperonimia, hiponimia, meronimia y holonimia. WordNet incorpora como términos no solo al sustantivo, sino también a los adjetivos y, algo sumamente importante en la etiquetación de imágenes móviles, los verbos.

Estas funcionalidades han sugerido a WordNet como instrumento para la construcción automática de tesauros, bien basados en la co-ocurrencia por identificación estadística automática de relaciones semánticas, bien basados en la asociación predicado-argumento, identificando las palabras más significativas de un entorno (predicado) y aquellas con las que se relaciona para ser argumentadas. En todo caso, la desambiguación es la más abundante y variada aplicación de WordNet (Morato; Marzal; Llorens y Moreiro, 2004).

WordNet también está teniendo una importante aplicación en el desarrollo de tesauros visuales, como Visual Thesaurus (<https://www.visualthesaurus.com/>) para apoyar su sistema de navegación. Existen ya experiencias de uso de WordNet para apoyar tesauros cuyo objetivo es una indexación para recuperación eficaz de repositorios audiovisuales (Hollink; Malaisé y Schreiber, 2010), así como estudios sobre el modo de implementar la asociatividad de lenguajes tan jerárquicos como los tesauros con el apoyo de WordNet (Ali Hadj Taleb; Ben Aouicha y Ben Hamadou, 2014).

El desarrollo y aplicaciones del fenómeno de las redes sociales ha impulsado también la eclosión de la *etiquetación social* en las redes sociales, siendo particularmente útiles las taxonomías y folksonomías. En definición de M. Centelles (2005): «Una taxonomía es un tipo de vocabulario controlado en que todos los términos están conectados mediante algún modelo estructural (jerárquico, arbóreo, facetado...) y especialmente orientado a los sistemas de navegación, organización y búsqueda de contenidos de los sitios web». Las taxonomías son, por tanto, un sistema de clasificación que permite agrupar un conjunto de términos dentro de unas categorías predefinidas, contenidas unas en otras o relacionadas. Las folksonomías implican una categorización colaborativa mediante etiquetas simples, sin jerarquías ni relaciones previamente establecidas, cuyo primer objetivo es realizar búsquedas particularmente activas en las redes sociales, debido precisamente al etiquetado colaborativo y de iniciativa individual del usuario. Las folksonomías generan un «vocabulario» muy apto y adaptado a la naturaleza de los hiperdocumentos en su más lato sentido y carácter, en su representación, pero para la recuperación no supera el grave inconveniente de la absoluta ambigüedad semántica.

4. La indización como fuente terminológica en la construcción de tesauros

Establecido y tratado el corpus, la depuración semántica al descriptor, debe proceder necesariamente del «término», un producto de la indización, proceso y método que se convierte en muy relevante para la construcción de tesauros.

La aplicación de los tesauros en web obliga a reubicar las fuentes de indización para su construcción, dentro de un nuevo «mapa semántico», en atención a distintas dimensiones que debe contemplar la funcionalidad de los tesauros en web:

- Una adaptación a la horizontalidad, por enlaces y vínculos hiperdocumentales, por lo que hay una redefinición de las tres categorías de relaciones jerárquicas reconocidas en los tesauros (Alexiev; Isac y Lindenthal, 2015).
- La incorporación de la multiseccionalidad en los dominios, mediante el método de enumerar clasificatoriamente las características de contenido, para definir una estructura terminológica dentro de un dominio (Gubanov; Makarenco y Novikov, 2014).
- La adaptación a etiquetar la dinamicidad y asociatividad de las etiquetas de los contenidos a través de verbos (Macura y Lima, 2014).
- La socialización en la definición de etiquetas y términos, candidatos a descriptores, mediante el desarrollo de clústeres semánticos y participación societaria de la comunidad de usuarios (Gibert; Valls y Batet, 2014).
- En entornos web, la construcción de tesauros deben desarrollar los mecanismos necesarios para acometer la actualización del control de su corpus documental, tendiendo a fortalecer su efectividad por medio de la cooperación con otros lenguajes documentales a través de la interoperabilidad de sus elementos (Neelameghan y Lalitha, 2013).
- La «explosión» de los datos en la web y su identificación mediante metadatos, el fenómeno denominado *big data*, que exigen un adecuado tratamiento y disponibilidad, en aras de la transparencia en la gestión administrativa y de entidades mediante una adecuada gestión de contenidos digitales, un desafío al que se intenta dar respuesta con la inclusión de vocabularios controlados y tesauros, capaces de anotar los contenidos y generar metadatos (Wright; Harrison y Whatkins, 2015).

Los procesos de indización, evidentemente la automática, para la construcción de tesauros también se adaptan a estas nuevas tendencias. Junto con modelos lingüísticos y complejos procesos de automatización mediante modelos cognitivos, de representación vectorial o redes neuronales, la Indización Automática pivotaba preferentemente sobre modelos de representación probabilística, como la Indización Estadística de Términos por Frecuencias o IDF. Es un sistema de filtrado basado en la ley de Zipf, que establece que las palabras de mayor frecuencia

no suelen ser significativas, ya que la importancia de un término es proporcional a la frecuencia de ocurrencia en el documento e inversamente proporcional al número de documentos en que aparece. Esta técnica se complementa con el Método N-grams, que calcula la frecuencia no sobre cada término o palabra compuesta, sino con cadenas de caracteres de longitud predefinida y fija. Para lograr la obtención de relaciones se aplicaban los *Clasificadores científico-métricos*, mediante el método de Chen, que analiza el uso de grupos de palabras que aparecen simultáneamente en varios documentos, como también el método *Algoritmos K-vecinos*, que consiste en mover cada vector al clúster cuyo centroide esté más cercano y actualizar después los centroides de los clústeres.

Surgen, por tanto, tanto herramientas para indización, como motores de búsqueda apoyados en sistemas de indización que, a las funcionalidades arriba anotadas, unen otras conforme a los nuevos requisitos. Es el caso de Carrot2, un motor de búsqueda cuya recuperación de recursos de información adquiere una semblanza de Google, y para refinar las búsquedas ofrece unos clústeres de términos, así como permite una visualización de resultados por imágenes (<<http://search.carrot2.org/stable/search>>); un sistema de indización automática, Aduna AutoFocus (<<http://www.iskouk.org/sites/default/files/aduna.pdf>>), que permite una indización probabilística, pero categoriza los resultados por propiedades, casi hipertextuales, y posibilita recuperar mediante la visualización de clústeres y enlaces; sistemas para el tratamiento automatizado de los textos que producen y visualizan los términos resultantes, en una red semántica, como un mapa semántico visual, destacando fundamentalmente la asociatividad y creación de redes semánticas, caso de Texttexture (<<http://texttexture.com/>>) o Gephi, más complejo y polifacético en su aplicación (<<https://gephi.org/>>). El control de tendencias y usos de etiquetas y términos en redes sociales también parece ser un factor muy relevante, para el que pueden ser especialmente útiles herramientas como NodeXL (<<https://nodexl.codeplex.com/>>).

Las tendencias en los requisitos solicitados para las fuentes de construcción de tesauros se ven reforzadas, sin duda, por las aplicaciones más diversas y polifacéticas de los tesauros en la web: se convierten en una herramienta muy útil para la navegación (tesauro USGS, <<http://www.usgs.gov/>>), como herramienta de etiquetación y recuperaciones «integradas» de textos e imágenes, muy positivas para museos (como el Virtual Museum of Pacific, <<http://australianmuseum.net.au/blogpost/science-bytes/virtual-museum-of-the-pacific>>), como excelentes instrumentos en gestión de contenidos digitales, por su «exportación» a la web para bibliotecas digitales y virtuales (como Tematres, o de Omnigator en el caso de su apoyo como topic map).

Es especialmente interesante la evolución aplicativa de los tesauros a modo de lenguaje documental icónico idóneo, como atestigua la investigación, que sigue tres líneas:

- La *anotación de imágenes* por etiquetación social, un crowdfunding terminológico, especialmente efectivo en las recuperaciones (Ionescu; Popescu; Radu; Müller, 2014). Esta etiquetación en «socialmedia» y en el contexto evolutivo de la Web, reclama que las anotaciones sean «enlazadas», razón por la que existen ya algunas experiencias para un «linked tag», como sistema semiautomático de anotación enlazada (Im; Park, 2014), reforzando así la tendencia en hacer eficaz la compartición de información de imágenes que proporcionan las folksonomías (Jeong; Hong y Lee, 2013).
- La *construcción de un lenguaje visual*, capaz de lograr un alto nivel de representación, mediante una analogía con el sistema de recuperación textual (El-Sayad; Martinet; Urruty y Djeraba, 2012).
- Una vía muy interesante para la *etiquetación de vídeos*, especialmente interesante por el menudeo de tutoriales, visitas virtuales, YouTube y el auge del fenómeno de la Khan Academy, para un uso didáctico de los vídeos de YouTube (Rudinac; Larson y Hanjalic, 2012).

Bibliografía

- ALEXIEV, V.; ISAC, A., y LINDENTHAL, J. (2015): «On Compositionality ISO 25964 Hierarchical Relationships (BTG, BTP, BTD)», en *Networked Knowledge Organization Systems (NKOS) Workshop*. Londres: DL2014 Conference.
- ALI HADJ TALEB, M.; BEN AOUICHA, M., y BEN HAMADOU, A. (2014): «A new semantic relatedness measurement using WordNet features», *Knowledge Information Systems*, 41, pp. 467-497.
- BARITÉ, M. (2011): «La garantía cultural como justificación en sistemas de organización del conocimiento: aproximación crítica», *Palabra Clave*, 1 (1), pp. 2-11.
- BERNABÉ, M. (2012): «Pluriculturalidad, multiculturalidad e interculturalidad, conocimientos necesarios para la labor docente», *Revista Educativa Hekademos*, n.º 11, pp. 67-76.
- CABRÉ CASTELLVÍ, M. (2000): «Terminología y Documentación», en *Documentación, Terminología y Traducción*. Edición de C. GONZALO GARCÍA. Madrid: Síntesis, pp. 31-82.
- CANTLE, T. (2013): «Interculturalism as a new narrative for the era of globalisation and super-diversity», en *Interculturalism and multiculturalism: similitudes and differences*. Edición de M. BARRETT. Estrasburgo: Consejo de Europa.
- CARR, N. (2011): *Superficiales: ¿Qué está haciendo Internet con nuestras mentes?* Madrid: Taurus.
- CELIS, M. A. (2012): *Patrones de comportamiento lingüístico en el discurso electrónico, una aproximación teórico-práctica con el análisis de dos obras literarias hipertextuales: Gabriella Infinita de Jaime Alejandro Rodríguez y Tierra de Extracción de Doménico Chiappe*. [Tesis Doctoral inédita]. Ciudad Real: Universidad de Catilla-La Mancha (Facultad de Letras).
- CENTELES, M. (2005): «Taxonomías para la categorización y la organización de la información en sitios web», *Hipertext.net*, 3, [en línea]: <<https://www.upf.edu/hipertextnet/numero-3/taxonomias.html>>. [Consulta: 9 de marzo de 2016].
- CEREZO (2011): «Identidad digital y reputación online», *Cuadernos de Comunicación Evoca*, n.º 5, <<http://www.evocaimagen.com/cuadernos/cuadernos5.pdf>>. [Consulta: 22 de febrero de 2016].
- DOORN, M. VON, y POLMAN, K. (2010): «From Classification to Thesaurus ...and back? Subject Indexing Tools at de Library of the Afrika-Studiecentrum Leiden», *Knowledge Organization*, 37 (3), pp. 203-208.
- EL-SAYAD, I.; MARTINET, J.; URRUTY, T., y DJERABA, CH., (2012): «Toward a higher level visual representation for content-based image retrieval», *Multimedia Tools and Applications*, 60 (2), pp. 455-482.
- GIBERT, K.; VALLS, A., y BATET, M. (2014): «Introducing semantic variables in mixed distance measures: Impact on hierarchical clustering», *Knowledge Information Systems*, 40, pp. 559-593.
- GÓMEZ DÍAZ, R., y AGUSTÍN LACRUZ, M. C. (eds.) (2010): *Polisemias visuales. Aproximaciones a la alfabetización visual en la sociedad intercultural*. Salamanca: Ediciones Universidad de Salamanca.
- GUBANOV, D. A.; MAKARENKO, A. V., y NOVIKOV, D. A. (2014): «Analysis methods for the terminological structure of a Subject area», *Automation and Remote Control*, 75 (12), pp. 2231-2247.
- HOLLINK, L.; MALAISÉ, V., y SCHREIBER, G. (2010): «Thesaurus enrichment for query expansion in audio-visual archives», *Multimedia Tools and Applications*, 49 (1), pp. 235-257.
- IM, D-H, y PARK, G-D. (2015): «Linked tag: image annotation using semantic relationships between image tags», *Multimedia Tools and Applications*, 74, pp. 2273-2287.
- INSUASTY, E. G. (2013): «Lectura y lecturabilidad icónica en objetos de aprendizaje soportados por plataformas virtuales», *Teoría de la educación*, 25 (1), pp. 239-241.
- IONESCU, B.; POPESCU, A.; RADU, A., y MÜLLER, H. (2014): «Result diversification in social image retrieval: a benchmarking framework», *Multimedia Tools and Applications*, 75 (2), pp. 1301-1331.

- JEONG, J-W; HONG, H-K., y LEE, D-H. (2013): «I-TagRanker: an efficient tag ranking system for image sharing and retrieval using the semantic relationships between tags», *Multimedia Tools and Applications*, 62, pp. 451-468.
- JUNIOR, J.; CARVALHO, R., y AZEVEDO, A. (2013): «Da recuperação da informação à recuperação do conhecimento: reflexões e propostas», *Perspectivas em Ciencia da Informacao*, 18 (4), pp. 2-17.
- MACURA, B. C.; LIMA, G.A.B.O. (2014): «Relacionamentos em tesauros: o valor semântico dos verbos», *Perspectivas em Ciencia da Informacao*, 19 (4), pp. 182-201.
- MALERBA, F.; NELSON, F.; ORSENIGO, L., y WINTER, S. (1999): «History friendly models of industry evolution: the computer industry», *Industrial and Corporate Change*, 8 (1), pp. 3-40.
- MARTÍN-MONCUNILL, D.; GARCÍA-BARRIOCANAL, E.; SICILIA, M. A. Y SÁNCHEZ-ALONSO, S. (2015): «Evaluating the Practical Applicability of Thesaurus-Based Keyphrase Extraction in the Agricultural Domain: Insights from the VOA3R Project», *Knowledge Organization*, 42 (2), pp. 76-89.
- MARZAL, M. A., y GONZALES, A. (2010): «Del documento al hiperdocumento: una visión “funcional” de un concepto en evolución». *Revista Signo y Pensamiento*, vol. 29, n.º 57, pp. 84-99.
- MORATO, J; MARZAL, M. A; LLORENS, J., y MOREIRO, J. (2004): «WordNet applications», en *GWC 2004: Proceedings of the Second International WordNet Conference*. Edición de P. SOJKA; K. PALA; P. SMRZ; CH. FELLBAUM Y P. VOSEN. Brno: Masaryk University, pp. 270-278.
- NEELAMEGHAN, A., y LALITHA, S. K. (2013): «Multilingual Thesaurus and Interoperability», *Journal of Library and Information Technology*, 33 (4), pp. 289-294.
- PIETERSE, V., y KOURIE, D. G. (2014): «Lists, Taxonomies, Lattices, Thesauri and Ontologies: Paving a Pathway through a Terminological Jungle», *Knowledge Organization*, 41 (3), pp. 217-229.
- RUDINAC, S.; LARSON, M., y HANJALIC, A. (2012): «Leveraging visual concepts and query performance prediction for semantic-theme-basic video retrieval», *Multimedia Tools and Applications*, 1, pp. 263-280.
- SANCHO, A.; FERNÁNDEZ, C., y BOULAT, P. (2012): «La búsqueda de información jurídica: de los tesauros a la inteligencia artificial», *Scire*, 18 (1), pp. 73-83.
- SIEMENS, G. (2010): *Conociendo el Conocimiento*. [s. l.]: Nodos Ele.
- SOWA, J. (2000): *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove: Brooks-Cole.
- VARÓ, E. A. (2006): «La sociedad del conocimiento, marco de las lenguas profesionales y académicas», *Las lenguas profesionales y académicas*. Barcelona: Ariel, pp. 3-12.
- WRIGHT, D.; HARRISON, K., y WHATKINS, J. (2015): «Automated tagging of environmental data using a novel SKOS formatted environmental thesaurus», *Earth Science Information*, 8, pp. 103-110.

Una ontología para la gestión del conocimiento: el Tesouro de la UNESCO y su aplicación en el Portal sobre Aprendizaje «Learning Portal» del Instituto Internacional de Planeamiento de la Educación de la UNESCO

Asunción Valderrama Pareja

UNESCO - Instituto Internacional de Planeamiento de la Educación

Cuando en 1946 se crea la UNESCO, la humanidad acaba de salir de uno de los peores conflictos de su larga historia. En noviembre de 1945, justamente al concluir la guerra, una Conferencia de las Naciones Unidas para el establecimiento de una organización educativa y cultural (ECO/CONF) reúne a los representantes de unos 40 Estados. Con el impulso de Francia y del Reino Unido, dos países muy afectados por el conflicto, los delegados deciden crear una organización destinada a instituir una verdadera cultura de paz. Dentro de su espíritu, esta nueva organización debe establecer la «solidaridad intelectual y moral de la humanidad» y, de esta manera, impedir que se desencadene una nueva guerra mundial.

En el momento de su creación la UNESCO va a recibir un amplísimo mandato, el de «Contribuir a la conservación de la paz y de la seguridad estrechando, mediante la educación, la ciencia y la cultura, la colaboración entre las naciones con el fin de asegurar el respeto universal de la justicia, de la ley, de los derechos humanos y de las libertades fundamentales para todos, sin distinción de raza, de sexo, de idioma o de religión, que la Carta de las Naciones Unidas reconoce a todos los pueblos». Este mandato queda resumido en el preámbulo de su Constitución: «Puesto que las guerras nacen en la mente de los hombres, es en la mente de los hombres donde deben erigirse los baluartes de la paz».

Durante sus setenta años de existencia la UNESCO ha funcionado como un laboratorio de ideas, un organismo normativo para lograr acuerdos mundiales en temas éticos, un centro de intercambio de información, un organismo de desarrollo de capacidades en los Estados miembros y un catalizador de la cooperación internacional. Se han añadido a su mandato inicial las ciencias sociales y humanas así como la comunicación y la información. Su misión hoy consiste en contribuir a la consolidación de la paz, la erradicación de la pobreza, el desarrollo sostenible y el diálogo intercultural mediante la educación, las ciencias, la cultura, la comunicación y la información. La Organización se centra particularmente en dos grandes prioridades: África y la igualdad entre hombres y mujeres; y en una serie de objetivos globales: lograr la educación de

calidad para todos y el aprendizaje a lo largo de toda la vida; movilizar el conocimiento científico y las políticas relativas a la ciencia con miras al desarrollo sostenible; abordar los nuevos problemas éticos y sociales; promover la diversidad cultural, el diálogo intercultural y una cultura de paz; construir sociedades del conocimiento integradoras recurriendo a la información y la comunicación.

Es interesante observar cómo en un plazo tan corto (si lo comparamos con la historia secular de las instituciones culturales representadas en este curso) los términos que se han ido utilizando al cabo de los años para definir la misión de la UNESCO se han ido modificando para adaptarse a los cambios de la sociedad. Las referencias al desarrollo sostenible y al diálogo intercultural son una clara muestra de esta evolución así como las metas de la erradicación de la pobreza o de la igualdad entre hombres y mujeres. En 1946 estas realidades ya existían pero, o no se tenía conciencia de ellas (como en el caso del desarrollo sostenible), o no se les daba la misma importancia (la igualdad entre hombres y mujeres es una de las dos prioridades de la Organización, mientras que la mención «sin distinción de sexo» no es más que una distinción entre otras en 1946), o no se imaginaba siquiera que se pudieran alcanzar (como en el caso de la erradicación de la pobreza). El elemento que no ha variado es la cobertura geográfica, que desde el principio fue la mundial, y la pluralidad de temas cubiertos por el mandato de la UNESCO, que ha ido incluso aumentando conforme se le añadían nuevas áreas, como la comunicación y las ciencias de la información¹.

El Tesoro de la UNESCO

Para poder describir los documentos generados por la UNESCO, se necesita un vocabulario enciclopédico y elaborado desde un punto de vista internacional e intercultural. A partir de 1969 se recopiló una lista de descriptores no estructurada partiendo de términos seleccionados en el Macrotesauro de la OCDE y en los documentos fundamentales de la UNESCO. Esta lista constituyó la base de la edición provisional del Tesoro de la UNESCO, que se terminó de elaborar en abril de 1974 y que contaba con unos 5000 términos. Se repartió esta lista entre los especialistas de los distintos programas, que, como era de esperar por parte de especialistas para quienes su materia nunca queda lo suficientemente bien reflejada, la ampliaron considerablemente haciéndole alcanzar unos 11 500 términos. La primera edición se publicó en inglés en 1977, en francés en 1983 y en español en 1984; pero la evolución rápida de la terminología empleada en los documentos y las necesidades de los usuarios (cuyo número iba creciendo a la vez que se diversificaba su origen) llevaron a revisarlo en profundidad en 1995. Se añadieron términos, se suprimieron muchos, se orientaron los poco utilizados hacia términos preferidos, se multiplicaron las notas explicativas y de alcance y se revisaron las relaciones de jerarquía y de asociación. Pero, sobre todo, se aplicó un enfoque que reflejaba menos las orientaciones de la investigación pura y que a partir de entonces describía el contenido de forma más pragmática, y se llegó a reducir el Tesoro a unos 6600 términos, pasando de dos volúmenes por cada idioma a un solo volumen que contenía las tres versiones lingüísticas, el famoso volumen rojo que sigue prestando servicios en algunas bibliotecas [1_Tesoro_Portada]. En esa misma época se crea ya una base de datos del Tesoro (en CDS-ISIS) y en el año 2000 aparece la primera versión *online* en inglés (web-ISIS), seguida por las versiones en francés (2003), español (2004) y ruso (2005).

¹ Una prueba visual de esta variedad se puede apreciar gracias al proyecto Mosaic of Change <<http://mosaicofchange.org>> y en particular disfrutando de las imágenes de la proyección <<https://vimeo.com/146507807>> a partir del minuto 3.

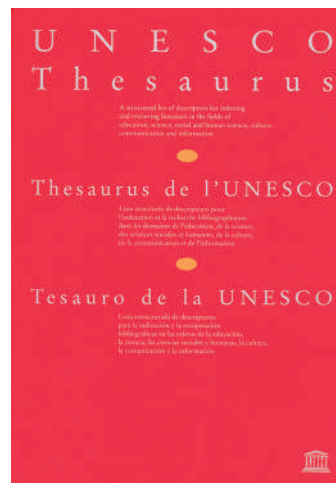


Figura 1. Tesauro de la UNESCO, edición de 1995.

El Tesauro de la UNESCO abarca los principales campos del conocimiento, pero ha sido diseñado para servir como instrumento de referencia para servicios de información que quieran elaborar sus propios vocabularios para responder a necesidades particulares. De allí que se haya utilizado y se siga utilizando en el mundo entero tanto en bibliotecas generales como especializadas². Es multidisciplinario, abarca las siete áreas principales del conocimiento incluidas en el mandato de la UNESCO: educación, ciencia, cultura, ciencias humanas y sociales, información y comunicación, y política, derecho y economía. A estas entradas hay que añadir un último grupo: los nombres de países y grupos geográficos, políticos o lingüísticos. Estas temáticas se dividen a su vez en 88 microtesauros. [2_pantallazo 1] Es monojerárquico excepto para los descriptores geográficos, ya que una misma entidad puede pertenecer a varios grupos (por ejemplo, Burkina Faso pertenece a los grupos África Occidental, África de habla francesa, Sahel y País islámico) [3_pantallazo 2].

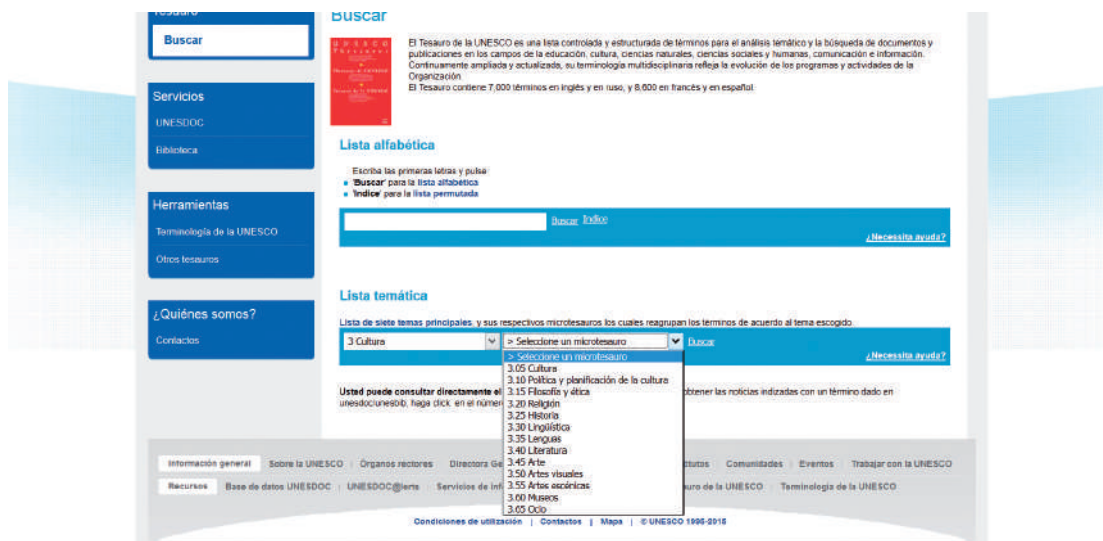


Figura 2. Microtesauros asociados al tema «Cultura».

² Véase en Garrod (2000) un ejemplo de las ventajas del Tesauro de la UNESCO en relación con otras clasificaciones.



Figura 3. Estructura multijerárquica de los descriptores geográficos.

El Tesoro de la UNESCO contiene: 4500 conceptos y 7075 términos en inglés y en ruso (8613 términos en francés y 8629 en español), unas 640 notas de alcance y 17050 relaciones. Las relaciones entre términos son: las de equivalencia entre un descriptor y un no-descriptor (UP/EMP); las jerárquicas (TG/TE); las de relación asociativa entre términos (TR) y la de inclusión en un microtesoro (MT). Además de las definiciones y notas de alcance, el Tesoro de la UNESCO contiene las equivalencias lingüísticas en los cuatro idiomas mencionados excepto para los no-descriptores.

Se ha ido ampliando y actualizando regularmente y hoy en día la versión disponible en la web recibe 15000 visitas mensuales, de las cuales más del 70% para las páginas en español. En el 2016 se pondrá a disposición del público una versión compatible con la norma ISO 25964 y que se publicará como Linked Open Data. Podrá incluir contenido en alfabeto chino y árabe (compatible UTF-8) y permitirá importar y exportar contenido en formatos SKOS, ZTHES y RDF. Después de muchos años de espera, este proyecto se ha podido al fin concretizar y se podrá acceder a la plataforma SKOS desde el sitio web del Tesoro <<http://vocabularies.UNESCO.org>>. El alineamiento del Tesoro con otros vocabularios y recursos está también previsto. Este recurso que tantos servicios ha prestado en tan distintos centros y lugares del mundo va a permitir nuevas realizaciones en la web semántica y continuar contribuyendo al enriquecimiento de la búsqueda de información.

El «Learning Portal» del Instituto Internacional de Planeamiento de la Educación:

Sin esperar a la publicación de la versión SKOS y basándose en el desarrollo que el equipo de la Universidad de Murcia ha publicado en 2013³, el Instituto Internacional de Planeamiento de la Educación de la UNESCO (IIEP) ha lanzado en 2015 un portal de recursos sobre la mejora de los logros educativos y la importancia del planeamiento en este proceso. El «Portal sobre el Aprendizaje» (o Learning Portal) del IIEP es una ventanilla única que permite acceder a información completa, actualizada, relevante y neutral sobre el aprendizaje, desde la educación pre-primaria hasta la secundaria. El IIEP tiene una larga experiencia en la creación y mantenimiento de portales temáticos, siendo los más conocidos *Planipolis* <<http://planipolis.iiep.UNESCO.org>>, un portal internacional de planes y políticas educativas y el *Servicio de Información sobre VIH, Salud y Educación* <<http://hivhealthclearinghouse.UNESCO.org/>>.

³ <<http://skos.um.es/unescothes/?l=es>>

un servicio de facilitación e intercambio de información para desarrollar políticas, programas y acciones de promoción en el área del VIH y SIDA, de la salud escolar y de la educación sexual. Cuando se planteó la posibilidad de crear un nuevo portal temático para cubrir el déficit de información neutral sobre el aprendizaje, se decidió utilizar las posibilidades tecnológicas que brinda la web semántica para desarrollar esta nueva herramienta que es mucho más que una biblioteca de más de 1300 recursos reunidos en una base de datos con un motor de búsqueda.

El Portal sobre el Aprendizaje se ha diseñado para responder a las necesidades de un público muy variado: los planificadores, responsables educativos, actores de la sociedad civil (padres de alumnos, periodistas, activistas) y donantes de todo el mundo; ofreciendo:

- Investigaciones e informes sobre los trabajos realizados para mejorar el aprendizaje, ejemplos de políticas llevadas a cabo sobre el tema, debates actuales y una amplia gama de experiencias sobre las cuestiones del aprendizaje.
- Resúmenes sobre investigaciones y estudios realizados en torno a formas de mejorar el aprendizaje.
- Una visión panorámica sobre los pasos a dar para elaborar un plan para el aprendizaje mejorado.
- Herramientas y métodos para monitorear el aprendizaje y utilizar de forma práctica los datos.
- Un blog semanal y una selección diaria de noticias y de artículos sobre el aprendizaje de todo el mundo.
- Un canal para conocer los temas de debate más importantes y participar en el e-foro.
- Un glosario de los términos más importantes y la posibilidad de poder pedir ayuda a un(a) bibliotecario(a) para encontrar recursos.

Para conseguir dar un acceso fácil y rápido a todas estas distintas informaciones, se decidió interconectar los distintos vocabularios estructurados que ya se venían utilizando en la Biblioteca del IPE, es decir:

- **Tesoro de la UNESCO** -> más de 8600 términos, entre descriptores y no-descriptores en los tres idiomas del portal (inglés, francés y español).
- **Clasificación del IPE** -> 435 temas. Esta clasificación se ha ido elaborando a lo largo de los años y permite a la Biblioteca del IPE poder estructurar conceptos de una manera más coherente con los distintos programas en los que trabaja el Instituto y que pueden no verse reflejados en el Tesoro. Este sistema ha permitido conservar la armonización con el Tesoro manteniendo al mismo tiempo una flexibilidad en la indización. Los temas de la clasificación del IPE suelen ser más genéricos que los descriptores. Un recurso se indiza con dos o tres temas como mucho y con entre cinco y quince descriptores.
- **Lista de proyectos** -> más de 600 entradas que corresponden a programas (Educación Para Todos), evaluaciones (PISA), sistemas de información (EMIS), leyes, etc. Se han normalizado las formas y los acrónimos.
- **Lista de instituciones** -> se han extraído y normalizado más de 1200 entidades para permitir acceder fácilmente a todos los recursos producidos por los distintos departamentos de estas instituciones, en particular en el caso de los ministerios y de las agencias internacionales para los que pueden existir más de 50 formas distintas para una misma entidad principal.
- **Lista de países** separada de los descriptores -> el Tesoro de la UNESCO incluye los países como una de sus subdivisiones, pero esta presentación no permite ofrecer una búsqueda específica por faceta. A partir de esta se pueden también crear agrupaciones *ad hoc* (como, por ejemplo, «países participantes en PISA») y es la que genera los matices de colores que se aprecian en el mapa que figura en esta página: <<http://learningportal.iiep.unesco.org/en/search/notice>>.

La interconexión entre estos distintos vocabularios se ha hecho con la solución ITM de Mondeca. Como se trata de un sistema propietario, se ha diseñado una arquitectura relativamente compleja pero que permite, en caso de querer recuperar la solución de entrada y el motor de indexación para otro proyecto, disociar los componentes de Mondeca y pasar del módulo de entrada de datos (Records management) a la indexación por SolR sin enriquecer la información con el tratamiento de ITM y a un portal de visualización (End-user portal). [4_Eschema_ITM_SolR_IRI]

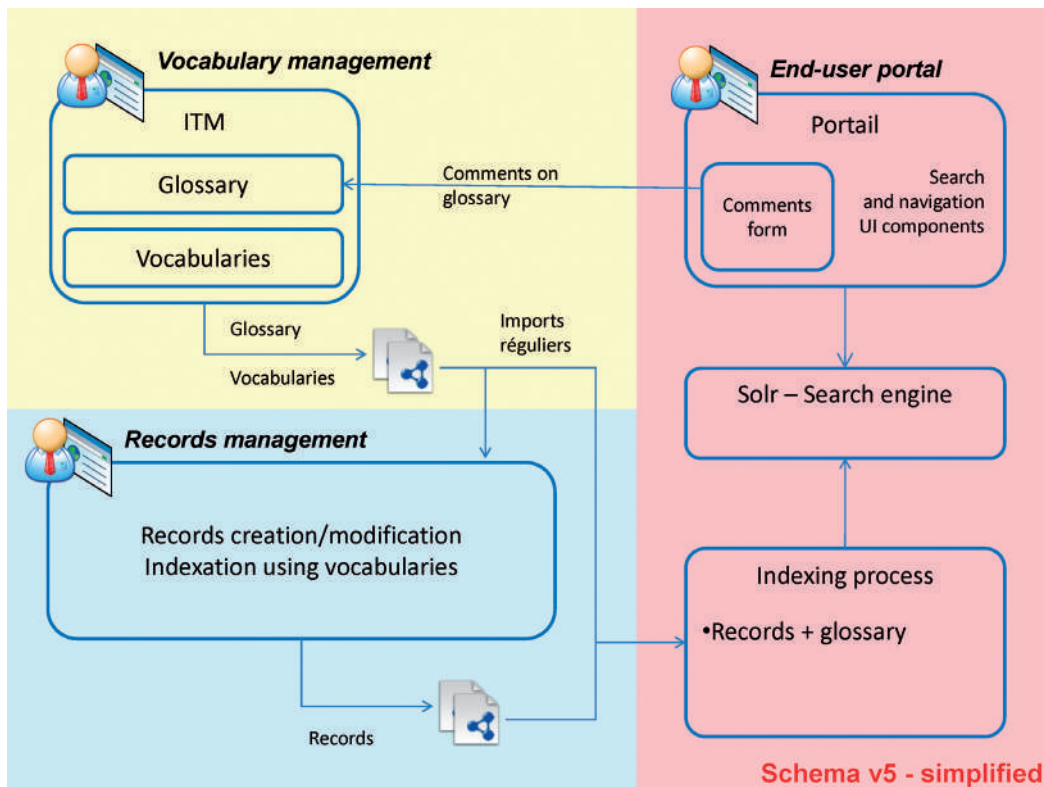


Figura 4. Esquema del sistema de entrada, enriquecimiento e indexación.

Un sistema de alineamiento de vocabularios basado en el Tesauro

A todas estas listas que ya se venían utilizando en la Biblioteca se decidió añadir un Glosario de casi 500 términos, extraídos de publicaciones o del material de formación del IIPE. Esta lista, al no estar estructurada jerárquicamente, se puede actualizar rápidamente, ya que se tiene más libertad para incluir nuevas entradas o modificar las existentes. Constituye la puerta de entrada en lenguaje «casi natural» para el portal al reunir las expresiones usadas más frecuentemente en los textos publicados por el IIPE. Cada entrada del glosario se ha alineado con uno o varios descriptores del Tesauro así como con un índice de la clasificación del IIPE [5_pantallazo 3]. De esta manera se ha alineado la Clasificación del IIPE con el Tesauro, pasando por el Glosario y evitando el tener que seleccionar entre los más de 8600 términos que este contiene, aquellos para los que tendría sentido una interconexión con la Clasificación. Este procedimiento se llevó a cabo manualmente en un periodo de dos meses (Chick, 2014), tras haber evaluado varias posibilidades y habiendo descartado un tratamiento semiautomático, que, de todas formas, hubiera necesitado una revisión posterior.

The screenshot displays a web application interface for a glossary. On the left, there is a sidebar with two sections: 'SUBJECT' containing 'Home', 'Alert management', 'Terminology management', and 'Workflow management'; and 'VOCABULARY' containing 'Countries', 'Tesauro de la UNESCO', and 'Theme Classification'. Below this is a 'NAVIGATION' section with a list of items including 'Home', 'SACMEQ', and 'Transferencia/traslado...'. The main content area is titled 'Transferencia/traslado de docentes/profesores' and shows an entry for 'Movilidad del docente'. The entry is structured with 'Attributes' and includes:

- corresponding concept:** Movilidad del docente
- definition:** Decision by the competent authority to assign teachers to an educational institution other than the one they are currently working at. (English), Decisión de la autoridad competente de asignar profesores a un centro educativo distinto de aquel en el que trabajan normalmente. (Spanish), and Décision de l'autorité compétente d'assigner des enseignants dans un établissement d'éducation autre que celui dans lequel ils travaillent habituellement. (French).
- definition source:** Eurydice, 2009. Thesaurus for Education Systems in Europe. Brussels: European Commission. (English), Eurydice, 2009. Tesauro Europeo de los Sistemas Educativos. Bruxelles: European Commission. (Spanish), and Eurydice, 2009. Thésaurus européen des systèmes éducatifs. Bruxelles: European Commission. (French).
- domain:** 470E - Condiciones de empleo docente
- example of use:** Teacher transfers have not been common in Indonesia, so the adoption of effective transfer systems at district level would be crucial. For instance, the Education Office in the district of Gorontalo identified 634 of its 5,000 teachers who could be redeployed, and implemented measures such as merging small schools, introducing multigrade teaching in schools with fewer than 50 students and providing incentives for teachers to move to remote schools (UNESCO, 2014: 253). (English), and El traslado de docentes no ha sido habitual en Indonesia, de modo que la adopción de un sistema eficaz de traslados a nivel de distrito sería de crucial importancia. Por ejemplo, la oficina de educación del distrito de Gorontalo individualizó a 634 de sus 5.000 docentes que podrían ser redistribuidos, y puso en marcha medidas como la fusión de escuelas pequeñas, la...

 On the right side of the main content area, there is an 'ACTINGS' menu with options like 'Edit', 'Edit in new window', 'Remove', 'Change type', 'Search', 'Metadata', 'Graph', 'Export xml', 'Export html', 'See history', 'New', and 'Duplicate from'. Below this is a 'CLIPPING' section with 'Select', 'Clear', and 'Deprecate' options.

Figura 5. Entrada del glosario conectada con el Tesauro (*corresponding concept*) y con la Clasificación (*domain*).

En un primer momento se pensó incluso en abandonar el uso del Tesauro, pero, tras discutir con Mondeca y con la agencia web que desarrolló el portal final, las ventajas de este instrumento quedaron patentes tanto para los que iban a desarrollar la aplicación como para el usuario final. Para el personal de la Biblioteca, muy acostumbrado al uso del Tesauro, este permitía además cubrir temáticas cuyo campo era en algunos casos más amplio que los habitualmente cubiertos por el IIPE. Quedaba, sin embargo, bien claro que el lenguaje del Tesauro, al alejarse en muchos casos del lenguaje natural, podría entorpecer la búsqueda, tratándose en particular de un portal que no está pensado sólo para especialistas. Para el etiquetado de las páginas de contenido, se decidió utilizar la Clasificación del IIPE, por su número más reducido de términos y su carácter más genérico [6_pantallazo 4]. El software permite también enlazar el contenido de las páginas con definiciones del glosario [7_pantallazo 5]. Al existir una interconexión entre las definiciones del Glosario, los índices de la Clasificación y los descriptores correspondientes del Tesauro, este sistema enriquece todas estas informaciones con las relaciones UP, TE y TG a las que están asociados los descriptores. Para resumir, partiendo de un etiquetado relativamente sencillo y leve de implementar como el que se consigue con la Clasificación y el Glosario, se añade la cadena semántica mucho más compleja y elaborada del Tesauro. Este tratamiento más sencillo se aplica a las distintas páginas de contenido, pero a los recursos de tipo bibliográfico (artículos, documentos, informes, etc.) se les sigue describiendo con una indización completa con descriptores [8_pantallazo 6]. Las búsquedas que se hacen en lenguaje natural utilizando el módulo sencillo devuelven resultados en la Biblioteca, en las páginas web y en el Glosario, 9_pantallazo 7]. La búsqueda en la biblioteca se hace en lenguaje natural o a través de facetas para quien quiera refinar los resultados.



Figura 6. Etiquetado de una página de contenido.

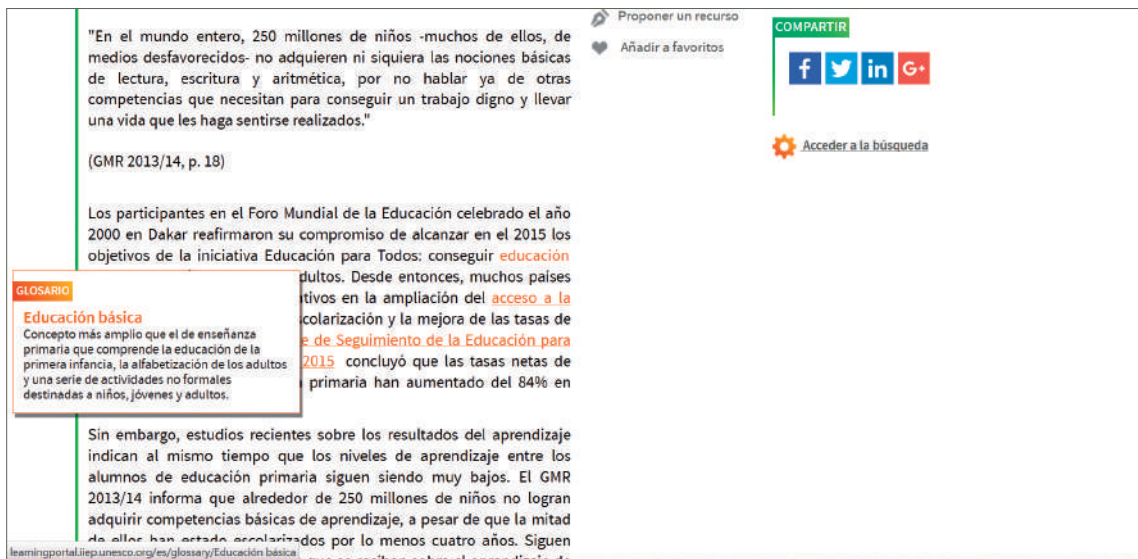


Figura 7. Enlace entre contenido de una página y definición del glosario, en este caso «educación básica».

Buscar en la biblioteca

DESCARGAR
Versión electrónica

Añadir a favoritos

AUTORES	ENTIDADES	PUBLICADO POR	COLECCIÓN	VOLUMEN	AÑO
Hart, Barbara C.	EBEAL	PREAL	Documentos	043	2009

PÁGINAS	IDIOMA	DISPONIBLE EN	PAÍSES	ISSN
36 p.	English	Spanish	América Latina	0719-6002

TEMAS

Perfeccionamiento docente

PALABRAS CLAVE

Actividad del docente, Actividad laboral, Ambiente educacional, Eficacia del docente, Formación de docentes, Inspección educativa, Motivación, Rendimiento escolar

Añadir filtros

See the most relevant results for the selected criteria

Perfeccionamiento docente

CONTRONERSA

Teacher Incentives

perfeccionamiento docente

LABELS SYNTHESIS

Códigos de conducta del docente

perfeccionamiento docente

BIBLIOTECA

Learning how to teach: the upgrading of unqualified primary teachers in sub-Saharan Africa: lessons from Tanzania, Malawi, and Nigeria

perfeccionamiento docente

BIBLIOTECA

Teacher certification in

Figura 8. Descripción de un documento en la biblioteca de recursos.

Importancia del aprendizaje | Mejorar el aprendizaje | Planificar para el aprendizaje | Seguimiento del aprendizaje

Buscar en la biblioteca

Añadir filtros

Añadir criterios:

- TEMAS
- PALABRAS CLAVE
- PAÍSES
- INSTITUCIONES
- AUTORES
- IDIOMAS
- TIPO DE DOCUMENTO
- PROYECTOS
- SEMINARIOS

2004 2015

200 elementos responden a uno o varios de los criterios seleccionados:

BIBLIOTECA (200) | SITIO WEB (24) | GLOSARIO (5)

BIBLIOTECA

Educadores de calidad: estudio internacional sobre las competencias y estándares para docentes

2011

Figura 9. Resultados de una búsqueda en lenguaje natural sobre la palabra «docentes» (200 respuestas en la Biblioteca, 24 páginas web y 5 entradas en el Glosario).

El Portal sobre el Aprendizaje del IPE es un proyecto ambicioso, ya que se ha pretendido combinar contenido e información. Permite pasar de las páginas descriptivas a los recursos utilizando el etiquetado con metadatos o la lista de recursos pertinentes propuesta por el motor de búsqueda. Al estar dirigido a públicos de muy distinta índole, se ha querido facilitar el acceso a los recursos bibliográficos a través de las páginas de contenido o del Glosario y, a su vez, permitir realizar búsquedas precisas utilizando facetas, navegación entre descriptores y criterios de pertinencia. El reto ha consistido en mantener al mismo tiempo una entrada sencilla con una navegación natural adaptada a los criterios actuales y una estructuración muy precisa de la información que proporciona resultados pertinentes y que se pueden refinar. Basándose en la estructura del Tesauro y en sus múltiples relaciones y asociaciones, y enriqueciendo un vocabulario sencillo, como el del Glosario, se han podido combinar estas exigencias. Vemos así cómo un instrumento terminológico que muchos consideraban totalmente desfasado en la actualidad puede servir de espina dorsal para un desarrollo en la web semántica.

Bibliografía

- CHIKH, S.-A. (2014): *L'exploitation des vocabulaires dans un portail web sémantique. Le cas du portail thématique «Plan4Learning» de l'IIEP-UNESCO*. INTD-CNAM. Mémoire, ingénierie documentaire <http://memic.ccsd.cnrs.fr/mem_01128428v2/document>. [Consulta: 16 de marzo de 2016].
- FRANCART, T. (2013): *Plan4Learning: Easing access to educational planning resources*. Paper presented to United Nations Library and Information Network for Knowledge Sharing, UN-LINKS, Ginebra. <[http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/50424CFEF79D053BC1257BF80040DF3F/\\$file/Plan4Learning+Easing+Access+to+Educational+Planning+Resources_Thomas+Francart.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/50424CFEF79D053BC1257BF80040DF3F/$file/Plan4Learning+Easing+Access+to+Educational+Planning+Resources_Thomas+Francart.pdf)>. [Consulta: 16 de marzo de 2016].
- GARROD, P. (2000): «Use of the UNESCO Thesaurus for Archival Subject Indexing at UK NDAD», *Journal of the Society of Archivists*, n.º 1, 21, pp. 37-54.
- Le Thésaurus de l'UNESCO ou la corne d'abondance du savoir*; <<http://gido.iut.u-bordeaux3.fr/evenements/le-thesaurus-de-lUNESCO-ou-la-corne-dabondance-du-savoir/>>. [Consulta: 16 de marzo de 2016].
- PASTOR SÁNCHEZ, J. A.; MARTÍNEZ MÉNDEZ, F. J.; LÓPEZ CARREÑO, R., y RODRÍGUEZ MUÑOZ, J. V. (2013): *UNESKOS. Publicación como Linked Open Data de la Nomenclatura Internacional de Ciencia y Tecnología y del Tesauro UNESCO*. *Actas I Congreso ISKO Espanha Portugal*, Porto.
- TESAURO DE LA UNESCO <<http://databases.UNESCO.org/thessp/>>. [Consulta: 16 de marzo de 2016].
- UNESCO THESAURUS <<http://skos.um.es/UNESCOthes/?l=es>>. [Consulta: 16 de marzo de 2016].
- VALDERRAMA PAREJA, A. (2013): *Plan for learning: a portal on Planning for Improved Learning Outcomes*. Paper presented to United Nations Library and Information Network for Knowledge Sharing, UN-LINKS, Ginebra. <[http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/BB5C8A9BA798F445C1257BF80040C5B9/\\$file/IIEP+Portal+Plan+for+Learning_UNESCO-IIEP.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/BB5C8A9BA798F445C1257BF80040C5B9/$file/IIEP+Portal+Plan+for+Learning_UNESCO-IIEP.pdf)>.

Construcción y uso de terminologías, categorías de descripción y estructuras semánticas vinculadas al patrimonio en la sociedad global de datos

Nuria Rodríguez Ortega

Universidad de Málaga

1. Las condiciones de la sociedad global de datos

La emergencia y progresiva consolidación de una sociedad construida y articulada en torno a los datos masivos que ella misma produce en un escenario global está dando lugar a profundas transformaciones en procesos relacionados con la construcción y distribución del conocimiento, así como en aquellos vinculados a la interpretación de sus artefactos –sociales, culturales y políticos–. Estas transformaciones también implican nuevas formas de acceso y de interacción con el material cultural digital (o digitalizado), situando el foco de atención en un orden de problemáticas distintas. Aunque estas son muy diversas, para los propósitos de este texto quiero llamar la atención sobre las siguientes.

- 1.º En primer lugar, la noción de «apertura» de los datos y de los dispositivos asociados a su descripción, clasificación y estructuración, noción que va aparejada a la idea de compartición y –de especial interés para la argumentación que quiero desarrollar– a las de uso, reutilización y aprovechamiento, claves del nuevo paradigma epistémico de nuestra contemporaneidad. Aunque con diferentes velocidades, es destacable la inversión de tiempo, dinero y recursos humanos que se está realizando para poner a disposición pública datos, metadatos, vocabularios, ontologías, etc.¹.

¹ Este texto es resultado del proyecto de investigación «Generación de conocimiento sobre exposiciones artísticas temporales para su reutilización y aprovechamiento multivalente», financiado por la FBBVA; y del proyecto de I+D HAR2014-51915-P: «Catálogos artísticos: gnoseología, epistemologías y redes de conocimiento. Análisis crítico y computacional». Europea, por ejemplo, tiene un programa específico para compartir y reutilizar los datos y metadatos registrados en sus sistemas. Consúltense en: <<http://labs.europeana.eu/>>. [Consulta: 12 de marzo de 2016]. En España, la Biblioteca Nacional acaba de anunciar la firma de un acuerdo con el Ministerio de Economía, Industria y Turismo para «favorecer el acceso y la reutilización del fondo digital de la BNE, fomentando la innovación tecnológica e impulsando el desarrollo y creación de nuevos productos y servicios a partir de los contenidos digitalizados y los conjuntos de datos generados por la Biblioteca». Véase en: <<http://www.bne.es/es/AreaPrensa/noticias2016/0310-convenio-red.html>>. [Consulta: 14 de marzo de 2016].

Esta circunstancia está dando lugar a la aparición de nuevas prácticas relacionadas con el uso de estos materiales que se alejan de los que habían sido hasta ahora sus objetivos tradicionales: registro, clasificación y descripción para el acceso y gestión de las colecciones digitales asociadas a los acervos patrimoniales.

- 2.º En segundo lugar, el tránsito de la web de los documentos a la web de los datos; esto es, la irrupción de una economía del conocimiento que ya no está basada solo en el valor de las informaciones contenidas en uno o varios documentos, sino en la potencialidad que tiene la recombinación de datos híbridos para generar un conocimiento inédito (descriptivo o predictivo). Este nuevo escenario lleva implícito el desarrollo de nuevas estrategias de análisis en las que la combinación de metodologías cuantitativas basadas en el procesamiento computacional de datos y metadatos mediante algoritmos más o menos complejos, y la exploración de nuevas formas de visualización desempeñan un papel central. Así pues, la vinculación entre sociedad digital y cultura algorítmica está dando lugar a una reformulación de las estrategias de acceso tradicionales, que ahora se entienden como procesos de «descubrimiento», sustancialmente diferentes de las técnicas de búsqueda y recuperación de información conocidas hasta ahora.
- 3.º En tercer lugar, hay que reseñar como una de las claves definitorias de nuestra sociedad digital la aparición de nuevos actores en el ecosistema global, circunstancia que está generando una diversificación de las prácticas relacionadas con el uso y construcción de materiales para la descripción de los acervos culturales. ¿Cuáles son estos nuevos actores?: comunidades a gran escala con capacidad para producir datos masivos (Twitter, Instagram, Flickr...); discursos colectivos basados en dichos datos (blogs, *wikis*, modalidades de escritura colaborativa); los *software* y algoritmos que los procesan; y los actores globales (Google, Facebook, GLAM, etc.) que los gestionan y los hacen accesibles (Kaplan, 2015). En consecuencia, debemos contar con un nuevo escenario transnacional y tecno-social como parte de las narrativas críticas, y examinar de qué modo estos actores condicionan y modelan los procesos ligados a la documentación digital de los acervos culturales y al uso que se realiza de esos materiales documentales. Dicho de otro modo, la construcción de terminologías y estructuras para el acceso y la descripción del patrimonio ya no es prerrogativa exclusiva de los documentalistas, expertos de dominio o instituciones de la memoria (es decir, los actores tradicionales); por el contrario, en la actualidad encontramos todo un universo de prácticas heterogéneas, descentralizadas y distribuidas que hay que tener en cuenta.

Dado este nuevo contexto, la pregunta a la que trata de responder este texto es la siguiente: ¿cuáles son las nuevas prácticas relacionadas con el uso y construcción de vocabularios, categorías de descripción y estructuras semánticas vinculadas a la sociedad global de datos que nos alejan tanto de los «usos» tradicionales como de los «actores» hasta ahora habituales?

Tomando pues como horizonte referencial esta pregunta, me propongo abordar desde la perspectiva del historiador del arte –que es el lugar desde el que yo hablo en este texto– tres aspectos: a) qué proyectos, investigaciones o iniciativas ejemplifican modelos de nuevas prácticas relacionadas con el uso y construcción de los lenguajes patrimoniales; b) a qué pensamiento y discurso crítico nos impelen estas nuevas prácticas; c) qué iniciativas se podrían emprender colaborativamente entre documentalistas, expertos de dominio, desarrolladores informáticos, públicos, etc., para expandir las potencialidades de estos materiales documentales e incrementar la calidad de su reaprovechamiento.

2. Construcción, uso y reaprovechamiento en contextos de investigación científico-académica

De acuerdo con lo dicho anteriormente, uno de los retos de nuestro tiempo desde la perspectiva del historiador del arte –es decir, de alguien que está comprometido con proporcionar un mejor y más profundo conocimiento de los procesos histórico-artísticos y de las condiciones del patrimonio–

reside en explorar cómo reutilizar todo este material que originariamente se generó para el acceso, registro y documentación de los acervos culturales en los procesos de investigación histórico-artística y patrimonial. Por tanto, me referiré, en primer lugar, a un grupo de prácticas que ilustra cómo los vocabularios y los metadatos que hemos utilizado durante largo tiempo para el registro y documentación de los artefactos culturales nos pueden servir ahora para llevar a cabo investigaciones que dibujan marcos de comprensión inéditos. Estas prácticas ejemplifican, además, cómo las condiciones materiales y tecnológicas de nuestro tiempo nos permiten estudiar «digitalmente» las colecciones patrimoniales desde puntos de vista innovadores. La idea que subyace –explícita o implícitamente– a todas estas prácticas es que el trabajo curatorial y documental realizado durante tanto tiempo en el proceso de registro y descripción de colecciones ha convertido a las instituciones de la memoria y a sus grandes bases de datos en repositorios de un conocimiento estructurado a través de metadatos que podemos «reutilizar» para fines diversos.

Las investigaciones llevadas a cabo por Matthew Lincoln en la Universidad de Maryland² constituyen un ejemplo de cómo indagar en colecciones específicas de museos a través del procesamiento de sus metadatos y estructuras semánticas para profundizar en determinados procesos artísticos. Particularmente, su objeto de investigación estuvo centrado en las colecciones de grabados holandeses del British Museum y del Rijksmuseum, que analizó con el objetivo de cartografiar las redes de relaciones profesionales y artísticas establecidas entre los actores implicados en sus procesos de producción y distribución (grabadores, dibujantes, diseñadores, impresores, etc.), para extraer los patrones sistémicos que rigieron en su establecimiento y evolución. La descripción de estas colecciones de grabados mediante un conjunto exhaustivo de metadatos semánticamente estructurados, y el hecho de que, estas estén accesibles mediante estructuras de Linked Open Data³ permitió a Matthew Lincoln, aplicando estrategias de *network analysis* y determinados índices estadísticos, construir representaciones dinámicas de la configuración de estas redes a lo largo del tiempo (Lincoln, 2015b).

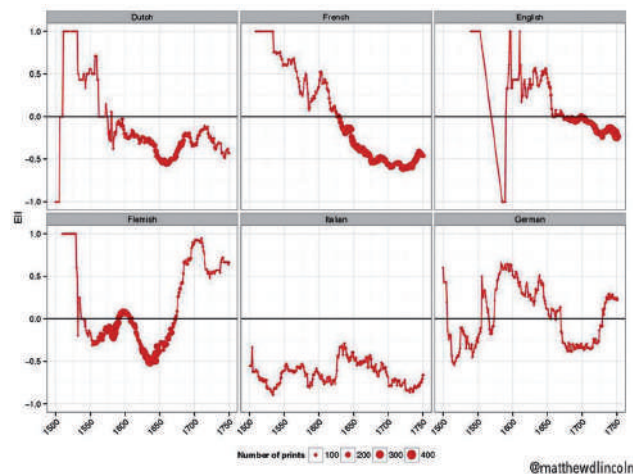


Figura 1. En la imagen superior se muestran las métricas que indagan sobre el carácter centralizado o descentralizado de las redes según nacionalidades. Los resultados por encima de 0 indican una mayor proporcionalidad de conexiones internacionales, mientras que los resultados por debajo de 0 señalan un mayor número de conexiones locales-nacionales. Obsérvese, por ejemplo, cómo las redes de Italia se definen durante todo el periodo por su carácter «doméstico». Disponible en: <<http://matthewlincoln.net/2015/05/21/the-art-historians-macroscopic.html>>. [Consulta: 1 de noviembre de 2015] (© Matthew Lincoln).

² <<http://matthewlincoln.net/>>. [Consulta: 1 de marzo de 2016].

³ El British Museum ha adoptado el modelo de referencia conceptual desarrollado por el CIDOC (CRM-CIDOC), el cual permite conectar objetos, personas, eventos y conceptos en un modelo de datos enriquecido (Le Boeuf *et alii*, 2012). Algunos proyectos que están utilizando este modelo de datos pueden consultarse en la sección de proyectos de la web *Research Space*. *Creating the Cultural Heritage Knowledge Graph*, espacio desarrollado por el British Museum con el apoyo de The Andrew W. Mellon Foundation. Disponible en: <<http://www.researchspace.org/home/crmprojects>>. [Consulta: 1 de noviembre de 2015].

Como se puede observar en la Figura 1, estas representaciones visibilizan aspectos que serían imposibles de obtener mediante los habituales procesos de búsqueda y recuperación de información, o simplemente mirando a través de listas de resultados. El enfoque que presidió esta investigación es ya por sí mismo innovador al centrar su estudio en las estructuras de relaciones y procesos sistémico a lo largo de amplios periodos de tiempo en vez de en los artefactos o las praxis individuales. No obstante, lo que aquí nos interesa destacar es cómo los metadatos utilizados para el registro y la documentación de las colecciones se transforman en «marcadores» que permiten establecer redes y construir métricas con la finalidad de averiguar cuáles son las estructuras que emergen a partir de las interrelaciones establecidas entre actores diversos⁴.

Otra de las iniciativas que ilustra este tipo de prácticas basadas en el uso de los conjuntos de metadatos existentes en los grandes repositorios son los análisis que se están llevando a cabo tomando como base el Getty Provenance Index⁵. El Getty Provenance Index constituye una de las mayores bases de datos relacionadas con el coleccionismo artístico, que incluye miles de inventarios, catálogos de venta y libros contables de marchantes en un periodo de tiempo que se extiende desde el siglo XVI hasta el siglo XX. Sus datos y metadatos, abiertos y accesibles se pueden utilizar para indagar sobre cuestiones relacionadas con comportamientos del consumo artístico, procesos de adquisición de obras, tendencias del mercado, redes entre marchantes y galeristas o modos de exhibición de colecciones.

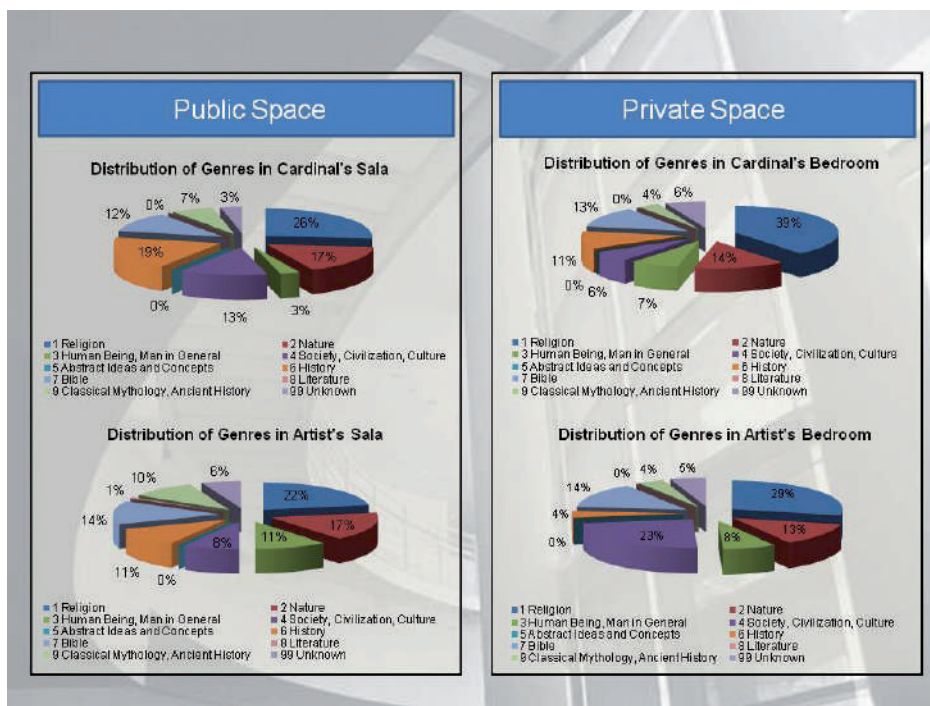


Figura 2. Análisis estadístico de la distribución de géneros pictóricos en los espacios palaciegos romanos a partir de los metadatos del Getty Provenance Index. Disponible en: <<http://www.getty.edu/research/tools/provenance/faq.html>>. [Consulta: 10 de noviembre de 2015] (©J. Paul Getty Trust and Christian Huemer).

⁴ Otros resultados derivados de estos análisis métricos pueden consultarse en Lincoln, 2015a.

⁵ El Getty Provenance Index, desarrollado durante más de tres décadas, forma parte del programa del Getty Research Institute «The Project for the Study of Collecting and Provenance (PSCP)». Se puede consultar en: <<http://www.getty.edu/research/tools/provenance/index.html>>. [Consulta: 10 de noviembre de 2015].

La Figura 2 ilustra, justamente, resultados de una investigación centrada en este último aspecto. Su objetivo era analizar la existencia de posibles patrones en la disposición de las colecciones pictóricas en los palacios de Roma entre 1550 y 1750. Para ello, los casi 300 inventarios romanos que comprenden un total de 65 000 registros individuales fueron procesados para mostrar la existencia –o no– de cambios en la distribución de los géneros pictóricos en los espacios palaciegos en función de su naturaleza –privados o públicos– o de los diferentes niveles sociales de sus propietarios⁶.

Esta investigación constituye, así, un ejemplo de cómo el procesamiento estadístico de los términos descriptivos y clasificatorios utilizados para documentar los objetos inventariados nos permiten explorar aspectos relacionados con la cultura visual y material de un momento dado, trascendiendo, de nuevo, los usos habituales de acceso y recuperación de información para los que fueron originariamente concebidos.

Sin abandonar este tipo de proyectos basados en el reaprovechamiento de materiales y estructuras documentales ya existentes, encontramos un conjunto de iniciativas que están centrando su atención en los metadatos y sistemas de descripción como objetos de investigación en sí mismos. En estos casos, los metadatos ya no representan solo un «instrumento» valioso para extraer conocimiento inédito sobre procesos artísticos –como hemos visto en los casos anteriores–, sino que ellos mismos se convierten en el objeto que hay que explorar y analizar bajo la premisa de que estos conjuntos de metadatos llevan inscritos discursos sociales, culturales, tecnológicos e institucionales al haber sido contribuidos y creados en diferentes eras, momentos y circunstancias (Battle y Loukissas, 2013). Este tipo de proyectos, pues, atiende a la idea de que los metadatos y las estructuras de clasificación son «artefactos y construcciones culturales» que derivan de distintos procesos interrelacionados⁷, y que nos cuentan historias de sí mismos, de las prácticas que han regido sus procesos de creación y uso; y, por tanto, de los individuos y de las instituciones que están detrás de su generación y gestión.

Una de las iniciativas que, a mi modo de ver, mejor ejemplifica este tipo de enfoque es el proyecto “The Life and Death of Metadata”⁸, que centra su atención en los metadatos de los registros de entrada del Arnold Arboretum, una extensa colección de especímenes arbóreos de muy heterogénea procedencia, que actualmente gestiona la Universidad de Harvard. Coordinado por Yanni Alexander Loukissas (Georgia Tech’s School of Literature, Media, and Communication), el proyecto combina procesamiento computacional avanzado, visualizaciones complejas y entrevistas con el personal del Arnold Arboretum para ilustrar cómo los metadatos registran los cambios locales que se producen en tecnología, personal y organización a lo largo del tiempo (1872-2012).

⁶ <<http://www.getty.edu/research/tools/provenance/faq.html>>. [Consulta: 10 de noviembre de 2015].

⁷ Según Battles y Loukissas (2013), los artefactos de datos (*data artifacts*) derivan de varios procesos interrelacionados: pueden ser artefactos de producción, resultantes de las condiciones materiales de los procesos de creación de datos; artefactos disciplinarios, que revelan los sistemas de ordenación específicos de las disciplinas; artefactos vernaculares, resultados de dialécticas locales o usos del lenguaje; artefactos de clasificación, ligados a cambios en los esquemas de ordenación; o artefactos ontológicos, producidos por conflictos entre clasificaciones culturalmente específicas.

⁸ Véase en: <<http://lifeanddeathofdata.org/>>. [Consulta: 1 de noviembre de 2015].

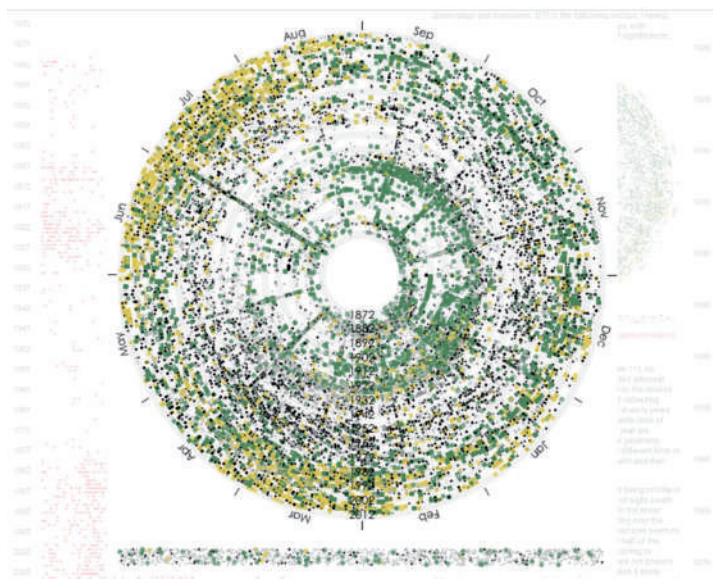


Figura 3. *Timeline* radial en el que se grafican los metadatos correspondientes a los registros de acceso del Arnold Arboretum (Universidad de Harvard) (captura de pantalla). Disponible en: <<http://lifeanddeathofdata.org/>>. [Consulta: 10 de noviembre de 2015].

Así, el *timeline* radial que se muestra en la Figura 3 nos cuenta una historia relacionada con las prácticas coleccionísticas de la propia institución. Véase, por ejemplo, el incremento del número de registros a largo de los años, pero también el cambio de temporalidad, pasando de un modelo de temporalidad en el que los registros se concentran en el periodo de invierno, a otro modelo de temporalidad en el que, a partir de la mediación del siglo, los registros se distribuyen a lo largo de todo el año. También se puede observar la coincidencia entre el incremento de una determinada especie arbórea (en amarillo en el gráfico original) y el inicio de la gestión de un nuevo director (véase Figura 4)⁹.



Figura 4. Captura de pantalla tomada de la visualización dinámica de *The Life and Death of Metadata*. Disponible en: <<http://lifeanddeathofdata.org/>>. [Consulta: 10 de noviembre de 2015].

⁹ Un análisis profundo de sus resultados se encuentra en la web del proyecto indicada en la nota anterior.

Por tanto, estos metadatos nos permiten comprender mejor no solo las colecciones que ellos describen, sino los propios procesos de registro y documentación, arrojando luz sobre cuáles son las prácticas sociales en torno a estos procesos y de qué manera las organizaciones y las instituciones construyen y sistematizan su conocimiento. En definitiva, nos permiten repensar los metadatos desde las condiciones de su producción material.

Por eso, este tipo de proyectos también nos confronta con algunas paradojas de la sociedad global de datos que merece la pena reseñar. Me refiero, en concreto, a la dialéctica entre la aparente condición universal de los datos, que implícitamente promueven las iniciativas de *open data*, las cuales ponen a disposición pública conjuntos de datos y metadatos que pueden utilizarse en cualquier lugar del mundo y para cualquier propósito –incluso, como hemos visto, para fines muy alejados de sus intenciones originarias–, y el hecho de que la relevancia y el valor de estos datos esté siempre local e históricamente situada en la medida en que son el resultado de las interacciones humanas –y de los objetivos de una comunidad– en un determinado contexto socio-cultural y político.

Otro conjunto de prácticas se orienta a la creación *ad hoc* de terminologías y vocabularios específicos, necesarios en aquellos proyectos basados, justamente, en el análisis de datos. Un interesante ejemplo a este respecto es el estudio llevado a cabo por el equipo del laboratorio CulturePlex¹⁰ (Western Ontario University) dirigido por Juan Luis Suárez en el marco del proyecto «The Hispanic Baroque: Complexity in the First Atlantic Culture». El objetivo de este proyecto era analizar el origen, la evolución y la transmisión de patrones de comportamiento y representación del mundo hispánico en el espacio transatlántico para examinar si efectivamente podíamos hablar de la existencia de una «cultura barroca» en este espacio geográfico y a través de un eje temporal amplio. Para ello se construyó una base de datos de más de 13 000 registros de obras pictóricas producidas a ambos lados del Atlántico durante tres siglos (1550-1850), que se describieron mediante 211 descriptores –o atributos– siguiendo el modelo de una ontología formal¹¹. El análisis consistió en observar la evolución de estos descriptores/atributos a lo largo del tiempo mediante un algoritmo que configuraba agrupaciones de pinturas (clústeres) definidas por atributos similares, demostrándose –entre otras cosas– que el principal elemento de conexión entre ellas eran los géneros o temas¹² (véase Figura 5).

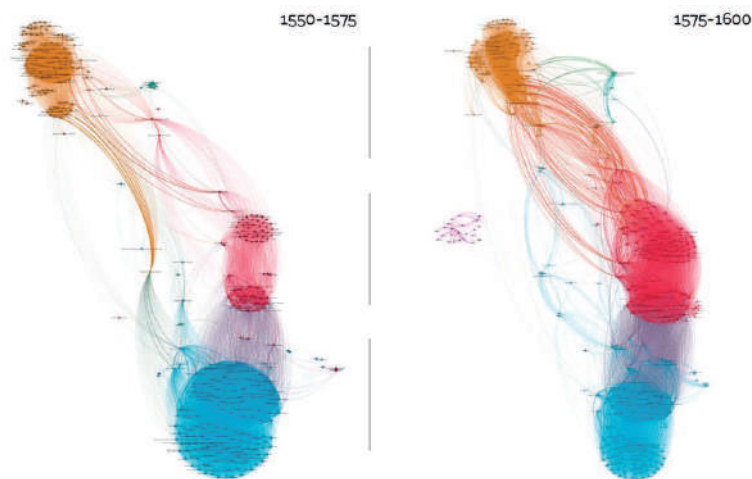


Figura 5. Los grafos que se representan en esta imagen muestran las dinámicas de cambio de los clústeres pictóricos para los dos periodos de estudio, especialmente la variabilidad de las interconexiones establecidas entre ellos, así como la desaparición/aparición de determinados clústeres en uno u otro periodo. Los colores hacen referencia a los distintos géneros pictóricos representados. (Suárez *et alii*, 2013) (imagen cortesía de Juan Luis Suárez).

¹⁰ Sitio web: <<http://www.cultureplex.ca/>>. [Consulta: 1 de noviembre de 2015].

¹¹ Véase en: <<http://ontologies.cultureplex.ca/baroqueart/spec/index.html>>. [Consulta: 10 de noviembre de 2015].

¹² El estudio completo con las distintas métricas utilizadas en el análisis se encuentra en Suárez *et alii*, 2013.

Independientemente del resultado de la investigación, lo que me interesa destacar es cómo estos vocabularios no se utilizan aquí con el objetivo específico de «documentar» el objeto en el sentido tradicional del término (acceder, describir, gestionar), sino como herramientas para encontrar similitudes o diferencias entre artefactos culturales y asociarlos mediante determinados algoritmos de agrupación –lo que llamamos procesos de *clustering*–, permitiéndonos, así, dar nuevas explicaciones sobre sus lógicas de funcionamiento.

Para cerrar esta primera parte, plantearé algunas cuestiones importantes que requieren consideración.

En primer lugar, parece claro que de nada sirve que se pongan a disposición pública datos y metadatos si no conocemos las tecnologías necesarias para extraerlos y procesarlos. Por tanto, el conocimiento de estas tecnologías específicas constituye un factor crítico para la viabilidad de este tipo de prácticas. Si estamos convencidos del enorme potencial que tiene el procesamiento de todo este material documental generado en torno a los acervos culturales, hemos de estar de acuerdo también en la necesidad de promover una adecuada formación.

A su vez –y sin menoscabo de los conocimientos técnicos imprescindibles que se deben adquirir–, también es preciso desarrollar una intensa colaboración entre documentalistas, desarrolladores y expertos de dominio para crear interfaces más amigables y usables, menos dirigidas a construir preguntas en un código complejo (Figura 6), y más encaminadas a funcionar como entornos de estimulación intelectual mediante la sugerencia de posibles vías de interrogación.

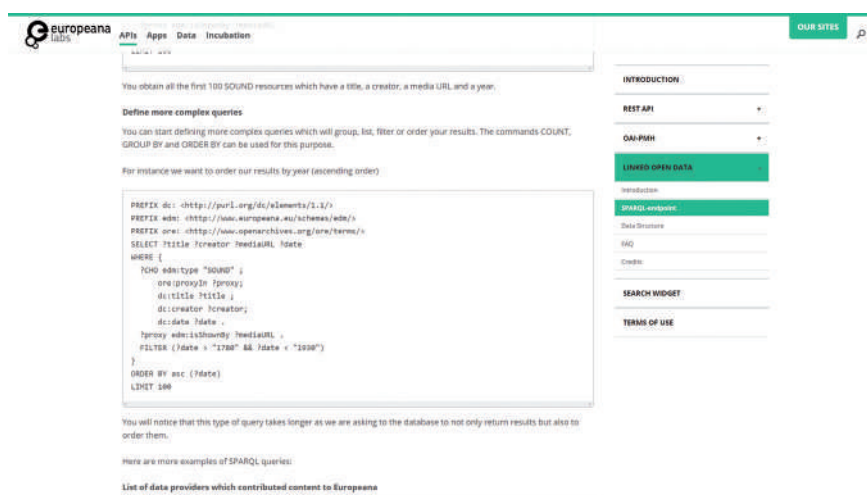


Figura 6. Consulta construida en SPARQL (*Protocol and RDF Query Language*). Captura de pantalla del SPARQL Endpoint del Europeana Lab. Disponible en: <<http://labs.europeana.eu/api/linked-open-data-SPARQL-endpoint>>. [Consulta: 1 de marzo de 2016].

En segundo lugar, promover y apoyar el uso de estándares y estructuras normalizadas en los proyectos específicos que acometemos los investigadores representa un factor crucial para poder reutilizar los materiales generados durante su desarrollo en iniciativas futuras. Instaurar esta cultura como requisito fundamental nos permitiría rentabilizar la inversión de tiempo, dinero y recursos humanos que se realiza cada vez que en un proyecto se afronta la construcción de un repositorio o una base de datos. Nuevamente, se requiere aquí una intensa labor de pedagogía y formación.

En este sentido una estrategia que también podría contribuir a este objetivo sería el desarrollo de plataformas en las que, a modo de recolectores o agregadores, estuviesen accesibles los diversos sistemas, estructuras, vocabularios, ontologías, etc., actualmente disponibles. De este

modo, los investigadores podríamos acudir a esta información para discriminar cuáles de ellos convienen más a nuestros propios proyectos. Esta propuesta incluiría la publicación en abierto de los tesauros que las instituciones de la memoria están desarrollando para la gestión de sus propias colecciones¹³ a fin de que pudiesen ser reutilizados en proyectos específicos, bien empleando las estructuras semánticas y los términos ya existentes, bien expandiéndolos con nuevas incorporaciones.

En tercer lugar, hemos de ser conscientes de que la publicación en abierto de los datos y metadatos y su codificación en estructuras semánticas normalizadas se está desarrollando a diversas velocidades –como ya apunté al inicio de este texto–, existiendo importantes desequilibrios entre países y comunidades. Esto no constituye un problema menor; por el contrario, genera un nuevo espacio desde el que seguir ahondando en las desigualdades que también están asociadas a la expansión de la cultura digital. Si son los países del bloque Norte los que están trabajando con mayor intensidad en políticas de datos abiertos y enlazados, las investigaciones volverán a centrarse en los recursos materiales y en los bienes patrimoniales de determinadas instituciones, territorios y contextos –aquellos sobre los que se puede investigar utilizando estas metodologías–, quedando el resto conminados a formar parte de una nueva marginalidad.

3. Nuevos contextos y escenarios

Como indiqué al inicio de este texto, nuevos actores se incorporan a la construcción y uso de vocabularios, categorías y estructuras semánticas para la descripción de los acervos culturales. The Art Genome Project, desarrollado por la empresa norteamericana Artsy¹⁴, constituye un buen punto de arranque para la reflexión que quiero proponer en este epígrafe. Artsy es una plataforma digital que nació para intermediar en procesos de compra-venta de obras artísticas. Sin abandonar este propósito, actualmente se ha convertido en un ingente yacimiento de información estratégica sobre el sector. Artsy registra obras artísticas –fundamentalmente en venta, pero no solo–, coleccionistas, artistas, galerías, museos, etc. En este contexto se explica el Art Genome, un vocabulario controlado sobre arte y diseño desarrollado internamente por Artsy que incluye estilos y movimientos artísticos, conceptos, atributos visuales, contextos geográficos, materiales y técnicas, etc. Actualmente cuenta con más de 1000 términos. Es de este vocabulario del que se toman los metadatos descriptivos que se agregan a las obras de arte y a los artistas registrados en la plataforma. Así pues, el Art Genome es el sistema de clasificación y la infraestructura tecnológica en la que se sustenta Artsy, siendo la principal vía de acceso a su información.

Sin duda, resulta interesante explorar este vocabulario para descubrir el establecimiento de nuevas dimensiones categoriales; por ejemplo, las relacionadas con los atributos visuales o los efectos sensoriales, que nos permiten acceder a los objetos artísticos desde otras perspectivas. La exploración de nuevas perspectivas de acceso no es ajena a los objetivos de Artsy: hacer sus registros más accesibles, más «discoverable» para sus clientes, de ahí que una de sus estrategias sea explorar las expectativas de sus usuarios, que son principalmente coleccionistas, galeristas y artistas.

Sin embargo, es necesario también enfocar este proyecto desde un discurso crítico. Su filosofía explícitamente declarada es detectar las características (llamadas «genes») que conectan entre sí a artistas y obras de arte a lo largo de la historia. Cada una de las palabras del Art Genome denomina una característica, entendida como «gen», por lo que la idea última sería reconstruir una especie de ADN de las obras artísticas¹⁵. Así pues, la clave del Art Genome Project no reside tanto en el vocabulario en sí, sino en el «Genome Connection», esto es, en la conexión

¹³ Como está haciendo ya el Ministerio de Cultura, Educación y Deporte con los Tesauros del Patrimonio Cultural de España: <<http://tesauros.mecd.es/tesauros/tesauros>>. [Consulta: 1 de marzo de 2016].

¹⁴ <<http://www.artsy.net>>. [Consulta: 10 de noviembre de 2015].

¹⁵ Véase en: <<https://www.artsy.net/article/theartgenomeproject-genome-connection-three-artists-take-on-the>>. [Consulta: 10 de noviembre de 2015].

que a través de este vocabulario se establece entre los distintos elementos registrados. De este modo, como dicen sus propios desarrolladores: «Art Genome Project aims to create a classification system that retains the nuances and mysteries within art and allows anyone with an Internet connection the opportunity to learn about art and art history»¹⁶. Este propósito resulta muy loable, pero si tenemos en cuenta el modelo de negocio de Artsy, el Genome Connection –que actúa estableciendo conexiones entre obras, artistas y coleccionistas que comparten atributos entre ellos («genes»)– constituye sobre todo y fundamentalmente un potente mecanismo de recomendación sobre el que se sustenta y expande el negocio de Artsy: la compra-venta de obras de arte. Por supuesto, el hecho de que los fines subyacentes al proyecto sean comerciales no es en sí mismo un factor «repudiable», pero sí que nos impele a desplegar un pensamiento crítico, independiente del valor que le podamos reconocer –y que de hecho le reconocemos– a este vocabulario para proponer nuevos conceptos o categorías de clasificación. En este sentido apunto dos líneas sobre las que, en mi modesta opinión, se debe reflexionar.

Por una parte, Artsy representa uno de estos nuevos actores globales a los que me refería al inicio de este texto ligados a los oligopolios del *software* y de los datos, pues la relevancia de Artsy radica, realmente, en el desarrollo de potentes dispositivos de computación para conectar elementos entre sí y en una acumulación ingente de datos «estratégicos», entre los que se engloban las interacciones de los usuarios y sus prácticas de búsqueda en la plataforma. Pues bien, estos nuevos actores, entre los que yo incluyo a Artsy, son susceptibles de convertirse en nuevas instituciones de legitimación cultural en convivencia paralela con las instituciones tradicionales. Esta cuestión no es baladí; es importante recordar que estos lenguajes descriptivos y clasificatorios de carácter normalizador implican discursos normativos sobre cómo se describen, organizan, ordenan y clasifican los artefactos culturales. En la medida en que estos emanan de instituciones a las que les reconocemos un estatuto de autoridad, se convierten en discursos en sí mismos legitimadores (Rodríguez Ortega, 2014). ¿Pero qué sucede cuando estas instituciones legitimadoras ya no son solo las instituciones tradicionales de la memoria, sino actores con intereses ajenos a la convencional «tutela» del patrimonio? Es necesario, pues, prestar atención a la posible influencia que puedan ejercer en la conformación de nuevas categorías de descripción y clasificación.

Por otra parte, sabemos que el uso del lenguaje y de los vocabularios como herramientas y estrategias «empresariales» no es nuevo. De hecho, toda la maquinaria de Google en cuanto «agencia de publicidad» está basada en ello, esto es, en su capacidad para conferir valor monetario a las palabras, que quedan así convertidas en mercancías¹⁷. La capacidad de Google para transformar el capital lingüístico –contribuido por todos– en un capitalismo lingüístico ha sido ya escrito y debatido (Kaplan, 2014). No son pocos los estudios que describen las estrategias de Google para vender sus palabras y que reflexionan sobre los efectos que esto –y otras estrategias lingüísticas, como el autocompletado– pueden tener en la futura evolución del lenguaje. No quiero redundar aquí sobre ello, pero sí considero necesario actualizar este debate en el terreno que nos interesa, pues, salvando las diferencias entre uno y otro caso¹⁸, la filosofía subyacente es la misma: el potencial de las palabras del Arte Genome deriva de su capacidad para dinamizar procesos de recomendación en el marco de una estrategia empresarial. Dicho de otro modo, se transforma un

¹⁶ [«El objetivo del Art Genome Project es crear un sistema de clasificación capaz de preservar los matices y misterios del arte, y permitir a todo aquel con una conexión a Internet el que pueda aprender sobre el arte y la historia del arte», traducción de la autora]. Véase en: <<https://www.artsy.net/article/theartgenomeproject-introducing-the-art-genome-project>>. [Consulta: 10 de noviembre de 2015].

¹⁷ Las palabras y expresiones lingüísticas monetarizadas por Google pueden ser compradas por las empresas en una especie de subasta en la que se dirime su mayor o menor posicionamiento en el sistema de anuncios de Google. Naturalmente, el valor de estas palabras varía a lo largo del año, por ejemplo durante los periodos vacacionales, en los que unas palabras adquieren mayor valor que otras. Véase la explicación del método en: <<https://support.google.com/adwords/answer/1704424?authuser=0>>, [Consulta: 1 de noviembre de 2015], especialmente lo relativo al «Planificador de palabras clave».

¹⁸ El funcionamiento es distinto, pues en Google media solo un determinado algoritmo que trabaja sobre el conjunto de palabras que el buscador reúne diariamente a través de las interacciones de los usuarios; y en Artsy trabaja un conjunto de «expertos» que deciden qué palabras incluir y cuáles no.

material lingüístico con un valor semántico y con una función epistémica –en el más amplio sentido del término– en un dominio dado (clasificación, descripción, generación de conocimiento, etc.) en un recurso potencialmente aprovechable desde un punto de vista económico.

Otro de los escenarios que hay que considerar como contexto diverso a los tradicionales –académicos e institucionales– es el representado por los colectivos de artistas, comisarios y expertos que actúan de manera autónoma y autogestionada, y que poseen un indudable potencial para provocar una revulsión en los procesos de construcción de vocabularios y categorías de descripción. Como ejemplo de este potencial podemos considerar el principio de intervención llevado a cabo en el marco del proyecto «Archivos en uso»¹⁹, comandado por la Red Conceptualismos del Sur²⁰ y el Grupo de Estudios sobre Arte, Cultura y Política en la Argentina Reciente. «Archivos en uso», cuyo objetivo es salvar la memoria de determinados procesos del arte latinoamericano, se enmarca dentro del interés creciente que la constitución, visibilidad, socialización y uso de los archivos relacionados con las prácticas artísticas en Latinoamérica ha experimentado en la última década. La preservación y acceso a los acervos documentales, muchos de ellos en peligro de pérdida, ha dado lugar a diversas iniciativas que se pueden consultar en la propia web de la Red Conceptualismos del Sur.

Pero, sobre todo, se ha alertado de la tensión que genera la constitución –o, mejor dicho, apropiación– de estos archivos por parte de las instituciones legitimadoras del sistema del arte contemporáneo (museos, coleccionistas privados, instituciones corporativas). Hay que tener en cuenta que estos archivos documentan principalmente prácticas de resistencia y disensión respecto de los discursos legitimadores, por lo que su potencia crítica corre peligro de desactivarse si estas prácticas son documentadas y descritas por las mismas instituciones frente a cuyos discursos hegemónicos se instituyeron. Así pues, la configuración de regímenes de archivo alternativos emerge como una necesidad crítico-política de primer orden. Este posicionamiento se evidencia en la propia construcción de los vocabularios utilizados para su descripción, que difieren en determinados casos de los empleados por las instituciones tradicionales. Encontramos, por ejemplo, categorías denominadas «conceptos fetiche» o «ideas fuerza» en las que se recoge un vocabulario que remite al carácter político, crítico y desinstitucionalizador de las prácticas que se documentan. Asimismo, se observa una mayor relación con los conceptos clave del discurso del arte del siglo XX, y con los que han dado forma a la teoría crítica de la cultura de nuestro tiempo.

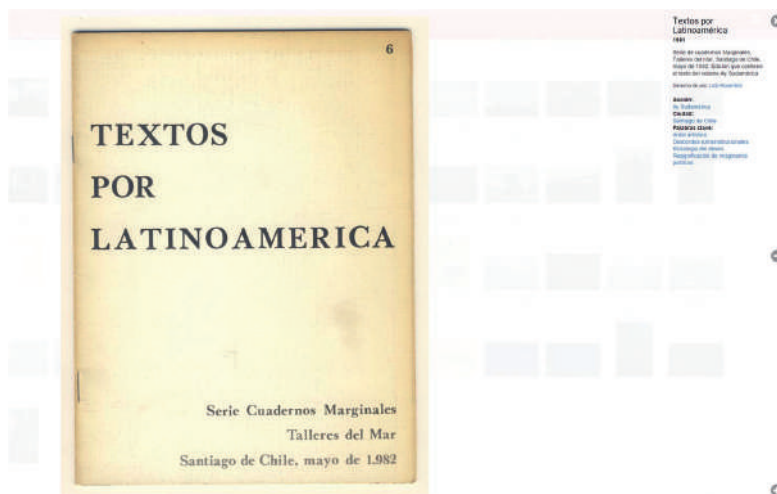


Figura 7. Captura de pantalla de Archivos en Uso. Disponible en: <<http://www.archivosenuso.org>>. [Consulta: 10 de noviembre de 2015].

¹⁹ <<http://www.archivosenuso.org>>. [Consulta: 10 de noviembre de 2015].

²⁰ Plataforma de investigación, discusión y toma de posición colectiva desde América Latina, fundada en 2007. Véase en: <<http://redcsur.net/>>. [Consulta: 1 de febrero de 2016].

La diferencia se hace más significativa si comparamos este tipo de vocabulario con el empleado por el MNCARS en sus prácticas de documentación. Así, si para buscar en el sistema del MNCARS utilizamos –pongamos por caso– los términos «apropiaciónismo», «formalismo», «formalidad», «poscolonialismo», «poscolonialidad», «ritual», «fetichismo», esto es, conceptos que se alejan de las categorías vinculadas a la descripción material y a la clasificación tradicional de los artefactos culturales, la búsqueda resulta infructuosa. Intuimos, por tanto, que estos términos no forman parte de su vocabulario documental²¹, pese a ser claves en los discursos del arte de los siglos XX y XXI, y a cuya luz las propuestas artísticas de este periodo cobran sentido.

Esta circunstancia me lleva a plantear una línea de reflexión importante: la necesidad de buscar un acuerdo entre lo que interesa para el registro y documentación de los artefactos culturales con vistas a su clasificación, gestión y acceso, y las «expectativas» de los usuarios –esto que Arsty ha visto tan claro–. Dicho de otro modo, es necesario buscar un reequilibrio entre lo que los usuarios esperan encontrar y los sistemas que las instituciones necesitan. Es en este escenario de confluencias en el que es posible establecer relaciones muy productivas entre documentalistas, investigadores, artistas, comisarios, desarrolladores informáticos, expertos en curaduría digital y ciudadanos en general con el objetivo de explorar nuevas vertientes de vocabulario que nos permitan ampliar el corpus de preguntas que se pueden plantear a las colecciones digitales de los acervos culturales. De este modo, estaremos trabajando en el enriquecimiento de sus posibilidades cognitivas y epistémicas; en definitiva, en el incremento de sus posibilidades como los entornos de descubrimiento que quieren llegar a ser.

Otro modelo de prácticas asociadas al uso y construcción de vocabularios y estructuras semánticas es el que se relaciona con las plataformas de producción social, esto es, aquellas en las que interviene una multiplicidad de actores, con bagajes formativos e intereses muy diversos, globalmente distribuidos y con dinámicas de participación flexibles y poco apriorísticas.

Fijémonos en que hasta ahora hemos estado hablando de bases de datos y repositorios –preexistentes o contruidos *ad hoc* para proyectos específicos– que han sido estructurados de manera consistente a través de procesos consensuados y estándares documentales. Pero si nuestro horizonte es la sociedad de datos, tenemos que referirnos irremediabilmente también a la web en su globalidad. Por tanto, cuando nos preguntamos cómo funcionan los vocabularios y los dispositivos utilizados para la descripción de los artefactos culturales, no podemos circunscribirnos solo al territorio de las bases de datos y los repositorios estructurados –es decir, ese territorio que podemos «controlar»–, sino que hemos de considerar la web en su totalidad, esto es, ese espacio de contenidos distribuidos, heterogéneos, semiestructurados o completamente desestructurados, y en fluencia continua. En este contexto, quiero aludir a dos tipos de proyectos que nos ofrecen claves interesantes de interpretación.

Uno de ellos es el llevado a cabo hace un par de años por un grupo de investigadores de la Universidad Tecnológica de Viena (Goldfarb *et alii*, 2012). Su objetivo era indagar en repositorios de información contruidos socialmente para examinar cómo se configura su estructura semántica subyacente a partir de la contribución descentralizada y distribuida de distintos actores, y especialmente cómo esta estructura contrasta con los repositorios contruidos siguiendo directrices normalizadas establecidas *a priori* por grupos de expertos. En concreto, el estudio compara las diferentes estructuras de conocimiento que emergen a partir del mismo conjunto de biografías según estas se encuentren en el Union List of Artist Names (ULAN)²² –desarrollado por el Getty Research Institute– o en Wikipedia. Lo que se analizan son las conexiones interpersonales que aparecen en ambos repositorios: unas, las del ULAN, decididas directamente por el

²¹ Digo «intuimos» porque el vocabulario empleado para la documentación de las prácticas artísticas y los artefactos culturales no está accesible y, por tanto, no se puede consultar.

²² Vocabulario controlado que ofrece información estructurada sobre artistas y otros actores del dominio artístico. Véase en: <<http://www.getty.edu/research/tools/vocabularies/ulan/>>. [Consulta: 1 de noviembre de 2015].

equipo editorial de acuerdo con unos criterios acordados *a priori* –según hemos dicho–; otras, las de Wikipedia, deducidas de las interconexiones establecidas sin un «plan» previo por la diversidad de sus participantes.

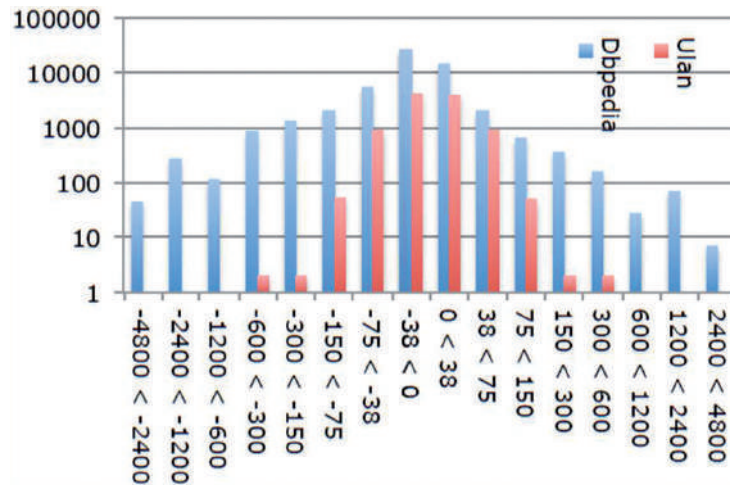


Figura 8. Distribución de las distancias de los enlaces establecidos entre los artistas (escala logarítmica) (© Doron Goldfarb).

En el gráfico de la Figura 8 se observa cómo el ULAN muestra una estructuración más regular y simétrica frente a la mayor «irregularidad» de los sistemas creados por *crowdsourcing*. Dado que los fenómenos culturales y sociales, considerados en su complejidad y diversidad, difícilmente responden a estructuras tan regulares, estas gráficas hacen visible el carácter «construido» de este tipo de repositorios, cuyas estructuras nos hablan más de las prácticas de los individuos y de las organizaciones que los crean, que de los fenómenos socio-culturales que están teóricamente «describiendo». Por su parte, la visualización de las conexiones en Wikipedia (Figura 9) nos proporciona una fotografía de la visión que sobre un determinado fenómeno tiene una multiplicidad de individualidades pertenecientes a una o varias comunidades.

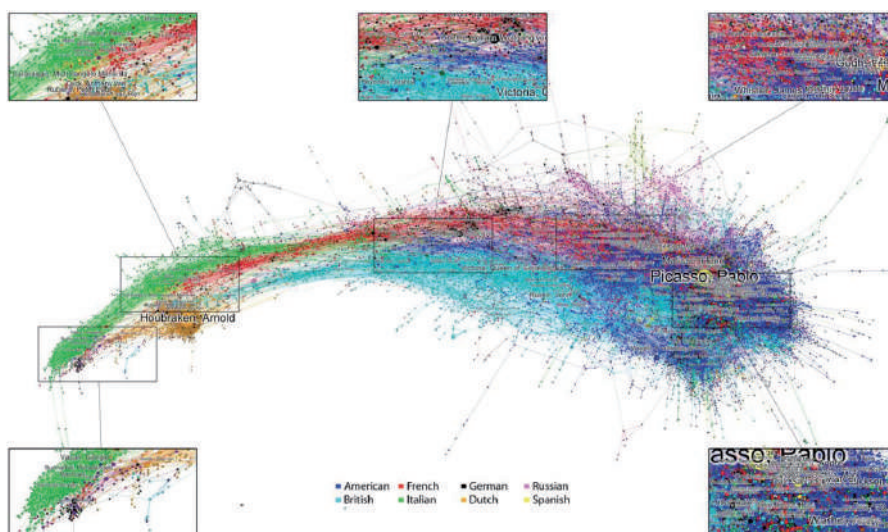


Figura 9. Visualización de los enlaces establecidos en Wikipedia (© Doron Goldfarb).

Nos encontramos, así, ante la confrontación de dos modos de construcción semántica que conviven en nuestro mundo contemporáneo, viabilizando, cada una de ellos, una determinada representación del conocimiento.

El otro proyecto al que quiero referirme es HuNi. El proyecto australiano HuNi indaga en la construcción específica de arquitecturas de la participación que tratan de cuestionar los modelos basados en la aplicación apriorística de determinadas estructuras o estándares²³. HuNi es una infraestructura tecnológica que congrega datos de múltiples fuentes (bases de datos, repositorios, sitios web, etc.) de contenido cultural. Está estructurada siguiendo los parámetros propios de una ontología (entidades y relaciones), pero en vez de partir de una ontología previamente diseñada a la que hay que acomodarse cuando se trata de registrar y documentar los datos, HuNi deja libertad para que sean los propios usuarios los que establezcan las relaciones entre las diversas entidades –que sí están definidas–. Podemos decir, por tanto, que nos encontramos ante una ontología producida «socialmente», de abajo-arriba, fluida y adaptativa, que transforma el objetivo tradicional de las ontologías al uso, pues en vez de tratar de ajustar los datos a un marco normativo preexistente, el propósito último de HuNi es capturar las diferentes perspectivas de los usuarios a medida que la ontología se va construyendo. Este planteamiento se inscribe en la emergente noción de «Social Linked Data», que viene a revisar las estrategias de Linked Open Data. Es decir, es la idea de que, frente al tradicional énfasis puesto en el *open data* –la apertura de los datos–, el LOD también debería ser explorado a través de un *open linking* –apertura de las relaciones– (Verhoeven, 2014: 213).

Sea como fuere, este tipo de proyectos vuelve a incidir en la reflexión planteada en párrafos anteriores: la necesidad de contar con las comunidades que están utilizando y produciendo esos datos –o que están directamente relacionados con ellos– en los procesos asociados a su construcción, uso y producción.

Con todo, este tipo de proyectos también tiene sus contrapartidas, que no residen tanto en la mayor o menor dificultad tecnológica que representa el desarrollo de las infraestructuras necesarias–para lo cual disponemos de un conocimiento avanzado–, sino en la capacidad para dinamizar la participación de los potenciales usuarios, sobre todo en comunidades que son poco dadas a ello. Nos encontramos, así, con un nuevo reto sobre el que trabajar.

La construcción social de vocabularios para la descripción de los acervos culturales no es, sin embargo, un hecho nuevo. El denominado etiquetado social (*social tagging*), entendido como la práctica descentralizada mediante la cual individuos o comunidades crean, gestionan y comparten términos para anotar y categorizar los recursos digitales en un entorno social, lleva muchos años desarrollándose. De hecho, en la actualidad ha llegado a convertirse en una praxis institucionalizada que se ha incorporado como una actividad más en no pocos museos; en algunos casos de una manera más libre; en otros, más controlada a partir de tesauros previos –como sucede en el Centre Pompidou–, lo cual, obviamente, limita su capacidad creativa, que es en donde radica la potencialidad del *social tagging* para enriquecer la construcción de vocabularios aportando puntos de vista alternativos.

Es cierto que sobre el *social tagging* se ha escrito mucho, y no querría yo redundar en cosas ya debatidas y que seguramente el lector conocerá, pero me gustaría mencionar dos aspectos que, a mi parecer, requieren todavía de una atenta consideración por parte de los historiadores del arte y de la cultura.

El primero tiene que ver con la formulación de dos preguntas: ¿hasta qué punto estas prácticas descentralizadoras han supuesto una alternativa a los vocabularios construidos por parte de las instituciones tradicionales?; y ¿hasta qué punto su potencia crítica, subversiva o simplemente alternativa se ha visto atemperada por los procesos de institucionalización que han experimentado? Ambas preguntas demandan todavía una respuesta.

²³ <<https://huni.net.au/#/search>>. [Consulta: 1 de noviembre de 2015].

El segundo aspecto concierne a la toma de conciencia de la riqueza que este material contribuido por los ciudadanos en sus interacciones «digitales» con los acervos culturales nos proporciona para analizar desde múltiples puntos de vista cómo estos se conceptualizan, entienden, receptionan y experimentan. Sinceramente creo que la aplicación de tecnologías de *social web mining* para analizar todo este material aportaría resultados inéditos en relación con estos procesos, teniendo en cuenta, además, que ahora podemos analizarlos en un ecosistema global y multicultural²⁴ (Rodríguez Ortega, 2013).

Finalmente, en este contexto que estamos considerando, quiero llamar la atención sobre la fuente extraordinaria que para la construcción de vocabularios desde otras perspectivas semánticas representan los blogs desarrollados por artistas individuales o colectivos, quienes ejercen una actividad continua de etiquetado de sus propias entradas y publicaciones. La eclosión que la blogosfera de contenidos artísticos ha experimentado en los últimos años se convierte, así, en un magnífico caudal en el que bucear mediante estrategias de *web mining*. Además, hay que tener en cuenta que estas etiquetas o términos provienen de los actores que están directamente implicados en los procesos de producción artística, por lo que integrarlos como parte de las fuentes y materiales que utilizamos en la construcción de vocabularios contribuiría también a esa necesidad de acuerdo a la que me he estado refiriendo de manera insistente a lo largo de este texto.

Bibliografía

- BATTLES, M., y LOUKISSAS, Y. (2013): «Data Artifacts: Visualizing Orders of Knowledge in Mega-Meta Collections», en *Proceedings of the UDC Seminar in Classification and Visualization*. Würzburg: Ergon Verlag.
- GOLDFARB, D. *et alii* (2012): «Comparing Art: Historical Networks», en *Arts, Humanities, and Complex Networks*. Edición de Maximilian Schich *et alii*. Cambridge: MIT, (Leonardo ebook Series).
- KAPLAN, F. (2014): «Linguistic Capitalism and Algorithmic Mediation», *Representations*, n.º 127, Summer, University of California, pp. 57-63.
- (2015): «A Map for Big Data Research in Digital Humanities», *Frontiers in Digital Humanities*, April 2015. Disponible en: <<http://journal.frontiersin.org/article/10.3389/fdigh.2015.00001/full>>. [Consulta: 20 de julio de 2015].
- LE BOEUF, P. *et alii* (2012): *Definition of the CIDOC Conceptual Reference Model. Version 6.2*. ICOM/CIDOC.
- LINCOLN, M. D. (2015a): «The Art Historian's Macroscope: Museum Data and the Academy», *Matthew Lincoln* (blog), 21 May 2015. Disponible en: <<http://matthewlincoln.net/2015/05/21/the-art-historians-macroscope.html>> [Consulta: 1 de noviembre de 2015].
- (2015b): «DH2015: Modelling the (Inter)national Printmaking Networks of Early Modern Europe», *Matthew Lincoln* (blog), 30 Jun 2015. Disponible en: <<http://matthewlincoln.net/2015/06/30/dh2015-modelling-the-international-printmaking-networks-of-early-modern-europe.html>>. [Consulta: 1 de noviembre de 2015].
- LU, C. (2013): *Exploiting Social Tagging Network for Web Mining and Search*, Philadelphia: Drexel University.
- RODRÍGUEZ ORTEGA, N. (2013): «Mediación digital en los procesos de recepción, construcción e interpretación de la herencia cultural», en *Memoria y significado. Uso y recepción de los vestigios del pasado*. Edición de Luis Arciniega. Valencia: Universidad de Valencia, pp. 443-461.
- (2014): «Canon, Value and Artistic Culture: Critical Inquiry about the New Processes of Assigning Value in the Digital Real», en *Digital Humanities Annual Conference 2014. Books*

²⁴ Sobre las potencialidades de exploración del etiquetado social puede consultarse: Lu, 2013.

of abstracts, Lausanne, 9-11 de julio de 2014: Ecole Polytechnique Fédérale de Lausanne y Universidad de Lausanne, pp. 331-333.

SUÁREZ, J. L. *et alii* (2013): «Towards a Digital Geography of Hispanic Baroque Art», *Literary and Linguistic Computing*, n.º 13, Oxford University Press, pp. 1-18.

VERHOEVEN, D. (2014): «Doing the Sheep Good. Facilitating Engagement in Digital Humanities and Creative Arts Research», en *Advancing Digital Humanities. Research, Methods, Theories*. Edición de Katherin Bode y Arthur Paul Longley. Hampshire: Palgrave MacMillan, pp. 206-220.

La normalización terminológica aplicada a la conservación y restauración del patrimonio cultural

Ana María Calvo Manuel

Universidad Complutense de Madrid. Facultad de Bellas Artes

A través de estas líneas nos proponemos ofrecer el panorama de los glosarios y de la normalización terminológica en materia de conservación y restauración del patrimonio cultural.

El área de conocimiento en conservación y restauración del patrimonio cultural es un ámbito de muy reciente implantación, por ello se hace todavía más necesaria la normalización terminológica. Sin un lenguaje común, en un espacio que cuenta con tan variadas profesiones intervinientes (conservadores, historiadores, arqueólogos, arquitectos, archiveros, artistas, químicos, biólogos, físicos, fotógrafos, etc.) debido a la multi y pluridisciplinariedad de los trabajos que se llevan a cabo, sería difícil el entendimiento y el avance científico.

A pesar de los esfuerzos que se han realizado a través de los diccionarios y glosarios elaborados en los últimos años, veremos como algunos matices terminológicos todavía no presentan la unidad necesaria. Y esto ocurre no solo en castellano, sino también en lenguas como la inglesa, que se impone actualmente en el contexto de las ciencias experimentales aplicadas al patrimonio cultural. Por ello, todavía estamos lejos de contemplar una normalización completa en este campo.

En relación con los primeros diccionarios vinculados a la conservación-restauración, podemos hablar de algunos precedentes en esta materia, como la obra de Gettens y Stout (1966), *Painting Materials. A Short Encyclopaedia*. Este libro, a pesar de ser una referencia fundamental de mediados del siglo XX, no aborda exactamente aspectos específicos de conservación, sino que trata fundamentalmente de técnicas pictóricas y materiales de pintura. Con todo, es evidente que el conocimiento de dichos aspectos es fundamental para la conservación de las pinturas. También en esa misma línea, el Getty Conservation Institute publicó en años sucesivos su serie de Looking at... A guide to technical terms, como, por ejemplo, *Looking at Paintings*, de Dawson W. Carr y Mark Leonard (1992).

Más cercano al tema que nos ocupa, el proyecto NARCISSE (Network of Art Research Computer Image Systems in Europe), que tuvo lugar entre los años 1990 y 1994, coordinado por

el Laboratoire de Recherche des Musées de France, tenía como objetivo crear un banco de datos de las obras más importantes de los museos de Europa a base de documentación técnica (fotografías con luz visible, ultravioleta e infrarroja, y radiografías), básicamente centrado en pintura de caballete. Como resultado del proyecto se editó un CD-ROM con un glosario especializado sobre el tema.

Sin embargo, desde el punto de vista específico de conservación y restauración, van a ser, sobre todo, los breves glosarios publicados en los apéndices de los libros técnicos sobre la materia los que darán las pautas de una terminología técnica más precisa.

Así pues, cuando en 1994 recibimos el encargo de realizar un diccionario de términos de conservación y restauración, para la colección «Cultura Artística» que iniciaba Ediciones del Serbal, dirigida por Joan Sureda, el primer gran problema al que nos enfrentábamos era el de la selección de las propias entradas, puesto que se trataba de un ámbito en el que debíamos contar con términos artísticos, químicos, físicos, biológicos y técnicos específicos. Basándonos en dichos glosarios que formaban apéndices de publicaciones en la materia, y en los libros precedentes citados, hicimos una propuesta que se concretó en el libro *Conservación y restauración. Materiales, técnicas y procedimientos. De la A a la Z* (Calvo, 1997). Esta obra ha tenido gran repercusión en el ámbito profesional y se han llegado a hacer cuatro reediciones (Figura 1).

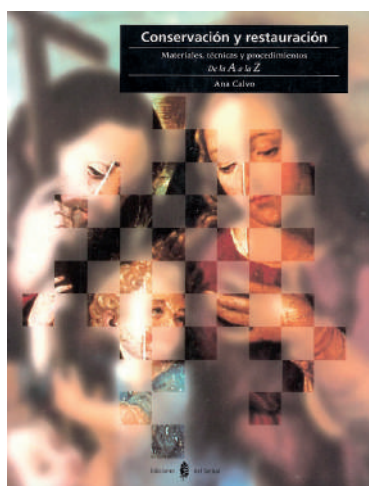


Figura 1. Ana CALVO. *Conservación y restauración. Materiales, técnicas y procedimientos. De la A a la Z*. (1997).

En italiano se había publicado, en 1992, *Lessico del restauro* por Cristina Giannini, pero no se trataba de un diccionario o glosario, sino de un ensayo que relataba la evolución de la profesión. Sin embargo, la misma autora, junto con Roberta Roani, editaron en el año 2000, la obra *Dizionario del restauro e della diagnostica*, que fue traducida y publicada en castellano por la editorial Nerea en 2008. También en italiano se publicó en 2005 el *Glossario delle tecniche artistiche e del restauro*, a cargo de Claudio Paolini y Manfredi Faldi, en una edición del Palazzo Spinelli de Florencia.

En el año 2003 se publica en España, en la colección de diccionarios técnicos de la editorial Akal, *Conservación y restauración de bienes culturales. Español - Alemán - Inglés - Italiano - Francés*, un diccionario multilingüe, dirigido por Celia Martínez Cabetas y Lourdes Rico Martínez, que, aunque no incluye definiciones sino únicamente la correlación de los términos en las lenguas citadas, supuso una afirmación de la terminología técnica en nuestro campo (Figura 2).



Figura 2. Celia MARTÍNEZ CABETAS y Lourdes RICO MARTÍNEZ (coords.). *Conservación y restauración de bienes culturales. Español - Alemán - Inglés - Italiano - Francés.* (2003).

En Barcelona, Mireia Xarrié (2005, 2006a, 2006b, 2007), comenzó a trabajar en unos glosarios especializados en conservación, publicándolos en primer lugar en inglés y, finalmente en 2007, un volumen en castellano.

Mucho más recientemente han aparecido el *Glosario de términos técnicos para restauración y museología*, de Marcos Rivadeneira y Guillermo Narváez, en edición latinoamericana, en 2013; y *Vademécum del conservador. Terminología aplicada a la conservación del Patrimonio Cultural*, de Javier Madrona Ortega, en 2015 (Figura 3).



Figura 3. Javier MADRONA ORTEGA. *Vademécum del conservador. Terminología aplicada a la conservación del Patrimonio Cultural.* (2015).

Los títulos citados hasta ahora eran de tipo generalista en torno a conservación y restauración. Además, han aparecido a través de los años glosarios especializados en algún tipo de material u objeto artístico y cultural específico. Por ejemplo, en 1997 se publicó *Conservación de*

libros y documentos. *Glosario de términos técnicos. Inglés - Español. Español - Inglés*, de John Mc. Cleary; en francés, en este caso, contamos con el *Petit glossaire à l'usage du conservateur-restaurateur de sculpture*, editado en 2006 por el Institut Royal du Patrimoine Artistique de Bruxelles, con la coordinación de Myriam Serck-Dewaide en el contexto de un proyecto europeo denominado CRISTAL. Más recientemente se ha publicado también: *Diccionario técnico Akal de materiales de restauración*, elaborado por Salvador Muñoz Viñas, Julia Osca Pons e Ignasi Gironés Sarrió (2014), centrado, como su nombre indica, en la descripción y características de algunos de los más importantes materiales que utilizamos en las intervenciones de conservación-restauración (Figura 4); y *EWAGLOS European Illustrated Glossary for Conservation Terms of Wall Painting and Architectonic Surfaces* (2015), referido a pinturas murales y superficies arquitectónicas, resultado también de un proyecto europeo, por lo que aparecen las definiciones en inglés con traducción a búlgaro, croata, francés, alemán, húngaro, italiano, polaco, rumano, español y turco. Este último se encuentra disponible gratuitamente en internet <<http://www.ewaglos.eu>> y cuenta asimismo con edición impresa en papel. Igualmente está accesible desde 2010 en la web la publicación de ICOMOS sobre las formas de alteración de la piedra: *Illustrated glossary on stone deterioration patterns*, en la colección Monuments and Sites XV, en versiones en inglés y en francés. Otro ejemplo, vinculado en este caso a la conservación de mosaicos, sería el *Illustrated Glossary: Technician Training for the Maintenance of In Situ Mosaics*, disponible desde 2013, también *online* en diferentes idiomas, entre las publicaciones de The Getty Conservation Institute.



Figura 4. Salvador MUÑOZ VIÑAS, Julia OSCA PONS e Ignasi GIRONÉS SARRIÓ. *Diccionario técnico Akal de materiales de restauración*. (2014).

En España contamos con las publicaciones, en papel pero también recientemente digitalizadas y disponibles *on line*, del Ministerio de Educación, Cultura y Deporte, en sus Colecciones en Red CERES, con libre acceso a todos los tesauros y diccionarios creados para normalizar el lenguaje de las bases de datos. Sin embargo, en el ámbito específico de la conservación-restauración, cabe destacar el libro *Proyecto COREMANS: «Criterios de intervención en materiales pétreos»* (Laborde, 2013), también del Ministerio pero gestionado por el Instituto del Patrimonio Cultural de España (Figura 5).



Figura 5. Ana LABORDE (coord.). *Proyecto COREMANS: «Criterios de intervención en materiales pétreos»*. (2013).

En relación con los materiales y las técnicas, es necesario citar asimismo dos glosarios o enciclopedias accesibles en Internet: CAMEO y Polyevart. La primera ha sido desarrollada por el Museum of Fine Arts de Boston en inglés: <http://cameo.mfa.org/wiki/Main_Page>; y la segunda está en español, y ha sido creada y ampliada por el Ministerio de Cultura a través de proyectos de investigación: <<http://www.mecd.gob.es/cultura-mecd/areas-cultura/patrimonio/mc/polyevart/glosario.html>>.

Para esas búsquedas de diccionarios y glosarios específicos sobre conservación y restauración, contamos actualmente con unas amplísimas bases de datos bibliográficas *online* que nos permiten «bucear» y encontrar títulos antiguos y modernos sobre este tema. Destacaríamos los *Abstracts of International Conservation Literature* del Getty - AATA online <<http://aata.getty.edu/Home>> (Figura 6) y *The Bibliographic Database of the Conservation Information Network - BCIN* <<http://www.bcin.ca/>> (Figura 7).

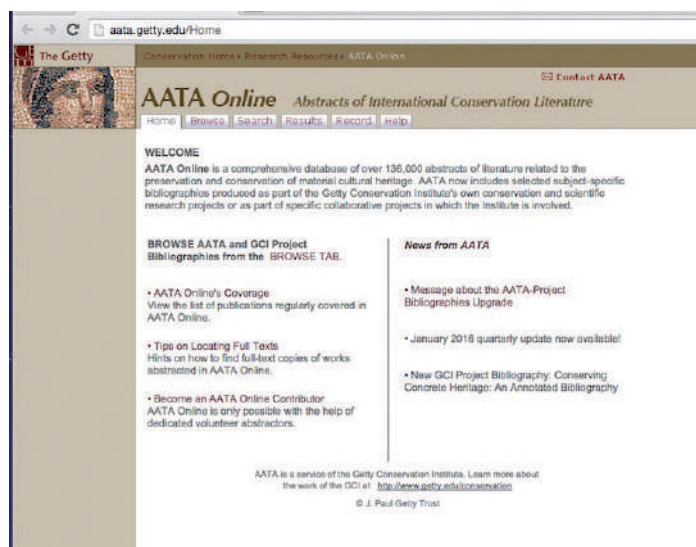


Figura 6. *Abstracts of International Conservation Literature* del Getty - AATA online <<http://aata.getty.edu/Home>>.

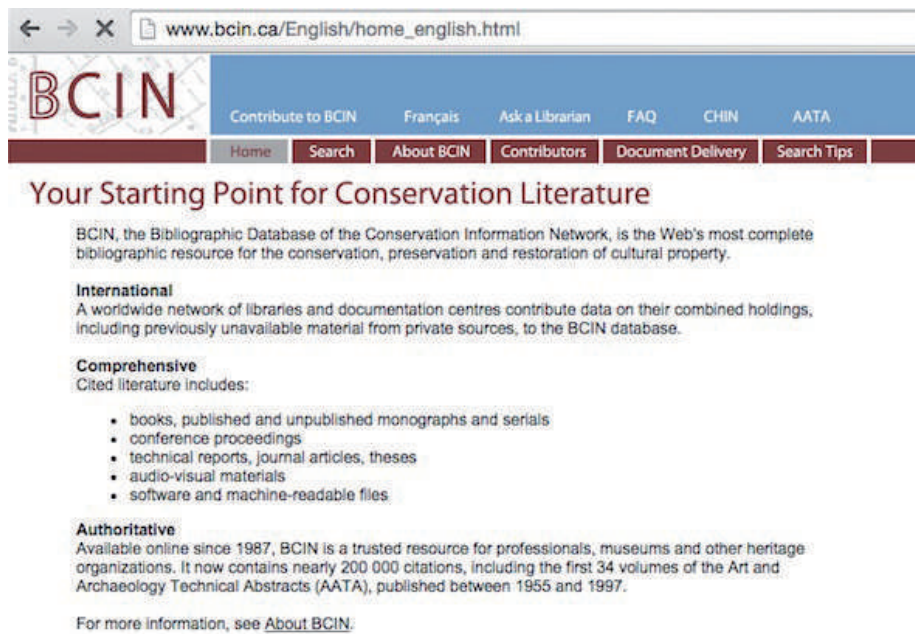


Figura 7. The Bibliographic Database of the Conservation Information Network -BCIN <<http://www.bcin.ca/>>.

Siguiendo la idea de la normalización en todos los campos, también en la Conservación del Patrimonio Cultural se ha creado, por parte del CEN –European Committee for Standardization–, un grupo de trabajo para el ámbito de la Conservación del Patrimonio Cultural, el CENT TC 346. Entre los diversos documentos ya aprobados, destacamos *Principales términos generales y definiciones*, norma española UNE-EN 15898, publicado en castellano en marzo de 2012 (AENOR, 2012).

A pesar de todos los trabajos citados, la terminología de conservación y restauración deber ser continuamente actualizada. Existen términos como «retoque», «repinte» y «reintegración cromática», cuyas interpretaciones han variado a lo largo del tiempo. Incluso contemporáneamente algunos vocablos, como «*retouching*» e «*inpainting*», que se utilizan como sinónimos en inglés, se han impuesto en países diferentes, en el primer caso en Gran Bretaña, y en el segundo en Estados Unidos (Calvo, Bailão, 2014). Publicaciones recientes lo confirman, como *Mixing and Matching. Approaches to Retouching Paintings* (Ellison; Smithen y Turnbull, 2010), editado en Londres; y *Painting Conservation Catalog. Volume III: Inpainting* (Metzger, 2011), en Estados Unidos.

Por la palabra «retoque» se entiende en el ámbito artístico como «aquellos toques sobre la obra original, realizados por el propio artista, a modo de toques finales para corregir o perfeccionar su obra» (Calvo, 1997). Esos toques originales, que decía Francisco de Goya, aportaban la genialidad a las pinturas. Como tal, siguiendo ese sentido, se emplea también para describir un aspecto de la técnica de reintegración del color, que se realiza sobre el barniz en la propia pintura original, para mitigar el efecto de desgastes que presentan las obras por deterioros del tiempo. Así lo empleaba también Vasari, cuando en sus *Vidas de artistas ilustres* (1568) decía que «sería mejor, algunas veces, quedarse con las cosas hechas por hombres excelentes un poco destruidas, que hacerlas retocar por quien sabe mucho menos». Sin embargo, en el contexto profesional denominamos «repinte» a la aplicación de color sobre un original con intención de reparar u ocultar daños, o de modificar el aspecto que realiza una persona diferente al artista y en una época posterior. Casi siempre tiene una connotación peyorativa. De esta forma, el término actualmente más apropiado para denominar la restitución del color en una zona perdida del original, que haga posible la integración estética del conjunto, sería «reintegración cromática». Dicho

término se impuso, según Ségolène Bergeon (1996), en el Instituto del Restauo de Roma por Cesare Brandi. Anteriormente, Paul Philippot (1959) había utilizado la expresión «integración».

En ámbitos más genéricos podríamos citar también el *Art & Architecture Thesaurus Online*, de The Getty Research Institute, <<http://www.getty.edu/research/tools/vocabularies/aat/>>, que está en fase de traducción a otras lenguas desde Holanda. Mientras que, en contextos más limitados, como en la docencia en el Grado de Conservación y Restauración del Patrimonio Cultural de la Facultad de Bellas Artes de la Universidad Complutense de Madrid, hemos trabajado en una base de datos para un glosario que sirviera de base terminológica para alumnos y docentes (Calvo *et alii*, 2015). Actualmente, se encuentra publicado *online* y accesible en pdf. en E-Prints Complutense <<http://eprints.ucm.es/35224/>> (Figura 8).



Figura 8. Ana CALVO *et alii*. *Terminología básica de conservación y restauración del Patrimonio Cultural 2. Español - Inglés - Francés - Italiano - Alemán*. (2016). <<http://eprints.ucm.es/35224/>>.

Así pues, como decíamos al inicio, a pesar de tantos avances en materia de publicaciones de glosarios en conservación y restauración, este trabajo debe continuar con objeto de afirmar los términos hoy empleados o de incorporar nuevos vocablos de nuestro ámbito de trabajo.

Bibliografía

- AENOR (2012): *Principales términos generales y definiciones*, norma española UNE-EN 15898, CENT TC 346 Conservación del Patrimonio Cultural.
- ALBERTI, L. *et alii* (2013): *Illustrated Glossary: Technician Training for the Maintenance of In Situ Mosaics*, Los Ángeles: The Getty Conservation Institute, disponible *online* en <http://www.getty.edu/conservation/publications_resources/pdf_publications/glossaire_f.html>. [Consulta: 2 de febrero de 2016].
- BERGEON, S. (1996): «Couleur et restauration», *Techne* n.º 4. París: LRMF, pp. 17-28.
- CALVO, A. (1997): *Conservación y restauración. Materiales, técnicas y procedimientos. De la A a la Z*. Barcelona: Ediciones del Serbal.

- CALVO, A. *et alii* (2016): *Terminología básica de conservación y restauración del Patrimonio Cultural 2. Español - Inglés - Francés - Italiano - Alemán*, disponible en pdf. *online* en E-Prints Complutense. <<http://eprints.ucm.es/35224/>> [Consulta: 2 de febrero de 2016].
- CALVO, A., y BAILÃO, A. (2015): «Reintegration, integration, inpainting retouching? Question around terminology», *II International Meeting on Retouching of Cultural Heritage* - RECH 2, Oporto (Portugal) 24-25 octubre 2014. Porto: RECH2 PROCEEDINGS, pp. 12-24.
- CARR, D. W., y LEONARD, M. (1992): *Looking at Paintings*, (Looking at... A guide to technical terms). Los Angeles: GCI.
- ELLISON, R.; SMITHEN, P., y TURNBULL, R. (2010): *Mixing and Matching. Approaches to Retouching Paintings*. Londres: Archetype.
- GETTENS, R. J., y STOUT, G. L. (1966): *Painting Materials. A Short Encyclopaedia*. Nueva York: Dover Publications.
- GIANNINI, C. (1992): *Lessico del restauro*, Florencia: Nardini Editore.
- GIANNINI, C., y ROANI, R. (2000): *Dizionario del restauro e della diagnostica*. Florencia, Nardini Editore. — (2008): *Diccionario de restauración y diagnóstico*. Guipúzcoa: Nerea.
- ICOMOS (2010): *Illustrated glossary on stone deterioration patterns* (Monuments and Sites XV), disponible *online* en: <<http://www.icomos.org/en/what-we-do/disseminating-knowledge/publicationall/monographic-series/116-english-categories/resources/publications/261-monuments-sites-xv>>. [Consulta: 2 de febrero de 2016].
- LABORDE, A. (coord.) (2013): *Proyecto COREMANS: «Criterios de intervención en materiales pétreos»*. Madrid: Secretaría General Técnica, Centro de Publicaciones, Ministerio de Educación, Cultura y Deporte. Disponible *online* en: <<https://sede.educacion.gob.es/publiventa/proyecto-coremans-criterios-de-intervencion-en-materiales-petres/conservacion-restauracion-patrimonio-historico-artistico/14516C>>. [Consulta: 2 de febrero de 2016].
- MADRONA ORTEGA, J. (2015): *Vademécum del conservador. Terminología aplicada a la conservación del Patrimonio Cultural*. Madrid: Tecnos.
- MARTÍNEZ CABETAS, C., y RICO MARTÍNEZ, L. (2003): *Conservación y restauración de bienes culturales. Español - Alemán - Inglés - Italiano - Francés*. Madrid: Akal.
- MC. CLEARY, J. (1997): *Conservación de libros y documentos. Glosario de términos técnicos. Inglés - Español. Español - Inglés*. Madrid: Clan Editorial.
- METZGER, C. A. (compiler) (2011): *Painting Conservation Catalog. Volume III: Inpainting*, USA: The Paintings Specialty Group of the American Institute for Conservation of Historic and Artistic Works.
- MUÑOZ VIÑAS, S.; OSCA PONS, J., y GIRONÉS SARRIÓ, I. (2014): *Diccionario técnico Akal de materiales de restauración*. Madrid: Akal.
- PAOLINI, C., y FALDI, M. (2005): *Glossario delle tecniche artistiche e del restauro*. Florencia: Edizioni Palazzo Spinelli.
- PHILIPPOT, P. (1959): «Le problème de l'integration des lacunes dans la restauration des peintures», *Bulletin de l'Institut royal du patrimoine artistique*, II, pp. 5-19.
- RIVADENEIRA, M., y NARVÁEZ, G. (2013): *Glosario de términos técnicos para restauración y museología*. Quito: Codeu.
- XARRIÉ, M. (2005): *Glossary of Conservation I*. Barcelona: Balaam. — (2006): *Glossary of Conservation II*. Barcelona: Balaam. — (2006): *Glossary of Conservation III*. Barcelona: Balaam.

— (2007): *Diccionario de conservación y restauración de obras de arte 1 (A-Z)*. Barcelona: Balaam.

VV. AA. (2015): *EWAGLOS European Illustrated Glossary for Conservation Terms of Wall Painting and Architectonic Surfaces*, disponible *online* en: <<http://projekte.hawk-hhg.de/ewaglos/pages/download.php>>. [Consulta: 2 de febrero de 2016].

Dimensión política y cultural en la armonización terminológica multilingüe: la experiencia del tesoro para la Red de Información del Patrimonio Europeo (HEREIN)

Teresa Abejón Peña

Consejo Superior de Investigaciones Científicas

El objetivo de este trabajo es comunicar la experiencia adquirida en el proceso de elaboración de un lenguaje documental del ámbito del Patrimonio Cultural y, en especial, destacar la dimensión política y cultural de la armonización terminológica en la elaboración de tesauros multilingües. El artículo se centra en el tesoro de la Red de Información del Patrimonio Europeo (HEREIN), y describe los criterios y la metodología que se siguió para resolver los problemas que el contexto político y la diversidad cultural plantean en la sistematización y normalización contrastiva o multilingüe de la terminología especializada.

En el caso que nos ocupa, el tesoro HEREIN, se planteaba como reto la construcción de un instrumento común en el que todas las lenguas tuvieran el mismo estatus. Por otra parte, y además de evitar en distintas etapas de elaboración del tesoro que hubiera lenguas que quedaran en una posición de superioridad o prestigio, era importante tratar, respetar y reflejar las especificaciones propias de cada cultura.

Introducción

La creación de la Red europea de información sobre el patrimonio cultural estuvo auspiciada por el Consejo de Europa y su Comité director de Patrimonio. El origen de la creación de la Red fue la 4ª conferencia europea de ministros responsables de patrimonio cultural organizada en Helsinki por el Consejo de Europa en 1996, que recomendó estudiar la creación de un sistema de información permanente que permitiera a las administraciones, los profesionales, los investigadores y los especialistas en formación conocer la evolución de la gestión y protección del patrimonio en los distintos países de Europa, utilizando informes nacionales sobre las políticas de patrimonio arquitectónico.

La red HEREIN (European Heritage Network o Red de Información del Patrimonio Europeo) debía ofrecer al público una fuente de datos y de experiencias fundamentales y contrasta-

das sobre la gestión del patrimonio cultural en Europa; servir como observatorio para analizar y prever las ventajas que ofrece el patrimonio cultural a una sociedad en rápida mutación; facilitar el seguimiento del desarrollo de las políticas de patrimonio y de su conformidad con las convenciones; mantener y desarrollar las redes de cooperación en materia de patrimonio cultural; y favorecer la cooperación transnacional.

Se había solicitado a todos los Estados un primer informe sobre políticas nacionales, pero esta documentación, que correspondía al principio a quince países, constaba únicamente en papel. Este soporte estático solo permitía una lectura lineal, una difusión limitada y una gestión centralizada de esta información; además, estaba disponible únicamente en las dos lenguas oficiales del Consejo de Europa. Para contar con un soporte dinámico, evolutivo y de acceso libre a esta información, se decidió crear un sitio web para la red HEREIN que fuera multilingüe y permitiera contrastar información sobre las políticas de protección y difusión del patrimonio que se llevaban a cabo en Europa consultando por país y tema en una base de datos de informes. En torno a este recurso se planteaba un sitio web más completo, mantenido por el Consejo de Europa, como sistema de información permanente e interactivo sobre patrimonio cultural.

Para este sistema de información se consideró importante elaborar un tesoro multilingüe que ofreciera una sistematización terminológica referida a las políticas nacionales relacionadas inicialmente con el patrimonio arquitectónico y arqueológico. El tesoro debía facilitar la navegación a través de los distintos informes nacionales accesibles en línea.

Al apostar por la consulta multilingüe y contemplar el desarrollo de un tesoro de estas características, el proyecto del Consejo de Europa encontró financiación en el programa IST de la Comisión Europea: el programa de Tecnologías de la Sociedad de la Información (IST) del V Programa Marco de IDT (1998-2002) de la Comisión para la comunidad investigadora de la UE. Este programa incluía líneas de acción como la gestión del conocimiento, un patrimonio para todos, interactividad natural y plurilingüe.

En el comienzo del proyecto, 1998-1999, la idea inicial era construir un tesoro bilingüe en inglés y francés, por ser las lenguas oficiales del Consejo de Europa. El Ministerio de Cultura de España se interesó rápidamente en que el sitio web de la Red tuviera interfaz en español y el tesoro fuera trilingüe. Así es como se desarrolló en una primera etapa (2000-2002) la primera versión en inglés, francés y español.

En una segunda etapa (2003-2004) se decidió ampliar el proyecto a los países NAS (Newly Associated States), los recientemente asociados a la Unión Europea, por eso se tradujo el tesoro HEREIN al húngaro, al polaco, al búlgaro, al rumano, al esloveno, etc. Con la financiación de la Comisión Europea, el tesoro se amplió incluyendo hasta diez lenguas. El Consejo de Europa se encargó después de mantenerlo como sistema permanente de información colaborando con los departamentos ministeriales correspondientes, y financiando la ampliación a otras lenguas y su actualización hasta el año 2012. La última versión incluye quince lenguas europeas: alemán, búlgaro, croata, esloveno, español, finés, francés, griego, húngaro, inglés, neerlandés, noruego, polaco, portugués y rumano.

En cuanto a la participación española en el tesoro HEREIN, el Ministerio de Cultura se implicó en el tesoro desde el inicio del proyecto, pero la Subdirección General de Protección del Patrimonio Histórico se dirigió al Centro de Información y Documentación Científica del CSIC (CINDOC) en busca de un socio institucional español con experiencia en el desarrollo de tesoros que colaborara con el Ministerio de Cultura. Así, el Consejo Superior de Investigaciones Científicas comenzó a colaborar en el tesoro desde 1999 y siguió participando en su actualización hasta el año 2013.

Objetivos y metodología del tesoro HEREIN

El tesoro HEREIN, como recurso lingüístico multilingüe que debía permitir el acceso a diferentes experiencias y actuaciones nacionales, se tenía que desarrollar con una meto-

dología apropiada y en un marco de cooperación internacional. Se concibió como un instrumento común en el que todas las lenguas debían tener el mismo estatus: la búsqueda de información no podía depender del conocimiento y la utilización de una lengua única dominante.

La incorporación de terminología específica relativa a las distintas experiencias y actuaciones nacionales era necesaria para contrastar información y disponer de un panorama general de la diversidad europea en la gestión y protección del patrimonio. Por esta razón, se admitían conceptos y designaciones específicos, aunque solo se utilizaran en una lengua o un país, y debían acompañarse de definiciones. El tesoro debía ser un verdadero reflejo de universos culturales y políticas diversas.

Quedaba así descartado partir de un solo tesoro monolingüe y traducirlo. Igualmente había que evitar que el nuevo tesoro multilingüe fuera una mera fusión de varios tesoros monolingües. Había que considerar además los aspectos 'políticos' y dar un tratamiento análogo y equilibrado a todas las lenguas y a todas las legislaciones nacionales. Por esta razón, y también debido a que no había un tesoro especializado monolingüe que cubriera específicamente las políticas y los aspectos legales del patrimonio, el tesoro HEREIN se construyó *ab initio* y respetando completamente todas las lenguas implicadas, intentando reflejar lo mejor posible las diferentes estructuras conceptuales y terminológicas con las cuales podían estar familiarizados los potenciales usuarios finales: los 'indizadores' y quienes buscan información. Esto no quiere decir que no se atendiera al examen de tesoros como, por ejemplo, el AAT o el Tesoro de Patrimonio Histórico Andaluz, que incluye un apartado A9 relativo al régimen jurídico del patrimonio histórico.

Se pretendía evitar que una lengua fuente se convirtiera naturalmente en la lengua dominante. Un tesoro monolingüe siempre está culturalmente sesgado, y una traducción directa podría conducir a un «imperialismo cultural» o «subordinación cultural» por el dominio de unas lenguas sobre otras.

Por otra parte, los términos considerados demasiado específicos de una lengua, o los términos que correspondían a las idiosincrasias regionales, no se excluyeron del tesoro, sino que se incluyeron como términos equivalentes a un concepto común en las distintas lenguas. Después de esto, el vocabulario inicial se completó con un proceso de retroalimentación entre lenguas.

La diversidad de políticas culturales descritas en los términos del tesoro debía contribuir a estructurar la herramienta. Los ajustes y las concesiones no se debían hacer siempre a costa de la misma parte: debía efectuarse una verdadera «armonización terminológica».

El corpus inicial para el desarrollo del tesoro estuvo constituido por los informes nacionales y por las convenciones y recomendaciones internacionales en inglés, francés y español. Se añadieron además algunos textos creados y difundidos por diferentes organizaciones internacionales relacionadas con el patrimonio (especialmente UNESCO y Consejo de Europa). Por otra parte, para la versión española se incluyó también la normativa específica estatal (especialmente la Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español). El Consejo de Europa, en todo el proceso de desarrollo y actualización del tesoro, también sugirió términos correspondientes a nuevos conceptos que consideraba importantes de cara a su implementación, como, por ejemplo, 'conservación integrada' o 'patrimonio inmaterial'.

En cuanto a las normas de elaboración de tesoros, para cada idioma, la construcción del tesoro se basó en el uso de relaciones básicas entre los términos (equivalencia, jerarquía y asociación), definidas de acuerdo con la norma ISO 2788 - Directrices para la creación y desarrollo de tesoros monolingües. Los diferentes niveles de equivalencias lingüísticas entre los distintos idiomas se establecieron según especifica la norma ISO 5964 - Directrices para la creación y el desarrollo de tesoros multilingües.

Los términos del tesoro se agruparon en nueve grupos o familias y se contempló la posibilidad de reflejar polijerarquías, de manera que un término podría ser incluido en dos grupos si se consideraba necesario para articular la estructura conceptual del tesoro.



Figura 1. Esquema general del tesoro.

Dificultades para establecer equivalencias interlingüísticas

En el proceso de construcción del tesoro HEREIN pudimos constatar muchas dificultades a la hora de determinar las equivalencias interlingüísticas. El establecimiento de este tipo de correspondencia entre términos es una operación, como poco, delicada y muchas veces controvertida.

En el grupo 2 del tesoro, que agrupa todos los términos relativos a los agentes, colectivos e individuales o, dicho de otro modo, a las instituciones y personas implicadas en la gestión del patrimonio, hubo dificultades especiales debidas a las diferentes estructuras territoriales de los países europeos y a la diversa organización administrativa de las competencias. Fue complicado, por ejemplo, unificar la definición y establecer la correspondencia entre ‘administración local y autonómica’ (español); ‘Lokalverwaltung’ (alemán), local administration (inglés) y administration décentralisée (francés). Otras dificultades parecidas hubo con el término inglés ‘non state administrations’, que solo es parcialmente equivalente a ‘collectivités territoriales’ (francés) y a ‘entes territoriales’ (español).

En caso del grupo 2, en el que se incluyen los tipos de bienes y los niveles de protección, los principales problemas para determinar las equivalencias se debieron principalmente a las diferencias en los niveles de protección. Así, para reflejar los distintos niveles europeos en la protección de los edificios, al término presente en la legislación española ‘edificios declarados de interés cultural’ (‘listed buildings’ en inglés y ‘édifices classés’ en francés), se le añadieron otros: ‘edificios históricos’ para reflejar la figura de protección inglesa ‘historic properties’, y ‘edificios declarados’ para que hubiera una correspondencia con ‘édifices inscrits’, concepto presente en la legislación francesa. Todos estos niveles se agruparon bajo el más general de ‘edificios protegidos’.

También se encontraron particularidades en la extensión y las restricciones en las «áreas protegidas» que hacían necesaria una armonización. Fue preciso definir y analizar el alcance de la protección y se comprobó que variaban los conceptos, y aunque se solapaban en gran medida, los términos no eran exactamente equivalentes. Se trabajó hasta encontrar la mejor correspondencia, aunque no fuera exacta, para términos como ‘zonas de respeto’ o ‘entorno protegido’. En algunos casos un concepto para el que en español se utilizaba un solo término, se correspondía con dos términos diferenciados en otros idiomas. Es el caso de ‘zonas de respeto’ y su relación de equivalencia con ‘contiguous zones’ y ‘curtilage’ en inglés. En caso de que en un mismo idioma la figura de protección de un área llevara distintas denominaciones según el país, se decidió que la correspondencia interlingüística reflejara esta diversidad denominativa añadiendo tras el término un código de país. Así, los equivalentes de ‘zonas de respeto’ para el francés son tanto ‘zones de protection’ (BE) como ‘abords’ (FR), con un código entre paréntesis para distinguir el término usado en Bélgica del utilizado en Francia.

En cuanto a elementos protegidos en los bienes inmuebles, en la mayoría de las lenguas se encontraba la distinción entre ‘elementos accesorios del inmueble’ y ‘elementos consustanciales

del inmueble' o se podía fácilmente ofrecer una traducción funcional. No obstante, en el caso del inglés había dos conceptos más específicos dentro de los 'elementos accesorios': 'movable fittings' y 'movable furnishings'. Este caso se resolvió de manera distinta para no forzar al resto de las versiones a incluir términos sin uso real: 'movable fittings' y 'movable furnishings' fueron considerados términos no preferentes y reenvían al término general 'fittings'.

Algunas denominaciones de categorías de bienes con particularidades en cuanto a la protección legal, como las 'reservas arquitectónicas' de Rumanía o los 'monumentos de la historia' de Polonia, finalmente se eliminaron del tesoro para asimilarlas a conceptos y denominaciones más genéricos, en un intento de no complicar excesivamente la correspondencia entre tipos de bienes.

Los términos agrupados en el grupo 3, relacionados con los sistemas de documentación, presentaron menos dificultades. Tras definir todos los conceptos se comprobó que, a pesar de la diversidad de denominaciones, había una mayor homogeneidad en los tipos de documentos e inventarios, y también en los procesos de recogida, análisis y recuperación de información sobre los bienes culturales.

En cambio, el trabajo de armonización terminológica tuvo que ser mucho mayor en el caso del grupo 4, que agrupa los términos referentes al sistema legal. Las causas principales de estas dificultades para establecer equivalencias interlingüísticas provenían de las particularidades de cada país en la propiedad y uso de los bienes, los distintos niveles de protección y procedimientos de declaración, las denominaciones diferenciadas según el ámbito geográfico, los distintos procedimientos de adjudicación de obras, las diferencias en los procesos de transmisión de la propiedad, y la inclusión de conceptos relevantes en algunas legislaciones nacionales que no existen otras. Así, hubo que encontrar soluciones en las otras lenguas para encontrar equivalentes a 'Patrimonio Nacional', pues solo en inglés se encontraba un término realmente funcional, 'crown property', y no tenía una correspondencia conceptual total con el término en español. En cuanto a procedimientos de declaración, hubo que incluir en la versión española el término 'exclusión'. La suspensión de la protección legal que sigue a la eliminación de un bien del patrimonio cultural del inventario o lista correspondiente no está contemplada o definida en la legislación española, pero había que proporcionar un término equivalente para este concepto que sí se utiliza en otros sistemas legales europeos: es 'delisting' en inglés y 'déclassement' o 'annulation de la protection' en francés, por ejemplo. En cuanto a los procesos de transmisión de la propiedad, los conceptos españoles 'derecho de tanteo' y 'derecho de retracto' se engloban en otros idiomas en un único concepto y término para designar el derecho reconocido en determinados casos a la administración o a entidades privadas que llevan a cabo una misión de servicio público para adquirir con preferencia la propiedad de un bien: 'droit de préemption' en francés, 'vorkaufsrecht' en alemán o 'right of preemptive purchase' en inglés.

Finalmente, es interesante destacar que hubo conflictos en casos en los que los términos tienen marcadas connotaciones culturales. Al abordar la terminología referida a actos ilícitos o delitos contra el patrimonio, se intentó en varias ocasiones que se eliminara el término 'expolio' del tesoro y se dejara únicamente 'pillaje'. Finalmente, dada la relevancia de este concepto en la legislación española, se logró su inclusión, tras largas discusiones, en todas las versiones lingüísticas: 'plundering' en inglés, 'spoliation' en francés y en alemán, 'espoliação' en portugués, etc. Quedó así de manifiesto que cada lenguaje natural es portador de sus propias implicaciones denotativas, connotativas, valorativas y emocionales.

Variación terminológica debida a neologismos y a calcos en las traducciones de los «textos internacionales»

En el periodo principal de desarrollo del tesoro (2000-2004), se estaba produciendo una paulatina creación de neologismos en las tipologías del patrimonio, influida de forma determinante por la evolución del pensamiento y el cambio en los valores sociales. La producción neológica

se concentraba especialmente en la denominación de nuevas figuras de protección, así como en los nuevos espacios y modelos para comunicar el patrimonio al público visitante. En estos casos era muy frecuente que la primera denominación del concepto fuera un calco del inglés o del francés que aparecía en traducciones de recomendaciones o convenciones internacionales. En ocasiones convivían ambos calcos, pero en el lenguaje natural de los textos escritos originariamente en castellano se podía observar como uno de los dos se iba imponiendo en el uso profesional. Como ejemplo, el calco del inglés ‘patrimonio intangible’ fue cediendo a favor del calco del francés ‘patrimonio inmaterial’. El tesoro HEREIN refleja estos fenómenos en las relaciones de equivalencia intralingüística: ‘patrimonio intangible’ USE ‘patrimonio inmaterial’.

La creación neológica no solo influyó en la necesidad de reflejar en el tesoro los casos en los que un concepto de introducción reciente se designa con demasiadas expresiones sinónimas, como era el caso de la ‘recreación histórica’ o ‘historia viva’. También la variación terminológica, junto con los calcos, dificultaba el establecimiento de las equivalencias interlingüísticas. El caso más conflictivo tuvo origen en una traducción del francés al inglés en un texto internacional: ‘archéologie préventive’ se tradujo como ‘preventive archaeology’ sin tener en cuenta cuál era la denominación real en el uso profesional en inglés. En muchas lenguas esta traducción al inglés se tomó como referencia y muchos miembros del equipo de desarrollo del tesoro querían que ‘preventive archaeology’ constara y fuera el término preferente en la versión inglesa. Finalmente, y tras dificultosas negociaciones, solo figura en el tesoro el término especializado ‘contract archaeology’.

Valoración de la experiencia y conclusiones

La experiencia adquirida en el proceso de elaboración del tesoro HEREIN corroboró que la elaboración de tesauros multilingües supone algo más que encontrar equivalentes para conceptos y términos. Hay una dimensión cultural en el proceso y, de hecho, cuando se aplica la metodología apropiada, podría ser más apropiado referirse a construcción de tesauros multiculturales más que a tesauros multilingües. Por otra parte, existe una dimensión política en la construcción de tesauros multilingües, especialmente en lo que concierne a lenguas que no están, contextualmente, en la misma condición de prestigio o reputación.

La evaluación externa del tesoro HEREIN en cuanto a metodología fue muy positiva. Se reconoció el esfuerzo considerable dedicado a asegurar que en el curso del desarrollo del Tesoro a todos los idiomas se les diera el mismo peso. La defensa de esta pauta supuso un desafío y fue una tarea compleja, pero el hecho de que ningún idioma fuera tratado como el idioma principal o de referencia se consideró esencial por los evaluadores externos, no solo con el fin de respetar la diversidad cultural y lingüística subyacente que existe dentro de Europa, sino también para garantizar un producto de mejor calidad.

En cambio, la evaluación externa en cuanto a implementación en el sitio web no fue tan positiva, pues se destacaron algunas mejoras necesarias con respecto al diseño y la lógica de la interfaz de usuario, así como el diseño gráfico de las propias páginas. Ciertamente la usabilidad es un aspecto muy importante en tesauros en línea multilingües tan complejos como el tesoro HEREIN.

Estos defectos de diseño se podrían haber evitado si se hubieran podido aprovechar en el momento de desarrollo las ventajas de un gestor de tesauros en el entorno digital. Una herramienta de estas características puede enriquecer la funcionalidad de la estructura de los tesauros a partir de la hipertextualidad (establecimiento de hipervínculos entre los términos y las partes del vocabulario), y reducir costes de actualización y mantenimiento. También estos gestores permiten integrar al usuario en el proceso de creación, gestión y optimización a través de tests.

En la experiencia de desarrollo del tesoro HEREIN y otros recursos similares, se ha visto que es recomendable atender también desde el inicio del proyecto a los aspectos de integración e interoperabilidad.

En cuanto a integración en otros sistemas, es importante contemplar la exportación en diferentes formatos: texto (txt), Zthes, Skos-Core, TopicMap (xtn), BS8723, IMS Vocabulary Definition Exchange (VDEX), Wordpress XML (WXR), Site Map, SQL, etc.

Por otra parte, los protocolos, marcos de metadatos, esquemas XML, etc., están ya diseñados para lograr interoperabilidad en la web semántica.

Finalmente, cabe señalar que es aconsejable en cualquier proyecto de elaboración de tesauros valorar el coste-beneficio del desarrollo de una herramienta tan compleja y estructurada, prever su sostenibilidad y mantenimiento, así como preparar desde el inicio su integración total con el sistema de información para el que se crea.

Bibliografía

- BENAVIDES, P. (1999): *Diccionario razonado de bienes culturales*. Sevilla: Padilla editores.
- ABEJÓN, T.; MALDONADO, Á., y FLORIÁN, L. (2006): «Contribución al estudio de la terminología de la gestión del patrimonio cultural y del turismo cultural a través de los datos del proyecto ESLEE», en *La terminología en el siglo XXI. Contribución a la cultura de la paz, la diversidad, la sostenibilidad*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 471-482.
- LAFUENTE, I. T. (2013): *Tesaurus y Diccionario de objetos asociados a la expresión artística. Tesaurus y Diccionario para la descripción y catalogación de bienes culturales*. Madrid: Ministerio de Educación, Cultura y Deporte.
- TEBÉ, C. (2001): «Elementos para la representación de los conceptos especializados en bases y bancos de datos», en *La Terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 203-208.
- Tesaurus HEREIN [en línea] <<http://www.european-heritage.net>>.

Creación, ejecución y desarrollo de un tesoro temático para los museos aragoneses: las tipologías arqueológicas

José Fabre Murillo

Conservador de Arqueología – Museo de Huesca. Coordinador DOMUS Aragón 2011-2015. Dir. Gral. Cultura y Patrimonio, Gobierno de Aragón

Resumen: Se recoge en la presente comunicación la evolución de un proceso de creación de un tesoro temático: el *Tesoro de tipologías arqueológicas en los museos de Aragón*. Desde el diagnóstico de un caos terminológico, pasando por el diseño de una herramienta que permitiera su uso a través del sistema DOMUS para, finalmente, ponerla en práctica en los museos arqueológicos aragoneses. Finaliza el mismo con el acercamiento a la publicación –2013– que presenta al público en general el trabajo realizado.

Abstract: We are going to explain how we have get to obtain a tematic thesaurus, we talk about the «Tesoro de Tipologías en los museos de Aragón», a thesaurus that includes all archeological types deposit in the Museums in Aragon. First, we talk about the requeriment, sencondly we will explain how we make it and apply in the museums. Finally we will finish this lecture doing a brief look to the book –published in 2013–that shows work done openly.

Sírvanos este texto para poner negro sobre blanco y de forma más o menos ordenada aquellas ideas, conceptos y reflexiones que se han expresado en torno al tesoro de tipos arqueológicos. Herramienta desarrollada para la correcta documentación de las colecciones de carácter arqueológico custodiadas en museos aragoneses en el marco del curso «El lenguaje del patrimonio: Estándares documentales para la descripción y gestión de colecciones».

Si lo que se ha pretendido ha sido mostrar el origen, desarrollo y ejecución de este proyecto, es preceptivo hacer un par de aclaraciones previas. En primer lugar, en cuanto a la autoría del mismo; como coordinador del proyecto me presento como voz autorizada, una de ellas, para relatar sus pormenores y detalles, pero sobre todo para cantar las bondades de un buen trabajo

en equipo¹. En segundo lugar, exponer que esta herramienta fue presentada a la comunidad museográfica y arqueológica de forma específica, desde el punto de vista de la disciplina arqueológica, personalizada, mediante la explicación de criterios seguidos por cada autor en cada una de las áreas de conocimiento, y en el ámbito –no excluyente– de aplicación del tesoro, estos es, para los museos radicados en Aragón².

Es por ello que asumimos gustosos la invitación a estas interesantes jornadas de trabajo con el fin de presentar al otro lado del *limes aragonensis* el presente proyecto, pero sobre todo contextualizarlo en un foro con amplias miras sobre la documentación y digitalización del saber cultural.

Antecedentes

Y como resulta cómodo empezar por el principio, hemos de referir que esta herramienta ha nacido para ser usada en un ámbito concreto y que, por tanto, su razón de ser es consecuencia de una necesidad. Hablamos de la implantación en los museos aragoneses del Sistema de Gestión y Documentación Museográfica DOMUS, que desde el año 2004 viene utilizándose de forma incremental³.

No es momento de realizar una inmersión en los pormenores de DOMUS, nosotros mismos, y otros antes, mejor que nosotros, lo han hecho ya (Carretero *et alii*, 1994; Alquézar, 2004; Fabre y Lorente, 2009, por citar solo algunos), sin embargo, sí es positivo destacar que estamos ante un sistema informático fundamentado en una profunda reflexión teórico-práctica que pretende dar cabida a toda la acción documental generada en un museo no solamente como mecanismo de salvaguarda de conocimiento museográfico, sino como herramienta de gestión diaria. Plantea una gestión integral [interdepartamental] a través del control de los instrumentos documentales de carácter técnico (inventarios y catálogos, archivo de documentación gráfica, archivo de conservación-restauración), administrativos (registros de entrada y salida de fondos museográficos, entradas temporales ajenas a la colección, directorio, gestión de tienda, material inventariable, taquilla...) y bibliográficos. Se basa en tres pilares fundamentales:

- Centralización e intercomunicación de esos instrumentos documentales de los distintos departamentos de un museo.
- Normalización y sistematización de los procesos de documentación de dichos departamentos.
- Normalización de la terminología utilizable en todo el Sistema.

Este tercer punto es el que nos interesa en este momento. A través de una categorización de los conceptos que ha de referir el sistema en todos sus módulos (áreas de trabajo de cada departamento: conservadores, fotógrafos, restauradores, administrativos, etc.), se establecen cuatro niveles de control terminológico (Asín y Fabre, 2009):

- 1.º Campos de libre escritura. Menor nivel de control lingüístico para aquellos campos que no requieran terminología específica, e ideados para aquellas parcelas de información que necesiten cierta redacción. En estos casos el cumplimentador es libre en la selección del lenguaje que se utiliza en el sistema.

¹ Los autores tanto del Tesoro tipológico en faceta de aplicación en el Sistema de Gestión y Documentación Museográfica DOMUS, así como en la publicación y tipoteca gráfica que se editaron en 2011 son en orden alfabético: C. Aguarod Otal (Unidad de Museos y Exposiciones Ayto. Zaragoza), V. Baldellou (Museo de Huesca), M. Beltrán Lloris (Museo de Zaragoza), R. Erixe Lacabe (Unidad de Museos y Exposiciones, Ayto. Zaragoza), C. Escriche Jaime (Museo de Teruel), B. Ezquerro Lebrón (Museo de Teruel), J. Fabre Murillo (Dir. Gral. Patrimonio Cultural del Gobierno de Aragón) y J. A. Paz Peralta (Museo de Zaragoza).

² En el marco de las III Jornadas de los Museos Aragoneses desarrolladas en mayo de 2013 y dedicadas a los museos arqueológicos, se presentaron hasta cuatro comunicaciones explicativas de este proyecto. Sus actas se encontrarán publicadas en formato digital a partir de 2016.

³ BOE n.º 132 de 1 de junio de 2004 y sucesivas adendas en 2007, 2008 y 2010, hasta el momento, dando cabida a más de cuatro decenas de museos y colecciones de la más diversa índole.

- 2.º Listas abiertas de términos. Campos de información conceptualizados (el título de una obra, lugares específicos de procedencia de una pieza...) que han de huir de la tan temida duplicidad de términos referentes (Retrato del rey, Retrato del rey Fernando VII, Fernando VII de Goya, Fernando VII el Deseado, etc.), pero que, sin embargo, requieren cierta capacidad de creación para el cumplimentador. Se materializa en una lista de términos alfabéticamente ordenados que permite a cualquier cumplimentador dar de alta un nuevo término si no es preexistente y lanza un mensaje al cumplimentador que le hace reflexionar si es estrictamente necesaria dicha creación del término.
- 3.º Listas cerradas de términos. A diferencia de las anteriores, estaríamos ante aquellas parcelas de información que se estiman normalizadas en cuanto a sus términos componentes, que se ordenan alfabéticamente, pero que pueden ser innovadas o completadas con nuevos términos si un usuario cualificado, no un simple cumplimentador, lo determina. Las formas jurídicas por las que una pieza puede ingresar en un museo (donación, depósito, hallazgo...) sería un claro ejemplo de aquellas parcelas sujetas, incluso por normativa jurídica, a una terminología muy ajustada. Sin embargo, cuenta con la posibilidad de ampliarse llegado el momento y la necesidad.
- 4.º Por último, el tesoro sería el nivel de mayor control que prevé el Sistema. El número de términos que componen una parcela o campo de conocimiento, previsiblemente mucho mayor que en los casos anteriores, haría necesaria una herramienta más potente. Sin embargo, la diferencia con una lista alfabética férreamente controlada no es el número de términos que contiene sino las relaciones existentes entre esa terminología.

Cada término del tesoro es dado de alta por el administrador del sistema en atención al control y convenciones ortográficas, lingüísticas, etc., y pasando los siguientes controles:

- 1.º Cada término debe estar relacionado jerárquicamente con el resto de términos del tesoro. A través de una categorización de términos genéricos, específicos o relacionados (subordinación conceptual). Muy ilustrativo nos resulta el ejemplo del tesoro de lugares geográficos en el que *Europa*, *España* y *Aragón* son tres términos del mismo tesoro que entre ellos mantienen la siguiente relación jerarquizada:

Europa
 España
 Aragón

En ella se han dado de alta tres términos en los que se establece una relación de jerarquía en la que *Europa* tiene un término específico que es *España*, el cual a su vez contiene otro término específico que es *Aragón*.

- 2.º El hecho de que un mismo concepto pueda ser identificado con distintos términos se resuelve en el tesoro adjudicando a su administrador la posibilidad de crear términos preferentes (USE) que recojan el concepto en cuestión sin perder las sinonimias que pueda presentar (USADO POR).

USE - *Huesca* USADO POR - *Ozca / Wasqa / Uesca*

- 3.º Por último, el término por crear debe estar bien determinado como término indicador-clasificador, es decir, términos que ayudan a ordenar conceptualmente el tesoro. O bien como términos descriptores, que son la parte activa del tesoro. De nuevo, un menguado tesoro geográfico nos ayudará a explicar este particular:

Europa - Término descriptor

[Monarquía parlamentaria] - Término indicador-clasificador

Bélgica - Término descriptor

España - Término descriptor

Reino Unido - Término descriptor

[República] Término indicador - Clasificador

Francia

Italia

Portugal

El Tesoro de tipologías arqueológicas

En el año 2008, la Coordinación DOMUS Aragón, asumiendo estas premisas, se enfrentó al reto de abordar la redacción y puesta en práctica en un tesoro que acogiera las distintas nomenclaturas recibidas por las variantes de objetos arqueológicos en nuestras colecciones (Fabre, 2016, en prensa –e.p.–). Para ello se creó un grupo de trabajo denominado Comisión de Redacción del Tesoro de Tipologías, entre un plantel de directores, conservadores y técnicos de museos (véase nota n.º 1) cuya característica en común era su específica formación y experimentado conocimiento de la cultura material procedente del amplio contexto cultural planteado: los objetos hallados o fabricados por las distintas sociedades desde la prehistoria (Baldellou y Fabre, 2016 e.p.), pasando por la Edad Antigua (Beltrán *et alii*, 2016 e.p.) y llegando hasta la Edad Media (Escriche, 2016 e.p.).

El resultado tras cuatro años de trabajo ha sido la creación de un tesoro con más de 4000 términos y más del doble de relaciones terminológicas. El objetivo de la Comisión ha sido desarrollar lenguajes de descripción homogéneos, y establecer claramente el orden en la jerarquía a la hora de nombrar esos objetos. Por lo tanto, no solamente se asumió un *tesoro de tipologías*, es decir, no se han ordenado los términos que definen la variante de este objeto, sino que para ello hemos tenido que hacer un ejercicio previo.

Fue necesario prever una estructura terminológica, un árbol, capaz de seguir creciendo y que permita albergar cualquier término que se pueda incorporar o modificar en un futuro.

Un reto imprevisto, pero predecible al que se enfrentó la llamada Comisión, fue crear el *tesoro de nombres comunes*. Si hablamos de variantes de objeto, es fundamental que esos objetos tuvieran una definición conocida y reconocible. Había que definir el objeto.



Figura 1. Cuenco TSG Dragendorff 29. Museo del Teatro de Caesar Augusta NIG00071 (foto José Garrido).

Si haciendo un breve experimento presentáramos la figura adjunta a un conjunto de arqueólogos cualquiera, sería muy probable que la nomenclatura adjudicada a ese objeto fuera múltiple. Muchos se inclinarían por definirla como *Terra sigillata*; si tal fuera el caso, estaríamos acercándonos al término que nos definiría un aspecto técnico, un momento concreto, una tradición cerámica histórica, pero en ningún caso nos define el objeto. Otros podrían pensar, probablemente, en que están ante una *Dragendorff 29*, porque conocen los estudios de este investigador, pero no estarían hablando del nombre del objeto que tienen delante. En ninguno de los casos de los que hemos formulado se habría dicho que estábamos ante un 'Cuenco'. Y esto, por muy obvio que pueda parecer *a priori*, fue uno de los primeros pasos atrás que tuvimos que dar para poder continuar en nuestro trabajo.

El objeto y su nomenclatura conocida y reconocida era la llave de acceso al ordenamiento del caos.

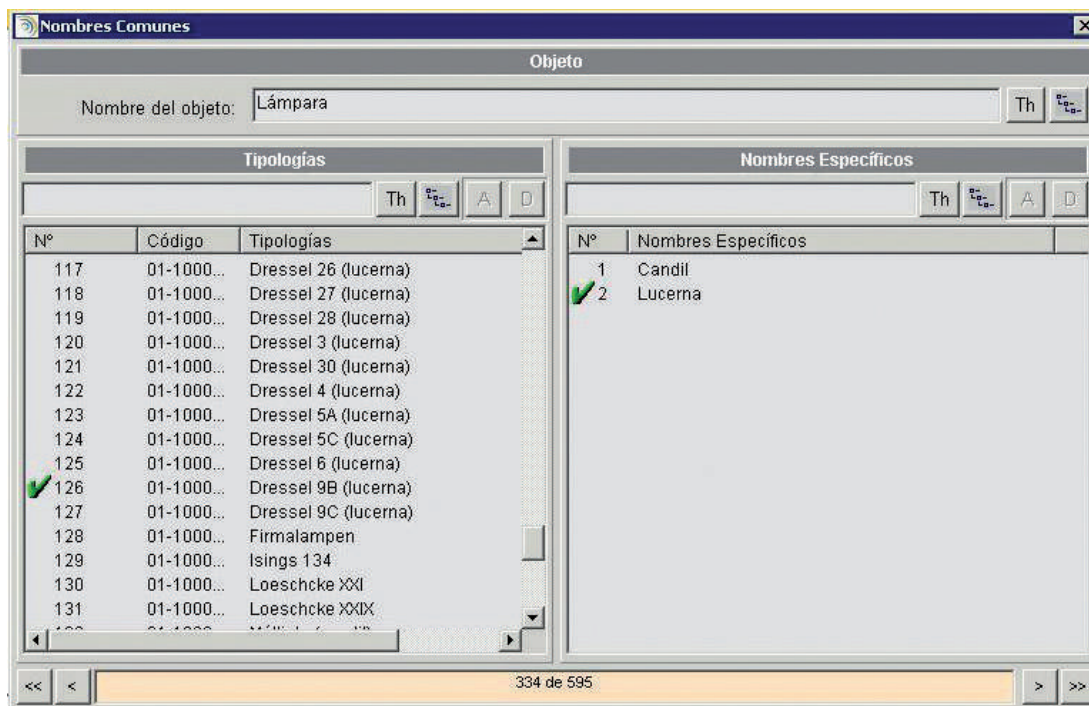


Figura 2. Tesaurus independientes pero conectados conceptualmente. La información relacionada en DOMUS.

Lógicamente, nació así el *tesaurus de nombres comunes* de objetos arqueológicos de los museos de Aragón. Se identifican así todas las formas tipo, todos los objetos que tenemos desde la prehistoria hasta la actualidad con una definición recogida, en los casos en los que ha sido posible, en el *Diccionario de la Lengua Española*, para, a continuación, recoger aquellas definiciones técnicas, consensuadas, que en cada una de las disciplinas se han podido desarrollar. Es decir, que hemos intentado tener una definición exacta por objeto.

Una vez alcanzada esa convención, lo que teníamos que hacer era crear una serie de clasificadores generales que pudieran albergar a las varias decenas de morfotipos arqueológicos identificados.

Se podrían haber clasificado genéricamente estos objetos en atención a muchos criterios, pero, dado que estamos hablando de artefactos, de objetos, que se crean por su utilidad, se consideró oportuno utilizar una serie de términos indicadores-clasificadores que atienden, precisamente, al uso que tuvieron en su creación. Se han consensuado los diecinueve indicadores:

1	Armamento
2	Indumentaria y complementos
3	Menaje de cocina
3.1.	Cocina. Cocción
3.2.	Cocina. Preparación
4	Menaje de mesa
5	Útil de aseo e higiene
6	Útil de caza, pesca o recolección
7	Útil de escritura y elaboración de documentos
8	Útil de farmacopea y perfumería
9	Útil de filtrado
10	Útil de función diversa
11	Útil de iluminación
12	Útil de lavado
13	Útil de perforación, corte o abrasión
14	Útil de trabajo doméstico
15	Útil de transporte y almacenaje
16	Útil de unión, sujeción, enganche o cierre
17	Útil lúdico
18	Útil ornamental
19	Útil ritual

Figura 3. Indicadores-clasificadores de morfotipos por función.

Además de ser ampliable con nuevos términos indicadores-clasificadores, permite que los morfotipos puedan ser relacionados con más de un clasificador. De esa manera, el término ‘Vaso’ se presenta vinculado a ‘Menaje de mesa’, ‘Útil ritual’, ‘Útil de aseo e higiene’, etc.

Definido y clasificado el objeto, era cuestión de recoger sus distintas voces a lo largo de la historia. Por ejemplo, encontramos ‘Lámpara’ desde la prehistoria hasta nuestros días, y en muchas ocasiones contamos con el saber necesario para conocer cómo eran denominados esos útiles por las sociedades que los produjeron, a saber, lámpara, lucerna, candil... Creábamos así el tesoro de nombres específicos.

Seguimos avanzando con el ejemplo de otro objeto: ‘Caja’. Le hemos dicho al catalogador qué es una caja, cómo va a identificar una caja de forma indudable, y, además, le hemos dicho que tiene un gran cajón clasificador, en este caso, que es un ‘*útil de transporte y almacenaje*’. Le hemos advertido que puede tener asociado a él el término ‘*Pixis*’, que hará referencia a la nomenclatura de ese objeto en un momento muy concreto de la Historia. Y dentro de ellos ya encontramos la tipología; las variantes de barniz negro, ibérica, *terra sigillata*, etcétera. Los términos preferentes así como sus sinónimos para que ningún término se perdiera en el camino.

algunos anexos: un anexo con unas pautas de catalogación en atención a las distintas tradiciones descriptivas por tipos de industria –la industria lítica, la industria ósea, la industria cerámica, la industria metálica–, que vienen a ser unos requisitos mínimos que el catalogador que no conozca la catalogación de objetos arqueológicos con fluidez va a tener como apoyo a la hora de describir una pieza arqueológica con rigor.

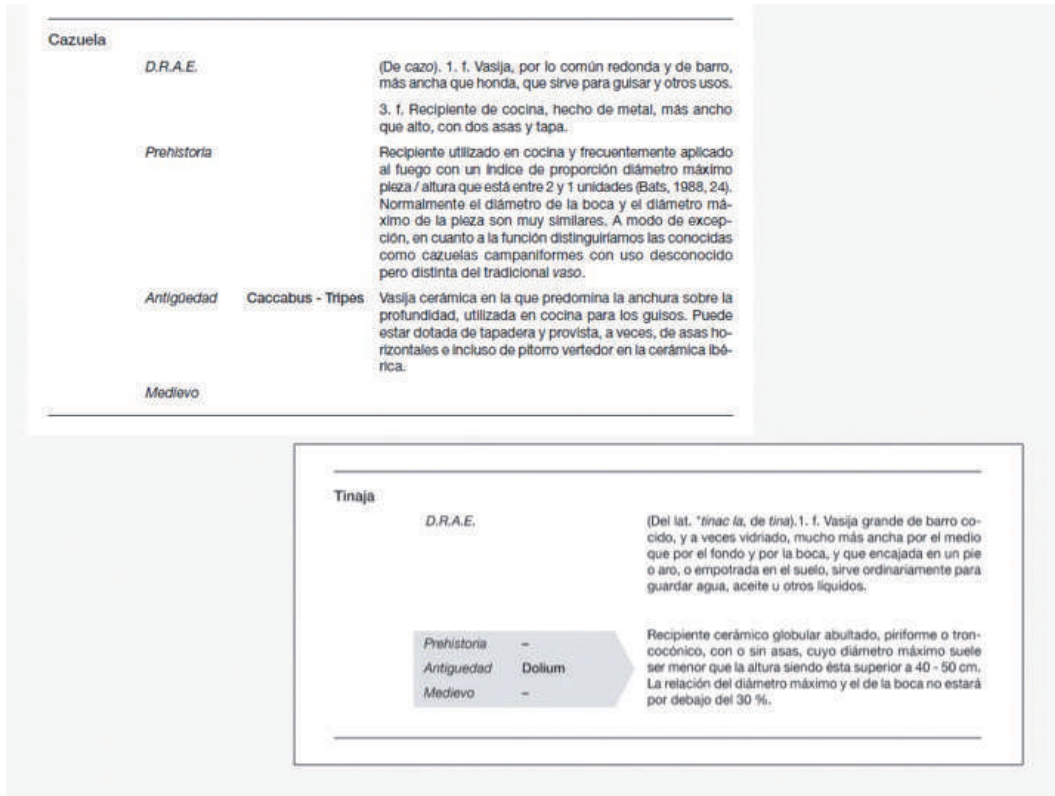


Figura 5. Dos ejemplos de definición de morfotipos.

Se han incluido todas las fuentes científicas de las que bebe la obra; en la mayoría de los casos, no hemos creado nuevas tipologías, lo que hemos hecho es decidir cuáles eran las fuentes tipológicas que considerábamos preferentes a la hora de ordenar nuestros objetos.

La obra se cierra con dos anexos gráficos: en primer lugar, lo que hemos introducido es un compendio de las piezas arqueológicas más significativas de las colecciones aragonesas, del orden de cuatrocientas imágenes. Y, además de ello, se consideró oportuno acompañar a esta publicación de un apéndice gráfico que reflejara todos y cada uno de esos términos que hemos normalizado. Es decir, los 4000 términos están ilustrados en el CD anexo que acompaña a la publicación en papel y que pretendemos en un futuro cercano sea accesible, al igual que el texto, a través de la red.

Lógicamente, acompañar la publicación de los 4000 dibujos era, desde el punto de vista del formato, bastante complicado. Por lo tanto, lo que se propuso fue incorporar una sencilla aplicación informática que viene a ser un navegador de tipos. Este navegador de tipos incorpora una tipoteca en la que, a través de la misma estructura conceptual, se puede recuperar hasta la última variante tipológica que aparece reflejada en el tesoro.



Figura 6. El navegador de tipos, anexo gráfico de la publicación.

A modo de conclusión, conviene recalcar la bondad de los trabajos en equipo cuyos objetivos son claros, definidos en el tiempo y cuya practicidad es inmediata. Este ha sido un claro ejemplo de ello.

En segundo lugar, y específicamente sobre aquellos tesauros de naturaleza temática amplia, como pudiera ser el referido, es imprescindible el consenso, el acercamiento de posiciones y la asunción de que son productos intelectuales orgánicos sujetos a cambios y actualizaciones constantes, pero, sin lugar a dudas, la cotidianeidad en su uso es la prueba irrefutable de su validez.

Bibliografía

- AGUAROD OTAL, C. *et alii* (2011): *Tesauro tipológico de los Museos Aragoneses. Colecciones arqueológicas*, vol. I. J. Fabre Murillo (coord.). Zaragoza: Gobierno de Aragón.
- ASÍN MARTÍNEZ, L., y FABRE MURILLO, J. (2009): «Tesauros y lenguajes documentales: su elaboración, razón de uso y operatividad en el Sistema de Documentación DOMUS», en *Actas de las I Jornadas de Museos de Aragón*. Zaragoza: Gobierno de Aragón.
- BALDELLOU, V., y FABRE MURILLO, J. (2016): «De la tipología de la prehistoriador a la normalización del museo», en *Actas de las III Jornadas de Museos Aragoneses. Los museos arqueológicos*. Museo de Zaragoza, 7 y 8 de mayo de 2013. Zaragoza: Gobierno de Aragón. En prensa.
- BELTRÁN LLORIS, M. *et alii* (2016): «De la tipología de la Arqueología Clásica a la normalización del museo», en *Actas de las III Jornadas de Museos Aragoneses. Los museos arqueológicos*. Museo de Zaragoza, 7 y 8 de mayo de 2013. Zaragoza: Gobierno de Aragón. En prensa.

- CARRETERO PÉREZ, A. *et alii* (1994): *La Normalización Documental de Museos*. Madrid: Ministerio de Cultura, Subdirección General de Museos Estatales.
- ESCRICHE JAIME, C. (2016): (e.p.). «De la tipología de la Arqueología Medieval a la normalización del museo», en *Actas de las III Jornadas de Museos Aragoneses. Los museos arqueológicos*. Museo de Zaragoza, 7 y 8 de mayo de 2013. Zaragoza: Gobierno de Aragón. En prensa.
- FABRE MURILLO, J., y LORENTE ALGORA, M. (2009): *Domus en los Museos Aragoneses. Convenciones del Sistema. V 1.0*. Zaragoza: Gobierno de Aragón.
- PADILLA MONTOYA, C. *et alii* (2002): *Diccionario de materiales cerámicos*. Madrid: Ministerio de Cultura.

El listado de autoridades y la normalización terminológica en el Museo del Prado

José Luis Cueto Martínez-Pontrémuli

Museo Nacional del Prado

Resumen: Se analiza el sistema de información del Museo del Prado y se aborda la cuestión del control terminológico en las bases de datos del Área de Biblioteca, Documentación y Archivo desde una perspectiva global. Mediante la realización de un mapeo que compara los campos de la norma ISAAR (CPF), el formato MARC21 de autoridades y las estructuras de datos del Servicio de Documentación, se estudian las posibilidades de convergencia de las autoridades de nombre de los tres departamentos. A continuación, se proponen unas pautas iniciales de normalización con el objetivo de diseñar en el futuro una tabla compartida de autoridades que nos permita ofrecer al usuario el acceso de forma conjunta a los libros, obras de arte y documentos que conserva el museo. Este proyecto ha avanzado en paralelo con el diseño de la nueva página web del Museo del Prado, según los estándares de la web semántica, y con la voluntad de la institución de ofrecer en Internet una información cada vez más completa y contextualizada sobre sus colecciones.

El Área de Biblioteca, Documentación y Archivo del Museo del Prado: hacia un sistema global de información

El sistema de información del Museo del Prado se sustenta sobre tres pilares: la Biblioteca, que contiene las publicaciones especializadas en historia del arte, abarcando el arco cronológico que afecta a las colecciones; el Archivo, donde se custodian los documentos producidos por el museo a lo largo de su actividad institucional, y el Servicio de Documentación, encargado de gestionar la información sobre las obras de la colección a través de un sistema integrado.

En 2008 se decidió la unificación de los tres departamentos dentro del Área de Biblioteca, Documentación y Archivo en el edificio del Casón del Buen Retiro, que abrió sus puertas en marzo de 2009, poniendo en práctica una concepción global de la gestión y difusión de la información.

Debido a la importancia creciente de la difusión social del conocimiento generado por el Museo del Prado, la actividad de sus servicios de información tiene una repercusión cada vez más notable en la página web del museo (Martín; Pantoja, 2013): en un primer momento, según las posibilidades de una página web creada en 2007; y, posteriormente, con un proceso de adaptación a las nuevas necesidades de normalización impuestas por la web semántica.

En los últimos años el Museo del Prado ha trabajado, junto con la empresa de software GNOSS, en el diseño de su nueva página web, según el modelo conceptual CIDOC-CRM, motivo por el que el Área de Biblioteca, Documentación y Archivo ha iniciado un proceso de estudio de sus autoridades de nombre en los tres departamentos, con el objetivo de diseñar una tabla única de autoridades que permita la recuperación de los contenidos de las tres bases de datos de forma conjunta. A partir del cumplimiento de esta primera fase de normalización de las autoridades de nombre a nivel interno, será posible plantear en el futuro la existencia en la página web de relaciones semánticas entre las fichas de catalogación de las obras, los registros de la Biblioteca y documentos digitalizados del Archivo.



Figura 1. Sala de lectura Luca Giordano.

Servicio de Biblioteca

La Biblioteca del Museo del Prado está formada por más de 75 000 títulos de monografías y 1300 títulos de revistas, junto con materiales audiovisuales y una importante colección de catálogos de subastas que está en proceso de catalogación, además de manuscritos y un fondo antiguo con más de 5500 volúmenes anteriores a 1900. Gracias a estos fondos, se consolida en la actualidad como una de las bibliotecas especializadas en arte más importantes de nuestro país.

El Servicio de Biblioteca gestiona, dentro del entorno AbsysNet 2.0, sus registros bibliográficos y de autoridad de acuerdo con los principales estándares internacionales de referencia en ámbito bibliotecario:

- En la descripción bibliográfica se siguen las vigentes Reglas de Catalogación Españolas, que, como es bien sabido, derivan de las pautas marcadas por la Descripción Bibliográfica Internacional Normalizada (ISBD).
- En el área del control de autoridades, se respeta la estructura definida por las Directrices para Registros de Autoridad y Referencia (GARE, Guidelines for Authority and Reference Entries) y su revisión publicada en 2001 como GARR, Guidelines for Authority Records and References.
- Se recurre a los formatos MARC21 (MACHINE-Readable Cataloging) para adaptar los estándares citados a una estructura informática comprensible por los ordenadores.

Para crear sus autoridades el Servicio de Biblioteca, utiliza el catálogo propio de autoridades, denominado AUTO. Además, a través del protocolo Z39.50, la Biblioteca utiliza la descarga de autoridades de algunas instituciones de referencia, como la Biblioteca Nacional de España y la Library of Congress. También se emplea como fuente de referencia el Union List of Artist Names (ULAN) del Getty Research Institute junto con otras importantes fuentes como el Diccionario Bénédiz de Artistas, accesible en línea a través de la plataforma Oxford Art Online.

Servicio de Documentación

Otro de los pilares del sistema de información del Museo del Prado es su Servicio de Documentación, cuya función es analizar todo tipo de documento que aporte información sobre las colecciones de la institución y volcarla posteriormente en una base de datos de consulta general. También se encarga de coordinar la incorporación de los datos que llevan a cabo los diferentes departamentos que están integrados en la herramienta de mantenimiento SIGMA, Sistema de Información y Gestión Museográfica (Docampo; Martín, 2009). De esta forma, las informaciones derivadas de la gestión de colecciones por parte de varios departamentos, que han sido previamente identificadas como relevantes para el usuario final, convergen en SAC, Sistema de Acceso a las Colecciones, último interfaz de consulta realizado a partir de 2010 por el Servicio de Informática.

La integración de información y contenidos procedentes de distintas bases de datos se ha llevado a cabo de dos formas:

- Mediante tablas de términos controlados que son compartidas por los departamentos implicados en la gestión de colecciones, como la Tabla de Autores utilizada por el Servicio de Documentación, Registro, Exposiciones Temporales, Archivo Fotográfico y Restauración.
- Con la creación de hipervínculos que conectan los registros del catálogo de Biblioteca y los documentos digitalizados del Archivo con las fichas de catalogación de las obras de arte.

En el diseño de las tablas de control terminológico del Servicio de Documentación se sigue un modelo de datos propio que tiene en cuenta la *Normalización Documental de Museos* publicada por el Ministerio de Cultura en 1996, así como los diccionarios terminológicos desarrollados por los Museos Estatales, por lo que existe una cierta afinidad con la línea seguida por estas instituciones con las adaptaciones que se han considerado necesarias para la naturaleza del Museo del Prado y sus colecciones.

El actual proyecto de normalización de autoridades de nombre pretende generalizar el trabajo con tablas compartidas del Servicio de Documentación hacia una visión más amplia en la que se incluyan la Biblioteca y el Archivo.



Figura 2. Depósito central de monografías.

Servicio de Archivo

El tercero de los pilares del actual sistema de información del Museo del Prado es el Archivo, formado por un conjunto de más de 4500 cajas que contienen la documentación producida por la institución desde su fundación, y 10 500 registros digitalizados que son accesibles al usuario desde los terminales de la Biblioteca. En la actualidad todo el fondo del Archivo está informatizado hasta nivel de expediente y se está trabajando desde 2010 en la descripción y digitalización de su fondo antiguo.

El mantenimiento y consulta del Archivo del museo, que también se realizaba en el pasado a través de una base de datos elaborada por el Servicio de Informática, se ha actualizado en el año 2012 al programa de archivo Albalá, y a finales de 2015 se produjo el cambio hacia Albalá-Net, todo lo cual ha permitido una progresiva adaptación a la Norma Internacional General de Descripción Archivística ISAD (G), al mismo tiempo que los campos que presenta este programa para el control de autoridades facilitan el trabajo de acuerdo con la Norma Internacional sobre Encabezamientos Autorizados Archivísticos para Entidades, Personas y Familias ISAAR (CPF).

Para la consulta del Archivo, se ha optado por el recolector MediaSearch, que está preparado para recuperar contenidos de diferentes entornos de datos mediante el protocolo OAI-PMH, y que ofrece para el Museo del Prado la ventaja de ser compatible tanto con AlbaláNet como con AbsysNet, lo que abre una vía en el futuro para la recuperación de las autoridades de nombre en los tres servicios de información una vez solventada la posible dificultad de conexión con el modelo de datos propio de Documentación.



Figura 3. Sala auxiliar de lectura.

Proyecto de Normalización de Autoridades

Durante los últimos años los Servicios de Biblioteca, Documentación y Archivo del Museo del Prado han avanzado en materia de normalización terminológica de forma independiente, si bien la interconexión de contenidos planteada por el proyecto de web semántica nos ha llevado a comprender la necesidad de unificar criterios a una escala mayor que la de cada una de estas tres bases de datos.

Con este objetivo, hemos iniciado el estudio del estado del control terminológico y los modelos de datos seguidos en los Servicios de Biblioteca, Documentación y Archivo, comenzando por las autoridades de nombre de persona. En cuanto al diferente estado del control terminológico en los tres departamentos y la gestión que realizan del control de autoridades, conviene tener presente los siguientes aspectos:

- El Servicio de Biblioteca muestra el mayor grado de normalización y adecuación a los estándares internacionales dentro del conjunto del sistema de información del Museo del Prado. El seguimiento de las Reglas de Catalogación Españolas y su adecuación a MARC21 lo convierten en punto de referencia para el proyecto de normalización.

- Hasta 2011 la Biblioteca creaba sus encabezamientos de materia con una cadena única y cerrada formada a partir de la suma de los distintos subencabezamientos necesarios para la autoridad, mientras que con posterioridad a esa fecha se ha empezado a trabajar con autoridades independientes, que se pueden unir para crear un nuevo encabezamiento de materia. Este hecho facilitaría la recuperación de la autoridad de nombre desde una tabla compartida y su incorporación en la secuencia deseada para el encabezamiento de materia.
- El Servicio de Documentación mantiene en la actualidad dos tablas compartidas de nombres de persona que parten de un momento previo al nuevo plan de normalización de autoridades, y que serán objeto de la primera fase de unificación de los nombres de persona en el museo.
- A pesar de que la utilización de Albalá en el Archivo está permitiendo un acercamiento a las normas ISAD (G) e ISAAR (CPF), la existencia de un volumen importante de autoridades procedentes de la migración de datos del antiguo programa conlleva la existencia de algunos descriptores compuestos que incluyen la mención de la actividad o cargo junto con el nombre de la persona.

Una vez examinada la gestión del control de autoridades que llevan a cabo los Servicios de Biblioteca, Documentación y Archivo, es posible plantear un modelo inicial de convergencia que responde a estas características:

- Realización de un mapeo con los campos de información de los registros de autoridad de nombre de los tres departamentos, gracias al cual estamos en disposición de entender los puntos de encuentro y las diferencias entre las tres bases de datos a la hora de tratar las autoridades en una tabla común.
- Los modelos comparados en el mapeo son la norma ISAAR (CPF) de Archivo, el formato MARC21 de autoridades de Biblioteca y las dos tablas del Servicio de Documentación para el control de los nombres de persona.
- El formato MARC21 de autoridades se ha completado con los campos añadidos en los últimos años para adaptarse al nuevo código de catalogación RDA (Lupe, 2012), gracias al trabajo realizado desde 2008 por el RDA/MARC Working Group. El hecho de que RDA esté pensado íntegramente para el entorno digital y la conexión de los registros bibliográficos y de autoridad con los modelos conceptuales FRBR y FRAD en la web semántica, lo convierten en objeto de estudio obligado dentro de nuestro proyecto, si bien por el momento no está prevista una adaptación al código RDA por parte del Servicio de Biblioteca.

La organización del mapeo se inicia con los campos de la norma ISAAR (CPF) por la claridad que presenta mostrando las áreas de Identificación, Descripción, Relación y Control que requiere un registro de autoridad. A continuación, se establecen las equivalencias con el formato MARC21 de autoridades, con la indicación diferenciada de los nuevos campos motivados por la adaptación a RDA. Por último, se disponen las tablas compartidas del Servicio de Documentación: la Tabla de Autores, desde la que se gestionan los nombres de los artistas; y la Tabla de Personas Físicas, donde se localiza el resto de nombres de persona que han sido incluidos en el sistema para su asignación como persona representada o fuente de ingreso.

Análisis de los resultados

El mapeo de los registros de autoridad de los tres servicios de información del Museo del Prado demuestra que existen unas necesidades similares en la gestión de las autoridades de nombre en la Biblioteca, el Servicio de Documentación y el Archivo. A continuación, destacamos los principales puntos de convergencia a tener en cuenta en la tabla común de autoridades:

Área de Identificación

Archivo: ISAAR (CPF)	Biblioteca: MARC21	Documentación: Autores	Documentación: Personas Físicas
Forma(s) autorizada(s) del nombre	100 Encabezamiento - Nombre personal	Nombre de autor	Nombre
Formas paralelas del nombre	400 Referencia de envío "véase" - Nombre personal	Variantes lingüísticas introducidas en los alias	
Otras formas del nombre	400 Referencia de envío "véase" - Nombre personal	Alias	Otros nombres

Figura 4. Mapeo de autoridades. Área de Identificación.

- En primer lugar, es necesario unificar la forma autorizada del descriptor de persona del Archivo que contempla ISAAR (CPF) con el campo 100 de MARC21 para el encabezamiento autorizado, y la forma principal del nombre en las tablas del Servicio de Documentación.
- A continuación, se podrán indicar las formas variantes del nombre en el Archivo y los denominados «alias» del Servicio de Documentación coincidiendo con el campo 400 de MARC21 de referencia de envío «véase».

Área de Descripción

Archivo: ISAAR (CPF)	Biblioteca: MARC21	Documentación: Autores	Documentación: Personas Físicas
Fechas de existencia (recomendable ISO 8601)	100 Encabezamiento - \$d Fechas asociadas con el nombre RDA: 046 para fechas codificadas - \$f nacimiento / \$g fallecimiento - ISO 8601	Año de nacimiento y muerte / Siglo	Fecha de nacimiento y defunción
Historia	678 Datos biográficos y/o Históricos	Biografía	Biografía
Lugar(es)	RDA: 370 Lugar Asociado - \$a - nacimiento - \$b - fallecimiento	Lugar de nacimiento y muerte	Lugar de nacimiento y defunción
Funciones, ocupaciones y actividades	RDA: 372 Actividad / 374 Ocupación	Tipo de autor y autoría	Cargos, Oficios y Profesiones / Carácter
Estructura(s) interna(s)/Genealogía (Instituciones y familias)	RDA: 376 Información de la familia (Títulos hereditarios asociados a familias)		Títulos (Realeza, nobiliarios y honoríficos)
Contexto general		Escuela	

Figura 5. Mapeo de autoridades. Área de Descripción.

- Los tres departamentos comparten la necesidad de indicar unas fechas conocidas o aproximadas de nacimiento y muerte de la persona. Es conveniente igualar la utilización de estos campos en las tablas del Servicio de Documentación hacia el modelo normalizado de ISAAR (CPF) y MARC21/RDA empleando la Norma ISO 8601 de Representación de fechas y horas, que especifica un formato AAAA-MM-DD.
- Los lugares de nacimiento y muerte, contemplados en los Servicios de Documentación y Archivo, pueden adaptarse en el formato MARC21/RDA con el nuevo campo 370 Lugar Asociado y los indicadores $-\$a-$ para el lugar de nacimiento y $-\$b-$ para el lugar de fallecimiento.
- Una información fundamental para el Servicio de Documentación es la biografía, que se corresponde con el campo Historia de ISAAR (CPF) y 678 Datos biográficos o históricos de MARC21, respectivamente.
- La designación del Tipo de Autor y Tipo de Autoría en el Servicio de Documentación está en clara correspondencia con las Funciones, ocupaciones y actividades en los descriptores de persona del Archivo y los nuevos campos MARC21/RDA 372 Actividad y 374 Ocupación.
- La asignación de Títulos de realeza, nobiliarios y honoríficos en la tabla de Personas Físicas del Servicio de Documentación tiene un planteamiento diferente a los modelos de Biblioteca y Archivo. En el Servicio de Documentación se trabaja con descriptores asociados al nombre de persona, mientras que ISAAR (CPF) y el nuevo código de catalogación RDA hacen un planteamiento basado en la entidad Familia tal y como aparece recogida en los Requisitos Funcionales de los Datos de Autoridad FRAD, que se suma a las dos entidades ya presentes en FRBR $-Persona$ y $Entidad$ corporativa-. La Genealogía de ISAAR (CPF) permite explicar un árbol genealógico con las relaciones existentes entre sus miembros, mientras que el campo 376 Información de la Familia en MARC21/RDA incluye datos del tipo de familia, el miembro prominente de la familia y el título hereditario.

Área de Relación

Archivo: ISAAR (CPF)	Biblioteca: MARC21	Documentación: Autores	Documentación: Personas Físicas
Nombre(s)/Identificadores de instituciones, personas o familias relacionadas			Personajes relacionados / Obras relacionadas
Naturaleza de la relación			Tipo de relación

Figura 6. Mapeo de autoridades. Área de Relación.

- Los campos para relacionar autoridades de nombre tienen un buen desarrollo en la norma ISAAR (CPF) y suponen una vía importante para enriquecer en el futuro nuestra tabla común de autoridades con las relaciones asociativas propias de un tesoro.
- En un primer momento, se identifica una clara convergencia de los campos de ISAAR (CPF) para indicar las autoridades relacionadas y el tipo de relación existente $-jerárquica$, $temporal$, $familiar$ o $asociativa-$ con los campos del Servicio de Documentación para asignar a los nombres otros personajes y obras relacionadas junto con el tipo de relación establecida, en este caso maestro, discípulo, familiar, etc.

Área de Control

Archivo: ISAAR (CPF)	Biblioteca: MARC21	Documentación: Autores	Documentación: Personas Físicas
Identificador del registro de autoridad	001 Número de control de registro	Código informático de identificación	Código informático de identificación
Identificador(es) de la institución	008 Elementos de longitud fija - 39 - Fuente de la catalogación	Se indica en texto libre	
Reglas y/o convenciones	008 Elementos de longitud fija - 10 - Reglas de la catalogación descriptiva		
Estado de elaboración	008 Elementos de longitud fija - 33 - Nivel de formulación		
Fechas de creación, revisión o eliminación	008 Elementos de longitud fija - 00-05 - Fecha de ingreso del registro - 38 - Registro modificado		
Fuentes	670 Fuente en la que se localizaron los datos - RDA: 368 Otros Atributos de Persona o Entidad Corporativa- \$v - Fuente de información - \$u - URI	Se indica en texto libre	

Figura 7. Mapeo de autoridades. Área de Control.

- Es necesario estudiar la unificación del identificador de persona del Archivo con el número de control de registro del campo 001 de MARC21 y el código informático que se genera en la base de datos de Documentación para los literales de los nombres de persona.
- Para avanzar en materia de normalización, conviene también adaptar el modelo de datos del Servicio de Documentación al planteamiento de ISAAR (CPF) y MARC21 con los campos de identificación de la institución que crea la autoridad (008-39), las reglas de catalogación empleadas (008-10) y otros aspectos relevantes, como el estado del registro de autoridad y sus fechas de creación y modificación.
- El último punto imprescindible para la tabla común de autoridades es la indicación de la fuente en la que se localizaron los datos, disponible tanto en ISAAR (CPF) como en MARC21 en su campo 670, y que hasta ahora se añade como texto libre por el Servicio de Documentación. Para adaptarnos a un entorno de web semántica, es recomendable tener presente el campo MARC21/RDA 368 y los subcampos –\$v– Fuente de información y, especialmente, –\$u– URI, que permite la vinculación a través de una URL con fuentes accesibles electrónicamente.

Interrelación de contenidos: la web semántica del Museo del Prado

Gracias al trabajo de la institución para adaptar sus estructuras de datos a las necesidades de información de la sociedad actual, la nueva página web del Museo del Prado se presentó al público el 10 de diciembre de 2015, y supuso la semantización de la base de datos de obras de

arte que gestiona el Servicio de Documentación, así como su relación con recursos multimedia procedentes del anterior canal de vídeos PradoMedia y las voces de la Enciclopedia *online* de la Fundación Amigos del Museo del Prado. En enero de 2016 la web ha alcanzado las 10 500 obras disponibles para su consulta en español, mientras que ha superado las 2000 obras en su versión en inglés, como resultado de un reciente proyecto de traducción tanto de las tablas de vocabularios controlados como de los amplios textos que explican las obras.

AUTOR	+
TEMA	+
ÉPOCA	+
ESCUELA	+
SOPORTE	+
TÉCNICA	+
TIPO DE OBRA	+
Nº DE CATÁLOGO	+
UBICACIÓN	+
Iconografía	
OBJETO	+
FAUNA	+

Pintura x Retrato Real x Armas x Elementos Heráldicos y Nobiliarios

Felipe II, con la armadura de San Quintín
Óleo sobre lienzo. Primera mitad del siglo XVII

El emperador Carlos V
Óleo sobre lienzo. 1605
PANTOJA DE LA CRUZ, JUAN

Figura 8. Web del Museo del Prado. Búsqueda combinada por Tipo de obra, Tema y Objeto representado.

Esta nueva web representa la información sobre las colecciones de acuerdo con el modelo conceptual CIDOC-CRM, con una estructura XML y tripletes RDF, que configuran un grafo global capaz de mostrar las relaciones entre las obras de arte, sus autores y todos los campos de información que se han considerado relevantes para su visibilidad en la web. La consulta se lleva a cabo mediante un buscador facetado que organiza las colecciones atendiendo a los conjuntos de autor, tema, época, escuela, materia, soporte, técnica, tipo de obra, número de catálogo, ubicación y una serie de descriptores iconográficos –objeto, fauna, flora, ubicación geográfica y personaje representado–, todos los cuales responden a las tablas de vocabularios controlados gestionadas desde el Servicio de Documentación. De esta forma el usuario puede navegar por las colecciones en función de sus intereses, combinando múltiples criterios de recuperación de la información e interactuando con un sistema que ofrece contenidos relacionados gracias a una descripción detallada de los recursos.



Figura 9. Web del Museo del Prado. Propuesta de obras de un mismo artista. Tiziano.

Entre los criterios más destacados para relacionar contenidos –además de la agrupación de resultados en función de las facetas seleccionadas por el usuario–, encontramos la propuesta de otras obras del mismo autor; la relación de artistas coetáneos; las obras pertenecientes a una misma serie o colección, así como los contenidos audiovisuales que tratan aspectos relativos al autor, obra o época de la ficha en que se encuentre el usuario en cada caso.



Figura 10. Web del Museo del Prado. Propuesta de contenidos multimedia relacionados con Tiziano.

El modelo CIDOC-CRM, empleado para representar la base de datos de obras de arte, y el resto de contenidos en la web del Museo del Prado, ha sido diseñado por el Comité para la Documentación del Consejo Internacional de Museos (ICOM) con la finalidad de expresar de forma comprensible por las máquinas el conocimiento relativo al patrimonio cultural que conservan los museos. Su objetivo principal es permitir el acercamiento entre muy diferentes conjuntos de datos, para lograr armonizar sus descriptores y estructuras de catalogación dentro de un dominio de conocimiento global sobre el patrimonio cultural, sin perder el carácter específico y la personalidad propia de cada tipo de colecciones. La elaboración del modelo se inició en el año 1994 –con una primera publicación en 1999–; en el año 2006 ha sido publicado como norma ISO 21127:2006, y revisado en la norma ISO 21127:2014.

Dentro del ámbito internacional, diversas instituciones han adaptado sus bases de datos a este modelo, como es el caso de la versión semántica del British Museum, lo que abre nuevas vías de interoperabilidad entre el Museo del Prado y otros museos nacionales e internacionales. Entendiendo la interoperabilidad desde una perspectiva más amplia, otro aspecto fundamental de este modelo es que está siendo el centro de una política de acercamiento entre los modelos conceptuales de museos, bibliotecas y archivos a nivel internacional, ya que desde el año 2003 se ha empezado a trabajar en la armonización entre CIDOC-CRM y FRBR_{OO}, a lo que se añade desde 2010 los primeros pasos en el diseño de un modelo conceptual para incluir los datos de archivo en esta ontología global del patrimonio cultural.

Gracias a la representación de sus datos mediante CIDOC-CRM, el Museo del Prado podrá explorar la posibilidad de incluir otros contenidos de la Biblioteca y el Archivo y relacionarlos con las fichas de catalogación de las obras de arte, para lo cual se hace imprescindible avanzar en proyectos de normalización a gran escala, como el que estamos desarrollando para las autoridades de nombre.

Conclusión

En los últimos años las instituciones museísticas están teniendo que hacer frente a nuevos retos en materia de gestión de la información para adaptarse con éxito a un entorno de web semántica. Este hecho viene motivado por la transformación de la relación establecida entre la sociedad y el museo, que ya no se reduce a la visita presencial a sus salas, sino que abarca toda una serie de peticiones de información sobre las colecciones, la historia y el funcionamiento de la institución. Para dar una respuesta satisfactoria a esta nueva comprensión del museo como un centro de conocimiento, el Museo del Prado ha optado de forma decidida por potenciar el papel de sus servicios de información así como su presencia en la página web.

En este contexto, el Área de Biblioteca, Documentación y Archivo está ampliando progresivamente su concepción de la normalización documental hacia una visión que agrupe los criterios seguidos en los tres departamentos. De esta forma, un proyecto como el de la unificación de las autoridades de nombre permitirá, en primer lugar, facilitar el trabajo y la entrega de materiales a nivel interno, así como abrir vías en el futuro hacia el establecimiento de nuevas relaciones semánticas que enriquezcan la consulta de las colecciones.

Bibliografía

- DOCAMPO CAPILLA, F. J., y MARTÍN BRAVO, A. (2009): «El Área de Biblioteca, Archivo y Documentación del Museo Nacional del Prado: hacia una integración de procesos y servicios», *XI Jornadas de Gestión de la Información: Servicios polivalentes, confluencia entre profesionales de archivo, biblioteca y documentación*, 19-20 de noviembre. Madrid: Biblioteca Nacional, pp. 53-66. <<http://eprints.rclis.org/13878/1/XIJGI-Docampo.pdf>> [Consulta: 15 de enero de 2016].
- ICOM.CIDOC (2015): *Definition of the CIDOC Conceptual Reference Model. Version 6.1*. CIDOC CRM Special Interest Group. <http://www.cidoc-crm.org/docs/cidoc_crm_version_6.1.pdf>. [Consulta: 15 de enero de 2016].
- IFLA-ICOM.CIDOC (2015): *FRBR^{OO}. Object-oriented definition and mapping from FRBR^{ER}, FRAD and FRISAD (version 2.1)*. International Working Group on FRBR and CIDOC CRM Harmonisation. <http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V2.1_2015February.pdf>. [Consulta: 20 de enero de 2016].
- INTERNATIONAL COUNCIL ON ARCHIVES (2004): *ISAAR (CPF). Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias*. Ministerio de Cultura. Traducción de *ISAAR (CPF). International Standard Archival Authority Records for Corporate Bodies, Persons and Families*. <<http://www.agn.gob.mx/menuprincipal/archivistica/normas/pdf/asaar.pdf>>. [Consulta: 20 de enero de 2016].
- LIBRARY OF CONGRESS (2006): *Formato MARC21 Conciso para Datos de Autoridad*. <<http://www.loc.gov/marc/authority/ecadspa.html>>. [Consulta: 27 de enero de 2016].
- LUPE CRISTÁN, A. (2012): *Elementos de RDA en Registros de Autoridad de Nombre MARC 21*. Library of Congress. Presentación. <www.loc.gov/catdir/cpsd/RDA/temas_especiales_elementos_spa.ppt> [Consulta: 27 de enero de 2016].
- MARTÍN BRAVO, A., y PANTOJA FERRARI, J. (2013): «Integración de servicios documentales: el nuevo website del Museo del Prado», *Segundas Jornadas sobre Bibliotecas de Museos. Estrategias e Innovación*, 6-8 de noviembre. Madrid: Museo Nacional del Prado, pp. 96-103. <<https://sede.educacion.gob.es/publiventa/detalle.action?cod=20083C>> [Consulta: 27 de enero de 2016].

Normalización terminológica en bibliotecas: autoridades de materia y términos de género/forma en la Biblioteca Nacional

Pascual Jiménez Huerta

Jefe del Servicio de Clasificación. Biblioteca Nacional de España

Introducción

La presentación pretende mostrar los trabajos que ha realizado la Biblioteca Nacional de España con relación a la normalización de vocabularios de materia y cómo este control terminológico ha dado como resultado el fichero de los encabezamientos de materia de la Biblioteca Nacional de España (EMBNE) así como la elaboración del fichero de autoridades de materia de dicha entidad. También hace referencia a los trabajos realizados con otros registros de autoridad, como los de subencabezamientos y de nombres geográficos. Por último se mencionan los términos de género/forma que han sido muy útiles en la elaboración de distintos tesauros en los distintos departamentos que componen la Biblioteca Nacional. Se muestran los pasos que se deben seguir hasta la elaboración de un único vocabulario de términos de género/forma.

1. El control terminológico. Vocabularios controlados

El lenguaje documental surge a raíz del intento de controlar los documentos para su localización dentro de un sistema de información. Entendemos por lenguaje documental todo sistema artificial de signos normalizados que facilita la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática de la información solicitada por los usuarios. Opera como un lenguaje intermediario entre las informaciones contenidas en los documentos y las solicitadas por los usuarios, controla la posible ambigüedad entre términos y resuelve problemas derivados del multilingüismo, como el caso de los tesauros multilingües o las clasificaciones de tipo numérico.

Una de las principales operaciones que se realizan para extraer la información del documento es la de la **indización**. La indización es una operación intelectual donde se aplican técnicas normalizadas a un conjunto de documentos para hacerlos controlables y utilizables, es decir,

para permitir su recuperación. En esta fase se realiza una operación de tipo analítica para representar los conceptos clave, que implica transformar la información para su almacenamiento.

Indización por materias

Consiste en la correlación sucesiva de términos llamados encabezamientos, que expresan el tema de un documento. Es, por tanto, un lenguaje precoordinado, de estructura combinatoria, donde las materias pueden expresarse bien en encabezamientos simples o compuestos (encabezamientos más subencabezamientos). Es verdad que en ocasiones es un sistema poco flexible a la hora de representar conceptos muy específicos; sin embargo, sus posibilidades de mostrar todas las características de una materia son innegables. El ejemplo más característico de este tipo de indización son las listas de encabezamientos de materia, que son ampliamente utilizadas en la mayoría de las bibliotecas.

2. EMBNE. Encabezamientos de materia

Los encabezamientos de materia de la Biblioteca Nacional de España son el fruto de una labor ardua y constante de normalización terminológica para asignar el contenido de los documentos. Estos encabezamientos son el fruto de una normalización exhaustiva en línea con las principales bibliotecas del mundo y que determina la elección de las materias asignadas a los documentos. Además, el Servicio de Clasificación ha elaborado un Manual de indización donde se recogen los acuerdos y directrices llevadas a cabo por los clasificadores en su trabajo de asignación de encabezamientos. Estas directrices son fundamentales para tener la mayor uniformidad posible a la hora de utilizar los encabezamientos siempre de la misma manera. El Manual está disponible en la página web de la Biblioteca Nacional en la dirección: <<http://www.bne.es/es/Micrositios/Publicaciones/MEMBNE/>>.

3. Registros de autoridad de materia

Los registros de autoridad reflejan la investigación llevada a cabo por la Biblioteca Nacional para establecer de forma normalizada el encabezamiento utilizado en los registros bibliográficos como puntos de acceso de:

- **Títulos** para identificación de la obra o expresión.
- **Personas y entidades corporativas** en función de la responsabilidad intelectual o artística de la obra o expresión; responsabilidad en la producción física y distribución de la manifestación; y responsabilidad en la custodia del ejemplar.
- **Puntos de acceso en función del tema o materia** tratado en esa obra.

Objetivos del control de autoridades

Entre los objetivos que persigue el control de autoridades está el de dar consistencia e integridad a la base de datos, conseguir mejorar la precisión en la recuperación de la información, ayudar en el trabajo de referencia de todo tipo de biblioteca y facilitar el intercambio de información bibliográfica, y con ello aligerar los costes de la catalogación.

Los registros de autoridad dan uniformidad a los datos de toda la base de datos de un catálogo y permite una mayor precisión en las búsquedas de información. Por ello, y más concretamente en el tema de las materias, es indispensable contar con un vocabulario lo más estructurado posible.

Registros de autoridad de materia

Actualmente el volumen de registros de autoridad de materia de la biblioteca cuenta con los siguientes datos:

– Número de registros de autoridad de materia validados	25 000
– Registros de autoridad de subencabezamientos	450
– Registros de autoridad de nombre geográfico	10 000
– Registros de autoridad de género/forma	300

Estos datos se refieren a registros de autoridad de materia completamente estudiados y validados, con todas sus referencias y fuentes de información. Existe, por tanto, una multitud de registros provisionales que son perfectamente válidos pero que todavía no han sido validados por los responsables de autoridades de la biblioteca.

Cada registro de autoridad de materia está conformado, como es lógico, con sus correspondientes términos alternativos, términos sinónimos o aquellos términos por los que un usuario podría buscar esa materia.

También se compone de relaciones con otros términos: Términos genéricos o más generales, términos específicos o más pequeños y términos relacionados con el primero que conforman una estructura de tesoro dentro del vocabulario, que permite navegar entre las diferentes materias para hacer búsquedas mucho más exhaustivas.

Otro de los elementos importantes dentro del registro de autoridad son los campos de notas, que permiten definir una materia ambigua, o diferenciarla de otras materias que son afines y que podrían confundir tanto al indizador como al usuario a la hora de buscar la información que necesita.

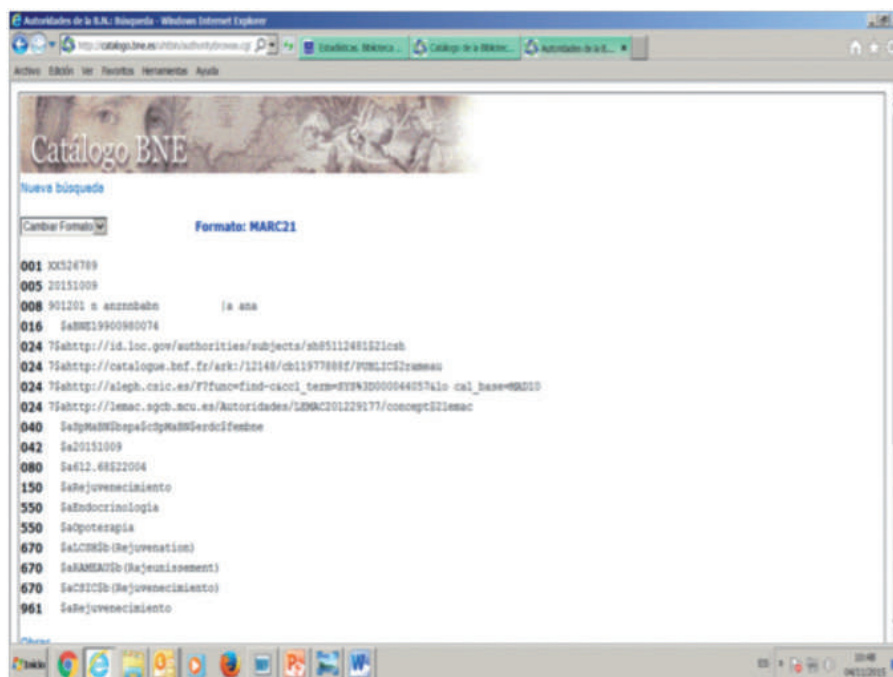
Por último, las fuentes de referencia dan consistencia al catálogo utilizando fuentes que nos parezcan rigurosas a la hora de establecer las materias. Usando las mismas fuentes, tanto de carácter general como específico, tendremos también un alineamiento más preciso con otras bibliotecas o centros de documentación.

Valor añadido

Hasta ahora la información que he descrito no difiere en gran medida de los registros de autoridad de materia de cualquier catálogo de bibliotecas. Lo que yo he llamado «valor añadido» es el resultado de un *enrichment* de información en los registros que proporciona un plus de posibilidades a la hora de obtener información sobre la materia buscada.

Campo 024: Enlaces a las URI LCSH, RAMEAU, CSIC y LEMAC

En el campo 024 del formato MARC21 se han añadido las URI de las materias equivalentes de los catálogos de la Library of Congress, la Biblioteca Nacional de Francia, el Consejo Superior de Investigaciones Científicas y los encabezamientos de materia de la Biblioteca de Cataluña. Estos enlaces nos envían a las autoridades de materia de estos catálogos y posibilitarán la búsqueda de información de la misma materia que hemos buscado nosotros en sus respectivos catálogos. Las posibilidades de la tecnología de Linked Data permiten enlazar distintos recursos de información con una única búsqueda desde nuestro catálogo.



CDU: Todas las autoridades de materia tienen incorporada en el campo 080 la notación de CDU que le corresponde, de su edición de 2004. Para el año 2016 se va a migrar a la última edición publicada de 2015, con lo que se actualizarán las notaciones de CDU de los registros de autoridad de materia y de nombre geográfico. En esta nueva edición la Biblioteca Nacional ha colaborado en la revisión de términos anteriores y en la traducción de términos nuevos.

Además, existe un Convenio de colaboración entre AENOR y BNE para incorporar los encabezamientos de materia de la Biblioteca Nacional con las notaciones de la nueva CDU On Line en español que va a salir publicada próximamente. En esta plataforma estarán todas las notaciones de la CDU original traducidas al castellano y con enlaces a las materias de la Biblioteca Nacional, por lo que la herramienta será muy útil para los clasificadores e indizadores que consulten dicha herramienta.

Transformación a SKOS: La BNE ha transformado a SKOS todas sus autoridades de materia, de forma que sean legibles por máquina y se puedan utilizar para diferentes aplicaciones informáticas. La transformación a SKOS permite traspasar las barreras de los catálogos cerrados y poder utilizar los datos en distintos ámbitos. El vocabulario de EMBNE está también subido a DataHub y se puede consultar desde la plataforma de Datos.bne que ofrece la biblioteca.

- Podemos ver un registro cualquiera de materia integrado en datos.bne.es en: <<http://datos.bne.es/resource/XX239890>>.
- Ver un registro geográfico: <<http://datos.bne.es/resource/XX450834>>.
- Ver un registro de Género-forma: <<http://datos.bne.es/resource/XX4843208>>.
- Ver un subencabezamiento de materia: <<http://datos.bne.es/resource/XX4990059>>.

4. Registros de subencabezamientos

Independientemente de los registros de autoridad de materia tanto simples como compuestos, la BNE también está creando registros de autoridad de subencabezamientos. Este tipo de registros de autoridad se nos antojan muy importantes, ya que permiten determinar la utilización de estos términos en los encabezamientos de materia principales. Posibilitan un control de uso de dichos subencabezamientos y unifican criterios de actuación a la hora de clasificar los documentos.

Existen, por tanto, registros de autoridad de subencabezamientos Generales, Cronológicos y de Forma. Como en el caso de las materias, los registros de autoridad de subencabezamiento tienen también términos alternativos; genéricos, específicos y relacionados. Fundamentales son las notas de uso que delimitan su aplicación y las fuentes consultadas para la realización de los registros de autoridad.

Al igual que las materias, también tienen sus enlaces en el campo 024 a LCSH, RAMEAU, etc., y están también transformados a SKOS.



5. Registros de autoridad de nombre geográfico

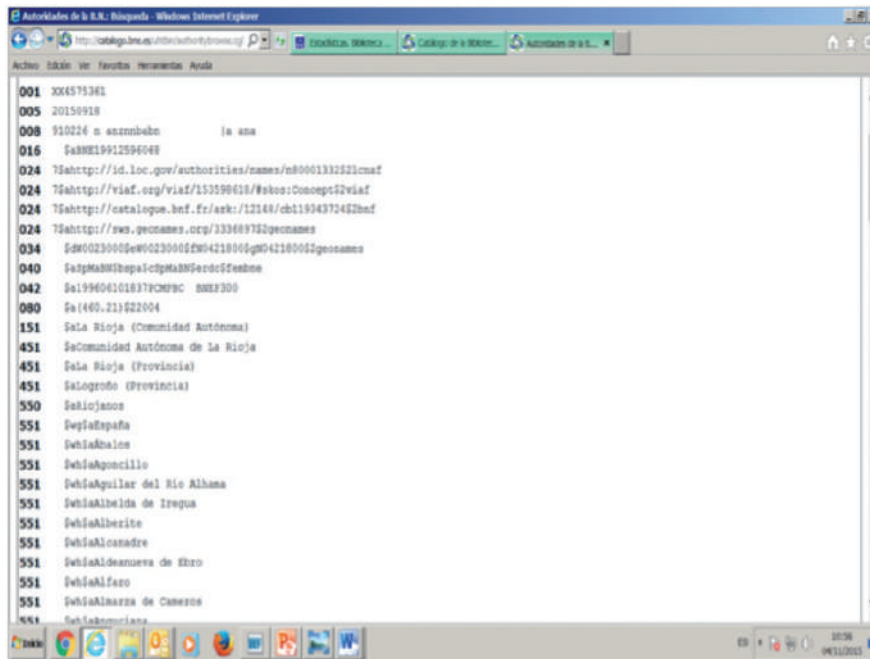
¿Elche o Elx? ¿Xátiva o Játiva? ¿Fuenterrabía u Hondarribia? ¿Lérida o Lleida?

En este tipo de registros ha habido una enorme labor de normalización. En primer lugar y dado el carácter plurilingüístico español, la elección del punto de acceso principal de los topónimos españoles requería de unas directrices claras a la hora de utilizar este tipo de encabezamientos. Al final, y tras muchas discusiones, se eligió como fuente normalizadora el Nomenclátor Geográfico Nacional, del Ministerio de Fomento, que es el organismo responsable de la toponimia en España.

Además de la elección del punto de acceso, estos registros se enriquecieron con enlaces a LCSH, RAMEAU, VIAF y GeoNames, que es la base de geográficos más grande y más utilizada en el mundo.

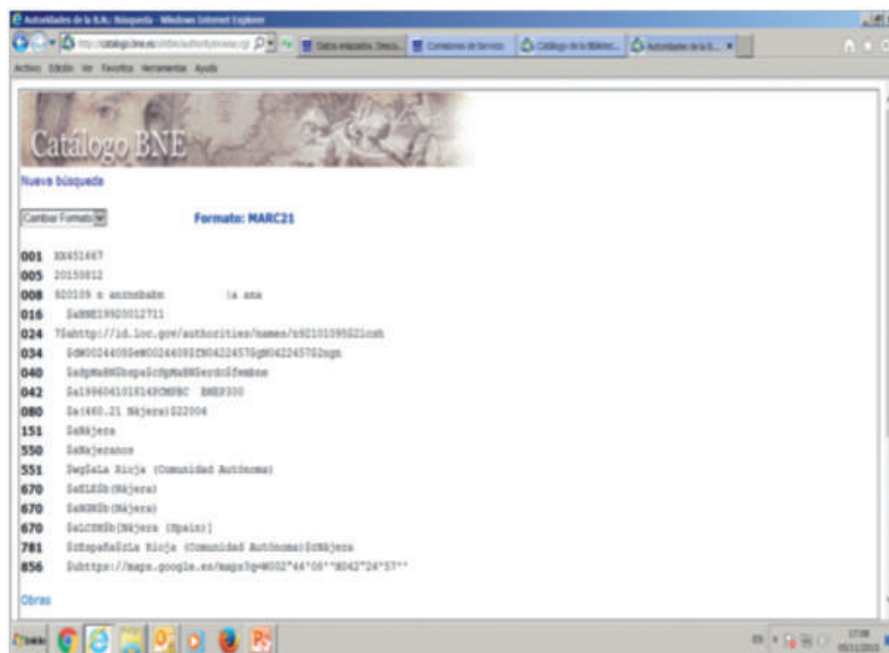
Otro de los puntos en los que se trabajó fue en la inclusión de las coordenadas geográficas en los topónimos. Las coordenadas geográficas no solo sirven para deshacer homonimias entre lugares con el mismo nombre, sino que además abre una puerta a distintos proyectos de georreferenciación. Las posibilidades de manejar información geográfica con las coordenadas son enormes y prueba de ello es la conexión que hemos realizado con Google Maps a la que me referiré más adelante.

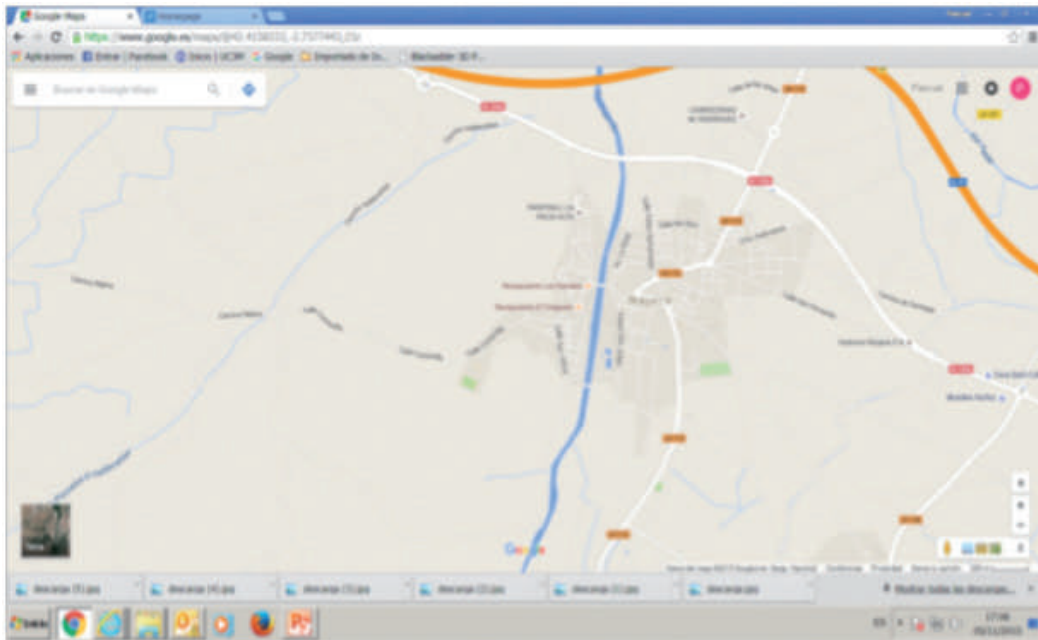
Por supuesto, los registros cuentan con su CDU actualizada y mantienen las relaciones genérico-específicas como las materias. En este sentido se ha hecho especial hincapié en dibujar la organización territorial del estado en este tipo de relaciones, es decir, que España tiene como términos específicos las diferentes comunidades autónomas, que estas a su vez tienen como específicos las provincias, de las que a su vez cuelgan los municipios y como último punto las entidades locales menores, de manera que se puede navegar de una entidad territorial a la superior y viceversa.



Un añadido más es que en cada topónimo se incluye en el campo 781 la estructura de País-CC. AA.-Provincia-Municipio y Entidad local menor que le corresponde.

Por último, y para terminar uno de los proyectos más interesantes, es la conexión con Google Maps, en el que se aprovechó la información de las coordenadas para establecer un link con Google Maps que muestra el mapa del topónimo en cuestión. Esto abre un abanico de posibilidades de trabajos de georreferenciación en los que las autoridades de nombre geográfico jugarían un papel intermediador entre los recursos y la visualización del topónimo en un mapa.





6. Registros de género/forma

¿Qué es un encabezamiento de género/forma?

Los encabezamientos de género/forma describen lo que el ejemplar **es**, no de lo que **trata**. La **forma** se define como una característica de las obras con un formato o propósito concreto y el **género** se refiere a categorías de obras que se caracterizan por tener argumentos, temas, escenarios, situaciones o caracteres similares.

Los términos de género/forma sirven para describir cómo es una publicación, no su contenido temático.

El objetivo de los encabezamientos de género/forma es describir la expresión, no la manifestación de la obra, es decir, los encabezamientos se refieren a la expresión intelectual o artística de la obra, no al soporte físico.

Aunque no suponen ninguna novedad, pues ya desde los años setenta se han venido utilizando en muchas bibliotecas, sobre todo en aquellas con amplias colecciones de libros antiguos o raros, materiales gráficos, audiovisuales, etc., su uso no estaba muy extendido fuera de dichos ámbitos. Es ahora cuando parece que muchas bibliotecas se han sumado a la utilización de estos términos.

Hasta ahora pocas bibliotecas se habían embarcado en la creación de sus propios registros de autoridad de términos de género/forma. Normalmente, aquellas que habían decidido utilizar este método de indización para determinados tipos de materiales especiales se habían decantado por el uso de tesauros específicos. Ahora las grandes bibliotecas han retomado el interés de utilizar este tipo de puntos de acceso como una herramienta importante en la recuperación de información que demandan los usuarios.



Se debería asignar el encabezamiento de materia «Películas de terror» a libros que **tratan** sobre películas de terror. Un catalogador que asigne un encabezamiento a la película «Frankenstein» también utilizará un encabezamiento llamado «Películas de terror» pero como encabezamiento de género/forma, ya que la película **es** una película de terror.

¿Qué **beneficios** supone la utilización del género/forma?

- Nuevo punto de acceso.
- Reducción de utilización de tesauros.
- Mayor control de vocabulario.
- Actualización de la terminología.

Beneficios para los usuarios

- Mejoras en la capacidad de búsqueda. Los encabezamientos compuestos a veces inhibían a los usuarios a descubrir las obras que buscaban. En ocasiones la búsqueda de los usuarios se refiere exclusivamente a un género o una forma más que a la materia propiamente dicha.
- Posibilidad de buscar por géneros o formas exclusivamente.

Es un nuevo punto de acceso, por lo que redundará en beneficio de la recuperación. La utilización de términos de género/forma reduce la posibilidad de tener que emplear muchos tesauros como fuentes. Mejora también la cooperación interbibliotecaria. La posibilidad de utilizar un tesoro único favorece el intercambio de información.

Principios para el desarrollo de términos de género/forma:

- Principio de autoridad literaria para asegurar el uso de la terminología adecuada dentro de los encabezamientos de género/forma.
- Sincronizar la terminología, significado y estilo dentro de los encabezamientos de género/forma con los de materia tanto como sea posible.
- Incluir una nota de uso en el encabezamiento de género/forma correspondiente incluyendo instrucciones sobre la aplicación del encabezamiento.
- Guardar una estructura consistente entre encabezamientos y también entre disciplinas.
- Examinar cada disciplina independientemente para determinar si la subdivisión geográfica es pertinente.
- Investigar la necesidad de generar un *top-term* para el tesoro en cada disciplina.

7. Diseño del plan de realización de registros de autoridad de género/forma en la Biblioteca Nacional

Planteamiento inicial

El Proyecto de Adopción de Términos de Género/Forma es un proyecto que intenta incorporar un nuevo punto de acceso a los registros bibliográficos con relación a su manifestación, es decir, en función de lo que los documentos son, en vez de en función de lo que tratan.

El proyecto pretende ofrecer nuevas formas de búsqueda a los usuarios de manera que puedan satisfacer sus demandas informativas a partir de la forma y el género del documento.

Antecedentes

Hasta 2009 en la base de datos de la BNE no existían registros de autoridad de términos de género/forma en su base de datos Unicorn, aunque sí se utilizaba el campo 655 de los registros bibliográficos como campo para describir los términos de género/forma de algunos departamentos. Estos términos no estaban normalizados, aunque se utilizaban diferentes listas y tesauros para justificar su uso. Por supuesto, estos campos aparecían en Unicorn como «Unauthorized», no autorizados, aunque permitían la recuperación por esos términos.

¿Qué teníamos?

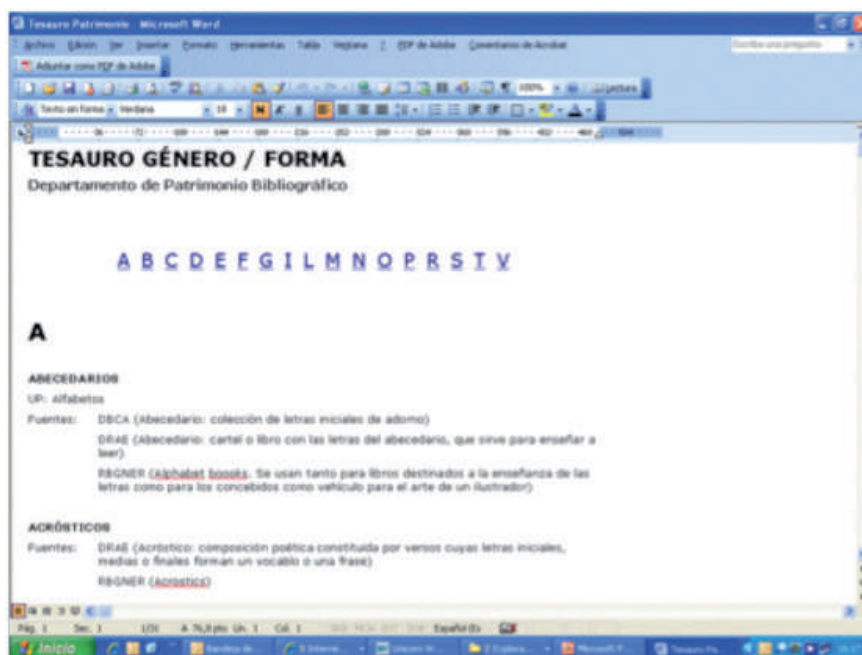
El tesauro más completo lo poseía el Departamento de Patrimonio, donde tenían un tesauro propio de términos en el que aparecían términos relacionados, alternativos y fuentes, pero funcionaba como un tesauro independiente del resto de la biblioteca.

En Dibujos y Grabados elaboraron una lista de términos válidos para utilizar como términos de género/forma.

En Música únicamente utilizaban el término «Archivo de la palabra» como término de género/forma. El resto aparecían como encabezamientos de materia aunque pudieran ser de género/forma.

En Clasificación no se utilizan este tipo de términos, siendo todos ellos de materia.

En Cartografía tampoco utilizaban este tipo de términos.



Parecía claro que, antes de abordar la elaboración de los registros de autoridad de género/forma, había que pulir y estudiar las diferentes listas y tesauros para ver si los términos eran coincidentes y no existía incoherencia entre dichos términos. En otras palabras, habría que cotejar los términos de las diferentes listas. Esto implica también estudiar los alternativos y compararlos también con los encabezamientos de materia.

Otro de los aspectos a tener en cuenta era que había que normalizar las fuentes que íbamos a utilizar en los términos de género forma. Para ello había que elaborar una lista de tesauros y fuentes autorizadas que nos permitiera tener un control de los términos para utilizar como registros de autoridad, prestar atención a las definiciones con el fin de acotar las formas y buscar en otros catálogos y tesauros sus correspondencias.

También había que tener en cuenta las aportaciones de la Library of Congress de los nuevos registros de autoridad que elaboraran y nos sirvieran de referencia para realizar nuestras autoridades.

Con todo ello, pedimos a los distintos departamentos que nos mandaran las fuentes que utilizaban para extraer sus términos de género/forma y que utilizaban en sus listas. Fue una labor de recopilación de fuentes de distintas áreas temáticas que nos permitió tener un control exhaustivo del vocabulario utilizado.

Con todas las fuentes disponibles, se estudiaron, cotejaron y se elaboró una lista única de fuentes para la elaboración de términos de género/forma. Era una buena base de partida para la creación de los registros de autoridad.

Referencias

- **Encabezamientos alternativos:** Se consignan en el Área de Cita de Referencia de «Véase» y se refieren a aquellos términos que sirven para expresar el mismo género o la misma forma. (Estos términos se consignan en el campo 455 del formato MARC). Se recomienda que los encabezamientos alternativos de género/forma sean coincidentes con los de materia.
- **Encabezamientos asociados:** El Área de Referencia de «Véase además» recoge los términos de género/forma relacionados con el encabezamiento principal. (Se consignan en el campo 555 del formato MARC). Pueden existir relaciones jerárquicas (tanto de tipo genérico como específico) y asociativas.
- **Notas de alcance:** Las notas de alcance se utilizan para especificar el alcance del encabezamiento, o bien definir el término cuando el significado sea poco conocido o ambiguo. Asimismo se normalizó el texto de la nota de uso para aquellos términos de género-forma que eran coincidentes con los términos de materia. Se procedió a separar aquellos registros bibliográficos que tenían asignado un encabezamiento de materia erróneo cuando deberían llevar un encabezamiento de género-forma.
- **Nota de fuentes de datos encontrados:** Con estas notas se indican las obras de referencia que se han consultado en las que aparece el término del encabezamiento de género/forma.

Vocabularios terminados:

Cine (Géneros cinematográficos): Fue el primer vocabulario en elaborarse. Tras un intenso trabajo, se especificaron los siguientes géneros cinematográficos:

- Acción y aventuras
- Animación
- Bélico
- Ciencia ficción
- Comedia cinematográfica

- Criminal
- Documental
- Drama cinematográfico
- Erótico
- Fantástico
- Histórico
- Mudo
- Musical
- Negro
- Oeste (*Western*)
- Pornográfico
- Religioso
- Suspense (*Thriller*)
- Terror

Los beneficios y utilidades de los encabezamientos de género/forma para los géneros cinematográficos son innegables: aportan mayor información al documento, favorecen la recuperación de información y contemplan las demandas informativas sobre géneros concretos.

Proyectos en Género/Forma

- Bellas Artes: Incluye Fotografías, Dibujos, Grabados y Ephemera. Como fuentes se han usado el AAT, TGM, el Tesoro de Isabel Trinidad Lafuente, el *Diccionario de Artes Gráficas* de la Real Academia de Bellas Artes de San Fernando y otros diccionarios de términos artísticos. Se han creado tanto términos de género como de forma, lo que hace que haya términos en singular y plural. Se han admitido subencabezamientos geográficos y cronológicos en los casos en los que se consideraba que aportaba información (aunque ha llevado a un uso generalizado de subencabezamientos).
- Fondo antiguo: Existía ya un tesoro creado por el departamento desde 2008/2009. Se usaba el 655 desde el catálogo con una tabla con unos pocos términos. La fuente principal era RBMS (vocabulario de género/forma de la Rare books and manuscript section de la ALA <<http://www.rbms.info/vocabularies/>>). Ahora se van actualizando fuentes y también la RBMS está revisando términos.
- Cartografía: Se utilizan como fuentes el AAT y TGM, pero sobre todo el *Diccionario de términos geográficos* del Instituto Cartográfico de Cataluña. Se han eliminado los subencabezamientos de forma de los distintos tipos de mapas, sustituyéndolos por los términos de género/forma.
- Literatura: Se usan como fuentes el LCGFT cuando es posible y el AAT. También libros españoles especializados en géneros literarios y teoría de la literatura, diccionarios de términos literarios y el de bibliología de Martínez de Sousa. No se ha implementado en los bibliográficos todavía, ya que se ha acabado este grupo en agosto. Se espera comenzar a utilizarlo en 2016.
- Música: Hemos empezado a trabajar con música, comparando nuestro vocabulario con el de la LC, pero acabamos de comenzar.
- Próximos proyectos: Publicaciones seriadas.

Con esto estarían cubiertos todos los vocabularios de la Biblioteca. Los términos de género/forma al igual que las materias y geográficos llevan su conexión de Linked data en el campo 024, incluyendo las URI de LCGFT y AAT. También este vocabulario está transformado a SKOS.

8. La cooperación en bibliotecas. Proyectos en funcionamiento

Ante la gran demanda informativa de los usuarios y la cantidad ingente de documentos existentes tanto físicos como electrónicos, cada vez se hace más patente la importancia de la cooperación. No tiene sentido catalogar muchas veces el mismo documento cuando ya lo han hecho otros. Por ello, hay que tratar de sumar esfuerzos y, para cooperar, es necesario desarrollar las tecnologías derivadas de datos enlazados que permitan orientar al usuario hacia donde está el recurso que necesita. En este sentido la plataforma de datos.bne intenta dar el salto de los catálogos cerrados y ofrecer la información del catálogo de la BNE más una variedad de recursos en otras bases de datos y recursos de información que posibilitan al usuario satisfacer sus necesidades informativas de la manera más adecuada. Datos.bne se encuentra todavía en una fase beta, pero esperamos que pronto esté operativo al cien por cien.



PARES hacia la web semántica: autoridades y tesauros en los archivos estatales

Aránzazu Lafuente Urién

Subdirección General de los Archivos Estatales. MECD

No cabe duda de que el desarrollo de las nuevas tecnologías en el entorno web ha modificado nuestro trabajo diario. En los archivos se ha pasado de trabajar de manera «aislada» atendiendo en persona a investigadores y usuarios, y estos consultando los originales en las salas de consulta, a encontrarnos inmersos en un entorno en continuo cambio, en ocasiones difícil de asimilar, con nuevos entornos y lenguajes de intercambio, plataformas colaborativas y gestores documentales, la consulta en línea desde entornos remotos y cada vez una mayor exigencia de inmediatez en la respuesta en unas organizaciones que avanzan hacia la e-administración sin papel. Sin embargo, estos cambios son a su vez un gran reto, nuestra gran oportunidad para que el ingente trabajo que se ha venido haciendo en nuestros archivos desde hace siglos, de manera lenta, constante, pero oculta, pueda ponerse en valor y darle la mayor visibilidad. Visibilidad e inmediatez. En este sentido, en esta ponencia se tratará de exponer brevemente los desarrollos que en los últimos años estamos desarrollando en el Portal de Archivos Españoles hacia la web semántica en torno a dos grandes ejes: las autoridades archivísticas y los vocabularios controlados, ambos imprescindibles para identificar en contexto e interoperar en la red en un entorno multilingüe, donde la reutilización, la transparencia o los datos enlazados no son posibles sin datos de calidad.

Frente al carácter voluntario del coleccionista en la formación de las colecciones de los museos y bibliotecas, los documentos que conforman un fondo¹ se acumulan de forma orgánica como reflejo escrito de las actividades de una institución –tanto da si es pública o privada– a lo largo de su vida y de acuerdo con el ejercicio de unas funciones y competencias. Los documentos creados, reunidos o acumulados o mantenidos para o por una organización o un individuo deben mantenerse unidos y conservados con el orden original en que se han creado o han sido mantenidos. Esta afirmación es el principio de procedencia, *provenance* en inglés

¹ Conjunto de documentos y expedientes generados por una institución a lo largo de su vida.

y *respect des fonds* en francés, –el principio fundamental en el que está basado todo trabajo archivístico–. Documentar el contexto de producción de un «archivo» e identificar y proceder a la descripción de los productores –*creators* en inglés– es una de las actividades esenciales de los archiveros o, lo que es lo mismo, supone documentar quién, cómo, cuándo y por qué (o de acuerdo a qué normas o a qué funciones) un fondo de archivo ha sido creado. Con estas premisas se puede comprender la importancia que para el tratamiento archivístico tiene el proceso de identificación y control de los nombres de los productores de fondos de archivo, que incluye el control de las autoridades archivísticas.

El volumen del Patrimonio Documental Español custodiado en los archivos es ingente. El art. 49.2 de la LPHE incluye los fondos archivísticos considerados de especial protección por formar parte del patrimonio documental español: todos los producidos por las administraciones públicas, los de más de cuarenta años de instituciones eclesiásticas, sindicatos y partidos, y los de entidades privadas de más de cien años. Solo en nuestros archivos nacionales o estatales, España posee uno de los patrimonios documentales más valiosos del mundo, siendo fundamental para la memoria de Europa, América y el Pacífico al menos hasta el siglo XVIII. La difusión del contenido de los Archivos Estatales es una de las funciones que tiene encomendada nuestra organización y se realiza a través del Portal de Archivos Españoles PARES. Sin embargo, PARES, un sistema propietario de la Subdirección General de los Archivos Estatales del MECD, es mucho más que un portal de difusión, es un sistema integral para la gestión de los centros y los fondos archivísticos, desarrollado en varios módulos para la gestión de usuarios y las consultas de fondos en línea y presencial, la gestión de sala de investigadores, control de préstamos, estadísticas y el módulo de descripción. PARES es una base de datos en Oracle, común para los siete archivos y los centros dependientes de la SGAE, que incluye además al Censo Guía de Archivos Estatales, instrumento jurídico de control y censado establecido en la LPHE (censo que desde 1990 se ha abierto a Iberoamérica, recogiendo el censo de los archivos de este continente). PARES es un OAI que puede ser abierto en un futuro próximo a archivos de otras administraciones mediante convenios².

En octubre de 2015 se han finalizado los desarrollos del nuevo módulo de autoridades y puntos de acceso, la adaptación al mismo del módulo de descripción, la mejora y el desarrollo de herramientas de revisión. En estos momentos estamos³ cerrando las pruebas y testeos del nuevo buscador del portal. Todos estos nuevos desarrollos obedecen a seis grandes objetivos:

- **Proporcionar datos de calidad** adaptando los campos descriptivos a las nuevas normas internacionales de intercambio, junto con un proceso paralelo de revisión y adaptación de los contenidos y operaciones de gestión.
- **Multilingüismo**, desarrollando vocabularios de puntos de acceso y, en su caso, autoridades –además de traducir etiquetas de elementos y herramientas– en cinco lenguas: inglés y francés, y catalán, gallego y euskera, con el objetivo de hacer los datos interoperables y más accesibles en la web.
- **Interoperabilidad**: como resultado del paso anterior, trabajamos en el desarrollo y mejora de las herramientas y en la conversión de los vocabularios multilingüe y las autoridades en LOD como conceptos **SKOS**.
- Además de cumplir con los requisitos legales de **transparencia y reutilización**,
- **Compartimos datos** tanto con Iberoamérica a través del Censo Guía, como con Europa por medio del Archives Portal Europe, utilizando los perfiles del portal Europeo en ficheros XML, y a través de este portal con Europeana mediante la conversión de estos ficheros en formato EAD y EAC-CPF en EDM.

² Véase <<http://www.mecd.gob.es/cultura-mecd/areas-cultura/archivos/portal-de-archivos-espanoles-pares.html>> o el buscador en «pares.mcu.es». Para ver la evolución desde 2010, véase el vídeo: <http://www.youtube.com/watch?v=9hbFV7yk_8o>.

³ Un equipo de técnicos de PARES y el CIDA y de la SGTI de la Secretaría de Estado de Cultura.



Figura 1. Objetivos. (Elaboración propia).

Quizá uno de los retos más interesantes en estos desarrollos ha sido aplicar los modelos teóricos a la práctica archivística en las estructuras y herramientas de descripción de PARES, adaptando la primera y segunda parte dedicada a los agentes y las relaciones del **Modelo Conceptual de Descripción Archivística de la Comisión de Normas Españolas de Descripción Archivística** (CNEDA), publicado en 2012⁴. Mientras se redactaba este texto, se ha puesto en circulación para información pública la propuesta de la tercera parte de estas normas (atributos)⁵, que deberá también testearse en su momento en PARES.

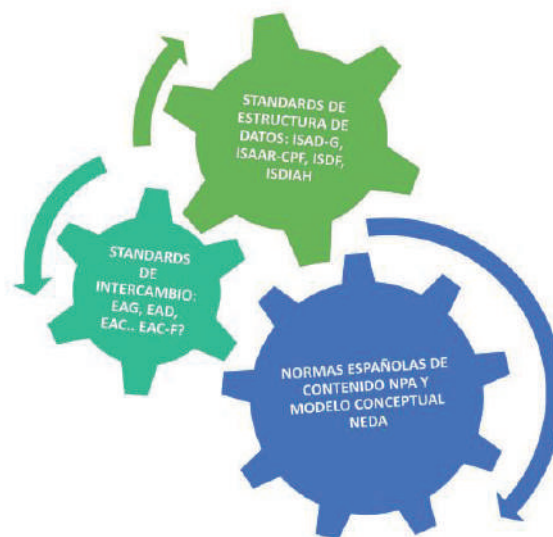


Figura 2. Normas de descripción archivística. (Elaboración propia).

⁴ Véase Bibliografía.

⁵ Disponible en <[http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MCDA-P3-15-01-2016-_borradorIP/MCDA-P3%20\(15-01-2016\)_borradorIP.pdf](http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MCDA-P3-15-01-2016-_borradorIP/MCDA-P3%20(15-01-2016)_borradorIP.pdf)>. [Consulta: 27 de enero de 2016].

En este proceso de adaptación del MCDA, la parte de las relaciones ha sido crítica, conscientes de que era un elemento esencial y necesario para poder interoperar en RDF. Sin embargo, se ha dado un paso más allá de los modelos de relaciones propuestos por la NEDA. Las normas EAC-CPF establecen cinco tipos de naturaleza de las relaciones (identidad, familia, asociativas, jerárquicas y temporales) y deja abierta la posibilidad de establecer «localmente» relaciones específicas. Durante el proceso de desarrollo del nuevo módulo de autoridades, consideramos que necesitábamos identificar dentro de cada tipo de relaciones las denominaciones de estas relaciones específicas de acuerdo con la dirección de las mismas (directa o inversa) cuando estas son asimétricas. Téngase en cuenta que únicamente los tipos de relación de identidad y asociativas son simétricas, es decir, una relación entre A y B se denomina igual que entre B y A –por ejemplo, A es socio de B, y B es socio de A–. Sin embargo, las relaciones jerárquicas, familiares o temporales implican mayor precisión: por ejemplo, A es padre de B, pero B es hijo, o hija, natural o legítima de B–, es decir, cuatro tipos diferentes de relaciones específicas de B con respecto a A. Así, durante este proceso se documentaron todas las relaciones existentes entre las autoridades en PARES, se cotejaron los ejemplos de relaciones de las NEDA, de las ISAAR-CPF, EAD-CPF y también los modelos funcionales FRAD y de otros modelos, como los apéndices J y K de las RDA⁶. El éxito de esta iniciativa nos permite en la actualidad presentar relaciones pertinentes y concretas, y generar materiales y grafos de gran calidad: genealogías ascendentes o descendentes de una persona (relaciones familiares), pero también organigramas para las relaciones jerárquicas entre instituciones, o esquemas de evolución de organismos para las relaciones temporales.

TIPO PRODUCTO	TIPO PRO	NATURALEZA RELACION	DESCRIPCIÓN ESPECÍFICA	DESCRIPCIÓN ESPECÍFICA-INVERSA
Persona	Persona	Identidad	Identidad alternativa	Identidad real
Persona	Persona	Identidad	Identidad real	Identidad alternativa
Persona	Persona	Familia (progenitor)	Es padre de	Es hijo/a de
Persona	Persona	Familia (progenitor)	Es madre de	Es hijo/a de
Persona	Persona	Familia (descendiente)	Es hijo/a de	Es padre de - Es madre de
Persona	Persona	Familia (colateral)	Es hermano/a de	Es hermano/a de
Persona	Persona	Familia (colateral)	Es sobrino/a de	Es tío/a de
Persona	Persona	Familia (colateral)	Es tío/a de	Es sobrino/a de
Persona	Persona	Familia (progenitor)	Es abuelo/a de	Es nieto/a de
Persona	Persona	Familia (progenitor)	Es bisabuelo/a de	Es bisnieto/a de
Persona	Persona	Familia (descendiente)	Es nieto/a de	Es abuelo/a de
Persona	Persona	Familia (descendiente)	Es bisnieto/a de	Es bisabuelo/a de
Persona	Persona	Familia (colateral)	Es yerno/nuera de	Es suegro/a de
Persona	Persona	Familia (colateral)	Es suegro/a de	Es yerno/nuera de
Persona	Persona	Familia (colateral)	Es primo/a de	Es primo/a de
Persona	Persona	Familia (colateral)	Es hermanastro/a de	Es hermanastro/a de
Persona	Persona	Familia (colateral)	Es cuñado/a de	Es cuñado/a de
Persona	Persona	Familia (matrimonio)	Esta casado/a con	Está casado/a con
Persona	Persona	Asociativa	Ejerce de procurador de	Su procurando de
Persona	Persona	Asociativa	Ejerce de tutor de	Es tutorando de
Persona	Persona	Asociativa	Es representante de	Está representado/a por
Persona	Persona	Asociativa	Es novio/a de	Es novio/a de
Persona	Persona	Asociativa	Es pareja de	Es pareja de
Persona	Familia	Familia (pertenencia)	Es miembro de	Tiene como miembro a
Familia	Persona	Familia (pertenencia)	Tiene como miembro a	Es miembro de
Persona	Familia	Familia (fundador)	Es fundador de la familia	Es fundada por
Familia	Persona	Familia (fundador)	Es fundada por	Es fundador de la familia
Persona	Familia	Asociativa	Es esponsor de la familia	La familia esta esponsorizada por
Familia	Persona	Asociativa	La familia esta esponsorizada por	Es esponsor de la familia
Persona	Institución	Asociativa	Es empleado de/ trabaja en	Tiene como empleado a
Institución	Persona	Asociativa	Tiene como empleado a	Es empleado de/ trabaja en
Persona	Institución	Asociativa	Es fundador de	Está fundada por
Institución	Persona	Asociativa	Esta fundada por	Es fundador de

Figura 3. Detalle de la lista controlada de relaciones específicas con las relaciones inversas de PARES. (Elaboración propia).

Otra de las claves en todo este complejo proceso de desarrollo del módulo de descripción de autoridades y puntos de acceso ha sido, sin duda, el encaje y la adaptación de los requerimientos de la nueva norma internacional de intercambio (EAC-CPF) con la norma internacional de estructura (ISAAR-CPF) y con nuestra norma propia de contenido para el caso de las

⁶ Véase Resource Description and Access: <<http://www.rda-jsc.org/archivedsite/docs/6JSC-CILIP-rep-3-Appendix-1.pdf>>.

autoridades de personas, familias, instituciones, materias y lugares. La Subdirección General de Archivos Estatales publicó en 2010 las ***Normas para la elaboración de puntos de acceso normalizados de instituciones, personas, familias, lugares y materias en el sistema de descripción archivística de los Archivos Estatales*** (NPA). Las *Normas* tienen como objetivos proporcionar reglas y/o recomendaciones para la elaboración de puntos de acceso normalizados de instituciones, personas, familias, lugares y materias, designados como *Forma(s) autorizada(s) del nombre*, en el caso de instituciones, familias y personas, o como *Término(s) preferente(s)*, en el caso de lugares y materias. Asimismo, pretenden facilitar determinadas reglas o recomendaciones para la cumplimentación de otros elementos o componentes estructurales de los registros de autoridad (de instituciones, familias y personas) y tesauros (lugares y de materias) que puedan desarrollarse con posterioridad. Estas *Normas* se han creado para su uso en el sistema de descripción archivística automatizado de los Archivos de titularidad estatal gestionados por el Ministerio de Educación, Cultura y Deporte: Portal de Archivos Españoles (PARES). Sin embargo, los propios desarrollos del nuevo módulo han supuesto tener que retocar y afinar aspectos de la NPA que resultaban inconsistentes. Estas normas están en proceso de revisión, ya que si en 2010 tenían como objetivo la formalización de puntos de acceso, ahora el objetivo es la elaboración de registros de autoridad como conceptos. Se va a incluir, además, una nueva parte de las normas dedicada a la formalización de nombres de normas y de actividades y funciones que no estaban contempladas en la primera edición, otro capítulo a los nombres de las relaciones, así como adaptar la parte de lugares y materias a la nueva ISO25964 de tesauros. Esperamos que la nueva edición de las NPA, cuya redacción está ya muy avanzada, vea la luz lo antes posible.

Los desarrollos del nuevo módulo han ido paralelos a los trabajos de colaboración internacional que hemos venido llevando a cabo en el Portal de Archivos Europeos (en adelante, APE), red de archivos que comenzó su andadura en 2009 bajo el nombre de APENet a iniciativa española y con el patrocinio de la Comunidad Europea⁷. Durante la fase del proyecto APEx Archives Portal Europe Network of Excellence, entre 2012 y 2015, hemos formado parte del grupo de trabajo WP4: Standard and Guidelines, que ha supuesto trabajar en la elaboración de los perfiles del portal de las normas internacionales de intercambio EAG, EAC-CPF y EAD, así como en los documentos *schemas*, mapeos, DTD y *deliverables* necesarios para APE; en especial, en el desarrollo de las apeEAC-CPF⁸ y las autoridades en el Portal y en la adecuación de los perfiles de EAD a las EAD3 y su interoperabilidad con EDM. En paralelo, hemos trabajado en los desarrollos para la elaboración en PARES de la herramienta para la generación automática de los ficheros XML en los perfiles de APEx: apeEAD, apeEAC-CPF y apeEAG2012, mientras adaptábamos los contenidos y se revisaban los datos y campos obligatorios de diversos fondos y colecciones de los archivos estatales, y se elaboraban ficheros de autoridades de personas, instituciones y familias. Se han realizado continuas pruebas de subida y visionado en las herramientas de APE para la preparación e intercambio de datos, el *dashboard* y el DPT. Como un gran engranaje, se han testado los errores, resolviendo inconsistencias, que en ocasiones eran problemas de los propios datos, en otras, errores en la construcción de los XML, en otras, del propio portal europeo... aprovechando todo este gran campo de pruebas para avanzar en el desarrollo del módulo y en la confección y construcción de nuestros registros de autoridad y vocabularios.

APE⁹. Es el agregador archivístico europeo de Comisión Europea y Europeana. Se han desarrollado diferentes estrategias, documentos y herramientas para facilitar el envío de datos a Europeana desde un entorno normalizado y resolver los problemas de interoperabilidad entre las normas archivísticas en contexto en un entorno común con bibliotecas, museos y otras institu-

⁷ Véase sobre Apenet: <<http://www.apenet.eu/>>, sobre APEx: <<http://www.apex-project.eu/index.php/en/>> y como fundación: <http://www.archivesportaleuropefoundation.eu/>.

⁸ Manuales y materiales del WP4 disponibles en: <<http://www.apex-project.eu/index.php/en/outcomes/standards>> y en http://wiki.archivesportaleurope.net/index.php/Main_Page>.

⁹ Para más información véase <<http://www.archivesportaleurope.net/>>.

ciones culturales. Si nuestros datos son interoperables con APE, en sus cuatro variables, nuestros datos serán interoperables con Europeana: apeEAD con EDM, y también apeEAC con EDM¹⁰.



Figura 4. Estado actual de los trabajos de desarrollo del portal PARES. (Elaboración propia).

En resumen, el nuevo módulo de autoridades de PARES contempla, tal y como hemos comentado, el control de autoridades de los tipos de entidad contemplados en la NEDA:

- Entidades del tipo agente (que incluye los subtipos institución, persona y familia), tanto productores de fondos o no, cuya estructura de datos está basada en ISAAR-CPF y EAC-CPF.
- Entidades de lugar y materias, con la estructura de datos ISO 25964.
- Entidades del tipo función y sus divisiones, y las entidades del tipo norma, cuya estructura de datos se basa en ISDF y en EAC-CPF (se utilizan etiquetas de campos equivalentes, ya que aún no está implementada la norma EAC-F).

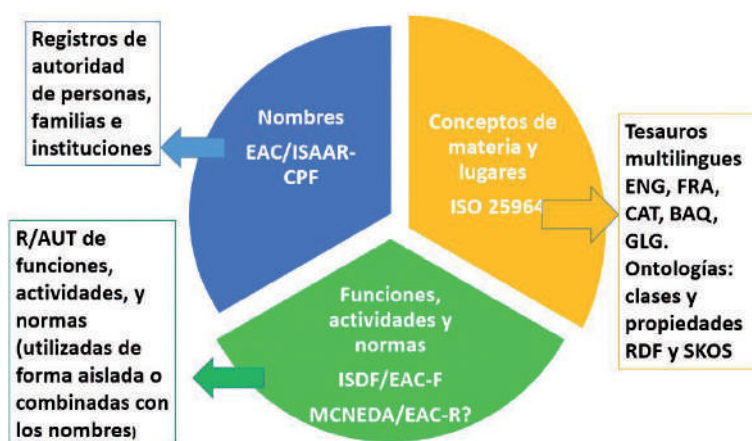


Figura 5. Tipos de registros de autoridad. (Elaboración propia).

¹⁰ El modelo de datos de Europeana (EDM) es una versión de 2013 del modelo de datos ESE (Europeana Semantic Elements, <<http://pro.europeana.eu/share-your-data/data-guidelines/ese-documentation>>) en desarrollo. Se reutilizan otras herramientas interoperables: RDF (Resource Description Framework), ORE (OAI Object Reuse and Exchange), SKOS (The Simple Knowledge Organization System) y DC (Dublin core namespaces for properties).

No nos es posible abordar por completo cada uno de estos tipos de registros de autoridad y cómo se están llevando a cabo los procesos de normalización de los diversos tipos (véase Lafuente y Muriel, 2014¹¹). Intentaremos sintetizar las novedades más interesantes y centrarnos en los tesauros:

Sobre los registros de autoridad de personas, familias e instituciones, ya hemos comentado que lo más novedoso de los desarrollos, además de la implementación del modelo de datos con las EAC-CPF, ha sido el desarrollo de las relaciones: entre registros de autoridades de personas, familias e instituciones, entre estas y sus funciones, con las materias (como actividades o eventos bibliográficos), con los lugares y con otros recursos, y lo más importante, con las unidades de descripción asociadas, diferenciando las relaciones como productor de las que son materia de un documento. Cuando las relaciones específicas de familia están bien informadas, el sistema genera automáticamente las genealogías ascendentes y descendentes de la persona.

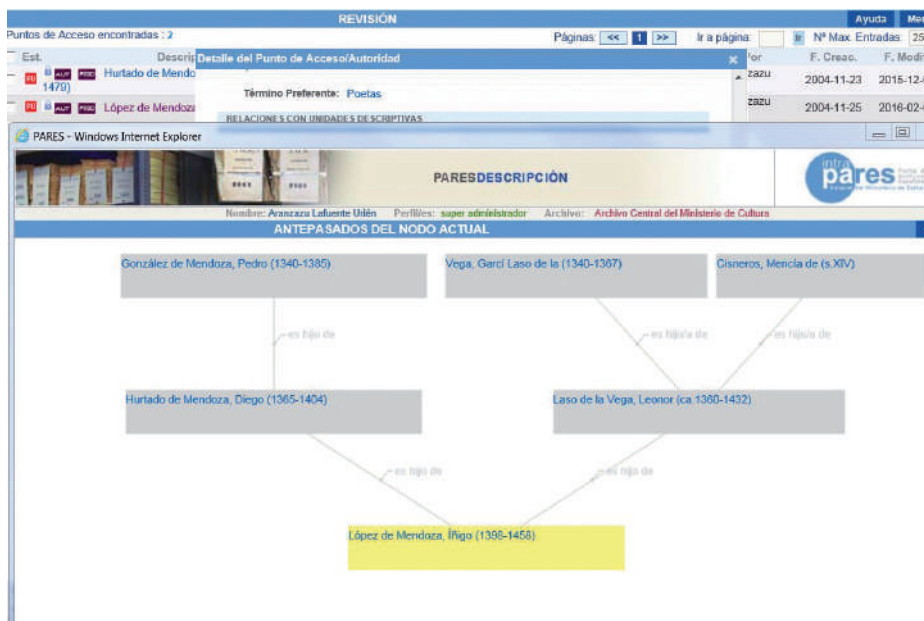


Figura 6. Detalle del árbol genealógico ascendente del I Marqués de Santillana. (Captura de INTRAPARES, 3 de febrero de 2016).

Otro caso de relaciones, las establecidas entre eventos (tipos de materias) utilizados para cumplimentar con listas estructuradas el campo historia o <BioHist> según las especificaciones de la etiqueta <chronList>, que también puede usarse de forma inversa, es decir, desde un evento ver las personas relacionadas. Figura 7.

¹¹ Remitimos al trabajo publicado con ocasión del Micrositio de la Fragata Mercedes, donde se comentan estos trabajos previos de manera más pormenorizada.

Nombre: Aranzazu Lafuente Urién Perfiles: super administrador Archivo: Archivo Central del Ministerio de Cultura

REVISIÓN

Puntos de Acceso encontradas : 1 Páginas: << 1 >> Ir a página: Nº Max. Er

Est.	Descripción	Tipo Autoridad	Creado Por	Modificado Por	F. Creac.
<input type="checkbox"/>	Batalla de Lepanto, 1571	Materias	Importacion ,SNAHN (SNAHN)	Lafuente Urién,Aranzazu (SGAE)	2015-09-25

Detalle del Punto de Acceso/Autoridad

<<<Volver

Punto de acceso seleccionado

Categoría Materias
Término Preferente: Batalla de Lepanto, 1571

Listado de elementos relacionados

Relación 1
Categoría Persona
Forma Autorizada del Austria, Juan de (ca.1545-1578)
Nombre:

Relación 2
Categoría Persona
Forma Autorizada del Bazán, Álvaro de (1520-1588)
Nombre:

Relación 3
Categoría Persona
Forma Autorizada del Cervantes, Miguel de (1547-1610)
Nombre:

Relación 4
Categoría Persona
Forma Autorizada del Doria, Juan Andrea (1540-1606)
Nombre:

Cancelar

Figura 7. Detalle del registro de autoridad del evento (materia)
«Batalla de Lepanto, 1571» y sus relaciones con autoridades de persona. (Captura de INTRAPARES, 3 de febrero de 2016).

Materias y lugares. A partir de haberse cerrado los desarrollos de la herramienta, hemos empezado a construir los tesauros de materias y de lugares a partir de los puntos de acceso que ya existían en la base de datos PARES –que se están normalizando, dividiendo o unificando para adecuarlos a las NPA–, y de todos aquellos nuevos que se van proponiendo por los técnicos que llevan a cabo los tratamientos archivísticos relacionados con la descripción de fondos en los archivos y, al igual que en todo el módulo, asociados a las unidades de descripción correspondientes.

El modelo aplica los principios de las ISO 25964, donde cada concepto puede englobarse en varias familias del tesauro, como un tesauro multifaceta. Otra de las líneas de trabajo es tratar de hacer compatibles al máximo las familias del tesauro (los *Top Term*) así como los términos más generales con los tesauros más conocidos y en uso, como el de la UNESCO o EUROVOC, o los vocabularios y *open data* de las grandes bibliotecas. Sin embargo, las temáticas que contemplan los fondos de nuestros archivos son tan complejos y extensos que hemos tenido que ampliar o desdoblar las familias clásicas de tesauros, por un lado, y ampliando los términos específicos mucho más allá, integrando tanto términos de uso actual con otros ya inexistentes pero de gran uso y pertinencia en los documentos de los fondos antiguos.

A la vez que vamos normalizando los vocabularios, se establecen las formas preferentes de acuerdo a las NPA, se indican sus equivalentes lingüísticos enlazados en su caso con sus equivalentes según otras reglas con los repositorios *open data* (BNE, LEM y sus equivalentes en gallego y catalán, y por supuesto con la Library of Congress y la BNF francesa) o recursos electrónicos con enlaces permanentes (DRAE, CSIC, etc.) en los cinco idiomas que se han comentado, se enriquece en la medida de lo posible con los recursos electrónicos y fuentes pertinentes. De momento, el tesauro multilingüe no es simétrico, aunque el objetivo es que lo llegue a ser.

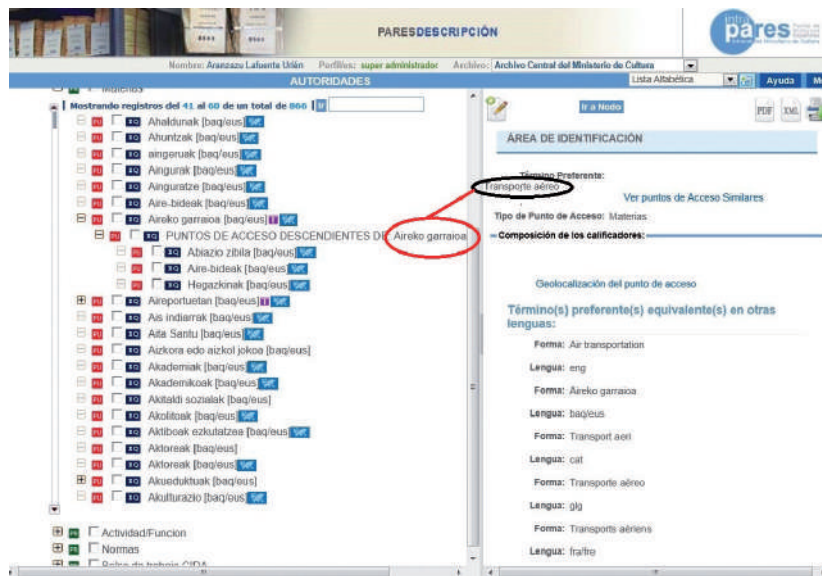


Figura 8. Detalle del tesoro multilingüe de materias en euskera: «Transporte aéreo» (Captura de INTRAPARES, 3 de febrero de 2016).

Como ocurre en todo tesoro multilingüe, la complejidad reside en los problemas semánticos, es decir, en determinar si los conceptos en dos lenguas son exactos o aproximados, y en este caso, cómo o de qué manera resolver la traducción de términos a otras lenguas, algunos, intraducibles, como los antiguos impuestos u oficios medievales, por ejemplo, «alcabala del viento», «almojarifazgo» o «almotacén».

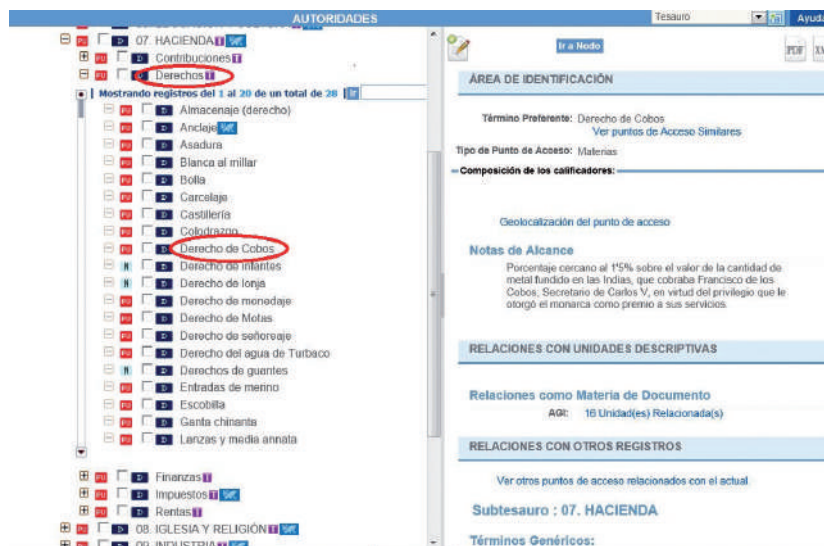


Figura 9. Términos específicos no traducibles (Captura de INTRAPARES, 4 de febrero de 2016).

En los tesauros de lugares se están formalizando los lugares de acuerdo con la toponimia oficial actual (importante en comunidades lingüísticas diferentes al castellano), georreferenciando, recogiendo también cualquier otro nombre alternativo o anterior, y tesaurizando.

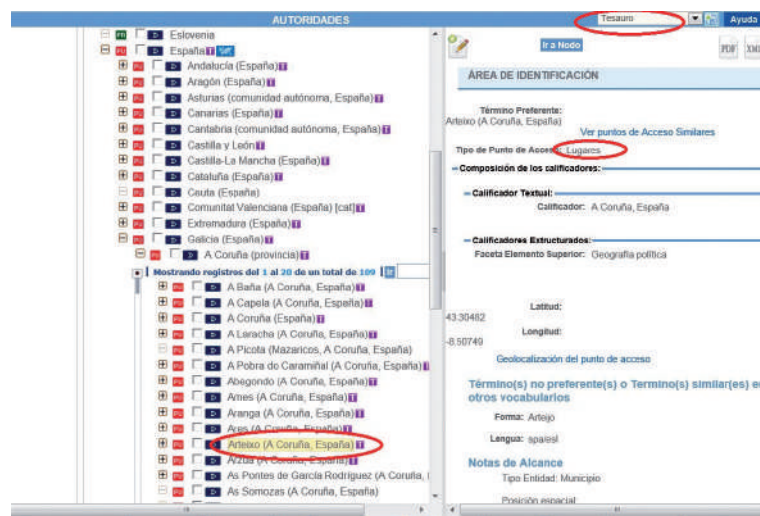


Figura 10. Tesaurus de lugares de geografía política. Arteixo (A Coruña). (Captura de INTRAPARES, 4 de febrero de 2016).

Este tesaurus consta –de momento– de varias facetas: geografía política, geografía natural, la artificial y, la más compleja pero necesaria tratándose de archivos históricos, las jurisdicciones antiguas: judicial, eclesiástica, militar y las entidades ya desaparecidas.

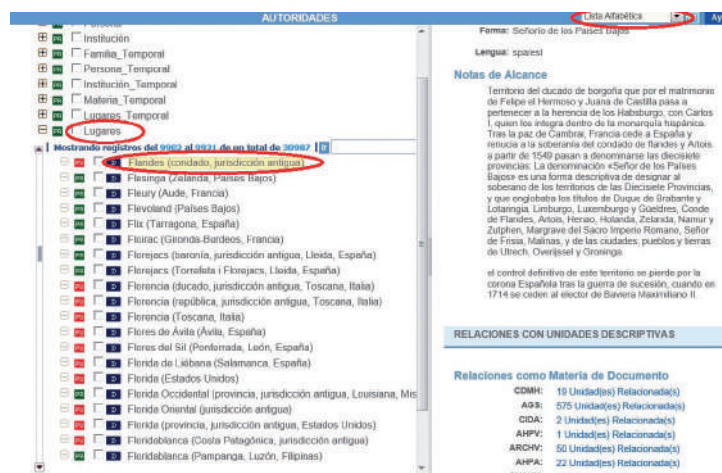


Figura 11. Lugares (lista alfabética): Jurisdicción antigua: Flandes. (Captura de INTRAPARES, 4 de febrero de 2016).

Funciones y actividades. Aquí nos hemos topado con una barrera importante, necesitamos avanzar en los desarrollos pero aún no se han desarrollado las normas que los establecen. Estas relaciones son básicas para documentar el contexto entre agentes. Los últimos meses hemos estado tratando de ver cómo poder implementar las funciones y también las normas en el módulo al mismo nivel que las personas, familias, instituciones, materias y lugares, si bien aún las EAC-F no existen (norma de intercambio de las ISDF). En este caso, hemos aplicado la misma lógica que se emplea en la EAC-CPF para los productores en aquellos campos similares (formas autorizadas, formas paralelas, otras formas, fechas de existencia, relaciones, etc.). Necesitamos establecer relaciones entre nombres/productores y sus funciones/actividades, para informar y documentar el contexto y, a su vez, entre funciones/actividades y normas (*mandate* en EAC-CPF).

The screenshot displays the PARES web interface. On the left, a navigation tree under 'AUTORIDADES' has 'Actividad/Funcion' selected. The main area shows the 'ÁREA DE IDENTIFICACION' with 'Término Preferente: Cabeza del linaje' and 'Tipo de Punto de Acceso: Actividad/Funcion'. Below this is the 'ÁREA DE DESCRIPCION' with a description: 'Función específica de las familias nobiliarias.' At the bottom, an XML view shows the following structure:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<doc xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="urn:isbn:1-931666-33-666">
  <Control>
    <recordId>ES-28804-CIDA100252140</recordId>
    <maintenanceStatus>revised</maintenanceStatus>
    <maintenanceAgency>
      <agencyCode>ES-28804-CIDA</agencyCode>
      <agencyName>CIDA</agencyName>
    </maintenanceAgency>
    <languageDeclaration>
      <language languageCode="spa"/>
      <script scriptCode="Latn"/>
    </languageDeclaration>
    <conventionDeclaration>
      <abbreviation>ISDF</abbreviation>
      <citation>ISDF - Norma internacional para la descripción de funciones [Recurso electrónico]. 1ª ed. 2007. (Disponible en http://www.mcu.es/archivos/docs/CE/ISDF_ESP_definitivo Consultado el 31 de mayo de 2010).</citation>
    </conventionDeclaration>
  </Control>
</doc>
```

Figura 12. Funciones de familias nobles. Detalle del XML.

El tesoro de *Normas* incluye todo tipo de disposición normativa, tesaurizada por tipos de normas (leyes, decretos, pragmáticas, etc.) que, o bien pueden ser los propios documentos de archivo (tratados internacionales, pragmáticas), o bien fuente legislativa de una entidad o disposición por la que se rige una función o competencia:

The screenshot displays the PARES web interface. On the left, a navigation tree under 'Normas' has 'Disposiciones normativas' selected. The main area shows 'Mostrando Unidad Descriptiva 52 de un total de 62' and 'ÁREA DE IDENTIFICACION' with 'Archivo: Archivo Histórico Nacional' and 'Código Referencia: ES.28079 AHN/1.1.7.11.3.22.118/ESTADO,2880_Exp.25Bis'. Below this is a list of records, with 'Paz de Westfalia, 1648' highlighted. The interface also shows a list of records on the left, with 'Paz de Westfalia, 1648' highlighted.

Figura 13. Normas. Detalle tesoro: Tratados internacionales. Paz de Westfalia y relación de UD asociada.

Obviamente, cada nuevo paso que hemos dado ha supuesto en ocasiones tener que revisar los pasos anteriores, revisar contenidos, realizar pruebas, revisar las normas en un trabajo continuo. En la actualidad, los esfuerzos en los desarrollos de PARES permiten visualizar todos estos trabajos, centrándonos en un nuevo buscador que posibilite buscar en autoridades y con-

ceptos y en una nueva interfaz del portal y nueva página web de PARES. Un nuevo buscador más sencillo, amigable pero a su vez más potente, con nuevas pantallas de resultados que presente filtros por facetas y permita búsquedas refinadas. Un segundo paso, cuando los tesauros están más avanzados y completos, será abrir los datos tanto de los vocabularios como de los registros de autoridad en SKOS.

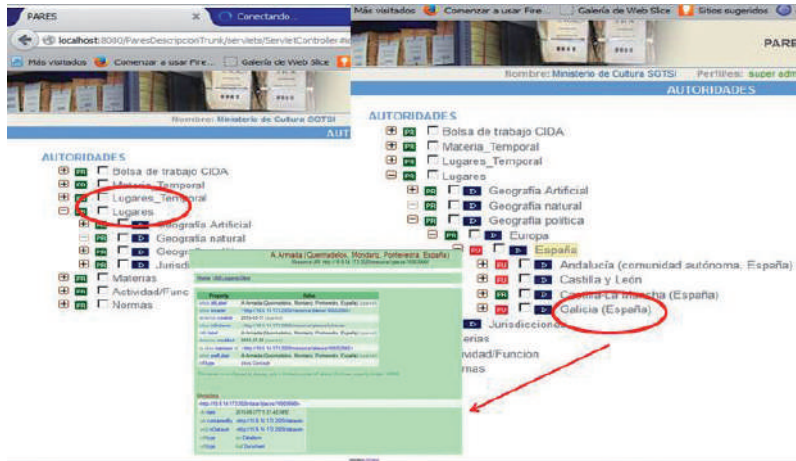


Figura 14. Conversión de lugares en SKOS.

Mientras tanto, seguimos trabajando en la mejora de nuestros datos y contenidos, implementando y enriqueciendo los vocabularios, para que las búsquedas resulten pertinentes y fiables, para que nuestros datos sean interoperables, transparentes y reutilizables; en definitiva, que nuestro patrimonio documental se revalorice, se aprecie y se disfrute. Muchas gracias.

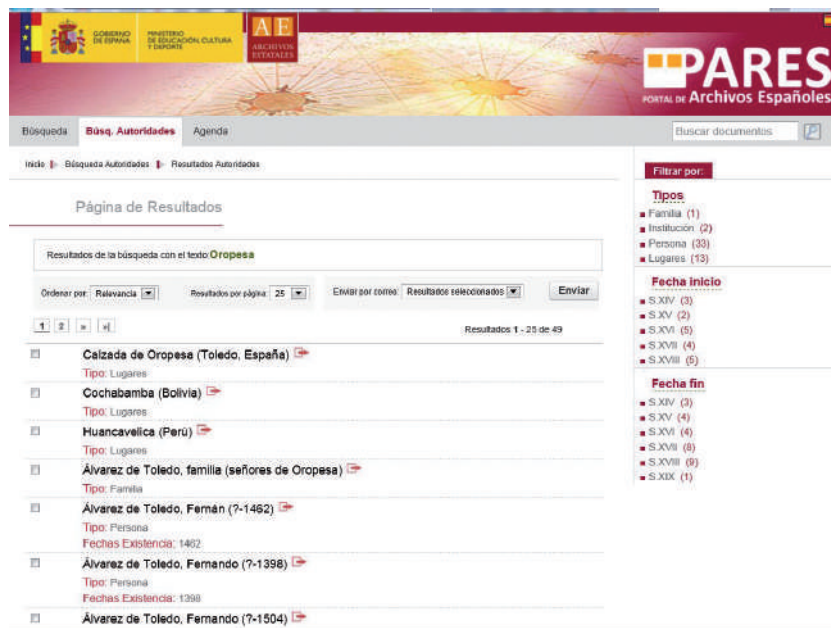


Figura 15. Nuevo buscador de PARES: búsqueda por autoridades (Oropesa). (Captura de INTRAPARES 4 de febrero de 2016).

Bibliografía

- HENNICKE, S. *et alii* (2011): *Conversion of EAD into EDM Linked Data*. First International Workshop on Semantic Digital Archives, TPDFL conference. Berlín. Disponible en: <<http://ceur-ws.org/Vol-801>>. [Consulta: el 31 de enero de 2016].
- LAFUENTE URIEN, A. (2014): *Archival authority control: an introduction text to Encoded Archival Context: corporate bodies, persons and families (EAC-CPF)*, julio 2014. Apex-project. (Recurso en línea). <<http://www.apex-project.eu/index.php/en/articles/184-archival-authority-control-an-introduction-to-encoded-archival-context-corporate-bodies-persons-and-families>>.
- LAFUENTE URIEN, A., y MURIEL HERNÁNDEZ, S. (2014): *Control de autoridades y puntos de acceso en el PORTAL DE ARCHIVOS ESPAÑOLES (PARES)*. Madrid: Ministerio de Educación Cultura y Deporte. (Recurso en línea): <<http://www.mecd.gob.es/fragatamercedes/dms/museos/fragatamercedes/historia/fuentes-historicas/Control-de-autoridades.pdf>>.
- SUGIMOTO, G. (2009): *Technical Report: Archival Digital Object Ingestion into Europeana (ESE-EAD harmonisation) v1.0*. Disponible en: <http://www.efgproject.eu/downloads/Technical_Report_Archival_Digital_Object_Ingestion_into_Europeana_%28ESE-EAD_harmonisation%29_v1.0.pdf>. [Consulta: 31 de enero de 2016].

Normas

- CNEDA (18-06-2012): *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones. Parte 1: Tipos de Entidad. Parte 2: Relaciones*. <http://www.mcu.es/archivos/docs/NEDA_MCDA_P1_P2_20120618.pdf> [Consulta: 31 de enero de 2016].
- CNEDA (16-01-2016): *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones. Parte 3: Atributos*. Documento en información pública. <[http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MCDA-P3--15-01-2016-_borradorIP/MCDA-P3%20\(15-01-2016\)_borradorIP.pdf](http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/MCDA-P3--15-01-2016-_borradorIP/MCDA-P3%20(15-01-2016)_borradorIP.pdf)>. [Consulta: 31 de enero de 2016].
- EAC-CPF: *Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF)*. Disponible en: <http://www.mcu.es/archivos/docs/EAC_CPF2010TL_SPA_2012.pdf2010>. [Consulta: 31 de enero de 2016].
- FRSAD (2010): *Functional requirements for subject authority data. A conceptual model*. IFLA, 2010. Disponible en: <<http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>>. [Consulta: 4 de febrero de 2016].
- FRSAD (2013): *Functional requirements for Authority data: a conceptual model*. IFLA, 2013. Disponible en: <<http://www.ifla.org/publications/functional-requirements-for-authority-data>>. [Consulta: 4 de febrero de 2016].
- ISAAR (CPF) (2004): *Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias* [Recurso electrónico], 2ª ed. Madrid: Ministerio de Cultura, Secretaría General Técnica. Disponible en: <<http://www.mcu.es/archivos/docs/isaar.pdf>>. [Consulta: 31 de enero de 2016].
- ISFD (2007): *Norma internacional para la descripción de funciones* [Recurso electrónico], 1ª ed. 2. Disponible en: <http://www.mcu.es/archivos/docs/CE/ISDF_ESP_definitiva.pdf>. [Consulta: 31 de mayo de 2010].
- NPA (2010): *Normas para la elaboración de puntos de acceso normalizados de instituciones, personas, familias, lugares y materias en el sistema de descripción archivística de los Archivos Estatales*.

Madrid: Subdirección General de los Archivos Estatales. Disponible en: <http://www.mcu.es/archivos/docs/Novedades/Norma_puntos_acceso2010.pdf>. [Consulta: 4 de febrero de 2016].

UNE-ISO 25964-1 (2014): *Información y documentación: Tesauros e interoperabilidad con otros vocabularios*. Parte 1: «Tesauros para la recuperación de la información».



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, CULTURA
Y DEPORTE